

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA SANITÁRIA E AMBIENTAL
CURSO DE ENGENHARIA SANITÁRIA E AMBIENTAL

Emanuel Goulart Farias

Aplicação de Rede Neural Artificial para Predição da Qualidade do Efluente Tratado
em *Wetland* Construído Vertical de Fundo Saturado

Florianópolis

2022

Emanuel Goulart Farias

Aplicação de Rede Neural Artificial para Predição da Qualidade do Efluente Tratado
em *Wetland* Construído Vertical de Fundo Saturado

Trabalho de Conclusão de Curso submetido ao curso de Engenharia Sanitária e Ambiental do Centro Tecnológico da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do título de Bacharel em Engenharia Sanitária e Ambiental.

Orientador(a): Prof. Dr. Pablo Heleno Sezerino
Coorientador(a): Me. Marcelo Demétrio de Magalhães

Florianópolis

2022

Farias, Emanuel Goulart

Aplicação de rede neural artificial para predição da qualidade do efluente tratado em wetland construído vertical de fundo saturado / Emanuel Goulart Farias ; orientador, Pablo Heleno Sezerino, coorientador, Marcelo Demétrio de Magalhães, 2022.

100 p.

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Santa Catarina, Centro Tecnológico, Graduação em Engenharia Sanitária e Ambiental, Florianópolis, 2022.

Inclui referências.

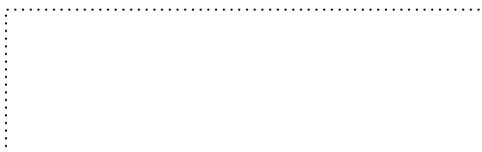
1. Engenharia Sanitária e Ambiental. 2. Wetlands Construídos. 3. Redes Neurais Artificiais. 4. Inteligência Artificial. 5. Tratamento de Efluentes. I. Sezerino, Pablo Heleno. II. de Magalhães, Marcelo Demétrio. III. Universidade Federal de Santa Catarina. Graduação em Engenharia Sanitária e Ambiental. IV. Título.

Emanuel Goulart Farias

Aplicação de Rede Neural Artificial para Predição da Qualidade do Efluente Tratado
em *Wetland* Construído Vertical de Fundo Saturado

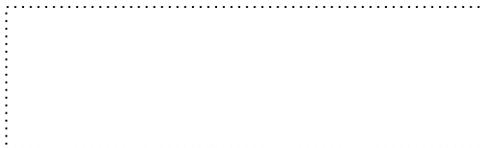
Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do título de Bacharel em
Engenharia Sanitária e Ambiental e aprovado em sua forma final pelo Curso Engenharia Sanitária e
Ambiental

Florianópolis, 12 de Dezembro de 2022.



Coordenação do Curso

Banca examinadora



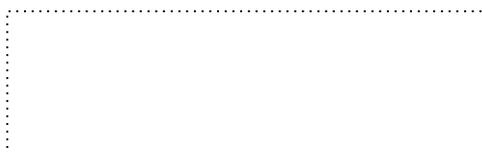
Prof. Pablo Heleno Sezerino

Orientador(a)



Alexandre Bach Trevisan

Casan



Ágata Bogado Dominguez

PPGEA - UFSC

Florianópolis, 2022

AGRADECIMENTOS

Findo esse ciclo como os dias findam as noites e os pássaros não hesitam na melodia das manhãs. Basta pequenos movimentos na alvorada para lembrar que sou amparado constantemente por todos aqueles que de alguma forma cruzaram por meu caminho.

Um recado especial para meus pais: Gerson e Jacque que não mediram esforços em minha formação como ser humano e me guiaram pé por pé com constância e paciência, me falta palavras para descrever tamanha bondade e compaixão. Meu querido irmão Gibram que me atura na mais tardar das horas, meus amáveis avós: Dona Maria, Seu Zanoni e Seu Lelo pela referência e exemplo dos gestos e ações estornam em mim. E o afeto de Bella, companhia ilimitada com sinceridade profunda nos dias de sol e nos dias de chuva.

Agradeço com franqueza a todos os amigos de minha terra natal -- Araranguá que consolidam uma ampla gama de visões e aprendizados, sendo um grande baú de riqueza diversa esses carinhosos amigos. Muito obrigado a todos os meus amigos e colegas construídos na Ilha do Desterro, há de ter sido dias mais leves diante da presença de gente querida e amada. Receio em citar nomes pois escrevo mediante prazos curtos e não gostaria de esquecer alguém, então a todos que leem e me conhecem por algum apelido saiba que nutro de reciprocidade pela companhia, felicidade por todas as trocas vividas e sorriso desaguado em mar mediante sua presença.

Agradeço ao professor Pablo que me instruiu e abraçou as ideias e causas, meu amigo e co-orientador Marcelo que mesmo atarefado se dispôs de peito aberto com o projeto, a Alexandre e Ágata pelas considerações, meu grande amigo e primo do peito Quincas Avelino por ensinar sobre concordância nominal de malandro e aos demais professores e colegas do departamento de Engenharia Sanitária e Ambiental bem como da UFSC por levarem a docência e pesquisa de ponta para todos.

Obrigado a Universidade Federal de Santa Catarina por seu ensino público e de qualidade, que todas as pessoas tenham a oportunidade de educação de qualidade nesse país tão bonito e diverso.

Ademais, deixo aqui meus sinceros agradecimentos sem mais citações nominiais pois entendo sobre a importância equânime de cada um na primavera de meus dias, há contribuição de todos para constante formação como pessoa, nutro de profundo carinho por cada um.

Assim como é, as coisas são como são.

Mestre Zen Yogo Roshi

RESUMO

A utilização da ecotecnologia dos *wetlands* construídos verticais de fundo saturado (WCV-FS) no tratamento de esgoto doméstico apresenta larga aplicabilidade para o cenário do saneamento na realidade brasileira, apresentando requisitos construtivos e operacionais atrativos aos sistemas descentralizados. Sua eficiência de remoção da demanda química de oxigênio (DQO) é consolidada e assegura um critério de qualidade operacional no tratamento do efluente. Nesse contexto, o presente trabalho avalia a aplicabilidade de Redes Neurais Artificiais (RNA) para a predição da DQO efluente do WCV-FS, utilizando como variáveis preditoras dados de qualidade da água do esgoto afluente e dados meteorológicos. Esta pesquisa utilizou técnicas de pré-processamento dos dados como Análise de Componentes Principais (ACP) e Fator de Influência de Predição, para seleção de componentes principais e redução de dimensionalidade. Quatro conjunto de dados foram formados para comparar como diferentes parâmetros de qualidade da água e meteorologia impactam no processo de predição da DQO através das RNAs, sendo o desempenho avaliado pelo Erro Quadrático Médio, Coeficiente de Regressão e Coeficiente de Regressão Ajustado (R^2 ajustado). O conjunto de dados transformados por ACP obteve a RNA os melhores resultados (R^2 ajustado = 0.676 e EQM = 244.81). Todos os modelos previram corretamente as tendências de desempenho de remoção de DQO nos WCV-FS, indicando uma viabilidade significativa do uso de RNA para predição de remoção de matéria orgânica dos WCV-FS, promovendo dessa forma, uma ferramenta capaz de auxiliar nas tomadas de decisão dentro dos processos operacionais do sistema de tratamento, bem como para efeitos de projeto.

Palavras-chave: *Wetlands*; Redes Neurais Artificiais. Análise de Componentes Principais; *Deep Learning*

ABSTRACT

The sanitation scenery in Brazil presents a very suitable opportunity for the use of constructed vertical wetlands with a saturated background in the treatment of its domestic sewage. This ecotechnology displays attractive operational and constructive requirements for decentralized systems. Its efficiency in the elimination of chemical oxygen demand (COD) is well established and it assures the quality criteria for effluent treatment operations. Given this background, the present work evaluates the applicability of Artificial Neural Networks (ANN) to predict the effluent COD from VSCW, taking in predictive variables from water data provided by the inflowing sewage and meteorological data. This research used data preprocessing techniques such as Principal Component Analysis (PCA) and Prediction Influence Factor to select the most relevant parameters and reduce the dimensionality. Four datasets were raised to compare how distinct water quality and meteorological parameters influence the COD prediction process through ANNs. The performance of the ANNs was evaluated using the Mean Squared Error, Regression Coefficient and Adjusted Regression Coefficient (Adjusted R^2). The ACP-transformed dataset obtained the best results for the ANN (adjusted $R^2 = 0.676$ and RMSE = 244.81). All models correctly predicted the trends of COD removal performance in VSCW, indicating a significant viability for the usage of ANNs in the evaluation and prediction of organic matter removal from VSCW, promoting a tool capable of aiding the decision making within the operational processes of the treatment system, as well as assisting in design purposes.

Keywords: Constructed Wetlands; Artificial Neural Network; Component Principal Analysis; Deep-Learning

LISTA DE FIGURAS

Figura 1 - Representação longitudinal de um WCV-FS.....	21
Figura 2 Modelo de um neurônio.	25
Figura 3 Representação do efeito do viés no termo uk	26
Figura 4 - Gráfico de uma função ReLU.....	27
Figura 5 - Modelo de uma Rede Neural Artificial de Multicamadas.....	28
Figura 6 - Mapa de Localização da EETE – GESAD/UFSC.....	36
Figura 7 - Arranjo dos <i>Wetlands</i> Construídos componentes da EETE – GESAD/UFSC.	37
Figura 8 - Controlador de nível de saída do WCV-FS componente da EETE – GESAD/UFSC.....	38
Figura 9 - Boxplot das diferenças percentuais entre as estações meteorológicas	44
Figura 10 – Distribuição dos dados para o parâmetro de DQO afluente ao WCV-FS componente da EETE – GESAD/UFSC, após e anterior a transformação boxcox.....	57
Figura 11 – Diagrama de caixas para efeito da parametrização no conjunto de dados para o parâmetro de DQO afluente ao WCV-FS componente da EETE – GESAD/UFSC	58
Figura 12 - Matriz de Correlação dos Parâmetros Selecionados referente a série histórica de dados afluente, efluente e climáticos influentes no módulo WCV-FS da EETE GESAD/UFSC.....	61
Figura 13 – Variância explicada e variância acumulada para cada componente principal.....	63
Figura 14 – Resultado das RNAs para os conjuntos de dados com diferentes variáveis de entrada ao módulo WCV-FS componente da EETE – GESAD/UFSC.....	67
Figura 15 - Diagrama de caixa para DQO efluente do módulo WCV-FS componente da EETE – GESAD/UFSC.. ..	68
Figura 16 - Erros Residuais das RNAs e suas distribuições para cada conjunto de dados referentes ao módulo WCV-FS componente da EETE – GESAD/UFSC.....	69

LISTA DE TABELAS

Tabela 1 - Características Construtivas do WCV-FS.....	37
Tabela 2 - Características Operacionais do WCV-FS.....	38
Tabela 3 – Parâmetros de Qualidade Analisados no afluente e efluente do WCV-FS componente da EETE – GESAD/UFSC.....	39
Tabela 4 - Percentagem de valores nulos para os parâmetros de qualidade avaliados junto a unidade WCV-FS componente da EETE – GESAD/UFSC.....	39
Tabela 5 - Trecho da tabela de Aferição da Vazão da Bomba dos WCV-FS	40
Tabela 6 - Parâmetros de qualidade incorporados no conjunto de dados.	41
Tabela 7 - Descrição dos parâmetros meteorológicos.....	42
Tabela 8 - Agregação dos Dados Meteorológicos.....	42
Tabela 9 - Porcentagem dos dados meteorológicos faltantes.....	43
Tabela 10 - Parâmetros Incorporados nos dados meteorológicos	45
Tabela 11 - Novos datasets a partir da ACPs.	50
Tabela 12 - Conjuntos de Dados utilizados para construção da RNA.	51
Tabela 13 - Composição dos Parâmetros que serão testados na RNA.....	54
Tabela 14 – Relação de importância de cada parâmetro no processo de predição do modelo XGBoost.....	59
Tabela 15 - Parâmetros Seleccionados.....	60
Tabela 16 - Parâmetros seleccionados da matriz de correlação.....	62
Tabela 17 – Lista de Parâmetros Seleccionados por CPs para 90% da variância explicada.....	64
Tabela 18 – Melhor Topologia Obtida para cada dataset.....	65
Tabela 19 - Resultado das Redes Neurais para cada dataset.....	65

LISTA DE ABREVIATURAS E SIGLAS

ACP	Análise de Componente Principal
ANN	Artificial Neural Network
CCA	Centro de Ciências Agrárias
CP	Componentes Principais
DBO	Demanda Bioquímica de Oxigênio
DQO	Demanda Química de Oxigênio
EETE	Estação Experimental de Esgoto (EETE)
EPAGRI	Empresa de Pesquisa Agropecuária e Extensão
EQM	Erro quadrático Médio
GESAD	Grupo de Estudos de Saneamento Descentralizado
INMET	Instituto Nacional de Meteorologia
LabClimAgri	Laboratório de Climatologia Agrícola da UFSC
LABHIDRO	Laboratório de Hidrologia da UFSC
OMS	Organização Mundial da Saúde
PLANSAB	Plano Nacional de Saneamento Básico
ReLu	Rectified Linear Activation Unit
RLN	Regressão Linear Múltipla
RNA	Redes Neurais Artificiais
SNIS	Sistema Nacional de Informações de Saneamento
UFSC	Universidade Federal de Santa Catarina
WC	<i>Wetlands</i> Construídos
WCH	<i>Wetland</i> Construído Horizontal
WCV	<i>Wetland</i> Construído Vertical
WCV-FS	<i>Wetland</i> Construído Vertical de Fundo Saturado
WCV-FS	<i>Wetlands</i> Construídos Verticais de Fundo Saturado

LISTA DE ABREVIATURAS E SIGLAS

gDQO/m ² .d	Gramas de DQO por metro quadrado dia
hPa	Hectopascal
L/d	Litros por dia
m ²	Metro Quadrado
m ³	Metro Cúbico
m ³ /m ² .d	Metros Cúbicos por metros quadrado dia
mg/L	Miligramas por Litro
mm	Milímetros
Seg	Segundos
W/m ²	Watt-hora por metro quadrado

SUMÁRIO

1	INTRODUÇÃO	18
2	OBJETIVOS	20
3	OBJETIVOS ESPECÍFICOS.....	20
4	REVISÃO BIBLIOGRÁFICA	21
4.1	<i>WETLAND</i> CONSTRUÍDO VERTICAL DE FUNDO SATURADO.....	21
4.2	TRANSFORMAÇÃO E REMOÇÃO DE MATÉRIA ORGÂNICA CARBONÁCEA EM WCV-FS	22
4.3	ANÁLISE DE DADOS APLICÁVEL À PREDIÇÃO E CÁLCULO DE DESEMPENHO DE TRATAMENTO EM WCV-FS	23
4.3.1	Análise de Componentes Principais (ACP).....	23
4.3.2	Redes Neurais Artificiais.....	24
4.3.2.1	<i>Função de Ativação.....</i>	26
4.3.2.2	<i>Arquitetura de Redes Neurais.....</i>	27
4.3.2.3	<i>Processo de Aprendizagem.....</i>	28
4.3.3	Validação do Modelo	32
4.3.3.1	<i>Erro Quadrático Médio (EQM).....</i>	32
4.3.3.2	<i>Erro Quadrático Médio Normalizado (EQMN)</i>	32
4.3.3.3	<i>Coefficiente de Determinação Múltipla R^2</i>	33
4.3.3.4	<i>Coefficiente Ajustado de Determinação Múltipla</i>	34
4.3.4	Testes Estatísticos de Distribuição Normal.....	34
4.3.4.1	<i>Shapiro-Wilk.....</i>	34
4.3.4.2	<i>Anderson-Darling.....</i>	35
5	METODOLOGIA.....	35
5.1	DADOS DE QUALIDADE DOS EFLUENTES	36
5.1.1	Característica Operacional do WCV-FS.....	38
5.1.2	Descrição dos Dados de Qualidade	38
5.1.3	Análise de Inconsistências e Preenchimento de Valores	39
5.1.4	Adequação de Novos Parâmetros de Qualidade.....	40
5.2	DADOS METEOROLÓGICOS	41
5.2.1	Análise de Inconsistências.....	42
5.2.2	Preenchimento de Valores Nulos.....	43
5.2.3	Adequação de Novas Variáveis	45

5.3	UNIÃO DOS DADOS DE QUALIDADE E METEOROLÓGICOS.....	45
5.4	PRÉ PROCESSAMENTO DOS DADOS.....	46
5.4.1	Análise Descritiva	46
5.4.2	– Padronização dos Dados	46
5.4.3	Transformação Linear dos Dados.....	46
<i>5.4.3.1</i>	<i>Transformação BoxCox.....</i>	<i>47</i>
5.5	ANÁLISE DE COMPONENTES PRINCIPAIS (ACPS).....	47
5.6	SELEÇÃO DE PARÂMETROS	50
5.7	SELEÇÃO DO CONJUNTO DE DADOS	51
5.8	REDE NEURAL ARTIFICIAL	52
5.8.1	–Divisão do Conjunto de dados	52
5.8.2	Topologia das RNAs	53
5.8.3	- Validação do Modelo.....	54
6	RESULTADOS	54
6.1	ANÁLISE DESCRITIVA	55
<i>6.1.1.1</i>	<i>Distribuição Normal.....</i>	<i>56</i>
<i>6.1.1.2</i>	<i>Transformação Linear - BoxCox.....</i>	<i>57</i>
6.2	PADRONIZAÇÃO DOS DADOS	58
6.3	REDUÇÃO DE DIMENSIONALIDADE	58
6.4	COMPONENTES PRINCIPAIS	63
6.5	TOPOLOGIA E RESULTADO DAS REDES NEURAS	65
7	CONCLUSÃO.....	71
	REFERÊNCIAS.....	74
	APÊNDICE A – MAPA CONCEITUAL DA METODOLOGIA UTILIZADA NO PRESENTE TRABALHO	78
	APÊNDICE B – CONJUNTO DE DADOS DE METEOROLOGIA E QUALIDADE HÍDRICA.....	80
	APÊNDICE C – ANÁLISE ESTATÍSTICA DO CONJUNTO DE DADOS....	85
	APÊNDICE D – TESTES DE DISTRIBUIÇÃO NORMAL PARA OS PARÂMETROS DO CONJUNTO DE DADOS.....	86
	APÊNDICE E – IMPORTÂNCIA DE CADA VARIÁVEL NO PROCESSO DE PREDIÇÃO DO MODELO XGBOOST.....	87

APÊNDICE F – PROPORÇÃO DE VARIÂNCIA EXPLICADA E VARIÂNCIA EXPLICADA ACUMULADA PARA CADA COMPONENTE PRINCIPAL.....	88
APÊNDICE G – VALORES DE <i>LOADINGS</i> DAS COMPONENTES PRINCIPAIS PARA CADA PARÂMETRO.....	89
ANEXO A – DESCRIÇÃO.....	94
ANEXO B – TABELA DE AFERIÇÃO DA VAZÃO DA BOMBA DOS <i>WETLANDS</i> CONSTRUÍDOS DO GESAD.	96

1. INTRODUÇÃO

Os sistemas de tratamento descentralizados possuem o uso viabilizado em áreas de menor densidade populacional e áreas rurais, segundo Dos Santos (2015), visto que a implantação, operação e manutenção de sistemas centralizadas possuem um alto custo e, para poucos usuários, acarreta em um aumento das taxas de serviço, bem como inviabilidade financeira de implantação.

Nesse sentido, há inúmeras tecnologias de sistemas descentralizados de tratamento de efluentes domésticos, tais como: fossas sépticas (decanto-digestores) e filtros anaeróbios, bacias de evapotranspiração, biodigestores e dentre essas tecnologias vem se destacando a utilização de *wetlands* construídos (WC).

Os *qwaw* construídos utilizam o conceito de biomimetismo ao replicar e otimizar processos de transformação de matéria orgânica e ciclagem de nutrientes que ocorrem naturalmente em ambientes alagados, como pântanos, banhados e mangues. Na biomimética há a validação da natureza como detentora de tecnologia intrínseca, sendo tal tecnologia desenvolvida e otimizada em processos evolutivos de milhões de anos (PAWLIN, 2011).

Os WCs incorporam os processos naturais que envolvem a interação entre vegetação, o solo e suas populações microbianas em sistemas de engenharia para tratamento de águas residuárias (SEZERINO, 2018).

Em sua pesquisa Schroeder (2019) destaca que devido a sua simplicidade operacional, resistência as variações de cargas hidráulicas e orgânicas, baixo custo quando comparado a outros sistemas e elevadas eficiências de tratamento, tem-se tornado cada vez mais presente, na realidade brasileira, o tratamento de águas residuárias.

Esses sistemas, de interações entre macrófitas e atividade microbiana favorecem a remoção em termos de matéria carbonácea como DQO e DBO, sólidos em suspensão, frações nitrogenadas e patógenos (PLATZER, 1999). O entendimento profundo do comportamento desses sistemas é um dos pilares para a otimização do seu funcionamento.

Diferentes abordagens tentam viabilizar a previsão das concentrações de saída dos *wetlands*, devido a um grande número de processos físicos, químicos e biológicos que ocorrem paralelamente e se influenciam mutualmente.

O *software* “*Hydrus Wetland Model*” desenvolvido por Langergraber (2009) na Universidade de Recursos Naturais e Ciências da Vida de Viena-Áustria, utiliza de modelagem conceitual, matemática e numérica, bem como calibração do modelo com dados existentes para implementar uma série de equações diferenciais a fim de detalhar os processos existentes nos

sistemas de *wetlands* e prever as concentrações de saída. Contudo, há abordagens mais simplificadas, denominadas *black-box*, que utilizam correlações entre concentrações de afluente e efluente para previsão numérica das concentrações de saída do sistema.

As Redes Neurais Artificiais (RNA) são modelos de redes neurais que utilizam uma abordagem *data-driven* dos dados, e não uma abordagem conceitual para construção de modelos de previsão, já que abordagens *data-driven* não requerem detalhes fundamentais e conhecimento fenomenológicos dos processos (Yang B; Xiao Z et al, 2020). Conforme Granata et al 2020 apud Bowen et al 2022 abordagens *data-driven* possuem amplas aplicações e fornecem resultados preditivos que podem auxiliar na previsão de qualidade da água para *wetlands*, desde que estejam disponíveis dados experimentais.

Segundo Hosseinzadeh et al (2018), RNA são capazes de realizar a predição de eventos devido a capacidade de se adaptarem as relações lineares e não-lineares entre as variáveis, paralelismo, multidisciplinaridade e tolerância a falhas, além de possuírem processos computacionais não tão custosos quando abordagens conceituais e modelagens numérico-matemáticas (Coelho; Mariani, 2006). Por exemplo, Yang B et al (2022) utilizaram redes neurais para realizar a predição da qualidade do efluente dos *wetlands* utilizando indicadores de qualidade da água e meteorológicos obtendo então resultados satisfatórios, com coeficientes de regressões maiores que 0.8 em seus modelos. Akratos et al (2007) utilizaram de redes neurais para prever a remoção de DBO e DQO em *wetlands* construídos horizontais de fundo saturado (WCH-FS) e a partir das redes neurais modelou equações hiperbólicas de zero e primeira ordem para predição de remoção de DQO e DBO.

Song et al (2013) aplicaram RNA e regressão linear múltipla para simular o efeito das taxas de desnitrificação em WC, e concluíram que RNA alcançaram resultados superiores aos algoritmos de múltipla regressão linear.

Dentro desse contexto, este trabalho tem como foco desenvolver uma rede neural artificial com a série histórica de dados de qualidade d'água, de junho de 2015 a setembro de 2020, de um *wetland* construído vertical de fundo saturado – (WCV-FS), pertencente ao Grupo de Estudos de Saneamento Descentralizado (GESAD), e a série histórica de dados meteorológicos da estação meteorológica da EPAGRI em Florianópolis, para realizar a predição da concentração de DQO na saída do WCV-FS.

2 . OBJETIVOS

Este trabalho visa analisar a aplicabilidade de predição de redes neurais artificiais sob o desempenho de tratamento de esgoto promovido por unidade de *wetland* construído vertical de fundo saturado.

3 . OBJETIVOS ESPECÍFICOS

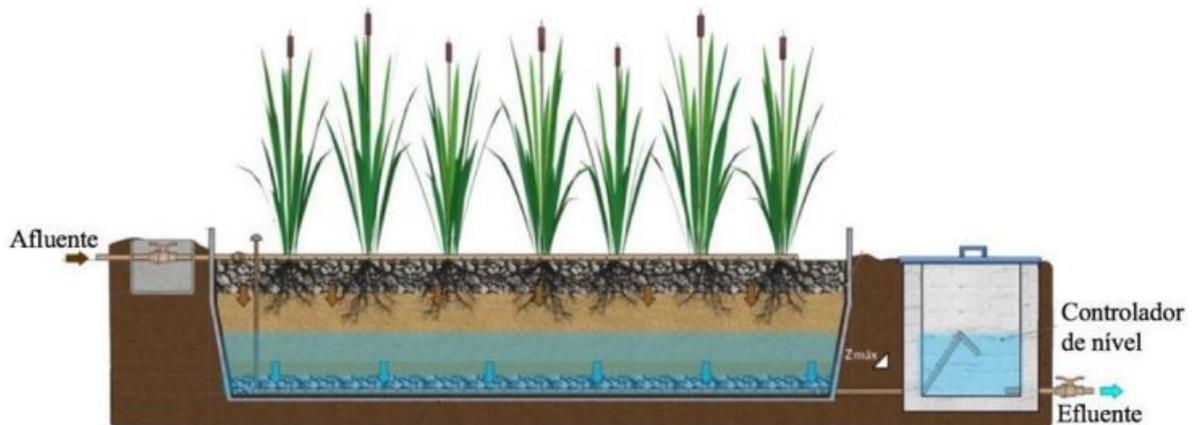
- Desenvolver uma rede neural artificial para predição da remoção de DQO em um *wetland* construído vertical com fundo saturado.
- Avaliar diferentes parâmetros que formam conjuntos de dados de entrada na eficiência do modelo.
- Realizar uma análise de principal componente (PCA) para redução da dimensionalidade do conjunto de dados e seleção de parâmetros.
- Produzir um mapa conceitual das etapas de desenvolvimento de uma rede neural artificial e disponibilizar os códigos abertos e estruturados.

4 REVISÃO BIBLIOGRÁFICA

4.1 WETLAND CONSTRUÍDO VERTICAL DE FUNDO SATURADO

A unidade de WCV-FS é uma ecotecnologia empregada para promover o tratamento secundário e avançado em efluentes biodegradáveis pós unidades de tratamento primário, sendo basicamente uma adaptação do *wetland* construído vertical (WCV). Esta adaptação refere-se, basicamente, a elevação da tubulação de saída do efluente, permitindo assim que haja uma zona de saturação dentro do maciço filtrante (Figura 1). A partir da saturação de uma parcela do maciço há o favorecimento de condições anóxicas e anaeróbias onde permite que ocorra processos de desnitrificação (BASSANI, 2022). Segundo Da Silva (2021), o funcionamento ideal depende de parâmetros tais como profundidade do fundo saturado, tempo de detenção hidráulica, pH, temperatura, carga orgânica aplicada e taxa hidráulica aplicada. Pelissari et al (2018) apud Bassani (2022) apontam que dado o fato dos WCV-FS apresentarem maiores tempos de detenção hidráulica (TDH) e maior umidade média no maciço, há um maior desenvolvimento do sistema radicular quando comparado ao WCV.

Figura 1 - Representação longitudinal de um WCV-FS



Fonte: Acervo GESAD.

4.2 TRANSFORMAÇÃO E REMOÇÃO DE MATÉRIA ORGÂNICA CARBONÁCEA EM WCV-FS

Efluentes domésticos apresentam frações maiores de matéria orgânica que quando em mistura com corpos hídricos levam a depleção de oxigênio dissolvido, impactando na qualidade das águas e nos serviços ecossistêmicos (LEITE, 2004). Segundo von Sperling (2005) está contido nos esgotos 40 a 60% de proteínas, 25 a 50% de carboidratos e 8 a 12% de óleos, além de ureia, fenóis, pesticidas e outros poluentes. O monitoramento da matéria orgânica é medido por métodos indiretos como a Demanda Bioquímica de Oxigênio (DBO) e Demanda Química de Oxigênio (DQO). Para von Sperling (2000) a matéria orgânica biodegradável pode ser dividida em particulada ou dissolvida, a fração particulada tem degradação mais lenta e ocorre através de processos de hidrólise após processos físicos de filtração e sedimentação. Já a fração dissolvida da matéria orgânica pode ser transformada a partir de processos aeróbios e anaeróbios/anóxicos, podendo ocorrer adsorção e biodegradação no biofilme.

A degradação aeróbia de carbono se dá principalmente por microrganismos quimioautotróficos e quimiheterotróficos, estas bactérias oxidam a matéria orgânica e utilizam oxigênio como acceptor final do elétron (Garcia et al, 2010) liberando dióxido de carbono, amônia e outros compostos químicos estáveis. Segundo Saeed (2012) a intensidade dessas rotas metabólicas é altamente dependente do fornecimento de matéria orgânica e oxigênio.

O desempenho de remoção da matéria orgânica de efluentes submetidos ao tratamento em WCV-FS é dependente de inúmeros fatores, tais como (Sezerino et al, 2018):

- carga orgânica aplicada;
- taxa hidráulica aplicada;
- regime de aplicação intermitente em pulsos;
- alternância de uso com unidade em operação e unidade em repouso;
- manejo das macrófitas;
- condições climáticas locais.

4.3 ANÁLISE DE DADOS APLICÁVEL À PREDIÇÃO E CÁLCULO DE DESEMPENHO DE TRATAMENTO EM WCV-FS

4.3.1 Análise de Componentes Principais (ACP)

Análise de componente principal é técnica estatística multivariada utilizada para redução de dimensionalidade do conjunto de dados, na qual permite que variáveis altamente correlacionáveis sejam reduzidas a um pequeno número de componentes ortogonais não-correlacionados que contenham a maior parte da informação do conjunto original. (Akratos et al 2008).

Segundo Hongyu, Martins, Oliveira (2015) cada componente principal é uma combinação linear de todas as variáveis originais, são independentes entre si e contém o máximo de informação da variância dos dados. A ACP tem por objetivo explicar a estrutura de variância e covariância de um vetor aleatório, composto por n variáveis aleatórias gerado por meio de combinações lineares das variáveis originais.

Há desvantagens na ACP perante a influência de *outliers*¹ e quando as matrizes possuem dupla-ausência (muitos valores zero). Além disso, ao reduzir o número de variáveis há uma perda na quantidade de informações. A parte explicada torna-se padrão de resposta e a outra parte torna-se o ruído, ou seja, erro de medida e redundância. Hongyu; Martins; Oliveira (2015)

É possível avaliar a importância de um componente dado a porcentagem de variância que ele retém. Dessa forma, o primeiro componente principal armazena mais variância que o segundo componente e assim sucessivamente. A independência ou ortogonalidade (ausência de correlação) é uma propriedade útil do ACP, significando que tais componentes medem diferentes “dimensões” dos dados (MANLY, 1984 apud BARBOSA, 2003).

