

UNIVERSIDADE FEDERAL DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM SANEAMENTO,
MEIO AMBIENTE E RECURSOS HÍDRICOS

PROPOSIÇÃO DE MODELO DE PREDIÇÃO DE
DESEMPENHO DE ESTAÇÕES
CONVENCIONAIS DE TRATAMENTO DE ÁGUA
UTILIZANDO REDES NEURAS ARTIFICIAIS

Juscelino Alves Henriques

Belo Horizonte
2019

**PROPOSIÇÃO DE MODELO DE PREDIÇÃO DE
DESEMPENHO DE ESTAÇÕES
CONVENCIONAIS DE TRATAMENTO DE ÁGUA
UTILIZANDO REDES NEURAIS ARTIFICIAIS**

Juscelino Alves Henriques

Juscelino Alves Henriques

**PROPOSIÇÃO DE MODELO DE PREDIÇÃO DE
DESEMPENHO DE ESTAÇÕES
CONVENCIONAIS DE TRATAMENTO DE ÁGUA
UTILIZANDO REDES NEURAIS ARTIFICIAIS**

Tese apresentada ao Programa de Pós-graduação em Saneamento, Meio Ambiente e Recursos Hídricos da Universidade Federal de Minas Gerais, como requisito parcial para a obtenção do título de Doutor em Saneamento, Meio Ambiente e Recursos Hídricos.

Área de concentração: Saneamento

Linha de pesquisa: Qualidade e Tratamento de Água para Consumo Humano

Orientador: Marcelo Libânio

Coorientador: Veber Afonso Figueiredo Costa

Belo Horizonte

Escola de Engenharia da UFMG

2019

H519p Henriques, Juscelino Alves.
Proposição de modelo de predição de desempenho de estações convencionais de tratamento de água utilizando redes neurais artificiais [recurso eletrônico]/ Juscelino Alves Henriques.- 2019.
1 recurso online (xv, 131 f. : il., color.): pdf.

Orientador: Marcelo Libânio.
Coorientador: Veber Afonso Figueiredo Costa.

Tese (doutorado) – Universidade Federal de Minas Gerais, Escola de Engenharia.

Bibliografia: f. 123-131.
Exigências do sistema: Adobe Acrobat Reader.

1. Engenharia sanitária- Teses. 2. Saneamento - Teses. 3. Predição (Lógica) - Teses. 4. Inteligência artificial - Teses. 5. Redes neurais (Computação) - Teses. 6. Água - Estações de tratamento - Teses.
I. Libânio, Marcelo, 1960-. II. Costa, Veber Afonso Figueiredo. III. Universidade Federal de Minas Gerais. Escola de Engenharia. IV. Título.

CDU: 628(043)

Ficha catalográfica: Biblioteca Profº Mário Werneck, Escola de Engenharia da UFMG



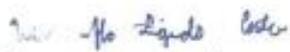
FOLHA DE APROVAÇÃO

Proposição de Modelo de Predição de Desempenho de Estações Convencionais de Tratamento de Água Utilizando Redes Neurais Artificiais

JUSCELINO ALVES HENRIQUES

Tese defendida e aprovada pela banca examinadora constituída pelos Senhores:

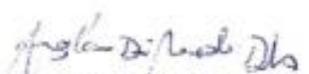

Prof. MARCELO LIBÂNIO


Prof. VEBER AFONSO FIGUEIREDO COSTA

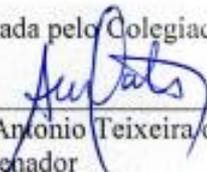

Prof. LUIZ RAFAEL PALMIER


Prof. CESAR ROSSAS MOTA FILHO


Prof. FREDERICO KEIZO ODAN


Prof.ª ANGELA DI BERNARDO DANTAS

Aprovada pelo Colegiado do PG SMARH


Prof. Antonio Teixeira de Matos
Coordenador

Versão Final aprovada por


Prof. Marcelo Libânio
Orientador

Belo Horizonte, 15 de abril de 2019.

AGRADECIMENTOS

A Deus, Pai de valor inestimável, que me sustenta e proporciona o melhor para a minha vida.

Aos meus pais Maria Lúcia e José, por todo esforço, carinho, apoio e dedicação.

À Rosália e à nossa filha Helena, por toda compreensão, dedicação e paciência nos dias difíceis, e pelo companheirismo nas duradouras madrugadas de estudos.

Às minhas irmãs, Elizabete (Betinha), Elidiane (Lí) e Jucimery (Jú), além dos meus sobrinhos e cunhados, por todo apoio e afeto.

À família Varotto, que me “adotou” e me acolheu quando cheguei em Minas Gerais; dividimos juntos os momentos de ganhos e perdas.

Ao Instituto Federal do Espírito Santo – IFES Campus Ibatiba, por todo apoio na realização deste doutorado; aos colegas professores e, especialmente, aos meus alunos, por toda compreensão e por me suportar nos momentos mais estressantes.

Aos irmãos da Paróquia Nossa Senhora do Rosário de Ibatiba-ES, pela acolhida e fraternidade.

A todos os colegas contemporâneos do PPG-SMARH, por todos os momentos vividos.

Aos professores Marcelo Libânio e Veber Afonso Figueiredo Costa, por toda amizade, confiança, conselhos e orientações.

Aos professores Angela Di Bernardo Dantas, Luiz Rafael Palmier, Cesar Rossas Mota Filho e Frederico Keizo Odan pela participação, na qualidade de avaliadores.

A todos os professores do SMARH, pelos ensinamentos e aconselhamentos indispensáveis.

Aos servidores do SMARH, DESA e EHR, pela atenção e disponibilidade de sempre.

À FAPEMIG, pela bolsa de estudos durante o ano de 2015.

RESUMO

As estações de tratamento de água (ETA) possuem papel fundamental e estratégico no controle de doenças de veiculação hídrica por meio da potabilização da água, com vistas ao atendimento dos anseios da população que por elas é abastecida. Nesse contexto, avaliar o desempenho dessas estações é primordial, particularmente para os entes responsáveis pela etapa de controle da qualidade da água, tendo em vista que a ETA deve apresentar e operar com condições mínimas necessárias que atendam aos padrões de potabilidade. Como forma de antever situações adversas, bem como na busca de auxiliar os tomadores de decisão e gestores responsáveis por tais operações, foram propostos diversos modelos de predição de desempenho de ETAs, fazendo uso de uma série de ferramentas, notadamente da modelagem computacional, importante área da Inteligência Artificial (IA). Nesse contexto, essa pesquisa teve por objetivo propor modelos de predição de desempenho de estações convencionais de tratamento de água utilizando Redes Neurais Artificiais (RNA), tendo como parâmetros de controle a turbidez e a cor aparente da água tratada. Para o seu desenvolvimento foram elaborados cinco modelos em RNA, dois para ETAs de pequeno porte, dois para ETAs de médio porte e um para ETAs de grande porte. Os dados utilizados referem-se à qualidade da água bruta e tratada, aspectos operacionais e parâmetros hidráulicos de 12 estações de tratamento de água, quatro com vazões de até 50 L.s^{-1} , três com vazões entre 50 L.s^{-1} a 500 L.s^{-1} e cinco com vazões acima de 500 L.s^{-1} . Os modelos foram desenvolvidos utilizando o *RStudio*®, a partir da rede neural artificial *multilayer backpropagation*, com função de ativação *logistic*. Os modelos apresentaram coeficientes de determinação (R^2) de 0,66 e 0,67 para a turbidez e a cor aparente da água tratada de estações de pequeno porte, respectivamente. Já para as ETAs de médio (Modelos 3 e 4) e grande (Modelos 5) porte os valores de R^2 foram 0,21, 0,33 e 0,60, respectivamente. Na comparação dos resultados, também foram desenvolvidos modelos em regressão linear múltipla, sendo constatado que os modelos em RNA apresentaram melhor desempenho. Os resultados corroboram para a aplicação de técnicas de IA, particularmente RNA, em ETAs, com vista à otimização de processo, menor consumo de produtos químicos e geração de resíduos, bem como para possibilitar operação mais eficiente e segura.

Palavras-chave: Predição de desempenho de ETA, Modelagem computacional, Redes Neurais Artificiais.

ABSTRACT

Water treatment plants (WTP) have a fundamental and strategic role in the control of waterborne diseases through the potabilization of water, in order to meet the needs of the population that is supplied by it. In this context, evaluating the performance of these stations is paramount, particularly for the entities responsible for the water quality control stage, since WTP must present and operate with minimum conditions necessary to achieve its objective. As a way of anticipating adverse situations, as well as in the search to help decision makers and managers responsible for such operations, several models of WTPs performance prediction were proposed, making use of a series of tools, notably computational modeling, an important area of Artificial Intelligence (AI). As a contribution to this, this PhD thesis aimed to propose performance prediction models of conventional water treatment stations using Artificial Neural Network (ANN), with turbidity and apparent color of the treated water as control parameters. For its development (five ANN models were developed, two for small WTPs, two for medium-sized WTPs and one for large WTPs), data on raw and treated water quality were used. and hydraulic parameters of 12 water treatment plants, four with flows of up to $50 \text{ L}\cdot\text{s}^{-1}$, three with flows between $50 \text{ L}\cdot\text{s}^{-1}$ - $500 \text{ L}\cdot\text{s}^{-1}$ and five with flows above $500 \text{ L}\cdot\text{s}^{-1}$. The models were developed using *RStudio*®, from the artificial neural network multilayer backpropagation, with logistic activation function. The models presented determination coefficients of 0.66 and 0.67 for the turbidity and the apparent color of the treated water of small stations, respectively. The R^2 values were 0.21, 0.33 and 0.60, respectively, for the medium (Models 3 and 4) and large (Models 5). In the comparison of the results, models were also developed in multiple linear regression, being found that the models in ANN presented better performance. The results corroborate the application of AI techniques, particularly ANN, in WTPs, in order to optimize the process, reduce consumption of chemicals and generate waste, as well as to make the operation more efficient and safer by offering a product more reliable to the population supplied.

Keywords: Predict performance of WTP, Computational modeling, ANN.

SUMÁRIO

LISTA DE FIGURAS	VIII
LISTA DE QUADROS E TABELAS	X
LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS	XII
1. INTRODUÇÃO.....	16
1.1 CONSIDERAÇÕES INICIAIS	16
1.2 JUSTIFICATIVA E RELEVÂNCIA	18
1.3 HIPÓTESES E PREMISSAS	20
1.4 OBJETIVOS	22
1.4.1 <i>Objetivo geral</i>	22
1.4.2 <i>Objetivos específicos</i>	22
2. REVISÃO DA LITERATURA.....	23
2.1 ANÁLISE DE DESEMPENHO DE ESTAÇÕES DE TRATAMENTO DE ÁGUA.....	23
2.1.1 <i>Ferramentas para a análise de desempenho de estações de tratamento de água</i>	27
2.2 SISTEMAS ARTIFICIAIS INTELIGENTES	33
2.2.1 <i>Redes Neurais Artificiais</i>	35
2.3 MODELAGEM DE PREDIÇÃO DE DESEMPENHO DE ETA.....	41
3. MATERIAL E MÉTODOS	44
3.1 DEFINIÇÃO DO OBJETO DE ESTUDO	44
3.1.1 <i>Caracterização das estações de tratamento de água</i>	48
3.2 DESENVOLVIMENTO DOS MODELOS DE PREDIÇÃO.....	50
3.2.1 <i>Tabulação dos dados e testes estatísticos básicos</i>	50
3.2.2 <i>Modelos de regressão linear múltipla (baseline) e escolha das variáveis</i>	52
3.2.3 <i>Modelos baseados em redes neurais artificiais</i>	53
3.3 ANÁLISE DOS MODELOS PROPOSTOS	56
4. RESULTADOS E DISCUSSÃO.....	59
4.1 MODELOS PARA ESTAÇÕES DE TRATAMENTO DE PEQUENO PORTE	59
4.1.1 <i>Análise exploratória para os modelos de ETAs de pequeno porte</i>	59

4.1.2	<i>Escolha das variáveis para compor os modelos de ETAs de pequeno porte</i>	63
4.1.3	<i>Modelos em regressão linear múltipla para ETAs de pequeno porte</i>	66
4.1.4	<i>Arquitetura e topologia da rede neural aplicada aos modelos</i>	70
4.1.5	<i>Modelo em RNA de estações de tratamento de pequeno porte para a turbidez da água tratada</i>	71
4.1.6	<i>Modelo em RNA de estações de tratamento de pequeno porte para a cor aparente da água tratada</i>	76
4.2	MODELOS PARA ESTAÇÕES DE TRATAMENTO DE MÉDIO PORTE	81
4.2.1	<i>Análise exploratória para os modelos de ETAs de médio porte</i>	81
4.2.2	<i>Escolha das variáveis para compor os modelos de ETAs de médio porte</i>	85
4.2.3	<i>Modelos em regressão linear múltipla para estações de tratamento de médio porte</i>	87
4.2.4	<i>Modelo em RNA de estações de tratamento de médio porte para a turbidez da água tratada</i>	91
4.2.5	<i>Modelo em RNA de estações de tratamento de médio porte para a cor aparente da água tratada</i>	96
4.3	MODELOS PARA ESTAÇÕES DE TRATAMENTO DE GRANDE PORTE	100
4.3.1	<i>Análise exploratória para os modelos de ETAs de grande porte</i>	100
4.3.2	<i>Escolhas das variáveis para compor os modelos de grande porte</i>	104
4.3.3	<i>Modelo em regressão linear múltipla para ETAs de grande porte</i>	106
4.3.4	<i>Modelo em RNA para ETAs de grande porte</i>	108
4.4	COMPARAÇÃO ENTRE OS MODELOS PROPOSTOS E DISCUSSÃO DOS RESULTADOS	111
5.	CONCLUSÕES	119
6.	RECOMENDAÇÕES	121
	REFERÊNCIAS	123

LISTA DE FIGURAS

FIGURA 1: Metodologia para a definição dos indicadores de desempenho.....	24
FIGURA 2: Fases de aplicação da Análise do Ciclo de Vida	31
FIGURA 3: Dimensões da Inteligência Artificial	34
FIGURA 4: Atuação entre agente e ambiente por meio de sensores e atuadores em IA	35
FIGURA 5: Cronologia do desenvolvimento das RNAs e seus principais responsáveis	37
FIGURA 6: Representação de um neurônio natural (A), artificial (B) e as sinapses que ocorrem entre ambos (C e D).....	38
FIGURA 7: Estrutura de uma rede MLP com duas camadas ocultas.....	39
FIGURA 8: Etapas metodológicas da pesquisa	44
FIGURA 9: Tela inicial da ferramenta utilizada para o desenvolvimento dos modelos em RNA.....	54
FIGURA 10: Composição de gráficos <i>boxplot</i> de cada uma das variáveis que compõem os modelos para as ETAs de pequeno porte.....	63
FIGURA 11: Rede neural gerada a partir do modelo proposto para a turbidez da água tratada em ETAs de pequeno porte.....	72
FIGURA 12: Comparação entre o comportamento dos valores de turbidez da água tratada durante o treinamento do modelo das estações de pequeno porte	73
FIGURA 13: Comportamento da correlação entre a turbidez calculada e observada durante o treinamento da rede neural para estações de pequeno porte	74
FIGURA 14: Intervalo de predição gerado a partir dos valores preditos de turbidez da água tratada para ETAs de pequeno porte.....	75
FIGURA 15: Rede neural gerada a partir do modelo proposto para a cor aparente da água tratada em ETAs de pequeno porte.....	77
FIGURA 16: Comparação entre o comportamento dos valores de cor aparente da água tratada durante o treinamento do modelo das estações de pequeno porte	78
FIGURA 17: Comportamento da correlação entre a cor aparente calculada e observada durante o treinamento da rede neural para estações de pequeno porte	79
FIGURA 18: Intervalo de predição gerado a partir dos valores preditos de cor aparente da água tratada para ETAs de pequeno porte	80
FIGURA 19: Composição de gráfico <i>boxplot</i> de cada uma das variáveis que compõem os modelos para as ETAs de médio porte	85

FIGURA 20: Rede neural gerada a partir do modelo proposto para a turbidez da água tratada em ETAs de médio porte	92
FIGURA 21: Comparação entre o comportamento dos valores de turbidez da água tratada durante o treinamento do modelo das estações de médio porte	93
FIGURA 22: Comportamento da correlação entre a turbidez calculada e observada durante o treinamento da rede neural para estações de médio porte	94
FIGURA 23: Intervalo de predição gerado a partir dos valores preditos de turbidez da água tratada para ETAs de médio porte	95
FIGURA 24: Rede neural gerada a partir do modelo proposto para a cor aparente da água tratada em ETAs de médio porte	97
FIGURA 25: Comparação entre o comportamento dos valores de cor aparente da água tratada durante o treinamento do modelo das estações de médio porte	97
FIGURA 26: Comportamento da correlação entre a cor aparente calculada e observada durante o treinamento da rede neural para estações de médio porte	98
FIGURA 27: Intervalo de predição gerado a partir dos valores preditos de cor aparente da água tratada para ETAs de médio porte	99
FIGURA 28: Composição de gráficos <i>boxplot</i> de cada uma das variáveis que compõem os modelos para as ETAs de grande porte	103
FIGURA 29: Rede neural gerada a partir do modelo proposto para a turbidez da água tratada em ETAs de grande porte	108
FIGURA 30: Comparação entre o comportamento dos valores de turbidez da água tratada durante o treinamento do modelo das estações de grande porte	109
FIGURA 31: Comportamento da correlação entre a turbidez calculada e observada durante o treinamento da rede neural para estações de grande porte	109
FIGURA 32: Intervalo de predição gerado a partir dos valores preditos de turbidez da água tratada para ETAs de grande porte	110

LISTA DE QUADROS E TABELAS

QUADRO 1: Principais ferramentas utilizadas na avaliação de desempenho de ETA.....	29
QUADRO 2: Análise das hipóteses propostas para a pesquisa, de acordo com o desenvolvimento dos modelos	116
TABELA 1: Descrição das estações de tratamento de água utilizadas para elaboração dos modelos	49
TABELA 2: Apresentação dos resultados da estatística exploratória das variáveis referentes às estações de pequeno porte	60
TABELA 3: Matriz de correlação entre as variáveis utilizadas no desenvolvimento dos modelos de estações de pequeno porte	65
TABELA 4: Análise de variância gerada para o modelo de regressão em ETAs de pequeno porte a partir da turbidez da água tratada.....	67
TABELA 5: Sumário do modelo para turbidez da água tratada em ETAs de pequeno porte ..	67
TABELA 6: Coeficientes que compõem o modelo de estações de pequeno porte para turbidez da água tratada	68
TABELA 7: Análise de variância gerada para o modelo de regressão em ETAs de pequeno porte a partir da cor aparente da água tratada	69
TABELA 8: Sumário do modelo para cor aparente da água tratada em ETAs de pequeno porte	69
TABELA 9: Coeficientes que compõem o modelo de estações de pequeno porte para cor aparente da água tratada.....	70
TABELA 10: Resultados das métricas durante a elaboração do Modelo 1	76
TABELA 11: Resultados das métricas durante a elaboração do Modelo 2	81
TABELA 12: Apresentação dos resultados da estatística exploratória das variáveis referentes às estações de médio porte.....	82
TABELA 13: Matriz de correlação entre as variáveis utilizadas no desenvolvimento dos modelos para ETAs de médio porte.....	86
TABELA 14: Análise de variância gerada para o modelo de regressão em ETAs de médio porte a partir da turbidez da água tratada.....	88
TABELA 15: Sumário do modelo para turbidez da água tratada em ETAs de médio porte ...	88

TABELA 16: Coeficientes que compõem o modelo de estações de médio porte para turbidez da água tratada	89
TABELA 17: Análise de variância gerada para o modelo de regressão em ETAs de médio porte a partir da cor aparente da água tratada	90
TABELA 18: Sumário do modelo para cor aparente da água tratada em ETAs de médio porte	90
TABELA 19: Coeficientes que compõem o modelo de estações de médio porte para cor aparente da água tratada.....	91
TABELA 20: Resultados das métricas durante a elaboração do Modelo 3	96
TABELA 21: Resultados das métricas durante a elaboração do Modelo 4	100
TABELA 22: Apresentação dos resultados da estatística exploratória das variáveis referentes às estações de grande porte	101
TABELA 23: Matriz de correlação entre as variáveis utilizadas no desenvolvimento dos modelos para ETAs de grande porte.....	105
TABELA 24: Análise de variância gerada para o modelo de regressão em ETAs de grande porte a partir da turbidez da água tratada.....	106
TABELA 25: Sumário do modelo para turbidez da água tratada em ETAs de grande porte .	107
TABELA 26: Coeficientes que compõem o modelo de estações de grande porte para turbidez da água tratada	107
TABELA 27: Resultados das métricas durante a elaboração do Modelo 5	111
TABELA 28: Comparação entre os resultados dos modelos desenvolvidos para cada porte de estação, de acordo com os valores dos coeficientes de determinação (R^2) gerados em cada um dos modelos e nos <i>baselines</i>	112
TABELA 29: Resumo dos resultados das métricas para todos os modelos propostos	114

LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

ABNT	Associação Brasileira de Normas Técnicas
AHP	Análise Hierárquica de Processo
ANFIS	Sistema de inferência adaptativo Neuro-Fuzzy
ANOVA	Análise de variância
β	Parâmetros da equação de regressão
BIM	Modelo da Informação da Construção
bR^2	Coefficiente de determinação (R^2) multiplicado pela inclinação da linha de regressão entre os resultados simulados e observados
CCME	Conselho Canadense de Ministros do Meio Ambiente
CCP	Programa de correção composta
$cm.min^{-1}$	Unidade de velocidade – centímetro por minuto
$cm.s^{-1}$	Unidade de velocidade – centímetro por segundo
Coef	Coefficientes do modelo (relação entre as variáveis dependentes e independentes)
CorB	Cor aparente da água bruta
CorT	Cor aparente da água tratada
CPE	Avaliação de desempenho abrangente
CPI	Processo de melhoria contínua
DosCoag	Dosagem de coagulante
EmE	Avaliação em energia – avaliação de fluxo de energia
EnsJar	Realização de ensaio em <i>Jar Test</i>
EP de Coef	Erro padrão dos coefficients
ETA	Estação de tratamento de água
GL	Graus de liberdade
GradFloc	Gradiente de velocidade de floculação
h	Unidade de tempo – hora
IA	Inteligência artificial
IBGE	Instituto Brasileiro de Geografia e Estatística
ICMA	Associação Internacional de Gestão dos Conselhos Municipais
IC de 95%	Intervalo de confiança ao nível de significância de 95%

InstOp	Grau de instrução do operador da estação de tratamento de água
IQA	Índice de qualidade de água
IQAB	Índice de qualidade de água bruta
IQABF	Índice de qualidade de água bruta Fuzzy
IQETA	Índice de qualidade de estação de tratamento de água
IQETAF	Índice de qualidade de estação de tratamento de água Fuzzy
ISO	Organização Internacional de Padronização
kg.dia ⁻¹	Unidade de vazão mássica – quilograma por dia
L.s ⁻¹	Unidade de vazão – litro por segundo
L.s ⁻¹ .m ⁻¹	Unidade de vazão linear – litro por segundo por metro
LCA	Avaliação de ciclo de vida
LCIA	Avaliação de impacto do ciclo de vida
m ³ .m ⁻² .dia ⁻¹	Unidade de taxa de filtração – metro cúbico por metro quadrado por dia
mg.L ⁻¹	Unidade de concentração – miligrama por litro
mNSE	Eficiência de Nash-Sutcliffe modificada
MSE	Algoritmo para o cálculo do desempenho tipo Quadrado médio do erro
NBR	Norma Brasileira Regulamentadora
NRMSE	Raiz quadrada do erro médio normalizada
NSE	Eficiência de Nash-Sutcliffe
PAC	Cloreto de polialumínio
PBIAS	Porcentagem de polarização, mede a tendência média dos valores calculados para serem maiores ou menores do que os observados
pH	Potencial hidrogeniônico
pHB	Potencial hidrogeniônico da água bruta
PRESQ	Soma dos quadrados preditos do erro (calcula o desvio do modelo)
ProdLodo	Produção de lodo gerado no tratamento da água
PSA	Plano de Segurança da Água
PPG-SMARH	Programa de Pós-Graduação em Saneamento, Meio Ambiente e Recursos Hídricos
Q	Vazão
Qcalha	Vazão linear na calha de coleta de água decantada
Qaf	Vazão afluyente à estação de tratamento de água
QM (Aj.)	Quadrado médio ajustado

r	Coeficiente de correlação
R ²	Coeficiente de determinação
R ² (aj.)	Coeficiente de determinação ajustado
R ² (pred)	Coeficiente de determinação predito (determina o quanto o modelo prediz para novas observações)
rd	Índice de concordância relativo
RMSE	Raiz quadrada do erro médio, apresenta o desvio padrão do erro de predição do modelo
RNA	Redes neurais artificiais
S	Desvio padrão
s	Unidade de tempo – segundo
s ⁻¹	Unidade inversa ao tempo – gradiente de floculação
SSE	Algoritmo para o cálculo do desempenho tipo Quadrado da soma do erro
SQ (Aj.)	Soma dos quadrados ajustados
SQ Seq	Soma dos quadrados sequenciais
Tc	Tempo de contato
TemFloc	Tempo de floculação
TempOp	Tempo diário de operação da estação de tratamento de água
Termo	Elementos que irão compor o modelo
TuB	Turbidez da água bruta
TuD	Turbidez da água decantada
TuT	Turbidez da água tratada
TxFilt	Taxa de filtração
USEPA	Agência de Proteção Ambiental dos Estados Unidos
uT	Unidade de turbidez
uC	Unidade de cor aparente
Valor F	Estatística utilizada para verificar se o modelo está associado com a resposta
Valor-P	Mede a evidência contrária à hipótese nula do teste
Valor-T	Razão entre o coeficiente e o respectivo erro padrão
VE	Eficiência volumétrica
VelSed	Velocidade de sedimentação na decantação
VelLong	Velocidade de escoamento longitudinal de escoamento na decantação

VIF	Fator de inflação da variância
WHO	Organização Mundial da Saúde
x	Variável independente da equação de regressão
\hat{y}	Variável dependente da equação de regressão

1. INTRODUÇÃO

1.1 Considerações iniciais

Segundo a Organização das Nações Unidas (ONU, 2016), dos aproximadamente 7 bilhões de habitantes do globo terrestre, 2,5 bilhões não têm acesso ao saneamento básico, sendo que desses, 783 milhões não têm acesso à água potável. É importante ressaltar que, dentre os Objetivos Globais para o Desenvolvimento Sustentável, também estabelecidos pela ONU, encontra-se a universalização do acesso à água potável até 2030, meta de grande repercussão mundial, tendo em vista a escassez hídrica que alguns países enfrentam, bem como os diversos problemas relacionados com o acesso ao saneamento básico.

Conforme o Instituto Brasileiro de Geografia e Estatística - IBGE, a população brasileira é de aproximadamente 206 milhões de habitantes, sendo que 160 milhões vivem no meio urbano, distribuídos em 5.570 municípios (IBGE, 2019). Ainda de acordo com o IBGE (2014), 83,6% da população brasileira têm acesso à água potável por rede geral. Conforme ilustrado pelo IBGE (2008), no Brasil existem 6.040 estações de tratamento de água. As unidades convencionais respondiam por 69,2% de toda a água distribuída por rede geral (dos 5.531 municípios) e as demais por 7,5%, incluindo ETAs compactas (embora a maioria destas seja do tipo convencional), filtração direta e dessalinização, entre outros.

O tratamento de água possui papel fundamental para a adoção de medidas que visam à universalização do saneamento básico. Nesse contexto, a potabilização da água para o abastecimento humano merece grande destaque, notadamente quando da redução das doenças de veiculação hídrica. Assim, é necessário o atendimento às normas/resoluções/portarias que garantam que esse produto seja oferecido com quantidade e qualidade aos anseios dos consumidores.

Diante disso, há a necessidade da implantação das atividades de controle e vigilância do produto que está sendo gerado, conforme estabelecido pelo Anexo XX da Portaria de Consolidação nº5/2017 (BRASIL, 2017). Essa portaria também estabelece valores máximos permissíveis para determinados parâmetros de qualidade de água, além das metas que as empresas responsáveis pelos sistemas devem perseguir para oferecer um produto com segurança à população. Além disso, a Organização Mundial de Saúde prevê a implantação de

Planos de Segurança da Água (PSA), que são de extrema importância quando da tomada de decisão mediante situações de crise e/ou catástrofes que podem atingir os sistemas de tratamento (WHO, 2017).

Como forma de antever tais situações e, de uma maneira mais ampla, realizar a avaliação de desempenho das estações de tratamento de água, uma série de métodos foi proposta, fazendo uso de uma ampla gama de ferramentas, dentre as quais se destacam a análise multicritério, estatística multivariada, modelos de regressão linear múltipla, avaliação de ciclo de vida, modelos matemáticos mecanicistas e os modelos computacionais (Lógica Fuzzy, algoritmos genéticos, Redes Neurais Artificiais - RNA e sistemas híbridos, dentre outros).

A avaliação de desempenho surge sob a égide de uma abordagem mais holística sobre a gestão dos sistemas e serviços prestados. É possível, contudo, verificar que não basta realizar medições pontuais, mas é necessário traçar estratégias tangíveis para a aplicação de novas proposições de ferramentas de gestão, visando sempre à melhoria do que está sendo oferecido à população. Isso enseja na possibilidade da aplicação de indicadores de desempenho para aferição desses serviços. De acordo com Kaufman (1988), há dois usos primordiais para tais indicadores: a identificação daquilo que deve ser realizado e a apresentação de critérios para que seja determinado falha ou sucesso.

Para consolidar a aplicação da modelagem computacional na avaliação de desempenho de estações de tratamento de água, pode-se mencionar a Inteligência Artificial (IA), particularmente com o uso de RNA (RUSSELL e NORVIG, 2013). Nesse contexto, as RNAs têm a capacidade de realizar processamento de dados mediante valores de entrada, utilizados na etapa de treinamento ou aprendizagem, e, a partir dessa rede, podem ser obtidas decisões lógicas, conforme um cérebro humano, gerando informações de saída geralmente acuradas e precisas. A capacidade das RNAs na busca para a aplicação na modelagem de predição é justificada pela sua flexibilidade e generalização, além de sua capacidade de trabalhar com grande quantidade de informações simultaneamente (GRAUPE, 2007). Isso endossa ainda mais o uso desse tipo de modelagem em diversas atividades relacionadas ao tratamento da água, desde a estimativa da dosagem de produtos químicos, da geração de resíduos e subprodutos, da predição da qualidade do afluente à ETA, até das características da qualidade da água tratada, de acordo com as condições de entrada, fatores operacionais e parâmetros hidráulicos, quais sejam objeto deste estudo.

1.2 Justificativa e relevância

A demanda pelo consumo de água vem aumentando em todo o mundo e, com isso, a preocupação internacional se intensifica na busca da otimização dos sistemas existentes e na descoberta de outros mecanismos que possam garantir a toda a população mundial acesso a este recurso com qualidade. Tal premissa é fundamentada pelo objetivo 6 da agenda 2030 proposto pela ONU, que trata da necessidade de garantir o acesso, com qualidade, de toda a população do globo aos serviços de saneamento básico, particularmente no que concerne ao abastecimento de água potável (ONU, 2019). Neste sentido, emerge cada vez mais a preocupação dos sistemas de abastecimento de água, notadamente das estações de tratamento, para garantia da qualidade dos processos e na oferta de um produto confiável.

Com a tendência de atendimento aos critérios de qualidade de serviço e do fornecimento de um produto confiável (água), diversos sistemas de abastecimento de água começaram a realizar avaliações de desempenho das estações de tratamento como forma de facilitar e melhorar a gestão dos mesmos, garantindo maior eficiência, eficácia do processo e diminuição do risco de doenças de veiculação hídrica. Nesse contexto, uma série de índices e indicadores foi criada e estudada por diversos centros de pesquisas, e aplicadas por diversas empresas responsáveis por tal serviço. Seguindo essa tendência, em 2007 a *Internacional Organization of Standardization* (ISO) criou a ISO 24.510 (ABNT, 2012), que trata das diretrizes para a avaliação e melhoria dos serviços prestados aos usuários por entidades que desenvolvem atividades relacionadas a potabilização da água e ao esgotamento sanitário.

A avaliação de desempenho de estações de tratamento de água é uma importante ferramenta de gestão, sendo possível antever certos acontecimentos que podem trazer sérios prejuízos às populações por elas atendidas. É importante ressaltar que tais ferramentas possuem importância quando da implantação do PSA, o que corrobora ainda mais a necessidade das empresas de saneamento realizar essa avaliação.

Ainda sob a ótica da segurança da água e de modo a tornar os sistemas de abastecimento de água para consumo humano cada vez mais eficientes e seguros, é emergente a necessidade das empresas de saneamento se modernizarem, de modo a atender as necessidades cada vez mais crescentes da utilização de métodos e metodologias tecnológicas para a otimização de suas atividades.

Dentre essas, pode-se destacar a automação de estações de tratamento de água que já vem ocorrendo há alguns anos, de modo que, além da execução das tarefas simples de abrir e fechar uma válvula, possam realizar medições de alguns parâmetros, indicar certas situações de riscos existentes e tomar decisões a partir do treinamento prévio por reconhecimento de padrões. Essas ferramentas estão inseridas no grupo dos modelos computacionais ou *machine learning* (aprendizado de máquina), possuindo papel fundamental no conceito da Indústria 4.0, que vem ganhando, cada dia mais, espaço no setor industrial e nos novos empreendimentos.

Os sistemas de IA, notadamente as RNA, vêm sendo amplamente utilizados na área do saneamento ambiental, particularmente nas operações de estações de tratamento de água, com exemplo no processo de automação na dosagem de produtos químicos e na geração de subprodutos da cloração. É notório o elevado número de trabalhos que vem sendo desenvolvidos com base nessa ferramenta. Dentre eles, podem-se destacar as pesquisas realizadas por Chang *et al.* (2007), Sarkar e Pandey (2015) e Gazzaz *et al.* (2012), que utilizaram as RNAs na modelagem da qualidade da água de rios, e Chou *et al.* (2018), que realizaram a aplicação de *Machine Learning* na predição da qualidade da água de 20 reservatórios em Taiwan; Wu *et al.* (2014), Maier *et al.* (2010) e Bagheri *et al.* (2019) apresentaram grande contribuição para a aplicação das redes neurais por meio de uma ampla revisão da literatura, a partir de diversos artigos com significativas relevância científica; Vásquez *et al.* (2019), Karadurmuş *et al.* (2018), Memon *et al.* (2011), Baxter *et al.* (2002), Maier *et al.* (2004) propuseram importantes modelos de predição de parâmetros de qualidade da água em ETAs, bem como modelos de predição para dosagem de coagulantes e remoção de compostos persistentes; por outro lado Cordoba *et al.* (2014) também desenvolveram modelos de predição da qualidade da água de abastecimento humano em redes de distribuição, estimando a concentração do cloro residual em cada nó; por sua vez, Marzouk e Elkadi (2016) e Andrade *et al.* (2016) aplicaram as RNAs na análise de layout de ETAs e na estimativa de fatores de custos de implantação de estações; Adamowski *et al.* (2012) compararam os resultados de modelos de previsão desenvolvidos em regressão linear, não-linear e em RNA de uma ETA em Montreal, Canadá. Esses trabalhos evidenciam a aplicação das RNA em vários aspectos relacionados aos sistemas de abastecimentos de água, iniciando pela proposição de modelos de predição da qualidade da água de mananciais utilizados para abastecimento humano, até a eficiência de remoção de compostos específicos ao longo do tratamento e a análise de redes de distribuição de água.

No que concerne às etapas que ocorrem em uma ETA convencional, esses e outros trabalhos são amplamente divulgados na literatura. No entanto, quando se trata da avaliação de desempenho de ETA, não foi possível verificar essa aplicação de forma direta. Em sua grande maioria as pesquisas são voltadas para análise em tempo real, a partir da utilização de parâmetros de qualidade de água ou, como já anteriormente citado, não realizam o desenvolvimento de modelos de predição para tal finalidade. Portanto, esta pesquisa pretende realizar uma aplicação de RNA como ferramenta para a proposição de modelos de predição, destinada à avaliação de estações de tratamento de água, de forma a apresentar resultados rápidos e confiáveis aos órgãos gestores responsáveis pelos sistemas de abastecimento e antever situações de riscos.

Dentre as lacunas já citadas, além da predição de valores de parâmetros de controle da qualidade da água tratada, os modelos que aqui foram desenvolvidos têm o objetivo de criar intervalos de predição, de modo a facilitar os gestores na tomada de decisão, verificar possíveis situações que podem ocorrer na estação de tratamento de água e, de modo particular, em algumas etapas do tratamento. Assim, esses modelos também podem ser utilizados como importantes ferramentas de planejamento de manutenção das unidades das ETAs, bem como na necessidade de ampliação, reparo e modificações para atendimento de algum critério específico. Além de possibilitar a crescente automatização que já vem ocorrendo em diversos setores da economia e área de saneamento básico.

Além das vantagens supracitadas, pesquisas como essa abrem a possibilidade do desenvolvimento de diversos estudos com a utilização de ferramentas computacionais, que muito ainda têm a contribuir dentro do setor do saneamento ambiental. Por exemplo, o desenvolvimento de softwares para o gerenciamento de estações de tratamento de águas ou sistemas de abastecimento, automação de manobras desses sistemas, entre outros.

1.3 Hipóteses e Premissas

- **Modelos de predição baseados em Inteligência Artificial podem gerar a perspectiva de melhor gestão da estação de tratamento de água, permitindo que as intervenções sejam realizadas de forma antecipada e pontual.**

Premissa: Segundo Nourani *et al.* (2014), a IA é uma importante ferramenta na modelagem de sistemas não-lineares, permitindo a manipulação de grande número de variáveis, bem como uma vasta dinamicidade quando de sua aplicação. Essas características favorecem sua aplicação na otimização de processos matemáticos, estatísticos, lógicos, probabilísticos, entre outros. No caso do estudo proposto, de acordo com os dados que compõem o modelo, é possível antever certos acontecimentos, o que contribui para uma melhor tomada de decisão e gestão da estação de tratamento.

- **Redes do tipo *backpropagation* apresentam boa capacidade de generalização e flexibilidade do modelo proposto.**

Premissa: Segundo Olawoyin (2016), as RNA do tipo *backpropagation* são uma importante ferramenta matemática funcional, notadamente, na aproximação e previsão de séries temporais em situações complexas que apresentam funções não-lineares. Somado ao exposto, esse algoritmo permite que a própria rede calcule o erro entre as saídas e essas se retroalimentem na busca de um erro tolerável.

- **Quanto maior o número de neurônios na camada oculta, maior será a otimização no processamento da rede e sua capacidade de generalização, embora o treinamento possa se tornar excessivo.**

Premissa: Há uma dicotomia quanto ao número de neurônios na camada oculta, bem como o número de camadas, tendo em vista que, quanto maior o número de neurônios, maior será a capacidade de modelagem ou de generalização da rede. Todavia, isso poderá implicar em um maior gasto de processamento na atividade de aprendizagem da rede (HE *et al.*, 2016).

- **Modelos desenvolvidos em RNA poderão apresentar melhores resultados quando comparados com modelos baseados em regressão linear múltipla.**

Premissa: De acordo com a literatura, os modelos baseados em RNA oferecem maior vantagem em relação à regressão linear múltipla, devido ao aspecto da aprendizagem, a qual permite maior generalização, bem como uma modelagem mais realista dos sistemas em estudo; outrossim, é importante salientar que cada modelo tem suas peculiaridades, particularmente seu processo matemático.

1.4 Objetivos

1.4.1 Objetivo geral

O objetivo geral desta tese é desenvolver modelos de predição de desempenho de estações de tratamento de água (para três portes distintos de ETAs), do tipo convencional, utilizando RNA, com base na turbidez e na cor aparente da água tratada, nos parâmetros hidráulicos e operacionais dessas estações.

1.4.2 Objetivos específicos

- Elaborar modelos de predição, a partir da regressão linear múltipla, com prévia análise de correlação entre as variáveis disponíveis;
- Elaborar modelos de predição em RNA e, a partir dos valores preditos, gerar intervalos de predição;
- Verificar a possibilidade de desenvolvimento de um modelo genérico, independente do porte da ETA, a partir dos resultados dos modelos; e
- Comparar os modelos desenvolvidos em RNA e com os modelos desenvolvidos em regressão linear múltipla.

2. REVISÃO DA LITERATURA

2.1 *Análise de desempenho de estações de tratamento de água*

A aplicabilidade da análise de desempenho como instrumento avaliativo e/ou como parte integrante de um processo de gestão já vem sendo utilizada há décadas, tendo sido empregada em alguns setores de serviços, particularmente daqueles sob responsabilidade de entes públicos. Segundo ICMA (1999), em 1938 já havia, por exemplo, o conceito de “*measuring municipal activities*” (medindo atividades municipais) – uma forma que governantes locais desenvolveram para poder analisar, a partir de vários tipos de informações, o acesso e a qualidade dos serviços prestados à população. No desenvolvimento de tal ação se faz necessária a adoção de indicadores representativos específicos, capazes de apontar evidências, com eficiência e eficácia, de acordo com os objetivos da análise que se pretende realizar (CHANG *et al.*, 2007).

Segundo Kaufman (1988), há dois usos primordiais para tais indicadores: a identificação daquilo que deve ser realizado e a apresentação de critérios para que seja determinado se há falha ou sucesso. Todavia, é importante ressaltar que o uso dos indicadores é apenas um, dentre tantos, dos métodos utilizados nesse tipo de análise.

No âmbito das estações de tratamento, a análise de desempenho é uma prática que já vem sendo bastante consolidada como ferramenta de controle e gerenciamento das mesmas, buscando a melhoria contínua do processo e da qualidade. É de senso comum a excelência de sua aplicação no saneamento ambiental, tendo-se como exemplo as estações de tratamento de águas residuárias, com vistas à redução da concentração de alguns parâmetros de qualidade, bem como para o atendimento aos padrões de lançamento preconizados pelos órgãos ambientais; para a otimização de processos e do consumo de produtos e insumos utilizados nas etapas de tratamento, bem como na minimização de resíduos e/ou subprodutos gerados (SWEETAPPLE *et al.*, 2014). Portanto, como forma de auxiliar os tomadores de decisões e gestores dos sistemas de abastecimento de água, tem-se a avaliação de desempenho de estações de tratamento de água.

Vieira *et al.* (2008) apresentam uma metodologia genérica a ser utilizada em estações de tratamento de água como forma de estabelecer indicadores de desempenho (Figura 1). Ainda segundo esses autores, o objetivo inicial dessa metodologia é a definição dos parâmetros mais

comuns entre as ETAs, sendo esses necessários para uma operação eficiente e eficaz. Elencaram-se sete componentes relacionados ao desempenho das ETAs, quais sejam: qualidade da água tratada, confiabilidade da ETA, uso de recursos naturais e matérias-primas, gerenciamento de subprodutos, segurança, recursos humanos, e recursos econômicos e financeiros.

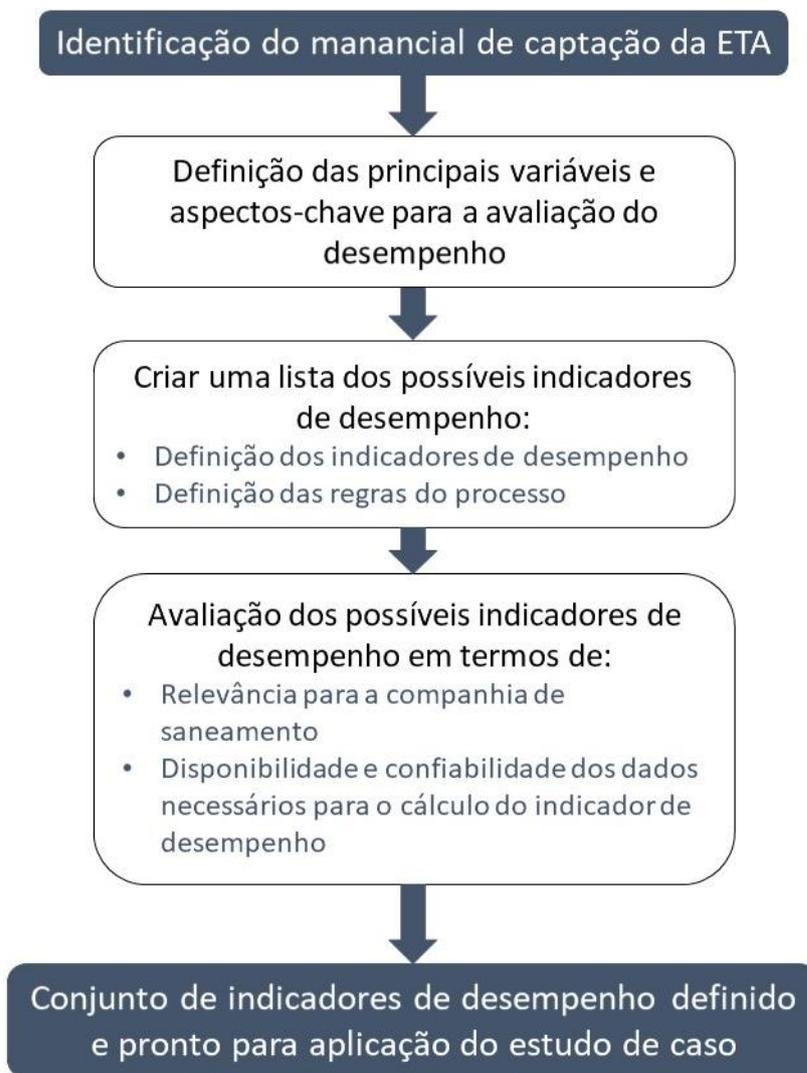


FIGURA 1: Metodologia para a definição dos indicadores de desempenho

Fonte: Adaptado de Vieira *et al.* (2008).

Nas estações de tratamento de água para abastecimento humano tal análise se reveste de grande importância, notadamente como forma de eliminar e/ou mitigar os riscos à saúde associados com a qualidade da água produzida. Segundo Alegre *et al.* (2006), os indicadores de desempenho podem ser divididos nos seguintes grupos: recursos hídricos, pessoal, aspectos físicos, operacionais, qualidade dos serviços e levantamento econômico e financeiro. Diante

do exposto, essa análise pode ser realizada a partir de diversos critérios, dentre os quais podem-se destacar:

- Atendimento aos padrões de potabilidade em vigor;
- Capacidade de remoção e/ou diminuição aos padrões aceitáveis de um determinado parâmetro de qualidade;
- Atendimento aos parâmetros hidráulicos de projetos, conforme recomendação das normas técnicas;
- Fatores operacionais, tais como ensaios em *Jar Test* para escolha da dosagem de coagulante;
- Grau de instrução dos operadores, que pode implicar em operação menos intuitiva e, por conseguinte, mais técnica;
- Consumo de energia elétrica;
- Quantidade de lodo produzido; e
- Quantidade de produto químico utilizado.

Em 1998, a Agência de Proteção Ambiental dos Estados Unidos da América (USEPA) publicou um documento técnico que traz em seu escopo formas de otimização das operações unitárias que envolvem o tratamento de água a partir do desempenho de uma ETA, com base em uma metodologia denominada *Composite Correction Program* (CCP), já aplicada desde 1988 em diversos sistemas de abastecimento de água dos Estados Unidos e Canadá. A finalidade básica desse programa era dividir todas as operações unitárias de uma ETA e analisá-las individualmente, de modo a estabelecer barreiras disponíveis eficazes, no processo de tratamento de água, contra a passagem de material particulado e/ou de agentes patogênicos na água tratada (USEPA, 1998). Apesar do objetivo desse programa, verifica-se que há um complexo estudo para o entendimento do processo de tratamento de água que ocorre em cada ETA, notadamente quando da priorização de critérios para o controle da qualidade da água produzida.

A utilização do conceito de múltiplas barreiras, para materiais persistentes e para organismos patogênicos, é citada por várias agências reguladoras como forma de avaliação do desempenho de estações de tratamento (CCME, 2004; ALBERTA, 2012). Isso se deve ao fato da diminuição dos riscos à saúde, pois quanto mais eficientes forem as barreiras, menor será a chance de ocorrência de um possível contaminante na água (ZHANG *et al.*, 2012).

O'Connor *et al.* (2009) realizaram uma avaliação do desempenho de ETAs com base na remoção de microrganismos e material particulado. Como forma de melhor realizar a avaliação, todas as ETAs escolhidas foram divididas em várias etapas de tratamento, de modo a identificar cada fase do sistema. Por exemplo, foram consideradas as etapas que envolvem a adição de produtos químicos, o decaimento do número de células de indicadores microbiológicos nas operações, entre outros. Nesse estudo, a avaliação de desempenho foi realizada separadamente, para cada operação unitária que compõe a ETA, analisando os resultados dos parâmetros de qualidade da água, na entrada e na saída, bem como sua capacidade de remoção de patógenos. Os autores apontam a necessidade de técnicas mais avançadas para otimização dos processos de tratamento e para o melhor controle desses.

Apesar das diversas formas de aplicação, essa atividade não é realizada de forma intuitiva, pois há diversos métodos descritos na literatura e toda uma exigência por parte das agências reguladoras e órgãos ambientais para sua execução. Dentre os marcos regulatórios que preveem a análise de desempenho de sistemas que produzem água potável, tem-se a ISO 24.510/2007, que norteia as atividades de avaliação e de melhorias em estações de tratamento de água e de esgotamento sanitário. Essa norma estabelece a necessidade da adoção de critérios para tal avaliação dos sistemas, bem como o uso de indicadores de desempenho atrelados a esses critérios (ABNT, 2012). No entanto, é importante ressaltar que essas atividades são para fins de ilustração, tendo em vista que esse tipo de avaliação não se limita à adoção de um grupo de indicadores de desempenho. Contudo, a aplicação dessa Norma aos sistemas de produção de água potável garante ainda mais sua confiabilidade e enfatiza as ações previstas no PSA estabelecido pela Organização Mundial de Saúde (WHO, 2017).

Segundo Oliveira (2014), inicialmente nessa avaliação levava-se em consideração apenas o atendimento aos aspectos legais vigentes para a potabilização da água, tendo como parâmetro fundamental a turbidez efluente. Contudo, outros conceitos foram introduzidos (GUPTA e SHRIVASTAVA, 2006; ZAKARIAN *et al.*, 2007; GUNDERSON e PRICHARD, 2002), tais como:

- Confiabilidade – capacidade de uma estação produzir água dentro dos padrões de qualidade;
- Robustez – capacidade de uma estação atender à legislação independentemente da qualidade do afluente; e

- Resiliência – tempo que uma ETA leva para recuperar sua operação ordinária após uma interferência abrupta.

A partir dos conceitos supracitados, diversos métodos de análise de desempenho para estações de tratamento de água foram desenvolvidos. Dentre eles, podem-se destacar os métodos: (i) fundamentados apenas nos valores das variáveis de controle, sendo esses os mais comuns e primitivos; (ii) baseados no uso de indicadores de desempenho, a partir de conjunto de parâmetros de maior relevância para a análise; (iii) os baseados na utilização de índices de desempenho, fazendo uso da estatística multivariada para a adoção de índices representativos que possam caracterizar a estação de tratamento; (iv) os fundamentados no uso da modelagem estatística, a partir de regressão linear múltipla; (v) definidos por modelagem matemática, a partir do desenvolvimento de equações matemáticas que simulam um sistema com base nos seus dados de entrada e de saída. Todas essas ferramentas expressam o desempenho da estação com base em uma, ou mais, variável resposta de interesse.

2.1.1 Ferramentas para a análise de desempenho de estações de tratamento de água

Para a realização da análise de desempenho das ETAs, diversas ferramentas foram desenvolvidas, aperfeiçoadas e aplicadas ao longo do tempo. Inicialmente, o objeto mais comum para avaliar o desempenho de uma ETA estava, indubitavelmente, no monitoramento de alguns parâmetros de qualidade da água, fato esse que ainda é imposto como forma de controle aos prestadores por meio do Anexo XX da Portaria de Consolidação nº5/2017 (BRASIL, 2017).

Perseguindo a supracitada analogia, uma gama de aplicações ainda vem sendo realizada em várias partes do mundo. Dentre esses trabalhos podem-se destacar os desenvolvidos por Ahmadli *et al.* (2016), Mota *et al.* (2015), Makungo *et al.* (2011), Ali *et al.* (2012), Al-Jeebory e Ghawi (2009), Farhaoui e Derraz (2016) e Gholikandi *et al.* (2012), em que os parâmetros de qualidade da água são o objeto principal para a avaliação de desempenho das estações em estudo. Em alguns desses trabalhos observa-se notório interesse pelos processos e operações unitárias típicas em ETA convencionais, tais como a etapa de filtração, por meio do estudo da granulometria da areia utilizada como meio filtrante. Também é relevante destacar o enfoque dado à remoção e/ou ao controle da turbidez, bem como aos indicadores de

organismos patogênicos, como forma de analisar o desempenho da ETA a partir de uma ótica do risco à saúde dos consumidores.

Além da utilização dos parâmetros de qualidade da água como ferramenta de análise de desempenho, diversos indicadores foram desenvolvidos em todo o mundo para tal finalidade, levando em consideração vários aspectos, tais como parâmetros de projetos estruturais de concepção das ETAs (principalmente os hidráulicos), aspectos operacionais das ETAs, qualidade de água bruta e tratada e atendimentos aos padrões de potabilidade em vigor, dentre outros (ZHANG *et al.*, 2012).

Alguns desses indicadores foram desenvolvidos no Programa de Pós-Graduação em Saneamento, Meio Ambiente e Recursos Hídricos (PPG-SMARH) da Universidade Federal de Minas Gerais, como o Índice de Qualidade de Estações Convencionais de Tratamento de Água – IQETA (LOPES, 2005; LOPES e LIBÂNIO, 2009; ALMEIDA, 2009) e o Índice de Qualidade da Estação de Tratamento de Água Fuzzy – IQETAF (OLIVEIRA, 2014). No desenvolvimento desses índices e modelos, diversos métodos foram utilizados, dentre eles podem-se destacar a Estatística Multivariada, a Análise de Regressão, a Lógica Fuzzy, a RNA e sistemas híbridos (Neuro-Fuzzy). Esses são baseados em métodos estatísticos clássicos, lógica booleana com aprendizagem humana e simulação do raciocínio humano, respectivamente. As técnicas estatísticas também são utilizadas como ferramenta de análise de desempenho de ETA, particularmente quando da necessidade de avaliar estações diferentes de maneira simultânea (MELO *et al.*, 2015).

Bereskie *et al.* (2017) desenvolveram e aplicaram mais uma ferramenta capaz de avaliar o desempenho de pequenas estações de tratamento de água. A metodologia é um processo de melhoria contínua (do inglês *Continuous Performance Improvement* - CPI), aplicada em sete ETAs de pequeno porte localizadas na província de *Newfoundland and Labrador*, Canadá. Para a realização da pesquisa, os autores dividiram a metodologia em cinco etapas, desde a coleta de dados até a implantação de ações de melhoria. De posse dos dados, foi possível estabelecer indicadores de qualidade da água. Em seguida foi implementado o *Benchmarking* de desempenho funcional, identificando-se as ações de prioridade em cada ETA, com vista ao processo de melhoria contínua. A partir dos resultados apresentados, foi verificado que o CPI, além de avaliar o desempenho atual das ETAs, é capaz de projetar melhorias para essas, possibilitando estratégias para a produção de água de melhor qualidade.

O Quadro 1 ilustra diversos trabalhos que utilizam outras ferramentas na análise de desempenho de ETAs.

QUADRO 1: Principais ferramentas utilizadas na avaliação de desempenho de ETA

Autores	Periódico	Escopo da pesquisa	Ferramenta utilizada	Conclusão / Recomendações
Stein <i>et al.</i> (2017)	<i>Water and Environmental Journal</i>	Uso de metodologia de apoio à decisão para avaliação do desempenho operacional de ETA de pequeno porte	AHP / Análise de Árvore de Falhas / Sistema de Apoio a Decisão	Apesar de robustas, tais ferramentas devem ser aplicadas de forma diferenciada em cada ETA
Chang <i>et al.</i> (2007)	<i>Practice Periodical of Hazardous, Toxic, and Radioactive Waste Management</i>	Implementação de mecanismos de avaliação de desempenho de ETA a partir da hierarquização dos processos, com base na informação prévia e/ou de especialistas sobre o sistema de tratamento (CPE)	AHP / CPE	De forma detalhada, foi possível avaliar cada item responsável pelo desempenho da estação, que resulta na qualidade da água tratada, minimizando os custos totais do tratamento
Tomar e Borad (2012)	<i>International Journal of Engineering Research and Development</i>	Uso da AHP aliada à CPE para adoção de pesos e avaliação dos parâmetros da ETA	AHP / CPE	O levantamento de ponderações relativas através da AHP foram úteis na análise de eficiência, identificando áreas que requerem maior atenção, com base em prioridades estabelecidas
Sadhvani e Bhawe (2016)	<i>International Advanced Research Journal in Science, Engineering and Technology</i>	Revisão de literatura sobre a aplicação de métodos para a avaliação do desempenho de ETAs, com vistas à otimização das mesmas	AHP / CPE / CCP	A CPE foi satisfatória, quando aliada aos métodos utilizados, visando à segurança da água de abastecimento
Borghetti <i>et al.</i> (2013)	<i>The International Journal of Life Cycle Assessment</i>	Desenvolvimento de Análise de Ciclo de Vida (LCA) em sistemas de abastecimento de águas na Sicília-Itália. Comparação entre diferentes tipos e metodologias de estudos	LCA	A LCA é uma ferramenta importante de tomada de decisão para dar suporte na seleção de processos, numa perspectiva de redução de impactos ambientais

Continua

QUADRO 1: Principais ferramentas utilizadas na avaliação de desempenho de ETA*Continuação*

Arbault <i>et al.</i> (2013)	<i>Ecological Engineering</i>	Aplicação da metodologia de Avaliação em Emergia (EmE), sendo a mesma utilizada em diversas etapas do tratamento de água. Comparação entre a EmE e LCA em ETA na França	EmE / LCA	A EmE apresentou baixa precisão quando comparada com a LCA, quando a mão de obra predominou.
Igos <i>et al.</i> (2013)	<i>Science of the Total Environmental</i>	Desenvolvimento de indicador de custo de desempenho para duas ETAs localizadas em Paris. O indicador relaciona o impacto ambiental gerado na produção de água com o ganho da qualidade total	LCA / Indicador de custo de desempenho / ReCiPe / Stewise / Eco-costs	A LCA foi capaz de calcular, com êxito, o custo de performance, sendo necessário aprofundamento do trabalho para o levantamento do ganho médio de qualidade
Garfí <i>et al.</i> (2016)	<i>Journal of Cleaner Production</i>	Avaliação de impacto ambiental causado pelo consumo de água em Barcelona. Cinco cenários foram avaliados entre tratamento e formas de envasamento	LCA – Uso de software	De modo geral, ETAs convencionais apresentaram menor impacto. Água envasada apresentou maior impacto, particularmente aquelas que utilizam garrafas de vidro
Rodriguez <i>et al.</i> (2016)	<i>Ambiente e Água - An Interdisciplinary Journal of Applied Science</i>	Aplicação da Análise do Ciclo de Vida para a avaliação do impacto ambiental em quatro ETAs localizadas no Nordeste da Colômbia, a partir da ISO 14040/2006	LCA	A proteção do manancial é a melhor forma de prevenção de deterioração da qualidade da água. Mudanças no processo e na técnica de tratamento podem influenciar a análise
Capitane scu <i>et al.</i> (2016)	<i>Journal of Environmental Management</i>	Modelagem computacional com seis algoritmos (SPEA2, NSGA-II, IBEA, MOEA/D, DE, PSO), a partir do custo operacional e LCA, na qualidade da água	LCA / Modelagem comp. de otimização	O algoritmo NSGA-II se apresentou mais eficiente para os tomadores de decisão, na redução de impactos e custo de operação através da LCA

Legenda: AHP – Análise Hierárquica de Processo, ETA – Estação de Tratamento de Água, CPE – Avaliação de Desempenho Abrangente, CCP – Programa de Correção Composta, EmE – Avaliação em Emergia, LCA – Análise de Ciclo de Vida.

Nessas pesquisas é possível verificar que as ferramentas são mais subjetivas, com uma visão mais ampla dos sistemas. Também é notório que sua utilização está vinculada ao objetivo de cada tipo de análise, podendo as mesmas serem mais superficiais ou mais específicas em determinadas partes que compõem as estações de tratamento de água.

Dentre as ferramentas citadas, pode-se destacar a LCA, do inglês *Life Cycle Assessment* (Avaliação do Ciclo de Vida), relacionada aos impactos ambientais causados por um determinado produto, sendo levadas em consideração suas etapas de produção e consumo. A LCA, comumente denominada de avaliação “do berço ao túmulo” – normatizada pela ABNT NBR ISO 14.040/2009, avalia o processo desde a aquisição da matéria-prima, linha de produção, utilização, descarte, reaproveitamento e/ou reutilização e disposição final. O estudo da LCA (Figura 2) apresenta quatro fases: (i) definição dos objetivos e escopo da avaliação; (ii) análise do inventário do sistema de produção; (iii) avaliação de impactos; e (iv) interpretação (ABNT, 2009). Outra maneira seria a aplicação da avaliação de impacto do ciclo de vida (LCIA), que se trata de um estudo mais aprofundado dos impactos ambientais do processo.

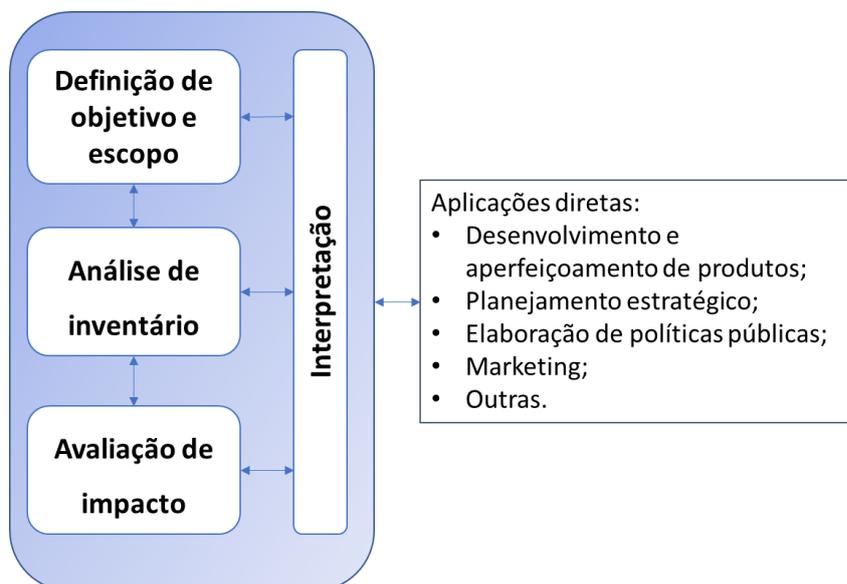


FIGURA 2: Fases de aplicação da Análise do Ciclo de Vida

Fonte: Adaptado de ABNT (2009).

No que se refere às ETAs, diversos autores destacam a LCA e suas variantes como importantes ferramentas, tendo em vista que sua grande área de aplicação está voltada para a análise de resíduos gerados no processo, do consumo de insumo e de energia elétrica, da gestão realizada nos mananciais, da comparação entre as etapas do tratamento e dos diversos

tipos de tratamento empregados, entre outros (BONTON *et al.*, 2012; AHMADI e TIRUTA-BARNA, 2015; AHMADI *et al.*, 2016).

Adicionada ao exposto, outra ferramenta com grande capacidade intuitiva é a AHP, *Analytic Hierarchy Process* – Análise Hierárquica de Processo, sendo bastante difundida no meio acadêmico e em diversas aplicações práticas nas áreas de recursos hídricos, de saneamento ambiental, entre outras. A AHP, desenvolvida por Thomas Saaty nos anos 1970, faz parte do grande grupo dos Métodos Multicritério de Apoio à Tomada de Decisão.

Essa ferramenta utiliza o processo lógico matemático, aliado a fatores de conhecimento técnico que irão produzir um conjunto hierárquico de informações ou variáveis, a julgar pelos pesos e/ou grau de importância dentro de um determinado processo. O objetivo a ser alcançado com a aplicação da AHP é o desenvolvimento de prioridades para adoção de alternativas na tomada de decisão, bem como criar critérios a serem utilizados no julgamento dessas alternativas (SCHMOLDT *et al.*, 2001).

No que concerne à avaliação de desempenho de estações de tratamento de água, o Quadro 1 apresenta diversas pesquisas que utilizam a AHP como forma de realizar o levantamento de indicadores / parâmetros / critério para tal processo avaliativo. Nas aplicações aqui ilustradas nota-se que essa ferramenta geralmente não é implementada sozinha, já que sempre há alguns métodos de fundamental importância para dar suporte ao resultado da avaliação. Dentre esses podem-se destacar o CCP, citado anteriormente, tendo como foco os mecanismos de múltiplas barreiras contra organismos patogênicos nas estações de tratamento de água; e *Comprehensive Performance Evaluation* (CPE) – uma parte integrante da metodologia CCP, que é utilizada para a realização de uma avaliação mais detalhada do desempenho da estação de tratamento. Inicialmente, a CPE parte de uma análise minuciosa da capacidade de projeto instalada, avaliação das operações unitárias, identificação e priorização dos fatores limitantes de desempenho, avaliação do acompanhamento e relatório da avaliação (USEPA, 1998).

Ainda sob a ótica da avaliação de desempenho das estações de tratamento de água, pode-se destacar o trabalho desenvolvido por Almeida e colaboradores, que realizaram a avaliação de desempenho de uma ETA convencional, a partir do monitoramento de parâmetros de qualidade da água (cor aparente, turbidez e pH), fazendo sua relação com os eventos de precipitação (dividindo a análise entre período chuvoso e seco). Foi verificado que a ETA não

consegue atender ao padrão de potabilidade em ambos os períodos, sendo necessárias adequações para que se possa ofertar água com qualidade aceitável (ALMEIDA *et al.* 2017).

De modo similar Khan e Ahmad (2018) também realizaram a avaliação de desempenho de uma ETA, na Índia. Eles utilizaram um número mais amplo de parâmetros físico-químicos, como oxigênio dissolvido, sólidos totais, sólidos dissolvidos, temperatura, cloro residual livre, além de indicadores microbiológicos. Eles concluíram que a turbidez da água tratada se manteve dentro dos padrões de potabilidade. Já Madhu e Soumyashree (2016) aplicaram a ideia similar para avaliação de unidade de tratamento de água em Kartanaka – Índia, concluindo que a água tratada atendeu aos padrões de potabilidade. Por outro lado, Mishra e Kadu (2014), também na Índia, tomaram como referência a capacidade da ETA de remover compostos metálicos e patógenos. Dohare e Gupta (2014) realizaram uma revisão de literatura para o desenvolvimento de índices de qualidade de água de 2007-2013, utilizados na avaliação de desempenho de estações de tratamento de água entre 2003 e 2014. Eles apontam que não basta a ETA ser bem dimensionada para garantia da qualidade da água, mas a operação é fator primordial do processo, assegurando ainda que a avaliação de desempenho é essencial para o entendimento do processo de tratamento e sua eficiência.

Esse tipo de abordagem, que vem sendo utilizado para a avaliação de desempenho das estações de tratamento de água é considerado moroso e, de certa forma, apresenta um custo relativamente elevado em detrimento de outras metodologias já citadas. Outra questão interessante é que a avaliação ocorre apenas de forma pontual e para uma determinada ETA específica, segundo suas características de projeto, a qualidade da água bruta e os aspectos operacionais em cada unidade de tratamento. Nesse contexto, verifica-se a necessidade da aplicação de ferramentas mais científicas e tecnológicas, já aplicadas em outras frentes de pesquisas, que sejam capazes de responder de maneira mais ampla e objetiva às demandas apresentadas no processo de gerenciamentos dos sistemas de abastecimento de água, notadamente nas ETAs. Dentre essas ferramentas podem-se destacar os sistemas artificiais inteligentes, amplamente utilizadas nas mais diversas áreas de pesquisa e desenvolvimento tecnológico e científico em todo o mundo.

2.2 Sistemas artificiais inteligentes

Com o desenvolvimento do cálculo matemático e a aplicação da lógica, passou-se a utilizar desses recursos como ferramental básico para a realização de diversas tarefas no dia a dia.

Inicialmente tratada apenas como uma simples realização de cálculos complexos, a IA perpassa essas aplicações, assumindo seu papel em um espaço mais amplo e em aplicações interdisciplinares, tais como sistemas econômicos e financeiros, sistemas educacionais, desenvolvimento industrial (automação e controle de diversos processos computacionais), sistemas político e institucional, fomento para cultura organizacional nas instituições públicas e privadas, ferramenta de gestão de análise e processo de sistemas de tratamento, entre outros.

Segundo Russell e Norvig (2013), a IA surgiu como novo ramo das ciências e das engenharias após a Segunda Guerra Mundial, tendo como premissa quatro dimensões básicas, quais sejam, pensando e agindo como seres humanos, pensando e agindo racionalmente, conforme ilustrado na Figura 3. A partir das dimensões utilizadas para a definição da IA encontra-se em comum o raciocínio ou a utilização de uma forma de processamento de informação, por meio de uma imitação do cérebro humano ou da própria atividade computacional, ou seja, a realização de um processo e/um processamento por meio de uma máquina.



FIGURA 3: Dimensões da Inteligência Artificial

Fonte: Adaptado de Russell & Norvig (2013).

Atualmente é cada vez maior a aplicação da IA nas diversas atividades rotineiras, tais como o processo de robotização em veículos, a partir de uma composição de sensores e/ou “agentes”

capazes de entender o ambiente ao seu redor e, a partir disso, apresentar respostas por meio de mecanismos denominados “atuadores”. Esses são responsáveis pela realização de tarefa pré-determinada (Figura 4), no caso dos automóveis, conseguem realizar algumas manobras para estacionamento de maneira autônoma. Outras aplicações que se pode destacar são o planejamento logístico, a tradução automática, o reconhecimento de voz, o planejamento autônomo, os processos aleatórios, a automatização de máquinas em geral, o reconhecimento de padrões para identificação de doenças em diagnóstico por imagens e a escolha de padrões de cultivo, dentre outros.

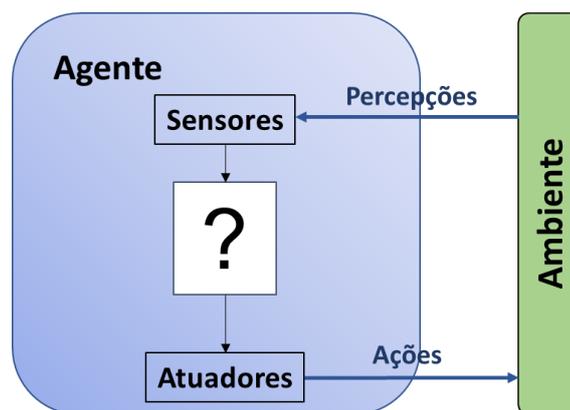


FIGURA 4: Atuação entre agente e ambiente por meio de sensores e atuadores em IA
Fonte: Adaptado de Russell & Norvig (2013).

A IA apresenta-se como ferramenta para a resolução de vários sistemas que requerem a inteligência humana, sendo definida como um sistema que apresenta habilidade de aprendizado efetivo (KASABOV, 1996). Dentre esses, podem-se destacar: a Lógica Fuzzy, uma evolução da Lógica Booleana, a partir da teoria dos conjuntos nebulosos, que permite valores das variáveis intermediários entre 0 e 1, dentro de um conjunto de regras previamente estabelecidas, as RNAs, sistemas que se baseiam no processamento cerebral humano para tomada de decisões a partir de um processo de aprendizado prévio, e os sistemas híbridos, utilizando RNA e Lógica Fuzzy simultaneamente (HAYKIN, 2001).

2.2.1 Redes Neurais Artificiais

Como parte da grande linha da modelagem computacional, a RNA é uma ferramenta funcional, dotada de aplicações múltiplas, notadamente nas soluções complexas e que requeiram acurácia e precisão nos seus resultados. Portanto, ressalta-se a importância de se trabalhar com dados confiáveis, oriundos de fontes seguras e de acordo com métodos

padronizados de monitoramento e análise. Isso se faz necessário tendo em vista que, segundo a literatura, quando se trabalha em modelagem de sistemas complexos, com informações que não atendem a esses critérios, seus resultados se tornam mera especulação, impossibilitando sua aplicação e fragilizando-os.

As RNAs possuem algumas características que as diferenciam dos demais tipos de ferramentas computacionais, sendo que estas estão intimamente relacionadas com a sua aplicação. Para tanto, Da Silva *et al.* (2016) elencaram algumas dessas características que podem facilitar quando da necessidade de sua escolha, quais sejam: Adaptação por experiência, Capacidade de aprendizagem, Habilidade de generalização, Organização de dados, Tolerância a falhas, Armazenamento distribuído e Facilidade de prototipagem.

Dentre as características anteriormente citadas, destacam-se: a adaptação a partir de padrões assimilados pelas RNA, os quais estão diretamente relacionados aos valores dos pesos sinápticos; a capacidade de aprendizado mediante o treinamento prévio, de acordo com os valores das variáveis apresentadas; a capacidade de generalização, conforme seu treinamento, uma vez que a rede é capaz de estimar valores a partir de dados de entrada desconhecidos; e a possibilidade de transformação do processamento da RNA em um hardware, permitindo um número maior de aplicação (DA SILVA *et al.*, 2016).

As primeiras informações sobre as RNA surgiram em 1943, com Warren Mc Culloch, psiquiatra e neuroanatomista, e Walter Pitts, matemático. A ideia desses pesquisadores era simular o cérebro humano e todo o seu processo de raciocínio, criando um modelo matemático com base no neurônio biológico, daí o termo “rede neural”. Esse modelo foi denominado de lógica limiar e precursor necessário para que fosse dado início aos estudos relacionado às RNAs até os dias atuais, conforme ilustrado na Figura 5. Posteriormente ao desenvolvimento do modelo matemático proposto por Mc Culloch e Pitts, diversos avanços ocorreram no campo das redes neurais artificiais, com vista à necessidade cada vez mais frequente do atendimento as aplicações no contexto do desenvolvimento científico e tecnológico. Ao passo que as tecnologias computacionais foram avançando, os pesquisadores foram buscando outras maneiras de aplicar e desenvolver os conhecimentos de RNA. Nesse contexto, outros tipos de redes neurais artificiais foram surgindo, assim também como sua forma de aplicação aliada a outras ferramentas (AGGARWAL, 2018).

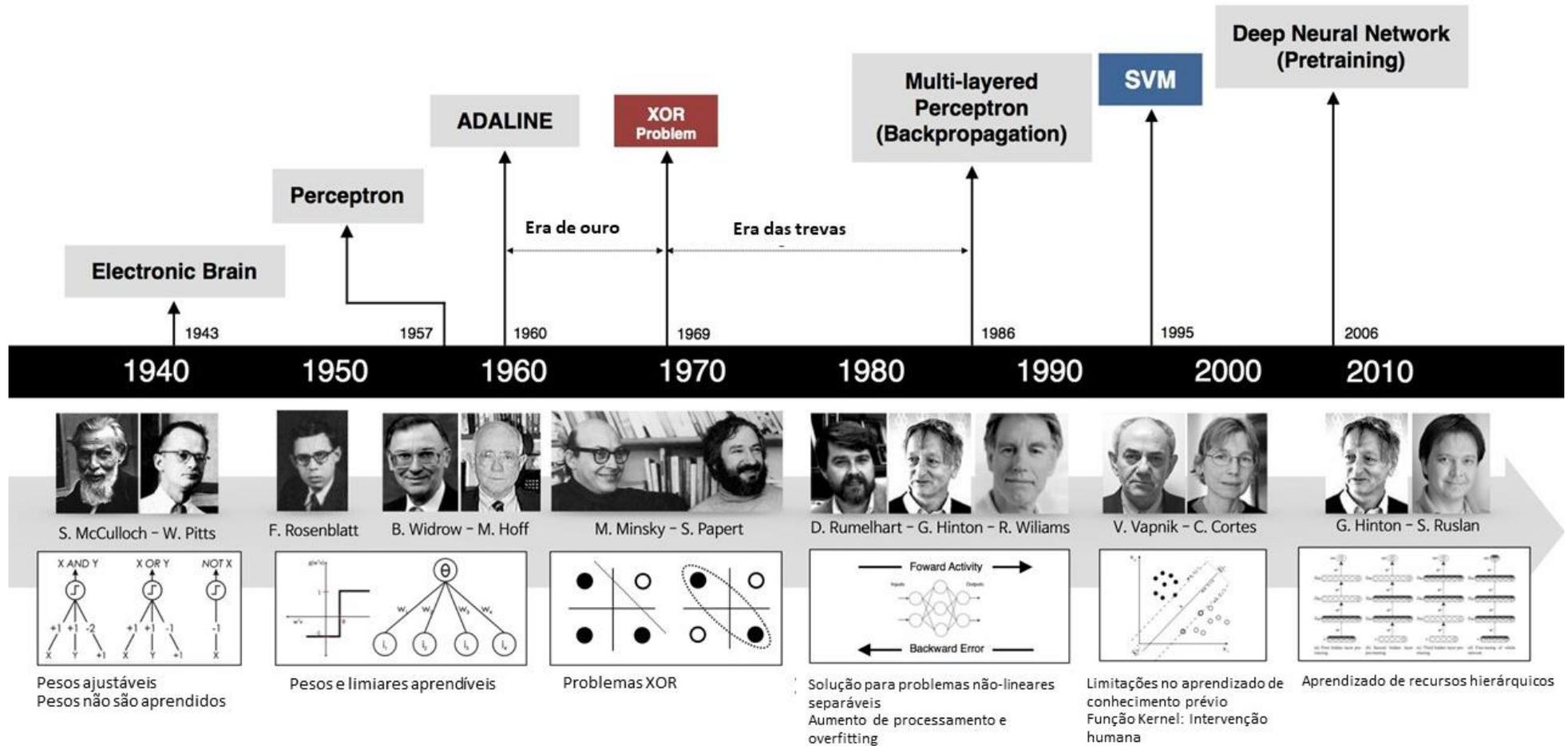


FIGURA 5: Cronologia do desenvolvimento das RNAs e seus principais responsáveis

Fonte: Adaptado de Deep Learning Book (2018).

Um sistema baseado em RNA é representado pela formação de uma rede composta por neurônios (Figura 6), sendo que cada um desses é formado por nós, sinapses ou vieses e função de ativação; em cada nó há um peso específico, responsável pelo efeito que terá sobre a etapa seguinte (ARAM *et al.*, 2017). Em sua aplicação, são utilizadas diversas formas de arquitetura, sendo que sua escolha está condicionada à aplicabilidade da rede, tendo em vista seu desempenho mediante à tarefa a ser realizada e o processamento dos dados de entrada.

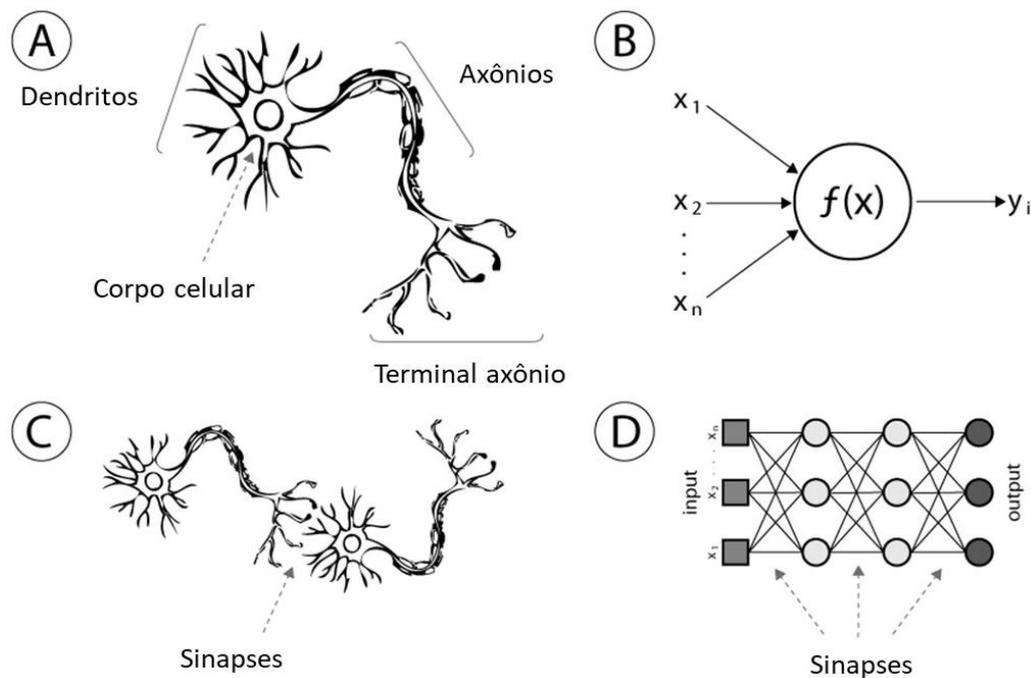


FIGURA 6: Representação de um neurônio natural (A), artificial (B) e as sinapses que ocorrem entre ambos (C e D)

Fonte: Adaptado de Medium (2018).

Outra importante vantagem da aplicação das RNA é a sua capacidade de aprendizagem e de generalização. Isso a distingue das demais aplicações de modelos, como, por exemplo, a Lógica Fuzzy. Tal característica lhes permite maior aplicabilidade e, de certa forma, autonomia na realização da operação a que se destina. Essas são apenas algumas das características que justificam a utilização de RNA na análise de desempenho de ETA, ainda mais quando do seu uso em modelos de previsão de desempenho para produção de água para consumo humano.

Em se tratando de redes neurais, dois aspectos são de fundamental importância para sua implementação: sua arquitetura – relacionada ao posicionamento dos neurônios que a compõem e a sua quantidade e distribuição; e a topologia – intimamente relacionada àquilo

que forma a RNA e dá sentido ao seu funcionamento, função de ativação, algoritmo de treinamento, cálculo de erro, entre outros.

No que concerne à arquitetura das RNA, pode-se destacar as seguintes: Adaline, Redes Neurais Convolucionais, Perceptron camada múltipla (Figura 7), Camada recorrente, Estrutura reticulada, Long, Short-Term Memory (LSTM), Redes de Hopfield, entre tantas outras arquiteturas que existem e outras que ainda se encontram em fase de desenvolvimento. Aqui será dada uma atenção às redes do tipo Perceptron camada múltipla – também denominadas de *Multilayer Perceptrons* (MLP). Esse tipo de rede possui em sua composição uma camada de entrada, uma ou mais camadas ocultas e uma de saída; sua característica principal é a utilização do algoritmo de retropropagação (*backpropagation*) para o cálculo do erro, utilizando treinamento do tipo supervisionado, que permite o monitoramento e o ajuste de pesos de acordo com o objetivo da rede proposta.

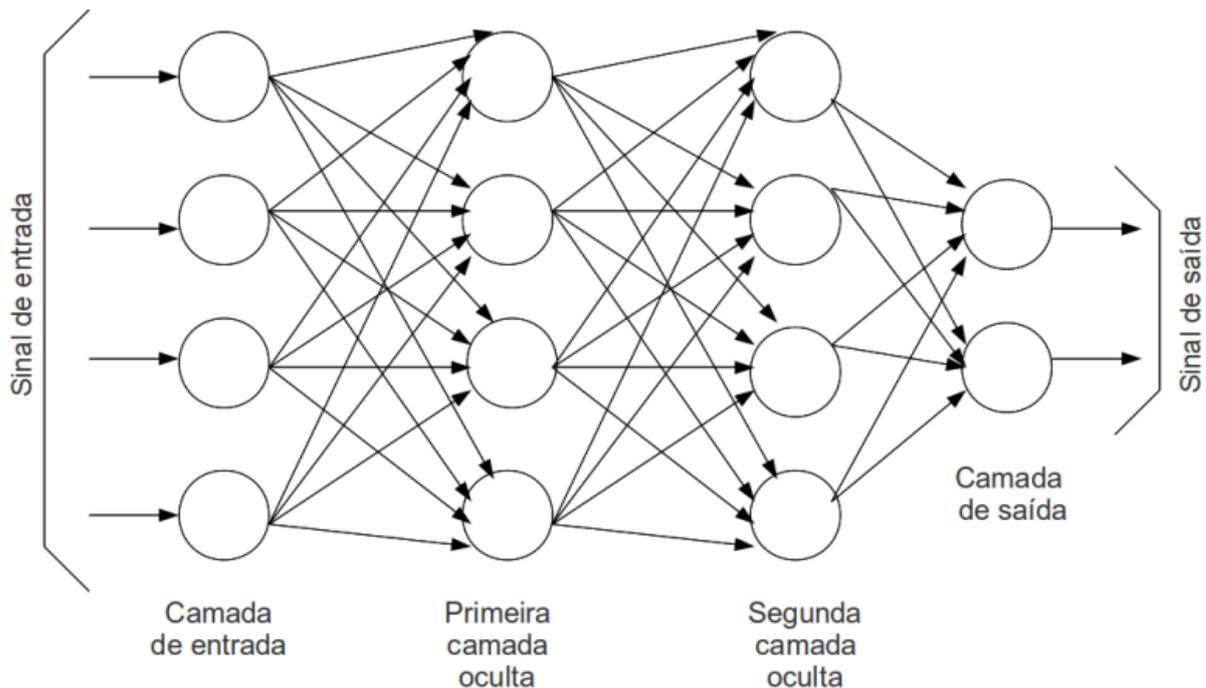


FIGURA 7: Estrutura de uma rede MLP com duas camadas ocultas

Fonte: Volpi (2018).

O processo de aprendizagem desse tipo de rede é de fundamental importância para seu funcionamento, particularmente na realização dos cálculos a que essa se aplica. Segundo Ferreira:

A aprendizagem por retropropagação do erro consiste de dois passos: um passo para frente, a propagação, e um passo para trás, a retropropagação. No passo para frente, um padrão de entrada (vetor de entrada) é aplicado aos

nodos e seu efeito se propaga através da rede camada por camada. Finalmente um conjunto de saídas é produzido com a resposta real da rede. Durante o passo de propagação, os pesos da rede são todos fixos. Durante o passo para trás, os pesos são todos ajustados de acordo com uma regra de correção de erros. Especificamente, a resposta real da rede é subtraída de uma resposta desejada (alvo) para produzir um sinal de erro. Esse sinal de erro é então propagado para trás através da rede, contra a direção das conexões sinápticas. Os pesos são ajustados de forma que a resposta real da rede se mova para mais perto da resposta desejada. (FERREIRA, 2004, p. 26)

É de grande importância destacar que a escolha da arquitetura de cada rede neural está condicionada a sua aplicação. Sendo assim, essa escolha parte do conhecimento prévio do pesquisador, com base na revisão de trabalhos sobre suas diversas aplicações e exemplos práticos de sua aplicação, perseguindo o objetivo a ser alcançado.