A variância antes dispersa entre as variáveis correlacionáveis é agora atribuída a um número menor de componentes, ocorrendo assim uma simplificação considerável referente aos cálculos estatísticos e interpretação dos resultados, tal fato corrobora melhor quando o número de indivíduos avaliados for grande (CRUZ, 1990).

¹ *Outlier*: ou valor atípico, é uma observação que apresenta um grande afastamento das demais observações do conjunto amostral.

4.3.2 Redes Neurais Artificiais

A inteligência artificial (IA) pode ser entendida como uma área da ciência da computação que visa o projeto de sistemas inteligentes, ou seja, sistemas que buscam copiar algum tipo de inteligência humana, em termos de processos computacionais (HAYKIN, 2001). Segundo Nilsson (1998) os sistemas inteligentes exibem características que são associadas ao comportamento inteligente como percepção, aprendizagem e raciocínio. Há áreas de estudo dentro da IA como aprendizado de máquina que através de técnicas estatísticas visam desenvolvimento de programas que realizam tarefas descritivas ou preditivas baseado no conjunto de dados. Por fim, aprendizado profundo (*deep learning*) é uma sub-área de aprendizado de máquina que permite a resolução de problemas por aprendizado repetitivos e organizado em sistemas de camadas denominados de Redes Neurais Artificiais (RNA). Segundo Haykin (2001) RNAs podem ser entendidas como sistemas computacionais estruturado dada sua capacidade de aprendizagem de padrões existentes em um conjunto de dados, além de capacidade de generalização.

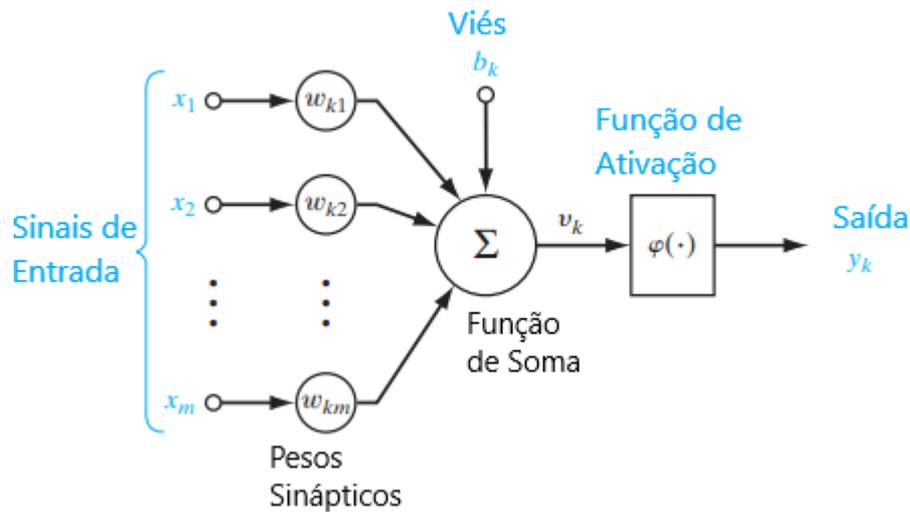
Dessa forma as redes neurais possuem características que melhor explicam seu potencial devido à adaptabilidade, paralelismo, multidisciplinariedade e tolerância a falhas, que segundo Coelho; Mariani (2006), gera uma ferramenta matemática potencial para aplicações práticas e complexas.

Redes Neurais Artificiais são bastante utilizadas em muitas áreas da ciência e engenharia, Naykak (2006) comenta as principais vantagens de utilizá-las é a não necessidade do entendimento da complexidade da natureza do processo, utilizando uma abordagem de “*black-box*” e também aprender relações lineares e não-lineares entre as variáveis. Soares (2015) cita que o poder de generalização das RNAs está na capacidade de aprender através de um pequeno conjunto de dados e fornecer respostas coerentes para resultados não conhecidos.

As redes neurais são formadas assim por neurônios e conexões entre eles. Uma das principais partes de RNA são seus processadores de informações chamados de neurônios, dessa forma, as RNAs são constituídas de neurônios interconectados que trabalham juntos para resolver um problema.

Para Haykin (2006) um neurônio é constituído de três elementos principais, são estes: peso sináptico, função de soma e função de transferência. A ilustração de um neurônio está definida na Figura 2.

Figura 2 Modelo de um neurônio.



Fonte: Haykin (2001), adaptado pelo autor.

Os pesos sinápticos são caracterizados por um fator multiplicativo. Um sinal x_m na entrada da sinapse m que está conectada a um neurônio k , e é multiplicada pelo peso sináptico w_{kj} . Onde k da entrada w_{kj} representa o neurônio em questão e j refere-se a entrada do sinal sináptico para qual o peso sináptico se refere.

A função de soma, adiciona os sinais sinápticos multiplicados pelos respectivos pesos sinápticos, a operação citada constitui uma combinação linear.

Já a função de ativação limita a amplitude de saída de um neurônio, ou seja, limita a possibilidade de saída do sinal para algum valor finito.

Além disso, um típico modelo de neurônio inclui um viés (b_k) que possui o efeito de aumentar ou diminuir a entrada na função de ativação.

Em termos matemáticos, um modelo de neurônio (k) pode ser descrito pela equação 1 e equação 2:

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

$$y_k = \varphi(u_k + b_k) \quad (2)$$

Onde:

x_1, x_2, \dots, x_m : representam os valores de entrada.

$w_{k1}, w_{k2}, \dots, w_{km}$: representam os pesos sinápticos de neurônio k .

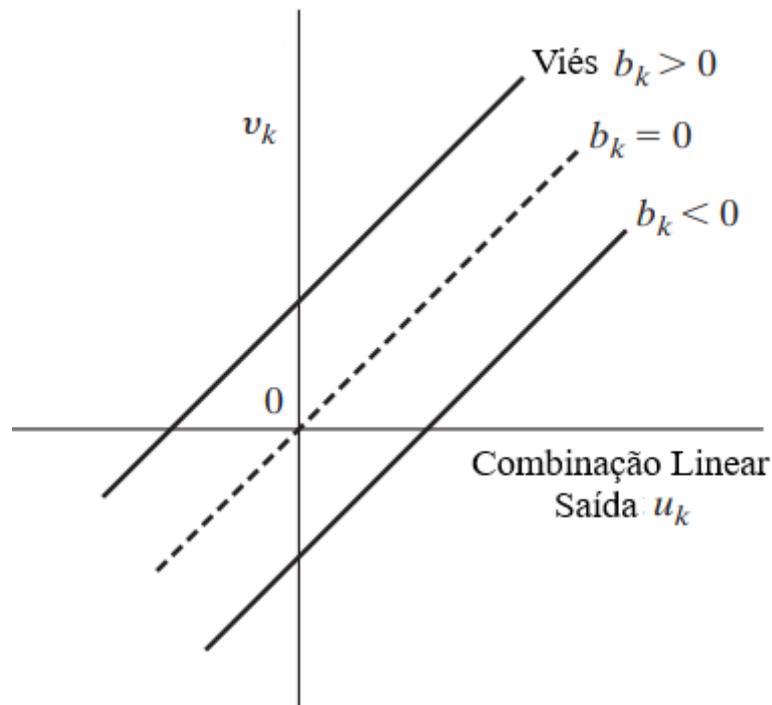
b_k : viés.

φ : função de ativação.

y_k : é sinal de saída do neurônio.

A parcela y_k correspondente da Figura 3, u_k é a saída da combinação linear e o termo de viés (b_k) (*Idem*, 2006) pode ser compreendido como um parâmetro externo do neurônio k , cujo efeito é de deslocar as combinações lineares referentes as parcelas de u_k , ilustrado na Figura 3.

Figura 3 Representação do efeito do viés no termo u_k



Fonte: Haykin (2001), adaptado pelo autor.

4.3.2.1 Função de Ativação

A função de ativação denotada por φ , define os limites de saída do neurônio para algum valor finito, geralmente no intervalo normalizado $[0,1]$ ou $[-1,1]$. Para Nogueira (2018) a camada de ativação adicionam robustez para as redes neurais gerando não-linearidades a partir das funções de ativação. As principais funções de ativação são demonstradas pelas equações a seguir:

- Função Sigmóide, descrita pela (3)

$$\varphi(v) = \frac{1}{1+\exp(-av)} \quad (3)$$

- Unidade Linear Retificada (ReLU)

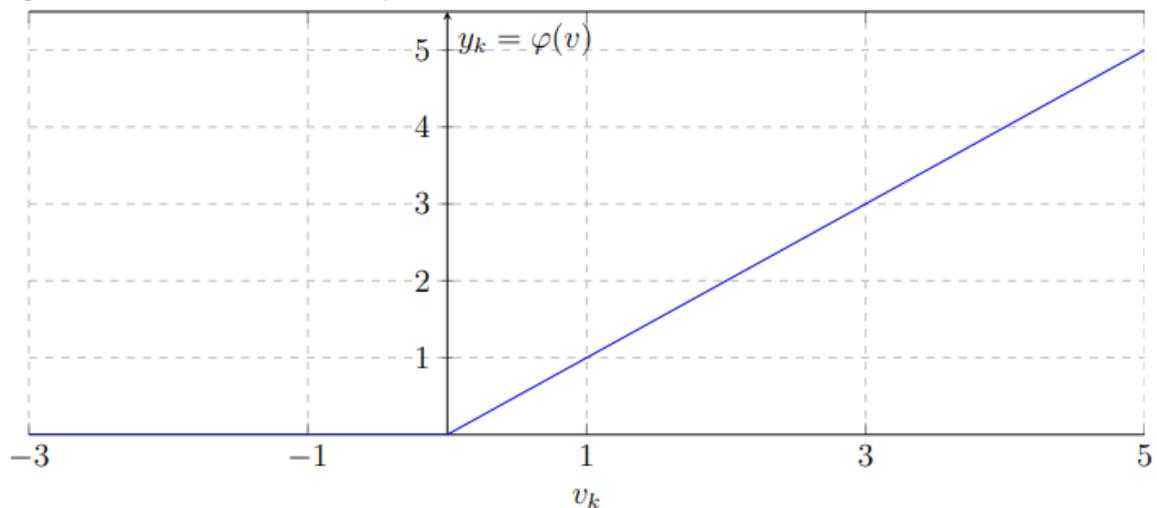
A função de Unidade Linear Retificada (ReLU - *Rectified Linear Unit*) apresenta aprendizado extremamente rápido para as redes de multi-camadas e processamento rápido, sendo bastante utilizada dentro do estado da arte de redes neurais.

$$\varphi = \text{ReLU}(x) = \max(0, x) \quad (4)$$

$$\text{Se } y_k = \begin{cases} x & \text{se } v_k > 0 \\ 0 & \text{se } v_k < 0 \end{cases}$$

O gráfico de uma função ReLU é ilustrado na Figura 4

Figura 4 - Gráfico de uma função ReLU



Fonte: Nogueira 2018 adaptado pelo Autor.

Bengio; Goodfellow (2016) citam que devido a função ReLU ser quase linear, preserva muitas propriedades que facilitam a otimização com métodos de gradiente descendente, e muitas outras propriedades as quais permitem RNAs generalizarem bem.

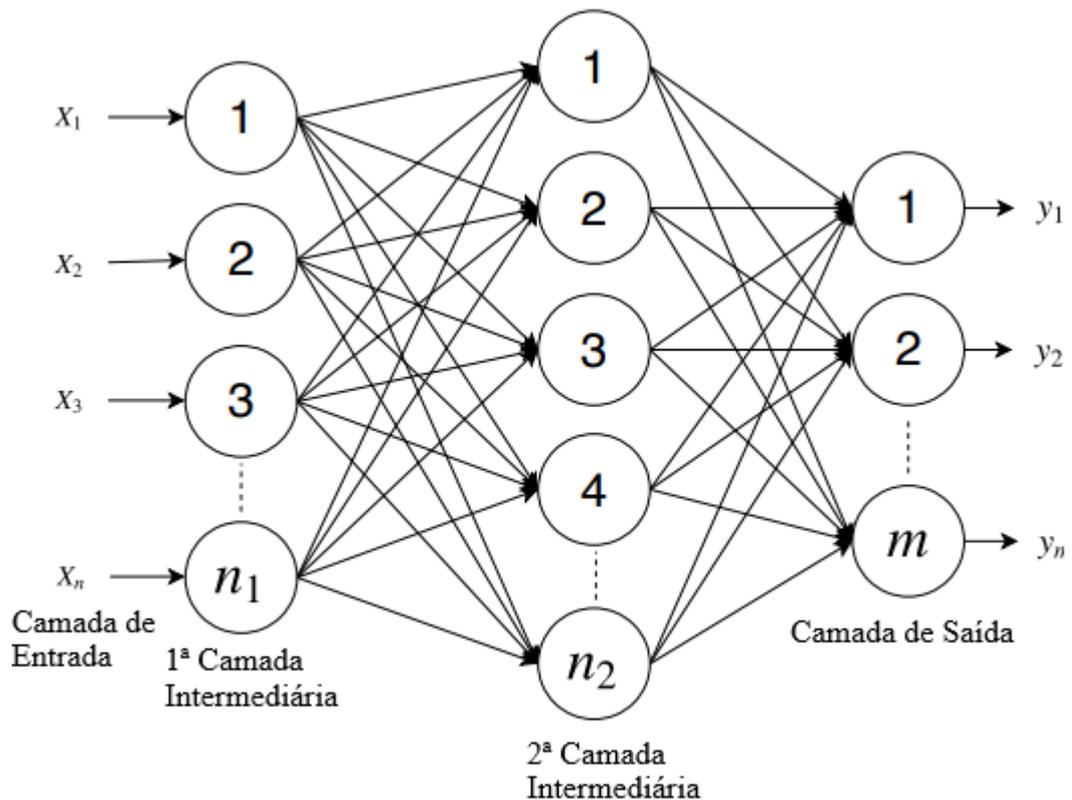
4.3.2.2 Arquitetura de Redes Neurais

Diferentes combinações de neurônios formam as arquiteturas das RNAs, podendo estes estarem dispostos de maneira que formem uma ou mais camadas.

As camadas são denominadas de: Camadas de Entrada (*Input*); Camadas Intermediárias (*Hidden Layers*); e Camadas de Saída (*Output*). Quando há pelo menos uma camada intermediária são chamadas de Redes Neurais Artificiais de Multicamadas.

Um exemplo das RNAs pode ser observado na Figura 5, onde cada círculo representa um neurônio artificial e pode ser interpretado como encapsulamento da Figura 2.

Figura 5 - Modelo de uma Rede Neural Artificial de Multicamadas



Fonte: Adaptado pelo autor de Manerichi (2014).

As RNAs propagam as informações entre suas camadas pelo método de *feedforward* ou seja, a informação disposta na camada de entrada, passa pelo neurônios das camadas intermediárias até a camada de saída, responsável por apresentar os resultados (MELLO, 2004).

4.3.2.3 Processo de Aprendizagem

O processo de treinamento das redes neurais, necessita do uso de algoritmo de aprendizagem, sendo o método mais utilizado o *Backpropagation*, ou algoritmo de propagação

de erro. O processo tem por objetivo definir os pesos sinápticos e viés de todos os neurônios inseridos na rede por meio de um processo recursivo (MORAIS, 2011).

O processo inicia-se com um vetor de entradas x_i à rede, que percorre os neurônios nas camadas intermediárias até chegar à camada de saída, realizando o processo de *feedforward*.

A seguir é feito o processo de *feedback* onde o valor de saída da rede é comparado a um vetor desejado d_i de respostas apresentado para a rede. Nesse passo ocorre a correção dos pesos sinápticos e vieses, onde o valor do erro é ajustado em cada neurônio, sendo um processo iniciado da camada de saída até a camada de entrada.

É necessário a utilização de uma função de custo para validar o erro total da rede chegar a um intervalo ideal desejado. A função mais comum para descrição do erro é o Erro Quadrático Médio, definido pela seguinte equação:

$$EQM(k) = \frac{1}{2} \sum_{j=1}^n (d_j(k) - Y_j(k))^2 \quad (5)$$

Onde:

EQM(k) representa o erro da k-ésima amostra de treinamento, $Y_j(k)$ é a saída do j-ésimo neurônio da camada de saída da k-ésima amostra e d_j é o respectivo valor desejado da k-ésima amostra.

De acordo com Lugwig Jr. e Montgomery (2007) os erros são elevados ao quadrado para evitar que erros negativos oriundos de um neurônio compensem erros positivos de outros.

O ajuste dos pesos é feito baseado na retropropagação do erro, assim o erro gerado pelos neurônios da camada de saída é distribuído para os demais neurônios da rede. Dado o erro global da rede, não é possível definir os pesos sinápticos exatos para cada neurônio para poder corrigi-los, porém, estabelece a direção para quais os pesos devem ser ajustados a fim de minimizar o erro quadrático total da saída da rede.

O algoritmo aplica o gradiente ao erro em relação à uma matriz W_{ji} . Utilizando-se da definição de gradiente e da regra da cadeia, tem-se a equação 6.

$$\nabla E = \frac{\partial E}{\partial W_{ji}^{(m)}} = \frac{\partial E}{\partial Y_j^{(m)}} \cdot \frac{\partial Y_j^{(m)}}{\partial I_j^{(m)}} \cdot \frac{\partial I_j^{(m)}}{\partial W_{ji}^{(m)}} \quad (6)$$

Onde o índice m representa a última camada de processamento, ou seja, a camada de saída. O peso sináptico $W_{ij}^{(m)}$ é tratado como uma matriz que comporta os pesos de todos os neurônios na camada de saída, onde os índices representam o j -ésimo neurônio da camada calculada em relação ao i -ésimo neurônio da camada imediatamente anterior.

I_j^m representa um vetor onde os valores de entradas do j -ésimo neurônio da camada m ponderado pelo peso sináptico correspondente. $Y_i^{(m)}$ é a saída do i -ésimo neurônio da camada m .

Matematicamente I_j^m e $Y_i^{(m)}$ podem ser representados pelas equações 7 e 8.

$$I_j^{(m)} = \sum_{i=0}^{n_{m-1}} W_{ij}^{(m)} \cdot Y_i^{m-1} \quad (7)$$

$$Y_i^m = g(I_j^m) \quad (8)$$

Onde n_{m-1} diz respeito ao neurônio da camada anterior a m . A função $g(\cdot)$ representa a função de ativação aplicada.

A partir dos termos da equações 9, 10 e 11 é possível obter a equação 12.

$$\frac{\partial I_j^m}{\partial W_{ji}} = Y_i^{m-1} \quad (9)$$

$$\frac{\partial Y_j^m}{\partial I_j^m} = g'(I_j^m) \quad (10)$$

$$\frac{\partial E}{\partial Y_j^m} = -(d_j - Y_j^m) \quad (11)$$

Onde $g'(\cdot)$ representa a derivada primeira da função de ativação e a equação 11 representa o erro na última camada. Por fim, a equação 12 diz respeito a representação do erro em relação aos pesos sinápticos.

$$\frac{\partial E}{\partial W_{ji}^m} = -(d_j - Y_j^m) \cdot g'(I_j^m) \cdot Y_i^{m-1} \quad (12)$$

Logo o ajuste da matriz pode ser definido pela equação 13, equação 14 e equação 15 e deve ser realizado na direção oposta ao gradiente, pois pretende-se reduzir a equação do erro:

$$\Delta W_{ji}^m = -\eta \cdot \frac{\partial E}{\partial W_{ji}^m} \quad (13)$$

$$\delta_j^m = (d_j - Y_j^m) \cdot g'(I_j^m) \quad (14)$$

$$\Delta W_{ji}^m = \eta \cdot \delta_j^m \cdot Y_i^{m-1} \quad (15)$$

Onde η é a taxa de aprendizagem do próprio algoritmo, Y_i^{m-1} é a saída do i -ésimo neurônio da camada anterior e δ_j^m é o gradiente local em relação ao j -ésimo neurônio da camada.

A equação pode ser descrita de forma iterativa como demonstra a equação 16.

$$W_{ji}^m(k+1) = W_{ji}^m(k) + \eta \cdot \delta_j^m \cdot Y_i^{m-1} \quad (16)$$

Por fim, enquanto o EQM da equação 5 não estiver contido no intervalo desejado, o algoritmo recalculará recursivamente os pesos sinápticos, e essa recursividade dá-se o nome de épocas (*epochs*).

O ajuste dos pesos sinápticos nas camadas intermediárias ocorre de forma parecida com o ajuste da camada de saída. A diferença é que os neurônios da camadas intermediárias não possuem um valor desejado de saída, por isso utilizam de estimativas de erros de saída dos neurônios da camada imediatamente posterior, que já foram ajustados na etapa.

A equação 17 demonstra isso:

$$\nabla E^{m-1} = \frac{\partial E}{\partial W_{ji}^{(m-1)}} = \frac{\partial E}{\partial Y_j^{(m-1)}} \cdot \frac{\partial Y_j^{(m-1)}}{\partial I_j^{(m-1)}} \cdot \frac{\partial I_j^{(m-1)}}{\partial W_{ji}^{(m-1)}} \quad (17)$$

Onde aplicando as mesmas substituições realizadas para as equações 5 a 15 encontra-se a forma iterativa descrita na equação 18.

$$W_{ji}^{(m-1)}(k+1) = W_{ji}^{(m-1)}(k) + \eta \cdot \delta_j^{(m-1)} \cdot Y_i^{(m-2)} \quad (18)$$

Esse procedimento possibilita o cálculo do peso das camadas ocultas, sempre aplicando da última camada para a primeira, ocorrendo o processo de forma iterativa até que se encontre um erro mínimo ou estabeleça-se um método de parada que avalia a variação desse EQM entre cada época.

Os pesos conectados as camadas de saída são ajustados para reduzir esse erro. Em seguida, o erro da camada de saída é utilizado para derivar estimativas de erro para as unidades da camadas ocultas, para que o erro seja então retro propagado até a conexão da camada de entrada.

Um dos possíveis problemas que podem ocorrer durante o processo de treinamento da rede neural é *overffiting*. Isso ocorre quando após um determinado ciclo de treinamento, uma época, a rede tende a diminuir sua taxa de acertos para padrões diferentes daqueles utilizados nos ajustes dos pesos. Segundo Morais (2011) considera-se que o *overffiting* é quando a rede memorizou os padrões de treinamento, suas peculiaridades e ruídos, ou seja, ela se adaptou somente a base de dados que foi utilizada para treiná-la e não possui capacidade de generalizar bem.

4.3.3 Validação do Modelo

4.3.3.1 Erro Quadrático Médio (EQM)

Para validação do modelo de redes neurais um dos métodos utilizados é o Erro Quadrático Médio, descrito em 4.3.2 Redes Neurais Artificiais, dado pela equação 5.

Métodos supervisionados, diferentes de métodos não supervisionados, avaliam os conjuntos de saída das RNAs e comparam com conjuntos de valores desejados, dessa forma por se tratar de uma rede de aprendizagem supervisionado, cujo objetivo é ajustar os parâmetros de pesos sinápticos e viés da rede a fim de minimizar o EQM, sendo esse um dos principais parâmetros de avaliação (MORAIS, 2011).

4.3.3.2 Erro Quadrático Médio Normalizado (EQMN)

O erro quadrático médio normalizado é uma variação do erro quadrático médio. O EQMN é utilizado para avaliar como os modelos predizem valores distante da média, e é dado pela equação 19.

$$EQMN = \sqrt{\frac{\sum_{i=1}^n (M_i - P_i)^2}{\frac{n}{\bar{M}}}} \quad (19)$$

Onde \bar{M} é a média dos valores de validação, ou dos valores medidos, M_i é o i -ésimo valor medido, P_i é o i -ésimo valor predito. Para o EQMN quanto mais próximo do valor zero melhor é o modelo, como é uma unidade de medida normalizada pode ser utilizada para comparar diferentes unidades nos *wetlands*.

4.3.3.3 Coeficiente de Determinação Múltipla R^2

O coeficiente é definido como:

$$R^2 = \frac{SQ_R}{SQ_T} = 1 - \frac{SQ_E}{SQ_T} \quad (20)$$

SQ_R é a soma quadrática da regressão, descrita por:

$$SQ_R = \sum_{i=1}^n (\hat{y} - \bar{y})^2 \quad (21)$$

sendo \hat{y} o valor previsto, e \bar{y} a média dos valores previstos n é o número de observações.

SQ_T é a soma quadrática total corrigida e apresenta a variância presente nos dados, dada por:

$$SQ_T = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (22)$$

sendo y_i o valor real e \bar{y} a média dos valores reais n é o número de observações.

SQ_E é a soma dos quadrados do erro, ou soma dos quadrados dos resíduos, e apresenta o desvio entre os valores reais e os preditos pelo modelo.

$$SQ_E = \sum_{i=1}^n (y_i - \hat{y})^2 \quad (23)$$

E por fim, n , é o número de amostras.

Este coeficiente mede o grau de correlação e a direção dessa correlação entre duas variáveis, baseado na qualidade de um ajuste linear dos dados. Este coeficiente assume valores entre 0 e 1, em que altos valores de R^2 representam boa correlação entre as variáveis de entrada e de saída

4.3.3.4 Coeficiente Ajustado de Determinação Múltipla

O coeficiente é dado pela equação 24.

$$R_{ajustado}^2 = 1 - \frac{SQE/(n-p)}{SQ_T/(n-1)} \quad (24)$$

Onde p representa o número de parâmetros a serem ajustados, no caso de RNA, os pesos sinápticos e viés.

Diferente do R^2 o $R_{ajustado}^2$ nem sempre aumenta quando uma variável é adicionado ao modelo, o $R_{ajustado}^2$ só aumenta se a adição da variável produzir uma redução na soma quadrática residual, que seja grande suficiente para compensar a perda de um grau de liberdade no resíduo

4.3.4 Testes Estatísticos de Distribuição Normal

4.3.4.1 Shapiro-Wilk

O teste de Shapiro-Wilk testa a hipótese nula de que uma amostra x_1, x_2, \dots, x_n veio de uma população normalmente distribuída (VON SPERLING et al, 2020). A estatística do teste é descrita pela equação 25.

$$W = \frac{\sum_{i=1}^n a_i x_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (25)$$

Onde:

\bar{x} : é a média da amostra

a_i : coeficientes dados pela equação 26.

$$(a_1, \dots, a_n) = \frac{m^T V^{-1}}{C} \quad (26)$$

Onde C é a norma de um vetor

4.3.4.2 Anderson-Darling

O teste Anderson-Darling é utilizado para comprovar se uma amostra de dados possui sua população com uma distribuição específica (MARSAGLIA, 2004). O teste de Anderson-Darling utiliza de distribuições específicas para calcular valores críticos. A hipótese do teste Anderson-Darling é que os dados vem de uma distribuição específica e caso o valor do teste seja maior que o valor crítico a hipótese é rejeitada, para dado nível de significância.

O teste de Anderson-Darling é dado pela equação 27:

$$A^2 = -N - S \quad (27)$$

Onde

A^2 : é o valor do teste estatístico

N : é o tamanho do conjunto

S : é descrito por:

$$S = \sum_{i=1}^N \frac{2i-1}{N} [\ln F(1 - F(Y_{N+1-i}))] \quad (28)$$

Sendo F a função de distribuição cumulativa da distribuição especificada.

5. METODOLOGIA

O presente trabalho utilizou dados qualitativos de efluentes provenientes do Grupo de Estudos em Saneamento Descentralizado (GESAD) constituindo um conjunto de dados com um período amostral do dia 25/06/2015 ao dia 15/07/2020, e dados meteorológicos, os quais foram cedidos pela EPAGRI, através de solicitação indexada no anexo A, com um período amostral idêntico ao dados de qualidade.

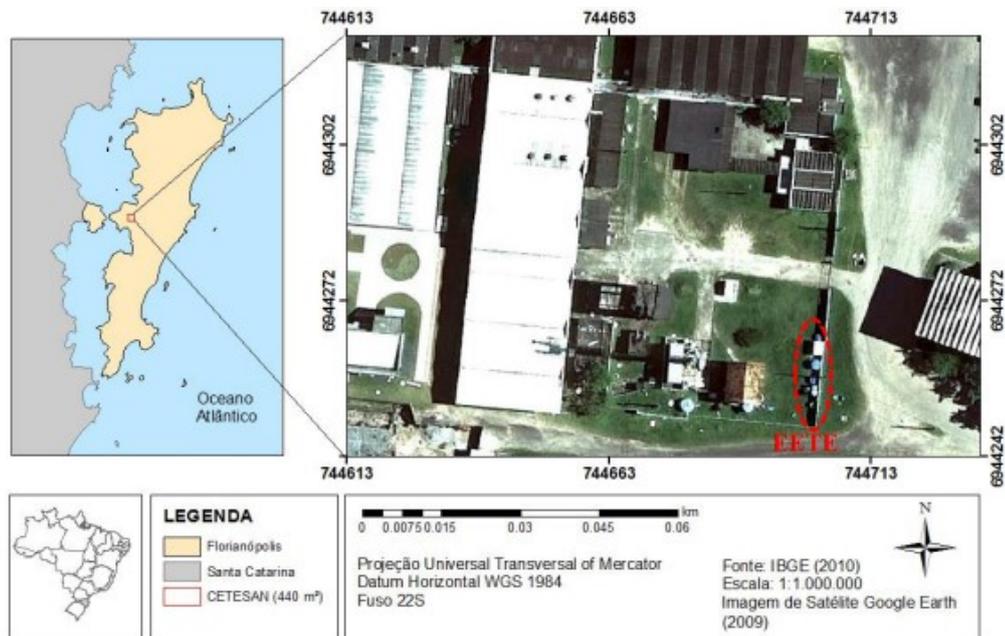
O mapa conceitual, apresentado no Apêndice A descreve os processos e etapas para consolidação do presente trabalho. Nele é feito a análise de inconsistência dos conjuntos de dados e preenchimento de valores nulos, unificados em um único conjunto de dados. Na ilustração, vista no mapa, é verificada a normalidade para cada parâmetro do conjunto, aplicando transformações lineares e autenticadas com testes estatísticos de assunção de normalidade. Esse conjunto de dados, transformados, passa pela etapa de análise de componentes principais e de influência através do algoritmo *XGBoost*. Gerando, então,

conjuntos de dados com diferentes parâmetros que compõem as Redes Neurais Artificiais é realizado a busca da melhor topologia da RNA e validado os resultados através do EQM e Coeficientes de Regressão.

5.1 DADOS DE QUALIDADE DOS EFLUENTES

Os dados foram provenientes da Estação Experimental de Esgoto (EETE) localizada dentro da Central Educacional de Tratamento de Esgoto Sanitário, nas coordenadas $27^{\circ}36'12.64''S$ de latitude e $48^{\circ}31'14.43''O$ de longitude. A Figura 6 apresenta o mapa de localização da estação.