As aplicações desse tipo de arquitetura de rede neural são encontradas nas mais diversas aplicações e área de conhecimento, e sua implementação possui grande destaque no cenário da inteligência artificial. Alguns desses podem ser encontrados nos trabalhos recentes desenvolvidos por Bayat *et al.* (2018) – da realização de sua aplicação no processamento computacional em detrimento de microprocessadores convencionais; Heidari *et al.* (2018) utilizaram as redes MLP para a proposição de algoritmos de otimização de busca flexíveis e adaptáveis; Ghorbani *et al.* (2018) utilizaram esta arquitetura na implementação de um modelo híbrido de predição de evaporação no Norte do Irã, entre outras que aqui não foram citadas. Esses exemplos só consubstanciam a importância da utilização da RNA, notadamente quando da necessidade de processamento de informações, predição e automação de tarefas.

No que concerne à aplicação das redes neurais artificiais, segundo Maier *et al.* (2010), nos últimos 15 anos foi maciça a utilização dessa ferramenta na elaboração de modelos de previsão e predição na área de recursos hídricos e outras áreas afins. A partir de 1992 o uso das RNA foi se tornando cada vez mais comum na área ambiental e um grande investimento foi realizado na busca de melhoria dos processos já existente e na implementação de outras metodologias a serem aplicadas. Segundo Wu *et al.* (2014), entre os anos 2000 e 2012 foram publicados 618 artigos em língua inglesa sobre o assunto em 307 periódicos.

2.3 Modelagem de predição de desempenho de ETA

Como ferramenta no auxílio à tomada de decisão, os modelos de predição são admitidos como importante aliados, tendo em vista sua capacidade de predição, antevendo situações e garantindo maior eficiência frente às situações que requerem respostas com certa urgência, como ocorre em sistemas de tratamento após catástrofe, por exemplo. Essa e outras aplicações fazem com que o uso de modelos de predição e predição seja cada vez mais requerido na busca de uma abordagem mais científica na gestão de sistemas de grande interesse, como as estações de tratamento de água para abastecimento humano e de tratamento de águas residuárias.

A modelagem de processos e sistemas já está bastante consolidada, seja no meio científico ou na sua aplicação nas diversas atividades industriais, financeiras, entre outras. Sendo assim, uma de suas aplicações é a predição de resultados e/ou comportamentos de certos parâmetros ou variáveis, sendo capazes de estimar valores a partir da simulação de determinados acontecimentos. Modelos assim concebíveis são denominados de Modelos de Simulação ou Modelos de Previsão/Predição, capazes de apresentar prognósticos desses sistemas a partir de cálculos estatísticos-matemáticos.

As bases de modelos de predição propostos pela literatura são diversas, assim também são as suas aplicações nas mais diferentes áreas do conhecimento. Segundo Raikar *et al.* (2016), as ferramentas computacionais muito têm contribuído para tais aplicações, graças à robustez, à precisão e ao baixo custo de solução. Na prática, algumas ferramentas computacionais simulam o cérebro humano na busca de resultado lógico-preciso.

Dentre as bases para proposição dos modelos de predição, podem-se destacar os baseados em estatística (regressão simples e múltiplas), Lógica Fuzzy, algoritmo genético, RNA e sistemas híbridos (utilização de mais de uma base). No entanto, considerando suas aplicações, as RNA estão se tornando cada vez mais populares no meio científico e nas aplicações práticas, particularmente pela sua robustez e diversidade de aplicação.

Na área do saneamento ambiental são diversas as aplicações de modelos de predição baseados em RNA. Para tanto, pode-se destacar o trabalho de Nasr *et al.* (2012), que utilizou esse tipo de modelagem na predição do desempenho de uma estação de tratamento de esgoto em El-Agamy (Egito); nesse trabalho foi obtido um coeficiente de correlação de 0,90 entre os

Programa de Pós-graduação em Saneamento, Meio Ambiente e Recursos Hídricos da UFMG

valores da predição e os valores reais de demanda biológica de oxigênio - DBO, demanda química de oxigênio - DQO e sólidos suspensos totais - SST para desempenho da estação de tratamento em estudo, sendo possível admitir que o modelo apresentou aceitável capacidade de generalização e precisão.

Outra importante contribuição foi o trabalho semelhante desenvolvido em uma estação de tratamento de esgoto da cidade de Konya, na Turquia. Nesse trabalho foram aplicados modelos de predição de desempenho da estação de tratamento utilizando RNA. Como resultado, foi verificado que a função de transferência que apresentou melhor resultado foi a *Tansig*, com coeficiente de correlação (r) acima de 0,99 entre os resultados do modelo e os resultados reais, sendo possível concluir que o modelo foi sofisticado e eficiente para a aplicação e que as RNAs foram efetivas na predição do desempenho da estação em estudo (TÜMER e EDEBALI, 2015).

Em estações de tratamento de água de abastecimento humano, mencionam-se os trabalhos desenvolvidos por Rak (2013) e Wu e Lo (2008). No primeiro, foi realizada a modelagem com o uso de RNA na predição da turbidez da água tratada durante o processo de tratamento de água do sistema de Sosnówka – Polônia, apresentando resultados favoráveis para a aplicação desta ferramenta. Já no segundo foi utilizada a RNA, e RNA com aprendizagem baseada na Lógica Fuzzy (ANFIS – *Adaptative network based fuzzy inference system*), na predição da dosagem de coagulante, sendo utilizados os parâmetros turbidez, pH, cor aparente e temperatura do afluente à ETA, em Taipei – Taiwan, da água pós-floculação, da água decantada e da água tratada. Nessa pesquisa chegou-se à conclusão que o modelo apresentou uma predição relativamente boa, particularmente para a dosagem de cloreto de polialumínio em tempo real. O modelo em RNA apresentou-se melhor que o ANFIS, resultando em valores elevados de R^2 e baixos valores de RMSE (raiz quadrada do erro médio).

A literatura também apresenta diversos trabalhos utilizando ferramentas baseadas em modelagem computacional em estações de tratamento de águas, com destaque para o trabalho de Kulkarni e Chellam (2010), que utilizaram RNA para prever a geração de subprodutos da desinfecção por cloro, e Sadiq *et al.* (2010), que, por meio da Lógica Fuzzy, desenvolveram índices (a partir de uma série de parâmetros de qualidade da água tratada, tais como cloro residual livre, trihalometanos, indicadores microbiológicos, entre outros; além de aspectos operacionais, fatores humanos, potencial poluidor, características do sistema de distribuição,

qualidade da água bruta, entre outros) para avaliação da performance para dez pequenas estações de tratamento de água.

As aplicações supracitadas e outras que podem ser consultadas na literatura corroboram com a aplicação e atestam a necessidade do uso dos modelos de predição para as diversas atividades desenvolvidas no âmbito do saneamento ambiental, notadamente a aplicabilidade das RNAs, por serem passíveis de grande flexibilidade, generalização e apresentarem resultados precisos nos diversos sistemas estudados.

3. MATERIAL E MÉTODOS

A presente pesquisa é fruto de uma série de estudos que já vêm sendo desenvolvidos desde 2005 no Programa de Pós-Graduação em Saneamento, Meio Ambiente e Recursos Hídricos da UFMG, iniciada por Lopes (2005). Nesse trabalho, foi proposto um Índice de Qualidade de Estações de Tratamento de Água - IQETA, tendo por fundamento a proposição do Índice de Qualidade da Água (IQA), utilizando-se a metodologia Delphi aplicada a estações convencionais de tratamento. Posteriormente, trabalhos semelhantes foram realizados, com vistas à otimização dos processos e dos métodos utilizados na formação desses índices, bem como na busca de modelos de predição de desempenho de estações de tratamento de água, conforme desenvolvido por Oliveira *et al.* (2014).

Na busca da melhoria dos modelos de predição desenvolvidos (baseado em Lógica Fuzzy), esta pesquisa apresenta a proposição de modelos de predição, por meio das Redes Neurais Artificiais. O uso dessa metodologia se deve ao fato de sua capacidade de otimização e generalização dos processos, além da possibilidade de trabalhar com sistemas não lineares e sua capacidade de aprendizado para a tomada de decisão em situações complexas. Para efetivação da referida proposta, a metodologia está dividida em três fases, conforme ilustrado na Figura 8.

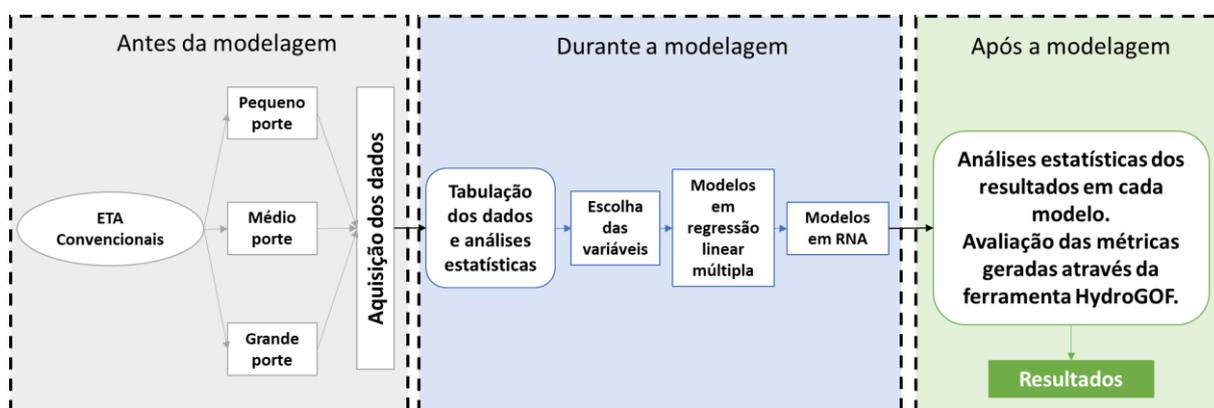


FIGURA 8: Etapas metodológicas da pesquisa

3.1 Definição do objeto de estudo

Para a realização desta pesquisa foram consideradas estações convencionais de tratamento de água com etapas de coagulação, floculação, decantação, filtração rápida e desinfecção. Essa

escolha se deve ao fato desse tipo de tratamento ser o mais utilizado na maioria das ETAs de todo o país.

A partir da adoção do tipo de tratamento, foi realizada a divisão das estações em três categorias, pequeno ($Q < 50 \text{ L/s}$), médio ($50 \text{ L/s} \leq Q \leq 500 \text{ L/s}$) e grande ($Q > 500 \text{ L/s}$) porte; essa classificação foi feita a partir da vazão mínima afluente a cada ETA e com base na população abastecida. Para adoção desses valores foi tomado por base o que é preconizado pela Deliberação Normativa COPAM nº 257/2017, que determina que as ETAs sejam divididas em pequeno ($20 \text{ L.s}^{-1} < \text{vazão de água tratada} < 100 \text{ L.s}^{-1}$), médio ($100 \text{ L.s}^{-1} \leq \text{vazão de água tratada} \leq 500 \text{ L.s}^{-1}$) e grande porte ($\text{vazão de água tratada} > 500 \text{ L.s}^{-1}$) (COPAM, 2017). No entanto, optou-se por diminuir a vazão para ETAs de pequeno porte, tendo em vista que vazões de até 50 L.s^{-1} podem abastecer cidades de 20 mil habitantes, que segundo o IBGE (2019) representam 68,4% dos municípios brasileiros. A classificação se faz necessária face à proposição de modelos mais realísticos, tendo em vista que algumas variáveis contempladas nos modelos podem apresentar pesos diferentes de acordo com o porte da unidade. Um exemplo desse fato é a variável “grau de instrução do operador”, que em ETAs de pequeno porte, comumente, é inferior à das de grande porte. Apesar da subjetividade, essa variável foi mantida baseando-se nas pesquisas anteriormente citadas e, por se tratar de uma pesquisa científica, analisar sua influência no desenvolvimento dos modelos propostos.

Nesse sentido, foram escolhidas 12 estações de tratamento de água de duas concessionárias da Região Sudeste do Brasil. Em todos os casos foram realizados acordos de cooperação técnica, entre as partes interessadas, para formalização da pesquisa.

Os parâmetros utilizados foram escolhidos de acordo com o preconizado pelo Anexo XX da Portaria de Consolidação nº5/2017 do Ministério da Saúde, referente à etapa de controle da qualidade da água realizada pelas concessionárias responsáveis pelas ETAs em estudo (BRASIL, 2017). Também foram utilizados parâmetros que compõem o IQETA, levando em consideração o que recomenda a NBR 12.216/1992 (ABNT, 1992), seguindo a metodologia utilizada por Oliveira (2014) para a elaboração do modelo de predição baseado na Lógica Fuzzy, e outros adotados nesta pesquisa, quais sejam:

- Qualidade de água bruta: turbidez, cor aparente e pH;
- Parâmetros hidráulicos: vazão afluente, gradiente de velocidade de floculação, tempo de floculação, velocidade de sedimentação, velocidade longitudinal de escoamento nos

decantadores de escoamento horizontal, vazão linear da calha de coleta de água decantada e taxa média de filtração;

- Parâmetros operacionais: tempo diário de operação da ETA, grau de instrução do operador, realização rotineira de ensaio em *Jar Test*, dosagem de coagulante (é importante destacar que, na prática, essa informação deve estar relacionada ao pH de coagulação);
- Qualidade de água na ETA: turbidez da água decantada;
- Qualidade de água tratada: turbidez e cor aparente (com exceção das ETAs de grande porte); e
- Produção de resíduos gerados no tratamento da água: estimativa de produção de lodo.

O gradiente de velocidade e o tempo de mistura rápida não entraram nessa pesquisa devido à falta de dados para seu cálculo em determinadas estações, e pela existência de sobrecarga em algumas delas.

No cálculo dos parâmetros hidráulicos, com exceção da vazão afluente, para a composição dos modelos foram utilizadas equações de projeto conforme Lopes (2005) e com base na NBR 12.216/1992 (ABNT, 1992). Já para as variáveis relativas ao “grau de instrução do operador” e “ensaio em *Jar Test*” foram atribuídos valores numéricos de 1 a 4¹ para o primeiro, e 1 ou 2 para o segundo (1 para sim e 2 para não).

Para estimar a produção de lodo de cada ETA foi utilizada a Equação 3.1.1, proposta por Cornwell (2006):

$$W = 0,0864 * Q * (0,44 * D + 1,5 * T + A) \quad (3.1.1)$$

na qual:

W - Massa de lodo produzida por dia (kg.dia⁻¹);

Q - Vazão de água bruta (L.s⁻¹);

D - Dosagem de sulfato de alumínio (mg.L⁻¹);

T - Turbidez da água bruta (uT); e

A - Dosagem de auxiliares ou outros produtos adicionados (mg.L⁻¹).

¹ Iniciando com curso técnico ou superior em exatas, técnico ou superior em outra área, ensino médio completo e ensino fundamental completo, em ordem decrescente.

O levantamento desses parâmetros para o modelo se justifica pelo fato de estarem diretamente ligados à produção de água potável e de reunirem um conjunto de informações relevantes acerca da qualidade, quantidade e, por conseguinte, da segurança da água produzida em cada ETA para a qual o modelo poderá ser implementado.

Outra variável de interesse na produção de água potável, que poderia ser utilizada na elaboração desses modelos, é o consumo de energia elétrica em cada estação de tratamento. No entanto, há uma notória dificuldade na obtenção do valor que represente verdadeiramente esse consumo no tratamento, tendo em vista que, frequentemente, as ligações contemplam outros dispositivos e equipamentos, por vezes vinculados à adução de água bruta e/ou tratada.

Na escolha desses parâmetros foi considerado o quantitativo de dados, bem como a frequência de análise e o tempo de monitoramento. Esses aspectos são de fundamental importância para a realização das análises estatísticas e para que se possam alcançar resultados mais realísticos a partir da proposição dos modelos. Para a efetivação desta etapa foram realizadas análises de consistência dos dados.

Os dados médios diários de monitoramento das ETAs foram adquiridos por solicitações realizadas e mediante visita técnica em cada estação para o preenchimento de formulários padrões estabelecidos pela companhia de saneamento e pelo acordo de cooperação entre as entidades envolvidas nesta pesquisa. As variáveis de qualidade de água foram enviadas por arquivos digitais em planilhas eletrônicas. Os dados utilizados nesta pesquisa são referentes aos anos de 2009 a 2014, inclusive; em quatro ETAs, os dados utilizados são relativos aos anos de 2003 a 2011, devido à dificuldade de atualização desses.

A aquisição dos parâmetros hidráulicos foi realizada mediante cálculo das mesmas, com base em equações de projetos das unidades que compõem o sistema. Nesses cálculos foi levada em consideração os tipos de unidades, suas dimensões, a vazão afluente de cada ETA, algumas constantes hidráulicas citadas na literatura técnica, entre outros.

Ainda com relação à utilização dos dados, é importante destacar que nas ETAs de grande porte não foi possível o desenvolvimento dos modelos com base no parâmetro cor aparente da água tratada. Esse fato ocorreu devido à falta de informação desse parâmetro na base de dados adquirida. Portanto, as comparações dos modelos entre os portes das estações de tratamento de água só foram possíveis a partir do parâmetro turbidez da água tratada.

3.1.1 Caracterização das estações de tratamento de água

Conforme mencionado anteriormente, nesta pesquisa foram utilizados dados de 12 estações de tratamento de água escolhidas com base no tipo de tratamento utilizado, número de dados disponíveis, facilidade na aquisição e no acesso às mesmas.

A Tabela 1 ilustra algumas características básicas das estações de tratamento de água utilizadas nessa pesquisa. Dentre as ETAs apresentadas, três estão operando com sua capacidade excedida, que pode ensejar na diminuição da qualidade da água produzida, bem como no pior desempenho delas; essas estações são caracterizadas como de pequeno porte e sua sobrecarga pode estar relacionada ao aumento populacional e à falta de ampliação dessas unidades, que não acompanharam tal crescimento. Com relação ao tempo diário de operação, todas as estações realizam operação ininterrupta.

Como forma de manter a ética e o sigilo da pesquisa, optou-se por utilizar nomenclaturas gerais ao invés dos nomes dos municípios, mananciais utilizados no abastecimento e nomes das estações de tratamento de água, que em sua grande maioria são homônimas aos municípios onde estão presentes ou responsáveis pelo tratamento da água ofertada aos seus habitantes. O afluente a todas as estações advém de um ou mais cursos d'água por meio de captação direta sem reservatório de acumulação.

TABELA 1: Descrição das estações de tratamento de água utilizadas para elaboração dos modelos

Município	Manancial	Estação de tratamento	Vazão de projeto (L.s ⁻¹)	Vazão média afluyente (L.s ⁻¹)	Porte
M1	Rio 1	ETA1	36	39	Pequeno
M2	Rio 2	ETA2	40	28	
M3	Rio 3	ETA3	11	47	
M4	Rio 4	ETA4	26	37	
M1	Rio 1	ETA1	110	65	Médio
M2	Rio 2 Rio 3 Rio 4	ETA2	440	355	
M3	Rio 5	ETA3	130	92	
M1	Rio 1	ETA1	3.300	2.757	Grande
	Rio 2	ETA2	750	658	
RM1	Rio 3	ETA3	2.700	2.280	
	Rio 4	ETA4	8.000	6.200	
M2	Rio 1	ETA5	1.000	747	

Fonte: Concessionárias responsáveis pelas estações de tratamento.

Verificou-se que algumas dessas estações passaram por algum tipo de reforma e/ou ampliação, fato que contribuiu para que elas mantivessem a qualidade do serviço prestado, embora as vazões outorgadas pelos órgãos reguladores de alguns desses mananciais já estão se igualando à vazão de demanda destes municípios.

Com relação as ETAs de pequeno porte, todas elas estão localizadas no interior em cidades com população inferior a 30 mil habitantes. Algumas dessas estações já passaram por algum tipo de reforma, no entanto, algumas informações não foram atualizadas, por exemplo a ETA 3. No entanto, verifica-se que as concessionárias mantêm todo o controle da operação desses sistemas e sua boa operacionalidade.

3.2 Desenvolvimento dos modelos de predição

3.2.1 Tabulação dos dados e testes estatísticos básicos

Nesta fase, após a aquisição dos dados e da digitação dos valores calculados em planilhas Excel®, eles foram separados por porte de ETA, ou seja, foi criada uma planilha para cada classificação de estação, de modo a facilitar a manipulação dos dados e sua aplicação nas etapas futuras. Posteriormente, iniciou-se a organização dos dados, a começar pelas estações de pequeno porte; nesta fase os dados foram dispostos por data, na primeira coluna, seguido pelos respectivos parâmetros (um total de 18):

1. Vazão afluente à ETA ($L.s^{-1}$) – **Qaf**;
2. Tempo diário de operação da ETA (h) – **TempOp**;
3. Cor aparente da água bruta (uC) – **CorB**;
4. Turbidez da água bruta (uT) – **TuB**;
5. Potencial hidrogeniônico da água bruta – **pHB**;
6. Turbidez da água decantada (uT) – **TuD**;
7. Dosagem de coagulante ($mg.L^{-1}$) – **DosCoag**;
8. Produção de lodo gerado na ETA ($kg.dia^{-1}$) – **ProdLodo**;
9. Grau de instrução do operador da ETA – **InstOp**;
10. Realização de ensaio de *Jar Test* – **EnsJar**;
11. Gradiente de velocidade de floculação (s^{-1}) – **GradFloc**;
12. Tempo de floculação (s) – **TemFloc**;
13. Velocidade de sedimentação na decantação ($cm.min^{-1}$) – **VelSed**;
14. Velocidade longitudinal de escoamento na decantação ($cm.s^{-1}$) – **VelLong**;
15. Vazão linear na calha de coleta de água decantada ($L.s^{-1}.m^{-1}$) – **Qcalha**;
16. Taxa de filtração ($m^3.m^{-2}.dia^{-1}$) – **TxFilt**;
17. Turbidez da água tratada (uT) – **TuT**; e
18. Cor aparente da água tratada (uC) – **CorT**.

As análises preliminares (verificação de dados censurados, ocorrência de *outliers*, estatística descritiva ou exploratória) foram realizadas, tendo em vista que essas são de grande importância para que se tenha um conhecimento inicial do comportamento dos dados e suas influências na proposição dos modelos.

Dentre as análises estatísticas preliminares, pode-se destacar a verificação de dados (erros de medição, valores fora da escala de medida – pH > 14, dentre outros), na qual foi observado se os dados adquiridos estavam de acordo com os comumente encontrados no monitoramento de sistemas similares e/ou com os parâmetros de projeto de ETAs convencionais, bem como valores faltosos. Essa etapa foi de fundamental importância para que não ocorressem resultados imprecisos na aplicação dos métodos estatísticos, bem como nos resultados apresentados a partir dos modelos de regressão linear múltipla (*baseline*) e RNA.

Quando os dados censurados e/ou lacunas na planilha foram detectados, optou-se pela exclusão de toda a linha de dado; fato similar também ocorreu para as variáveis faltosas por períodos e/ou não monitoradas.

A investigação de *outliers* também se reveste de grande importância quando da necessidade de identificar números discrepantes dentro dos conjuntos de dados de todas as variáveis analisadas. Para tanto, nesta pesquisa foi utilizado o método exploratório dos quartis e amplitude interquartis, conforme descrito por Triola (2013). Após a constatação desses valores, foi analisada sua consistência, de acordo com os erros de medição ou anotações que divergem daqueles comumente encontrados na literatura, ou fora de valores estabelecidos por norma padrão. Os casos não consistentes foram excluídos do conjunto de dados, e os demais foram mantidos.

Em seguida, foi realizada a junção dos dados das ETAs, formando-se um conjunto de dados para cada porte. Como tentativa de generalização dos modelos propostos, foi aplicada a aleatorização dos mesmos, linha a linha (trata-se de um processo de embaralhamento dos dados). Para a efetivação da etapa de aleatorização, foi gerada uma coluna com números aleatórios maiores que 0 e menores que 1; em seguida as linhas foram classificadas de acordo com os valores da coluna gerada, optando-se por iniciar por ordem crescente.

Como forma de obter um conhecimento preliminar dos dados, foram estimadas as estatísticas descritivas amostrais, quais sejam: medida de tendência central (média), sua dispersão (desvio padrão) máximo, mínimo, entre outros. A partir de seus resultados foram elaborados tabelas e gráficos do tipo *boxplot*, para melhor apresentação do comportamento deles.

Posteriormente, foi aplicado o teste de normalidade (Shapiro-Wilk), ao nível de significância de 5% ($\alpha = 0,05$), com vistas a verificar se a distribuição de probabilidade do conjunto de

dados se aproxima da distribuição normal. Após a referida etapa foram aplicados análise de correlação de Pearson ($\alpha = 0,05$); os métodos a serem utilizados foram adotados conforme a distribuição dos dados (paramétricos ou não-paramétricos). Em todos os testes estatísticos foi utilizado o software Minitab® 18.

O teste de correlação teve por objetivo definir as variáveis mais explicativas para os dois parâmetros de resposta dos modelos (turbidez e cor aparente da água tratada), com base nos valores do coeficiente de correlação de Pearson (r) e no *p-valor*. Em dois casos, em particular, em que a análise de correlação não foi suficiente para a escolha das variáveis, foi aplicado um método de transformação (foi aplicado o logaritmo na base 10 em todas as observações) das mesmas, visando obter maior valor do coeficiente de determinação nos modelos de regressão linear múltipla.

Com os resultados dos referidos testes, foram escolhidas as variáveis para a elaboração dos modelos em regressão linear múltipla (*baseline* da pesquisa). A partir de seus resultados, elas foram utilizadas para os modelos em RNA, objeto deste estudo.

3.2.2 Modelos de regressão linear múltipla (*baseline*) e escolha das variáveis

Para o desenvolvimento dos modelos baseados em regressão linear múltipla, foram utilizadas as mesmas componentes escolhidas na elaboração dos modelos em RNA. Tal metodologia se deve ao fato da necessidade de comparação entre os modelos, a partir das mesmas condições de implementação, ou seja, mesmas variáveis (dependentes e independentes) e mesma quantidade de dados utilizados.

Esse tipo de modelagem (mecanicista), diferente da modelagem computacional, é bem mais difundida no meio científico e possui grande aplicabilidade nas diversas atividades industriais, agropecuária, educacional e financeira, que envolvem estudos econômicos, produtivos, biomédicos, dentre outras, buscando prever resultados, inferir e selecionar determinadas variáveis. A regressão linear múltipla relaciona diversas variáveis de entrada (independentes), a partir de uma variável de saída (dependente), para gerar uma função linear representativa de determinado fenômeno. Apresenta sua capacidade de predição, de acordo com o coeficiente de determinação gerado (R^2) e com as características das variáveis que compõem o modelo. Um modelo em regressão linear múltipla pode ser expresso da seguinte maneira:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon,$$

No qual:

Y – Variável dependente a ser prevista;

X_i – Variáveis independentes;

β_0 – Constante;

β_i – Coeficientes parciais de regressão (parâmetros do modelo); e

ε – Erro ou perturbação.

Antecipadamente à referida etapa, verificou-se o atendimento das variáveis às condições básicas de aplicação da regressão linear, quais sejam, ausência de multicolinearidade entre as variáveis independentes do modelo, presença de correlação estatisticamente significativa entre as variáveis independentes e as variáveis dependentes (cor aparente e turbidez da água tratada), e homocedasticidade na distribuição do conjunto dos resíduos (TRIOLA, 2013).

3.2.3 Modelos baseados em redes neurais artificiais

Inicialmente, foi elaborado o modelo baseado na resposta da turbidez da água tratada, tendo em vista que esse parâmetro possui grande relevância quando da segurança da qualidade da água, conforme já citado pelo Anexo XX da Portaria de Consolidação nº5/2017; e com base nos resultados desse, a elaboração do segundo modelo (cor aparente) foi iniciada.

Para a elaboração dos modelos foi utilizado um quantitativo de 17.784 observações para cada parâmetro que compõe os modelos, sendo realizada uma divisão de 70/30, ou seja, 70% dos dados foram utilizados para calibração e 30% para a simulação e validação dos modelos propostos para cada porte de ETA.

Os modelos desenvolvidos nesta pesquisa (dois modelos para cada porte de ETA, com exceção das ETAs de grande porte) têm como função básica realizar o processamento de cálculos entre os valores dos parâmetros adotados e compará-los com os valores de saída, de modo a gerar valores de erros e pesos para cada uma das variáveis de entrada dos modelos. Esses últimos serão de fundamental importância na etapa de simulação.

Para facilitar o entendimento e roteiro metodológico foi realizada a seguinte divisão entre os modelos propostos:

- ETAs de pequeno porte – Modelo 1 (turbidez) e Modelo 2 (cor aparente);
- ETAs de médio porte – Modelo 3 (turbidez) e Modelo 4 (cor aparente); e
- ETAs de grande porte – Modelo 5 (turbidez);

A modelagem tem por finalidade relacionar às diversas variáveis que envolvem o funcionamento das estações de tratamento de água, atribuindo pesos a cada uma delas, e gerando como resultado valores (estimados ou calculados) de turbidez e cor aparente da água tratada (variáveis resposta), escolhidos para indicar o desempenho das ETAs, por porte. Com o desenvolvimento do modelo é possível prever o desempenho das estações de tratamento de água, tendo como parâmetros de interesse aqueles supracitados. Além de gerar os valores preditos, também é objetivo dessa modelagem gerar intervalos de predição, facilitando os tomadores de decisão com base em intervalos de 95% de confiança.

Os modelos foram desenvolvidos com o uso do software R[®]. Nesta etapa, utilizou-se a ferramenta *RStudio* versão 3.5 (Figura 9). Essa ferramenta é um console para o desenvolvimento e implementação de diversos tipos de modelagem, estudos estatísticos, *machine learning*, entre outras aplicações; fazendo uso da linguagem R e oferecendo um ambiente interativo interessante, a partir de sua concepção GUI (*Graphical User Interface*). Através dessa interface é possível que o usuário possa interagir com os sistemas de maneira mais clara, simples e objetiva, utilizando ícones, por exemplo.

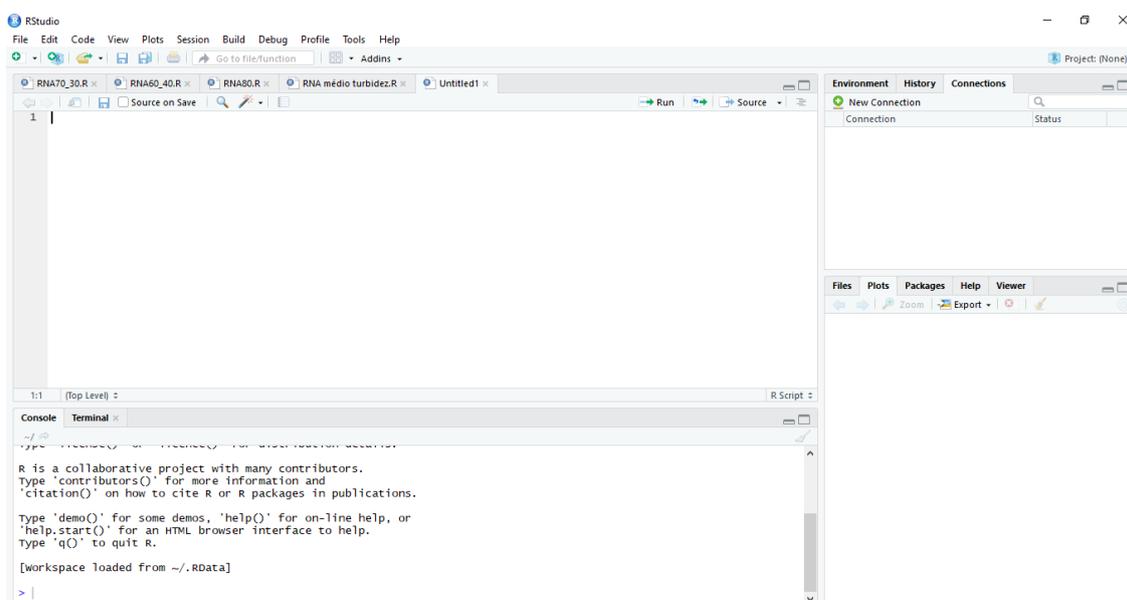


FIGURA 9: Tela inicial da ferramenta utilizada para o desenvolvimento dos modelos em RNA

No *RStudio* foram implementados os códigos de programação para as rotinas de comando necessários no processamento da RNA e na geração dos intervalos de predição, bem como no cálculo das métricas utilizadas na avaliação da qualidade da resposta do modelo. Nesta etapa da pesquisa foram atribuídos os pontos de contorno, tais como topologia e arquitetura da rede, número de neurônios na camada oculta, número de camadas, algoritmo de treinamento, função de ativação na camada oculta e na camada de saída, função de adaptação de aprendizagem, função para o cálculo da performance e percentual de dados utilizados na etapa de aprendizagem da rede (treinamento e teste).

Iniciando a etapa de desenvolvimento dos modelos, foram utilizados parâmetros de treinamento comumente citados na literatura técnica sobre o assunto. Posteriormente, as avaliações foram aplicadas na etapa de treinamento da rede. Nessa etapa, a rede neural foi submetida ao processo de aprendizagem, ou seja, assim como o cérebro humano, a rede teve que “aprender a pensar” para que suas decisões fossem tomadas de forma coerente e precisa; nesse caso, as decisões foram os resultados das saídas com base nos valores dos pesos. Após esse processo, foi possível mensurar o desempenho da RNA em resposta às expectativas do modelo para que se possa realizar sua aplicação.

Na utilização do *RStudio* para tal finalidade, uma das etapas de grande importância é adoção das bibliotecas, nas quais estão contidos os pacotes necessários para a aplicação dos vários métodos de modelagem aplicados. Para a realização dessa pesquisa foram utilizados os pacotes *neuralnet* (necessário para todas as etapas de implementação e funcionamento da rede neural) (FRITSCH *et al.*, 2019), *hydroGOF* (utilizado na aplicação de algumas funções e métricas estatísticas) (BIGIARINI, 2017) e o *nnetpredint* (responsável por toda a criação dos intervalos de predição a partir da saída da rede neural) (LEEPER e GANZ, 2019). Todos os pacotes aqui utilizados, e outros que fazem interface com o *RStudio* ou o próprio R, nas diversas plataformas de sistemas operacionais, podem ser instalados a partir do repositório padrão CRAN (*Comprehensive R Archive Network*).

O *package* *nnetpredint* tem por finalidade criar, a partir de um conjunto de valores preditos/simulados/calculados, intervalos, no qual uma observação futura se encontre, dado o modelo estimado. Esses intervalos são gerados a partir de um nível de confiança. Como forma de exemplificar esse método, pode-se destacar o trabalho desenvolvido por Kasiviswanathan *et al.* (2013), que propuseram o uso de intervalos de predição, utilizando RNA, em modelos

de estimativa de *runoff*, apresentando resultados de 97,17% dos pontos observados dentro do intervalo de predição, chegando-se à conclusão que essa metodologia é mais acurada do que o método tradicional, baseando-se apenas num conjunto único de valores de predição.

No desenvolvimento desta pesquisa foi iniciada a elaboração do modelo para a turbidez da água tratada, tendo como arquitetura inicial a rede do tipo *Feed-Forward Backpropagation*, com função de treinamento *Levenberg-Marquardt (trainln)*, cálculo da performance pela soma dos quadrados do erro (*SSE*), duas camadas ocultas com oito neurônios cada e função de transferência na camada oculta do tipo logística ou logarítmica, não sendo utilizada função de transferência na camada de saída. Essas configurações foram adotadas por serem as mais comumente utilizadas em modelagens semelhantes às desenvolvidas (QADERI e BABANEZHAD, 2017; GIWA *et al.*, 2016; FAN *et al.*, 2018; CHEN *et al.*, 2018), bem como pelos resultados dos coeficientes de determinação alcançados pelos sucessivos treinamentos realizados mediante modificação da arquitetura e topologia da RNA.

Na elaboração dos Modelos 2 e 4 (saídas com base na cor aparente da água tratada), foi aplicada a última topologia da rede neural artificial utilizada para os Modelos 1 e 3, tomando-se como métrica principal o coeficiente de determinação.

3.3 Análise dos modelos propostos

Conforme mencionado, foram gerados 10 modelos de predição de desempenho de estações de tratamento de água para cada método de modelagem (cinco em regressão linear múltipla – utilizadas como o modelo-base, e cinco em RNA com geração de intervalos de predição) para os três portes das unidades. Para tanto, foi necessária a avaliação entre os tipos de modelos (ferramenta aplicada para o desenvolvimento dos modelos), bem como entre os portes das ETAs. Essa análise foi conduzida com base nas métricas estatísticas da relação entre os valores estimados e observados dos parâmetros turbidez e cor aparente da água tratada.

Inicialmente, foi realizada a análise das métricas geradas pelos modelos para verificação de diferença entre os seus resultados, sendo um para cada porte de estação, tendo por objetivo averiguar se o porte poderia causar alguma influência no modelo e quão generalista é o modelo; a realização dessa etapa ocorreu como forma de poder aplicar os referidos modelos em outras estações de mesma tecnologia de tratamento. Essa análise foi realizada tanto com os modelos propostos em RNA, como aqueles em regressão linear múltipla, separadamente.

Posteriormente, realizou-se a avaliação entre os modelos em RNA e regressão. Nessa etapa, os resultados dos modelos foram para cada porte de ETA. O objetivo nesta etapa da pesquisa foi verificar se há diferença entre os valores estimados dentre os modelos e qual o modelo apresenta melhor resultado. Assim, ele torna-se passível de ser aplicado, quando da necessidade de antever situações de fragilidade nas ETAs, e auxiliar os gestores na tomada de decisão de forma mais precisa e acurada. O objetivo dessa etapa metodológica também foi medir o ganho de um modelo em detrimento do outro, a partir dos coeficientes de determinação calculados entre os valores dos parâmetros de controle observados e preditos.

Para a execução dessas etapas, foram utilizadas as métricas geradas a partir da execução de cada um dos tipos de modelo. Por exemplo, durante o desenvolvimento dos modelos em regressão linear múltipla foi gerado como resultado os valores dos coeficientes de cada variável, o erro de cada um desses, seus intervalos de predição, o valores F e P, de fundamental importância para verificação da significância de cada uma das variáveis que compõem o modelo; o Índice de Inflação da Variância – VIF, que irá analisar a existência ou não de multicolinearidade, com vistas a possibilitar a utilização do modelo; além dos coeficientes de determinação, o R^2 ajustado, que irá indicar quão significativo é o ajuste do modelo. Já no desenvolvimento dos modelos em RNA, através da ferramenta HydroGOF, foram gerados uma série de métricas capazes de avaliar o desempenho dos modelos.

Como forma de avaliar a qualidade das saídas relativas ao parâmetro turbidez da água tratada (Modelos 1, 3 e 5), foram calculadas algumas métricas, presentes no *package* HydroGOF (BIGIARINI, 2017), entre os valores gerados pelos modelos (turbidez calculada) e os valores observados (turbidez observada) para cada porte de ETA. Dessa relação, avaliaram-se os valores dos coeficientes de determinação (R^2), buscando-se o modelo que apresentasse os maiores valores de R^2 . Posteriormente, passou-se para etapa de implementação de melhoria do modelo. Para tanto, buscou-se realizar modificações na etapa de treinamento da rede, sendo realizadas diversas iterações nos modelos até a convergência do mesmo, ou seja, procurou-se encontrar um valor de R^2 maior que o anterior, caso contrário, o modelo anterior era considerado “melhor” que o subsequente e/ou os subsequentes.

Segundo Bigiarini (2017), HydroGOF é uma biblioteca ou *package* de soluções estatísticas e gráficas desenvolvida para o ambiente R e *RStudio*. Possui uma gama de funcionalidades na avaliação, análise e testes de modelos, como forma de medir a relação entre valores

observados e estimados, nas etapas de calibração, validação e aplicação de modelos hidrológicos. No que se refere às métricas apresentadas pelo HydroGOF, tem-se:

- MSE – Erro médio quadrático, apresenta o desvio padrão do erro de predição do modelo, quando menor o erro melhor será o ajuste do modelo;
- RMSE – Raiz quadrada do erro médio;
- NRMSE – Raiz quadrada do erro médio normalizada;
- PBIAS – Percentagem de polarização, mede a tendência média dos valores calculados para serem maiores ou menores do que os observados;
- NSE – Eficiência de Nash-Sutcliffe, estatística normalizada que determina a magnitude relativa da variância residual em relação à variância dos dados medidos;
- mNSE – Eficiência de Nash-Sutcliffe modificada;
- r – Coeficiente de correlação de Pearson entre valores observados e calculados;
- R^2 – Coeficiente de determinação; e
- bR^2 – Coeficiente de determinação (R^2) multiplicado pela inclinação da linha de regressão entre os resultados simulados e observados.

Nesta pesquisa foram escolhidas as métricas que foram capazes de gerar informações necessárias para a análise dos modelos, de modo a subsidiar a escolha desses. Nessa escolha, observou-se a capacidade de avaliação de cada uma das métricas com relação ao ajuste e aderência do modelo, medição do erro gerado, verificação da perfeição do modelo, existência de viés de subestimação ou superestimação, avaliação entre os valores observados e calculados, através da relação entre eles. Assim as métricas escolhidas foram: MSE, RMSE, NRMSE, NSE, mNSE, PBIAS, r , R^2 , bR^2 .

4. RESULTADOS E DISCUSSÃO

4.1 Modelos para Estações de Tratamento de Pequeno Porte

Na elaboração desses modelos, foram utilizadas 2.747 observações de 18 parâmetros distintos, totalizando 49.446 valores observados de quatro estações de tratamento de água.

4.1.1 Análise exploratória para os modelos de ETAs de pequeno porte

Quanto à análise de *outliers*, foi verificado que a maioria dos conjuntos de variáveis apresentou considerável quantidade de valores extremos ou fora do padrão do conjunto de dados. No entanto, todos os valores foram mantidos por representarem a realidade das estações de tratamento de água utilizadas na pesquisa. Tal feito também se deve ao fato de algumas ETAs operarem com sobrecarga, além das próprias características das águas afluentes às mesmas; por serem oriundas de mananciais lóticos, em que a qualidade de suas águas sofre grande influência dos eventos de chuva no aumento de cor aparente e turbidez, por exemplo.

Dentre os conjuntos das 18 variáveis, foi constatado que todas seguem uma distribuição normal ao nível de significância de 5% para o teste de Shapiro-Wilk.

A Tabela 2 ilustra o resultado da estatística descritiva para as 18 variáveis estudadas nos modelos. Com relação à medida de tendência central, verificou-se que a maioria dos valores se encontra dentro do limite comumente encontrado para estações de tratamento de água que realizam o mesmo tipo de operação, bem como para as características da qualidade da água afluente e efluente; os valores que se encontram fora desse limite podem estar relacionados à sobrecarga das estações, além de possíveis equívocos no dimensionamento de alguma operação unitária que compõe o tratamento, apresentando valores que divergem daqueles estabelecidos pela NBR 12.216/1992, caso, por exemplo, do gradiente de floculação, cujo valor máximo é de 70 s^{-1} (ABNT, 1992). É possível observar ainda que a média e mediana são muito próximas, para maioria das variáveis, fato que evidencia uma possível simetria na distribuição de frequência da maioria das amostras.

Em se tratando de medidas de dispersão, é possível verificar que há grande variabilidade de alguns conjuntos de variáveis, conforme apresentado pelos valores de variância, mínimo e

máximo. Tal variabilidade, quando verificada nas variáveis que representam a qualidade da água, pode estar relacionada aos eventos de chuvas que ocorreram na bacia hidrográfica na qual o manancial de captação se encontra, bem como aos fatores que caracterizam o uso e ocupação do solo, tendo em vista que esses irão implicar no escoamento das águas que afluem aos referidos mananciais.

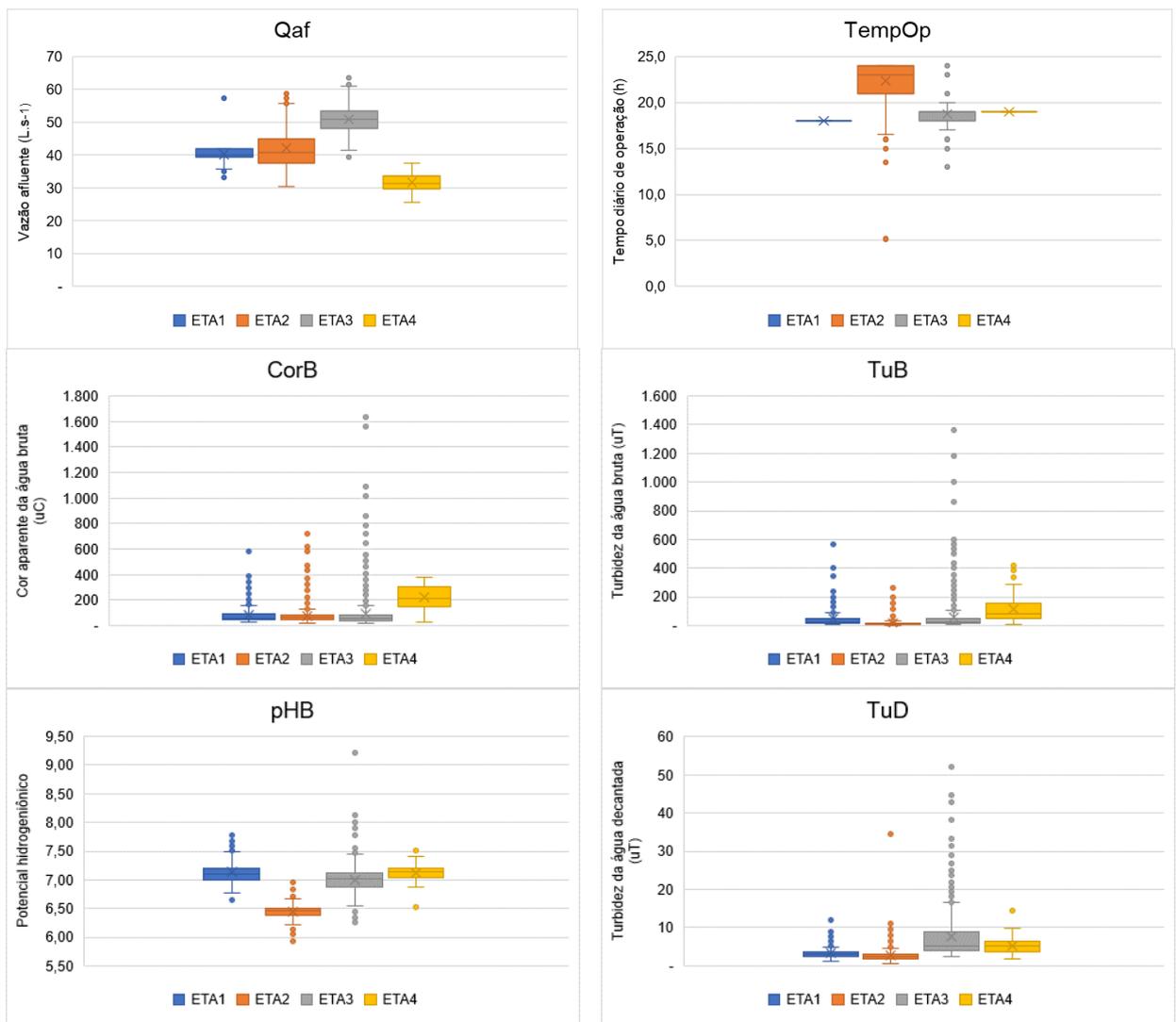
TABELA 2: Apresentação dos resultados da estatística exploratória das variáveis referentes às estações de pequeno porte

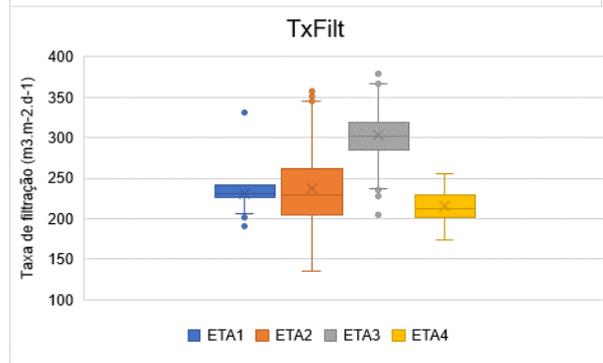
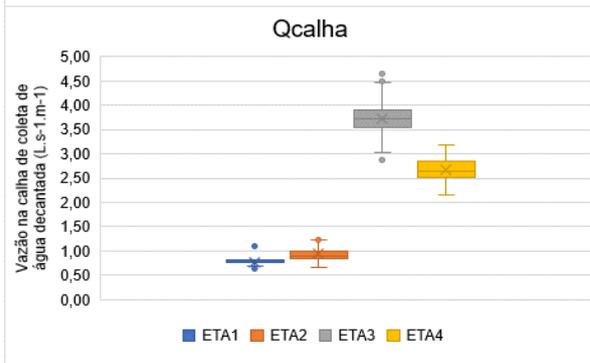
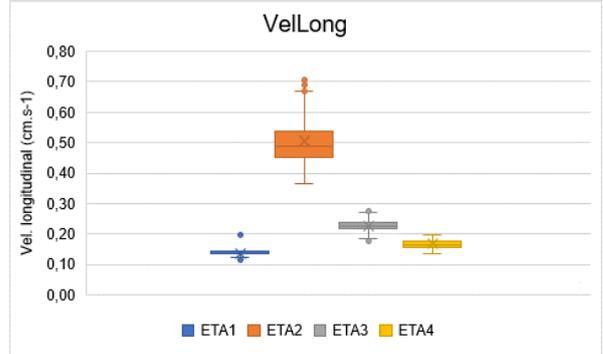
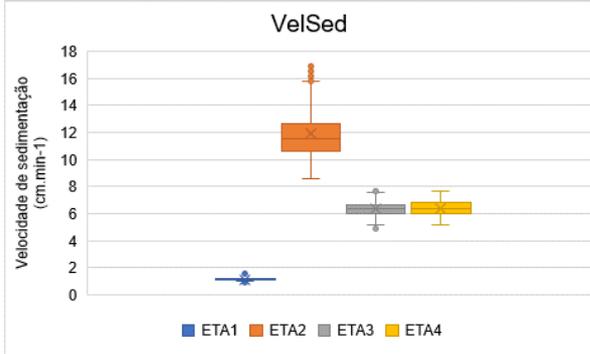
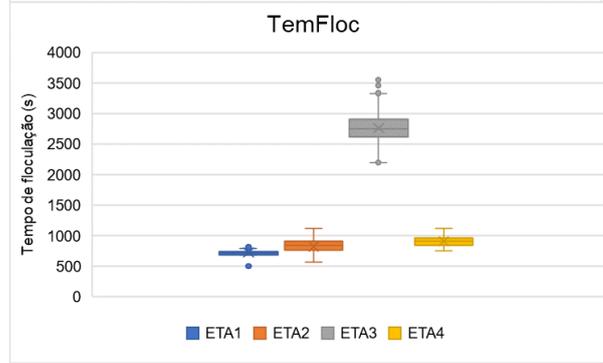
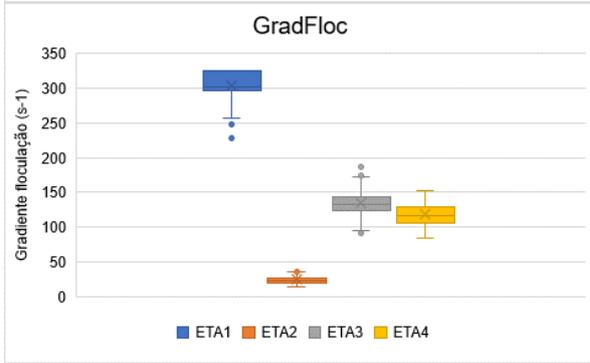
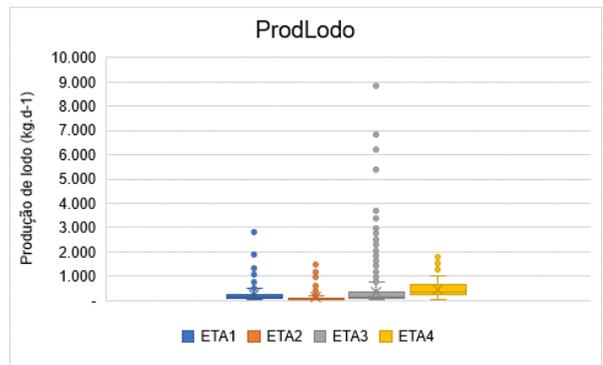
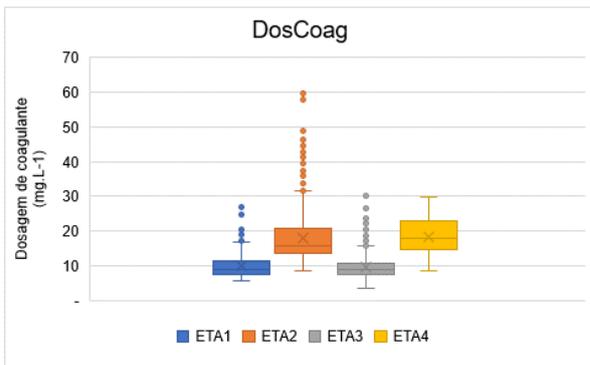
VAR	MÉDIA	MEDIANA	MODA	MÍNIMO	MÁXIMO	DESV.PAD.
Qaf (L.s⁻¹)	46,44	47,37	42,00	25,56	63,63	5,81
TempOp (h)	19,81	19,00	19,00	5,16	24,00	1,75
CorB (uC)	85,89	55,00	86,00	21,80	1.631,50	53,37
TuB (uT)	42,42	16,41	63,00	3,26	1.363,67	41,65
pHB	6,84	6,93	7,10	5,92	9,22	0,28
TuD (uT)	5,41	3,74	3,00	0,71	51,88	3,21
DosCoag (mg.L⁻¹)	12,58	10,71	7,50	3,98	159,69	4,58
ProdLodo (kg.dia⁻¹)	278,96	119,44	435,81	34,37	8.846,01	255,33
InstOP	2,46	2,00	2,00	2,00	3,00	0,50
EnsJartest	1,32	1,00	1,00	1,00	2,00	0,43
GradFloc (s⁻¹)	120,67	125,99	324,98	14,92	518,32	62,82
TemFloc (s)	1.858,25	2.475,04	692,57	507,35	3.555,59	976,15
VelSed (cm.min⁻¹)	7,44	6,54	1,16	0,92	16,93	2,84
VelLong (cm.s⁻¹)	0,30	0,23	0,14	0,11	0,72	0,13
Qcalha (L.s⁻¹.m⁻¹)	2,44	3,34	0,81	0,64	4,65	1,39
TxFilt (m³.m².dia⁻¹)	270,97	278,99	241,92	60,10	382,86	40,02
CorT (uC)	2,79	2,20	1,00	0,01	28,83	1,49
TuT (uT)	0,52	0,24	0,10	0,05	15,43	0,48

Como forma de ilustrar o comportamento das variáveis em cada uma das quatro estações de tratamento de água de pequeno porte, foi elaborada a Figura 10. Nela é possível observar os *outliers* presentes na maioria dos conjuntos de parâmetros. Além disso, é possível constatar a variabilidade dos valores de cada parâmetro em cada uma das ETAs. Por exemplo, a vazão afluente apresenta valores médios muito próximos, entre 20 e 60 L.s⁻¹ e o tempo diário de operação também, por se tratar de estações que abastecem menores populações.

Também se observa que outros parâmetros, particularmente aqueles característicos das operações unitárias, apresentam grande divergência de média e de mediana. Por exemplo, o

tempo de floculação - TemFloc, que por sua vez também apresentou expressiva variabilidade, notadamente na ETA3. Conforme mencionado, esses valores podem estar relacionados a fatores operacionais, a erros de dimensionamento das unidades ou na própria cessão dos dados pela concessionária, tendo em vista que alguns valores divergem consideravelmente daqueles apresentados pela ABNT NBR 12.216/1992 (BRASIL, 1992). Quando analisado o atendimento dos parâmetros de qualidade da água tratada (pH, cor aparente e turbidez) ao Anexo XX da Portaria de Consolidação nº 5/2017, foi verificado que as ETAs 3 e 4 apresentaram maior frequência de violação ao longo do período analisado.





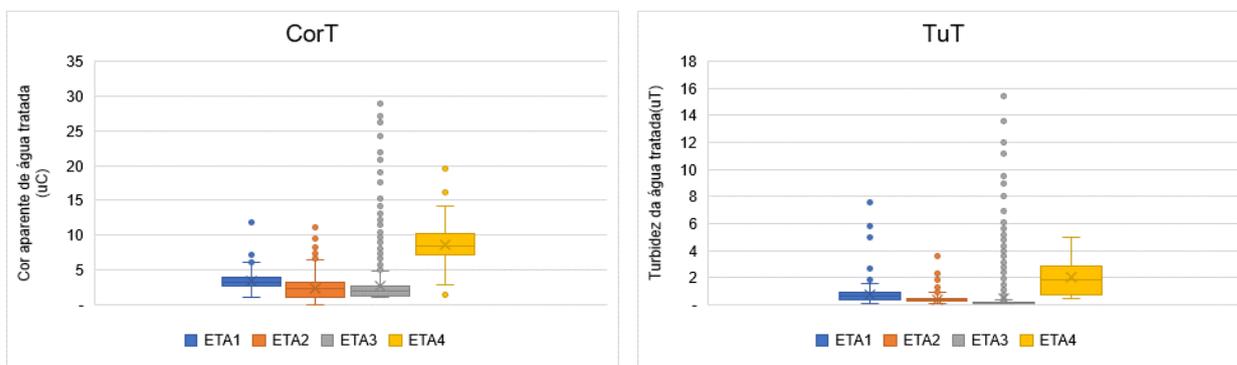


FIGURA 10: Composição de gráficos *boxplot* de cada uma das variáveis que compõem os modelos para as ETAs de pequeno porte

O parâmetro cor aparente foi mantido dentro do limite estabelecido pelo Padrão de Potabilidade, com alguns valores de *outliers* na ETA3. Já a turbidez apresentou um número maior de violações, tendo vista que o Padrão de Potabilidade estabelece que para filtração lenta o valor de turbidez de água filtrada seja de até 1,0 uT e para filtração rápida (que ocorre nas ETAs aqui estudadas) é de até 0,5 uT (BRASIL, 2017). Considerando o exposto, houve violação em todas ETAs, embora a ETA4 tenha apresentado valores médios e medianos acima do valor de referência, já na ETA3, apenas os *outliers* ficaram acima do permitido.

4.1.2 Escolha das variáveis para compor os modelos de ETAs de pequeno porte

Nesta etapa o objetivo foi escolher o número mínimo de variáveis, dentre as 18, que pudesse representar as ETAs aqui estudadas e prever, com acurácia o desempenho dessas, a partir dos valores da cor aparente e da turbidez da água tratada que eles geraram.

De acordo com a distribuição da série do banco de dados analisados, foi aplicada a análise de correlação de Pearson entre as 18 variáveis, ao nível de significância estatística de 5%. A partir dos valores das correlações e com base no *p-valor* foram selecionadas as variáveis para a composição dos modelos.

Como etapa inicial foram verificadas as variáveis que possuem coeficiente de correlação (*r*) acima de 10% com a variável TuT, sendo a mesma positiva ou negativa e com significância estatística *p*. Posteriormente, realizou a mesma análise para a variável CorT.

A Tabela 3 apresenta os resultados da matriz de correlação, com destaque para as variáveis cujos valores de *r* foram superiores a 0,1 para as variáveis CorT e TuT. De modo geral,

verifica-se que as variáveis responsáveis pelos parâmetros cor aparente e turbidez da água tratada são intimamente relacionadas ao controle da qualidade da água nas ETAs como parte integrante do processo de tratamento de água. Com relação às variáveis hidráulicas, para a turbidez, só a vazão afluente e as velocidades de escoamento no decantador apresentou correlação, fato não esperado, tendo em vista que as etapas anteriores, que também são responsáveis pela clarificação, não apresentaram esse comportamento; para a cor aparente, além das velocidades de sedimentação e longitudinal, o GradFloc também foi capaz de justificar a existência da cor na água tratada. A divergência entre as variáveis para composição dos modelos pode estar relacionada ao que é responsável por gerar cada um desses parâmetros, no caso da cor aparente, material dissolvido presente na massa líquida, e da turbidez a existência de material particulado em suspensão ou coloidal.

TABELA 3: Matriz de correlação entre as variáveis utilizadas no desenvolvimento dos modelos de estações de pequeno porte

	<i>Qaf</i>	<i>TempOp</i>	<i>CorB</i>	<i>TuB</i>	<i>pHB</i>	<i>TuD</i>	<i>DosCoag</i>	<i>ProdLodo</i>	<i>InstOP</i>	<i>EnsJartest</i>	<i>GradFloc</i>	<i>TemFloc</i>	<i>VelSed</i>	<i>VelLong</i>	<i>Qcalha</i>	<i>TxFilt</i>	<i>CorT</i>	<i>TuT</i>	
<i>Qaf</i>	1																		
<i>TempOp</i>	-0,23	1																	
<i>CorB</i>	-0,06	-0,05	1																
<i>TuB</i>	0,03	-0,16	0,93	1															
<i>pHB</i>	0,29	-0,70	-0,07	0,05	1														
<i>TuD</i>	0,30	-0,27	0,64	0,74	0,17	1													
<i>DosCoag</i>	-0,29	0,44	0,43	0,25	-0,53	0,02	1												
<i>ProdLodo</i>	0,10	-0,16	0,91	0,99	0,05	0,77	0,23	1											
<i>InstOP</i>	-0,71	0,53	-0,02	-0,15	-0,51	-0,47	0,46	-0,20	1										
<i>EnsJartest</i>	-0,43	0,82	-0,06	-0,21	-0,83	-0,39	0,55	-0,21	0,74	1									
<i>GradFloc</i>	0,10	-0,68	0,02	0,12	0,72	0,11	-0,40	0,11	-0,16	-0,75	1								
<i>TemFloc</i>	0,61	-0,52	0,03	0,16	0,48	0,48	-0,45	0,20	-0,99	-0,71	0,11	1							
<i>VelSed</i>	0,05	0,77	-0,05	-0,17	-0,76	-0,20	0,45	-0,14	0,33	0,84	-0,90	-0,33	1						
<i>VelLong</i>	-0,10	0,82	-0,07	-0,20	-0,80	-0,31	0,51	-0,19	0,55	0,93	-0,80	-0,55	0,96	1					
<i>Qcalha</i>	0,76	-0,51	0,04	0,15	0,51	0,46	-0,43	0,19	-0,97	-0,73	0,15	0,94	-0,27	-0,52	1				
<i>TxFilt</i>	0,95	-0,13	-0,03	0,06	0,35	0,33	-0,32	0,13	-0,72	-0,50	0,15	0,63	-0,03	-0,19	0,78	1			
<i>CorT</i>	-0,11	-0,08	0,65	0,64	0,00	0,54	0,28	0,63	0,05	-0,11	0,12	-0,04	-0,12	-0,13	-0,01	-0,06	1		
<i>TuT</i>	-0,10	-0,06	0,64	0,68	-0,02	0,58	0,24	0,67	0,03	-0,08	0,07	-0,01	-0,10	-0,10	-0,01	-0,06	0,86	1	

Conforme os valores de R supracitados, foram escolhidas as seguintes variáveis:

- Turbidez da água tratada (TuT) – Qaf, CorB, TuB, TuD, DosCoag, ProdLodo, VelSed e VelLong; e
- Cor aparente da água tratada (CorT) – Qaf, CorB, TuB, TuD, DosCoag, ProdLodo, EnsJartest, GradFloc, VelSed e VelLong.

4.1.3 Modelos em regressão linear múltipla para ETAs de pequeno porte

Como forma de verificar a qualidade dos modelos propostos elaborados em RNA, foram desenvolvidos modelos em regressão linear múltipla. Esta etapa foi necessária pelo fato dessa metodologia ser amplamente utilizada no desenvolvimento de modelos de predição nas mais diversas áreas de aplicação. O referido modelo servirá de base na comparação com o modelo elaborado em RNA. É importante destacar que, na elaboração dos modelos de ETAs de turbidez da água tratada, foi necessária aplicar a transformação das variáveis antes de aplicar a regressão linear múltipla.

Para a verificação da existência da multicolinearidade dos resíduos observaram-se os valores do Fator de Inflação de Variância (VIF). Nesta etapa, para que a multicolinearidade seja considerada inexistente é necessário que o valor do VIF seja menor que 10.

Inicialmente, foi elaborado o Modelo 1, tomando a turbidez da água tratada como variável de interesse. Nessa etapa aplicou-se a regressão linear múltipla com todas as variáveis escolhidas para a composição do modelo, no entanto constatou-se a presença de multicolinearidade, através dos valores do VIF, além dos valores-P acima de 0,05. Nesse caso, optou-se pela retirada de algumas variáveis até se chegar em um resultado aceitável com relação ao VIF. As variáveis que permaneceram para elaboração do referido modelo foram: Qaf, CorB, TuB, TuD, DosCoag e VelLong.

As tabelas 4, 5 e 6 ilustram os resultados do modelo de regressão, apresentando a estatística do teste, o coeficiente de determinação para o modelo proposto e a equação que representa o modelo. Na Tabela 4 destaca-se o percentual de contribuição de cada variável para o modelo, onde se verifica que a cor da água bruta é responsável por quase 40% da predição da turbidez da água tratada. Com relação à significância de cada variável para o modelo, verifica-se que a cor aparente da água bruta apresentou valor acima de 0,05, no entanto a mesma foi mantida devido a sua contribuição. Já na Tabela 5 são apresentados os valores dos coeficientes de

determinação ($R^2 = 0,51$), indicando que o modelo é capaz de responder por aproximadamente 50% da predição de turbidez da água tratada.

TABELA 4: Análise de variância gerada para o modelo de regressão em ETAs de pequeno porte a partir da turbidez da água tratada

Fonte	GL	SQ Seq	Contribuição	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	6	1397,25	51,49%	1397,25	232,875	484,34	0,000
Qaf	1	29,52	1,09%	82,78	82,783	172,18	0,000
CorB	1	1082,97	39,91%	1,75	1,750	3,64	0,057
TuB	1	167,99	6,19%	63,00	62,995	131,02	0,000
TuD	1	99,24	3,66%	110,16	110,155	229,11	0,000
DosCoag	1	13,85	0,51%	4,87	4,869	10,13	0,001
VelLong	1	3,67	0,14%	3,67	3,673	7,64	0,006
Erro	2738	1316,44	48,51%	1316,44	0,481		
Falta de ajuste	2678	1316,44	48,51%	1316,44	0,492	*	*
Erro puro	60	0,00	0,00%	0,00	0,000	*	*
Total	2744	2713,69	100,00%				

Legenda: GL – Graus de liberdade, SQ Seq – Soma dos quadrados sequenciais, Contribuição – Percentual que cada variável representa para o modelo, SQ (Aj.) – Soma dos quadrados ajustados, QM (Aj.) – Quadrado médio ajustado, Valor F – Estatística utilizada para verificar se o modelo está associado com a resposta, Valor-P – Mede a evidência contrária a hipótese nula do teste, Erro puro – erro calculado a partir do número de repetições de observações utilizadas no modelo, Falta de ajuste – calcula a falta de ajuste gerada pela presença de réplicas no conjunto de dados.

TABELA 5: Sumário do modelo para turbidez da água tratada em ETAs de pequeno porte

S	R2	R2(aj)	PRESQ	R2(pred)
0,693400	51,49%	51,38%	1348,85	50,29%

Legenda: S – Desvio padrão, R^2 – Coeficiente de determinação, $R^2(aj)$ – Coeficiente de determinação ajustado, PRESQ – Soma dos quadrados preditos do erro (calcula o desvio do modelo), $R^2(pred)$ – Coeficiente de determinação predito (determina o quanto o modelo prediz para novas observações).

Na Tabela 6 são apresentados os coeficientes gerados a partir da aplicação da regressão linear múltipla para a proposição do modelo. Na primeira coluna da tabela são apresentados os coeficientes para a elaboração da equação, em seguida o erro padrão de cada coeficiente (são responsáveis por medir a precisão do modelo na estimativa de valores desconhecidos dos coeficientes). Posteriormente verifica-se o intervalo de confiança gerado ao nível de

significância de 5%, os valores de T (razão entre os coeficientes e seus valores de erro padrão) e o valor-P (representa a significância de cada uma das variáveis independentes na predição da variável dependente do modelo), e por fim o VIF. O resultado da regressão é a equação do modelo de predição para a turbidez da água tratada.

TABELA 6: Coeficientes que compõem o modelo de estações de pequeno porte para turbidez da água tratada

Termo	Coef	EP de Coef	IC de 95%	Valor-T	Valor-P	VIF
Constante	1,099	0,107	(0,889; 1,310)	10,23	0,000	
Qaf	-0,02915	0,00222	(-0,03350; -0,02479)	-13,12	0,000	1,31
CorB	-0,000841	0,000441	(-0,001706; 0,000023)	-1,91	0,057	10,79
TuB	0,006285	0,000549	(0,005209; 0,007362)	11,45	0,000	11,06
TuD	0,06800	0,00449	(0,05919; 0,07681)	15,14	0,000	2,76
DosCoag	0,00907	0,00285	(0,00348; 0,01467)	3,18	0,001	2,28
VelLong	0,317	0,115	(0,092; 0,542)	2,76	0,006	1,65

Legenda: Termo – Elementos que irão compor o modelo, Coef – Coeficientes do modelo (relação entre as variáveis dependentes e independentes), EP de Coef – Erro padrão dos coeficientes, IC de 95% - Intervalo de confiança ao nível de significância de 95%, Valor-T – Razão entre o coeficiente e o respectivo erro padrão, Valor-P – Mede a evidência contra a hipótese nula, VIF – Fator de inflação da variância (demonstra como a variância de cada coeficiente eleva-se pela correlação entre as variáveis independentes ou preditoras).

Equação de Regressão

$TuT = 1,099 - 0,02915 Qaf - 0,000841 CorB + 0,006285 TuB + 0,06800 TuD + 0,00907 DosCoag + 0,317 VelLong$
--

No desenvolvimento do modelo de regressão linear múltipla para a cor aparente de água tratada, inicialmente, foram realizados os mesmos procedimentos para a turbidez (com exceção da transformação dos dados), fazendo uso das variáveis e as excluindo de acordo com os valores do VIF até definir as variáveis: Qaf, CorB, TuD, DosCoag, GradFloc e VelSed.

Conforme ilustrado na Tabela 7, a variável cor aparente da água bruta (CorB) foi a que mais contribuiu para o modelo, sendo responsáveis por mais de 40% da predição. Já na Tabela 8 são apresentados os coeficientes de determinação dos modelos, com destaque para o R² ajustado (0,50), que indica a capacidade do modelo para a predição da cor aparente efluente de ETAs de pequeno porte.

TABELA 7: Análise de variância gerada para o modelo de regressão em ETAs de pequeno porte a partir da cor aparente da água tratada

Fonte	GL	SQ Seq	Contribuição	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	6	8541,7	51,04%	8541,69	1423,62	475,63	0,000
Qaf	1	204,6	1,22%	660,93	660,93	220,82	0,000
CorB	1	6931,7	41,42%	781,98	781,98	261,26	0,000
TuD	1	814,1	4,86%	1138,27	1138,27	380,29	0,000
DosCoag	1	29,7	0,18%	20,22	20,22	6,75	0,009
GradFloc	1	269,7	1,61%	498,60	498,60	166,58	0,000
VelSed	1	291,8	1,74%	291,83	291,83	97,50	0,000
Erro	2738	8195,2	48,96%	8195,19	2,99		
Falta de ajuste	2684	8195,2	48,96%	8195,19	3,05	*	*
Erro puro	54	0,0	0,00%	0,00	0,00	*	*
Total	2744	16736,9	100,00%				

Legenda: GL – Graus de liberdade, SQ Seq – Soma dos quadrados sequenciais, Contribuição – Percentual que cada variável representa para o modelo, SQ (Aj.) – Soma dos quadrados ajustados, QM (Aj.) – Quadrado médio ajustado, Valor F – Estatística utilizada para verificar se o modelo está associado com a resposta, Valor-P – Mede a evidência contrária a hipótese nula do teste, Erro puro – erro calculado a partir do número de repetições de observações utilizadas no modelo, Falta de ajuste – calcula a falta de ajuste gerada pela presença de réplicas no conjunto de dados.

TABELA 8: Sumário do modelo para cor aparente da água tratada em ETAs de pequeno porte

S	R2	R2(aj)	PRESQ	R2(pred)
1,73007	51,04%	50,93%	8292,93	50,45%

Legenda: S – Desvio padrão, R² – Coeficiente de determinação, R²(aj) – Coeficiente de determinação ajustado, PRESQ – Soma dos quadrados preditos do erro (calcula o desvio do modelo), R²(pred) – Coeficiente de determinação predito (determina o quanto o modelo prediz para novas observações).

Os coeficientes do modelo estão apresentados na Tabela 9. Além dos coeficientes de cada variável independente, pode-se observar o seu erro padrão e os intervalos de confiança. Verifica-se que os valores do VIF foram abaixo de 10, indicando inexistência de multicolinearidade entre os resíduos. Posteriormente, para a aplicação do modelo tem-se a equação linear de regressão.

TABELA 9: Coeficientes que compõem o modelo de estações de pequeno porte para cor aparente da água tratada

Termo	Coef	EP de Coef	IC de 95%	Valor-T	Valor-P	VIF
Constante	1,568	0,311	(0,958; 2,177)	5,05	0,000	
Qaf	-0,09716	0,00654	(-0,10998; -0,08434)	-14,86	0,000	1,82
CorB	0,008802	0,000545	(0,007734; 0,009870)	16,16	0,000	2,64
TuD	0,2144	0,0110	(0,1929; 0,2360)	19,50	0,000	2,65
DosCoag	0,01784	0,00686	(0,00438; 0,03130)	2,60	0,009	2,12
GradFloc	0,01332	0,00103	(0,01130; 0,01535)	12,91	0,000	7,40
VelSed	0,2665	0,0270	(0,2136; 0,3195)	9,87	0,000	8,88

Legenda: Termo – Elementos que irão compor o modelo, Coef – Coeficientes do modelo (relação entre as variáveis dependentes e independentes), EP de Coef – Erro padrão dos coeficientes, IC de 95% - Intervalo de confiança ao nível de significância de 95%, Valor-T – Razão entre o coeficiente e o respectivo erro padrão, Valor-P – Mede a evidência contra a hipótese nula, VIF – Fator de inflação da variância (demonstra como a variância de cada coeficiente eleva-se pela correlação entre as variáveis independentes ou preditoras).

Equação de Regressão

$\text{CorT} = 1,568 - 0,09716 \text{ Qaf} + 0,008802 \text{ CorB} + 0,2144 \text{ TuD} + 0,01784 \text{ DosCoag} + 0,01332 \text{ GradFloc} + 0,2665 \text{ VelSed}$

4.1.4 Arquitetura e topologia da rede neural aplicada aos modelos

Após a definição das variáveis para composição dos modelos, iniciou-se a fase de teste com a RNA, de modo a verificar a arquitetura e topologia que iria atender melhor na predição das variáveis de controle ou dependentes. No entanto, antes que se fizessem quaisquer alterações nas estruturas da RNA, foi necessário efetuar alguns testes durante sua etapa de treinamento ou aprendizagem. Nesta, iniciaram-se as modificações na função ou algoritmo de treinamento, no algoritmo de medição do erro, quantidade de repetições para o treinamento, número máximo de iterações e erro tolerável.

Realizada a definição das configurações de treinamento, partiu-se para a escolha da melhor topologia da rede. A primeira iniciativa foi mudar o número de camadas ocultas e o número de neurônios para cada uma dessas camadas. Também foram realizadas modificações com a ausência e presença da função linear na camada de saída da rede.

A métrica escolhida para a apresentação dos resultados dos modelos foi o coeficiente de determinação R^2 , entre os valores preditos e observados durante a etapa de treinamento, bem como na capacidade dos valores preditos estarem inseridos, com $\alpha = 0,05$, no intervalo de predição. A escolha da referida métrica se deu pelo fato da mesma ser amplamente utilizada como ferramenta de avaliação da capacidade de predição de vários tipos de modelos.

Para os maiores valores de R^2 , tanto na fase de treinamento como na simulação, chegou-se às seguintes configurações:

- Topologia perceptron multicamada, algoritmo de treinamento – rprop+ (*backpropagation* resiliente com retrocesso dos pesos);
- Função de erro – sse (soma dos quadrados do erro);
- Duas camadas ocultas em cada modelo;
- Número de neurônios por camada – de acordo como o melhor ajuste de cada modelo;
- Número máximo de um milhão de interações;
- Sem utilização da camada de saída linear;
- Repetição única;
- Limite de erro – 0,0025; e
- Função de ativação – logistic (logarítmica).

É importante ressaltar que estas configurações da RNA foram utilizadas para os modelos com base na cor aparente e na turbidez da água tratada, bem como para os modelos das estações de tratamento de água de médio e grande porte, como forma de uniformizar e poder comparar os modelos.

4.1.5 Modelo em RNA de estações de tratamento de pequeno porte para a turbidez da água tratada

Durante essa etapa de desenvolvimento do modelo foram realizados diversos testes até que se chegasse à topologia apresentada na Figura 11. Nessa figura, é possível observar que a estrutura da rede é relativamente simples. Também é possível perceber os valores dos pesos em cada uma das camadas (de entrada, oculta e de saída) – em azul, além dos valores de saída de cada neurônio após cada sinapse, até a geração do vetor de saída com o conjunto dos valores preditos da variável de controle.

Na Figura 12 são ilustrados os valores de turbidez observada (em preto) e calculada (em vermelho), obtidos com o Modelo 1, durante a etapa de treinamento da rede neural. Conforme apresentado, os valores são relativamente próximos e constata-se que o modelo apresenta predisposição para a estimativa de valores extremos inferiores para a variável em questão, sendo possível realizar estudos de predição para a turbidez da água tratada a partir da estimativa das variáveis de entrada.

Com relação ao atendimento ao padrão de potabilidade em vigor, verificaram-se que os resultados gerados pelo Modelo 1 atenderam ao Anexo XX da Portaria de Consolidação nº5/2017, tendo em vista que a maioria dos valores foi inferior a 0,5 uT. Esse resultado justifica a aplicabilidade do referido modelo, pelo fato de o mesmo ser capaz de estimar resultados que possuem grande similaridade com os valores reais observados para o parâmetro em questão.

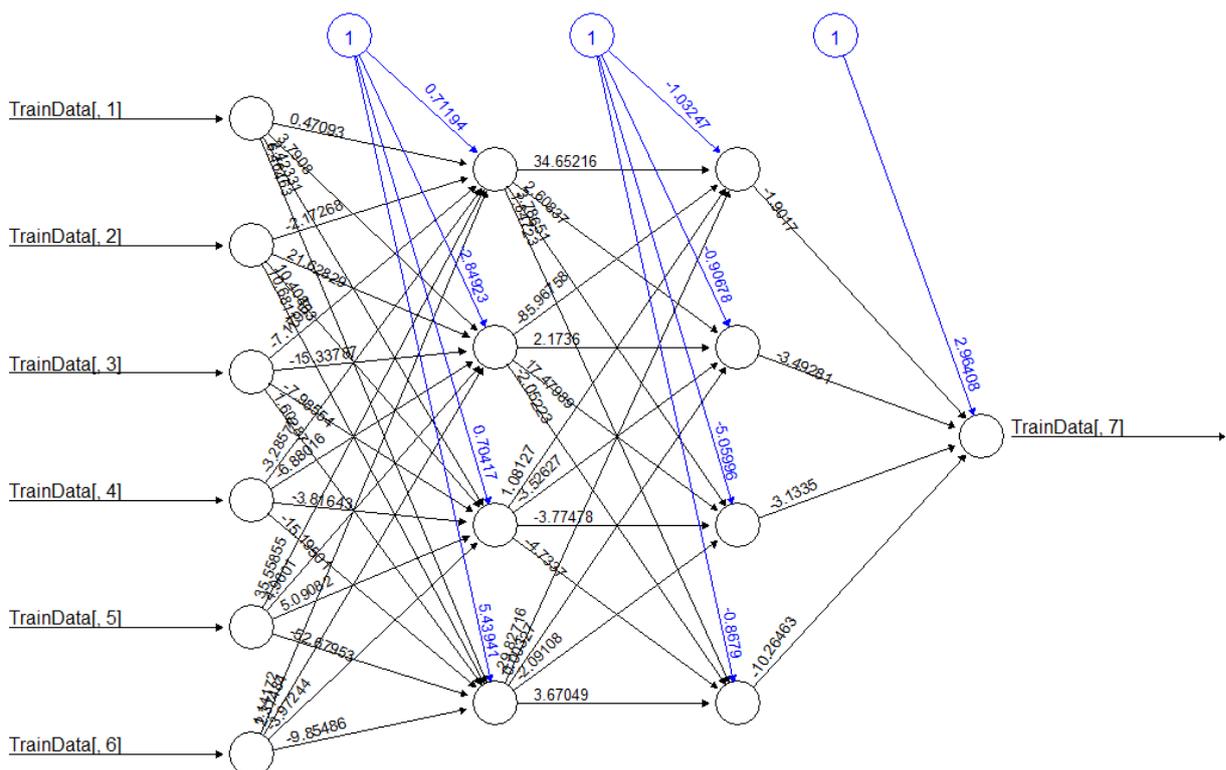


FIGURA 11: Rede neural gerada a partir do modelo proposto para a turbidez da água tratada em ETAs de pequeno porte

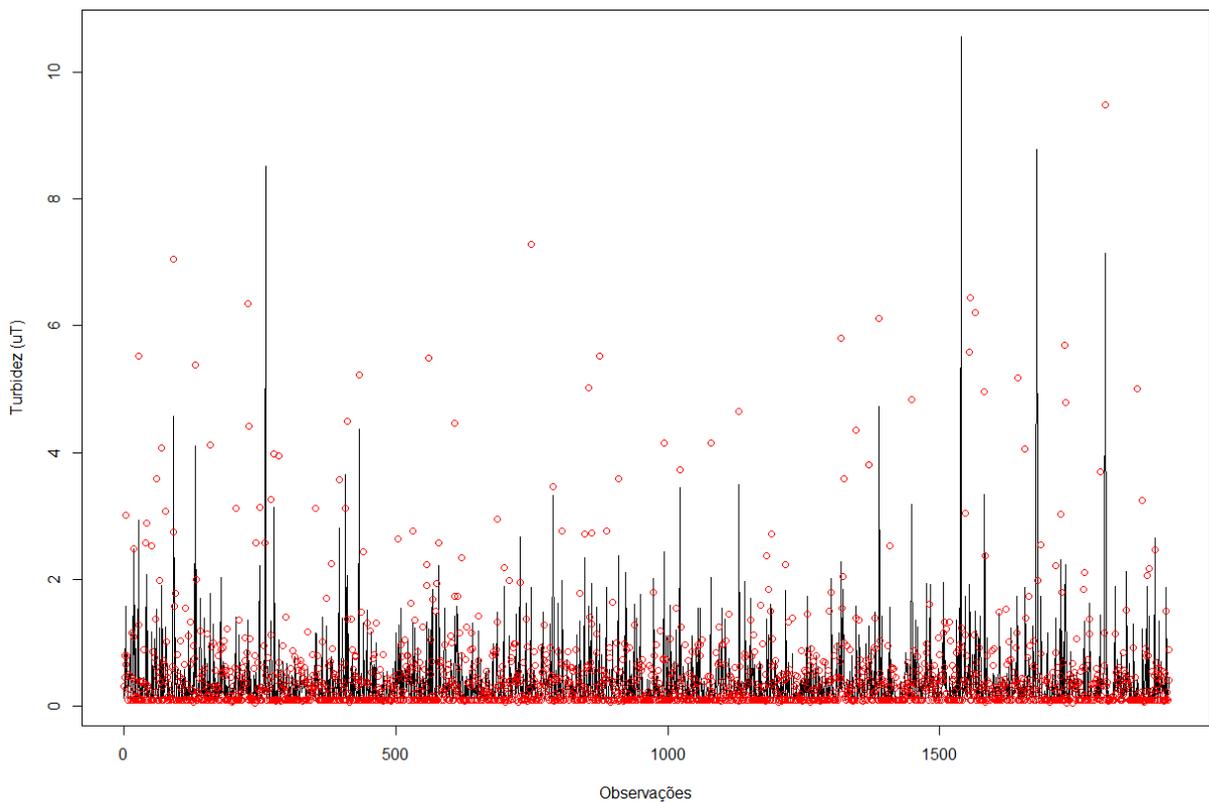


FIGURA 12: Comparação entre o comportamento dos valores de turbidez da água tratada durante o treinamento do modelo das estações de pequeno porte

Como forma de atestar ainda mais a aderência ou a resposta do Modelo 1, foi aplicada a análise de correlação entre os valores de turbidez observada e calculada. Foi gerado o modelo de regressão com o resultado, o qual apresentou coeficiente de determinação de 0,71 durante a fase de treinamento, conforme apresentado na Figura 13. Esse resultado indica que o Modelo 1 foi capaz de prever, mesmo com a presença de viés de superestimação (distribuição dos pontos acima da reta diagonal a 45°), o desempenho das estações de tratamento de água de pequeno porte, a partir dos valores de turbidez de água tratada.

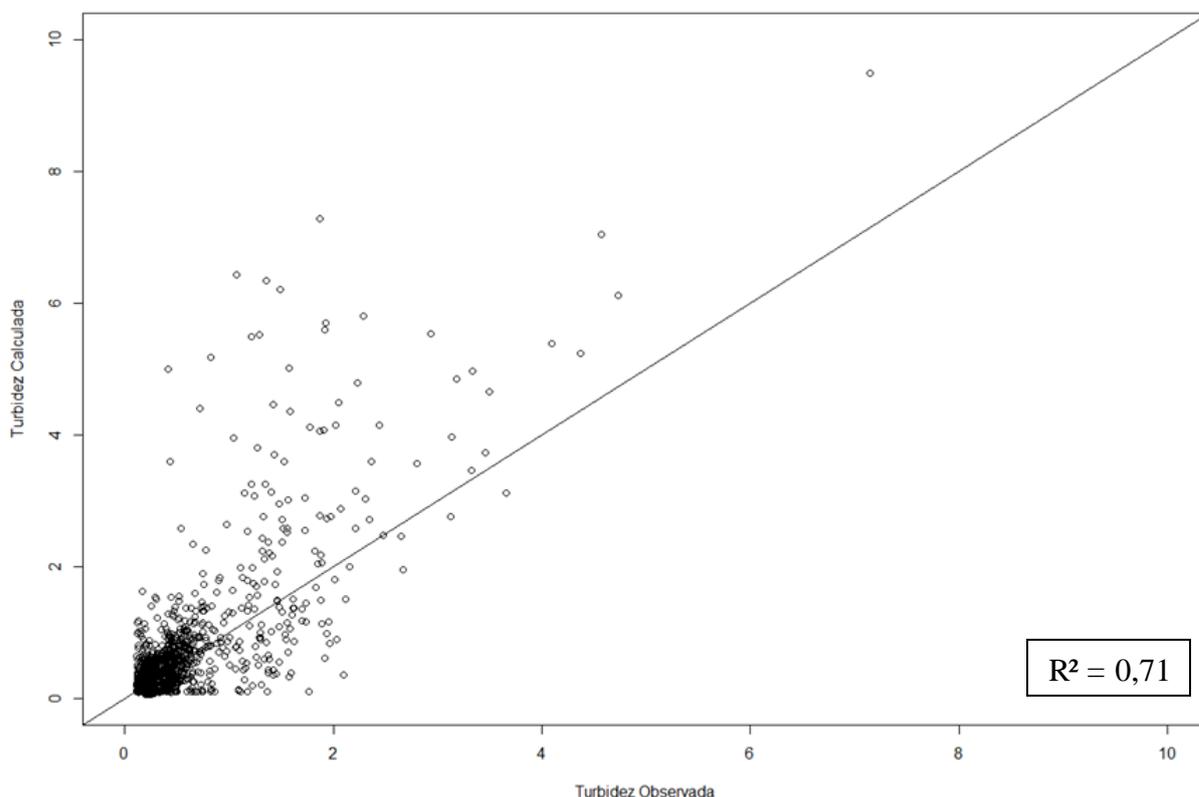


FIGURA 13: Comportamento da correlação entre a turbidez calculada e observada durante o treinamento da rede neural para estações de pequeno porte

A Figura 14 ilustra o resultado do Modelo 1, no qual se verificam os valores preditos (em vermelho) e intervalo de predição (em azul). Conforme observado, o coeficiente de determinação calculado para os valores preditos de turbidez da água tratada para as ETAs estudadas foi de 0,66 ou 66%. Esse resultado corrobora com a possibilidade do desenvolvimento de aplicações desse tipo de modelo, devido a sua capacidade de aderência. Outro aspecto de grande importância no desenvolvimento desse modelo é que o gestor do sistema tem acesso a um intervalo, com 95% de certeza, que ele poderá tomar decisão, e não necessariamente um valor isolado, que dificulta as chances de acertos.