Figura 6 - Mapa de Localização da EETE – GESAD/UFSC.



Fonte: Freitas (2017).

A EETE é composta por um tanque séptico, um tanque equalizador e três WCs inseridos em dois arranjos: primeiro é constituído por um WCV seguido de um WCH compondo assim um arranjo híbrido, e o segundo é constituído de um WCV-FS. O esgoto bruto é recalado da rede coletora de esgoto da concessionária do município de Florianópolis para realizar um tratamento primário no tanque séptico e em seguida, do tanque equalizador é bombeada para os arranjos tecnológicos. A Figura 7 ilustra o os arranjos dos WCs na EETE.

Figura 7 - Arranjo dos *Wetlands* Construídos componentes da EETE – GESAD/UFSC.



Fonte: Arquivo GESAD.

O módulo de WCV-FS apresenta, como demonstrado na Tabela 1, as seguintes características construtivas:

Tabela 1 - Características Construtivas do WCV-FS.

Dimensões	
Área Superficial	7,5 m ²
Altura Construtiva	0,83 m
Maciço Filtrante	
Altura Total Filtrante	0,75 m
Brita no topo	0,03 m
Areia Grossa (d ₁₀ = 0,29mm; d ₆₀ = 1,16mm)	0,62 m
Brita de Fundo	0,10 m
Saturação de Fundo	
Altura de Saturação	0,40 m
Vegetação	
Macrófitas	<i>Typha domingensis</i> (Taboa)

A altura de saturação do WCV-FS foi definida por Santos (2015) a partir de estudos de mestrado, sendo essa altura regularizada a partir de um controlador de nível de saída, como observado na Figura 8.

Figura 8 - Controlador de nível de saída do WCV-FS componente da EETE – GESAD/UFSC.



Fonte: Arquivo GESAD.

5.1.1 Característica Operacional do WCV-FS

O WCV-FS foi dimensionado para receber a contribuição de uma residência unifamiliar de cinco pessoas (cerca de 600 L/d), onde a carga orgânica aplicada prevista é de 41 gDQO/m².d, seguindo as recomendações de dimensionamento propostas por Sezerino (2006). A Tabela 2 aborda as características operacionais dos WCV-FS

Tabela 2 - Características Operacionais do WCV-FS

Parâmetro	Unidade	Valor
Carga Orgânica Aplicada	gDQO/m ² .d	36,23
Carga Hidráulica Aplicada	mm/d	83
Alimentação Semanal	dias/semana	3,5

5.1.2 Descrição dos Dados de Qualidade

A fim do desenvolvimento dos estudos científicos no GESAD houve coletas do esgoto desde o ano de 2015 amostrados quinzenalmente durante um período e posteriormente as coletas viraram mensais, até o ano de 2020. As coletas eram realizadas na saída do Tanque Séptico e entrada do WCV-FS, sendo denominadas de afluentes e na saída do WCV-FS, denominadas de efluentes.

Os dados de qualidade de efluente e afluente possuem os parâmetros de monitoramento descritos na Tabela 3.

Tabela 3 – Parâmetros de Qualidade Analisados no afluente e efluente do WCV-FS componente da EETE – GESAD/UFSC

Parâmetro	Abreviação	Unidade
Sólidos Suspensos	SS	mg/L
Sólidos Totais	ST	mg/L
Sólidos Totais Voláteis	STV	mg/L
Sólidos Totais Fixos	STF	mg/L
pH	pH	Adimensional
Alcalinidade	-	mg/L
Demanda Química de Oxigênio	DQO	mg/L
Demanda Bioquímica de Oxigênio	DBO	mg/L
Nitrogênio Amoniacal	NH ₄ ⁺	mg/L
Nitrito	NO ₂ ⁻	mg/L
Nitrato	NO ₃ ⁻	mg/L
Fósforo	PO ₄ ³⁻	mg/L

O período de datas contido nos dados está relacionado entre o dia 25/06/2015 e 15/07/2020, totalizando 108 registros de parâmetros de qualidade para afluente e efluente.

5.1.3 Análise de Inconsistências e Preenchimento de Valores

Devido a possibilidade de haver falhas na série histórica foi realizado uma análise em busca de valores nulos e valores inconsistentes, com um total de 108 dias contendo dados, foi possível verificar as falhas na série, a Tabela 4 apresenta a percentagem de valores nulos para cada parâmetro de qualidade.

Tabela 4 - Percentagem de valores nulos para os parâmetros de qualidade avaliados junto a unidade WCV-FS componente da EETE – GESAD/UFSC.

Parâmetro	Percentagem de Valores Nulos
Sólidos Suspensos Afluente	2,88 %
Sólidos Totais Afluente	7,69 %
Sólidos Totais Voláteis Afluente	6,73 %
Sólidos Totais Fixos Afluente	6,73 %
pH Afluente	14,42 %
Alcalinidade Afluente	16,35%
DQO Afluente	0 %
DBO Afluente	8,65 %
Nitrogênio Total Afluente	41,35 %
Nitrogênio Amoniacal Afluente	14,35 %

Nitrito Afluyente	3,85 %
Nitrato Afluyente	4,81 %
Fósforo Afluyente	2,88 %
DQO Efluyente	0 %

Para o presente trabalho, foi estabelecido um limiar de ausência de dados de 25%, logo, caso exista uma falha na série maior que 25%, o parâmetro de qualidade é excluído por inteiro.

Dessa forma, o parâmetro Nitrogênio Total Afluyente, que possui 41,35% dos dados nulos não será considerada no presente trabalho.

As variáveis referentes a sólidos, principalmente os Sólidos Suspensos Afluyente, possuem dois registros com valores não congruentes para a série histórica, pois há valores menores que zero, os quais apontam para possível erro de análise laboratorial. Por isso esses valores incongruentes foram excluídos.

Sendo assim os demais parâmetros foram preenchidos com a média da série histórica de cada parâmetro, uma vez que os valores nulos (vazio) impedem o funcionamento da rede neural.

5.1.4 Adequação de Novos Parâmetros de Qualidade

No processo de bombeamento, devido a manutenções periódicas da bomba, há um monitoramento das vazões que estão sendo bombeadas para controle de qualidade das taxas hidráulicas bem como do carregamento orgânico superficial condizerem com os valores de projeto. Dessa forma, o GESAD possui valores do período de tempo que a bomba demora para encher um volume específico em uma determinada data, como exemplificado pela Tabela 5.

Tabela 5 - Trecho da tabela de Aferição da Vazão da Bomba dos WCV-FS

Data	Tempo (seg)	Volume (L)
17/05/2017	75.89	38
14/06/2017	75.00	38
29/06/2017	74.63	38

A tabela completa de aferições da vazão da bomba está disponível no Anexo A.

A partir desses valores de bombeamento, foi calculado mais três parâmetros de qualidade que compõem o conjunto de dados, a taxa hidráulica superficial que é dada pela equação 29 o carregamento orgânico superficial, representado pela equação 30.

$$TH_S = \frac{Q}{A} \quad (29)$$

Onde :

TH_S = Taxa Hidráulica Superficial ($m^3/m^2.d$)

Q = Vazão (m^3/d)

A = Área Superficial (m^2)

$$\lambda_s = \frac{DQO_{afluente} * Q}{A} \quad (30)$$

Onde:

λ_s = Carregamento Orgânico Superficial ($gDQO/m^2.d$)

$DQO_{afluente}$ = DQO afluente (mg/L)

Q = Vazão (m^3/d)

A = Área Superficial (m^2)

Com os valores de DQO e de DBO é possível calcular a relação entre DQO/DBO Afluente, sendo esse incorporado no conjunto de dados de qualidade. A Tabela 6 descreve os novos parâmetros incorporados.

Tabela 6 - Parâmetros de qualidade incorporados no conjunto de dados.

Parâmetro	Unidade
Carregamento Orgânico Superficial	$gDQO/m^2.d$
Taxa de Aplicação Hidráulica	$m^3/m^2.d$
Relação DQO/DBO	adimensional

5.2 DADOS METEOROLÓGICOS

Os dados meteorológicos foram disponibilizados pela EPAGRI/CIRAM via e-mail no dia 02/08/2022 enviado por Paulo Bezerra. No corpo do e-mail há a informação de que pode haver falhas devido a equipamento avariado, estação inoperante ou dado reprovado no controle de qualidade. Os dados são pertencentes a da estação meteorológica da EPAGRI, localizada no bairro Itacorubi em Florianópolis, nas coordenadas 27°34'53.04"S; 48°30'25.92"O.

O período de dados vai do dia 01/01/2015 ao dia 31/12/2020, já os parâmetros meteorológicos dos dados recebidos estão descritos na Tabela 7

Tabela 7 - Descrição dos parâmetros meteorológicos.

Variável Meteorológica	Unidade	Medida descritiva Horária
Temperatura do Ar	°C	Média
Umidade Relativa	%	Média
Molhamento Foliar	Adimensional	Média
Pressão Atmosférica Instantânea	hPA	Média
Radiação Solar	W/m ²	Média
Velocidade do Vento	Km/h	Máxima Horária
Precipitação	mm	Acumulado
Molhamento Foliar	-	Instantâneo Horário

Foi realizado o agrupamento das variáveis de granularidade horária para diária e gerado a partir desses dados novos parâmetros meteorológicos como demonstrado na Tabela 8.

Tabela 8 - Agregação dos Dados Meteorológicos

Parâmetro	Medida Descritiva
Temperatura do Ar	Média Diária
Máxima Temperatura do Ar	Máxima Diária
Mínima Temperatura do Ar	Mínima Diária
Umidade Relativa	Média Diária
Umidade Relativa Máxima	Máxima Diária
Umidade Relativa Mínima	Mínima Diária
Molhamento Foliar Médio	Média Diária
Molhamento Foliar Mediano	Mediana Diária
Radiação Solar Média	Média Diária
Radiação Solar Máxima	Máxima Diária
Radiação Solar Mínima	Mínima Diária
Velocidade Máxima do Vento	Média Diária
Máxima Velocidade Máxima do Vento	Máxima Diária
Mínima Velocidade Máxima do Vento	Mínimo Diária
Pressão Atmosférica	Média Diária
Precipitação	Somatória

5.2.1 Análise de Inconsistências

Devido a possibilidade de haver falhas na série histórica foi realizado uma análise em busca de valores nulos e valores inconsistentes. Com um total de 2186 dias de dados no *dataset*, obteve-se a percentagem de dados faltantes para cada variável apresentado na Tabela 9.

Tabela 9 - Porcentagem dos dados meteorológicos faltantes.

Parâmetro	Porcentagem de dados Faltantes
Temperatura do Ar	0.14 %
Máxima Temperatura do Ar	0.14 %
Mínima Temperatura do Ar	0.14 %
Umidade Relativa	54,47 %
Umidade Relativa Máxima	54,47 %
Umidade Relativa Mínima	54,47 %
Molhamento Foliar Médio	0,23 %
Molhamento Foliar Mediano	0,23 %
Radiação Solar Média	3.56 %
Radiação Solar Máxima	3.56 %
Radiação Solar Mínima	3.56 %
Velocidade Máxima do Vento	0,14 %
Máxima Velocidade Máxima do Vento	0,14 %
Mínima Velocidade Máxima do Vento	0,14 %
Pressão Atmosférica	0,14 %
Precipitação	0,14 %

5.2.2 Preenchimento de Valores Nulos

O preenchimento de valores nulos é um campo vasto na estatística e matemática, podendo-se utilizar de métodos como preenchimento por médias, medianas, replicação, substituição por meio da utilização de dados provenientes de estações meteorológicas próximas, entre outros (GERON, 2019).

Para o presente trabalho, foi estabelecido um limiar de ausência de dados de 25% de dados faltantes, logo, caso exista uma falha na série maior que 25%, o parâmetro de dado meteorológico é excluído por inteiro.

Outras estações meteorológicas com proximidades a estação da EPAGRI foram consultadas e encontrou-se, como dados abertos e disponíveis, as estações provenientes do Laboratório de Hidrologia (LABHIDRO) da UFSC nas coordenadas 27°35'57.768"S 48°31'9.12"O, o laboratório de Climatologia Agrícola (LabClimAgri) da UFSC nas coordenadas 27°34'54.106"S, 48°30'15.88"O e uma estação pertencente ao INMET referente ao aeroporto Hercílio Luz nas coordenadas 27°39'39.924"S, 48°33'1.188"O, contudo nenhuma das estações possui série histórica no período de agosto de 2015 a setembro de 2020 com dados de pressão atmosférica que pudessem preencher a lacuna dos dados faltantes referentes a estação da EPAGRI, dessa forma, optou-se pela exclusão do parâmetro pressão atmosférica do conjunto de dados meteorológicos da presente pesquisa.

Para o parâmetro de precipitação, houve três dias com dados faltantes, assim para preencher esses valores comparou-se a série histórica com a série histórica das estações do LABHIDRO e do LabClimAgri. Foi comparado a diferença de precipitação para um mesmo dia entre as estações a fim de verificar a proximidade absoluta dos valores de precipitação.

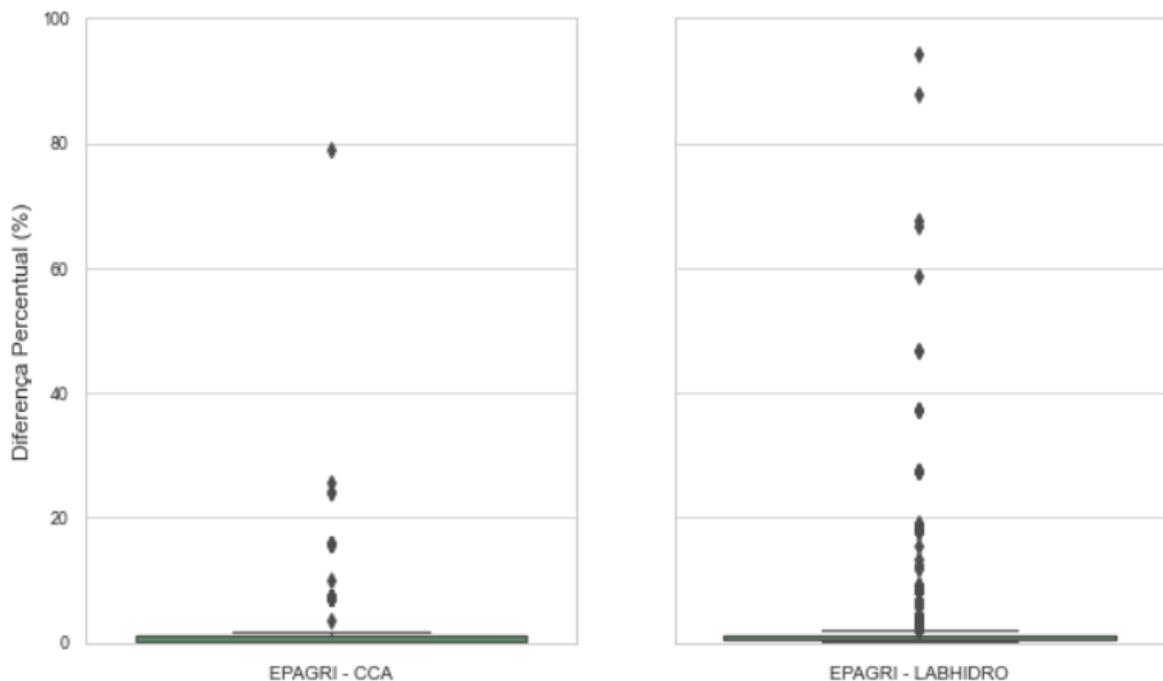
A diferença percentual foi calculada pela equação 31.

$$Dif = \frac{y_{epagri} - y_{outro}}{y_{epagri}} \quad (31)$$

Onde y_{epagri} representa o valor de precipitação da estação referente EPAGRI e y_{outro} representa o valor de precipitação da outra estação a ser comparada.

As diferenças percentuais entre a estação do LabClimAgri e da EPAGRI foram menores do que em comparação com as outras estações, portanto optou-se por preencher esses valores de precipitação com os dados provenientes da estação meteorológica do LabClimAgri. A Figura 9 demonstra as diferenças absolutas entre os valores de precipitação das estações.

Figura 9 - *Boxplot* das diferenças percentuais entre as estações meteorológicas



Fonte: Autor.

Foi realizada as mesmas verificações para o parâmetro de Temperatura e dada a pequena diferença entre as estações, por coerência, decidiu-se preencher com os valores provenientes da estação meteorológicas do CCA. Já para os demais parâmetros, não foi possível identificar valores de preenchimentos para as falhas na série em outras estações e, portanto, os valores ausentes foram preenchidos com a média da série histórica.

5.2.3 Adequação de Novas Variáveis

A fim de explorar melhor as características dos parâmetros dos dados meteorológicos, e gerar novos parâmetros que disponham boa influência na capacidade de generalização da RNA, foi criado novos parâmetros a partir da reunião de parâmetros já existentes, como média móvel de 3 dias, média móvel de 7 dias e somatórias, a Tabela 10 apresenta os novos parâmetros que foram incorporados aos dados meteorológicos.

Tabela 10 - Parâmetros Incorporados nos dados meteorológicos

Parâmetro	Medida de Cálculo
Precipitação Acumulada 3 dias	Somatório das precipitações nos 3 dias anteriores
Precipitação Acumulada 7 dias	Somatório das precipitações nos 7 dias anteriores
Temperatura do Ar 3 dias	Média Móvel de 3 dias
Temperatura do Ar 7 dias	Média Móvel de 7 dias
Temperatura do Ar Máxima de 3 dias	Média Móvel de 3 dias
Temperatura do Ar Mínima de 3 dias	Média Móvel de 3 dias
Radiação Solar Média 3 dias	Média Móvel de 3 dias
Velocidade Máxima do Vento Média de 3 dias	Média Móvel de 3 dias

5.3 UNIÃO DOS DADOS DE QUALIDADE E METEOROLÓGICOS

Por fim, com ambos os conjuntos de dados já devidamente corrigidos suas falhas foi feito a união dos conjuntos a partir das datas referentes aos dados de qualidade do WCV-FS, gerando assim um dataset de 103 linhas por 42 colunas (103x42), apresentado no Apêndice A que será a base de dados utilizada no presente estudo.

5.4 PRÉ PROCESSAMENTO DOS DADOS

5.4.1 Análise Descritiva

O *dataset* avaliado é descrito perante suas características estatísticas a partir de medidas de tendência central e testes de normalidade, bem como suas distribuições e correlações entre os dados.

5.4.2 Padronização dos Dados

A padronização dos dados é uma transformação linear aplicada durante o modelamento de redes neurais, com o objetivo de Y. Durante a padronização bastante utilizada para modelagem pois uma amplitude de valores de cada parâmetro do *dataset* nas redes neurais e na análise de componentes principais são processos que podem interferir no correto funcionamento dos algoritmos, dessa forma é realizado para garantir que todas os parâmetros e variáveis tenham a mesma importância (AKRATOS; PAPASPYROS; TSIHRINTZIS, 2008).

Dada a matriz $X (i x j)$, a padronização das variáveis é realizada a partir da equação 32.

$$Z_{ij} = \frac{x_{ij} - \bar{X}_j}{\sigma(X_j)} \quad (32)$$

Onde:

Z_{ij} = Variável Padronizada

\bar{X} = Media da j-ésimo parâmetro

$\sigma(X_j)$ = Desvio-padrão da j-ésimo parâmetro

5.4.3 Transformação Linear dos Dados

A transformação linear dos dados é uma técnica para ajustar a distribuição dos dados há distribuição normal.

von Sperling et al (2020) afirmam que se a assumption de normalidade não é satisfeita os níveis de significância podem apresentar distorção, e os métodos estatísticos podem sofrer perdas. Portanto as predições realizadas não possuirão robustez nos resultados.

A transformação linear logarítmica é dada pela equação 33:

$$\tilde{x}_i = \ln x \quad (33)$$

Onde: \tilde{x}_i é a x-ésima variável transformada.

5.4.3.1 Transformação BoxCox

A transformação Box-Cox foi proposta por Box e Cox em 1964 e é dado pela equação 34.

$$y(\lambda) = \begin{cases} \frac{y^{\lambda}-1}{\lambda} & \text{se } \lambda \neq 0 \\ \log y & \text{se } \lambda = 0 \end{cases} \quad (34)$$

Dessa forma, há um valor de λ que maximiza a normalidade da distribuição da amostra, ou seja, há um valor de λ que transforma o conjunto amostral de forma a estar mais próximo de uma distribuição normal teórica.

5.5 ANÁLISE DE COMPONENTES PRINCIPAIS (ACPS)

Na ACP os dados são transformados para um novo sistema de coordenadas com eixos ortogonais. A n-variáveis originais geram, através de suas combinações lineares n-CPs, cujas principais características são: ortogonalidade; não correlação e redução de dimensionalidade

Sejam as variáveis $X_1, X_2, X_3 \dots X_p$ e cada variável com n indivíduos. Esse conjunto $n \times p$ origina a matriz X ($n \times p$):

$$X = \begin{matrix} X_{11} & \dots & X_{1p} \\ \vdots & \ddots & \vdots \\ X_{n1} & \dots & X_{np} \end{matrix}$$

A interdependência entre as variáveis é representada pela matriz de covariância 'S', seja a covariância descrita pela equação 35:

$$\text{cov}(X, Y) = \frac{1}{n} \sum (X - \bar{X})(Y - \bar{Y}) \quad (35)$$

Onde:

\bar{X} e \bar{Y} são as médias de cada variável

A partir da matriz X ($n \times p$) matriz de correlação S ($p \times p$) é representada por:

$$S = \begin{matrix} Var(x_1) & Cov(x_1, x_2) & Cov(x_1, x_p) \\ Cov(x_2, x_1) & \ddots & Cov(x_2, x_p) \\ Cov(x_p, x_1) & Cov(x_p, x_2) & Var(x_p) \end{matrix}$$

Os componentes principais são determinados pela resolução da equação 36, característica de matrizes.

$$\det[S - \lambda I] = 0 \quad (36)$$

Onde:

I = Autovetores

λ = Autovalores da matriz

S = Matriz ($p \times p$)

Se a matriz S ($p \times p$), não apresentar nenhuma coluna que seja combinação linear da outra, a equação 36 terá 'p' raízes denominadas de autovalores.

Sejam $\lambda_1, \lambda_2, \lambda_3 \dots \lambda_p$ as raízes da equação característica da matriz S , então:

$$\lambda_1 > \lambda_2 > \lambda_3 \dots \lambda_p$$

Para cada Autovalor existe um Autovetor \tilde{a}_i :

$$\tilde{a}_i = \begin{matrix} a_{i1} \\ \vdots \\ a_{ip} \end{matrix}$$

Os autovetores são normalizados, isto é, a soma dos quadrados dos coeficientes é igual a 1, e são ortogonais entre si

Os componentes principais apresentam as seguintes propriedades:

- 1) A variância do componente principal Y_i é igual ao valor do autovalor λ_i :

$$Var(Y_i) = \lambda_i$$

- 2) O primeiro componente apresenta maior variância
- 3) O total de variância das variáveis originais é igual ao somatório dos autovalores que é igual ao total de variância dos componentes principais

$$\sum Var(X_i) = \sum \lambda_i = \sum Var(Y_i)$$

A contribuição C_i de cada componente principal Y_i é calculada dividindo a variância de Y_i pela variância total. Representa assim, a proporção da variância total explicada, representada pela equação (37).

$$C_i(\%) = \frac{\text{Var}(Y_i)}{\sum_{i=1}^p \text{Var}(Y_i)} = \frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \quad (37)$$

Onde C_i corresponde a variância explicada pela componente.

Assim, a porcentagem de variância nos dados explicada pelas CPs é igual ao acúmulo das variâncias explicadas por cada CP.

Johnson; Wichern, (1998) comentam que os últimos componentes principais são responsáveis por direções que não estão associadas a muita variabilidade, que identificam relações lineares entre as variáveis originais, próximos de constante.

Tomenko (2007) modelando a performance de sistemas de *wetlands* construídos para predição de DBO excluiu todas as componentes que representam menos de 2% da variabilidade dos dados.

Akratos; Papaspyros; Tsihrintzis (2008) desenvolvem uma RNA pra prever a remoção de DBO e DQO em WCHs utilizaram um número de componentes que explicam 75 % da variância dos dados para compor a RNA.

Q. Chen, A.E. Mynett (2003) descrevendo eutrofização no lago Taihu cita que um critério típico para o número de principais componentes é uma variância explicada de 75-95 % da variância total do *dataset*.

Para realizar a análise de componentes foi utilizado a linguagem de programação *Python* com as bibliotecas Scikit-Learn, Stats, Numpy e Pandas, as quais são conjuntos de módulos e funções úteis que reduzem o uso de código no programa, facilitando a programação e tornando a leitura do código facilmente legível.

A ACP é utilizada com dois propósitos no presente trabalho: a primeira para redução da dimensionalidade do conjunto de dados e, a segunda, para descarte de parâmetros.

Para redução da dimensionalidade, escolheu-se as componentes principais que representam 90% da variância, ou seja, compôs-se um novo *dataset* transformado que possui n-parâmetros. A matriz de transformação de componentes principais foi salva e usada para obter um conjunto de teste transformado do conjunto de teste original.

Para descarte das variáveis será utilizado o método B4, o qual consiste em selecionar uma variável por componente principal, e a redução da dimensionalidade será por adoção de critérios de variância total explicada.

Jolliffe (1972) desenvolveu o método B4 que utiliza as componentes principais para seleção dos parâmetros, sendo o método baseado no uso das primeiras componentes principais selecionadas, onde a variável de maior valor absoluto de *loading* correspondente a primeira componente principal é selecionada. A seguir, a próxima variável a ser selecionada será o maior valor de *loading* da segunda componente principal e assim sucessivamente até a CP selecionada. Caso a mesma variável apresente maior valor absoluto de *loading* em diferentes CPs então a variável a ser selecionada será a de segundo maior valor absoluto da componente principal correspondente. Os *loadings* são os coeficientes da combinação linear das variáveis originais a partir da qual são construídas as CPs.

Na tabela 11 apresentam-se os conjuntos de dados que serão construídos a partir da análise de componentes principais.

Tabela 11 - Novos datasets a partir da ACPs.

Apelido do Conjunto de Dados	Critério de Escolha	Descrição
ACP Dataset Reduzido	Loadings de cada Componente Principal	n-parâmetros a partir de n-CPs com 90 % da variância explicada
ACP Transformado Dataset 90	90 % de Variância Explicada	Transformação do dataset e redução da dimensionalidade para n-CPs

5.6 SELEÇÃO DE PARÂMETROS

Florestas Aleatórias é um método para classificação e regressão que combina a performance de várias árvores de Decisão - *Decision Trees* (DTs) para classificar ou prever os valores de uma variáveis. (BREIMAN, 2001, Guo et Al, 2011, RODRIGUEZ-GALIANO et al. 2012. Isto é, quando RF recebe um vetor de entrada (x_n) então RF constrói um número de K de árvores de regressão

Florestas Aleatórias são modelos do tipo *ensemble methods* que combinam a predição de um conjunto de árvores de decisão a fim de obter uma única resposta de saída, que segundo Alvarenga (2018) tendem a apresentar melhor resultado que as árvores de decisão devido a redução da variância. A equação a seguir descreve a saída para o modelo

$$\hat{f}_{fa}^B(x_i) = \frac{1}{B} \sum_{b=1}^B T(x_i, \Theta_b)$$

Onde, B é o número total de árvores, T() representa a resposta de uma árvore b para um vetor de entrada x_i e Θ_b representa o parâmetro dessas árvores.

Dessa forma, as florestas aleatórias utilizam um técnica de *bootstrap* para criar um subconjunto de dados que são utilizados para crescimento da árvore.

No presente trabalho, é utilizado o algoritmo XGBoost que utiliza de árvores de decisão e gradiente e, partir da modelagem do algoritmo, é possível observar como cada variável está influenciando o modelo na previsão dos dados, por fim, uma análise da influência dessas variáveis na previsão de modelo é feita de forma que variáveis expressam pouco influência na predição serão excluídas e portanto a *dataset* será reduzido então a apenas as variáveis que ponderam com maiores valores no processo preditivo.

5.7 SELEÇÃO DO CONJUNTO DE DADOS

A partir do pré-processamento dos conjuntos de dados detalhados nas seções anteriores, serão utilizados quatro conjuntos de dados para construção das RNA, sendo esses *datasets* exemplificados na Tabela 12.

Tabela 12 - Conjuntos de Dados utilizados para construção da RNA.

Conjunto de Dados	Apelido do Conjunto de Dados	Descrição
<i>Dataset</i> original	df_original	<i>Dataset</i> Original pós processo de limpeza e padronização
<i>Dataset</i> com parâmetros reduzidos	Df_reduzido	Advindo da escolha de parâmetros proposta por árvores de decisão
<i>Dataset</i> Transformado por Componentes Principais	Pca_df_90	Derivado do CP com uma variância explicada de 90 % da variância total do <i>Dataset</i>
<i>Dataset</i> com parâmetros reduzidos por Componentes Principais	Pca_df_reduzido	Método B4 através de análise de <i>loadings</i> das CPs

5.8 REDE NEURAL ARTIFICIAL

Para a modelagem da RNA será utilizado a linguagem computacional *Python* através das bibliotecas TensorFlow, Keras, Scikit-Learn, Pandas e Numpy.

5.8.1 Divisão do Conjunto de dados

Para realizar o processo de modelagem da RNA é necessário dividir a base de dados para que uma parcela seja utilizada como dados de treinamento da RNA e a outra parcela utilizada para teste e validação do modelo. Os valores de divisão mais utilizados são entre 20-30 % da base para teste, e 70-80 % dos dados para treino, com as porcentagens variando conforme a escolha do modelador e o tamanho total do conjunto de dados.

Akratos C; Papaspyros J; Tsihrintzis V (2008) utilizaram uma divisão dos dados de 2:1:1 onde 50% dos dados são usados para treino, 25% para teste do modelo e o restante é usado para validação. Para um trabalho dos mesmos autores publicado em 2009 também foi utilizado as mesmas taxas na repartição do conjunto de dados.

Naz; Uyanik; Yesilnacar (2009) comparando WCH e *Wetlands* de Superfície Livre com modelagem através de RNAs utilizaram uma taxa de divisão entre treino e validação de 2:1:1 também.

Para o presente trabalho utilizou-se as porcentagem relativas a 75 % para base de dados de treino e teste da RNA e 25 % para validação do modelo, a partir dos 75% de dados para treino é utilizado 25 % dos dados de treino para teste da própria RNA.

O processo de divisão realizado com a biblioteca Scikit-Learn em python é realizado através da seleção de uma *seed*, que é um número inteiro comumente entre [0,100] a qual cada número da *seed* representa uma escolha de índices do dataset que irão compor os dados de treino e os dados de validação, sendo assim, a partir da seleção de uma *seed*, independente do conjunto de dados (desde que não tenha sido feita alterações no ordenamento dos valores) será dividido de forma idêntica pois as escolhas dos índices de divisão são os mesmos, sendo estes atrelados ao gerador *seed*.

5.8.2 Topologia das RNAs

A escolha dos valores dos parâmetros da RNA foram baseados conforme os estudos apresentados a seguir.

Naz et al (2009) realizando um estudo de a comparação de WCH e WCs de superfície livre e modelagens de RNAs utilizou como valores para sua rede 1 camada intermediária com 20 neurônios e com função de ativação de *tan-sigmoide*, na camada de saída utilizou uma função de ativação ReLU e 3 neurônios de saída inferindo sobre 3 parâmetros de qualidade de efluente diferente. Para o algoritmo de *backpropagation* foi testado 12 diferentes *backpropagations* algoritmos.

Akratos; Paspaspyros; Tsihrintzis (2008) modelando RNAs para predição de remoção de DBO e DQO em WCH utilizaram 1 camada intermediária com função de ativação hiperbólica tangente (*tan h*) que é uma variação da função sigmoide de ativação e o número de neurônios na camada intermediária foi testado de 10 a 20 neurônios na camada, na camada de saída apenas 1 neurônio com função de ativação ReLU. Foi utilizado um algoritmo tradicional de *backpropagation* e rodado por 100 épocas (*epochs*) com uma taxa de aprendizagem de 0.01.

De acordo com Minghao; Xuelei; Houijian (2009) que utilizaram uma RNA para avaliar a qualidade do ecossistema de uma lagoa na China, uma rede neural com pelo menos uma camada intermediária que consegue capturar qualquer função não-linear, assim foi utilizado na RNA 1 camada intermediária com um número de neurônios dados por $n_{ci} = 2 * n_{ce} + 1$ onde n_{ci} que resulta no número de variáveis na camada de entrada. Como funções de ativação foi usado tan-sigmoide na camada de entrada e log-sigmoide na camada intermediária. No seu conjunto de dados de treino há 60 valores que representam ao treino e 15 valores que representam teste.

Tomenko et al (2007) testou de 6 a 20 neurônios na camada intermediária, com um total de 1000 epochs e 0.15 de taxa de aprendizagem. Após encontrar o melhor número de neurons foi feito um ajuste fino dos parâmetros da RNA utilizando taxa de aprendizagem de 0.09.

Yang et al (2022) testou de 3 a 30 neurônios e de 1 a 3 camadas intermediárias, encontrou a topologia da rede neural com melhor acurácia de três camadas intermediárias, com 18 neurônios na primeira camada, 14 neurônios na segunda camada intermediária, 6 neurônios na terceira camada intermediária.

Ozengin; Elmaci; Yonar (2016) estudaram a aplicação de uma rede neural para previsão de remoção de nutriente em um WCH e encontraram como topologia uma camada intermediária com função de ativação *log-sigmoid* em ambas as camadas.

A fim de encontrar a melhor topologia da RNA será testado os seguintes parâmetros (Tabela 13) através de uma busca exaustiva de parâmetros para validação dos dados de treino.

Tabela 13 - Composição dos Parâmetros que serão testados na RNA.

Parâmetro da RNA	Valores
Camadas Intermediárias	1 – 3 camadas
Neurônios por camadas	3 – 30 neurônios
Função de Ativação – Camada Intermediária	ReLU; Sigmoid
Função de Ativação na Camada de Saída	Linear; ReLU
Taxa de Aprendizado	0.09
Algoritmo de BackPropagation	Adam
<i>Epochs</i>	1000

A busca dos melhores parâmetros para o modelo será rodando uma série de modelos para cada possibilidade de conjunto de variáveis de entrada, dessa forma seleciona-se os melhores modelos segundo as métricas de avaliação propostas por esse trabalho.

5.8.3 - Validação do Modelo

A parte do conjunto de dados de treino, que representam 75% da parcela total do conjunto de dados são validados com as métricas propostas em 25 % do conjunto total de dados e comparados para fins de análises e verificações.

6. RESULTADOS

Todos os códigos utilizados para desenvolvimento do presente trabalho, estão abertos e estão disponibilizados através do link:

https://github.com/emanuel-gf/Wetlands_Artificial_Neural_Network_TCC.git

Ou através do github: @emanuel-gf – Wetlands_Artificial_Neural_Network_TCC

6.1 ANÁLISE DESCRITIVA

Os parâmetros de qualidade da água para efluentes e para os dados meteorológicos estão apresentados no Apêndice B.

Baseando-se na proposta apresentada por Warrick & Nielson (1980) para classificação da variabilidade dos dados, a partir do coeficiente de variação (CV) os parâmetros Molhamento Foliar Médio, Precipitação, Radiação Solar Mínima, Radiação Solar Mediana, Molhamento Foliar Mediano, Mínima Velocidade horária do Vento Máxima, Precipitação Acumulada 3 dias, Precipitação Acumulada 7 dias, Sólidos Suspensos Afluente, Relação DQO/DBO, Nitrato afluente e DQO afluente, apresentaram alta variabilidade ($CV \geq 60$) e os demais parâmetros apresentaram variabilidade menores.

Quanto a assimetria da distribuição dos dados em relação a uma distribuição normal, o parâmetro Skewness revela um indicador de normalidade, quando mais próximo de 0 mais “ideal” é a assimetria; para valores > 0 existe uma assimetria positiva, e negativa para valores < 0 .

Os parâmetros que apresentaram valores de Skewness elevados, maiores que 1, são: Precipitação, Radiação Solar Mediana, Mínima Velocidade do Vento Máxima, Precipitação Acumulada 3 dias, Precipitação Acumulada 7 dias, Sólidos Totais Afluente, Sólidos Totais Voláteis Afluente, Sólidos Totais Fixos Afluente, Relação DQO/DBO, Nitrato Afluente e Carregamento Orgânico Superficial. Portanto, apresentam uma assimetria da distribuição normal positiva, ou seja, os dados possuem uma assimetria a direita. Segundo von Sperling et al (2020) em plantas de tratamento as concentrações dos parâmetros de qualidade são normalmente com assimetria positiva, o que corrobora com as observações das variáveis de qualidade encontradas.

Já os parâmetros com assimetria negativa e menores que 1 são: Umidade Ar máxima, Alcalidade Afluente e Eficiência. von Sperling et al (2020) citam que assimetrias negativas são normais para parâmetros de eficiência de remoção, dessa forma o parâmetro de eficiência de remoção corrobora com as observações descritas.

Os parâmetros que possuem assimetria próxima de zero e, portanto, se assemelhando a distribuição normal são: Temperatura do ar médio, Umidade relativa, Molhamento Foliar Medio, Pressão Atmosférica Média, Radiação Solar média, Média da Velocidade Máxima do Vento, Umidade do Ar mínima, Temperatura Máxima, Temperatura Mínima, Radiação Solar Máxima, Molhamento Foliar Mediano, Máxima Velocidade do Vento horária máxima, Temperatura Mediana, Média 3 dias da Temperatura do ar, Temperatura do Ar média 7 dias,

Temperatura Máxima do Ar média de 3 dias, Temperatura mínima do ar média de 3 dias, Radiação Solar Média de 3 dias, Velocidade máxima do vento média de 3 dias, pH, DBO, Amônia, Nitrito, Fósforo e Fluxo Superficial.

6.1.1.1 Distribuição Normal

Utilizando os testes de normalidade de Anderson-Darling, Shapiro-Wilk, verificou-se a hipótese de normalidade na distribuição dos dados com um nível de significância de 5%. O Apêndice C apresenta os resultados dos testes de normalidade bem como seus respectivos parâmetros.

Para os parâmetros originais do conjunto de dados, os parâmetros que segundo os testes estatísticos de Shapiro-Wilk e Anderson-Darling apresentam distribuição normal são os seguintes: Pressão Atmosférica Média, Temperatura Máxima, Máxima Velocidade Máxima do Vento, Temperatura do Ar 7 dias, Temperatura do Ar máxima média 3 dias, Velocidade Máxima do Vento média 3 dias e Fósforo para parâmetro de qualidade.

Há parâmetros que para um teste foram atestados como normalmente distribuídos porém no outro não, são esses os parâmetros: Temperatura do Ar Média e Velocidade Máxima do Vento.

Assim, observando os valores do valor crítico para o teste de Anderson-Darling no parâmetro de Temperatura do Ar Média o valor de A^2 foi de 0.790 sendo o valor crítico de 0.759, o que nega a hipótese e, portanto, resulta em uma distribuição não normal, mas já para o teste de Shapiro-Wilk o p-value apresenta valor de 0.054 para um nível alpha de 0.05 o que em tese aceita a hipótese da distribuição normal, porém devido a proximidades de valores entre p-value e alpha entende-se que o parâmetro tende a não ser normalmente distribuído já que ambos os testes corroboraram para valores muito próximos aos valores de corte dos testes.

Já para o parâmetro de Velocidade Máxima do Vento atestou confirmando a normalidade no teste de Anderson-Darling porém para Shapiro-Wilk não foi atestado como normalmente distribuído e possuindo um valor do p-value de 0.017 o que de fato é bem menor que o valor de significância, podendo-se concluir dessa forma que esse parâmetro também não possui distribuição normal dos dados para a série histórica referente.

Segundo Ladim (2003) apud Abrahão (2012) não sendo satisfeita a condição de normalidade devem ser testadas transformações das variáveis, especialmente do tipo logarítmica, quadrática ou *boxcox*, contudo a partir de qualquer transformação há a necessidade

de realizar posteriormente a transformação inversa, trazendo os resultados para a distribuição original.

Assim foi aplicado a regularização através da transformação linear para os parâmetros que não possuíam distribuição normal nos testes de distribuição realizados logo acima e portanto foi refeito os testes para os parâmetros.

6.1.1.2 Transformação Linear - BoxCox

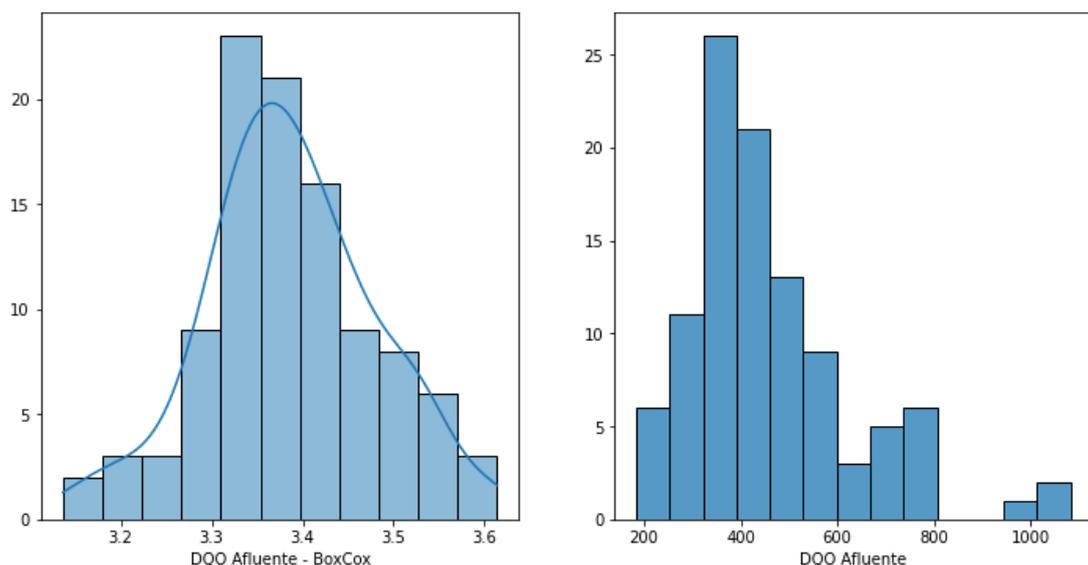
A partir da transformação, foi realizado novamente os testes estatísticos de distribuição normal para as variáveis e os seguintes parâmetros transformados foram validados com distribuição normal, são eles: Temperatura do Ar média, radiação solar média, Velocidade Máxima do Vento, Temperatura Mediana, Radiação Solar Média 3 dias, DQO, Relação DQO/DBO.

Os parâmetros que atestaram normalidade para um teste e para outro não, após a transformação linear passaram a corroborar com a hipótese de normalidade para ambos os testes estatísticos.

Os demais parâmetros não foram validados nos testes estatísticos de distribuição normal mesmo após transformação *boxcox* e portanto são considerados não normais.

A Figura 10 exemplifica a transformação *boxcox* na distribuição do parâmetro de DQO.

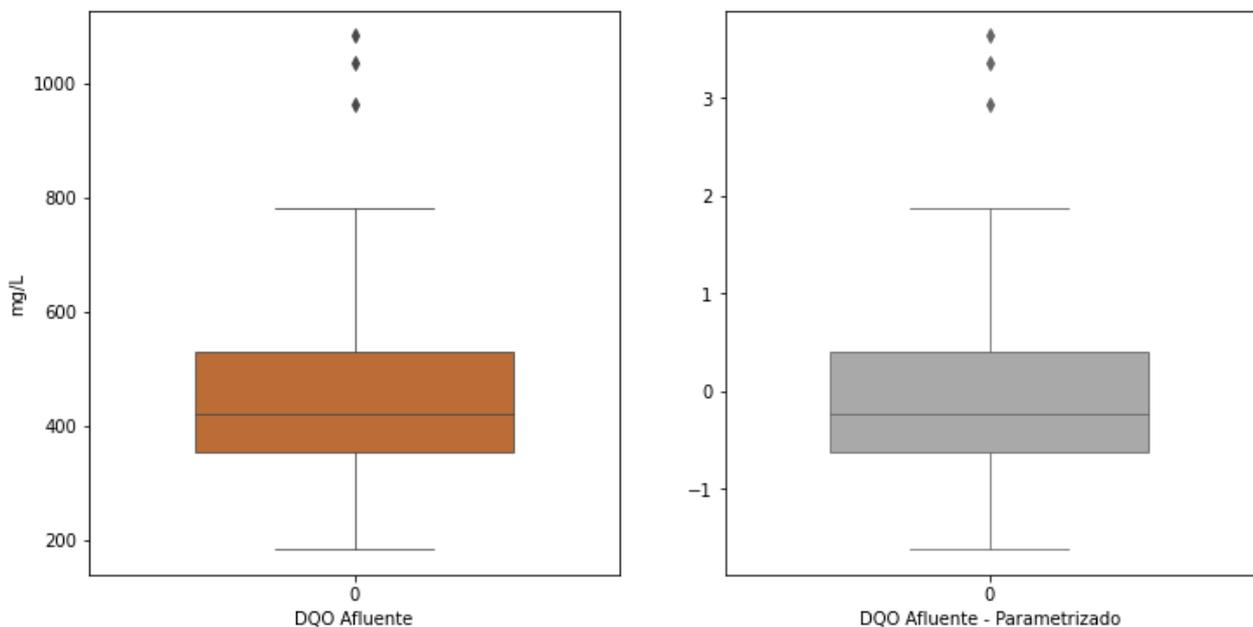
Figura 10 – Distribuição dos dados para o parâmetro de DQO afluente ao WCV-FS componente da EETE – GESAD/UFSC, após e anterior a transformação *boxcox*.



6.2 PADRONIZAÇÃO DOS DADOS

Foi aplicado a padronização dos dados referente a média e ao desvio padrão, como indica a Figura 11, a padronização não altera a distribuição dos dados, apenas diminui a amplitude e altera para cada parâmetro do conjunto de dados a média para zero e desvio padrão para 1.

Figura 11 – Diagrama de caixas para efeito da parametrização no conjunto de dados para o parâmetro de DQO afluente ao WCV-FS componente da EETE – GESAD/UFSC



6.3 REDUÇÃO DE DIMENSIONALIDADE

Segundo Tomenko (2007) um problema que afeta a performance das redes neurais é chamada de curso de dimensionalidade (“curse of dimensionality”) proposto por (Friedman, 1995) que quando a dimensionalidade do espaço aumenta há um incremento no volume do espaço dimensional maior que o os dados disponíveis e portanto, os dados tornam-se esparsos, dessa forma a diminuição da dimensionalidade dos dados pode superar os resultados do modelo em questão de acurácia.

Foi utilizado o algoritmo *XGBoost* para predição do conjunto de dados pois as árvores de decisão suportam ponderações de cada variável na determinação dos valores de DQO, a importância é calculada pelo número de vezes que uma variável é dividida e multiplicado pelo quadrado do erro absoluto, assim é possível observar a importância de cada variável na variável

preditora. Hastie et al (2010) citam que as estruturas das árvores de decisão dependem do conjunto de dados e pequenas mudanças nos dados resultam em diferentes divisões das árvores, onde esses fatores introduzem incertezas e limites preditivos.

O Apêndice D contém os resultados da influência de cada parâmetro no processo de predição do modelo.

Tabela 14 sintetiza como cada variável influenciou na predição do resultado para um vetor singular de entrada, dessa forma os parâmetros de qualidade da água como nitrato, DQO, sólidos suspensos, nitrito, pH bem como os parâmetros de critério de dimensionamento carregamento orgânico superficial e taxa aplicação hidráulica superficial obtiveram maior peso no processo de predição. Já os parâmetros meteorológicos umidade mínima do ar, radiação solar de média móvel de 3 dias, umidade máxima do ar, temperatura mediana, pressão atmosférica mediana, velocidade máxima do vento diária e precipitação ponderam com maior influência quando comparado as demais variáveis meteorológicas.

Tabela 14 – Relação de importância de cada parâmetro no processo de predição do modelo *XGBoost*.

Parâmetro	Importância do Parâmetro para <i>XGBoost</i>
Nitrato	0.142
DQO	0.093
Carregamento Orgânico Superficial	0.073
DBO	0.062
Umidade do Ar mínima	0.056
Alcalinidade	0.041
Radiação Solar Média de 3 dias	0.040
Sólidos em Suspensão	0.036
Taxa Hidráulica Superficial	0.033
Nitrito	0.028
Relação DQO/DBO	0.027
Umidade Máxima do Ar	0.024
Temperatura Mediana	0.023
Pressão Atmosférica Média	0.020
Sólidos Totais Voláteis	0.019
Umidade Relativa	0.019
Máxima Velocidade do Vento	0.018
pH	0.018
Precipitação	0.017

Os parâmetros mais insignificantes, apontado no Apêndice D, foram Molhamento Foliar Mediano, Precipitações acumuladas de 7 dias e temperaturas médias do ar de 3 e 7 dias. As variações de parâmetro da radiação solar com radiação solar mínima e radiação solar máxima diária não forneceram bons resultados quanto a importância da predição. A

precipitação acumulada de 3 e 7 dias também não foi significativa dentro do processo de previsão, e a precipitação diária ficou no limiar do limite de influência adotado.

De forma geral, os parâmetros de qualidade obtiveram maiores ponderações na influência do processo de previsão do que comparado a parâmetros meteorológicos.

Segundo Prasetiyowati. et al (2021) o limite de importância para redução da dimensionalidade é utilizado comumente como 0.05, ou determinado independente. Foi assumido, arbitrariamente, para o presente estudo um valor de limite de 0.016. Dessa forma, os parâmetros considerado estão descritos na Tabela 15:

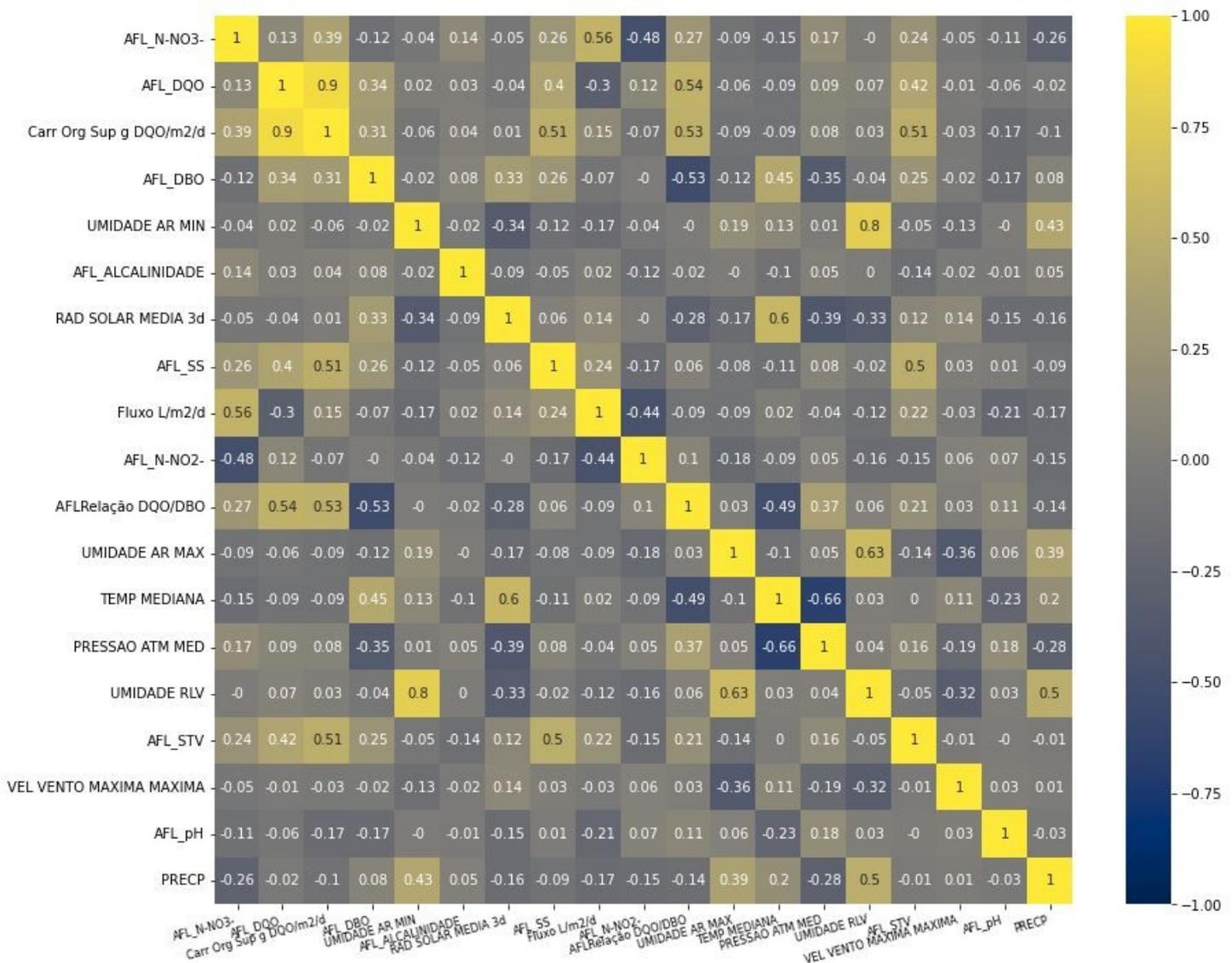
Tabela 15 - Parâmetros Seleccionados.

Parâmetro	“Alias” do Parâmetro	Importância do Atributo
Nitrato	AFL_N-NO3	0.142
DQO	AFL DQO	0.093
Carregamento Orgânico Superficial	Carr Org Sup DQO/m2/d	0.073
DBO	AFL DBO	0.062
Umidade Ar Mínima	Umidade Ar Min	0.056
Alcalinidade	AFL ALCALINIDADE	0.041
Radiação Solar Média de 3 dias	RAD SOLAR MEDIA 3D	0.040
Sólidos Suspensos	AFL SS	0.036
Taxa hidráulica superficial	Fluxo L/m2/d	0.033
Relação DQO/DBO	AFL Relacao DQO/DBO	0.028
Umidade do ar máxima	UMIDADE AR MAX	0.027
Temperatura Mediana	TEMP MEDIANA	0.024
Pressão Atmosférica	PRESSAO ATM MED	0.023
Umidade Relativa	UMIDADE RLV	0.020
Sólidos Totais Voláteis	AFL STV	0.019
Máxima Velocidade do Vento	VEL VENTO MAXIMA MAXIMA	0.018
pH	AFL PH	0.018
Precipitação	PRECP	0.017

Akratos (2008) utilizou parâmetros meteorológicos com médias móveis iguais ao tamanho do tempo de detenção do WC, assim, seus parâmetros meteorológicos de média móvel 6 dias foram umidade do ar, pressão barométrica, radiação solar, pressão barométrica, velocidade do vento e precipitação e encontrou que radiação solar média de 6 dias foi o parâmetro de maior influência. Dado o tempo de alimentação contínua do WCV-FS do GESAD é de 3,5 dias, sendo os demais 3,5 dias da semana em repouso, infere-se que a radiação solar de média de 3 dias foi um dos melhores parâmetros meteorológicos que influenciaram o processo preditivo.

Com um propósito de diminuir a correlação entre os parâmetros de qualidade e meteorológicos e buscar montar um conjunto de dados de valores independentes, foi descrita a correlação dos mesmos para exclusão dos parâmetros que podem ser considerados como altamente correlatos, e parâmetros observados como redundantes. A matriz de correlação é descrita na Figura 12.

Figura 12 - Matriz de Correlação dos Parâmetros Seleccionados referente a série histórica de dados afluente, efluente e climáticos influentes no módulo WCV-FS da EETE GESAD/UFSC



Observa-se um alta correlação entre DQO e carregamento orgânico superficial (índice 0.9). Como o carregamento depende da variável DQO em seu cálculo, já era esperado um maior grau de correlação entre os mesmos. Para umidade do ar e umidade do ar mínima a correlação de 0.8 é bastante alta, dessa forma decidiu-se por excluir o parâmetro de carregamento orgânico

superficial e umidade relativa do conjunto de dados devido a correlação e certa redundância dos dados. A relação entre DQO/DBO possui uma correlação razoável com as variáveis DBO, DQO e carregamento, portanto decidiu-se por excluí-la já que seu cálculo é dependente da DBO e DQO. A radiação solar média de 3 dias possui uma correlação alta temperatura medianas indicando que dias mais quentes recebem mais radiação solar. Há uma correlação razoável entre sólidos totais, sólidos totais voláteis bem como sólidos suspensos, contudo optou-se por manter os mesmos presentes no conjunto, já que estes são esperados para serem degradados por heterotróficos e, portanto, contribuirão para a redução da DQO do sistema.

A Tabela 16 - Parâmetros selecionados da matriz de correlação. indica os parâmetros restantes que serão utilizados para compor o *dataset* e utilizado para treinamento e teste da rede neural.

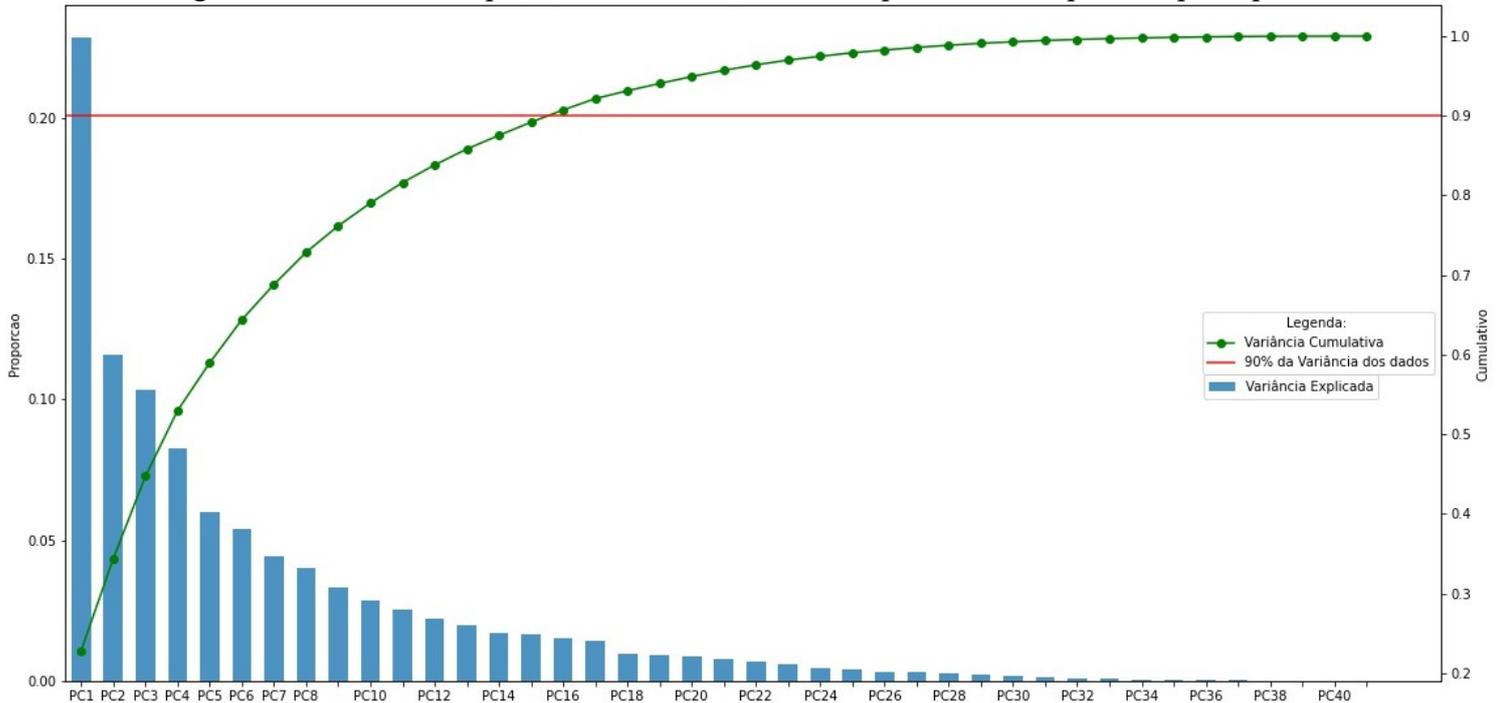
Tabela 16 - Parâmetros selecionados da matriz de correlação.

Parâmetro	“Alias” do Parâmetro
Nitrato	AFL_N-NO3
DQO	AFL DQO
DBO	AFL DBO
Umidade Ar Mínima	Umidade Ar Min
Alcalinidade	AFL ALCALINIDADE
Radiação Solar Média de 3 dias	RAD SOLAR MEDIA 3D
Sólidos Suspensos	AFL SS
Taxa hidráulica superficial	Fluxo L/m2/d
Umidade do ar máxima	UMIDADE AR MAX
Temperatura Mediana	TEMP MEDIANA
Pressão Atmosférica	PRESSAO ATM MED
Sólidos Totais Voláteis	AFL STV
Máxima Velocidade do Vento	VEL VENTO MAXIMA
pH	AFL PH
Precipitação	PRECP

6.4 COMPONENTES PRINCIPAIS

As componentes principais foram aplicadas para reduzir a dimensão do conjunto de dados de entrada da RNA, a Figura 13 apresenta os resultados em gráfico de barras mostrando a variância explicada de cada CP.

Figura 13 – Variância explicada e variância acumulada para cada componente principal.