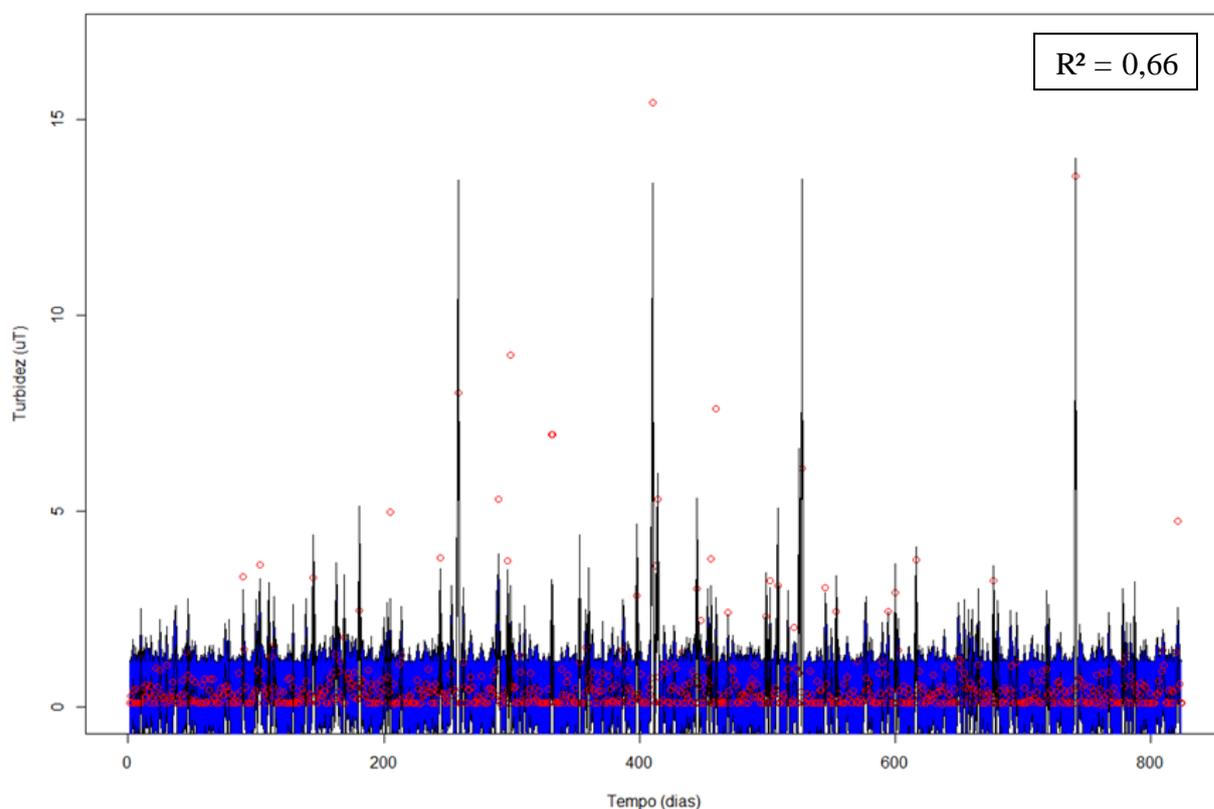


FIGURA 14: Intervalo de predição gerado a partir dos valores preditos de turbidez da água tratada para ETAs de pequeno porte

Na Tabela 10 são apresentados os resultados das métricas geradas a partir da aplicação do *package* HydroGOF, com vista à avaliação de desempenho do Modelo 1. Os valores apresentados pelos coeficientes de determinação, tanto no treinamento como no teste, são de interesse para esse tipo de pesquisa (R^2 de 0,71 e 0,66, respectivamente), tendo em vista o critério de eficiência de Nash-Sutcliffe foi acima de 0,6, demonstrando que o modelo apresenta concordância com o valor observado. Embora o erro apresentou valores abaixo de 0,60 uT, podendo gerar valores preditos fora do padrão de potabilidade de 0,5 uT (BRASIL, 2017), é importante ressaltar que na interpretação do RMSE seu valor estar elevado ao quadrado. Verifica-se ainda que os valores estimados pelo modelo dão indícios de vieses de subestimação em relação ao parâmetro observado, conforme as grandezas do percentual de BIAS possuem sinal negativo. Essa subestimação pode estar relacionada a diversos fatores, dentre eles o tipo de variável utilizada, as características das estações de tratamento, o tratamento prévio dos dados, entre outros. Também é possível inferir que o treinamento ocorreu de forma esperada, pelo fato de as métricas da etapa de teste estarem próximas as da etapa de treinamento, enfatizando a necessidade desta tarefa no resultado da performance do modelo.

Apesar da presença do viés de subestimação, é importante ressaltar que no intervalo de predição estão contidos os valores preditos, que mesmo apresentado incertezas, oferecem um limite para que as decisões sejam tomadas a partir de um intervalo.

TABELA 10: Resultados das métricas durante a elaboração do Modelo 1

MÉTRICAS	TREINAMENTO	TESTE
MSE	0,30	0,44
RMSE	0,55	0,66
NRMSE (%)	58,00	59,60
PBIAS (%)	-19,60	3,50
NSE	0,66	0,64
mNSE	0,50	0,46
r	0,84	0,81
R ²	0,71	0,66
bR ²	0,43	0,53

Legenda: MSE – Erro médio quadrático, RMSE – Raiz do erro médio quadrático, NRMSE (%) – Raiz do erro médio quadrático normalizada, PBIAS (%) – Percentagem de polarização, NSE – Eficiência de Nash-Sutcliffe, mNSE – Eficiência de Nash-Sutcliffe modificada, r – Coeficiente de correlação de Pearson, R² – Coeficiente de determinação, bR² – Coeficiente de determinação (R²) multiplicado pela inclinação da linha de regressão.

4.1.6 Modelo em RNA de estações de tratamento de pequeno porte para a cor aparente da água tratada

Como forma de assegurar que a predição do desempenho das estações de tratamento de água não estivesse sob o controle de um parâmetro apenas, como ocorre na maioria dos modelos propostos, a cor aparente da água tratada também foi escolhida como variável dependente ou de controle. Ambos os parâmetros reúnem informações importantes sobre a qualidade da água, além de, usualmente, apresentar significativa correlação quando a cor verdadeira é baixa. O modelo de predição de desempenho com base na cor aparente da água tratada foi denominado de Modelo 2, desenvolvido a partir de todos os parâmetros de treinamento já admitidos na composição do Modelo 1, conforme informações supracitadas.

A Figura 15 ilustra a configuração da RNA com três neurônios, em cada uma das duas camadas ocultas. Também se observam os vetores de entrada de cada uma das variáveis independentes do modelo, as saídas de cada neurônio em suas camadas e o valor dos pesos. Quando comparado com o Modelo 1, verifica-se que as configurações são bastante similares.

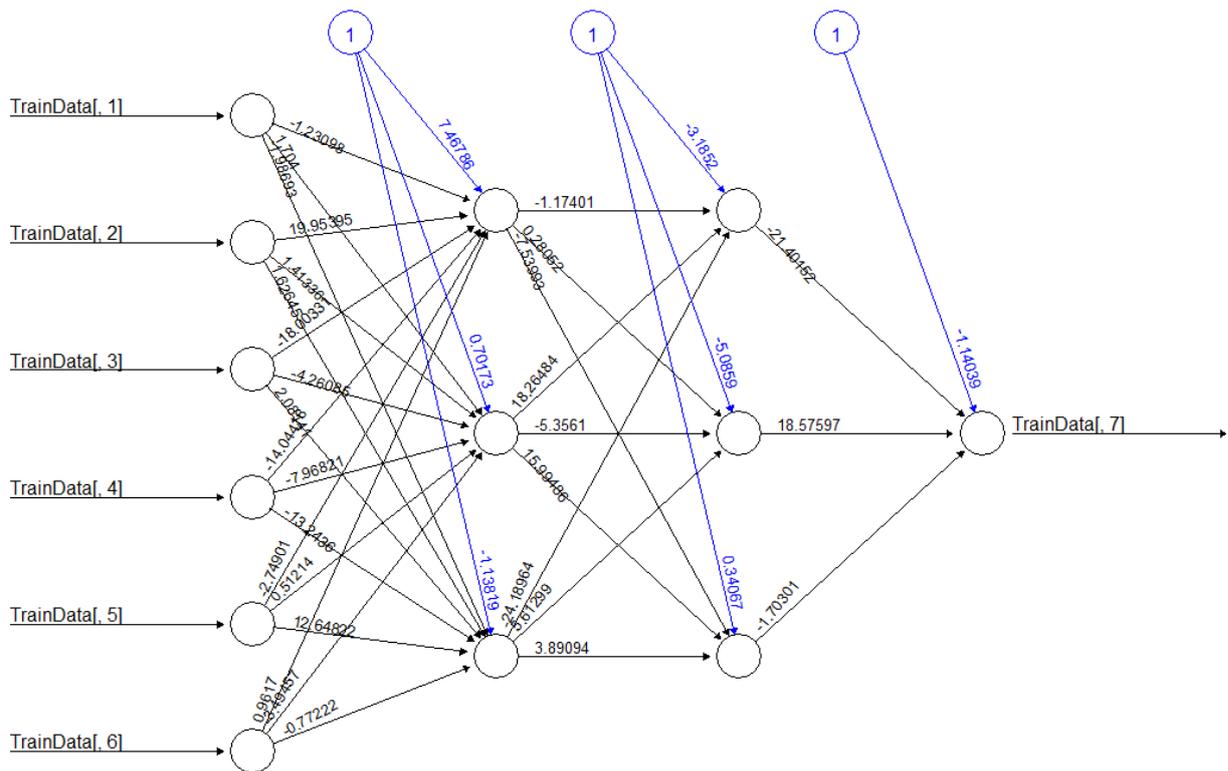


FIGURA 15: Rede neural gerada a partir do modelo proposto para a cor aparente da água tratada em ETAs de pequeno porte

Na Figura 16 é ilustrado o comportamento da cor aparente observada e calculada, para a mesma distribuição temporal, durante o treinamento da RNA. Constata-se que, assim como ocorreu com o Modelo 1, o Modelo 2 conseguiu prever o comportamento do parâmetro cor aparente, notadamente nos valores mais extremos, sejam eles superiores ou inferiores. Também é possível verificar que o Modelo 2 apresenta dificuldades na predição dos valores mais baixos, pois é perceptível ao longo tempo que os valores menores preditos de cor aparente foram subestimados. Esse fato pode estar relacionado à própria característica do parâmetro em questão, que diferente da turbidez, tem sua origem a partir da presença de material dissolvido, que dificulta sua remoção no processo de tratamento, particularmente quando os valores de entrada são reduzidos; adicionado ao exposto, a amplitude dos valores de cor aparente é bem maior quando comparada à da turbidez, inclusive nos limites estabelecidos pelo Anexo XX da Portaria de Consolidação nº 5/2017 (BRASIL, 2017).

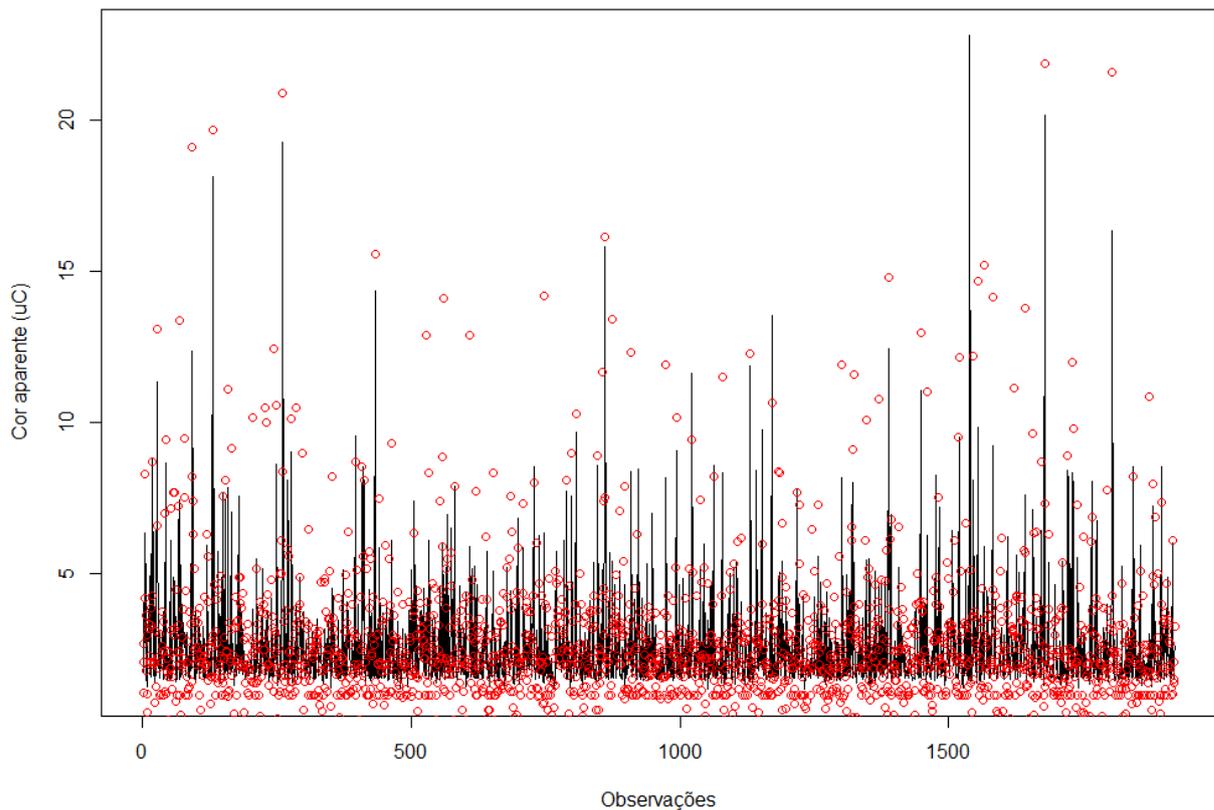


FIGURA 16: Comparação entre o comportamento dos valores de cor aparente da água tratada durante o treinamento do modelo das estações de pequeno porte

De maneira geral, é comum nas estações de tratamento de água que o parâmetro cor aparente apresente maior variabilidade que outros parâmetros monitorados, como por exemplo o pH, a turbidez, entre outros. Esse comportamento pode estar relacionado aos diversos aspectos que dão origem à cor, como a presença de matéria orgânica, ferro, manganês, poluentes oriundos de processos industriais, entre outros. Muitas vezes a cor aparente não é adequadamente removida pelas etapas convencionais de tratamento, particularmente em ETAs que apresentam baixo desempenho, e parte dessa cor persiste até a saída das estações, embora o parâmetro se apresente dentro do padrão de potabilidade.

Conforme apresentado na Figura 17, apesar de inferior ao Modelo 1, mesmo sendo modelos de ETAs de mesmo porte, o Modelo 2 apresentou aceitável capacidade de predição do ao longo do treinamento, com coeficiente de determinação (R^2) de 0,62, quando da comparação entre os valores observados e calculados de cor aparente. Além do valor de R^2 , verifica-se a distribuição da cor aparente e sua aderência à reta a 45°.

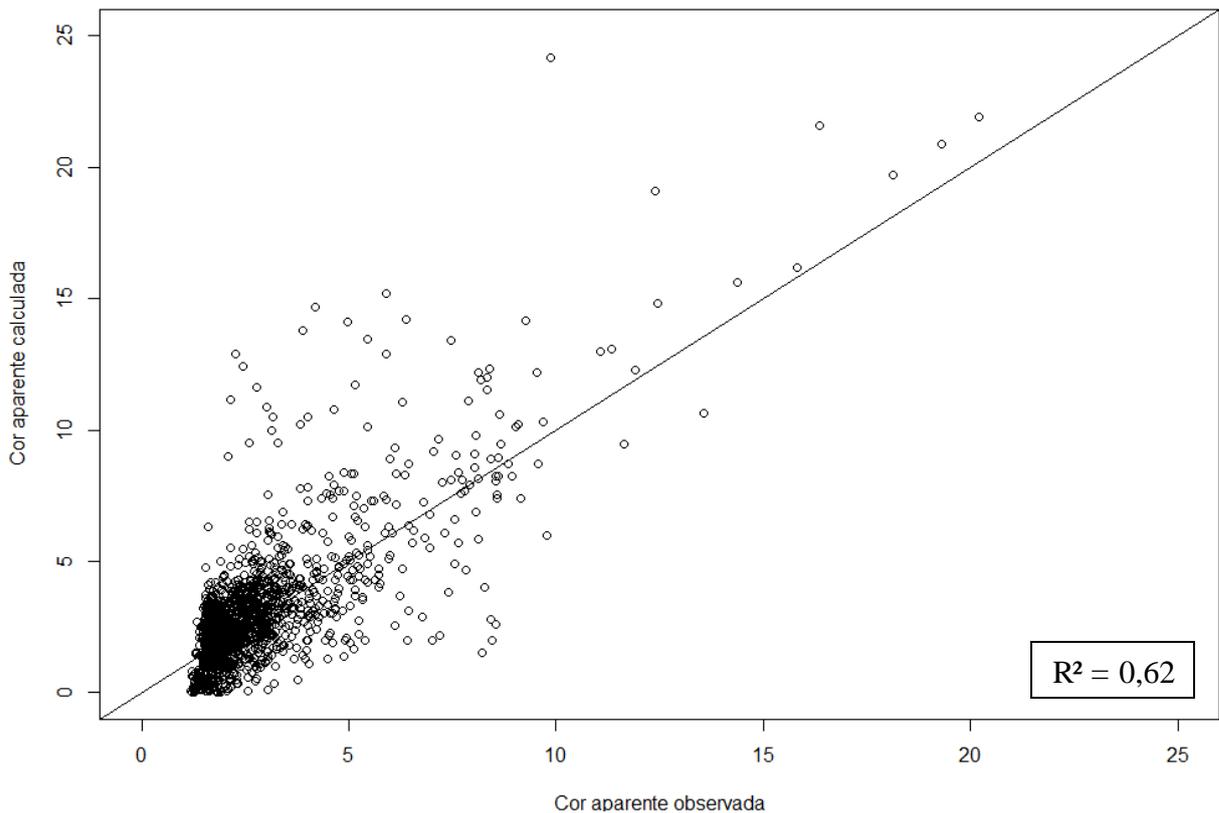


FIGURA 17: Comportamento da correlação entre a cor aparente calculada e observada durante o treinamento da rede neural para estações de pequeno porte

A Figura 18 apresenta o resultado do Modelo 2. Nela é observado, em azul, o intervalo de predição, os valores observados (em preto) e preditos (em vermelho), com seu coeficiente de determinação de 0,67. Esses resultados possibilitam estudos para a aplicação do referido modelo para a finalidade pretendida, de modo a facilitar os tomadores de decisão, além da possibilidade de melhorar os mecanismos de controle da qualidade da água nas estações de tratamento, que, em sua grande maioria e de forma exclusiva, lançam mão de métodos analíticos tradicionais, que requerem grandes aparatos experimentais e pessoal qualificado.

Em se tratando do atendimento ao Padrão de Potabilidade em vigor, os valores estimados pelo Modelo 2 para parâmetro cor aparente da água tratada, tanto na etapa de treinamento como na simulação, em sua maioria, mantiveram-se abaixo de 15,0 uC. Isso pode demonstrar a capacidade do modelo na geração de valores relativamente acurados para as estações de pequeno porte e sua potencial aplicação para estações de tratamento de água.

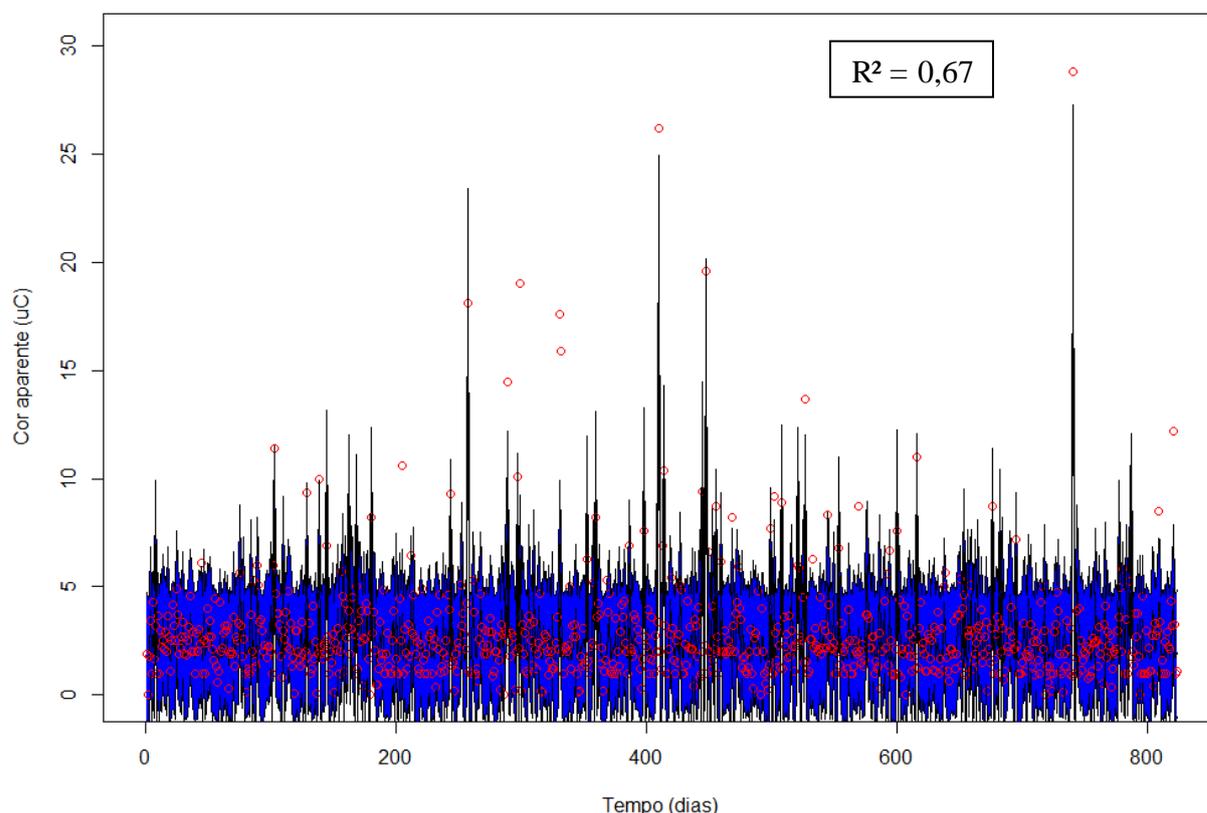


FIGURA 18: Intervalo de predição gerado a partir dos valores preditos de cor aparente da água tratada para ETAs de pequeno porte

No Modelo 2, conforme ilustrado na Tabela 11, seu comportamento também se apresentou de forma esperado, com valores de $R^2 = 0,67$, o que evidencia a relação de causalidade entre os preditores e a variável de interesse, bem como a capacidade explicativa do modelo frente aos valores de cor aparente observados. Observou-se que no modelo há concordância entre os valores observados e calculados, com NSE acima de 0,65. Com relação ao erro, constata-se a baixa variabilidade com valor de RMSE próximo de 1,4 uC durante o teste. Tendo em vista que o valor máximo permissível de cor aparente é 15,0 uC, esse erro representa baixo impacto na estimativa deste parâmetro, particularmente quando do atendimento ao padrão de potabilidade. Outra informação relevante é o fato da possibilidade de superestimação do modelo, com os valores de PBIAS positivo. De modo geral, verifica-se que, apesar da existência de superestimação, o Modelo 2 possui indícios de médio desempenho para a finalidade a que se destina, além abrir uma série de possibilidades para sua aplicação nas mais diversas áreas de aplicação.

TABELA 11: Resultados das métricas durante a elaboração do Modelo 2

MÉTRICAS	TREINAMENTO	TESTE
MSE	2,35	2,04
RMSE	1,53	1,43
NRMSE (%)	62,40	57,10
PBIAS (%)	-5,90	2,80
NSE	0,61	0,67
mNSE	0,36	0,35
r	0,79	0,82
R ²	0,62	0,67
bR ²	0,48	0,59

Legenda: MSE – Erro médio quadrático, RMSE – Raiz do erro médio quadrático, NRMSE (%) – Raiz do erro médio quadrático normalizada, PBIAS (%) – Percentagem de polarização, NSE – Eficiência de Nash-Sutcliffe, mNSE – Eficiência de Nash-Sutcliffe modificada, r – Coeficiente de correlação de Pearson, R² – Coeficiente de determinação, bR² – Coeficiente de determinação (R²) multiplicado pela inclinação da linha de regressão.

4.2 Modelos para Estações de Tratamento de Médio Porte

As etapas utilizadas na elaboração dos modelos para ETAs de médio porte (Modelo 3 e Modelo 4) foram similares às mencionadas anteriormente para os modelos de pequeno porte. Na proposição destes modelos foram utilizadas 17 variáveis, com 3.343 observações de cada, totalizando 56.831 dados de três ETAs.

4.2.1 Análise exploratória para os modelos de ETAs de médio porte

Como forma de avaliar o comportamento dos dados, foi aplicada a estatística exploratória, a qual pode ser observada na Tabela 12. Inicialmente, verifica-se que as vazões afluentes às ETAs estão próximas de 260 L.s⁻¹, embora haja grande variabilidade dessa variável, assim como com a produção de lodo (ProdLodo) e o tempo de floculação (TemFloc), sendo esse o que apresenta maior variabilidade dos dados. Em se tratando de floculação, os valores do gradiente de velocidade se mantiveram acima do recomendado pela ABNT NBR 12.2016/1992, que estabelecem valor de até 70 s⁻¹ (ABNT, 1992). Essas variáveis estão relacionadas aos aspectos hidráulicos de dimensionamento das ETAs, bem como a qualidade da água bruta afluente às mesmas. Também é verificado que não há variabilidade da variável EnsJartest, tendo em vista que todas as estações comumente realizam o ensaio em *Jar Test*. Com relação à qualidade da água tratada, é possível constatar que, de modo geral, as estações atenderam ao padrão de potabilidade para os parâmetros cor aparente e turbidez, apresentando apenas algumas violações isoladas ao longo do período de monitoramento.

TABELA 12: Apresentação dos resultados da estatística exploratória das variáveis referentes às estações de médio porte

VAR	MÉDIA	MEDIANA	MODA	MINIMO	MÁXIMO	DESV.PAD.
Qaf (L.s⁻¹)	262,55	329,75	70,00	39,00	663,25	151,24
TempOp (h)	22,37	24,00	24,00	7,50	24,00	2,06
CorB (uC)	79,22	45,50	42,75	3,70	2469,90	55,48
TuB (uT)	38,98	9,82	6,00	0,18	1441,00	45,20
pHB	6,59	6,45	6,60	5,21	8,91	0,32
DosCoag (mg.L⁻¹)	14,62	10,00	10,00	2,00	189,00	9,83
ProdLodo (kg.dia⁻¹)	788,64	447,12	579,07	61,56	15573,12	616,94
InstOP	2,42	2,00	2,00	2,00	3,00	0,49
EnsJartest	1,00	1,00	1,00	1,00	1,00	0,00
GradFloc (s⁻¹)	94,16	105,77	25,06	10,42	301,73	52,15
TemFloc (s)	1277,62	1304,47	1434,73	624,81	2819,65	166,58
VelSed (cm.min⁻¹)	1,69	1,74	1,47	0,30	3,49	0,44
VelLong (cm.s⁻¹)	0,32	0,30	0,51	0,05	0,71	0,10
Qcalha (L.s⁻¹.m¹)	1,66	1,61	2,33	0,33	3,22	0,42
TxFilt (m³.m².dia⁻¹)	228,43	249,64	150,78	40,39	502,32	80,05
CorT (uC)	2,67	2,40	5,00	0,01	20,00	1,50
TuT (uT)	0,46	0,34	0,10	0,01	7,10	0,30

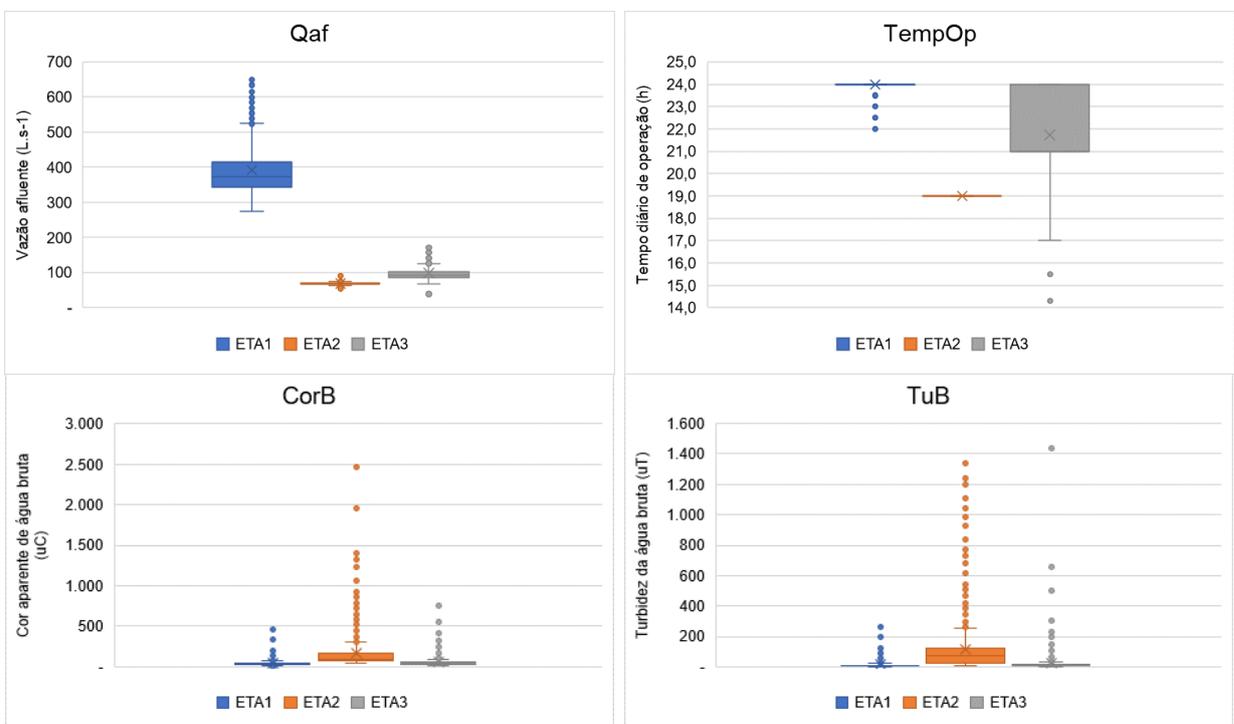
Para a verificação da distribuição dos dados, também foi aplicado o teste de normalidade, constatando-se que não há evidência de rejeitar a hipótese nula de normalidade ao nível de significância estatisticamente aceitável ($\alpha=0,05$).

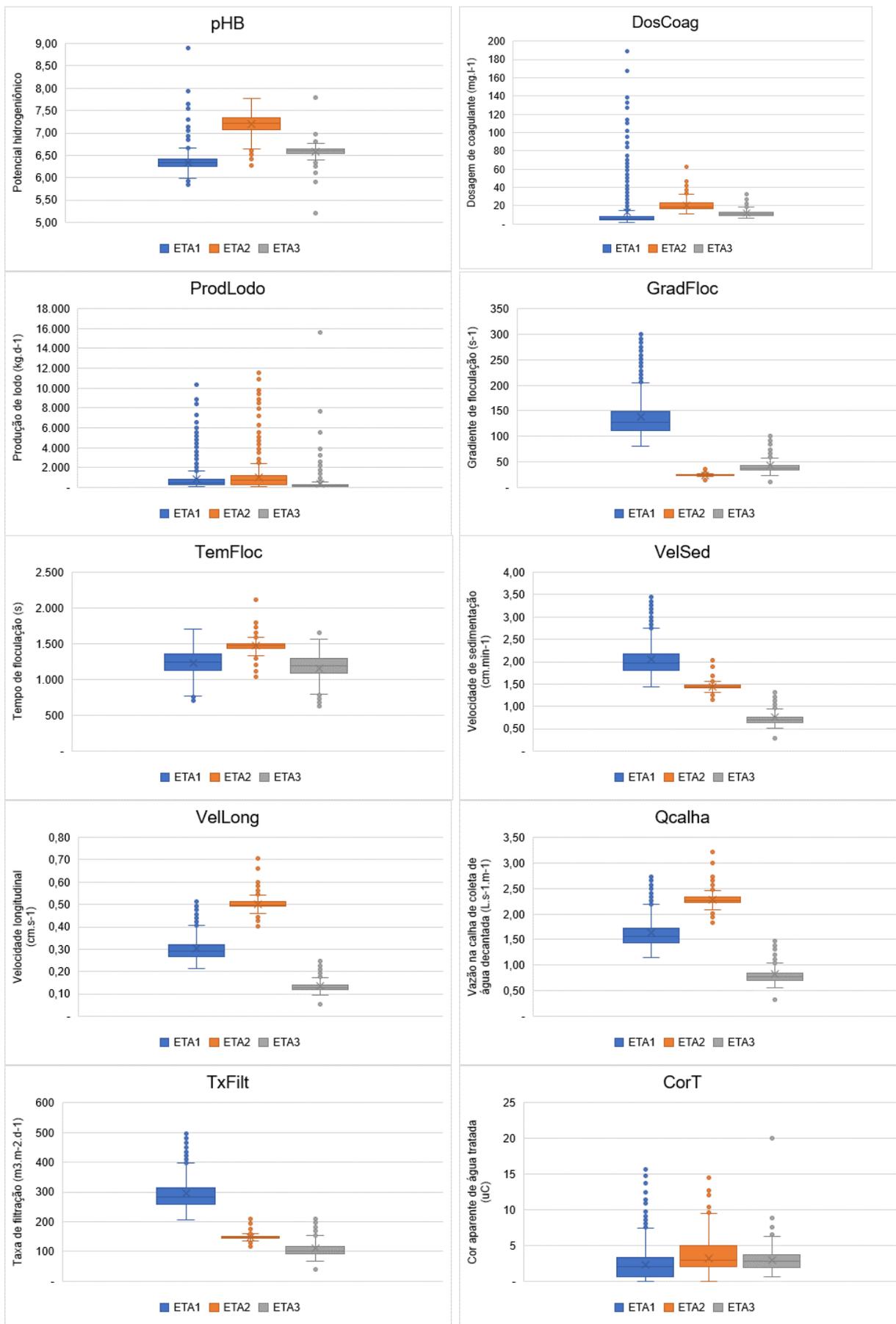
A Figura 19 apresenta o comportamento das variáveis que compõem os modelos na três ETAs de médio porte utilizadas no desenvolvimento dessa pesquisa. É possível observar que todos os parâmetros e em todas as estações de tratamento de água apresentam *outliers*. Nesse caso, optou-se por mantê-los, tendo em vista que esse comportamento é condizente com a realidade do funcionamento dessas estações, por diversos motivos. Dentre estes, a qualidade das águas que afluem às ETAs, os aspectos operacionais, aspectos de dimensionamento das unidades, qualidade dos produtos químicos, entre outros.

Em se tratando dos indicadores hidráulicos representados no modelo, observa-se que a maioria dos parâmetros apresenta grande diferença entre as ETA. Neste sentido, pode-se destacar a vazão, que na ETA1 é bem maior que as demais; corrobora ainda com esse fato os

parâmetros gradiente de velocidade de floculação, velocidade de sedimentação e longitudinal no decantador, vazão na calha de coleta de água decantada e a taxa de filtração. Já para os parâmetros de qualidade de água, observa-se que a água bruta apresenta valores relativamente próximos de pH, turbidez e cor parente, nas três estações, com exceção da ETA2, na qual a presença de valores extremos para esses dois últimos parâmetros é maior. Esse comportamento pode estar relacionado aos efeitos dos fenômenos de precipitação, a montante, na bacia hidrográfica onde o manancial está localizado, bem como as atividades antrópicas nela desenvolvidas.

No que concerne ao atendimento da norma técnica para projetos de estações de tratamento de água, pode-se observar que algumas etapas não funcionam de acordo com que é estabelecido. Por exemplo, o gradiente de velocidade da floculação (GradFloc) na ETA1, em todo o período de monitoramento analisado, se manteve acima de 70 s^{-1} (valor máximo admitido pela ABNT NBR 12.216/1992). Com relação à taxa média de filtração (TxFilt), segundo a mesma norma, para filtros rápidos de camada simples os valores devem ser de $180 \text{ m}^3 \cdot \text{m}^{-2} \cdot \text{d}^{-1}$ e, para camada dupla, $360 \text{ m}^3 \cdot \text{m}^{-2} \cdot \text{d}^{-1}$. Nesse caso a ETA1 também foi a única que apresentou valores superiores aos recomendados, o que pode ensejar o aumento da perda de carga, maior desgaste do meio filtrante, perspectiva de redução da duração das carreiras de filtração, entre outras consequências.





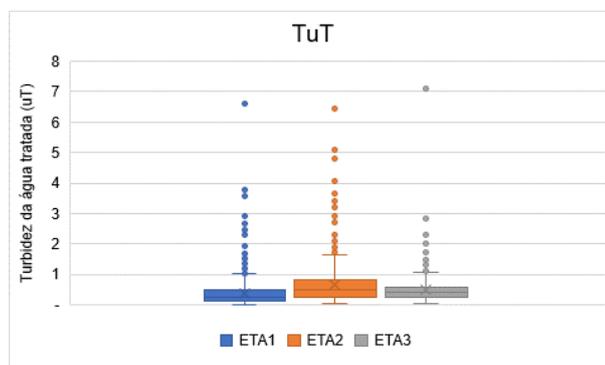


FIGURA 19: Composição de gráfico boxplot de cada uma das variáveis que compõem os modelos para as ETAs de médio porte

Para os parâmetros de qualidade da água na saída das ETAs, constata-se que eles também são bastante próximos e, em sua maioria, se apresentam dentro do Padrão de Potabilidade ao longo de período de monitoramento analisado. Com relação à cor aparente, só alguns valores extremos ultrapassaram o valor máximo permissível de 15,0 uC. Para o parâmetro turbidez, todas as ETAs apresentaram violação, tendo em vista que a Portaria de Consolidação nº 5/2017 permite um valor máximo de 0,5 uT para água filtrada, após filtração rápida. Em relação a tais violações, cabe destacar que quando parcela dos dados operacionais foi gerada vigia o limite máximo de turbidez de 1,0 uT, consoante a Portaria 518/2004. Fato também constatado para as ETAs de pequeno e grande porte.

4.2.2 Escolha das variáveis para compor os modelos de ETAs de médio porte

A escolha inicial das variáveis para a composição dos modelos de regressão linear múltipla para estações de médio porte seguiu o mesmo processo para as ETAs de pequeno porte. Foi aplicada a análise de correlação de Pearson, ao nível de significância de 0,05, a partir do seu resultado, tomando como base o p-valor, foram observadas as variáveis que apresentaram correlação acima de 10% com as variáveis dependentes dos modelos (cor aparente e turbidez da água tratada). A Tabela 13 ilustra os valores de da correlação em as variáveis, com destaque para aqueles que apresentam valores acima de 10%, que serão utilizadas na elaboração dos modelos.

TABELA 13: Matriz de correlação entre as variáveis utilizadas no desenvolvimento dos modelos para ETAs de médio porte

	<i>Qaf</i>	<i>TempOp</i>	<i>CorB</i>	<i>TuB</i>	<i>pHB</i>	<i>DosCoag</i>	<i>ProdLodo</i>	<i>InstOP</i>	<i>GradFloc</i>	<i>TemFloc</i>	<i>VelSed</i>	<i>VelLong</i>	<i>Qcalha</i>	<i>TxFilt</i>	<i>CorT</i>	<i>TuT</i>
<i>Qaf</i>	1															
<i>TempOp</i>	0,84	1														
<i>CorB</i>	-0,30	-0,37	1													
<i>TuB</i>	-0,33	-0,40	0,90	1												
<i>pHB</i>	-0,74	-0,83	0,24	0,28	1											
<i>DosCoag</i>	-0,11	-0,18	0,27	0,25	0,19	1										
<i>ProdLodo</i>	0,05	-0,05	0,69	0,75	-0,01	0,44	1									
<i>InstOP</i>	-0,94	-0,86	0,31	0,33	0,75	0,13	-0,02	1								
<i>GradFloc</i>	0,98	0,79	-0,28	-0,31	-0,70	-0,10	0,07	-0,86	1							
<i>TemFloc</i>	-0,54	-0,48	0,20	0,23	0,46	0,10	-0,01	0,29	-0,66	1						
<i>VelSed</i>	0,87	0,54	-0,13	-0,15	-0,40	-0,01	0,15	-0,77	0,88	-0,42	1					
<i>VelLong</i>	-0,13	-0,51	0,30	0,32	0,59	0,20	0,21	0,20	-0,09	0,23	0,37	1				
<i>Qcalha</i>	0,04	-0,37	0,25	0,26	0,47	0,18	0,23	0,06	0,09	0,09	0,52	0,98	1			
<i>TxFilt</i>	0,97	0,74	-0,23	-0,26	-0,60	-0,07	0,10	-0,88	0,97	-0,54	0,96	0,10	0,27	1		
<i>CorT</i>	-0,10	-0,19	0,17	0,14	0,12	0,23	0,19	0,20	-0,04	-0,11	0,00	0,18	0,17	-0,05	1	
<i>TuT</i>	-0,17	-0,24	0,41	0,44	0,13	0,12	0,38	0,22	-0,14	0,03	-0,06	0,20	0,18	-0,12	0,43	1

Conforme os valores em destaque na supracitada tabela, as variáveis escolhidas para os modelos são:

- Turbidez da água tratada (TuT) – Qaf, TempOp, CorB, TuB, pHB, DosCoag, ProdLodo, InstOP, GradFloc, VeLong, Qcalha e TxFilt; e
- Cor aparente da água tratada (CorT) – Qaf, TempOp, CorB, TuB, pHB, DosCoag, ProdLodo, InstOP, TemFloc, VeLong e Qcalha.

4.2.3 Modelos em regressão linear múltipla para estações de tratamento de médio porte

O desenvolvimento dos modelos em regressão linear múltipla para as estações de tratamento de água de médio porte deu início a partir do modelo de turbidez da água tratada. Seguindo os mesmos procedimentos utilizados para as estações de pequeno porte, foi aplicada a regressão linear múltipla ao nível de significância de 5%. As tabelas 14, 15 e 16 ilustram os resultados do teste e, posteriormente a equação que resume o modelo.

Na Tabela 14 é apresentada a análise de variância do modelo. Inicialmente, é possível verificar os valores do Qui-quadrado utilizados no ajuste do modelo; em seguida observa-se a contribuição de cada variável para a predição da turbidez, destacando-se a TuB como a variável de maior contribuição (16,33%). Posteriormente, são apresentados os valores do Qui-quadrado ajustado e o do quadrado médio, utilizados na medição do erro; por fim, constatam-se os valores de F e P, com destaque para o p-valor que indica se a variável é significativa ou não para o modelo.

Com relação à capacidade de predição do modelo, é percebido que o mesmo é bastante impreciso, apresentando coeficiente de determinação ajustado de 0,34 (Tabela 15), o que, a princípio, poderia inviabilizar o uso do modelo, aliado ao valor de erro acima de 77%. A Tabela 15 apresenta um resumo breve do resultado do modelo, com destaque para a medida de desvio padrão, os coeficientes de determinação, notadamente o R² ajustado que representa a qualidade do modelo com relação ao ajuste dos resultados dos valores preditos.

TABELA 14: Análise de variância gerada para o modelo de regressão em ETAs de médio porte a partir da turbidez da água tratada

Fonte	GL	SQ Seq	Contribuição	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	5	185,347	22,30%	185,347	37,0694	191,54	0,000
Qaf	1	25,099	3,02%	5,218	5,2181	26,96	0,000
TuB	1	135,730	16,33%	17,169	17,1694	88,72	0,000
ProdLodo	1	7,764	0,93%	6,871	6,8711	35,50	0,000
InstOP	1	15,551	1,87%	11,966	11,9660	61,83	0,000
Qcalha	1	1,202	0,14%	1,202	1,2021	6,21	0,013
Erro	3337	645,809	77,70%	645,809	0,1935		
Falta de ajuste	3309	645,809	77,70%	645,809	0,1952	*	*
Erro puro	28	0,000	0,00%	0,000	0,0000	*	*
Total	3342	831,156	100,00%				

Legenda: GL – Graus de liberdade, SQ Seq – Soma dos quadrados sequenciais, Contribuição – Percentual que cada variável representa para o modelo, SQ (Aj.) – Soma dos quadrados ajustados, QM (Aj.) – Quadrado médio ajustado, Valor F – Estatística utilizada para verificar se o modelo está associado com a resposta, Valor-P – Mede a evidência contrária a hipótese nula do teste, Erro puro – erro calculado a partir do número de repetições de observações utilizadas no modelo, Falta de ajuste – calcula a falta de ajuste gerada pela presença de réplicas no conjunto de dados.

TABELA 15: Sumário do modelo para turbidez da água tratada em ETAs de médio porte

S	R ²	R ² (aj)	PRESQ	R ² (pred)
0,439920	22,30%	22,18%	651,543	21,61%

Legenda: S – Desvio padrão, R² – Coeficiente de determinação, R²(aj) – Coeficiente de determinação ajustado, PRESQ – Soma dos quadrados preditos do erro (calcula o desvio do modelo), R²(pred) – Coeficiente de determinação predito (determina o quanto o modelo prediz para novas observações).

A Tabela 16 apresenta os coeficientes utilizados para a elaboração da equação que descreve o modelo. Inicialmente, são ilustrados os valores dos coeficientes para cada variável independente que compõem o modelo. O destaque dessa tabela são os valores dos intervalos de confiança de 95%, apresentando a variabilidade dos coeficientes para garantir o R² ajustado. Na última coluna são apresentados os valores inflacionados da variância. Neste caso, observa-se que estão abaixo de 10, atendendo à inexistência da multicolinearidade para o uso do modelo. Por fim, é apresentada a equação linear que escreve o modelo de predição para a turbidez da água tratada.

TABELA 16: Coeficientes que compõem o modelo de estações de médio porte para turbidez da água tratada

Termo	Coef	EP de Coef	IC de 95%	Valor-T	Valor-P	VIF
Constante	-0,826	0,146	(-1,113; -0,540)	-5,65	0,000	
Qaf	0,000777	0,000150	(0,000484; 0,001070)	5,19	0,000	10,30
TuB	0,001398	0,000148	(0,001107; 0,001689)	9,42	0,000	3,51
ProdLodo	0,000070	0,000012	(0,000047; 0,000093)	5,96	0,000	3,01
InstOP	0,3752	0,0477	(0,2817; 0,4688)	7,86	0,000	9,55
Qcalha	0,0400	0,0160	(0,0085; 0,0714)	2,49	0,013	1,20

Legenda: Termo – Elementos que irão compor o modelo, Coef – Coeficientes do modelo (relação entre as variáveis dependentes e independentes), EP de Coef – Erro padrão dos coeficientes, IC de 95% - Intervalo de confiança ao nível de significância de 95%, Valor-T – Razão entre o coeficiente e o respectivo erro padrão, Valor-P – Mede a evidência contra a hipótese nula, VIF – Fator de inflação da variância (demonstra como a variância de cada coeficiente eleva-se pela correlação entre as variáveis independentes ou preditoras).

Equação de Regressão

$$TuT = -0,826 + 0,000777 Qaf + 0,001398 TuB + 0,000070 ProdLodo + 0,3752 InstOP + 0,0400 Qcalha$$

Para a elaboração do modelo de regressão linear da cor aparente da água tratada foram realizados os mesmos procedimentos do modelo para turbidez, com exceção do número de variáveis independentes e a necessidade de transformação dos dados (logaritmo na base 10). A Tabela 17 ilustra os resultados da análise de variância do modelo, na qual se verifica que a variável de maior contribuição para o modelo é o grau de instrução do operador da ETA, com 6,14%. É possível constatar que, com o baixo valor de F (0,12), há evidência de que não há relação entre essa preditora e a variável de controle do modelo.

Como resultado do modelo proposto, observa-se na Tabela 18 que o valor do coeficiente de determinação é muito baixo ($R^2 = 0,17$), inclusive inferior ao encontrado para o modelo de turbidez da água tratada, evidenciando a baixa capacidade de predição do mesmo.

Na Tabela 19 são apresentados os coeficientes de cada variável independente, os intervalos de confiança e, principalmente o VIF, no qual se verifica que apenas a variável instrução do operador (InstOP) apresenta valor pouco superior a 10. Mesmo assim optou-se por manter a variável devido a sua importância para o modelo, em se tratando de aspectos ligados à operação das ETAs. Posteriormente, é apresentada a equação representativa do modelo de

regressão linear múltipla para a cor aparente da água tratada. Nela é possível verificar o termo independente e os coeficientes de cada variável.

TABELA 17: Análise de variância gerada para o modelo de regressão em ETAs de médio porte a partir da cor aparente da água tratada

Fonte	GL	SQ Seq	Contribuição	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	8	2296,6	17,41%	2296,6	287,078	87,87	0,000
Qaf	1	141,3	1,07%	625,4	625,437	191,44	0,000
TempOp	1	465,1	3,53%	74,5	74,505	22,80	0,000
CorB	1	160,9	1,22%	88,4	88,403	27,06	0,000
TuB	1	61,4	0,47%	151,7	151,652	46,42	0,000
pHB	1	26,5	0,20%	62,8	62,763	19,21	0,000
DosCoag	1	444,4	3,37%	175,0	174,983	53,56	0,000
ProdLodo	1	187,2	1,42%	134,0	134,034	41,03	0,000
InstOP	1	809,9	6,14%	809,9	809,862	247,89	0,000
Erro	3334	10892,5	82,59%	10892,5	3,267		
Falta de ajuste	3307	10889,1	82,56%	10889,1	3,293	26,13	0,000
Erro puro	27	3,4	0,03%	3,4	0,126	*	*
Total	3342	13189,1	100,00%				

Legenda: GL – Graus de liberdade, SQ Seq – Soma dos quadrados sequenciais, Contribuição – Percentual que cada variável representa para o modelo, SQ (Aj.) – Soma dos quadrados ajustados, QM (Aj.) – Quadrado médio ajustado, Valor F – Estatística utilizada para verificar se o modelo está associado com a resposta, Valor-P – Mede a evidência contrária a hipótese nula do teste, Erro puro – erro calculado a partir do número de repetições de observações utilizadas no modelo, Falta de ajuste – calcula a falta de ajuste gerada pela presença de réplicas no conjunto de dados.

TABELA 18: Sumário do modelo para cor aparente da água tratada em ETAs de médio porte

S	R ²	R ² (aj)	PRESQ	R ² (pred)
1,80751	17,41%	17,21%	11014,3	16,49%

Legenda: S – Desvio padrão, R² – Coeficiente de determinação, R²(aj) – Coeficiente de determinação ajustado, PRESQ – Soma dos quadrados preditos do erro (calcula o desvio do modelo), R²(pred) – Coeficiente de determinação predito (determina o quanto o modelo prediz para novas observações).

TABELA 19: Coeficientes que compõem o modelo de estações de médio porte para cor aparente da água tratada

Termo	Coef	EP de Coef	IC de 95%	Valor-T	Valor-P	VIF
Constante	0,18	1,72	(-3,19; 3,55)	0,10	0,917	
Qaf	0,008180	0,000591	(0,007021; 0,009339)	13,84	0,000	9,52
TempOp	-0,1664	0,0348	(-0,2347; -0,0981)	-4,78	0,000	6,19
CorB	0,003073	0,000591	(0,001915; 0,004231)	5,20	0,000	5,58
TuB	-0,006290	0,000923	(-0,008101; -0,004480)	-6,81	0,000	8,06
pHB	-0,643	0,147	(-0,931; -0,356)	-4,38	0,000	3,48
DosCoag	0,01851	0,00253	(0,01355; 0,02347)	7,32	0,000	1,42
ProdLodo	0,000350	0,000055	(0,000243; 0,000457)	6,41	0,000	3,86
InstOP	3,216	0,204	(2,815; 3,616)	15,74	0,000	10,36

Legenda: Termo – Elementos que irão compor o modelo, Coef – Coeficientes do modelo (relação entre as variáveis dependentes e independentes), EP de Coef – Erro padrão dos coeficientes, IC de 95% - Intervalo de confiança ao nível de significância de 95%, Valor-T – Razão entre o coeficiente e o respectivo erro padrão, Valor-P – Mede a evidência contra a hipótese nula, VIF – Fator de inflação da variância (demonstra como a variância de cada coeficiente eleva-se pela correlação entre as variáveis independentes ou preditoras).

Equação de Regressão

CorT =	$0,18 + 0,008180 \text{ Qaf} - 0,1664 \text{ TempOp} + 0,003073 \text{ CorB} - 0,006290 \text{ TuB}$ $- 0,643 \text{ pHB} + 0,01851 \text{ DosCoag} + 0,000350 \text{ ProdLodo} + 3,216 \text{ InstOP}$
--------	---

4.2.4 Modelo em RNA de estações de tratamento de médio porte para a turbidez da água tratada

A Figura 20 ilustra a configuração da RNA para estações de tratamento de água de médio porte. Nessa RNA são observadas duas camadas ocultas, com cinco neurônios cada, bem como os pesos de cada neurônio. Essa topologia de rede, devido ao reduzido número de neurônio nas camadas ocultas, permite um processamento mais rápido do modelo, reduzindo o tempo de treinamento e simulação.

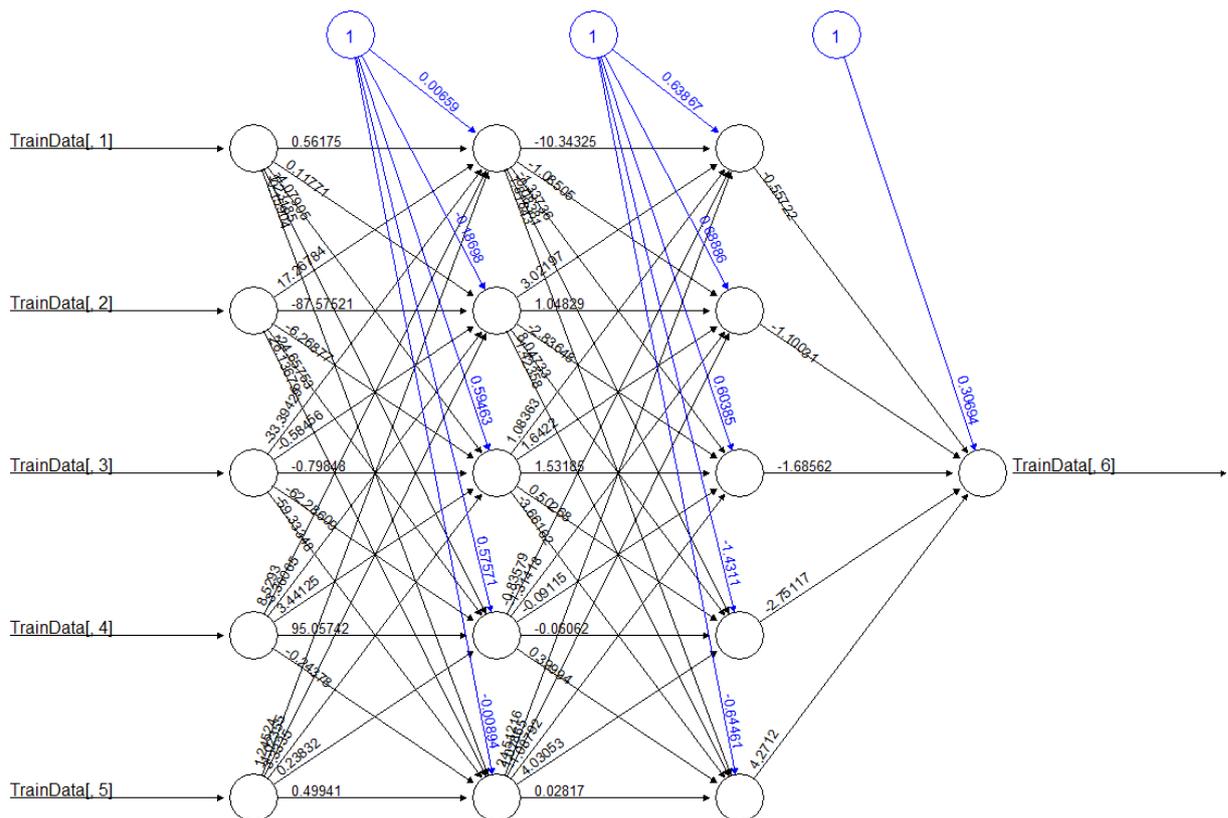


FIGURA 20: Rede neural gerada a partir do modelo proposto para a turbidez da água tratada em ETAs de médio porte

Quanto à etapa de treinamento da RNA, observa-se nas Figura 21 e 22 que a predição do modelo não ocorreu como o esperado, no que condiz à predição dos valores de turbidez da água tratada. Na Figura 21 há grande flutuação nos valores preditos (em vermelho) quando comparados com os valores observados (em preto), particularmente nos níveis mais baixos de turbidez, na qual se constata que o modelo subestimou os valores observados. Esse comportamento durante a etapa de treinamento já sinaliza a baixa capacidade do modelo para a predição.

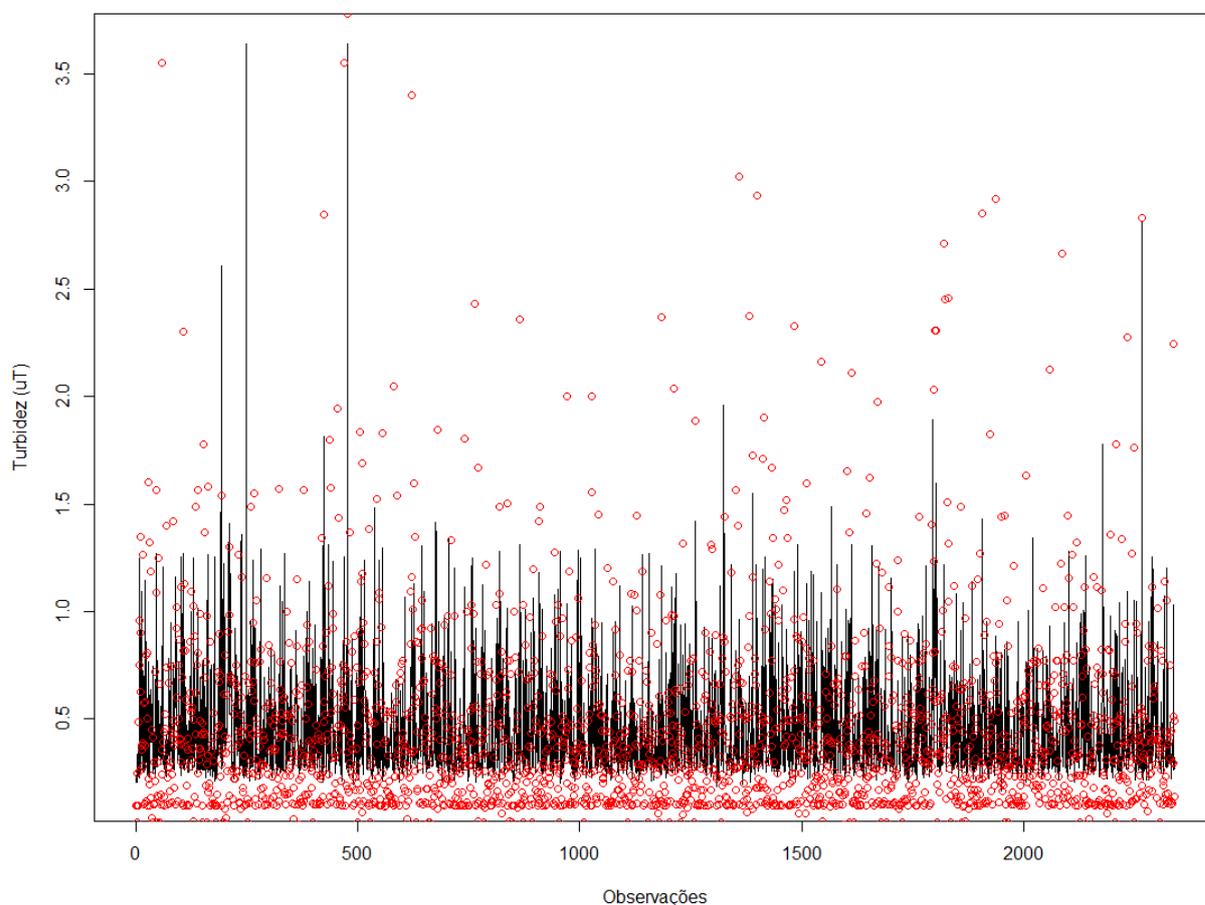


FIGURA 21: Comparação entre o comportamento dos valores de turbidez da água tratada durante o treinamento do modelo das estações de médio porte

Na Figura 22 observa-se a não aderência do modelo de regressão à reta de 45°, apresentando coeficiente de determinação de 0,32 para a aplicação pretendida. Essa dificuldade de ajuste, embora se perceba uma relação entre a variável observada e calculada, pode ser explicada pela variabilidade dos parâmetros que compõem o modelo entre as ETAs. Por se tratar de estações de tratamento de água localizadas em áreas diferentes, onde os mananciais e as bacias hidrográficas onde os mesmos estão inseridos apresentam características distintas, a influência de fenômenos de precipitação também ocorre de maneira desigual. Outro fato importante são as condições operacionais de cada ETA, tendo em vista seu tempo de funcionamento, as necessidades ou não de ampliação dos sistemas, a substituição de alguns materiais, tais como meio filtrante dos filtros, entre outros aspectos operacionais que influenciam diretamente na qualidade da água tratada.

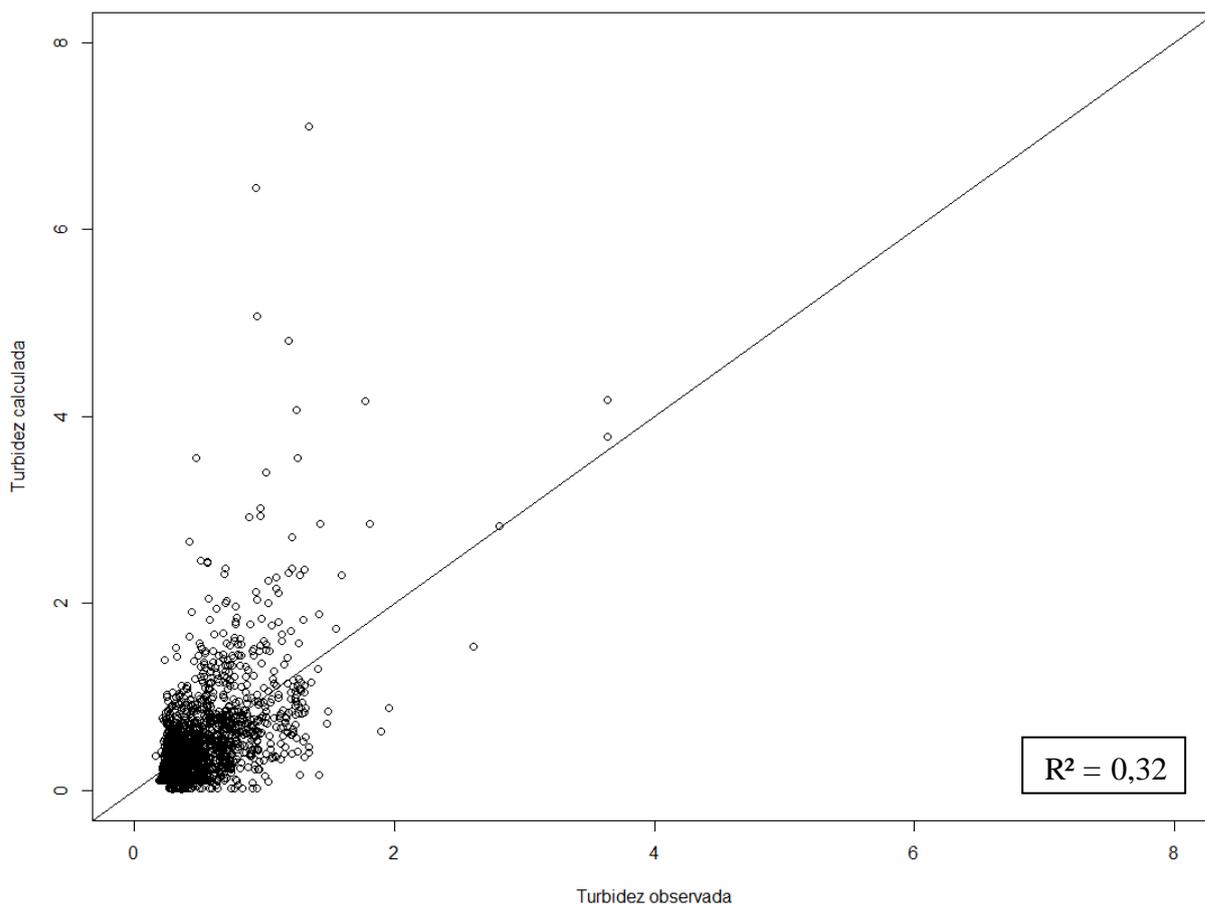


FIGURA 22: Comportamento da correlação entre a turbidez calculada e observada durante o treinamento da rede neural para estações de médio porte

O reflexo do comportamento do Modelo 3 durante a fase de treinamento é notório no seu resultado, conforme ilustrado na Figura 23, com $R^2 = 0,21$ (valor inferior ao encontrado para as mesmas variáveis no modelo em regressão linear múltipla), o que enseja na baixa aderência do modelo e na capacidade de predição. Embora a referida figura não consiga ilustrar com tanta precisão, o intervalo de predição admite muitos valores de turbidez negativa, o que não reflete a realidade. Aliado ao exposto, a amplitude do intervalo de predição é relativamente grande, influenciando na capacidade de predição do modelo, ensejando na sua baixa acurácia. Esse fato pode ser explicado pelo treinamento não ter ocorrido da maneira esperada. Outro fato interessante é a turbidez predita, em sua maioria, se manteve abaixo do valor observado e do preconizado pelo padrão de potabilidade, indicando viés de subestimação do modelo. Mesmo quando a normalização dos dados para a elaboração do modelo foi utilizada, não houve eficiência considerável com relação à elevação do coeficiente de determinação.

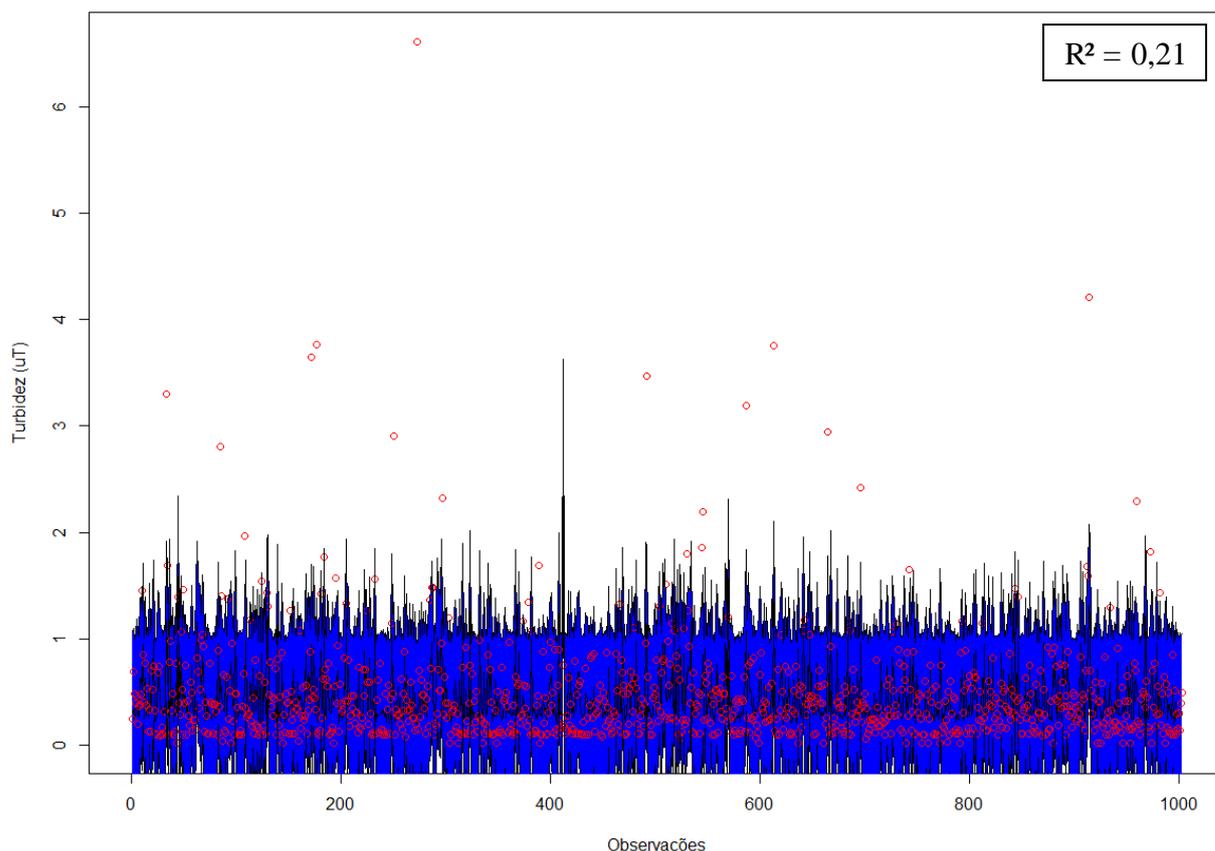


FIGURA 23: Intervalo de predição gerado a partir dos valores preditos de turbidez da água tratada para ETAs de médio porte

De maneira geral, os resultados do Modelo 3 indicam baixo desempenho do mesmo. Ao analisar as métricas para o Modelo (Tabela 20), observam-se indícios de falta de ajuste do modelo, e baixa concordância entre valores preditos e observados, tendo em vista que o coeficiente de Nesh-Scutcliffe foi de 0,32 para o treinamento e 0,21 para o teste (bem distante de 1, que seria o desejado para um modelo com bom desempenho). Também é possível verificar a existência de viés de subestimação no modelo, pelo fato os valores do PBIAS serem negativos, embora durante a etapa de treinamento tenha se mantido próximo de 0 (que dá indício de um modelo perfeito). Um aspecto relevante é que o erro foi relativamente baixo, com valores abaixo de 0,45 uT (RMSE) e MSE de 0,20 para o teste.

TABELA 20: Resultados das métricas durante a elaboração do Modelo 3

MÉTRICAS	TREINAMENTO	TESTE
MSE	0,17	0,20
RMSE	0,41	0,45
NRMSE (%)	82,20	88,90
PBIAS (%)	-0,10	-3,50
NSE	0,32	0,21
mNSE	0,18	0,17
r	0,57	0,46
R ²	0,32	0,21
bR ²	0,21	0,12

Legenda: MSE – Erro médio quadrático, RMSE – Raiz do erro médio quadrático, NRMSE (%) – Raiz do erro médio quadrático normalizada, PBIAS (%) – Percentagem de polarização, NSE – Eficiência de Nash-Sutcliffe, mNSE – Eficiência de Nash-Sutcliffe modificada, r – Coeficiente de correlação de Pearson, R² – Coeficiente de determinação, bR² – Coeficiente de determinação (R²) multiplicado pela inclinação da linha de regressão.

4.2.5 Modelo em RNA de estações de tratamento de médio porte para a cor aparente da água tratada

O desenvolvimento do modelo para cor aparente da água tratada (Modelo 4) encontra-se ilustrado na Figura 24. Nesse modelo utilizaram-se três neurônios em cada uma das camadas ocultas. Conforme já mencionado, a topologia a que se chegou nos modelos levou em consideração as diversas tentativas de interação das redes neurais artificiais, tanto no treinamento, como geração do intervalo de predição, na busca de valores de R² cada vez mais altos, indicando a possibilidade de aplicação dos modelos. Nessa topologia também se conseguiu reduzir os esforços computacionais, criando um modelo mais simplificado.

A Figura 25 ilustra o comportamento dos parâmetros de controle (cor aparente da água tratada) observados e calculados durante a etapa de treinamento do modelo. De modo similar ao ocorrido no Modelo 3, houve subestimação da cor aparente, sendo que os menores valores preditos foram inferiores aos observados. Essa característica encontrada na concepção desses modelos pode estar relacionada a uma falta de ajuste no mesmo, sendo necessária uma compreensão maior do fenômeno, bem como uma revisão mais acurada dos métodos empregados. Ainda com relação à cor aparente, verifica-se que a maioria dos registros está abaixo de 15,0 uC, conforme preconizado pelo Padrão de Potabilidade em vigor (BRASIL, 2017). Também se observa na referida figura que os valores de acima de 5,0 uC apresentam uma variabilidade maior que os demais, havendo uma concentração menor desses registros.

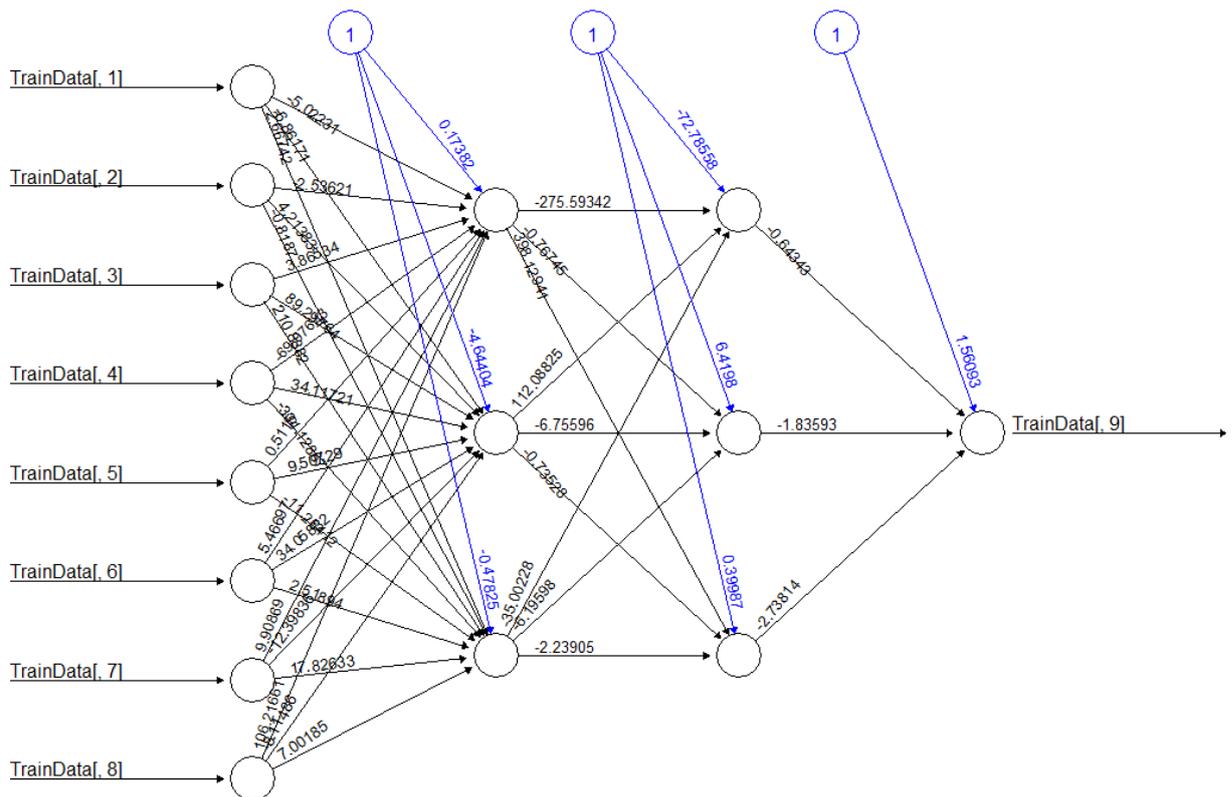


FIGURA 24: Rede neural gerada a partir do modelo proposto para a cor aparente da água tratada em ETAs de médio porte

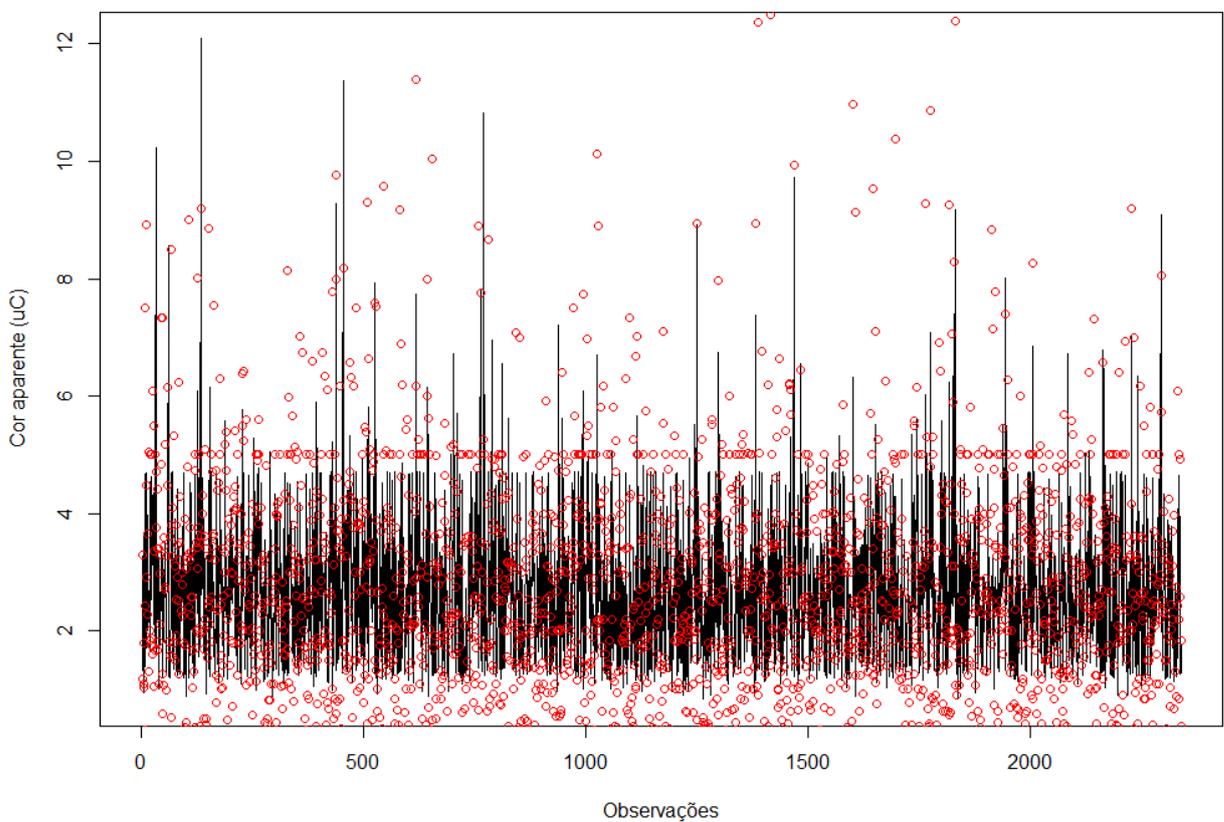


FIGURA 25: Comparação entre o comportamento dos valores de cor aparente da água tratada durante o treinamento do modelo das estações de médio porte

Quando verificada a relação entre os valores observados e calculados da cor aparente durante o treinamento (Figura 26) constata-se que há baixa aderência entre os valores preditos e a linha diagonal em 45°, apresentando R^2 de 0,40.

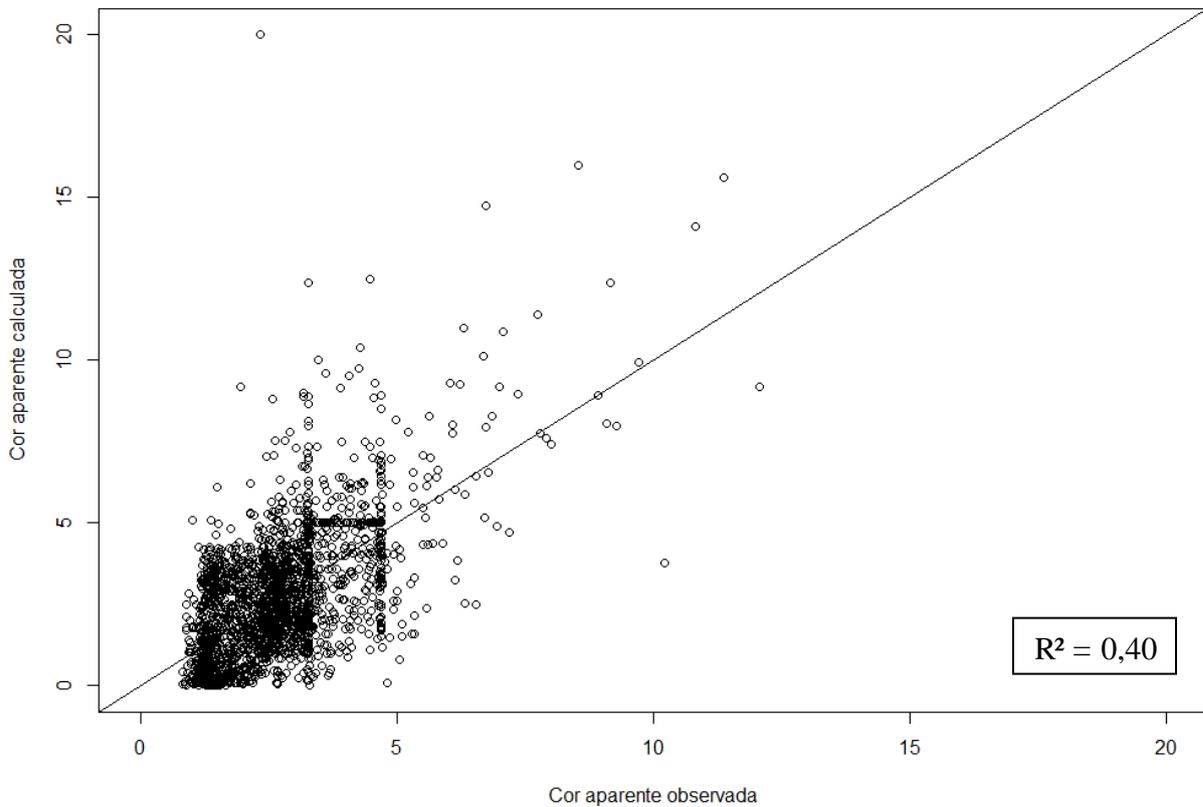


FIGURA 26: Comportamento da correlação entre a cor aparente calculada e observada durante o treinamento da rede neural para estações de médio porte

De forma geral, quando se compara os resultados do Modelo 3 com os do Modelo 4 verifica-se um pequeno ganho deste último com relação ao primeiro, notadamente pelos resultados dos coeficientes de determinação. No entanto, quando se observam os resultados do intervalo de predição do Modelo 4, assim como no Modelo 3, há a geração de valores negativos embora o R^2 do modelo tenha sido de 0,33, o que, evidencia a baixa capacidade do modelo em explicar os valores preditos para a variável de controle (Figura 27). Aliado ao exposto, verifica-se a grande amplitude do intervalo de predição, que justifica a baixa acurácia do modelo, sendo justificada também pela capacidade de treinamento da rede neural.

Mesmo que o Modelo 4 não atenda os anseios para sua aplicação sem a necessidade de vários ajustes, é possível admitir que o mesmo foi capaz de realizar a predição de valores de cor aparente da água tratada dentro dos padrões de potabilidade exigidos pela Portaria de Consolidação nº 5/2017 (BRASIL, 2017), de maneira similar aos valores observados.

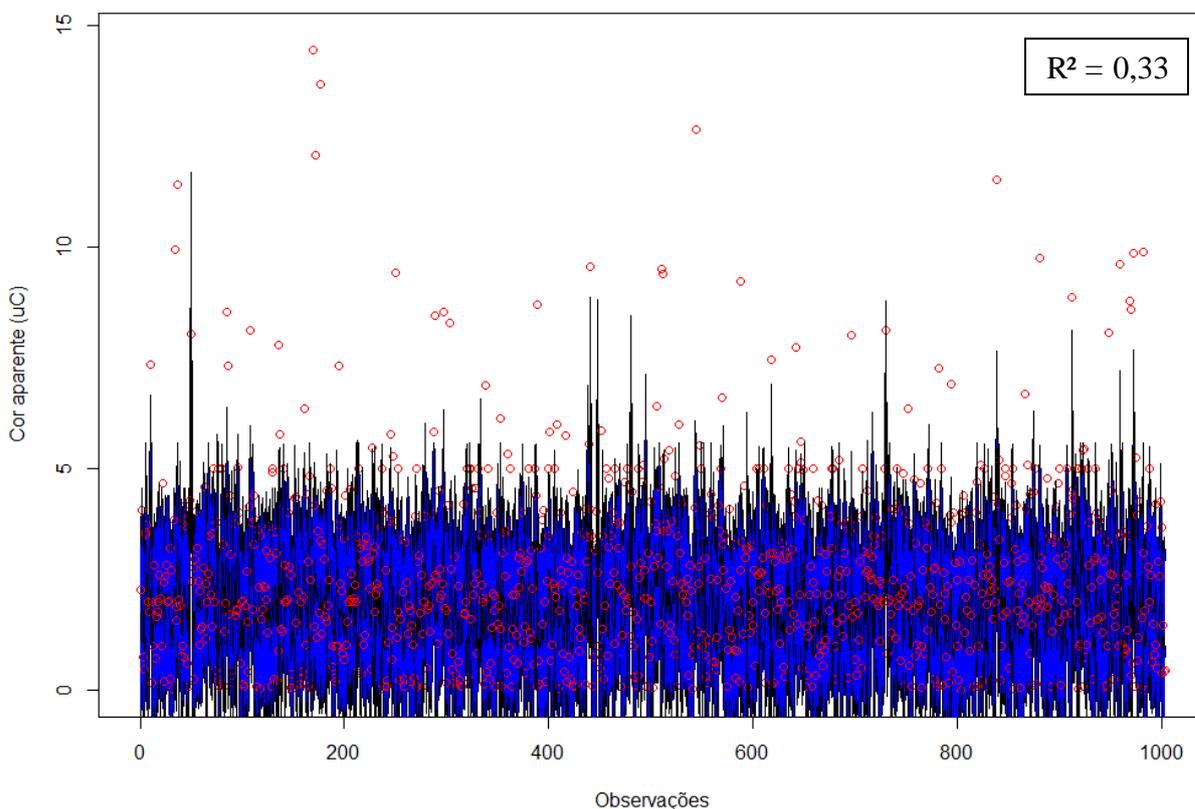


FIGURA 27: Intervalo de predição gerado a partir dos valores preditos de cor aparente da água tratada para ETAs de médio porte

A Tabela 21 ilustra os resultados das análises de desempenho do Modelo 4. Verifica-se que a relação entre os valores preditos e observados é muito reduzida. Esse fato enseja a baixa capacidade de ajuste do modelo, com coeficiente de determinação abaixo de 0,33 na etapa de teste, sendo confirmado por meio do NSE abaixo de 0,18. Esses resultados corroboram com baixa concordância entre os preditores e o valor predito. Já o erro apresentado para o modelo se manteve acima de 1,8 uC, considerado moderado pela oscilação da cor aparente da água tratada. Por outro lado, o PBIAS resultou em valores negativos para a fase de teste, o que atesta o comportamento do modelo em subestimação dos valores preditos, particularmente na etapa de teste com valor de -27,10%. Assim como ocorreu na elaboração do Modelo 3, verifica-se a baixa performance do Modelo 4, evidenciada pela dificuldade dos preditores na estimativa da variável de controle. Esses resultados enfatizam a necessidade de ajustes dos modelos, através de modificações nas características da RNA, reavaliação da seleção de variáveis de entrada do modelo, entre outros.