Através do gráfico é possível verificar que a maior quantidade de variância explicada dos dados é representada pela primeira componente principal, a qual explica 23% da variância total dos dados originais e cada componente representa menos a variância explicada dos dados do que comparado a primeira componente. Para explicar 75% da variância total dos dados é necessário 8 componentes principais e 90% da variância dos dados é explicada por 15 componentes principais.

O Apêndice E apresenta a tabela de componentes principais, proporção explicada por cada variância e variância cumulativa explicada por cada componente.

Através das componentes principais são extraídas as principais variáveis originais pelo valor absoluto dos *loadings* como descrito pelo método B4. O Apêndice F apresenta, os valores dos *loadings* para cada variável dada componente principal. Como critério de seleção uma variância explicada de 90% foi adotada o que infere 15 parâmetros que irão compor a rede

neural, dessa forma os parâmetros selecionados pelo método de componentes principais estão apontados na Tabela 17.

Tabela 17 – Lista de Parâmetros Selecionados por CPs para 90% da variância explicada

Parâmetro Selecionados	Componente Principal
Carregamento Orgânico Superficial	1ª Componente Principal
DQO	2ª Componente Principal
pH	3ª Componente Principal
Precipitação Acumulada de 3 dias	4ª Componente Principal
Nitrito	5ª Componente Principal
Umidade Relativa	6ª Componente Principal
Molhamento Foliar Mediano	7ª Componente Principal
Pressão Atmosférica	8ª Componente Principal
Nitrato	9ª Componente Principal
Amônia	10ª Componente Principal
Sólidos Suspensos	11ª Componente Principal
Temperatura do Ar média de 3 dias	12ª Componente Principal
Alcalinidade	13ª Componente Principal
Velocidade Máxima do Vento	14ª Componente Principal
Temperatura Mediana	15ª Componente Principal

Akratos; Papaspyros e Tsihrintzis (2008) encontraram como parâmetros utilizando o método de componentes principais temperatura, tempo de detenção hidráulico, porosidade e mês para expressar mudanças sazonais devido através da análise dos *loadings*. Como infere a Tabela 17, foi encontrado no presente trabalho temperatura do ar média de 3 dias e Temperatura Mediana como parâmetro, além de precipitação acumulada de 3 dias, pressão atmosférica, molhamento foliar mediano e máxima velocidade do vento. Os parâmetros de qualidade foram Carregamento Orgânico superficial, DQO, pH, Nitrito, Nitrato, Amônia, Sólidos Suspensos e Alcalinidade.

O método de redução de dimensionalidade através dos *loadings* indicou um maior número de parâmetros meteorológicos do que parâmetros de qualidade. Como o método avalia principalmente a variância dos dados e não processos bioquímicos de consumo de matéria orgânica no WC, entende-se que tais parâmetros para o método possuem maiores influência na variância do conjunto de dados do que real ponderação de influência no processo de remoção da demanda química de oxigênio dos efluentes dentro de um *wetland* construído.

6.5 TOPOLOGIA E RESULTADO DAS REDES NEURAIIS

Para os 4 conjuntos de dados existentes foram testadas as topologias propostas na Tabela 13 da metodologia do presente trabalho, assim conforme os critérios de validação por erro quadrático médio, R^2 e R^2 ajustado, foi escolhida a melhor topologia dos modelos de redes neurais artificiais que estão apresentados na Tabela 18.

Tabela 18 – Melhor Topologia Obtida para cada *dataset*.

Conjunto de Dados	Número de Camadas Intermediárias	Neurônios por camada	Função de Ativação das Camadas Intermediárias	Função de Ativação da Camada de Saída
Dataset original	3	16,8,4	ReLU	Linear
Dataset com parâmetros reduzidos	3	16,16,8	ReLU	Linear
Dataset Transformado por Componentes Principais	2	16,2	ReLU	Linear
Dataset com parâmetros reduzidos por Componentes Principais	2	4,2	ReLU	Linear

A partir das topologias encontradas para cada conjunto de dados foi realizado um aperfeiçoamento do modelo para ajuste fino dos parâmetros da rede neural como o tamanho do lote, número de épocas e taxa de aprendizado.

Após o ajuste fino foi rodado novamente as redes neurais e os resultados das redes neurais diante dos critérios de avaliação estabelecidos estão demonstrado na Tabela 19.

Tabela 19 - Resultado das Redes Neurais para cada *dataset*.

Conjunto de Dados	Apelido Conjunto de Dados	Erro Quadrático Médio	R^2	R^2 ajustado
Dataset original	Df original	273.89	0.801	0.642
Dataset com parâmetros reduzidos	Df reduzido	398.47	0.711	0.505
Dataset Transformado por Componentes Principais	Df PCA 90	244.81	0.822	0.676
Dataset com parâmetros reduzidos por Componentes Principais	Df PCA Reduzido	421.22	0.694	0.482

Com base na tabela acima é possível observar que o conjunto de dados transformado por componentes principais apresentou-se como o que melhor generalizou o conjunto de dados de validação, obtendo um erro quadrático médio de 244.81 e um coeficiente de regressão ajustado de 0.676. O conjunto de dados originais possuiu uma performance significativa

atingindo um erro quadrático médio de 273.89 e um R^2 ajustado de 0.642, possuindo assim um bom ajuste a base de dados de validação. O conjunto de dados com parâmetros reduzidos e selecionados por meio de árvores de decisão apresentou um erro quadrático médio de 398.47 e um R^2 ajustado de 0.505.

Por fim, o conjunto reduzido através das componentes principais e do método b4 apresentou um erro quadrático de 421.22 e um R^2 ajustado de 0.482. Considera-se que devido ao processo de seleção dos parâmetros não ter passado por um critério de escolha baseado em funcionamentos real do sistema de WCs era esperado um menor desempenho para a base de dados. Akrtos; Papaspyros e Tsihrintzis (2008) utilizaram o método de *loadings* para seleção dos melhores parâmetros de qualidade e meteorológicos para compor o modelo porém para cada componente principal o parâmetro era escolhido comparando valores de *loadings* próximos e justificando escolhas através de literatura consolidada como parâmetros de operação e parâmetros de qualidade.

Tomenko; Ahmed e Popop (2007) utilizaram como critério de seleção as componentes que possuíam uma variância explicadas com mais de 2%, comparando conjunto de dados originais e conjunto de dados transformados por componentes principais encontraram um erro quadrático médio um pouco maior para o conjunto transformado por componentes principais Para um conjunto de validação de 90 entradas o erro quadrático médio e o R^2 do conjunto de dados total foram, respectivamente, de 486 e 0.882 e o do transformado foi de 442.06, 0.842 indicando resultados próximos entre os conjuntos de dados e um ajuste adequado de generalização das RNAs.

Os resultados da previsão das redes neurais para cada conjunto de dados estão apresentados em gráficos na Figura 14.

Figura 14 – Resultado das RNAs para os conjuntos de dados com diferentes variáveis de entrada ao módulo WCV-FS componente da EETE – GESAD/UFSC.

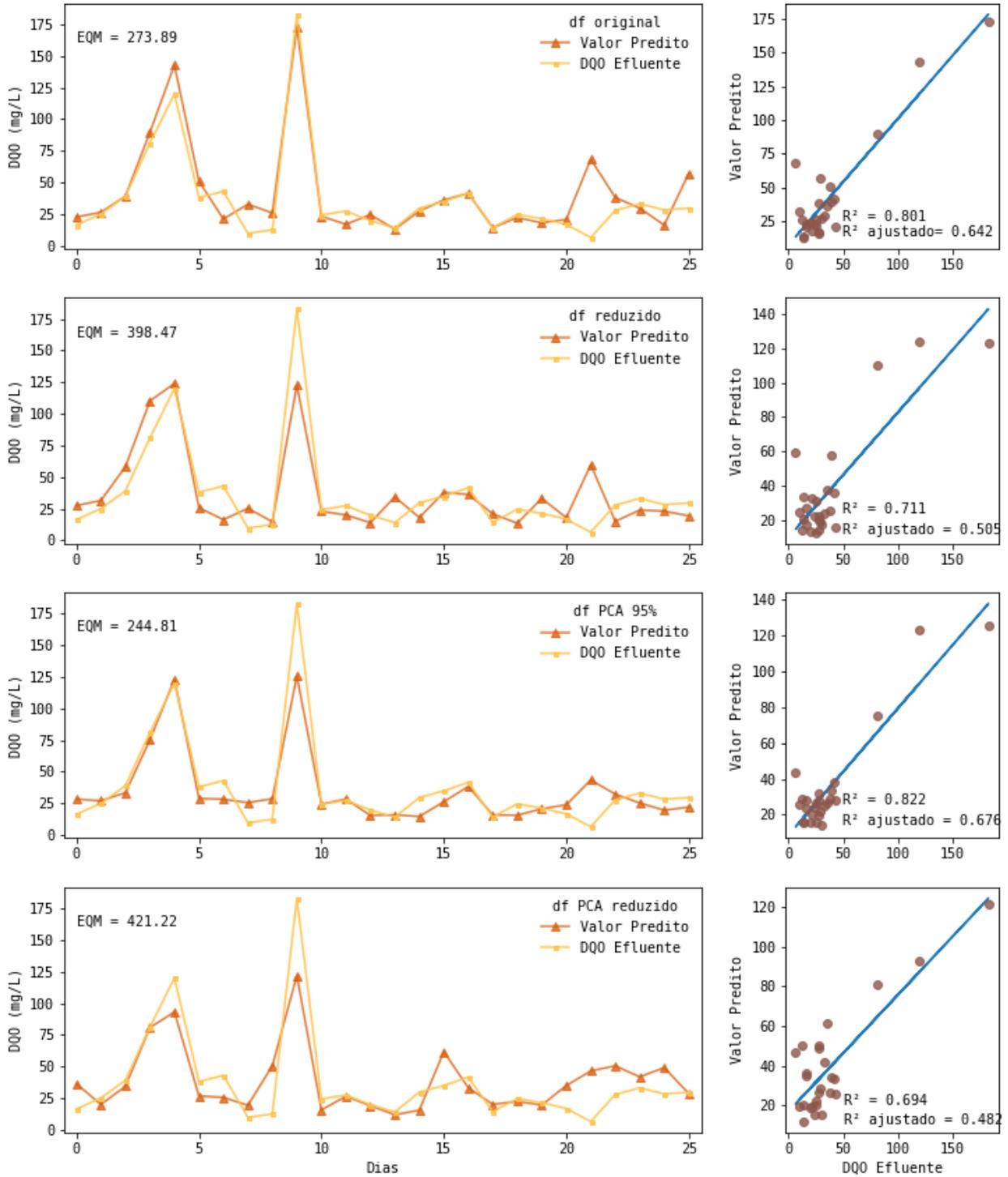
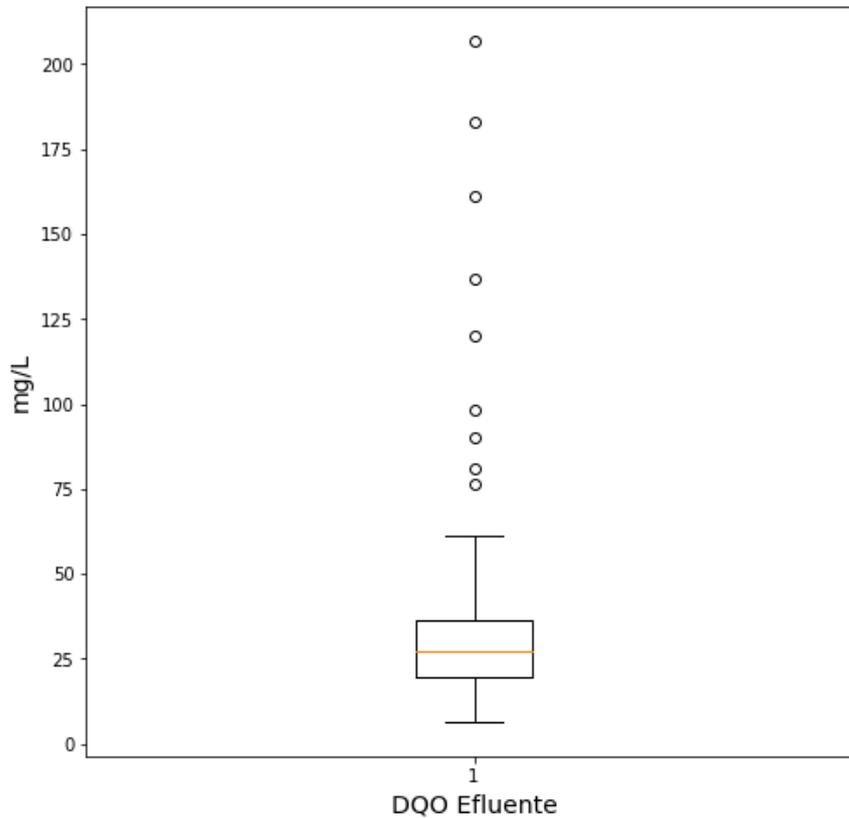


Figura 15 - Diagrama de caixa para DQO efluente do módulo WCV-FS componente da EETE – GESAD/UFSC.



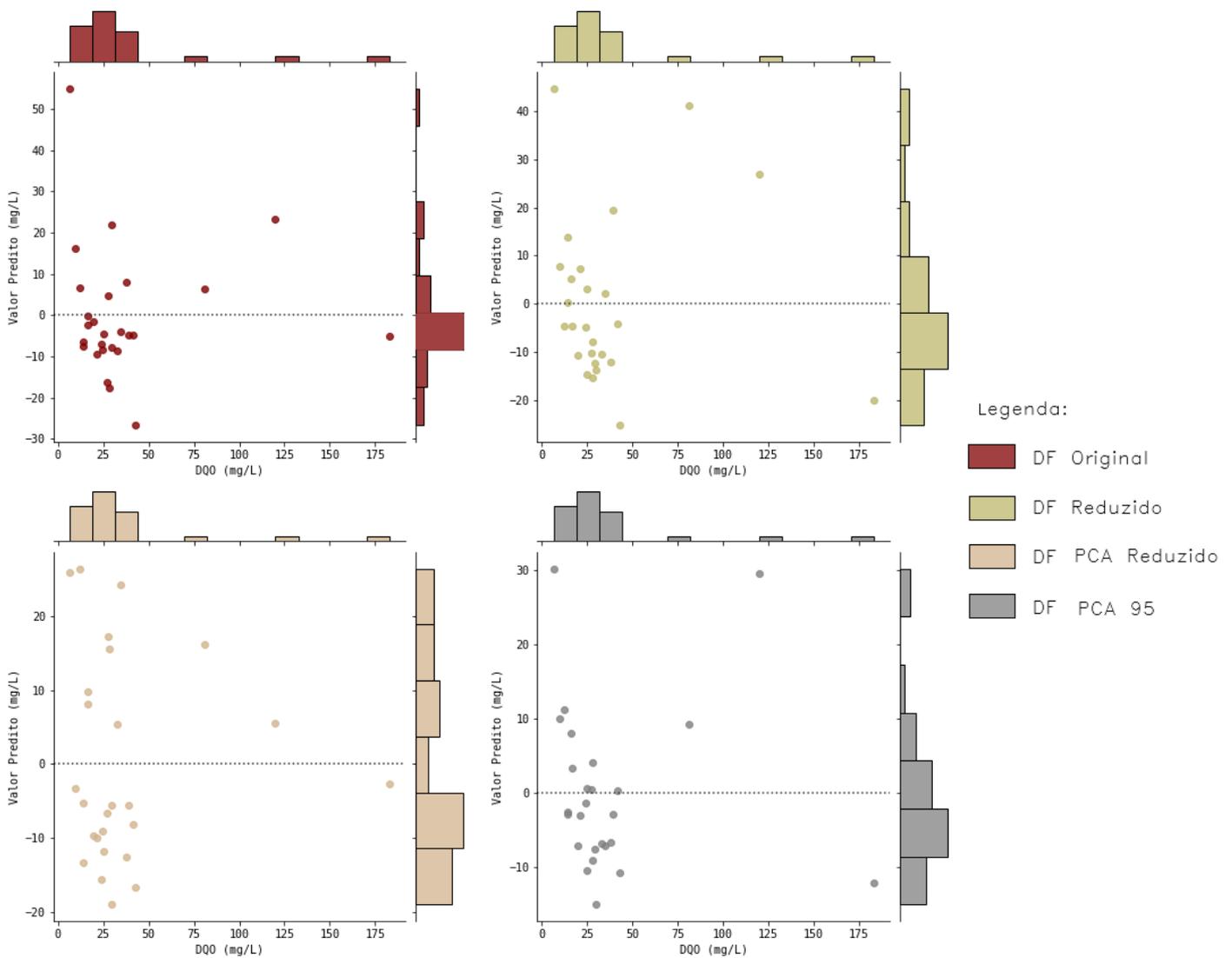
Com base nos resultados apresentados na Figura 14 e da variação dos dados de DQO apresentados na Figura 15, é possível observar que os valores *outliers* de DQO efluente, ou seja, valores possuem uma menor frequência amostral nos dados, foram previstos com maiores erros significativos. As RNAs para quaisquer conjuntos de dados previu os mesmos aumentos da DQO na saída como são os valores reais observados, contudo obteve para esses *outliers* as maiores distâncias absolutas entre o valor previsto e o valor real. Nenhuma rede neural conseguiu generalizar com exatidão esses valores de DQO maiores, porém, se destacou a rede neural do conjunto de dados originais que previu bem os valores mais altos de DQO quando comparado as outras redes.

É possível observar também que para valores mínimos as RNAs não conseguiram atingir bons resultados, indicando essa dificuldade de previsão para valores *outliers* da série histórica. Como observado na figura, todas as redes neurais superestimaram o valor mínimo de DQO de saída, sendo esse um erro em comum para as quatro redes neurais o que indica que não há dados suficientes com valores de DQO de saída pequenos (< 19 mg/L) para treinar a RNA a fim de generalizar tais valores.

Devido as necessidades reais de operação das WCs estarem em prever sobrecargas nos sistemas, entende-se que há uma maior grau de importância na acurácia de maiores valores de DQO, seja por processos legais e necessidades de atingir uma eficiência de remoção estabelecida por órgãos reguladores bem como de diminuir o impacto dos efluentes com elevada concentração de matéria orgânica nos ecossistemas locais.

Por fim foi observado como os erros residuais - apresentados na Figura 16 se comportam, para averiguar que a não ocorrência de viés de generalização do modelo, subestimando ou superestimando os valores.

Figura 16 - Erros Residuais das RNAs e suas distribuições para cada conjunto de dados referentes ao módulo WCV-FS componente da EETE – GESAD/UFSC.



Em geral é possível observar uma tendência em prever os valores de previsão para menos, ou seja, negativamente, como mostra a maior distribuição dos dados no quadrante negativo do gráfico de erros. De maneira que os modelos apresentaram uma boa homocedasticidade dos erros residuais, isto é, a variância dos erros não está condicionada a valores constantes.

Quanto a distribuição dos erros residuais, ao RNA do conjunto de dados originais, apresentaram visualmente uma melhor distribuição, contudo nenhum modelo apresentou uma distribuição que fosse genuinamente normal. Os erros residuais apresentaram uma não constante variância residual, pois os conjuntos de dados que apresentaram menores valores de EQM possuem a maioria dos erros residuais entre 10 e -10, indicando assim uma boa resposta de generalização do modelo para os dados de validação, contudo houve erros absolutos na casa de 40 e 50 mg/L principalmente para valores de DQO que era excessivamente altos ou excessivamente baixos.

É necessário ressaltar sobre as limitações da pesquisa, que envolvem uma base de dados com uma série histórica, resultando em 103 amostras somente para um período de três anos. As redes neurais possuíram um resultado diante das capacidades de generalização das mesmas.

7. CONCLUSÃO

Foi construído na presente pesquisa modelos de predição da quantidade de matéria orgânica representada pela DQO de saída dos *Wetlands* Construídos Verticais de Fundo Saturado. Assim, foram construídos modelos de redes neurais artificiais e como técnica de pré-processamento dos dados, normalizações, árvores de predições e análise de componentes principais.

De forma a avaliar o desempenho dos modelos com a aplicação de tais técnicas, quatro modelos foram comparados. Utilizando o conjunto de dados inteiro para treinar um modelo (M1), a redução de dimensionalidade a partir de ponderações de influência de árvores de decisão (M2), e dois modelos que utilizaram a análise de componentes principais com os propósitos de seleção das CPs e descarte de parâmetros do conjunto de dados original, (M3 e M4 respectivamente).

De acordo com os resultados, os modelos advindos do conjunto de dados originais (M1), obtida de um conjunto de dados transformados composto por 15 componentes principais que contêm 90 % da variância dos dados originais (M3), apresentaram melhores resultados do que os outros dois modelos (M2 e M4), isso considerando o EQM, R^2 bem como o R^2 ajustado.

O M1 apresentou um EQM e um R^2 ajustado, respectivamente de 273.89,0.64. Já o M3 apresentou um EQM e R^2 ajustado de 244.81 e 0.676, respectivamente. O modelo advindo da seleção de parâmetros através de árvores de decisão (M2) apresentou 398.47, 0.507 e, por fim, a RNA com propósito de descarte de parâmetros através de CPs (M4) apresentou 421.22 e 0.482 de EQM e R^2 ajustado, respectivamente.

Os resultados apontam para a necessidade de inferir diferentes variáveis que compoñham o conjunto de dados de treino da redes neurais, a fim de melhorar a acurácia de predição. Embora existam diversas técnicas de pré-processamento dos dados, o PCA apresentou um aumento de desempenho no modelo quando transformou o conjunto de dados e reduziu a dimensionalidade, contribuindo para um modelo mais robusto.

Apesar do conjunto de dados originais conter muitas variáveis, sendo algumas com um alto coeficiente de correlação entre elas, atingiu resultados satisfatórios para as métricas de avaliação das RNAs estabelecidas, não sendo possível observar o nível de ruído que essas variáveis redundantes impactaram no modelo.

Os resultados dos quatro modelos previram corretamente os aumentos da DQO efluente dos WCV-FS, apesar dos erros absolutos maiores para valores de DQO mais altos houve a tendência correta de previsão dos modelos o que indica a possibilidade de uso de redes

neurais artificiais para prever sobrecargas nos sistemas de tratamento, baseado nas características de qualidade do esgoto afluente e parâmetros meteorológicos.

É importante ressaltar que mesmo com uma população não tão robusta, com apenas 103 indivíduos compondo cada parâmetro do conjunto de dados, houve resultados interessantes que destacam a viabilidade de uso das RNAs para predição de concentração de matéria orgânica na saída dos WCs.

Esta pesquisa buscou avaliar a aplicação de RNAs para previsão de saída dos esgotos efluentes dos WCs e utilizou diferentes métodos de pré-processamento dos dados para otimizar os resultados e a acurácia de previsão dos modelos. Embora algumas técnicas utilizadas nesse estudo não tenham mostrado eficiência em todas as suas explicações, existem outras técnicas presentes na literatura que podem ser exploradas.

Recomendações para próximas pesquisas

Durante o desenvolvimento deste estudo foram identificados pontos que necessitam de uma investigação mais profunda e servem como continuação do trabalho, sendo:

- Aumentar o número de indivíduos do conjunto de dados para atingir maiores variâncias do conjunto e mais robustez para o modelo.
- Testar diferentes transformações que visam adequar a distribuição dos parâmetros para uma distribuição normal, ou aumentar o conjunto de dados até que se obtenha distribuições normais para cada parâmetro.
- Utilização de limites de influência para seleção de parâmetros por Prasetiyowati (2021) utilizando o modelo XGBoost
- Trabalhar com escolhas mais criteriosas e embasadas na literatura para a seleção dos parâmetros através dos métodos de CPS por *loadings* e do método de influência de previsão das árvores de decisão.
- Utilizar validação cruzada (*cross-validation*) na construção dos modelos de RNA para formar uma maior variação de dados no treinamento do modelo e escolha do melhor modelo por média dos critérios de avaliação para cada parcela do conjunto de dados.
- Testar como diferentes porcentagens de variância explicada que compõem o conjunto de dados transformados afetam os resultados dos modelos de RNA.
- Comparar o desempenho do modelo de *deep-learning* com modelos de *machine-learning* para predição de DQO efluente em WCV-FS.

- Utilização de variáveis de qualidade que podem ser aferidas a partir de processos em monitoramento de tempo real (exemplo: pH, condutividade, espectrometria, turbidez, oxigênio dissolvido).

REFERÊNCIAS

- AGUIAR, Silvana dos Reis. **Projeto executivo de um sistema de *wetland* construído para tratamento de esgoto doméstico**. 2020. 66 f. TCC (Graduação) - Curso de Engenharia Sanitária e Ambiental, Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2020. Disponível em: <https://repositorio.ufsc.br/handle/123456789/212047>. Acesso em: 16 out. 2022.
- Akratos, C. S., Papaspyros, J. N. E., & Tsihrintzis, V. A. *Total nitrogen and ammonia removal prediction in horizontal subsurface flow constructed wetlands: Use of artificial neural networks and development of a design equation*. **Elsevier**. Bioresource Technology, p. 586–596. Set. 2009.
- Akratos, C. S., Papaspyros, J. N. E., & Tsihrintzis, V. A. *An artificial neural network model and design equations for BOD and COD removal prediction in horizontal subsurface flow constructed wetlands*. **Elsevier**. Chemical Engineering Journal, p. 96–110. Dez 2008.
- ALVARENGA JÚNIOR, Wagner José de. **Métodos de otimização hiperparamétrica: um estudo comparativo utilizando árvores de decisão e florestas aleatórias na classificação binária**. 2018. 82 f. Dissertação (Mestrado) - Curso de Engenharia Elétrica, Universidade Federal de Minas Gerais, Belo Horizonte, 2018.
- B. Yang, Z. Xiao, Q. Meng, Y. Yuan, W. Wang, H. Wang, Y. Wang, X. Feng. *Deep learning-based prediction of effluent quality of a constructed wetland*. **Environmental Science and Ecotechnology**. Chinese Society for Environmental Sciences, Harbin Institute of Technology, Chinese Research Academy of Environmental Sciences. P. 1-11. Jan. 2022.
- BASSANI, Leandro. **Influência do modo de alimentação sobre o desempenho de *wetland* construído vertical de fundo saturado no tratamento de esgoto sanitário**. 2022. 258 f. Tese (Doutorado) - Curso de Engenharia Sanitaria e Ambiental, Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2022.
- BOX, G. E. P. and Cox, D. R.. *An analysis of transformations*, **Journal of the Royal Statistical Society**, Series B, 26, 211-252, London, England, 1964.
- BRASIL. Ministério da Saúde - Fundação Nacional de Saúde. **Wetlands Construídos Aplicados no Tratamento de Esgoto Sanitário: recomendações para implantação e boas práticas de operação e manutenção**. Florianópolis: Copiart, 2018.
- BRASIL. Lei nº. 11.445, de 05 de janeiro de 2007. Estabelece diretrizes nacionais para o saneamento básico; altera as Leis nos 6.766, de 19 de dezembro de 1979, 8.036, de 11 de maio de 1990, 8.666, de 21 de junho de 1993, 8.987, de 13 de fevereiro de 1995; revoga a Lei no 6.528, de 11 de maio de 1978; e dá outras providências.
- BRASIL. Sistema Nacional de Informação sobre Saneamento. Diagnóstico Temático Serviços de Água e Esgotos 2021. Disponível em:

http://www.snis.gov.br/downloads/panorama/PANORAMA_DO_SANEAMENTO_BASICO_NO_BRASIL_SNIS_2021.pdf . Acesso em: 12 out. 2022.

BARBOSA, Leandro. **Utilização de técnicas de análise multivariada na avaliação de características quantitativas de uma população f2 de suínos**. 2003, 92 f. Tese (Doutorado). Pós-Graduação em Zootecnia, Universidade Federal de Viçosa, Viçosa, 2003.

COELHO, L. dos S.; MARIANI, V. C. Sistema híbrido neuro-evolutivo aplicado ao controle de um processo multivariável. **Revista Controle & Automação**. Campinas, São Paulo, Fevereiro/Março 2006.

CRUZ, Cosme Damião. **Aplicação de algumas técnicas multivariada no melhoramento de plantas**. 1990. 188p. Tese (Doutorado em Genética e Melhoramento) – Escola Superior de Agricultura, Universidade de São Paulo, Piracicaba, 1990.

DE PAOLI, A. C., VON SPERLING, M. Avaliação das condições hidrodinâmicas de wetlands construídos de escoamento horizontal subsuperficial. **Gesta**, v.1, p.213-222, 2013.

DOS SANTOS, Rubens. Francisco; PIERRE, Silvia; PELOSI, Elisabeth; TADEU, Francisco; Abordagem descentralizada para concepção de sistemas de tratamento de esgoto doméstico. **Revista Eletrônica de Tecnologia e Cultura**. Volume 16; p. 35-44, São Paulo, 2015.

FREITAS, M. N. de. **Avaliação da hidrodinâmica e desempenho de um wetland construído horizontal componente de um sistema híbrido empregado no tratamento de esgoto sanitário**. 2017. Dissertação (Mestrado). Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia Ambiental, Florianópolis, 2017.

Garcia, J., Rousseau, D.P.L., Morato, J., Lesage, E., Matamoros, V., Bayona, J.M. Contaminant removal process in subsurface-flow constructed wetlands: a review. **Critical Reviews in Environmental Science and Technology** 40 (7), Londres, 2010.

GÉRON, Aurélien. *Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems*. **O'Reilly Media**, California, 2019.

GRANATA, F., GARGANO, R., & de Marinis, G. *Artificial intelligence based approaches to evaluate actual evapotranspiration in wetlands*. **Science of the Total Environment**, pp.1–15. Califórnia, 2022

HONGYU, Kuang. LÚCIA, Vera Lucia Martins. M. OLIVEIRA Gilmar Junior. Análise de Componentes Principais: resumo teórico, aplicação e interpretação *Principal Component Analysis: theory, interpretations and applications*. **Engineering and Science**, pp. 1–8, Mato Grosso, 2015.

JOHNSON, R.A.; WICHERN, D.W. *Applied multivariate statistical analysis*. **Madison: Prentice Hall International**, Londres, 1998.

LARSEN, Tove. UDERT, Kai. LIENERT, Judit. Source separation and decentralization for wastewater management. **IWA Publishing**, Londres, 2013.

LEITE, Alfredo Estevão de Barros. **Simulação do lançamento de esgotos domésticos em rios usando um modelo de qualidade d'água, SisBAHIA**. 2014. 94 f. Dissertação (Mestrado). Fiocruz, Escola Nacional de Saúde Pública, Rio de Janeiro, 2004.

MANERICHI, Guilherme Moser. **Sistema de Classificação de Terrenos baseado em Redes Neurais e Visão Computacional**. 2014. 52 f. TCC (Graduação) - Curso de Engenharia de Controle e Automação e Computação, Centro de Blumenau, Universidade Federal de Santa Catarina, Blumenau, 2014.

MARSAGLIA, George. MARSAGLIA, John. Evaluating the Anderson-Darling Distribution. *Journal of Statistical Software*, 9(2), pp. 1–5. Londres, 2004.

MELLO, Marília Terra de. **Aplicação de redes neurais artificiais no processo de precificação de ações**. 2004. 49 f. TCC (Graduação). Curso de Ciência da Computação, Instituto de Física e Matemática, Universidade Federal de Pelotas, Pelotas, 2004.

MORAIS, Jacina Tabita Gurgel. **Análise de Componentes Principais Integrada a Redes Neurais Artificiais para Predição de Matéria Orgânica**. 2011. 68 f. Dissertação (Mestrado). Curso de Engenharia Industrial, Universidade Federal da Bahia, Salvador, 2011.

Nayak, P.C., Satyaji Rao, Y.R., Sudheer, K.P.,. Groundwater level forecasting in ashallow aquifer using artificial neural network approach. **Water Resources Management** 20, pp. 77–90, Londres, 2006.

Naz, M., Uyanik, S., Yesilnacar, M. I., & Sahinkaya, E.. Side-by-side comparison of horizontal subsurface flow and free water surface flow constructed wetlands and artificial neural network (ANN) modelling approach. *Ecological Engineering*, 35(8), pp. 1255–1263, Londo, 2009.

NILSSON, N. J. *Artificial Intelligence. A New Synthesis*. Morgan Kaufmann Publishers, San Francisco, 1998.

NOGUEIRA, Angelo Baruffi. **Análise de viabilidade no uso de Deep Learning para contagem de pessoas com câmeras de segurança**. 2018. 78 f. TCC (Graduação). Curso de Engenharia de Controle e Automação, Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2018.

PAWLYN, Michael, **Biomimicry in architecture**. London, RIBA. 2011

PLATZER, Christoph. *Design recommendations for subsurface flow constructed wetlands for nitrification and denitrification*. **Water Science and Technology**, v. 40, n. 3, pp. 257- 263, Londres, 1999

QIUWEN, Chen. MYNETT, Artur. *Integration of data mining techniques and heuristicknowledge in fuzzy logic modelling of eutrophication in Taihu Lake*. **Ecological Model**. 162. pp. 55–67, Delft, 2003.

SAEED, Tanveer. SUN, Guangzhi. . *A review on nitrogen and organics removal mechanisms in subsurface flow constructed wetlands: Dependency on environmental parameters*,

operating conditions and supporting media. In *Journal of Environmental Management* .Vol. 112, pp. 429–448, Londres, 2012.

SANTOS, M. O. dos. **Wetlands Construídos verticais modificados e intensificados aplicados no tratamento de esgoto sanitário**. 2017. Dissertação (Mestrado). Curso de Engenharia Sanitária e Ambiental, Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2017.

SEZERINO, P. H et al. **Wetlands Construídos Aplicados no Tratamento de Esgoto Sanitário**. GESAD, UFSC, Florianópolis, 2018.

SILVA, Arieleen Reis da. **Influência do Regime Hidráulico no Desempenho de Wetlands Construídos Verticais Empregados no Tratamento de Esgoto Sanitário**. 2021. 65 f. Dissertação (Mestrado). Curso de Engenharia Sanitária e Ambiental, Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2021.

SILVA, I. N. d. SPATTI, D. H. FLAUZINO, R. A. Redes neurais artificiais: para engenharia e ciências aplicadas. [S.l.]: ARTLIBER, São Paulo, 2010.

SOARES, I. S. ALENCAR, R. J. N. D. Identificação de correntes de intrush na proteção diferencial de transformadores de potência através de redes neurais artificiais MLP. **XII Simpósio Brasileiro de Automação Inteligente (SBAI)**, Natal, 2015.

Song, Keunyea. Park, Young-Seuk. Zheng, Fawen. Kang, Hojegong. *The application of Artificial Neural Network (ANN) model to the simulation of denitrification rates in mesocosm-scale wetlands*. **Ecological Informatics** 16, pp. 10-16. Toronto, 2013

SCHROEDER, Amanda Kempt. **Avaliação de tecnologias normatizadas e wetlands construídos empregados no tratamento descentralizado de esgoto por meio de análise multicriterial**. 2020. 139 f. Dissertação (Mestrado) - Curso de Engenharia Ambiental, Engenharia Sanitária e Ambiental, Universidade Federal de Santa Catarina, Florianópolis, 2020.

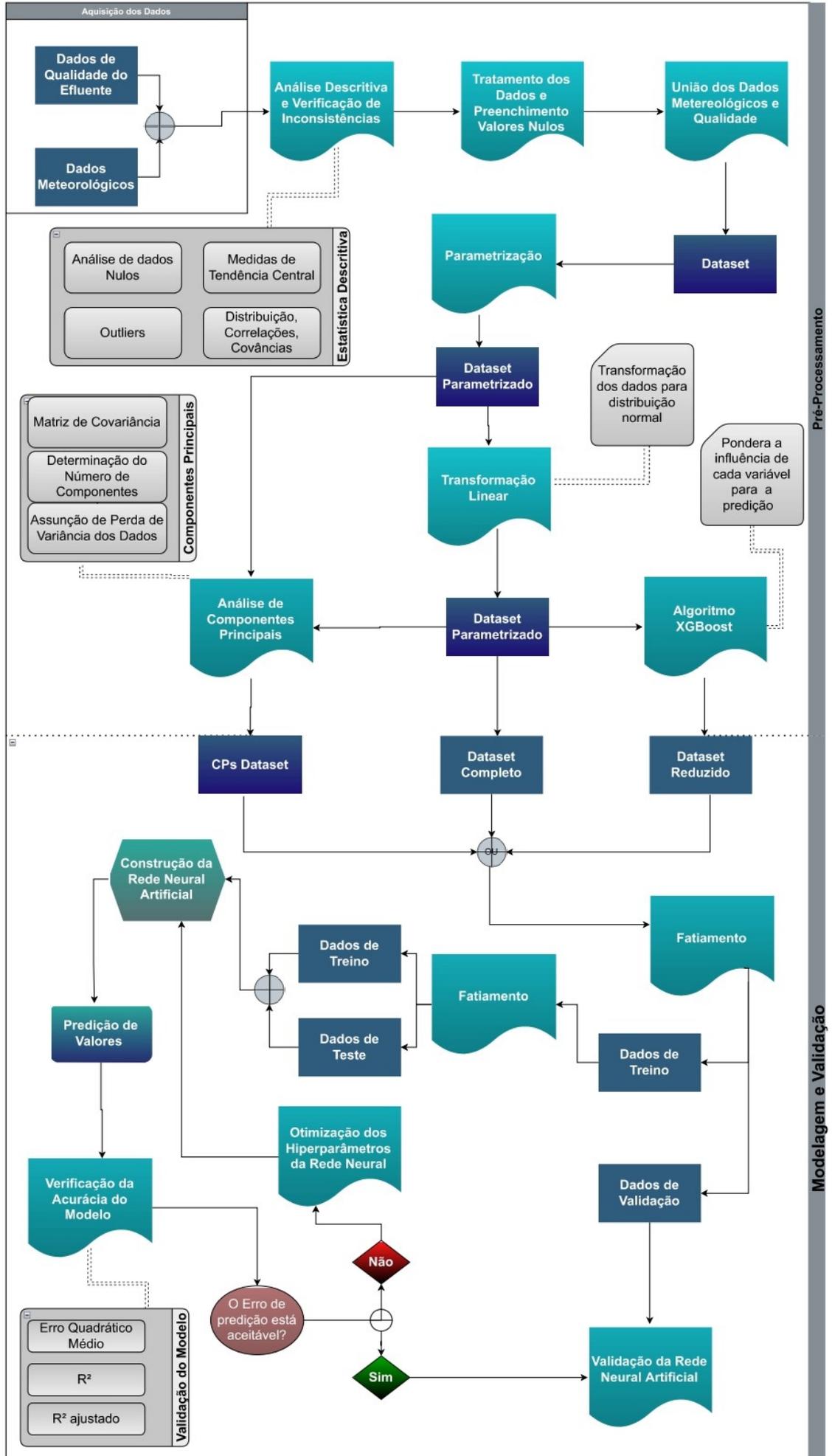
TOMENKO, V., Ahmed, S., & Popov, V.. Modelling constructed wetland treatment system performance. **Ecological Modelling**, 205(3–4), 355–364. Londres, 2007

VARELLA, Carlos Alberto Alves. **Análise de Componentes Principais: análise multivariada aplicada a ciências agrárias**. 2018. 12 f. NOTAS DE AULA - Curso de Ciências Agrárias, Universidade Federal Rural do Rio de Janeiro, Seropédica, 2018.

VON SPERLING, Marcos *et al.* **Assessment of Treatment Plant Performance and Water Quality Data: a guide for students, researchers and practitioners**. Londres, Iwa Publishing, 2020.

WORLD HEALTH ORGANIZATION – Water, Sanitation and Hygiene Links to Health. November, 2004.

**APÊNDICE A – MAPA CONCEITUAL DA METODOLOGIA UTILIZADA
NO PRESENTE TRABALHO**



**APÊNDICE A – CONJUNTO DE DADOS DE METEOROLOGIA E
QUALIDADE HÍDRICA.**

Data	TEMP AR MÈDIO	UMIDADE RLTV	MOLHAMENTO FOLIAR MÈDIO	PRESSAO ATM	RADIAÇAO SOLAR MÈDIA	VEL MAX VENTO MÈDIA	UMIDADE AR MAX	UMIDADE AR MIN	TEMP MAX	TEMP MIN	PRECP	RADIAÇAO SOLAR MAX	RADIAÇAO SOLAR MIN	RADIAÇAO SOLAR MÈDIANA	MOLHAMENTO FOLIAR MÈDIANO	VEL VENTO MAXIMA MAXIMA	TEMP MÈDIANA	VEL VENTO MAXIMA MIN	PRECP ACUM 3d	PRECP acum 7d	TEMP AR 3d	TEMP AR 7d	TEMP AR MAX 3d	TEMP AR MIN 3d
2015-06-25 C	16,24	86,94	0,60	1031,19	105,39	9,71	98,43	62,24	20,63	13,51	0,53	417,40	0,09	4,24	0,90	23,29	14,95	2,12	0,58	0,68	16,11	16,49	19,34	13,67
2015-07-02 C	17,30	81,46	0,00	1019,21	68,30	9,26	90,35	71,69	19,57	15,85	0,02	426,45	0,00	0,05	0,00	13,07	17,10	3,89	0,57	1,66	18,01	17,95	22,03	14,62
2015-07-23 C	15,68	95,02	0,57	1024,36	40,45	8,66	100,00	86,57	17,59	13,52	24,41	194,76	0,00	0,48	0,93	35,64	16,54	3,53	27,25	48,17	16,12	17,64	18,86	13,72
2015-08-06 C	20,88	85,89	0,17	1014,21	140,69	21,27	94,35	74,28	23,19	19,26	0,02	556,12	0,00	0,57	0,00	41,62	20,13	4,25	0,07	0,42	20,27	20,01	23,55	17,45
2015-08-13 C	21,28	85,09	0,35	1019,38	146,01	23,86	92,44	70,60	24,78	19,83	0,02	598,73	0,00	0,53	0,05	34,24	20,21	19,04	0,07	5,75	21,26	21,36	24,17	19,79
2015-08-20 C	17,94	74,85	0,17	1019,44	105,53	24,36	96,23	62,07	20,59	15,20	0,02	481,34	0,00	2,70	0,00	40,57	17,87	4,25	2,60	2,70	19,67	20,28	22,96	17,20
2015-08-27 C	19,40	83,82	0,43	1013,40	113,11	17,46	100,00	57,87	24,44	13,07	18,30	471,74	0,00	1,68	0,27	39,17	19,71	3,53	48,27	72,70	18,45	18,50	21,20	15,15
2015-09-03 C	16,66	94,47	0,99	1015,50	52,41	6,82	98,31	84,56	19,44	15,15	0,27	185,30	0,38	9,87	1,00	13,75	16,33	2,81	4,88	4,97	16,72	18,32	20,00	14,40
2015-09-10 C	18,91	90,48	0,59	1009,99	125,09	21,47	100,00	73,56	23,17	13,90	13,48	629,48	0,20	4,21	1,00	36,00	19,38	2,48	28,51	91,28	18,92	17,77	22,36	15,17
2015-09-24 C	23,12	86,59	0,09	1006,67	180,33	24,81	96,89	69,58	28,76	20,45	0,27	723,01	0,01	11,33	0,00	34,92	21,95	14,47	0,32	27,84	22,65	22,00	27,18	20,32
2015-10-01 C	18,75	90,01	0,41	1016,95	63,10	7,49	100,00	76,78	21,21	17,49	2,56	291,33	0,00	4,76	0,22	12,71	18,22	3,89	2,86	111,89	19,14	19,22	22,34	16,52
2015-10-08 C	19,40	95,76	0,52	1011,79	42,83	14,18	100,00	91,21	22,02	18,73	10,17	270,88	0,01	24,05	0,47	27,50	20,48	3,89	15,28	39,23	19,53	18,53	21,40	17,32
2015-10-15 C	23,07	86,92	0,14	1008,22	172,87	21,24	97,29	66,41	27,98	20,48	0,02	599,68	0,00	22,27	0,00	40,93	21,95	8,46	30,28	65,83	20,17	20,02	23,98	17,77
2015-10-29 C	21,32	88,70	0,24	1019,02	136,49	27,90	95,29	83,20	22,91	20,02	5,35	577,53	0,01	24,86	0,16	36,00	21,36	15,52	45,69	53,64	20,55	19,69	21,99	19,49
2015-11-19 C	24,42	88,16	0,47	1002,32	264,32	12,94	97,96	71,89	28,82	21,53	1,79	939,25	0,27	7,49	0,21	26,82	23,37	4,57	11,22	11,31	23,79	22,82	27,65	21,03
2015-11-26 C	21,88	93,33	0,20	1014,82	45,17	21,52	99,66	86,78	23,03	20,57	54,11	230,66	0,00	15,50	0,00	30,35	21,91	9,18	90,21	93,09	21,46	21,86	23,27	19,63
2015-12-03 C	21,28	95,92	0,25	1014,44	99,36	13,25	99,79	90,05	22,49	20,24	5,57	407,02	0,20	26,46	0,10	26,46	21,41	2,12	8,12	12,25	21,22	21,63	23,10	19,23
2015-12-17 C	23,61	86,44	0,34	1013,07	149,22	18,45	97,55	72,72	26,42	21,29	1,80	562,56	0,00	40,88	0,02	30,71	23,86	2,48	9,20	19,20	24,52	25,01	28,10	22,00
2016-01-14 C	24,62	90,23	0,63	1010,84	174,53	14,94	99,40	71,26	28,09	21,86	13,73	781,68	0,01	29,16	0,83	28,22	24,19	2,12	13,77	52,68	25,76	25,29	29,65	22,04
2016-01-27 C	22,45	81,85	0,39	1013,13	70,10	16,94	97,39	61,05	24,34	20,89	3,06	408,21	0,00	9,36	0,00	31,75	22,26	7,78	7,94	8,03	25,19	24,90	29,29	22,04
2016-02-11 C	24,98	81,79	0,12	1014,15	219,80	11,64	98,22	62,27	28,86	21,25	0,02	827,64	0,01	57,71	0,00	27,86	24,58	2,81	27,76	27,86	26,06	25,56	29,95	22,96
2016-02-25 C	27,59	80,29	0,09	1006,31	209,22	10,35	96,05	57,79	33,03	23,93	0,02	912,39	0,00	33,21	0,00	24,34	26,60	2,48	2,10	109,85	26,89	26,13	31,60	24,00
2016-03-17 C	25,56	84,37	0,35	1017,83	253,67	13,68	99,70	62,58	29,60	21,08	0,02	837,01	0,00	9,01	0,07	28,22	24,59	1,04	0,57	20,47	24,83	23,28	29,38	20,94
2016-03-31 C	24,67	90,83	0,56	1016,78	117,92	11,82	98,11	77,63	28,44	22,56	1,28	589,56	0,02	3,72	0,71	31,39	23,99	3,17	1,33	106,02	24,19	23,74	28,43	19,83
2016-04-19 C	27,19	86,18	0,28	1012,61	183,40	14,85	98,75	70,64	31,06	23,63	0,02	674,73	0,00	4,61	0,00	26,82	26,79	2,81	0,07	5,73	27,45	26,39	31,53	24,05
2016-04-28 C	12,53	72,74	0,41	1019,91	198,14	10,36	95,75	42,45	17,78	8,19	0,02	688,49	0,00	2,87	0,05	27,50	11,42	3,17	30,01	44,02	16,26	21,47	20,76	11,56
2016-05-12 C	18,84	76,61	0,05	1020,52	77,15	19,83	80,44	72,73	20,08	17,65	0,02	281,86	0,00	0,25	0,00	28,94	19,03	9,18	0,07	10,32	20,06	19,83	22,42	18,49
2016-05-24 C	14,90	75,59	0,15	1027,09	151,46	15,04	95,84	56,80	19,57	8,56	0,02	570,20	0,00	0,77	0,00	30,71	15,97	4,25	19,12	32,92	15,66	15,06	19,27	11,18
2016-06-08 C	10,59	80,78	0,68	1018,66	146,28	9,33	100,00	45,25	17,85	4,62	0,02	552,67	0,00	1,14	1,00	26,46	9,54	3,17	0,07	0,67	12,24	13,42	16,44	8,55
2016-06-30 C	15,54	95,23	0,38	1023,48	53,22	11,81	99,35	87,96	17,46	14,61	0,02	268,80	0,00	0,21	0,00	21,17	15,43	5,29	0,07	0,17	15,35	15,49	18,24	13,79
2016-07-28 C	13,05	76,35	0,65	1022,05	141,21	17,08	92,06	68,55	15,67	9,10	0,02	582,71	0,00	0,42	1,00	28,94	12,84	4,25	0,07	0,17	15,73	15,07	19,27	12,38
2016-08-11 C	15,64	70,25	0,07	1021,85	135,19	23,18	88,76	55,73	19,97	10,94	0,02	507,87	0,00	0,72	0,00	38,09	15,35	4,57	2,82	18,88	14,94	16,15	19,02	10,84
2016-08-25 C	19,40	78,71	0,01	1019,55	175,46	22,86	92,28	63,93	23,06	17,23	0,02	670,38	0,00	1,50	0,00	37,04	18,52	12,71	0,07	50,89	15,00	14,47	21,30	9,73
2016-09-08 C	16,10	76,11	0,38	1017,91	228,35	12,33	98,51	48,91	22,65	9,26	0,02	780,97	0,00	4,59	0,00	22,93	16,55	2,12	21,41	128,18	17,59	17,45	23,26	13,10
2016-09-22 C	17,67	78,60	0,67	1021,00	238,06	17,51	98,79	54,63	21,82	13,35	0,02	852,77	0,00	19,94	1,00	32,47	17,82	2,12	0,32	14,37	17,25	18,04	21,77	12,12
2016-10-13 C	21,25	91,41	0,12	1009,53	108,30	30,03	99,22	83,80	23,09	19,45	18,55	757,18	0,01	6,97	0,00	44,82	20,63	5,29	18,60	19,44	20,34	18,85	22,79	18,37
2016-11-10 C	24,03	76,49	0,26	1007,98	309,51	18,66	94,41	58,46	27,87	20,00	0,27	941,72	0,00	42,18	0,00	34,92	23,73	4,93	4,63	4,72	23,57	22,37	26,34	20,72
2016-11-24 C	19,65	72,29	0,28	1015,24	310,15	12,25	98,36	46,06	26,18	12,05	0,02	988,99	0,00	75,89	0,00	29,63	20,60	2,12	0,57	1,16	19,56	19,95	23,47	15,46
2017-01-12 C	24,92	77,18	0,15	1014,09	273,32	21,35	93,91	66,54	27,70	21,62	0,02	1014,30	0,00	97,62	0,00	34,24	24,58	4,57	2,10	18,92	25,60	25,78	30,07	22,14
2017-01-26 C	23,71	87,40	0,29	1010,81	258,58	18,26	100,00	67,60	28,20	19,70	24,15	990,23	0,20	8,93	0,00	34,56	22,29	4,57	40,94	41,55	25,06	24,81	30,42	21,21
2017-02-09 C	24,87	81,55	0,16	1014,36	199,00	14,79	92,74	66,51	28,53	22,41	4,58	910,57	0,00	24,59	0,00	33,52	24,68	2,48	4,63	9,03	24,46	25,70	27,18	22,47
2017-02-23 C	25,77	83,93	0,70	1011,00	158,91	11,67	94,70	60,15	32,06	22,99	0,77	748,94	0,00	14,33	1,00	24,70	24,98	4,25	0,82	0,91	26,82	27,25	31,03	24,30
2017-03-08 C	24,25	90,07	0,16	1014,74	140,63	11,97	99,87	77,20	27,85	20,64	12,70	555,69	0,00	7,57	0,02	22,57	24,72	1,76	19,61	74,05	24,80	26,21	28,74	21,57
2017-04-05 C	23,83	90,52	0,57	1010,18	61,43	17,64	99,06	83,01	26,12	22,22	3,05	386,48	0,00	2,61	0,91	29,63	23,75	6,70	25,68	64,33	22,72	22,02	25,46	20,40
2017-04-19 C	20,80	87,41	0,11	1018,71	102,76	11,45	100,00	66,11	23,95	18,60	8,90	342,41	0,00	2,87	0,00	22,57	20,50	3,89	17,30	29,55	20,60	21,42	23,77	18,67
2017-05-03 C	20,43	88,63	0,54	1019,08	142,41	9,99	99,99	69,75	25,25	16,79	0,02	651,02	0,00	1,36	0,84	24,34	20,01	2,12	0,07	0,17	19,56	17,48	24,38	15,55
2017-05-17 C	18,23	88,72	0,59	1017,76	132,23	9,45	100,00	64,76	23,79	13,54														

Data	TEMP AR MEDIO	UMIDADE RLTV	MOLHAMENTO FOLIAR MEDIO	PRESSAO ATM	RADIACAO SOLAR MEDIA	VEL MAX VENTO MEDIA	UMIDADE AR MAX	UMIDADE AR MIN	TEMP MAX	TEMP MIN	PRECP	RADIACAO SOLAR MAX	RADIACAO SOLAR MIN	RADIACAO SOLAR MEDIANA	MOLHAMENTO FOLIAR MEDIANO	VEL VENTO MAXIMA MAXIMA	TEMP MEDIANA	VEL VENTO MAXIMA MIN	PRECP ACUM 3d	PRECP acum 7d	TEMP AR 3d	TEMP AR 7d	TEMP AR MAX 3d	TEMP AR MIN 3d
2018-03-21 C	23,11	75,89	0,39	1012,96	245,37	16,68	93,70	54,87	26,61	20,53	0,02	914,00	0,00	9,43	0,00	33,70	22,77	5,29	4,32	4,66	23,69	24,46	27,26	20,96
2018-04-11 C	24,41	81,66	0,40	1016,79	212,35	14,42	98,70	51,18	29,17	20,79	0,27	811,00	0,00	6,94	0,00	29,99	23,81	4,57	0,32	5,91	24,24	23,48	28,96	20,36
2018-04-25 C	23,42	83,06	0,62	1018,27	206,49	9,36	99,70	54,63	29,75	19,66	0,02	743,90	0,00	1,77	1,00	20,30	21,36	1,22	0,07	0,17	23,30	23,12	29,04	19,04
2018-05-09 C	20,12	96,36	0,99	1020,41	54,80	7,59	100,00	90,20	22,03	17,85	2,27	224,50	0,00	1,47	1,00	13,39	20,49	2,66	12,06	19,90	21,03	22,49	23,87	18,35
2018-05-23 C	15,83	77,50	0,60	1021,92	179,18	10,38	96,90	47,78	20,11	11,64	0,02	673,10	0,00	0,13	0,97	26,46	14,72	4,07	0,07	34,15	15,23	17,44	21,25	10,08
2018-06-27 C	17,14	83,76	1,00	1016,90	71,49	8,45	96,80	65,69	20,24	15,40	0,02	309,60	0,00	0,00	1,00	14,29	16,76	1,76	9,46	13,75	17,14	17,58	20,50	14,86
2018-07-11 C	15,30	83,76	0,38	1024,19	163,86	18,30	96,80	65,69	21,03	8,65	0,02	654,50	0,00	0,00	0,01	29,12	14,93	3,71	0,07	0,17	16,09	16,72	20,78	11,81
2018-08-01 C	14,75	83,76	0,54	1022,57	153,85	17,45	96,80	65,69	18,40	12,21	0,02	686,40	0,00	0,00	0,75	27,68	14,24	5,47	0,07	0,17	15,66	17,21	19,07	13,57
2018-08-15 C	17,41	83,76	0,14	1021,50	112,72	22,77	96,80	65,69	20,38	12,50	0,02	472,00	0,00	0,60	0,00	32,80	17,18	4,57	0,07	0,17	17,14	16,36	22,60	11,74
2018-08-29 C	19,56	83,76	0,00	1020,90	132,75	21,71	96,80	65,69	23,50	17,76	0,02	596,10	0,00	2,53	0,00	33,88	18,59	12,71	0,07	13,56	15,73	15,82	22,02	10,16
2018-09-12 C	17,48	83,76	0,99	1021,27	88,72	15,87	96,80	65,69	18,88	15,71	7,01	393,30	0,00	4,78	1,00	34,74	17,73	4,25	31,45	31,54	17,72	17,83	20,04	15,20
2018-09-26 C	22,86	83,76	0,77	1011,95	155,83	11,35	96,80	65,69	27,31	20,44	0,02	818,00	0,00	6,45	1,00	33,29	22,12	2,66	0,07	0,17	22,89	22,66	27,51	19,93
2018-10-10 C	22,86	83,76	0,39	1012,68	200,45	24,77	96,80	65,69	27,04	20,59	0,42	854,00	0,00	10,55	0,00	39,71	22,28	6,88	3,47	10,76	22,74	20,27	26,68	20,28
2018-10-31 C	21,26	83,76	0,46	1012,13	56,12	18,18	96,80	65,69	23,40	19,71	19,42	422,80	0,00	8,03	0,19	33,88	21,48	4,93	19,47	66,92	21,03	21,11	23,94	18,48
2018-11-14 C	27,18	83,76	0,24	1005,90	225,40	17,14	96,80	65,69	31,77	23,49	0,02	889,00	0,00	20,55	0,00	25,74	27,13	7,06	2,27	20,74	25,93	24,03	29,68	23,05
2018-11-28 C	23,04	83,76	0,90	1015,95	113,79	14,62	96,80	65,69	25,02	20,81	0,02	369,80	0,00	30,76	1,00	26,10	23,20	5,29	0,47	14,76	22,40	23,05	25,06	19,28
2018-12-12 C	27,51	83,76	0,71	1013,10	291,91	17,92	96,80	65,69	35,13	21,15	0,02	1053,00	0,00	65,13	1,00	36,00	26,77	4,93	0,07	0,17	26,11	23,23	31,89	21,58
2019-03-07 C	26,67	83,76	0,56	1012,03	149,09	15,34	96,80	65,69	29,68	23,58	11,82	586,90	0,00	18,23	0,88	26,82	26,97	2,12	13,67	14,16	27,25	26,56	30,92	23,87
2019-03-20 C	23,60	83,76	0,76	1011,61	146,19	24,72	96,80	65,69	26,64	21,54	1,82	684,50	0,00	6,36	1,00	38,81	23,17	8,10	3,27	4,56	25,16	24,91	28,90	22,77
2019-04-03 C	25,93	83,76	0,71	1009,42	223,41	12,79	96,80	65,69	30,98	22,06	0,02	852,00	0,00	5,81	1,00	18,00	25,16	5,47	0,07	28,15	24,79	23,45	29,63	20,88
2019-04-17 C	22,35	83,76	0,07	1010,54	184,38	15,49	96,80	65,69	27,22	17,59	0,22	797,00	0,00	2,22	0,00	31,75	22,76	4,43	7,86	7,96	23,49	23,73	27,60	19,66
2019-05-09 C	21,64	83,76	0,84	1016,15	121,54	10,32	96,80	65,69	24,65	19,48	4,02	530,60	0,00	1,25	1,00	20,99	21,53	2,48	4,07	5,76	21,69	23,67	24,23	19,55
2019-05-22 C	21,13	83,76	0,83	1012,82	52,55	7,91	96,80	65,69	24,58	19,69	1,62	288,30	0,00	0,00	1,00	21,53	20,33	4,07	1,67	21,94	21,14	20,75	25,93	18,06
2019-06-12 C	19,62	83,76	0,91	1017,84	137,98	11,85	96,80	65,69	24,60	15,57	0,02	594,80	0,00	0,21	1,00	26,10	19,47	2,81	0,27	0,96	19,99	19,72	25,13	15,84
2019-06-26 C	14,88	83,76	1,00	1014,26	19,61	21,83	96,80	65,69	19,52	12,37	12,21	90,40	0,00	1,53	1,00	37,94	13,50	3,71	12,46	12,76	18,91	18,70	25,81	13,68
2019-07-10 C	16,34	83,76	0,69	1019,42	151,18	13,32	96,80	65,69	22,62	10,22	0,02	621,10	0,00	0,00	1,00	30,89	15,77	1,76	0,27	0,37	14,33	12,71	21,79	7,87
2019-07-24 C	19,06	83,76	0,48	1017,98	149,89	15,56	96,80	65,69	22,72	15,37	0,02	601,90	0,00	0,12	0,41	26,82	19,80	3,17	0,07	6,36	20,10	18,99	25,18	16,20
2019-08-21 C	14,48	83,76	0,03	1024,47	147,41	19,16	96,80	65,69	16,90	12,90	0,02	814,00	0,00	1,63	0,00	31,03	14,19	7,06	3,07	14,56	14,09	16,30	16,11	12,56
2019-10-02 C	21,90	82,99	0,68	1010,74	239,04	23,79	95,10	64,08	25,50	19,77	0,02	952,00	0,00	14,92	1,00	37,22	20,88	4,43	1,87	1,96	21,26	20,09	24,12	19,44
2019-10-23 C	20,46	82,81	0,64	1015,63	207,02	14,73	97,80	54,27	24,10	17,41	1,02	649,20	0,00	40,24	1,00	26,28	20,53	4,93	3,27	18,35	19,83	20,92	22,75	16,92
2019-11-06 C	21,89	69,06	0,00	1016,28	293,04	21,83	79,00	55,67	24,60	20,18	0,02	948,00	0,00	69,91	0,00	30,35	21,23	6,70	27,86	41,74	22,61	23,33	25,26	20,25
2019-11-20 C	24,76	76,21	0,00	1015,86	330,40	23,39	87,49	60,37	27,76	22,30	0,02	1080,00	0,00	72,50	0,00	32,11	24,16	15,34	0,07	15,15	22,95	22,30	27,57	17,89
2019-12-04 C	22,43	77,62	0,35	1011,39	90,22	20,84	87,84	68,24	23,94	20,95	0,02	297,60	0,00	61,58	0,00	30,53	22,34	12,53	0,07	1,17	21,50	22,33	24,21	17,66
2020-01-29 C	24,91	92,65	0,77	1010,17	84,90	9,75	99,22	73,42	28,45	22,97	7,62	403,80	0,00	23,50	1,00	21,71	24,52	3,17	13,86	41,55	25,38	24,26	29,42	22,70
2020-02-27 C	21,66	67,00	0,20	1013,13	322,89	18,09	99,20	45,44	26,62	15,01	0,02	1066,00	0,00	31,02	0,00	30,89	21,72	3,17	29,66	30,16	23,16	22,67	26,70	19,83
2020-03-12 C	25,78	81,00	0,53	1014,20	284,66	19,42	96,69	58,29	29,81	23,10	0,02	944,00	0,00	16,29	0,72	32,62	24,79	7,42	0,07	1,57	25,42	24,45	30,31	21,73
2020-04-30 C	22,33	80,30	0,37	1014,42	118,05	15,08	95,33	67,87	24,73	20,67	0,01	516,90	0,00	1,26	0,00	28,58	21,76	4,75	0,03	0,12	23,38	23,17	26,79	20,83
2020-05-13 C	21,02	93,07	0,77	1010,06	70,35	10,61	99,83	76,59	24,37	18,63	12,62	408,30	0,00	0,05	1,00	20,30	20,31	3,35	12,66	12,76	21,16	19,91	26,36	17,18
2020-05-27 C	13,72	85,16	0,63	1016,13	169,93	8,80	99,71	56,91	19,66	8,06	0,02	670,60	0,00	0,13	1,00	21,71	12,79	2,12	0,07	1,17	16,01	19,11	22,05	10,54
2020-06-17 C	21,32	88,08	0,17	1021,22	92,27	29,63	94,50	77,98	22,84	20,45	0,02	436,40	0,00	2,40	0,00	39,17	21,02	20,99	6,06	6,55	18,90	19,26	20,96	17,23
2020-07-01 C	17,72	58,99	0,00	1005,11	162,25	29,84	86,19	43,56	21,61	11,39	0,02	633,40	0,00	0,00	0,00	52,38	18,15	5,11	18,06	30,55	17,47	16,64	20,68	13,15
2020-07-15 C	12,06	84,05	0,06	1023,63	168,08	9,41	100,00	62,81	16,39	7,91	0,02	669,80	0,00	0,00	0,00	24,88	11,10	1,76	4,27	29,35	13,95	15,26	16,87	11,12