TABELA 21: Resultados das métricas durante a elaboração do Modelo 4

MÉTRICAS	TREINAMENTO	TESTE
MSE	2,34	3,31
RMSE	1,53	1,82
NRMSE (%)	77,40	90,40
PBIAS (%)	-0,10	-27,10
NSE	0,40	0,18
mNSE	0,22	0,15
r	0,63	0,57
R ²	0,40	0,33
bR ²	0,32	0,18

Legenda: MSE – Erro médio quadrático, RMSE – Raiz do erro médio quadrático, NRMSE (%) – Raiz do erro médio quadrático normalizada, PBIAS (%) – Percentagem de polarização, NSE – Eficiência de Nash-Sutcliffe, mNSE – Eficiência de Nash-Sutcliffe modificada, r – Coeficiente de correlação de Pearson, R² – Coeficiente de determinação, bR² – Coeficiente de determinação (R²) multiplicado pela inclinação da linha de regressão.

4.3 Modelos para Estações de Tratamento de Grande Porte

Devido à ausência de informações, na proposição dos modelos para estações de grande porte apenas a turbidez da água tratada foi escolhida como variável dependente. Das ETAs inclusas na pesquisa, foi empregado um total de 14 variáveis, com 11.695 observações de cada, totalizando um quantitativo de 163.730 dados.

4.3.1 Análise exploratória para os modelos de ETAs de grande porte

Na realização da análise exploratória, conforme apresentado na Tabela 22 e na Figura 28, foi possível constatar que houve grande variabilidade de algumas variáveis. Dentre elas, pode-se destacar a vazão afluente às ETAs (Qaf), com valores de 263-6.000 Ls⁻¹. Esse fato se deve à classificação das ETAs (Q > 500 Ls⁻¹). A produção de lodo também se destaca por sua dispersão. No entanto, é importante ressaltar que esta variável depende da vazão afluente, da dosagem do coagulante utilizado e dos valores de turbidez do afluente às estações. A única variável que se destaca pela inexistência de variância é o tempo de operação, tendo em vista que todas as ETAs realizam operação de forma ininterrupta.

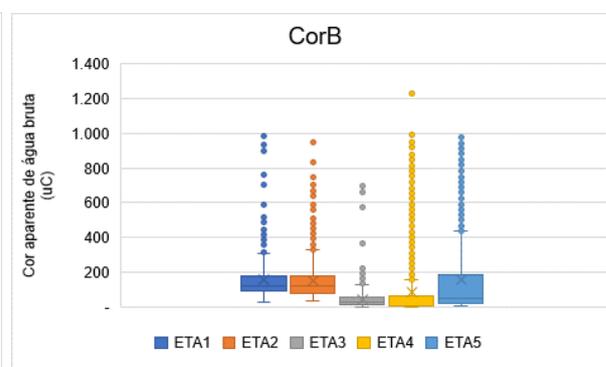
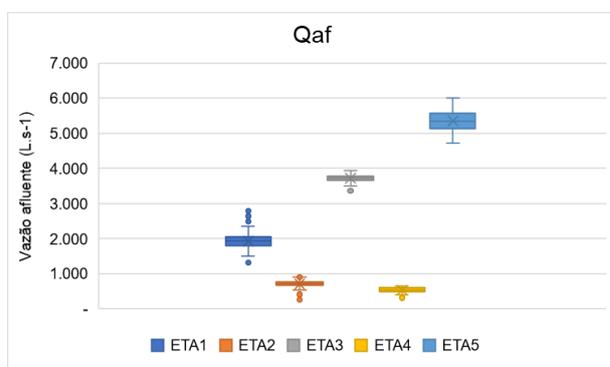
Com relação à qualidade da água bruta afluente às estações, observa-se que o pH se manteve próximo à neutralidade. Já a turbidez e a cor aparente apresentaram maior variação, sendo esses parâmetros influenciados pelos eventos de precipitação e pelas características de uso e ocupação do solo em cada bacia hidrográfica, onde o manancial está localizado. Por outro lado, a qualidade da água tratada apresentou baixa variabilidade, com valores médios e

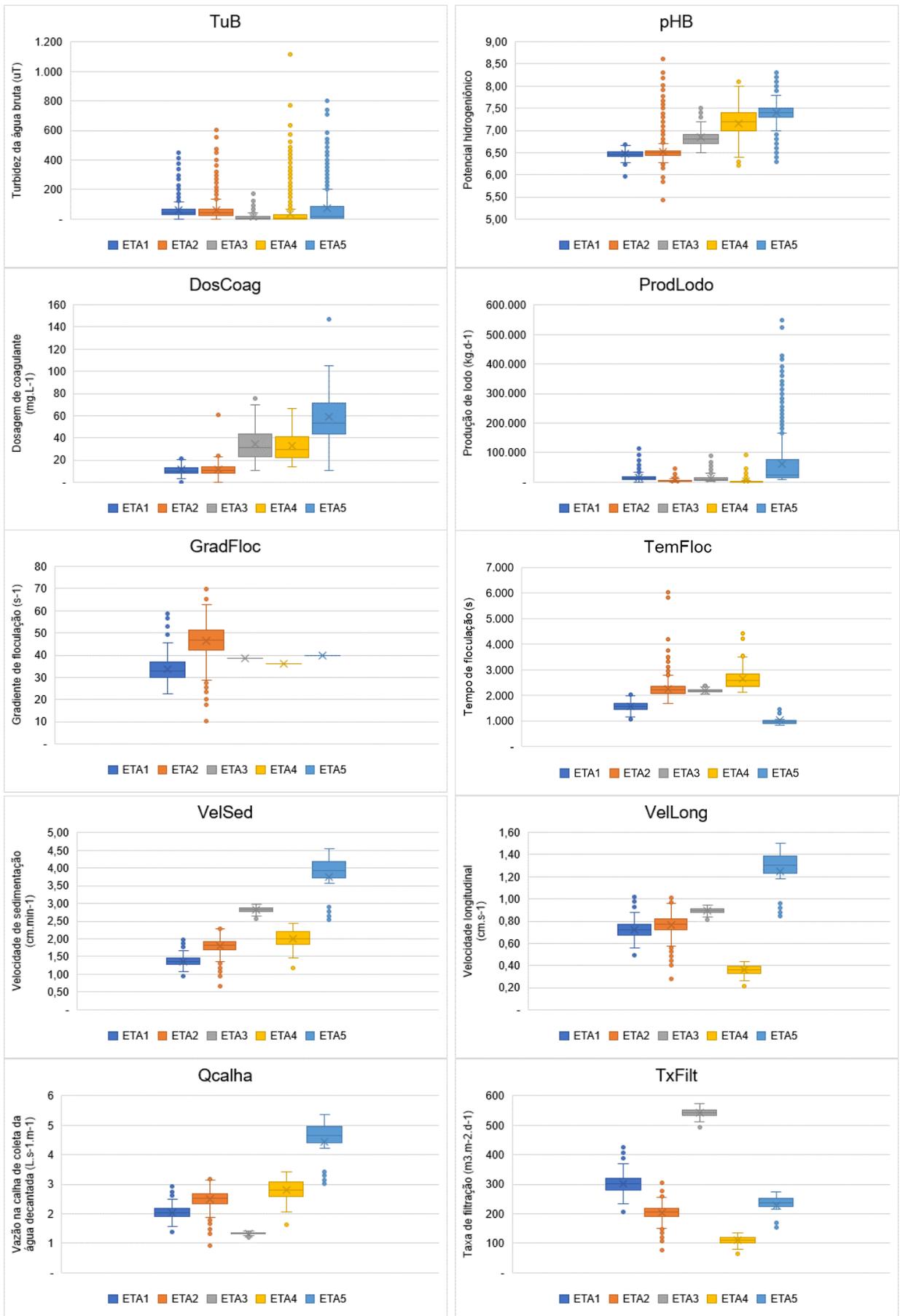
medianos de turbidez dentro dos padrões de potabilidade, o que evidencia a qualidade da etapa de controle realizada pelos operadores das estações.

Ao avaliar a distribuição do conjunto de dados, foi verificado que a hipótese nula de normalidade (ao nível de significância de 0,05) não foi rejeitada. Além disso, há *outliers* no conjunto de dados, em alguns parâmetros analisados, em todas as ETAs. Assim como ocorreu nas análises para as estações de pequeno e médio porte, os valores foram mantidos.

TABELA 22: Apresentação dos resultados da estatística exploratória das variáveis referentes às estações de grande porte

VAR	MÉDIA	MEDIANA	MODA	MÍN.	MÁX.	D.PAD.
Qaf (L.s⁻¹)	2.802,23	3.608,90	3.773,30	263,54	6.005,90	1.835,17
TempOp (h)	24,00	24,00	24,00	5,75	34,00	0,01
CorB (uC)	105,16	45,00	5,00	1,00	1.979,92	103,20
TuB (uT)	44,08	15,00	1,45	0,55	1.112,79	47,08
pHB	7,01	7,00	7,40	5,43	8,61	0,34
DosCoag (mg.L⁻¹)	35,62	33,30	8,10	0,04	146,70	17,11
ProdLodo (kg.dia⁻¹)	22.067,86	8.267,27	2.795,03	356,46	548.417,42	23.804,77
GradFloc (s⁻¹)	38,68	38,55	38,55	10,34	69,96	2,54
TemFloc (s)	1.973,24	2.156,43	2.149,84	843,44	6.026,09	547,16
VelSed (cm.min⁻¹)	2,61	2,69	2,86	0,67	4,53	0,72
VelLong (cm.s⁻¹)	0,81	0,87	0,91	0,21	1,50	0,27
Qcalha (L.s⁻¹.m⁻¹)	2,72	2,59	1,35	0,92	5,37	1,00
TxFilt (m³.m².dia⁻¹)	286,00	235,70	550,37	64,51	573,36	144,39
TuT (uT)	0,29	0,21	0,18	0,02	7,36	0,16





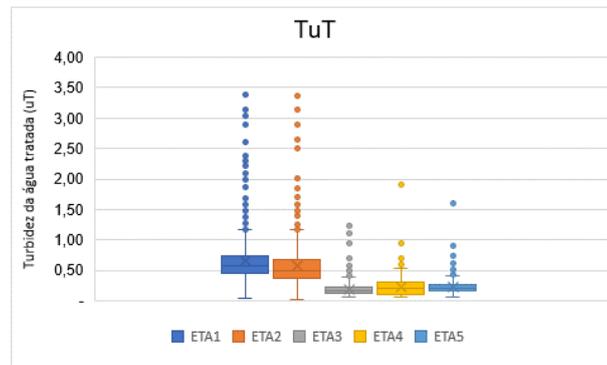


FIGURA 28: Composição de gráficos *boxplot* de cada uma das variáveis que compõem os modelos para as ETAs de grande porte

A diferença entre as vazões das ETAs é significativa, conforme citado anteriormente. Além dela, diversos outros parâmetros hidráulicos podem se destacar por apresentar tal comportamento. Por exemplo, o gradiente de velocidade de floculação, tempo de floculação, velocidades de sedimentação e longitudinal no decantador, velocidade da calha de coleta de água decantada e a taxa de filtração apresentaram variabilidade.

Com relação ao atendimento desses parâmetros à Norma Técnica vigente, observa-se que, para o gradiente velocidade de floculação, todas as ETAs apresentaram valores abaixo de 70 s^{-1} ; já para a taxa de filtração, apenas as ETAs 1 e 3 se apresentaram acima de $360 \text{ m}^3 \cdot \text{m}^{-2} \cdot \text{d}^{-1}$. Tal pode ensejar o mau funcionamento do sistema, particularmente na qualidade da água, tendo em vista que a filtração é a última barreira físico-química e microbiológica para a remoção de materiais particulados, podendo apresentar riscos à saúde da população abastecida.

Os parâmetros de qualidade da água bruta afluentes às estações de tratamento de água se apresentaram muito próximos, com considerável número de valores elevados, particularmente a turbidez e a cor aparente.

Com relação aos parâmetros de qualidade na saída das ETAs, observa-se que a turbidez, em todas estações de tratamento de água, se manteve dentro dos limites máximos permissíveis, estabelecido pelo Anexo XX da Portaria de Consolidação nº5/2017. É possível observar a existência de *outliers*, causando violação ao referido padrão de potabilidade, no entanto, são situações isoladas, que refletem as atividades rotineiras nas estações de tratamento de água de todo o país, onde não há tempo hábil para que os operadores consigam modificar os protocolos operacionais em meio a emergências. Por exemplo, pode-se citar o aumento

inesperado da turbidez afluyente à ETA devido a evento de precipitação extremo a montante do ponto de captação, entre outras situações similares.

4.3.2 Escolhas das variáveis para compor os modelos de grande porte

Nesta etapa, seguiram-se os mesmos procedimentos já adotados anteriormente para as estações de pequeno e de médio porte. Com base nos valores dos coeficientes de correlação de Pearson acima de 10% foi realizada a escolha das variáveis, conforme constatado na Tabela 23. Portanto, as variáveis escolhidas para a proposição do modelo de regressão linear múltipla da turbidez da água tratada para as estações de grande porte foram: Qaf, CorB, TuB, pHB, DosCoag, GradFloc, VelSed e TxFilt. Essas variáveis mantêm a representatividade quanto à qualidade da água bruta e tratada, dosagem de produtos químicos, fatores operacionais e aspectos hidráulicos.

TABELA 23: Matriz de correlação entre as variáveis utilizadas no desenvolvimento dos modelos para ETAs de grande porte

	<i>Qaf</i>	<i>TempOp</i>	<i>CorB</i>	<i>TuB</i>	<i>pHB</i>	<i>DosCoag</i>	<i>ProdLodo</i>	<i>GradFloc</i>	<i>TemFloc</i>	<i>VelSed</i>	<i>VelLong</i>	<i>Qcalha</i>	<i>TxFilt</i>	<i>TuT</i>
<i>Qaf</i>	1													
<i>TempOp</i>	-0,01	1												
<i>CorB</i>	0,04	0,03	1											
<i>TuB</i>	0,07	0,01	0,94	1										
<i>pHB</i>	0,37	-0,02	-0,04	0,01	1									
<i>DosCoag</i>	0,60	-0,01	0,30	0,34	0,49	1								
<i>ProdLodo</i>	0,44	0,00	0,70	0,77	0,20	0,59	1							
<i>GradFloc</i>	0,12	0,04	0,10	0,09	-0,10	-0,02	0,08	1						
<i>TemFloc</i>	-0,78	0,01	-0,17	-0,18	-0,34	-0,44	-0,47	-0,22	1					
<i>VelSed</i>	0,86	0,00	0,05	0,08	0,56	0,68	0,44	0,20	-0,73	1				
<i>VelLong</i>	0,90	0,00	0,12	0,13	0,23	0,49	0,45	0,36	-0,87	0,85	1			
<i>Qcalha</i>	0,36	0,00	0,23	0,26	0,62	0,52	0,42	0,16	-0,67	0,60	0,48	1		
<i>TxFilt</i>	0,46	0,00	-0,17	-0,19	-0,31	-0,02	-0,04	0,06	-0,09	0,27	0,40	-0,57	1	
<i>TuT</i>	-0,25	0,09	0,22	0,16	-0,35	-0,30	-0,04	0,16	-0,07	-0,32	-0,05	-0,04	-0,10	1

4.3.3 Modelo em regressão linear múltipla para ETAs de grande porte

As tabelas 24, 25 e 26 ilustram os resultados do desenvolvimento do modelo em regressão linear múltipla. Inicialmente, na Tabela 20, verifica-se que as variáveis com maior contribuição para o modelo são pH e cor aparente da água bruta, e a vazão afluyente da água bruta. De acordo com os valores F e P, todas as variáveis são estatisticamente significantes para o referido modelo.

TABELA 24: Análise de variância gerada para o modelo de regressão em ETAs de grande porte a partir da turbidez da água tratada

Fonte	GL	SQ Seq	Contribuição	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	8	209,772	26,96%	209,772	26,2215	539,11	0,000
Qaf	1	47,230	6,07%	9,246	9,2465	190,11	0,000
CorB	1	41,025	5,27%	16,032	16,0317	329,61	0,000
TuB	1	8,324	1,07%	3,434	3,4337	70,60	0,000
pHB	1	52,812	6,79%	19,237	19,2368	395,51	0,000
DosCoag	1	26,245	3,37%	18,831	18,8310	387,17	0,000
GradFloc	1	9,941	1,28%	10,736	10,7365	220,74	0,000
VelSed	1	7,514	0,97%	4,541	4,5407	93,36	0,000
TxFilt	1	16,681	2,14%	16,681	16,6810	342,96	0,000
Erro	11686	568,385	73,04%	568,385	0,0486		
Falta de ajuste	11404	568,092	73,00%	568,092	0,0498	47,99	0,000
Erro puro	282	0,293	0,04%	0,293	0,0010	*	*
Total	11694	778,157	100,00%				

Legenda: GL – Graus de liberdade, SQ Seq – Soma dos quadrados sequenciais, Contribuição – Percentual que cada variável representa para o modelo, SQ (Aj.) – Soma dos quadrados ajustados, QM (Aj.) – Quadrado médio ajustado, Valor F – Estatística utilizada para verificar se o modelo está associado com a resposta, Valor-P – Mede a evidência contrária a hipótese nula do teste, Erro puro – erro calculado a partir do número de repetições de observações utilizadas no modelo, Falta de ajuste – calcula a falta de ajuste gerada pela presença de réplicas no conjunto de dados.

Os coeficientes de determinação apresentados (Tabela 25), particularmente o R^2 ajustado, sugerem que o modelo apresenta baixa capacidade de predição, sendo necessária a adoção de outros procedimentos para a elaboração de novos modelos em regressão ou lançar mão de outras metodologias para tal proposição, capazes de proporcionar maior performance para ele.

TABELA 25: Sumário do modelo para turbidez da água tratada em ETAs de grande porte

S	R2	R2(aj)	PRESQ	R2(pred)
0,220541	26,96%	26,91%	569,952	26,76%

Legenda: S – Desvio padrão, R² – Coeficiente de determinação, R²(aj) – Coeficiente de determinação ajustado, PRESQ – Soma dos quadrados preditos do erro (calcula o desvio do modelo), R²(pred) – Coeficiente de determinação predito (determina o quanto o modelo prediz para novas observações).

Na Tabela 26 são apresentados os coeficientes para a formação da equação do modelo, em seguida observa-se o erro padrão dos coeficientes (são responsáveis por medir a precisão do modelo na estimativa de valores desconhecidos dos coeficientes); posteriormente são apresentados os intervalos de confiança, os valores de T (razão entre Coef e EP de Coef), o valor de P, e o VIF (neste caso o modelo não apresentou multicolinearidade entre os resíduos). Finalmente, é apresentada a equação matemática que explica o modelo proposto.

TABELA 26: Coeficientes que compõem o modelo de estações de grande porte para turbidez da água tratada

Termo	Coef	EP de Coef	IC de 95%	Valor-T	Valor-P	VIF
Constante	1,3808	0,0643	(1,2548; 1,5069)	21,48	0,000	
Qaf	0,000033	0,000002	(0,000028; 0,000037)	13,79	0,000	5,30
CorB	0,000658	0,000036	(0,000587; 0,000729)	18,16	0,000	8,19
TuB	-0,000640	0,000076	(-0,000789; -0,000490)	-8,40	0,000	8,27
pHB	-0,16781	0,00844	(-0,18435; -0,15127)	-19,89	0,000	2,59
DosCoag	-0,003051	0,000155	(-0,003355; -0,002747)	-19,68	0,000	2,55
GradFloc	0,007987	0,000538	(0,006933; 0,009040)	14,86	0,000	1,26
VelSed	-0,05908	0,00611	(-0,07106; -0,04709)	-9,66	0,000	6,88
TxFilt	-0,000337	0,000018	(-0,000373; -0,000302)	-18,52	0,000	2,29

Legenda: Termo – Elementos que irão compor o modelo, Coef – Coeficientes do modelo (relação entre as variáveis dependentes e independentes), EP de Coef – Erro padrão dos coeficientes, IC de 95% - Intervalo de confiança ao nível de significância de 95%, Valor-T – Razão entre o coeficiente e o respectivo erro padrão, Valor-P – Mede a evidência contra a hipótese nula, VIF – Fator de inflação da variância (demonstra como a variância de cada coeficiente eleva-se pela correlação entre as variáveis independentes ou preditoras).

Equação de Regressão

$$TuT = 1,3808 + 0,000033 Qaf + 0,000658 CorB - 0,000640 TuB - 0,16781 pHB - 0,003051 DosCoag + 0,007987 GradFloc - 0,05908 VelSed - 0,000337 TxFilt$$

4.3.4 Modelo em RNA para ETAs de grande porte

O modelo em RNA para a turbidez da água tratada nas estações de grande porte é ilustrado na Figura 29, na qual é verificada a configuração da RNA. A topologia da rede é composta por três neurônios em duas camadas ocultas. Essa rede apresenta velocidade de processamento relativamente alta, embora o número de observações de cada variável utilizada para a proposição do modelo cause um pequeno retardo durante as interações, em relação às respostas dos demais modelos supracitados.

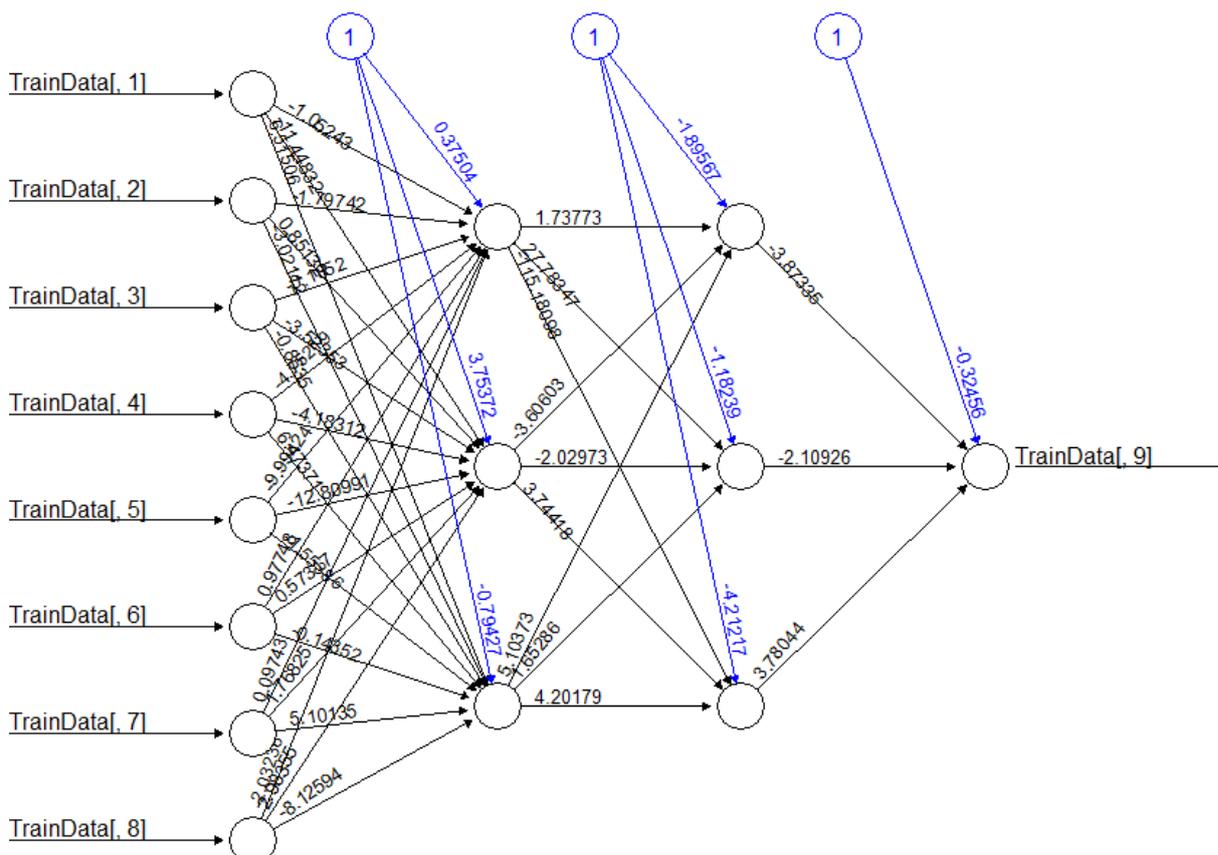


FIGURA 29: Rede neural gerada a partir do modelo proposto para a turbidez da água tratada em ETAs de grande porte

A Figura 30 ilustra o comportamento da turbidez da água tratada observada e calculada após o treinamento da RNA. Verifica-se que acima de 0,5 uT ocorre grande variabilidade no comportamento do parâmetro de interesse, notadamente para os valores preditos (vermelho), sendo esse comportamento justificado pela falta de ajuste do modelo.

Na Figura 31 é possível constatar que há aderência dos valores preditos ao modelo, quando comparados com a reta a 45° ($R^2 = 0,54$) durante a etapa de treinamento. No entanto, esse

valor pode ensejar na baixa capacidade de predição do modelo e no seu desempenho, além da possibilidade de gerar vieses de superestimação ou subestimação.

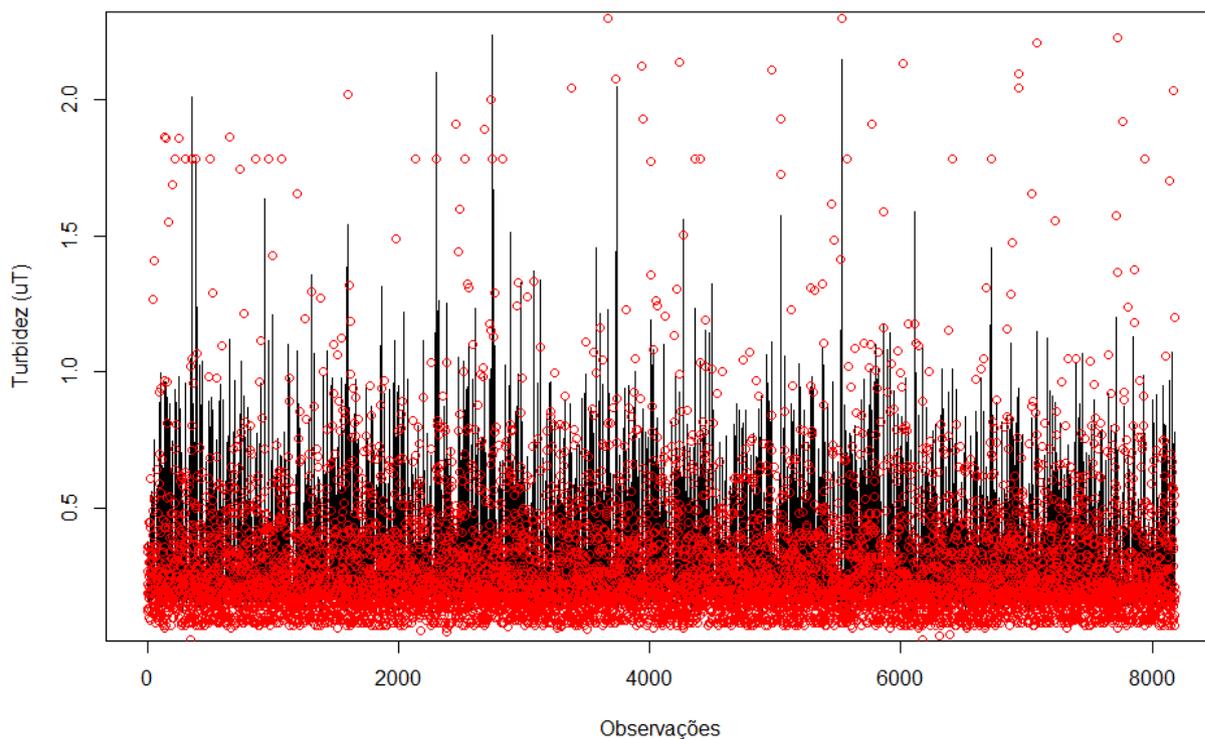


FIGURA 30: Comparação entre o comportamento dos valores de turbidez da água tratada durante o treinamento do modelo das estações de grande porte

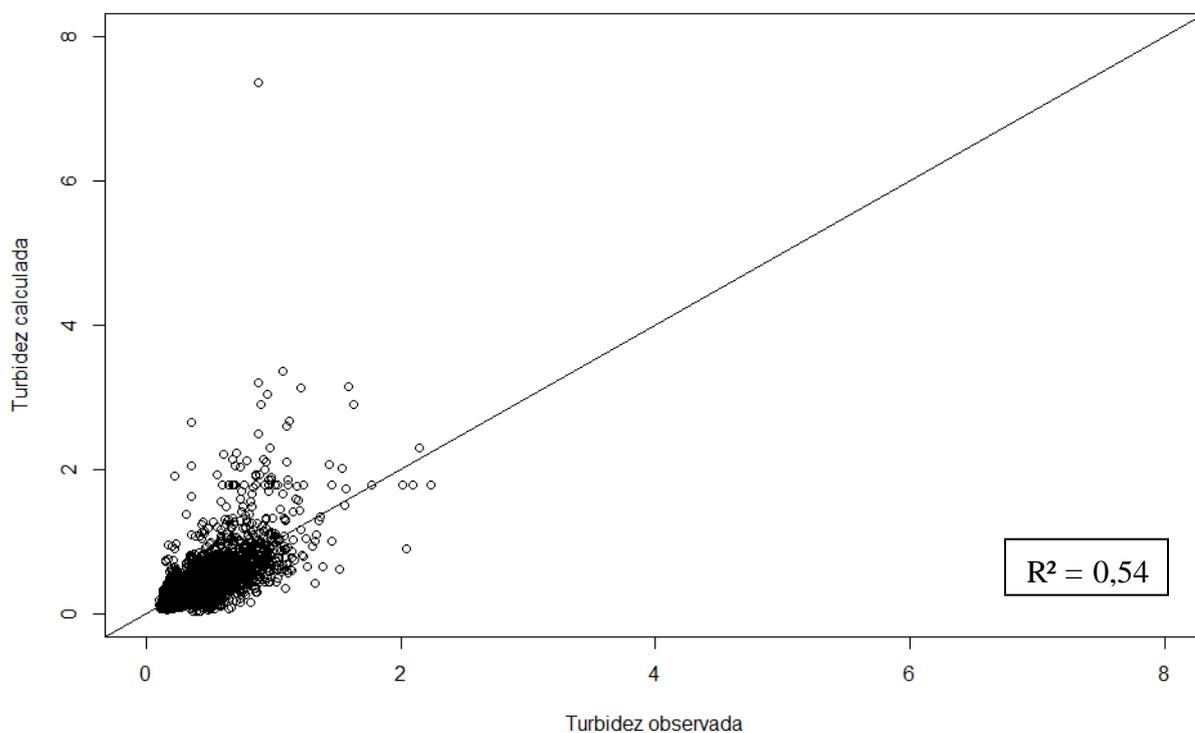


FIGURA 31: Comportamento da correlação entre a turbidez calculada e observada durante o treinamento da rede neural para estações de grande porte

O resultado do Modelo 5 é apresentado na Figura 32, no qual se observa o intervalo de predição criado através do modelo em RNA. O coeficiente de determinação alcançado pelo referido modelo foi de 0,60. Esse valor dá indícios do desempenho relativamente bom para o modelo, considerando o número de observações de cada variável utilizada e as características específicas de cada ETA. Além disso, o modelo foi capaz de explicar 60% da relação entre valores observados e preditos da turbidez da água tratada nas estações de grande porte. No entanto, a amplitude do intervalo predito enseja grande variabilidade desses valores.

Na perspectiva do atendimento ao Anexo XX da Portaria de Consolidação nº 5/2017, conforme já observado durante a fase de aprendizagem ou treinamento, a rede neural conseguiu manter a grande maioria dos valores preditos dentro do limite estabelecido pelo padrão de potabilidade. Contudo, verifica-se a existência de um número considerável de variações dos valores acima de 0,3 uT. A partir desse, há um verdadeiro espalhamento dos dados em sentido crescente dos níveis de turbidez.

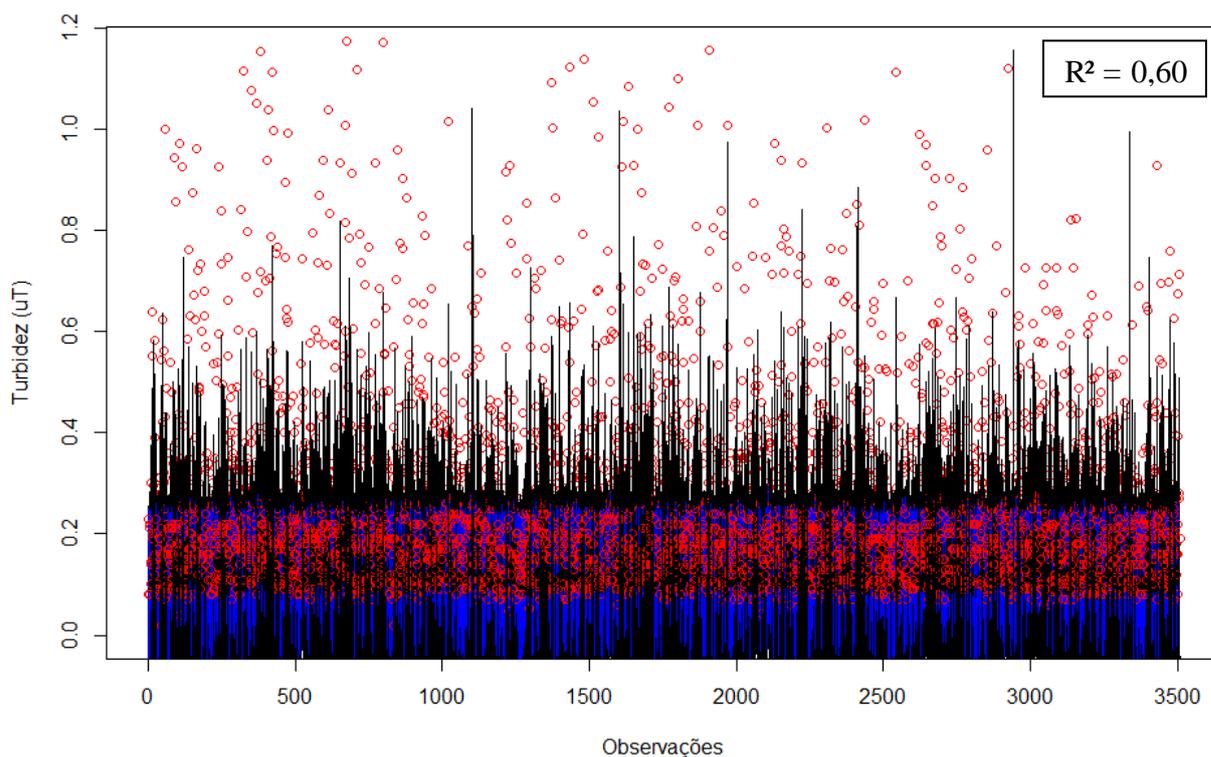


FIGURA 32: Intervalo de predição gerado a partir dos valores preditos de turbidez da água tratada para ETAs de grande porte

As métricas geradas a partir do desenvolvimento do modelo para as estações de grande porte estão apresentadas na Tabela 27, na qual se observa que o mesmo apresentou erro relativamente baixo (abaixo de 0,25 uT), notadamente quando comparado com os demais

modelos já apresentados. Esse resultado, aliado ao R^2 de 0,60, poderia ser considerado aceitável para tal aplicação, embora necessite de melhor ajuste. Contudo, o coeficiente de Nash-Sutcliffe se apresentou muito próximo de 0, durante a fase de teste. Adicionado ao exposto, o PBIAS apresentou resultado indicativo de viés de subestimação (-50,60) para os valores preditos. Esses resultados corroboram para a baixa performance do modelo, bem como para a necessidade de modificações na configuração da RNA e/ou na escolha das variáveis de entrada, tendo em vista que há grande variabilidade dos dados para a classificação desse porte de ETA ($Q > 500 \text{ L.s}^{-1}$).

TABELA 27: Resultados das métricas durante a elaboração do Modelo 5

MÉTRICAS	TREINAMENTO	TESTE
MSE	0,03	0,05
RMSE	0,18	0,23
NRMSE (%)	63,10	97,00
PBIAS (%)	-0,10	-50,60
NSE	0,54	0,06
mNSE	0,40	0,03
r	0,73	0,77
R^2	0,54	0,60
b R^2	0,42	0,24

Legenda: MSE – Erro médio quadrático, RMSE – Raiz do erro médio quadrático, NRMSE (%) – Raiz do erro médio quadrático normalizada, PBIAS (%) – Percentagem de polarização, NSE – Eficiência de Nash-Sutcliffe, mNSE – Eficiência de Nash-Sutcliffe modificada, r – Coeficiente de correlação de Pearson, R^2 – Coeficiente de determinação, b R^2 – Coeficiente de determinação (R^2) multiplicado pela inclinação da linha de regressão.

4.4 Comparação entre os modelos propostos e discussão dos resultados

Foram realizadas análises como forma de comparação entre os modelos propostos e, para avaliar seu desenvolvimento e sua possibilidade de aplicação nas atividades de gerenciamento de sistemas de abastecimento de água, particularmente, no que diz respeito à operação das estações de tratamento de água para consumo humano.

Inicialmente, foi conduzido um estudo comparativo entre os modelos propostos e aqueles desenvolvidos em regressão linear múltipla (*baseline*), conforme apresentado na Tabela 33, tomando-se como instrumento avaliativo apenas o coeficiente de determinação, pois ele foi a única métrica comum entre as duas ferramentas de elaboração dos modelos. O desempenho dos modelos em RNA foi superior ao *baseline* (com exceção do Modelo 3). O maior ganho obtido entre eles foi de 0,34 para o Modelo 5.

De modo geral, os resultados apresentados mostram a vantagem que se tem na utilização das RNA em sistemas possivelmente não-lineares, em detrimento do modelo tradicional (*baseline*), principalmente pelo fato de se utilizar uma ferramenta de aprendizagem, capaz de ser treinada e “aprender”, por meio de padrões, a tomar decisões de forma autônoma, necessitando da supervisão durante a etapa de treinamento.

TABELA 28: Comparação entre os resultados dos modelos desenvolvidos para cada porte de estação, de acordo com os valores dos coeficientes de determinação (R^2) gerados em cada um dos modelos e nos *baselines*

PORTE	MODELO/ PARÂMETRO	BASELINE	MODELO EM RNA (teste)
Pequeno	Modelo 1 - Turbidez	0,51	0,66
	Modelo 2 - Cor aparente	0,50	0,67
Médio	Modelo 3 - Turbidez	0,22	0,21
	Modelo 4 - Cor aparente	0,17	0,33
Grande	Modelo 5 - Turbidez	0,26	0,60

É importante destacar que os modelos em regressão linear múltipla apresentaram resultados próximos aos comumente relatados na literatura durante sua aplicação, particularmente, aqueles desenvolvidos para estações de tratamento de água de pequeno porte, com valores de R^2 acima de 0,50. Esse tipo de ferramenta ainda possui grande capacidade de aplicação, conforme apresentado pelas diversas pesquisas que o utilizaram. Dentre elas, pode destacar os trabalhos recentes desenvolvidos por Heddham *et al.* (2016), Wallace *et al.* (2016) e Wu *et al.* (2016).

Apesar do uso das RNAs, muitas vezes, ser mais complexo, devido às etapas que envolvem a elaboração de modelos e à necessidade de pessoal com certo conhecimento na área, é importante ressaltar que o desenvolvimento dos modelos em regressão linear múltipla necessita de um considerável número de ajustes, além de obedecer a sua série de critérios para que seus resultados sejam estatisticamente confiáveis. Por exemplo, inexistência de multicolinearidade entre as variáveis independentes do modelo, correlação significativa entre as variáveis independentes e a variável controle, número mínimo de observações para aplicação, homocedasticidade entre os resíduos gerados a partir do teste de regressão, entre outros (HEDDAM *et al.*, 2016).

Na avaliação de cada modelo individualmente, quando comparado com outros modelos de avaliação e/ou previsão de desempenho citados na literatura, os Modelos 1, 2 e 5 se apresentaram promissores, com coeficiente de determinação de 0,66, 0,67 e 0,60, respectivamente. Como exemplo, podem ser destacados os trabalhos de Rak (2013), que desenvolveu um modelo semelhante em RNA e com base na turbidez, com $R = 0,84$, e de Alsaqqar *et al.* (2016), que aplicaram RNA no desenvolvimento de índices para três ETAs em Bagdá (Al-Karakh, Al-Karama e Al-Qadisiya), apresentando coeficiente de correlação de 0,92, 0,83 e 0,79, respectivamente. Todos esses resultados ainda podem ser comparados com os apresentados por Maier *et al.* (2004), com $R^2 = 0,90$, na estimativa da turbidez da água tratada para um modelo de previsão da dosagem ótima de coagulante para águas superficiais do Sudeste da Austrália.

Ao se comparar os resultados dos modelos entre os diferentes portes de ETAs (Tabela 34) verifica-se que, para estações de pequeno porte, as redes neurais artificiais foram bem mais eficientes do que para os demais portes de ETA. De fato, houve grande diferença entre os valores preditos pelos modelos. Nesse contexto, pode-se constatar que há diversos fatores que podem influenciar nesses resultados. Dentre eles pode-se destacar a variância do conjunto de dados utilizados para a elaboração do modelo, mesmo que se admitam estações com valores de vazões relativamente próximas, há outros parâmetros que podem apresentar grande variabilidade. Por exemplo, os parâmetros de qualidade da água bruta, os parâmetros hidráulicos e operacionais nas unidades de floculação, decantação, filtração e desinfecção, os quais têm relação direta com o desempenho da estação de tratamento. Essa influência da variância na elaboração e no resultado dos modelos pode ser mais bem visualizada nas ETAs de médio e de grande porte, nas quais as vazões utilizadas para seus agrupamentos tinham uma amplitude relativamente grande. Para ETAs de grande porte, por exemplo, há, dentro do conjunto de dados, valores de vazão que variam de aproximadamente $650 \text{ L}\cdot\text{s}^{-1}$ até mais de $5.000 \text{ L}\cdot\text{s}^{-1}$. Assim, mesmo utilizando um banco de dados volumoso, do ponto de vista quantitativo e qualitativo, há uma notória dificuldade da RNA para “aprender” tantas situações diferentes e poder gerar resultados mais precisos, tendo em vista que o funcionamento dessa ferramenta parte do conceito de aprendizagem de padrões.