Data	RAD SOLAR MEDIA 3d	VEL MAX VENTO MEDIA 3d	EFL DQO	AFL SS	AFL ST	AFL_STV	AFL_STF	AFL pH	AFL ALCALINIDADE	AFL DQO	AFL DBO	DQO/DBO	AFL N-NH4+	AFL N-NO2-	AFL N-NO3-	AFL P-PO43-	Fluxo (L/m2/d)	Carr Org Sup (g DQO/m2/d)
2015-06-25 C	81,02	15,94	183,00	66,00	548,33	306,63	241,70	7,32	340,00	782,00	58,00	13,48	63,15	0,10	4,70	15,70	79,71	62331,13
2015-07-02 C	92,33	12,35	207,00	81,33	581,67	215,00	366,67	7,31	310,00	753,00	268,03	2,81	74,70	0,00	6,50	28,20	77,95	58696,45
2015-07-23 C	105,04	15,40	120,00	53,00	480,00	360,00	120,00	7,37	320,00	540,00	268,03	2,01	73,15	0,00	5,90	25,80	79,53	42943,75
2015-08-06 C	120,42	14,00	161,00	45,50	566,67	86,67	480,00	7,09	315,00	726,00	268,03	2,71	53,40	0,16	13,70	35,40	80,79	58653,15
2015-08-13 C	149,89	24,56	137,00	55,00	561,67	283,33	278,33	7,10	290,00	750,00	309,00	2,43	81,60	0,18	0,39	34,11	78,15	58612,40
2015-08-20 C	131,16	20,10	81,00	29,50	781,67	260,00	521,67	7,14	245,00	732,00	293,00	2,50	79,90	0,14	0,89	28,22	77,16	56482,10
2015-08-27 C	73,58	16,73	98,00	43,50	323,33	131,67	191,67	7,12	325,00	564,00	280,00	2,01	68,87	0,00	0,23	25,07	79,66	44925,53
2015-09-03 C	51,92	15,22	90,00	49,00	501,67	251,67	250,00	7,15	275,00	773,00	319,00	2,42	79,80	0,15	0,71	30,79	79,09	61134,41
2015-09-10 C	129,68	18,51	58,00	51,50	455,00	190,00	265,00	7,20	325,00	630,00	254,00	2,48	61,20	0,15	1,02	28,39	79,27	49938,25
2015-09-24 C	167,01	19,63	43,00	49,67	0,00	0,00	0,00	7,05	260,00	444,00	296,00	1,50	72,15	0,18	0,00	34,76	79,11	35126,08
2015-10-01 C	120,56	13,21	36,57	22,50	421,67	86,67	335,00	7,05	312,50	356,00	212,00	1,68	54,22	0,14	0,00	26,44	78,45	27928,72
2015-10-08 C	67,80	19,23	18,60	48,00	511,04	254,40	251,37	7,06	274,54	443,00	267,00	1,66	72,90	0,18	0,00	33,32	77,90	34509,88
2015-10-15 C	130,98	17,75	13,90	79,68	511,04	254,40	251,37	7,23	195,00	479,00	238,00	2,01	64,93	0,14	1,18	36,12	77,78	37254,83
2015-10-29 C	91,13	18,90	40,00	27,33	288,33	76,67	211,67	7,12	330,00	217,00	261,00	0,83	37,40	0,15	0,00	25,03	78,05	16936,82
2015-11-19 C	205,19	16,72	30,70	61,16	561,67	246,67	315,00	7,35	310,00	631,00	286,00	2,21	72,95	0,19	0,00	35,48	78,05	49249,47
2015-11-26 C	83,14	20,05	33,40	29,67	458,33	180,00	278,33	7,23	325,00	294,00	190,00	1,55	27,49	0,17	0,00	33,05	78,05	22946,66
2015-12-03 C	97,75	11,48	21,80	31,50	443,33	236,67	206,67	7,21	250,00	631,00	254,00	2,48	69,88	0,18	0,00	31,66	76,68	48382,72
2015-12-17 C	212,91	20,06	26,60	34,00	496,67	215,00	281,67	7,08	260,00	732,00	322,00	2,27	71,45	0,20	0,00	35,42	77,19	56499,96
2016-01-14 C	220,38	13,25	20,70	21,50	413,33	226,67	186,67	7,10	285,00	577,00	341,00	1,69	72,30	0,14	0,00	26,26	77,19	44536,17
2016-01-27 C	202,49	21,35	21,50	23,00	511,67	198,34	313,33	7,21	285,00	675,00	302,00	2,24	83,63	0,16	0,54	35,17	77,19	52100,37
2016-02-11 C	177,09	15,28	29,50	16,00	516,67	236,67	280,00	7,15	225,00	568,00	270,00	2,10	81,33	0,14	0,20	29,07	77,19	43841,50
2016-02-25 C	195,92	13,21	19,00	15,00	375,00	170,00	205,00	7,12	260,00	524,00	331,00	1,58	62,35	0,13	0,00	23,18	77,19	40445,32
2016-03-17 C	249,11	12,38	16,90	34,00	410,00	208,33	201,67	7,44	240,00	588,00	257,00	2,29	73,85	0,17	0,50	27,42	76,01	44692,39
2016-03-31 C	175,34	11,91	14,10	26,67	576,67	363,33	213,33	7,25	260,00	511,00	264,00	1,94	69,50	0,15	0,23	25,50	76,01	38839,82
2016-04-19 C	187,08	15,13	29,00	61,53	533,33	280,00	253,33	7,30	245,00	418,00	418,00	1,00	89,07	0,18	0,30	33,48	79,37	33176,77
2016-04-28 C	141,50	17,03	19,00	37,33	320,00	176,67	143,33	7,41	270,00	381,00	254,00	1,50	78,55	0,20	0,26	32,08	75,56	28788,46
2016-05-12 C	91,71	17,76	50,00	46,00	518,33	311,67	206,67	7,43	285,00	439,00	268,03	1,64	87,10	0,20	0,59	35,87	75,16	32997,23
2016-05-24 C	108,56	13,96	31,00	52,00	535,00	305,00	230,00	7,32	280,00	468,00	261,00	1,79	85,83	0,21	0,00	37,66	77,23	36145,76
2016-06-08 C	101,85	11,66	30,00	77,00	586,67	246,67	340,00	7,62	260,00	523,00	209,00	2,50	90,87	0,25	0,00	37,91	75,89	39690,12
2016-06-30 C	77,00	10,40	39,00	69,50	303,33	268,33	350,00	7,54	260,00	482,00	270,00	1,79	91,60	0,19	0,00	33,86	77,06	37144,78
2016-07-28 C	137,75	18,75	30,50	51,00	561,67	268,33	293,33	7,55	265,00	721,00	232,00	3,11	77,00	0,26	0,00	40,00	76,39	55075,94
2016-08-11 C	118,50	16,80	39,00	88,00	695,00	318,33	376,67	7,10	275,00	546,00	232,00	2,35	90,40	0,25	0,00	41,51	74,39	40614,39
2016-08-25 C	191,28	15,55	36,00	27,00	1088,33	275,00	813,33	7,13	250,00	431,00	199,00	2,17	92,75	0,22	0,00	34,73	74,39	32060,08
2016-09-08 C	211,88	19,38	34,00	79,68	511,04	254,40	251,37	7,23	274,54	420,00	268,03	1,57	64,93	0,14	1,18	36,12	75,26	31608,02
2016-09-22 C	230,06	16,95	28,00	55,56	236,67	246,67	335,00	7,20	250,00	338,00	196,00	1,72	90,30	0,16	0,00	30,86	75,26	25436,93
2016-10-13 C	125,72	22,96	33,00	57,22	436,67	298,33	138,33	7,59	255,00	459,00	254,00	1,81	78,95	0,18	0,00	33,83	74,82	34341,83
2016-11-10 C	226,32	17,84	34,82	56,50	398,33	65,00	333,33	7,64	330,00	428,73	348,00	1,23	96,50	0,19	0,00	39,03	72,98	31289,89
2016-11-24 C	183,72	14,95	28,22	51,00	548,33	338,33	210,00	7,31	310,00	467,76	273,00	1,71	97,90	0,21	0,00	39,80	72,98	34138,44
2017-01-12 C	228,59	17,91	42,86	37,50	516,67	170,00	346,67	7,66	275,00	218,29	322,00	0,68	86,00	0,10	0,00	24,38	74,86	16341,83
2017-01-26 C	233,49	14,51	24,12	86,00	660,00	293,33	366,67	7,75	275,00	357,31	312,00	1,15	69,25	0,14	0,00	25,80	72,98	26077,41
2017-02-09 C	162,69	14,58	26,05	45,67	558,33	291,67	266,67	7,75	225,00	379,30	315,00	1,20	82,45	0,16	0,00	34,64	74,70	28334,95
2017-02-23 C	231,07	16,79	25,58	60,86	511,67	323,33	188,33	7,04	255,00	469,80	351,00	1,34	72,10	0,16	0,00	31,24	74,70	35096,06
2017-03-08 C	178,78	12,96	19,66	23,89	428,33	190,00	238,33	7,37	274,54	336,70	322,00	1,05	77,50	0,16	0,00	29,71	75,19	25315,33
2017-04-05 C	97,67	14,58	21,34	36,00	291,67	261,67	300,00	7,23	205,00	317,84	254,00	1,25	51,75	0,12	1,18	20,65	102,66	32629,75
2017-04-19 C	101,43	11,46	76,19	41,67	351,67	136,67	215,00	6,99	274,54	368,90	238,00	1,55	53,75	0,11	1,00	21,10	102,26	37722,29
2017-05-03 C	118,09	8,88	27,89	72,22	490,00	300,00	190,00	7,47	274,54	389,54	261,00	1,49	78,10	0,17	1,80	31,24	73,93	28800,46
2017-05-17 C	150,50	11,06	37,91	143,33	511,04	254,40	251,37	7,23	245,00	468,19	296,00	1,58	75,10	0,09	0,70	20,28	102,01	47758,93
2017-06-14 C	102,97	8,09	27,84	94,29	440,00	170,00	270,00	7,38	280,00	341,17	222,00	1,54	75,65	0,12	0,90	21,65	101,36	34582,04
2017-06-29 C	124,79	11,54	41,45	70,71	431,67	223,33	208,33	7,80	305,00	425,26	177,00	2,40	91,20	0,15	0,00	32,77	101,04	42969,69
2017-07-13 C	112,50	12,21	39,92	129,64	495,00	188,33	306,67	6,74	320,00	567,08	268,03	2,12	89,25	0,15	0,80	31,84	100,42	56947,53
2017-08-02 C	147,28	11,01	34,50	285,00	775,00	441,67	333,33	7,40	300,00	744,95	335,00	2,22	89,00	0,00	3,00	31,38	104,00	77471,72
2017-08-23 C	142,03	19,20	27,70	90,63	568,33	300,00	268,33	8,04	270,00	403,68	203,00	1,99	76,75	0,14	2,70	29,00	102,74	41474,22
2017-09-13 C	166,63	17,99	22,44	76,50	393,33	223,33	170,00	7,47	274,54	360,49	302,00	1,19	82,10	0,13	0,90	28,47	98,98	35682,17
2017-09-27 C	166,63	16,98	21,20	62,00	426,67	243,34	183,33	7,23	274,54	427,60	302,00	1,42	73,65	0,17	2,00	32,29	100,10	42801,05
2017-10-18 C	166,63	23,50	19,67	61,50	366,67	150,00	216,67	7,23	274,54	315,87	251,00	1,26	70,10	0,13	0,45	25,61	98,68	31169,25
2017-11-08 C	233,15	21,94	11,73	72,67	410,00	168,33	241,67	7,23	274,54	314,49	254,00	1,24	76,95	0,12	1,40	25,28	102,33	32182,68
2017-11-22 C	271,47	17,11	21,11	66,00	411,67	233,34	178,33	7,23	274,54	332,99	25,00	13,32	79,20	0,12	1,90	26,18	102,33	34075,86
2017-12-13 C	334,03	23,68	52,15	71,00	535,00	281,67	253,33	7,23	274,54	533,20	344,00	1,55	76,20	0,12	2,30	45,30	98,98	52777,42
2018-01-31 C	310,85	22,08	49,61	59,83	486,67	263,34	223,33											

Data	RAD SOLAR MEDIA 3d	VEL MAX VENTO MEDIA 3d	EFL DQO	AFL SS	AFL ST	AFL_STV	AFL_STF	AFL pH	AFL ALCALINIDADE	AFL DQO	AFL DBO	DQO/DBO	AFL N- NH4+	AFL N-NO2-	AFL N- NO3-	AFL P- PO43-	Fluxo (L/m2/d)	Carr Org Sup (g DQO/m2/d)
2018-03-21 C	166,95	17,08	32,29	68,33	516,67	236,67	280,00	7,16	250,00	363,66	331,00	1,10	60,50	0,14	0,25	42,75	97,79	35560,32
2018-04-11 C	217,50	12,51	26,65	58,00	511,04	254,40	251,37	7,33	250,00	363,15	393,00	0,92	60,00	0,13	0,25	42,35	96,22	34943,41
2018-04-25 C	209,52	10,91	22,90	69,67	383,33	193,33	190,00	6,47	285,00	376,45	409,00	0,92	59,00	0,16	0,60	47,80	98,06	36915,14
2018-05-09 C	117,53	9,81	19,62	56,33	528,33	300,00	228,33	7,18	275,00	422,27	325,00	1,30	73,00	0,12	1,25	42,70	95,71	40413,07
2018-05-23 C	174,23	8,90	24,51	53,00	465,00	223,33	241,67	7,09	274,54	417,17	277,00	1,51	68,00	0,17	0,70	50,55	94,35	39360,94
2018-06-27 C	71,31	12,08	42,33	69,67	485,00	231,67	253,33	7,23	274,54	332,31	196,00	1,70	62,00	0,14	1,35	44,80	106,08	35250,12
2018-07-11 C	130,83	19,21	29,48	68,00	546,67	283,34	263,33	7,23	265,00	451,18	235,00	1,92	66,50	0,13	2,10	44,05	105,38	47544,51
2018-08-01 C	131,92	17,82	26,69	52,67	745,00	320,00	425,00	6,90	65,00	348,55	17,70	19,69	68,50	0,14	3,10	41,65	100,32	34967,14
2018-08-15 C	146,61	14,88	35,52	42,67	500,00	171,67	328,33	7,46	335,00	361,42	19,90	18,16	104,50	0,17	1,00	46,16	97,91	35384,46
2018-08-29 C	180,94	16,79	18,97	187,66	706,67	361,67	345,00	6,99	255,00	584,93	20,60	28,39	63,00	0,12	2,35	39,30	108,62	63536,81
2018-09-12 C	94,70	14,61	20,15	193,00	690,00	420,00	270,00	6,82	227,50	494,93	275,00	1,80	50,50	0,11	0,05	30,30	107,33	53119,45
2018-09-26 C	175,32	13,74	16,28	51,67	390,00	321,67	68,33	6,97	245,00	245,88	248,00	0,99	59,50	0,08	0,50	29,60	107,91	26531,32
2018-10-10 C	177,55	16,27	14,94	62,67	478,33	243,33	235,00	7,19	300,00	326,29	299,00	1,09	58,50	0,09	0,75	34,10	102,98	33600,66
2018-10-31 C	141,68	17,11	21,70	1466,67	1418,33	986,67	431,67	7,37	240,00	1082,05	377,00	2,87	61,50	0,10	1,15	30,50	105,21	113844,50
2018-11-14 C	269,85	19,31	19,14	59,33	341,67	180,00	161,67	6,93	315,00	316,37	290,00	1,09	60,50	0,11	1,30	38,25	105,49	33373,58
2018-11-28 C	191,76	15,05	25,29	66,67	436,67	203,33	233,33	6,85	275,00	489,46	348,00	1,41	68,00	0,16	0,45	49,95	105,34	51557,49
2018-12-12 C	326,74	18,55	24,96	56,67	483,33	303,33	180,00	7,15	180,00	355,62	312,00	1,14	71,00	0,11	0,70	41,40	105,89	37657,93
2019-03-07 C	233,58	14,75	11,09	79,68	511,04	254,40	251,37	7,23	274,54	420,00	268,03	1,57	64,93	0,14	1,18	36,12	105,89	44475,37
2019-03-20 C	174,09	17,20	13,14	52,33	405,00	245,00	160,00	6,60	300,00	368,01	319,00	1,15	54,00	0,13	1,18	37,65	104,35	38401,16
2019-04-03 C	236,13	12,64	19,68	74,33	696,67	385,00	311,67	6,91	305,00	515,44	399,00	1,29	70,50	0,14	1,05	47,70	100,05	51567,71
2019-04-17 C	170,61	14,23	17,23	53,18	478,33	306,67	171,67	6,18	280,00	401,91	319,00	1,26	71,00	0,13	1,20	47,75	100,05	40209,49
2019-05-09 C	112,69	13,40	31,07	63,67	463,33	243,33	220,00	7,10	275,00	380,33	319,00	1,19	70,50	0,13	1,45	39,80	100,08	38064,24
2019-05-22 C	110,97	8,41	12,11	67,00	550,00	258,33	291,67	7,30	260,00	426,13	335,00	1,27	66,50	0,15	1,10	44,10	100,08	42648,57
2019-06-12 C	143,52	10,84	39,20	211,81	530,00	325,00	205,00	7,70	240,00	524,07	302,00	1,74	37,40	0,10	1,80	38,25	101,56	53221,88
2019-06-26 C	100,34	13,47	22,74	72,33	436,67	211,67	225,00	7,40	280,00	373,55	241,00	1,55	31,45	0,11	0,55	36,75	99,46	37153,98
2019-07-10 C	153,72	11,47	29,10	437,00	876,67	600,00	276,67	7,20	260,00	961,72	257,00	3,74	39,05	0,08	1,50	33,00	101,58	97693,06
2019-07-24 C	126,81	14,62	25,85	69,67	511,04	246,67	213,33	7,20	215,00	400,97	286,00	1,40	35,45	0,16	1,15	52,05	102,20	40980,23
2019-08-21 C	89,00	18,37	35,34	47,67	406,67	186,67	220,00	7,80	315,00	276,80	180,00	1,54	33,55	0,08	1,00	34,60	98,57	27283,66
2019-10-02 C	193,77	27,18	61,21	92,92	511,04	254,40	251,37	7,30	300,00	747,80	547,00	1,37	48,55	0,18	2,95	71,00	105,92	79207,74
2019-10-23 C	189,61	15,79	16,40	60,33	565,00	331,67	233,33	7,70	320,00	455,10	341,00	1,33	40,20	0,14	1,45	52,00	105,92	48204,98
2019-11-06 C	170,88	18,84	12,35	49,67	571,67	340,00	231,67	6,96	320,00	358,40	273,00	1,31	41,35	0,13	2,30	48,00	107,03	38357,54
2019-11-20 C	330,44	18,51	15,61	208,00	1513,33	1115,00	398,33	7,00	345,00	1035,05	600,00	1,73	36,35	0,09	1,75	35,85	109,13	112951,00
2019-12-04 C	265,26	20,40	16,29	44,00	538,33	288,33	250,00	6,90	285,00	379,25	283,00	1,34	40,60	0,15	0,75	56,40	108,77	41250,63
2020-01-29 C	203,62	12,74	33,54	49,33	571,67	360,00	211,67	7,00	205,00	413,40	190,00	2,18	34,20	0,13	5,45	45,65	107,03	44244,49
2020-02-27 C	188,23	19,73	57,45	38,67	365,00	186,67	178,33	7,20	280,00	293,50	199,00	1,47	24,40	0,13	2,55	43,25	110,01	32287,27
2020-03-12 C	287,25	14,97	29,55	49,33	486,67	270,00	216,67	7,10	310,00	398,65	196,00	2,03	34,25	0,13	1,40	48,14	110,01	43854,58
2020-04-30 C	134,53	15,81	16,70	33,67	418,33	196,66	221,67	7,00	315,00	287,50	264,00	1,09	35,75	0,13	1,15	52,00	109,22	31399,45
2020-05-13 C	125,10	12,15	9,60	35,33	511,67	246,67	265,00	7,00	320,00	304,50	268,03	1,14	36,05	0,11	1,05	48,70	108,15	32932,73
2020-05-27 C	167,51	12,23	16,35	34,33	600,00	176,67	423,33	7,00	320,00	333,85	268,03	1,25	39,50	0,14	2,65	56,65	108,15	36107,03
2020-06-17 C	69,64	16,19	11,55	23,67	485,00	258,33	226,67	7,40	300,00	254,40	164,00	1,55	42,00	0,14	1,65	51,00	111,01	28240,81
2020-07-01 C	98,73	23,73	16,10	31,67	498,33	205,00	293,33	7,50	270,00	282,05	112,00	2,52	34,00	0,14	1,55	53,25	113,69	32065,02
2020-07-15 C	118,64	13,75	6,35	6,67	403,33	191,66	211,67	7,60	274,54	200,50	22,00	9,11	28,50	0,10	1,25	43,70	114,87	23031,27

APÊNDICE B – ANÁLISE ESTATÍSTICA DO CONJUNTO DE DADOS.

Parâmetro	Média	Desvio Padrão	Mínimo	25 %	50 %	75%	Máximo	Variância	Skewness	Kurtosis	CV (%)
TEMP AR MEDIO	20.69	3.80	10.59	17.96	21.28	23.48	27.59	14.46	-0.41	-0.49	18.38
UMIDADE RLV	83.70	6.40	58.99	81.27	83.77	86.93	96.36	40.98	-0.64	1.59	7.65
MOLHAMENTO FOLIAR MEDIO	0.42	0.28	0.00	0.16	0.40	0.61	1.00	0.08	0.27	-0.80	66.37
PRESSAO ATM MED	1015.73	5.15	1002.32	1012.08	1015.24	1019.30	1031.19	26.50	0.16	0.00	0.51
RADIACAO SOLAR MEDIA	158.58	75.81	19.61	106.91	149.22	203.47	366.20	5747.50	0.54	-0.20	47.81
VEL MAX VENTO MEDIA	16.39	5.85	6.07	11.74	15.80	21.10	30.03	34.18	0.39	-0.56	35.67
UMIDADE AR MAX	96.62	3.78	79.00	96.27	96.80	99.29	100.00	14.30	-2.31	6.76	3.91
UMIDADE AR MIN	66.02	10.39	42.45	61.01	65.69	70.62	91.21	108.01	0.15	0.36	15.74
TEMP MAX	24.42	3.99	15.67	21.72	24.51	27.37	35.13	15.93	0.04	-0.41	16.34
TEMP MIN	17.56	4.48	4.62	13.96	19.45	20.88	23.93	20.09	-0.75	-0.37	25.52
PRECP	3.20	7.43	0.00	0.02	0.02	2.04	54.11	55.24	3.98	20.86	232.51
RADIACAO SOLAR MAX	619.97	238.15	90.40	424.63	601.90	804.00	1132.00	56716.08	0.08	-0.70	38.41
RADIACAO SOLAR MIN	3.58	36.08	0.00	0.00	0.00	0.00	366.20	1301.81	10.00	98.01	1008.48
RADIACAO SOLAR MEDIANA	16.95	40.11	0.00	1.17	5.81	17.26	366.20	1608.88	6.70	54.26	236.62
MOLHAMENTO FOLIAR MEDIANO	0.39	0.45	0.00	0.00	0.05	1.00	1.00	0.20	0.44	-1.69	114.77
VEL VENTO MAXIMA MAXIMA	29.54	7.30	12.71	25.31	29.99	34.24	52.38	53.21	-0.07	0.39	24.70
TEMP MEDIANA	20.36	3.78	9.54	17.84	20.89	23.14	27.13	14.28	-0.56	-0.10	18.56
VEL VENTO MAXIMA MIN	5.34	4.67	0.00	2.66	4.25	5.38	29.27	21.76	2.50	7.45	87.32
PRECP ACUM 3d	8.56	13.93	0.03	0.07	2.10	12.56	90.21	194.12	2.77	10.70	162.76
PRECP acum 7d	23.30	28.09	0.12	2.33	13.75	31.05	128.18	788.87	1.76	2.83	120.54
TEMP AR 3d	20.73	3.66	12.24	17.86	21.03	23.63	27.45	13.41	-0.18	-0.92	17.67
TEMP AR 7d	20.73	3.34	12.71	18.18	20.92	23.39	27.25	11.17	-0.18	-0.78	16.12
TEMP AR MAX 3d	24.66	3.76	16.11	22.01	24.23	27.59	31.89	14.13	-0.06	-0.75	15.24
TEMP AR MIN 3d	17.45	4.06	7.87	14.51	18.48	20.46	24.30	16.52	-0.43	-0.81	23.30
RAD SOLAR MEDIA 3d	162.86	61.86	51.93	117.81	162.69	203.06	334.03	3827.05	0.66	0.15	37.99
VEL MAX VENTO MEDIA 3d	15.83	3.79	8.09	13.09	15.55	18.44	27.18	14.39	0.37	-0.02	23.96
EFL DQO	35.88	33.36	6.35	19.38	27.29	36.29	207.00	1112.94	3.20	11.10	92.98
AFL_SS	80.08	148.81	6.67	42.17	55.56	70.19	1466.67	22144.63	8.14	72.01	185.83
AFL_ST	511.93	190.56	0.00	420.00	500.00	549.17	1513.33	36313.17	2.56	11.46	37.22
AFL_STV	262.44	138.32	0.00	195.00	246.67	299.17	1115.00	19130.92	3.70	19.46	52.70
AFL_STF	257.67	97.33	0.00	209.17	250.00	292.50	813.33	9472.17	1.96	9.64	37.77
AFL_pH	7.23	0.28	6.18	7.10	7.23	7.37	8.04	0.08	-0.23	2.01	3.82
AFL_ALCALINIDADE	274.97	38.60	65.00	260.00	274.54	300.00	345.00	1490.11	-1.64	7.23	14.04
AFL_DQO	460.27	171.84	184.30	354.93	420.00	528.63	1082.05	29527.57	1.28	1.89	37.33
AFL_DBO	267.73	90.24	17.70	238.00	268.03	315.00	600.00	8142.64	-0.25	3.24	33.70
AFLRelação DQO/DBO	2.57	3.95	0.68	1.26	1.57	2.17	28.40	15.58	4.60	22.27	153.88
AFL_N-NH4+	65.14	18.75	24.40	53.58	69.25	77.80	104.50	351.72	-0.37	-0.69	28.79
AFL_N-NO2-	0.14	0.05	0.00	0.12	0.14	0.16	0.26	0.00	-0.54	2.03	33.31
AFL_N-NO3-	1.16	1.76	0.00	0.00	0.75	1.40	13.70	3.09	4.24	24.64	151.96
AFL_P-PO43-	36.17	9.56	15.70	29.65	34.76	42.53	71.00	91.40	0.60	0.63	26.43
Fluxo L/m2/d	91.96	13.72	72.98	77.19	98.54	103.49	114.87	188.26	-0.13	-1.69	14.92
Carr Org Sup g DQO/m2/d	41792.33	16338.82	16341.83	32964.98	37722.29	47651.72	113844.50	266956941.43	2.16	6.60	39.10

APÊNDICE C – TESTES DE DISTRIBUIÇÃO NORMAL PARA OS PARÂMETROS DO CONJUNTO DE DADOS.

Parâmetro	Anderson-Darling				Shapiro-Wilk		
	A ²	Valor Critico	Nível de Significancia	Distribuição Normal?	P-Value	Alpha	Distribuição Normal?
TEMP AR MEDIO	0.790	0.759	0.05	false	0.054	0.05	true
UMIDADE RLV	1.915	0.759	0.05	false	0.001	0.05	false
MOLHAMENTO FOLIAR MEDIO	0.865	0.759	0.05	false	0.004	0.05	false
PRESSAO ATM MED	0.266	0.759	0.05	true	0.910	0.05	true
RADIACAO SOLAR MEDIA	1.004	0.759	0.05	false	0.009	0.05	false
VEL MAX VENTO MEDIA	0.717	0.759	0.05	true	0.017	0.05	false
UMIDADE AR MAX	6.650	0.759	0.05	false	0.000	0.05	false
UMIDADE AR MIIN	2.087	0.759	0.05	false	0.002	0.05	false
TEMP MAX	0.264	0.759	0.05	true	0.796	0.05	true
TEMP MIN	2.943	0.759	0.05	false	0.000	0.05	false
PRECP	18.815	0.759	0.05	false	0.000	0.05	false
RADIACAO SOLAR MAX	0.428	0.759	0.05	true	0.253	0.05	true
RADIACAO SOLAR MIN	39.311	0.759	0.05	false	0.000	0.05	false
RADIACAO SOLAR MEDIANA	16.682	0.759	0.05	false	0.000	0.05	false
MOLHAMENTO FOLIAR MEDIANO	13.222	0.759	0.05	false	0.000	0.05	false
VEL VENTO MAXIMA MAXIMA	0.349	0.759	0.05	true	0.340	0.05	true
TEMP MEDIANA	0.871	0.759	0.05	false	0.022	0.05	false
VEL VENTO MAXIMA MIN	8.938	0.759	0.05	false	0.000	0.05	false
PRECP ACUM 3d	11.270	0.759	0.05	false	0.000	0.05	false
PRECP acum 7d	6.894	0.759	0.05	false	0.000	0.05	false
TEMP AR 3d	0.765	0.759	0.05	false	0.031	0.05	false
TEMP AR 7d	0.497	0.759	0.05	true	0.183	0.05	true
TEMP AR MAX 3d	0.413	0.759	0.05	true	0.206	0.05	true
TEMP AR MIN 3d	1.465	0.759	0.05	false	0.002	0.05	false
RAD SOLAR MEDIA 3d	0.871	0.759	0.05	false	0.004	0.05	false
VEL MAX VENTO MEDIA 3d	0.306	0.759	0.05	true	0.386	0.05	true
EFL DQO	12.707	0.759	0.05	false	0.000	0.05	false
AFL_SS	23.139	0.759	0.05	false	0.000	0.05	false
AFL_ST	6.470	0.759	0.05	false	0.000	0.05	false
AFL_STV	7.498	0.759	0.05	false	0.000	0.05	false
AFL_STF	3.076	0.759	0.05	false	0.000	0.05	false
AFL_pH	1.367	0.759	0.05	false	0.003	0.05	false
AFL_ALCALINIDADE	1.978	0.759	0.05	false	0.000	0.05	false
AFL_DQO	2.956	0.759	0.05	false	0.000	0.05	false
AFL_DBO	3.480	0.759	0.05	false	0.000	0.05	false
AFLRelação DQO/DBO	24.735	0.759	0.05	false	0.000	0.05	false
AFL_N-NH4+	1.561	0.759	0.05	false	0.003	0.05	false
AFL_N-NO2-	1.788	0.759	0.05	false	0.000	0.05	false
AFL_N-NO3-	9.536	0.759	0.05	false	0.000	0.05	false
AFL_P-PO43-	0.603	0.759	0.05	true	0.055	0.05	true
Fluxo L/m2/d	7.285	0.759	0.05	false	0.000	0.05	false
Carr Org Sup g DQO/m2/d	4.434	0.759	0.05	false	0.000	0.05	false

**APÊNDICE D – IMPORTÂNCIA DE CADA VARIÁVEL NO PROCESSO DE
PREDIÇÃO DO MODELO XGBOOST**

Parâmetro	Importância
AFL_N-NO3-	0.142
AFL_DQO	0.093
Carr Org Sup g DQO/m2/d	0.073
AFL_DBO	0.062
UMIDADE AR MIN	0.056
AFL_ALCALINIDADE	0.041
RAD SOLAR MEDIA 3d	0.04
AFL_SS	0.036
Fluxo L/m2/d	0.033
AFL_N-NO2-	0.028
AFLRelação DQO/DBO	0.027
UMIDADE AR MAX	0.024
TEMP MEDIANA	0.023
PRESSAO ATM MED	0.02
AFL_STV	0.019
UMIDADE RLV	0.019
VEL VENTO MAXIMA MAXIMA	0.018
AFL_pH	0.018
PRECP	0.017
TEMP MIN	0.015
AFL_P-PO43-	0.015
RADIACAO SOLAR MEDIA	0.015
AFL_ST	0.014
TEMP AR MEDIO	0.014
TEMP AR MAX 3d	0.014
VEL MAX VENTO MEDIA 3d	0.012
AFL_N-NH4+	0.011
TEMP AR MIN 3d	0.009
MOLHAMENTO FOLIAR MEDIO	0.009
VEL MAX VENTO MEDIA	0.009
AFL_STF	0.009
VEL VENTO MAXIMA MIN	0.009
TEMP MAX	0.009
PRECP ACUM 3d	0.008
RADIACAO SOLAR MIN	0.008
RADIACAO SOLAR MAX	0.007
TEMP AR 3d	0.006
RADIACAO SOLAR MEDIANA	0.006
TEMP AR 7d	0.005
PRECP acum 7d	0.004
MOLHAMENTO FOLIAR MEDIANO	0.003

**APÊNDICE E – PROPORÇÃO DE VARIÂNCIA EXPLICADA E
VARIÂNCIA EXPLICADA ACUMULADA PARA CADA COMPONENTE
PRINCIPAL**

Componente Principal	Proporção de Variância Explicada	Proporção de Variância Explicada Cumulativa
PC1	0,228	0,228
PC2	0,116	0,344
PC3	0,103	0,447
PC4	0,083	0,530
PC5	0,060	0,590
PC6	0,054	0,644
PC7	0,044	0,688
PC8	0,040	0,729
PC9	0,033	0,762
PC10	0,029	0,790
PC11	0,026	0,816
PC12	0,022	0,838
PC13	0,020	0,858
PC14	0,017	0,875
PC15	0,016	0,892
PC16	0,015	0,907
PC17	0,014	0,921
PC18	0,010	0,931
PC19	0,009	0,940
PC20	0,009	0,949
PC21	0,008	0,957
PC22	0,007	0,964
PC23	0,006	0,970
PC24	0,005	0,975
PC25	0,004	0,979
PC26	0,003	0,982
PC27	0,003	0,986
PC28	0,003	0,988
PC29	0,002	0,991
PC30	0,002	0,993
PC31	0,002	0,995
PC32	0,001	0,996
PC33	0,001	0,997
PC34	0,001	0,998
PC35	0,001	0,998
PC36	0,001	0,999
PC37	0,000	0,999
PC38	0,000	1,000
PC39	0,000	1,000
PC40	0,000	1,000
PC41	0,000	1,000

**APÊNDICE F – VALORES DE *LOADINGS* DAS COMPONENTES
PRINCIPAIS PARA CADA PARÂMETRO.**

Parâmetro	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
TEMP AR MEDIO	-0,312	0,031	0,046	0,237	-0,179	-0,079	0,033	0,011
UMIDADE RLV	-0,044	-0,312	-0,248	0,039	0,222	0,301	-0,293	-0,204
MOLHAMENTO FOLIAR MEDIO	0,041	0,051	0,218	-0,004	-0,034	-0,168	0,024	-0,036
PRESSAO ATM MED	0,058	0,219	-0,146	-0,048	-0,215	0,204	-0,007	0,327
RADIACAO SOLAR MEDIA	-0,001	0,046	0,141	-0,042	-0,069	0,129	-0,037	0,057
VEL MAX VENTO MEDIA	0,052	0,238	0,032	0,122	0,001	-0,028	0,062	0,214
UMIDADE AR MAX	-0,081	-0,063	0,052	0,011	0,303	-0,139	0,257	-0,265
UMIDADE AR MIN	0,078	0,066	-0,207	0,080	0,021	-0,033	0,145	-0,002
TEMP MAX	0,005	-0,053	-0,293	0,088	0,017	-0,150	-0,151	-0,031
TEMP MIN	-0,023	0,215	0,197	-0,022	0,123	0,213	0,274	0,077
PRECP	-0,003	0,032	0,050	-0,058	0,126	-0,065	0,110	-0,031
RADIACAO SOLAR MAX	0,049	0,237	-0,050	0,070	0,139	-0,009	0,141	0,213
RADIACAO SOLAR MIN	-0,015	-0,155	0,100	-0,188	-0,056	0,017	-0,110	-0,069
RADIACAO SOLAR MEDIANA	0,023	0,149	-0,039	0,024	0,026	0,185	0,267	0,012
MOLHAMENTO FOLIAR MEDIANO	-0,019	0,043	0,095	0,417	0,134	-0,049	-0,322	0,251
VEL VENTO MAXIMA MAXIMA	-0,030	0,201	-0,231	-0,172	0,129	-0,104	0,280	0,172
TEMP MEDIANA	0,055	-0,016	0,144	-0,297	-0,113	0,046	-0,116	-0,037
VEL VENTO MAXIMA MIN	0,121	0,003	-0,042	0,062	0,107	0,080	-0,084	-0,069
PRECP ACUM 3d	0,008	0,009	-0,002	-0,453	0,110	-0,048	-0,141	0,103
PRECP acum 7d	-0,011	-0,096	0,107	0,406	-0,057	0,073	0,293	-0,239
TEMP AR 3d	-0,045	-0,006	0,148	-0,021	0,120	-0,001	-0,078	0,173
TEMP AR 7d	-0,077	-0,046	-0,058	-0,090	-0,098	0,076	-0,065	0,246
TEMP AR MAX 3d	0,108	-0,050	0,106	0,289	0,116	0,123	-0,018	0,165
TEMP AR MIN 3d	-0,146	0,055	0,046	-0,063	-0,022	0,015	-0,095	-0,037
RAD SOLAR MEDIA 3d	0,098	-0,022	-0,099	0,169	0,077	0,057	-0,304	0,137
VEL MAX VENTO MEDIA 3d	-0,180	0,090	0,034	-0,193	0,129	0,184	-0,092	0,107
AFL_SS	-0,013	0,017	0,333	-0,102	0,061	-0,108	0,016	0,091
AFL_ST	-0,160	0,253	-0,028	0,114	0,045	0,172	-0,174	0,122
AFL_STV	0,042	-0,125	0,233	0,067	-0,067	0,192	0,020	0,109
AFL_STF	-0,032	0,357	-0,230	0,047	-0,082	-0,161	-0,133	-0,151
AFL_pH	-0,036	-0,230	-0,427	-0,041	-0,212	0,352	0,280	0,090
AFL_ALCALINIDADE	-0,033	-0,153	0,255	-0,026	0,095	0,277	0,086	0,014
AFL_DQO	0,015	-0,426	-0,072	0,000	0,084	-0,414	0,137	0,462
AFL_DBO	-0,044	0,007	0,034	0,024	0,117	-0,139	-0,044	0,033
AFLRelação DQO/DBO	0,122	0,005	0,128	-0,016	0,090	0,258	-0,018	0,023
AFL_N-NH4+	0,048	-0,090	0,057	0,051	-0,328	-0,173	0,049	0,011
AFL_N-NO2-	-0,045	0,081	0,136	0,005	-0,587	-0,024	-0,100	-0,082
AFL_N-NO3-	-0,034	0,282	-0,050	-0,032	0,179	-0,084	-0,095	-0,256
AFL_P-PO43-	-0,328	-0,053	0,009	-0,005	-0,010	-0,032	0,014	0,047
Fluxo L/m2/d	-0,458	-0,028	-0,011	0,006	0,041	0,022	0,014	0,025
Carr Org Sup g DQO/m2/d	0,635	0,030	0,002	-0,011	-0,018	-0,012	-0,014	-0,029

Parâmetro	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16
TEMP AR MEDIO	-0,296	-0,282	-0,013	-0,180	-0,036	-0,117	0,060	-0,059
UMIDADE RLV	-0,048	-0,055	-0,156	0,170	0,108	0,152	-0,196	0,262
MOLHAMENTO FOLIAR MEDIO	0,078	0,028	-0,040	0,005	-0,101	-0,083	0,220	-0,117
PRESSAO ATM MED	-0,066	0,146	0,360	-0,218	0,068	0,048	-0,157	0,157
RADIACAO SOLAR MEDIA	-0,058	0,060	0,035	-0,118	0,145	0,099	0,167	0,035
VEL MAX VENTO MEDIA	-0,019	0,132	-0,227	-0,183	0,458	0,393	0,055	-0,200
UMIDADE AR MAX	0,008	-0,166	0,176	0,167	0,350	0,357	0,068	-0,079
UMIDADE AR MIN	0,114	0,058	-0,052	0,120	-0,122	-0,127	-0,161	0,033
TEMP MAX	-0,004	0,011	-0,121	-0,003	-0,037	-0,030	-0,267	-0,163
TEMP MIN	0,040	-0,074	0,069	0,212	-0,089	-0,072	0,224	0,351
PRECP	0,010	-0,022	0,125	0,044	0,091	0,098	0,135	-0,071
RADIACAO SOLAR MAX	0,003	0,059	0,184	0,168	-0,087	-0,005	-0,043	-0,094
RADIACAO SOLAR MIN	-0,041	0,008	-0,021	-0,083	0,128	0,087	0,129	0,188
RADIACAO SOLAR MEDIANA	0,099	-0,029	-0,172	0,102	-0,070	-0,147	-0,046	0,293
MOLHAMENTO FOLIAR MEDIANO	-0,081	0,049	-0,086	0,213	-0,175	0,001	0,205	-0,085
VEL VENTO MAXIMA MAXIMA	-0,014	-0,043	-0,243	0,162	0,018	0,003	-0,297	-0,130
TEMP MEDIANA	0,091	0,079	-0,283	-0,087	-0,069	-0,097	0,257	-0,052
VEL VENTO MAXIMA MIN	0,217	0,063	-0,004	0,080	0,072	-0,019	0,189	0,232
PRECP ACUM 3d	0,064	0,022	-0,082	0,179	0,007	0,059	0,032	-0,154
PRECP acum 7d	-0,068	0,067	-0,131	-0,155	0,012	-0,065	-0,091	0,021
TEMP AR 3d	-0,115	0,043	-0,098	0,218	-0,043	-0,018	0,027	-0,140
TEMP AR 7d	-0,202	-0,083	0,076	-0,140	0,087	0,215	-0,102	0,308
TEMP AR MAX 3d	-0,085	0,138	-0,255	0,143	0,065	0,167	0,057	0,161
TEMP AR MIN 3d	-0,160	0,079	0,161	0,148	0,080	-0,079	0,002	0,059
RAD SOLAR MEDIA 3d	0,102	0,027	0,216	-0,031	-0,059	0,023	0,044	-0,137
VEL MAX VENTO MEDIA 3d	-0,273	-0,071	-0,144	0,039	-0,017	-0,060	-0,023	0,001
AFL_SS	-0,106	0,006	-0,380	-0,172	-0,236	0,175	-0,280	0,150
AFL_ST	0,137	-0,362	-0,090	0,133	0,177	-0,108	0,019	-0,067
AFL_STV	0,246	-0,209	-0,135	-0,161	0,090	-0,010	-0,200	-0,164
AFL_STF	0,290	-0,149	-0,128	-0,136	-0,323	0,449	0,170	0,181
AFL_pH	-0,143	-0,081	-0,183	0,025	-0,144	0,134	0,414	-0,262
AFL_ALCALINIDADE	-0,008	-0,152	0,290	0,057	-0,471	0,402	-0,178	-0,216
AFL_DQO	0,012	-0,333	0,029	0,000	0,018	-0,058	0,137	0,221
AFL_DBO	0,164	-0,125	0,091	-0,097	-0,020	0,027	-0,066	0,092
AFLRelação DQO/DBO	0,103	-0,039	0,015	-0,024	0,158	-0,162	-0,075	-0,157
AFL_N-NH4+	-0,179	0,402	0,008	0,393	-0,022	0,190	-0,067	0,057
AFL_N-NO2-	0,080	-0,369	-0,014	0,437	0,140	0,075	-0,132	0,012
AFL_N-NO3-	-0,529	-0,154	0,018	-0,054	-0,050	-0,052	0,064	0,006
AFL_P-PO43-	0,089	0,084	-0,006	0,005	0,028	-0,021	-0,009	0,025
Fluxo L/m2/d	0,102	0,147	-0,001	-0,031	-0,025	0,012	0,013	0,004
Carr Org Sup g DQO/m2/d	-0,240	-0,263	-0,011	-0,005	-0,012	0,005	0,004	0,012

Parâmetro	PC17	PC18	PC19	PC20
TEMP AR MEDIO	-0,307	-0,052	-0,016	-0,031
UMIDADE RLV	-0,046	0,234	-0,104	-0,098
MOLHAMENTO FOLIAR MEDIO	0,037	-0,141	-0,140	-0,095
PRESSAO ATM MED	0,077	0,209	0,300	0,242
RADIACAO SOLAR MEDIA	-0,002	0,190	0,039	-0,090
VEL MAX VENTO MEDIA	0,052	0,270	-0,338	-0,215
UMIDADE AR MAX	-0,074	-0,060	0,233	0,247
UMIDADE AR MIN	0,067	-0,080	-0,068	0,040
TEMP MAX	0,014	-0,042	-0,079	-0,047
TEMP MIN	-0,034	-0,062	-0,052	-0,272
PRECP	-0,018	-0,040	0,167	0,216
RADIACAO SOLAR MAX	0,034	0,018	-0,026	-0,144
RADIACAO SOLAR MIN	-0,011	-0,178	0,026	-0,032
RADIACAO SOLAR MEDIANA	0,025	0,143	-0,129	0,039
MOLHAMENTO FOLIAR MEDIANO	-0,022	0,089	0,057	0,166
VEL VENTO MAXIMA MAXIMA	-0,032	-0,048	-0,093	0,252
TEMP MEDIANA	0,021	0,162	-0,264	0,568
VEL VENTO MAXIMA MIN	0,087	0,167	0,065	0,021
PRECP ACUM 3d	-0,046	-0,015	0,139	-0,337
PRECP acum 7d	-0,061	0,137	0,027	-0,040
TEMP AR 3d	-0,056	-0,081	-0,073	-0,116
TEMP AR 7d	-0,030	-0,497	-0,322	0,038
TEMP AR MAX 3d	0,180	-0,365	0,062	0,220
TEMP AR MIN 3d	-0,269	0,223	-0,268	-0,007
RAD SOLAR MEDIA 3d	0,100	-0,054	0,120	-0,060
VEL MAX VENTO MEDIA 3d	-0,198	0,023	0,235	0,091
AFL_SS	0,041	0,171	0,380	-0,106
AFL_ST	-0,213	0,042	0,068	0,113
AFL_STV	0,055	-0,132	0,096	-0,051
AFL_STF	-0,098	-0,061	0,026	-0,036
AFL_pH	0,054	0,004	0,140	-0,093
AFL_ALCALINIDADE	0,045	0,158	-0,291	0,119
AFL_DQO	0,165	0,274	-0,052	0,005
AFL_DBO	-0,093	0,050	-0,105	0,043
AFLRelação DQO/DBO	0,191	-0,107	-0,057	-0,023
AFL_N-NH4+	-0,144	0,029	0,059	0,027
AFL_N-NO2-	0,309	0,066	0,039	-0,009
AFL_N-NO3-	0,573	0,063	-0,059	-0,009
AFL_P-PO43-	0,177	-0,007	-0,004	-0,011
Fluxo L/m2/d	0,274	-0,022	0,016	0,007
Carr Org Sup g DQO/m2/d	-0,133	0,000	-0,001	-0,001

ANEXO A – DESCRIÇÃO



OF. 010/GESAD/2022.

Florianópolis, 28 de julho de 2022.

De: Prof. Pablo Heleno Sezerino – GESAD/ENS-UFSC.

Para: Senhor Gerente da Epagri/Ciram.

Assunto: Informações para o desenvolvimento do Trabalho de Conclusão de Curso em Engenharia Sanitária e Ambiental, da UFSC.

Prezado Senhor.

Vimos por meio desta solicitar dados da estação de código 1006, para a localidade de Florianópolis no período de 01/01/2015 até 31/12/2020, os quais serão utilizados para fins acadêmicos através do trabalho de conclusão de curso do requerente Emanuel Goulart Farias, intitulado “Aplicabilidade de redes neurais artificiais para predição de remoção da demanda química de oxigênio (DQO) sob o desempenho de tratamento de esgoto promovido por unidade de Wetland Construído Vertical de Fundo Saturado (WCV-FS)”. As variáveis que precisamos estão listadas na Tabela 1, e serão usadas exclusivamente para fins não comerciais.

Estamos cientes de que a Epagri não permite o repasse dos dados para terceiros assim como solicita que seja citada como fonte dos dados. Todas as publicações que usarem os dados cedidos pela Epagri será citado nas referências o banco de dados da Epagri:

Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina. **Banco de dados de variáveis ambientais de Santa Catarina**. Florianópolis: Epagri, 2020. 20p. (Epagri, Documentos, 310) - ISSN 2674-9521 (On-line)

Durante o desenvolvimento dos trabalhos está previsto o análise de dados ambientais e de qualidade de efluente tais como demanda química de oxigênio, sólidos totais, nitrogênio total, fósforo, alcalinidade e pH coletados durante o período de 2015 a 2020. Os dados coletados pelo trabalho de conclusão de curso estarão à disposição da Epagri caso sejam solicitados por e-mail. A Epagri se compromete a usar os dados disponibilizados estritamente para pesquisa, citar a fonte de dados e não disponibilizá-los para terceiros.



GESAD – Grupo de Estudos em Saneamento Descentralizado
 ENS – Departamento de Engenharia Sanitária e Ambiental
 CTC – Centro Tecnológico
 UFSC – Universidade Federal de Santa Catarina
 Campus João Reitor David Ferreira Lima, s/n – Trindade, Florianópolis – SC. CEP: 88040-900
 Telefone: +55 (48) 3721-7696 www.gesad.ufsc.br



Tabela 1. Variáveis solicitadas para a Epagri

Variável	Frequência
Temperatura do ar	Horária
Umidade relativa do ar	Horária
Chuva	Horária
Molhamento foliar	Horária
Pressão atmosférica	Horária
Radiação solar	Horária
Velocidade Máxima do Vento	Horária
Temperatura do Solo (20, 30, 50cm)	Horária
Temperatura da Água	Horária
Insolação	Horária

Concordamos que a omissão da citação da fonte de dados, o repasse dos dados a terceiros ou a negativa em disponibilizar dados ambientais para a Epagri acarretará na desqualificação da instituição para receber dados sem custos.

Certos de vossa colaboração, agradecemos antecipadamente.

Atenciosamente.



Documento assinado digitalmente
 Pablo Heleno Sezerino
 Data: 28/07/2022 15:53:47-0300
 CPF: 952.175.849-04
 Verifique as assinaturas em <https://v.ufsc.br>

Prof. Dr. Pablo Heleno Sezerino
 Departamento de Engenharia Sanitária e Ambiental – ENS/CTC-UFSC
 Supervisor GESAD-ENS/UFSC



GESAD – Grupo de Estudos em Saneamento Descentralizado
 ENS – Departamento de Engenharia Sanitária e Ambiental
 CTC – Centro Tecnológico
 UFSC – Universidade Federal de Santa Catarina
 Campus João Reitor David Ferreira Lima, s/n – Trindade, Florianópolis – SC. CEP: 88040-900
 Telefone: +55 (48) 3721-7696 www.gesad.ufsc.br

**ANEXO A – TABELA DE AFERIÇÃO DA VAZÃO DA BOMBA DOS
WETLANDS CONSTRUÍDOS DO GESAD.**

Data	Tempo (s)	Volume (L)	Vazão da bomba (L/s)	Volume do pulso	Vazão Aplicada (L/dia)
07/04/2015	32.38	39.100	1.208	141.282	565.127
07/04/2015	32.00	39.100	1.222	142.959	571.838
07/04/2015	32.65	39.100	1.198	140.113	560.453
07/04/2015	32.25	39.100	1.212	141.851	567.405
07/04/2015	32.22	39.100	1.214	141.983	567.933
07/04/2015	30.78	39.100	1.270	148.626	594.503
07/04/2015	32.28	39.100	1.211	141.719	566.877
13/04/2015	30.51	39.100	1.282	149.941	599.764
13/04/2015	30.31	39.100	1.290	150.930	603.722
13/04/2015	30.23	39.100	1.293	151.330	605.319
13/04/2015	30.08	39.100	1.300	152.084	608.338
13/04/2015	30.38	39.100	1.287	150.583	602.330
14/04/2015	31.16	39.100	1.255	146.813	587.253
14/04/2015	30.70	39.100	1.274	149.013	596.052
14/04/2015	30.54	39.100	1.280	149.794	599.175
14/04/2015	31.30	39.100	1.249	146.157	584.626
14/04/2015	31.03	39.100	1.260	147.428	589.713
14/04/2015	30.81	39.100	1.269	148.481	593.924
14/04/2015	31.08	39.100	1.258	147.191	588.764
16/04/2015	31.74	39.100	1.232	144.130	576.522
16/04/2015	31.10	39.100	1.257	147.096	588.386
16/04/2015	31.69	39.100	1.234	144.358	577.431
17/04/2015	30.51	39.100	1.282	149.941	599.764
17/04/2015	30.71	39.100	1.273	148.965	595.858
17/04/2015	30.61	39.100	1.277	149.451	597.805
17/04/2015	30.70	39.100	1.274	149.013	596.052
17/04/2015	30.56	39.100	1.279	149.696	598.783
17/04/2015	30.71	39.100	1.273	148.965	595.858
21/05/2015	31.30	39.100	1.249	146.157	584.626
21/05/2015	30.86	39.100	1.267	148.240	592.962
21/05/2015	30.88	39.100	1.266	148.144	592.578
21/05/2015	31.05	39.100	1.259	147.333	589.333
25/06/2015	30.61	39.100	1.277	149.451	597.805
02/07/2015	31.30	39.100	1.249	146.157	584.626
16/07/2015	31.06	39.100	1.259	147.286	589.144
23/07/2015	30.68	39.100	1.274	149.110	596.441
03/08/2015	30.52	39.100	1.281	149.892	599.567
06/08/2015	30.20	39.100	1.295	151.480	605.921
13/08/2015	31.22	39.100	1.252	146.531	586.124
20/08/2015	31.62	39.100	1.237	144.677	578.710

27/08/2015	30.63	39.100	1.277	149.354	597.414
03/09/2015	30.85	39.100	1.267	148.288	593.154
10/09/2015	30.78	39.100	1.270	148.626	594.503
24/09/2015	30.84	39.100	1.268	148.337	593.346
01/10/2015	31.10	39.100	1.257	147.096	588.386
08/10/2015	31.32	39.100	1.248	146.063	584.253
15/10/2015	31.37	39.100	1.246	145.830	583.322
29/10/2015	31.26	39.100	1.251	146.344	585.374
03/12/2015	31.82	39.100	1.229	143.768	575.072
17/12/2015	31.61	39.100	1.237	144.723	578.893
17/03/2016	32.10	39.100	1.218	142.514	570.056
12/04/2016	30.51	39.100	1.282	149.941	599.764
19/04/2016	30.74	39.100	1.272	148.819	595.277
28/04/2016	32.29	39.100	1.211	141.675	566.702
12/05/2016	32.46	39.100	1.205	140.933	563.734
24/05/2016	31.59	39.100	1.238	144.815	579.259
08/06/2016	32.15	39.100	1.216	142.292	569.170
30/06/2016	31.66	39.100	1.235	144.495	577.979
28/07/2016	31.94	39.100	1.224	143.228	572.912
11/08/2016	32.80	39.100	1.192	139.473	557.890
08/09/2016	32.42	39.100	1.206	141.107	564.429
13/10/2016	32.61	39.100	1.199	140.285	561.141
20/10/2016	32.47	39.100	1.204	140.890	563.560
27/10/2016	32.78	39.100	1.193	139.558	558.231
03/11/2016	33.43	39.100	1.170	136.844	547.377
12/01/2017	32.59	39.100	1.200	140.371	561.485
26/01/2017	33.43	39.100	1.170	136.844	547.377
09/02/2017	32.66	39.100	1.197	140.070	560.282
23/04/2017	33.00	39.100	1.185	138.627	554.509
08/03/2017	32.45	39.100	1.205	140.977	563.908
22/03/2017	78.00	39.100	0.501	192.492	769.969
19/04/2017	78.31	39.100	0.499	191.730	766.921
17/05/2017	78.50	39.100	0.498	191.266	765.065
14/06/2017	79.00	39.100	0.495	190.056	760.223
29/06/2017	79.25	39.100	0.493	189.456	757.825
13/07/2017	79.74	39.100	0.490	188.292	753.168
02/08/2017	77.00	39.100	0.508	194.992	779.969
23/08/2017	77.94	39.100	0.502	192.640	770.562
13/09/2017	80.90	39.100	0.483	185.592	742.368
27/09/2017	80.00	39.100	0.489	187.680	750.720
18/10/2017	81.15	39.100	0.482	185.020	740.081
08/11/2017	78.25	39.100	0.500	191.877	767.509
13/12/2017	80.90	39.100	0.483	185.592	742.368
17/01/2018	81.30	39.100	0.481	184.679	738.716
31/01/2018	81.26	39.100	0.481	184.770	739.079
22/02/2018	82.47	39.100	0.474	182.059	728.236

07/03/2018	81.09	39.100	0.482	185.157	740.629
21/03/2018	81.89	39.100	0.477	183.348	733.394
11/04/2018	83.22	39.100	0.470	180.418	721.673
25/04/2018	81.66	39.100	0.479	183.865	735.459
09/05/2018	83.67	39.100	0.467	179.448	717.791
23/05/2018	84.87	39.100	0.461	176.911	707.642
27/06/2018	75.49	39.100	0.518	198.893	795.570
11/07/2018	75.99	39.100	0.515	197.584	790.336
01/08/2018	79.82	39.100	0.490	188.103	752.413
15/08/2018	81.79	39.100	0.478	183.573	734.290
29/08/2018	73.72	39.100	0.530	203.668	814.672
12/09/2018	74.61	39.100	0.524	201.238	804.954
26/09/2018	74.21	39.100	0.527	202.323	809.293
10/10/2018	77.76	39.100	0.503	193.086	772.346
31/10/2018	76.11	39.100	0.514	197.272	789.089
14/11/2018	75.91	39.100	0.515	197.792	791.168
28/11/2018	76.02	39.100	0.514	197.506	790.024
12/12/2018	75.62	39.100	0.517	198.551	794.203
07/03/2018	80	39.100	0.489	187.680	750.720
20/03/2019	76.74	39.100	0.510	195.653	782.611
03/04/2019	80.04	39.100	0.489	187.586	750.345
17/04/2018	81.95	39.100	0.477	183.214	732.857
09/05/2019	80.01	39.100	0.489	187.657	750.626
12/06/2019	78.85	39.100	0.496	190.417	761.669
26/06/2019	80.51	39.100	0.486	186.491	745.964
10/07/2019	78.83	39.100	0.496	190.466	761.862
24/07/2019	78.35	39.100	0.499	191.632	766.530
21/08/2019	81.24	39.100	0.481	184.815	739.261
02/10/2019	75.6	39.100	0.517	198.603	794.413
06/11/2019	74.82	39.100	0.523	200.674	802.694
20/11/2019	73.38	39.100	0.533	204.612	818.446
04/12/2019	73.62	39.100	0.531	203.945	815.778
29/01/2020	74.82	39.100	0.523	200.674	802.694
26/02/2020	72.79	39.100	0.537	206.264	825.058
30/04/2020	73.32	39.100	0.533	204.779	819.116
13/05/2020	74.04	39.100	0.528	202.788	811.151
17/06/2020	73.98	40.100	0.542	208.143	832.571
01/07/2020	74.04	41.100	0.555	213.160	852.642
15/07/2020	75.06	42.100	0.561	215.380	861.519