Portanto, constata-se que, na prática, não é condizente realizar a aplicação de um modelo de um determinado conjunto de ETAs em outro diferente, tendo em vista que ele foi desenvolvido apenas para aqueles grupos específicos de estações. Nesse contexto, também se

verifica a necessidade do emprego do número maior de modelos, com vista à diminuição do intervalo entre os valores de vazão, entre outras variáveis. Seguindo essa tendência, espera-se que o número maior de estações de tratamento de água permita a obtenção de dados mais consistentes, com menor variância e mais uniformes, proporcionando o desenvolvimento de modelos mais realísticos.

TABELA 29: Resumo dos resultados das métricas para todos os modelos propostos

FASES	MÉTRICAS	PEQUENO		MÉDIO		GRANDE
		Mod. 1	Mod. 2	Mod. 3	Mod. 4	Mod. 5
Trein.	MSE	0,30	2,35	0,17	2,34	0,03
	RMSE	0,55	1,53	0,41	1,53	0,18
	NRMSE (%)	58,00	62,40	82,20	77,40	63,10
	PBIAS (%)	-19,60	-5,90	-0,10	-0,10	-0,10
	NSE	0,66	0,61	0,32	0,40	0,54
	mNSE	0,50	0,36	0,18	0,22	0,40
	r	0,84	0,79	0,57	0,63	0,73
	R ²	0,71	0,62	0,32	0,40	0,54
	bR ²	0,43	0,48	0,21	0,32	0,42
Teste	MSE	0,44	2,04	0,20	3,31	0,05
	RMSE	0,66	1,43	0,45	1,82	0,23
	NRMSE (%)	59,60	57,10	88,90	90,40	97,00
	PBIAS (%)	3,50	2,80	-3,50	-27,10	-50,60
	NSE	0,64	0,67	0,21	0,18	0,06
	mNSE	0,46	0,35	0,17	0,15	0,03
	r	0,81	0,82	0,46	0,57	0,77
	R ²	0,66	0,67	0,21	0,33	0,60
	bR ²	0,53	0,59	0,12	0,18	0,24

Legenda: MSE – Erro médio quadrático, RMSE – Raiz do erro médio quadrático, NRMSE (%) – Raiz do erro médio quadrático normalizada, PBIAS (%) – Percentagem de polarização, NSE – Eficiência de Nash-Sutcliffe, mNSE – Eficiência de Nash-Sutcliffe modificada, r – Coeficiente de correlação de Pearson, R² – Coeficiente de determinação, bR² – Coeficiente de determinação (R²) multiplicado pela inclinação da linha de regressão, Mod. – Modelo, Trein. – Treinamento.

Durante a fase de treinamento todos os modelos apresentaram viés de subestimação pelos valores negativos de percentual de Bias. Além de avaliar a possibilidade de subestimação ou superestimação, essa métrica é capaz de verificar a existência de erros sistemáticos que podem ocorrer no modelo, o que inviabilizaria, inicialmente, sua aplicação e necessitaria

lançar mão de outras rotinas para que houvesse essa correção. Já durante o teste, foram os Modelos 1 e 2 que se destacaram por serem os únicos a não apresentar viés de subestimação.

No que concerne ao erro, durante o treinamento e teste, o Modelo 5 foi o que apresentou menor valor (RMSE de 0,18 uT e 0,23 uT, respectivamente). Esse resultado, juntamente com os valores de R^2 de 0,54 e 0,60, poderia corroborar para um bom resultado do modelo. No entanto, o valor do PBIAS na etapa de teste se manteve acima de 50% e de forma negativa, isso sugere a existência de viés de subestimação da variável de interesse, reduzindo o desempenho do modelo.

Apesar das incertezas existentes nos modelos, bem como a presença dos vieses de subestimação e superestimação, observa-se que os intervalos de predição são importantes ferramentas no auxílio a tomada de decisão, tendo em vista que não se trata de uma observação isolada, mas de um limite onde o tomador de decisão poderá fazer sua “escolha”. Os resultados apresentados pelas métricas sugerem a necessidade de melhor ajuste dos modelos, com vistas à redução das incertezas e melhor ajuste deles, particularmente quando se tenciona sua aplicação.

Na análise dos resultados dos modelos, quando se compara com outros trabalhos realizados para a predição de parâmetros de qualidade de água, particularmente da turbidez, os resultados apresentados podem ser comparados com os de Cordoba *et al.* (2014), que obtiveram valores em torno de $R^2 = 0,90$, nas etapas de treinamento, teste e validação de modelo de qualidade de água em um sistema de abastecimento de água em Brno, República Checa. Gazzaz *et al.* (2012) desenvolveram um índice de qualidade de água para o Rio Kinta - Malásia, também utilizando RNA e alcançaram correlação de 0,97, sendo que o modelo de predição conseguiu explicar 95,4% dos valores medidos.

Kim e Parnichkun (2017) desenvolveram modelos semelhantes para a estimativa da qualidade da água bruta de uma estação de tratamento em Changwon – Coreia do Sul, com vista a aplicação da dosagem ótima de coagulante, sendo que eles elaboraram modelos em RNA e híbridos (ANFIS). Os resultados alcançados em redes neurais artificiais foram de $R^2 = 0,84$, com RMSE de 0,0633 uT. Esses resultados corroboram para ainda mais com os resultados dos modelos propostos nesta pesquisa, embora não tenha sido possível se chegar a valores de erros tão baixos. No entanto, é necessário verificar que essas pesquisas possuem particularidades distintas.

Como forma de avaliar as hipóteses sugeridas para o desenvolvimento da pesquisa, foi elaborado o Quadro 2, que as apresenta e qual a sua situação de acordo com os resultados. Essas hipóteses versam sobre a melhoria no gerenciamento de sistemas de abastecimento de água, particularmente do tratamento, quando da aplicação de ferramentas computacionais desenvolvidas com base nas RNAs; melhorias dos modelos propostos, a partir de mudanças nas suas configurações, e a comparação entre os modelos propostos e modelos clássicos, amplamente divulgada no meio científico, que apresentam algumas limitações, como o caso da regressão linear múltipla.

QUADRO 2: Análise das hipóteses propostas para a pesquisa, de acordo com o desenvolvimento dos modelos

HIPÓTESES	SITUAÇÃO/JUSTIFICATIVA
Modelos de predição baseados em Inteligência Artificial podem gerar a perspectiva de melhor gestão do sistema de tratamento de água, permitindo que as intervenções sejam realizadas de forma antecipada e pontual	Aceita , devido à grande capacidade dos modelos de predição desenvolvidos em RNA de predizer, com considerada precisão e acurácia, os valores dos parâmetros aqui analisados
Redes do tipo <i>backpropagation</i> apresentam maior capacidade de generalização e flexibilidade do modelo proposto	Aceita , pelo fato deste tipo de algoritmo de treinamento apresentar maior flexibilidade aos modelos. Com relação à generalização, não foi possível realizar a verificação
Quanto maior o número de neurônios na camada oculta, maior será a otimização no processamento da rede e sua capacidade de generalização, muito embora o treinamento possa se tornar excessivo	Refutada , tendo em vista que, ao longo das várias tentativas para a obtenção de um modelo cada vez mais otimizado, essa hipótese não foi confirmada
Modelos desenvolvidos em RNA poderão apresentar melhores resultados quando comparados com modelos baseados em regressão linear múltipla	Aceita , os resultados apresentam de forma clara e objetiva, notadamente entre os Modelos 1 e 2, que houve melhoria significativa na predição dos parâmetros de tratamento de água quando se utilizou RNA

Das quatro hipóteses apresentadas, apenas uma foi refutada. Ela apresentava a necessidade do uso de um número maior de neurônios nas camadas ocultas para o desenvolvimento de modelos, quando se dispunha de um número elevado de variáveis. No entanto, após o desenvolvimento da pesquisa, observou-se que, para a situação aqui estudada, a hipótese não foi válida, tendo em vista que o número menor de neurônios conseguiu gerar resultados mais otimizados do que quando se utilizou uma maior quantidade desses. Conforme já ilustrado pelo quadro, as demais hipóteses foram aceitas, com todas as justificativas válidas.

Apesar de alguns dos resultados dessa pesquisa, em alguns momentos, apontarem alguns vieses de subestimação ou superestimação, é perceptível sua contribuição para o desenvolvimento de trabalhos futuros com a aplicação do mesmo método, bem como a necessidade de se trabalhar com um número maior de estações de tratamento de água. É notória a escassez da aplicação de ferramentas desse tipo nos processos e aplicações do saneamento ambiental, particularmente, onde muitas vezes faltam investimentos na elaboração de tecnologias que garantam processos de automação e sofisticação dos procedimentos operacionais, otimizando custos e garantindo maior precisão na tomada de decisão.

Embora a literatura apresente diversos trabalhos sobre a utilização de RNA na área do saneamento ambiental, durante as fase de desenvolvimento dessa pesquisa, não foi constatada a existência de pesquisas que se utilizassem dessa ferramenta de maneira similar à que aqui se apresenta, para o desenvolvimento de modelos de predição em ETAs, notadamente quando agrupadas em portes e os modelos são capazes de gerar não apenas valores preditos, mas intervalos de predição.

Ainda com relação à aplicabilidade dos modelos, é possível verificar que há uma área enorme de aplicação, particularmente, na operação das estações convencionais de tratamento de água. No campo da operacionalidade das ETAs, há uma carência desse tipo de ferramenta, tendo em vista que os métodos de monitoramento/controle utilizados em cada unidade de tratamento estão fundamentados, meramente, na análise de parâmetros de qualidade de água, estabelecidos pelos órgãos de saúde pública de cada país, sendo essas norteadas pelas diretrizes estabelecidas pela Organização Mundial de Saúde.

Outro tipo de aplicação interessante para esses modelos é o planejamento de ações de melhoria das estações, tendo em vista que, apesar da resposta dos mesmos ser apresentada por um único parâmetro de qualidade de água, sua elaboração envolve uma série de outros parâmetros ou variáveis relacionadas a aspectos hidráulicos das unidades, fatores operacionais, uso de produtos químicos, entre outros. Isso permite sua utilização de maneira mais direcionada para a realização de prognósticos das estações de tratamento de água, podendo ser alinhado a outras ferramentas de construção. Por exemplo, o *Building Information Modeling* – BIM (Modelagem da Informação da Construção), auxiliando na elaboração de um modelo físico, a partir da modelagem computacional.

5. CONCLUSÕES

O objetivo geral desta pesquisa foi propor modelos de predição de desempenho de estações convencionais de tratamento de água utilizando redes neurais artificiais, tendo como parâmetros de controle a turbidez e a cor aparente da água tratada. A partir dos resultados apresentados, pode-se concluir que:

- Dos modelos desenvolvidos a partir dos testes de regressão linear múltipla, apenas os de estações de pequeno porte apresentaram resultados com indícios para a sua aplicação, após melhorias necessárias. Os demais, apresentaram coeficientes de determinação relativamente baixos, o que demonstra a dificuldade do modelo em prever o comportamento da saída da qualidade da água nas ETAs. Desse modo, os modelos para os grupos das ETAs estudados não são recomendados para a aplicação proposta, no entanto, eles foram de fundamental importância para a comparação entre os resultados dos modelos em RNA;
- De modo geral, os modelos desenvolvidos em RNA, apesar de apresentarem um pouco de dificuldade na estimativa dos picos de turbidez e cor aparente, foram capazes de simular, com precisão e eficiência, o desempenho das estações de tratamento de água analisadas, particularmente os Modelos 1 e 2, apresentando coeficientes de determinação de 0,66 e 0,67, respectivamente; com valores de erros relativamente baixos, além de importante capacidade de gerar intervalos de predição em atendimento ao Padrão de Potabilidade ora em vigor;
- Na perspectiva da geração de um modelo generalizado, capaz de ser utilizado em qualquer estação convencional de tratamento de água, foi percebido, durante o desenvolvimento dessa pesquisa, que o mesmo não seria eficiente, tendo em vista os resultados alcançados. Nesse caso, recomenda-se investir em modelos mais específicos, para um determinado grupo de estações de tratamento de água (inclusive modelos específicos para cada ETA), do que desenvolver modelos mais generalistas, muito embora a literatura destaque o potencial de generalização das RNAs como grande vantagem, mas nessa aplicação de forma particular não foi possível constatá-la diretamente;

- Os modelos desenvolvidos em RNA, de maneira geral, foram superiores aos desenvolvidos a partir da regressão linear múltipla. Isso se deve ao fato de as RNA serem capazes de se adaptar, de maneira mais adequada, às situações que naturalmente ocorrem nas ETAs, ou seja, a capacidade de sua aplicação na resolução de problemas e de se adaptar a situações não-lineares; e
- Os resultados apresentados ao longo dessa pesquisa corroboram o desenvolvimento de outras pesquisas voltadas para o uso da modelagem computacional, notadamente por meio das RNA, no gerenciamento de sistemas de abastecimento de água, com destaque para as etapas que compõem o tratamento convencional. Essa aplicação poderá refletir na garantia de maior qualidade para os serviços prestados pelas empresas de saneamento, na otimização das atividades operacionais de controle da qualidade da água – de responsabilidade dessas empresas, na redução do número de análises físico-químicas e microbiológicas utilizadas na etapa de controle, na possível otimização e viabilidade econômica no que se refere aos custos operacionais dos sistemas; na possibilidade de antever situações de riscos, dando maior tempo para que os tomadores de decisão possam gerenciar a situação de maneira mais equilibrada e confiável. Todos esses aspectos são capazes de gerar um produto cada vez mais confiável a ser ofertado à população.

6. RECOMENDAÇÕES

Ao longo do desenvolvimento dessa pesquisa foi possível alcançar os objetivos propostos. No entanto, também se verificou que os resultados poderiam ser melhorados e outras aplicações poderiam ser realizadas, abrindo possibilidades para uma série de pesquisas semelhantes. Portanto, recomendam-se com essa tese:

- A utilização de outras ferramentas de modelagem computacional, tais como algoritmo genético, entre outros, que sejam capazes de otimizar os modelos de interesse dessa pesquisa, ou de outros parâmetros que possam corroborar com as atividades de controle e monitoramento ou vigilância da qualidade da água. Muito embora, recomenda-se também que sejam utilizados bancos de dados confiáveis, com prévia análise de confiabilidade deles;
- Na tentativa de otimizar ainda mais os modelos aqui desenvolvidos, pode-se antecipadamente realizar uma análise estatística básica, com vista à verificação da variabilidade dos conjuntos dos parâmetros monitorados, de modo que se agrupem elementos – nesse caso, ETAs que apresentem parâmetros com valores mais aproximados, como vazões afluentes de mesma magnitude. Dessa forma, a quantidade de conjuntos das ETAs será maior, mas os modelos poderão ser mais realísticos, precisos e acurados;
- Tendo em vista que essa pesquisa abordou apenas o desenvolvimento de modelos em RNA para estações convencionais de tratamento de água, é necessária a aplicação das mesmas metodologias, seguindo as etapas aqui apresentadas em estações não convencionais ou em outras tecnologias de tratamento. Como por exemplo, em sistemas simplificados de tratamento de água, utilizados em comunidades rurais difusas; em sistemas mais complexos, planejados para a remoção de compostos persistentes ao tratamento convencional; em estações operadas sob pressão como osmose inversa, entre outros. Objetiva-se que essa ferramenta possa subsidiar na gestão das ETAs, independente de seu porte e/ou tecnologia de tratamento utilizada, oferecendo maior segurança, predição e acurácia ao processo;

- Como o desenvolvimento dos modelos propostos nessa pesquisa está fundamentado na utilização de ferramentas computacionais, particularmente as RNA, é possível sugerir a possibilidade da criação de uma ferramenta computacional, tal como um software ou um aplicativo, de modo a facilitar a aplicação desses modelos no dia a dia dos profissionais que atuam na área e criar um ambiente interativo entre o modelo e o usuário;
- Como a criação de modelos que geraram intervalos de predição só foi possível quando da utilização da RNA, é necessário que se busquem mecanismos e/ou métodos estatísticos capazes de desenvolver tais intervalos, a partir de valores preditos em modelos desenvolvidos por regressão linear múltipla;
- Com base nos resultados apresentados pelos modelos em regressão linear múltipla, percebe-se a carência da proposição de modelos a partir da regressão não-linear, tendo em vista a quantidade já existente citada na literatura. O objetivo é realizar um estudo comparativo entre os modelos para a aplicação que aqui se destina; e
- Pelo fato das concessionárias responsáveis pelos sistemas de abastecimento de água terem dificuldades no levantamento de informações relacionadas ao consumo energético, despesa com pessoal e com produtos químicos, e tendo em vista que o desempenho de uma estação de tratamento de água não se restringe apenas aos aspectos relacionados ao tratamento propriamente dito, há a possibilidade do desenvolvimento de um modelo mais completo, capaz de estimar ou realizar predição a partir de um número maior de variáveis.

REFERÊNCIAS

- ADAMOWSKI, J.; CHAN, H. F.; PRASHER, S. O.; OZGA-ZIELINSKI, B.; SLIUSARIEVA, A. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *WATER RESOURCES RESEARCH*, v. 48, p. 1-14, 2012.
- AGGARWAL, C. C. *Neural Networks and Deep Learning: A Textbook*. USA: Springer, 2018.
- AHMADLI, Z.; FAHIMINIA, M.; JANG, S. A.; ANSARI, R.; SAVADKOUHI, M.; ANBAZ, A.; ESCANDARI, A. Performance Evaluation of Qom Water Treatment of Plant during 2005 to 2014, Iran. *Archives of Hygiene Sciences*, v. 6, n. 1, p. 58-65, 2016.
- AHMADI, A.; TIRUTA-BARNA, L. A Process Modelling-Life Cycle Assessment-MultiObjective Optimization tool for the eco-design of conventional treatment processes of potable water. *Journal of Cleaner Production*, v. 100, p. 116-125, 2015.
- AHMADI, A.; TIRUTA-BARNA, L.; BENETTO, E.; CAPITANESCU, F.; MARVUGLIA, A. On the importance of integrating alternative renewable energy resources and their life cycle networks in the eco-design of conventional drinking water plants. *Journal of Cleaner Production*, v. 135, p. 872-883, 2016.
- ALBERTA. *Standards and Guidelines for Municipal Waterworks, Wastewater and Storm Drainage Systems, Part 1*. Alberta: Queen's Printer, 2012.
- ALEGRE, H.; BAPTISTA, J. M.; CABRERA JUNIOR, E.; CUBILLO, F.; DUARTE, P.; HIRNER, W.; MERKEL, W.; PARENA, R. *PERFORMANCE INDICATORS FOR WATER SUPPLY SERVICES*. 2 ed. London: IWA Publishing, 2006.
- ALI, A.; HASHMI, H. N.; BAIG, N.; IQBAL, S.; MUMTAZ, K. Performance evaluation of the water treatment plants of Islamabad – Pakistan. *ARCH. ENVIRON. SCI.*, v. 6, p. 111-117, 2012.
- AL-JEEBORY, A. A.; GHAWI, A. H. PERFORMANCE EVALUATION OF AL-DEWANYIA WATER TREATMENT PLANT IN IRAQ. *Al-Qadisiya Journal For Engineering Sciences*, v. 2, n. 4, 2009.
- ALMEIDA, M. C.; SILVA, M. M.; DE PAULA, M. AVALIAÇÃO DO DESEMPENHO DE UMA ESTAÇÃO DE TRATAMENTO DE ÁGUA EM RELAÇÃO A TURBIDEZ, COR E PH DA ÁGUA. *Revista Eletrônica de Gestão e Tecnologias Ambientais*, v. 5, n. 1, p. 25-40, 2017.
- ALMEIDA, J. M. S. *Otimização do índice de qualidade de estação convencional de tratamento de água (IQETA) por meio de análise estatística multivariada*. 2009. 80f. Dissertação (Mestrado em Saneamento, Meio Ambiente e Recursos Hídricos) - Escola de Engenharia, UFMG, Belo Horizonte, 2009.

ALSAQQAR, A. S.; KHUDAIR, B. H.; ALI, S. K. Prediction of Ryznar Stability Index for Treated Water of WTPs Located on Al-Karakh Side of Baghdad City using Artificial Neural Network (ANN) Technique. *Journal of Engineering*, v. 22, n. 5, p. 1-10, 2016.

ANDRADE, M. A.; CHOI, C. Y.; LANSEY, K.; JUNG, D. Enhanced Artificial Neural Networks Estimating Water Quality Constraints for the Optimal Water Distribution Systems Design. *Journal of Water Resources Planning and Management*, v. 142, n. 9, p. 1-14, 2016.

ARAM, Z.; JAFARI, S.; MA, J.; SPROTT, J. C.; ZENDEHROUH, S.; PHAM, V.-T. Using chaotic artificial neural networks to model memory in the brain. *Commun Nonlinear Sci Numer Simulat*, v. 44, p. 449-459, 2017.

ARBAULT, D.; RUGANI, B.; TIRUTA-BARNA, L.; BENETTO, E. Emergy evaluation of water treatment processes. *Ecological Engineering*, v. 60, p. 172-182, 2013.

Associação Brasileira de Normas Técnicas (ABNT). *NBR 12.216, 30 de abril de 1992*. Projeto de estação de tratamento de água para abastecimento público. Rio de Janeiro: ABNT, 1992.

Associação Brasileira de Normas Técnicas - Internacional Organization of Standardization (ABNT - ISO). *NBR ISO 14.040, 21 de junho de 2009*. Gestão Ambiental – Avaliação de Ciclo de Vida – Princípios e estrutura. Rio de Janeiro: ABNT, 2009.

Associação Brasileira de Normas Técnicas - Internacional Organization of Standardization (ABNT - ISO). *NBR ISO 24.510, 10 de dezembro de 2012*. Atividades relacionadas aos serviços de água potável e de esgoto — Diretrizes para a avaliação e para a melhoria dos serviços prestados aos usuários. Rio de Janeiro: ABNT, 2012.

BAGHERI, M.; AKBARI, A.; MIRBAGHERI, S. A. Advanced control of membrane fouling in filtration systems using artificial intelligence and machine learning techniques: A critical review. *Process Safety and Environmental Protection*, v. 123, p. 229-252, 2019.

BAGHERI, M.; MIRBAGHERI, S. A.; EHTESHAMI, M; BAGHERI, Z. Modeling of a sequencing batch reactor treating municipal wastewater using multi-layer perceptron and radial basis function artificial neural networks. *Process Safety and Environmental Protection*, v. 93, p. 111-123, 2015.

BAXTER, C. W.; STANLEY, S. J.; ZHANG, Q.; SMITH, D. W. Developing artificial neural network models of water treatment processes: a guide for utilities. *J. Environ. Eng. Sci.*, v. 1, p. 201-211, 2002.

BAYAT, F. M.; PREZIOSO, M.; CHAKRABARTI, B.; NILI, H.; KATAEVA, I.; STRUKOV, D. Implementation of multilayer perceptron network with highly uniform passive memristive crossbar circuits. *NATURE COMMUNICATIONS*, v. 9, n. 2331, p. 1-7, 2018.

BERESKIE, T.; HAIDER, H.; RODRIGUEZ, M. J.; SADIQ, R. Framework for continuous performance improvement in small drinking water systems. *Science of the Total Environment*, v. 574, p. 1405-1414, 2017.

BIGIARINI, M. Z. Package 'hydroGOF' - 2017. Disponível em: < <https://cran.r-project.org/web/packages/hydroGOF/hydroGOF.pdf>>. Acesso em: 12 fev. 2019

BONTON, A.; BOUCHARD, C.; BARBEAU, B.; JEDRZEJAK, S. Comparative life cycle assessment of water treatment plants. *Desalination*, v. 284, p. 42-54, 2012.

BORGHI, A. D.; STRAZZA, C.; GALLO, M.; MESSINEO, S.; NASO, M. Water supply and sustainability: life cycle assessment of water collection, treatment and distribution service. *The International Journal of Life Cycle Assessment*, v. 18, p. 1158-1168, 2013.

BRASIL. Ministério da Saúde. *PORTARIA DE CONSOLIDAÇÃO Nº 5*. Consolidação das normas sobre as ações e os serviços de saúde do Sistema Único de Saúde. Brasília: Ministério da Saúde, 2017.

Canadian Council of Ministers of the Environment - CCME. *From Source to Tap: Guidance on the Multi-Barrier Approach to Safety Drinking Water*. Winnipeg: Canadian Council of Ministers of the Environment, 2004.

CAPITANESCU, F.; REGE, S.; MARVUGLIA, A.; BENETTO, E.; AHMADI, A.; GUTIÉRREZ, T. N.; TIRUTA-BARNA, L. Cost versus life cycle assessment-based environmental impact optimization of drinking water production plants. *Journal of Environmental Management*, v. 177, p. 278-287, 2016.

CHANG, E.-E.; CHIANG, P.-C.; HUANG, S.-M.; LIN, Y.-L. Development and Implementation of Performance Evaluation System for a Water Treatment Plant: Case Study of Taipei Water Treatment Plant. *PRACTICE PERIODICAL OF HAZARDOUS, TOXIC, AND RADIOACTIVE WASTE MANAGEMENT*, v. 11, n. 1, p. 36-47, 2007.

CHEN, Y.; CAI, K.; TU, Z.; NIE, W.; JI, T.; HU, B.; CHEN, C.; JIANG, S. Prediction of benzo[a]pyrene content of smoked sausage using back-propagation artificial neural network. *J Sci Food Agric*, v. 98, p. 3022-3030, 2018.

CHOU, J.-S.; HO, C.-C.; HOANG, H.-S. Determining quality of water in reservoir using machine learning. *Ecological Informatics*, v. 44, p. 57-75, 2018.

Conselho Estadual de Política Ambiental – COPAM. Deliberação Normativa Copam nº 217, de 06 de dezembro de 2017. Disponível em: <http://www.udop.com.br/download/legislacao/meio/institucional_site_juridico/deliberacao_normativa_n217.pdf>. Acesso em: 18 fev. 2019.

CORDOBA, G. A. C.; TUHOVČÁKA, L.; TAUŠ, M. Using artificial neural network models to assess water quality in water distribution networks. *Procedia Engineering*, v. 70, p. 399-408, 2014.

CORNWELL, D. A. *Water treatment residuals engineering*. Denver: AWWA Research Foundation and American Water Works Association, 2006.

DA SILVA, I. N.; SPATTI, D. H.; FLAUZINO, R. A. *Redes Neurais Artificiais: Para Engenharia e Ciências Aplicadas*. 2 ed. São Paulo: Artliber, 2016.

DOHARE, D.; GUPTA, V. REVIEW ON PERFORMANCE EVALUATION OF WATER TREATMENT PLANT AND WATER QUALITY INDEX. *INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH*, v. 4, n. 2, p. 101-113, 2014.

FAN, M.; HU, J.; CAO, R.; RUAN, W.; WEI, X. A review on experimental design for pollutants removal in water treatment with the aid of artificial intelligence. *Chemosphere*, v. 200, p. 330-343, 2018.

FARHAOUI, M.; DERRAZ, M. Review on Optimization of Drinking Water Treatment Process. *Journal of Water Resource and Protection*, v. 8, p. 777-786, 2016.

FERREIRA, A. A. *COMPARAÇÃO DE ARQUITETURAS DE REDES NEURAIS PARA SISTEMAS DE RECONHECIMENTO DE PADRÕES EM NARIZES ARTIFICIAIS*. 2004. 119f. Dissertação (Mestrado em Ciência da Computação) – Centro de Informática, UFPE, Recife, 2004.

FRITSCH, S.; GUENTHER, F.; WRIGHT, M. N.; SULING, M.; MUELLER, S. M. *Package 'neuralnet'*. Disponível em: <<https://cran.r-project.org/web/packages/neuralnet/neuralnet.pdf>>. Acesso em: 14 fev. 2019.

GARFÍ, M.; CADENA, E.; SANCHEZ-RAMOS, D.; FERRER, I. Life cycle assessment of drinking water: Comparing conventional water treatment, reverse osmosis and mineral water in glass and plastic bottles. *Journal of Cleaner Production*, v. 137, p. 997-1003, 2016.

GAZZAZ, N. M.; YUSOFF, M. K.; ARIS, A. Z.; JUAHIR, H.; RAMLI, M. F. Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. *Marine Pollution Bulletin*, v. 64, p. 2409-2420, 2012.

GHOLIKANDI, G. B.; DEGHANIFARD, E.; SEPEHR, M. N.; TORABIAN, A.; MOALEJ, S.; DEHNAVI, A.; YARI, A. R.; ASGARI, A. R. Performance Evaluation of Different Filter Media in Turbidity Removal from Water by Application of Modified Qualitative Indices. *Iranian Journal of Public Health*, v. 41, n. 4, p. 87-93, 2012.

GHORBANI, M. A.; DEO, R. C.; YASEEN, Z. M.; KASHANI, M. H.; MOHAMMADI, B. Pan evaporation prediction using a hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model: case study in North Iran. *Theor Appl Climatol*, v. 133, p. 1119-1131, 2018.

GIWA, A.; DAER, S.; AHMED, I.; MARPU, P. R.; HASAN, S. W. Experimental investigation and artificial neural networks ANNs modeling of electrically- enhanced membrane bioreactor for wastewater treatment. *Journal of Water Process Engineering*, v. 11, p. 88-97, 2016.

GRAUPE, D. *PRINCIPLES OF ARTIFICIAL NEURAL NETWORKS*. 2 ed. Chicago: World Scientific, 2007.

GUNDERSON, L.H.; PRITCHARD, L. *Resilience and the Behavior of Large-scale Systems*. Island Press, Washington – DC, 2002.

GUPTA, A.K.; SHRIVASTAVA, R.K. Uncertainty analysis of conventional water treatment plant design for suspended solids removal. *Journal of Environmental Engineering*, v. 132, n.11, p. 1413-121, 2006.

HAYKIN, S. *REDES NEURAIS, Princípios e práticas*. 2 ed. Porto Alegre: Bookman, 2001.

HE, Y.-L.; XU, Y.; ZHU, Q.-X. Soft-sensing model development using PLSR-based dynamic extreme learning machine with an enhanced hidden layer. *Chemometrics and Intelligent Laboratory Systems*, v. 154, p. 101-111, 2016.

HEDDAM, S.; LAMDA, H.; FILALI, S. Predicting Effluent Biochemical Oxygen Demand in a Wastewater Treatment Plant Using Generalized Regression Neural Network Based Approach: A Comparative Study. *Environmental Processes*, v. 3, p. 153-165, 2016.

HEIDARI, A. A.; FARIS, H.; ALJARAH, I.; MIRJALILI, S. An efficient hybrid multilayer perceptron neural network with grasshopper optimization. *Soft Computing*, v. 22, n. 233, p. 1-18, 2018.

IGOS, E.; BENETTO, E.; BAUDIN, I.; TIRUTA-BARNA, L.; MERY, Y.; ARBAULT, D. Cost-performance indicator for comparative environmental assessment of water treatment plants. *Science of The Total Environment*, v. 443, p. 367-374, 2013.

Internacional City Council Management Association (ICMA). *Beyond data: Current uses of comparative performance measurement in local government*. Washington: ICMA, 1999.

Instituto Brasileiro de Geografia e Estatística (IBGE). *PESQUISA NACIONAL DE SANEAMENTO BÁSICO 2008*. Rio de Janeiro: IBGE, 2008.

_____. *Censo demográfico*. Disponível em: <<http://www.ibge.gov.br/home/>>. Acesso em: 02 out 2014.

_____. *Agência IBGE Notícias*. Disponível em: <<https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/22374-ibge-divulga-as-estimativas-de-populacao-dos-municipios-para-2018>>. Acesso em: 18 fev. 2019.

KARADURMUŞ, E.; TAŞKIN, N.; GÖZ, E.; YÜCEER, M. Prediction of Bromate Removal in Drinking Water Using Artificial Neural Networks. *OZONE: SCIENCE & ENGINEERING*, v. 40, p. 1-11, 2018.

KASABOV, N. K. *Foundations of Neural Networks, Fuzzy System, and Knowledge Engineering*. Londres: The MIT Press, 1996.

KAUFMAN, R. Preparing useful performance indicators. *Training Development*, v. 42, n. 9, p. 80-83, 1988.

KASIVISWANATHAN, K. S.; CIBIN, R.; SUDHEER, K. P.; CHAUBEY, I. Constructing prediction interval for artificial neural network rainfall runoff models based on ensemble simulations. *Journal of Hydrology*, v. 499, p. 275-288, 2013.

KHAN, A. A.; AHMAD, S. K. PERFORMANCE EVALUATION OF WATER TREATMENT PLANT AT NANGLOI, NEW DELHI: A CASE STUDY. *International Journal of Research in Engineering and Technology*, v. 7, n. 8, p. 49-60, 2018.

KIM, C. M.; PARNICHKUN, M. Prediction of settled water turbidity and optimal coagulant dosage in drinking water treatment plant using a hybrid model of k-means clustering and adaptive neuro-fuzzy inference system. *Applied Water Science*, v. 7, n. 7, p. 3885–3902, 2017.

KULKARNI, P.; CHELLAM, S. Disinfection by-product formation following chlorination of drinking water: Artificial neural network modes and changes in speciation with treatment. *Science of the total environment*, v. 408, p. 4202-4210, 2010.

LEEPER, T. J.; GANZ, C. *Package 'prediction'*. Disponível em: < <https://cran.r-project.org/web/packages/prediction/prediction.pdf>>. Acesso em: 14 fev. 2019.

LOPES, V. C. *Índice de qualidade de desempenho de estações convencionais de tratamento de água*. 2005. 217f. Dissertação (Mestrado em Saneamento, Meio Ambiente e Recursos Hídricos) - Escola de Engenharia, UFMG, Belo Horizonte, 2005.

LOPES, V. C. & LIBÂNIO, M. Assessing the feasibility of a water treatment plant quality index. *Journal of Water Supply: Research and Technology – AQUA*, v. 58, n. 5, p. 354-362, 2009.

MADHU, K. M.; SOUMYASHREE, S. H. Performance Evaluation of a Water Treatment Plant at Davangere-(Karnataka). *International Research Journal of Engineering and Technology*, v. 3, n. 9, p. 1467-1468, 2016.

MAIER, H. R.; JAIN, A.; DANDY, G. C.; SUDHEER, K. P. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Environmental Modelling & Software*, v. 25, p. 891-909, 2010.

MAIER, H. R.; MORGAN, N.; CHOW, C. W. K. Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters. *Environmental Modelling & Software*, v. 19, p. 485-494, 2004.

MAKUNGO, R.; ODIYO, J. O.; TSHIDZUMBA, N. Performance of small water treatment plants: The case study of Mutsheds Water Treatment Plant. *Physics and Chemistry of the Earth*, v. 36, p. 1151-1158, 2011.

MARZOUK, M.; ELKADI, M. Estimating water treatment plants costs using factor analysis and artificial neural networks. *Journal of Cleaner Production*, v. 112, p. 4540-4549, 2016.

MELO, L. V.; OLIVEIRA, M. D. de; LIBÂNIO, M.; OLIVEIRA, S. C. Applicability of statistical tools for evaluation of water treatment plants. *Desalination and Water Treatment*, v. 57, n. 30, p. 1-10, 2015.

MEMON, N. A.; UNAR, M. A.; ANSARI, A. K. pH Prediction by Artificial Neural Networks for the Drinking Water of the Distribution System of Hyderabad City. *Mehran University Research Journal of Engineering & Technology*, v. 31, n. 1, p. 137-146, 2012.

MISHRA, A. R.; KADU, P. A. Performance Evaluation of Water Treatment Plant at Yavatmal (M.S.): Case Study. *International Journal of Research in Advent Technology*, v. 2, n. 5, p. 455-458, 2014.

MOTA, M. H.; CHOUGULE, S. S.; VATKAR, Y. S. Performance Evaluation of Urban Water Treatment Plant. *International Journal of Science and Research*, v. 4, n. 4, p. 2821-2825, 2015.

NASR, M. S.; MOUSTAFA, M. A. E.; SEIF, H. A. E.; EL KOBROSY, G. Application of Artificial Neural Network (ANN) for the prediction of EL-AGAMY wastewater treatment plant performance-EGYPT. *Alexandria Engineering Journal*, v. 51, p. 37-43, 2012.

NOURANI, V.; BAGHANAM, A. H.; ADAMOWSKI, J.; KISI, O. Applications of hybrid wavelet–Artificial Intelligence models in hydrology: A review. *Journal of Hydrology*, v. 514, p. 358-377, 2014.

O'CONNOR, J. T.; O'CONNOR, T.; TWAIT, R. *WATER TREATMENT PLANT PERFORMANCE EVALUATIONS AND OPERATIONS*. New Jersey: John Wiley & Sons, 2009.

OLAWOYIN, R. Application of backpropagation artificial neural network prediction model for the PAH bioremediation of polluted soil. *Chemosphere*, v. 161, p. 145-150, 2016.

OLIVEIRA, M. D. de. *DESENVOLVIMENTO DE MODELOS DE PREVISÃO DE DESEMPENHO DE ESTAÇÕES CONVENCIONAIS DE TRATAMENTO DE ÁGUA*. 2014. 225f. Tese (Doutorado em Saneamento, Meio Ambiente e Recursos Hídricos) - Escola de Engenharia, UFMG, Belo Horizonte, 2014.

OLIVEIRA, M. D. de; REZENDE, O. L. T. de; OLIVEIRA, S. M. A. C.; LIBÂNIO, M. Nova abordagem do Índice de Qualidade de Água Bruta utilizando a Lógica Fuzzy. *Engenharia Sanitária e Ambiental*, v. 19, n. 4, p. 361-372, 2014.

Organização das Nações Unidas (ONU). *17 Objetivos para transformar nosso mundo*. Disponível em: <<https://nacoesunidas.org/pos2015/>>. Acesso em: 01 fev 2019.

Organização das Nações Unidas (ONU). *Notícia e Mídia, Rádio ONU*. Disponível em: <<http://www.unmultimedia.org/radio/portuguese/2015/03/onu-alerta-que-783-milhoes-nao-tem-acesso-a-agua-no-mundo/#.WASrVMmYKwk>>. Acesso em: 17 out 2016.

QADERI, F.; BABANEZHAD, E. Prediction of the groundwater remediation costs for drinking use based on quality of water resource, using artificial neural network. *Journal of Cleaner Production*, v. 161, p. 840-849, 2017.

RAK, A. Water Turbidity Modelling During Water Treatment Processes Using Artificial Neural Networks. *International Journal of Water Sciences*, v. 2, n. 3, p. 1-10, 2013.

- RAIKAR, R. V.; WANG, C.-Yi.; SHIH, H.-P.; HONG, J.-H. Prediction of contraction scour using ANN and GA. *Flow Measurement and Instrumentation*, v. 50, p. 26-34, 2016.
- RODRIGUEZ, O. O. O.; VILLAMIZAR-GALLARDO, R. A.; GARCÍA, R. G. Life cycle assessment of four potable water treatment plants in northeastern Colombia. *Ambiente & Água – A Interdisciplinary Journal of Applied Science*, v. 11, n. 2, p. 268-278, 2016.
- RUSSELL, S.; NORVIG, P. *Inteligência Artificial*. 3ed. Rio de Janeiro: Elsevier, 2013.
- SADHWANI, K. M.; BHAVE, P. P. A review of Performance Evaluation and Performance Optimisation of Water Treatment Plant. *International Advanced Research Journal in Science, Engineering and Technology*, v. 3, n. 7, p. 157-162, 2016.
- SADIQ, R.; RODRIGUES, M. J.; TEFAMARIAM, S. Integrating indicators for performance assessment of small water utilities using ordered weighted averaging (OWA) operators. *Expert systems with applications*, v. 37, p. 4881-4891, 2010.
- SARKAR, A.; PANDEY, P. River Water Quality Modelling using Artificial Neural Network Technique. *Aquatic Procedia*, v. 4, p. 1070-1077, 2015.
- SCHMOLDT, D.; KANGAS, J.; MENDOZA, G. A.; PESONEN, M. *The Analytic Hierarchy Process in Natural Resource and Environmental Decision Making*. v. 3. Netherlands: Springer, 2001. (Managing Forest Ecosystems).
- STEIN, D.; ACHARI, G.; LANGFORD, C. H.; DORE, M. H. I.; HAIDER, H.; ZHANG, K.; SADIQ, R. Performance management of small water treatment plant operations: a decision support system. *Water and Environment Journal*, v. 31, n. 3, p. 330-344, 2017.
- SWEETAPPLE, C.; FU, G.; BUTLER, D. Multi-objective optimisation of wastewater treatment plant control to reduce greenhouse gas emissions. *Water Research*, v. 55, p. 52-62, 2014.
- TOMAR, M. M.; BORAD, N. N. Use of AHP Method in Efficiency Analysis of Existing Water Treatment Plants. *Internacional Journal of Engineering Research and Development*, v. 1, n. 7, p. 42-51, 2012.
- TRIOLA, M. F. *INTRODUÇÃO À ESTATÍSTICA Atualização da Tecnologia*. 11 eds. Rio de Janeiro: LTC, 2013.
- TÜMER, A. E.; EDEBALI, S. An Artificial Neural Network Model for Wastewater Treatment Plant of Konya. *International Journal of Intelligent Systems and Applications in Engineering*, v. 3, n. 4, p. 131-135, 2015.
- United States Environmental Protection Agency - USEPA. *Optimizing Water Treatment Plant Performance Using the Composite Correction Program*. Washington: USEPA, 1998.
- VÁSQUEZ, E. A. O.; MALDONADO, J. L. B.; DELGADO, D. E. J.; MALDONADO, J. S. T.; ALFARO, J. F. O. Determinación del modelo matemático basado en redes neuronales del

sistema dedosificação de sulfato de alumínio y cal de la planta de tratamiento de agua potable Carigán. *Brazilian Journal of Development*, v. 5, n. 3, p. 2162-2172, 2019.

VIEIRA, P.; ALEGRE, H.; ROSA, M. J.; LUCAS, H. Drinking water treatment plant assessment through performance indicators. *Water Science & Technology: Water Supply*, v. 8, n. 3, p. 245-253, 2008.

VOLPI, A. *Ferramentas de Python para Aprendizado de Máquina*. Disponível em: <<https://alexandrevolpi.wordpress.com/2015/10/02/ferramentas-de-python-para-aprendizado-de-maquina/>>. Acesso em: 09 out. 2018.

WALLACE, J.; CHAMPAGNE, P.; HALL, G. Multivariate statistical analysis of water chemistry conditions in three wastewater stabilization ponds with algae blooms and pH fluctuations. *Water Research*, v. 96, p. 155-165, 2016

World Health Organization – WHO. *Guidelines for drinking-water quality: fourth edition incorporating the first addendum*. Switzerland, 2017.

WU, H.; LIN, L.; ZHANG, J.; GUO, W.; LIANG, S.; LIU, H. Purification ability and carbon dioxide flux from surface flow constructed wetlands treating sewage treatment plant effluent. *Bioresource Technology*, v. 219, p. 768-772, 2016.

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

WU, W.; DANDY, G. C.; MAIER, H. R. Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modelling. *Environmental Modelling & Software*, v. 54, p. 108-127, 2014.

ZAKARIAN, A., KNIGHT, J., BAGHDASARYAN, L. Modelling and analysis of system robustness. *Journal of Engineering Design*, v. 18, n. 3, p. 243-263, 2007.

ZHANG, K.; ACHARI, G.; SADIQ, R.; LANGFORD, C. H.; DORE M. H. I. An integrated performance assessment framework for water treatment plants. *Water Research*, v.46, n. 6, p. 1673-1683, 2012.