



MODELAGEM POR REDES NEURAIS ARTIFICIAIS PARA PREDIZER O TEOR DE FÓSFORO E NITROGÊNIO PRESENTES EM EFLUENTES

Giulia Rosado de Oliva Maya⁽¹⁾

Bacharel em Engenharia Química pelo Centro Universitário do Leste de Minas Gerais - UNILESTE; Estágio e pesquisa realizados na Companhia de Saneamento de Minas Gerais - COPASA

Tayla Luiza Pereira Borges⁽²⁾

Bacharel em Engenharia Química pelo Centro Universitário do Leste de Minas Gerais - UNILESTE; Mestranda em Engenharia Química pela Universidade Federal de Viçosa - UFV;

Débora Luiza Correia Alves⁽³⁾

Bacharel em Engenharia Química pelo Centro Universitário do Leste de Minas Gerais - UNILESTE

Letícia Fabri Turetta⁽⁴⁾

Doutora em Engenharia Química pela Universidade Federal de Minas Gerais - UFMG; Docente do curso de Engenharia Química na Universidade Federal de Itajubá - UNIFEI

Endereço⁽¹⁾: Rua Mário de Andrade, 297, Apartamento 3C, Cidade Nobre - Ipatinga - Minas Gerais - CEP: 35162-390 - Brasil - Tel: +55 (31) 98871-0352 - e-mail: giulia.maya@hotmail.com

RESUMO

O presente trabalho teve como objetivo propor modelos de Redes Neurais Artificiais (RNA) para prever o teor de fósforo e nitrogênio na saída de reatores UASB em uma Estação de Tratamento de Esgoto (ETE). Foi realizado testes de normalidade em todas as variáveis independentes, recorrendo aos modelos de Kolmogorov-Smirnov (K-S) e Shapiro-Wilk (S-W), posteriormente foi feita a análise de correlação aplicando os testes de Pearson e Spearman, além da modelagem por Redes Neurais Artificiais. Nos testes de K-S e S-W obteve-se um comportamento de dados não normais descrito por K-S. Após, foram realizadas análises de correlação aplicando os testes de Pearson e Spearman, onde visualizou-se melhores resultados sob o modelo de Spearman, encontrando que a variável de condutividade tem uma forte correlação negativa com nitrogênio, enquanto fósforo contém uma forte correlação com nitrogênio. As RNA's de fósforo mostraram-se com uma eficiência de 98,5% utilizando variáveis com correlações acima de 0,3 enquanto as de nitrogênio obtiveram 99,2% de eficiência ao utilizar correlações acima de 0,6. Os testes mostraram-se eficazes, sendo uma alternativa para a predição de efluentes.

PALAVRAS-CHAVE: Redes Neurais Artificiais, Normalidade, Correlação.

INTRODUÇÃO

O Brasil é um dos países com maior abundância de recursos hídricos naturais, porém, não possui uma distribuição homogênea da utilização da água, fazendo com que áreas industrializadas utilizem da demanda hídrica mais do que zonas rurais, ocasionando também, o descarte exagerado de efluentes sem o devido tratamento. Assim como alertado pela Agência Nacional de Águas (ANA, 2018), “a poluição hídrica ocasionada pela alta demanda, compromete a qualidade das águas superficiais pela falta de tratamento de esgoto adequado”.

O desordenado aumento do destino incorreto de efluentes (sendo esses domésticos ou industriais), contribui com a perda de ecossistemas por consequência da alteração de sua estrutura trófica. “Além dessas consequências, traz consigo diversos problemas indesejáveis, tais como contaminação da água destinada ao abastecimento público, diminuição do oxigênio dissolvido, maus odores e crescimento de macrófitas” (SCHINDLER, 2012).

A adição do nitrogênio (N) e fósforo (P) em ecossistemas aquáticos, além de causarem o aumento da produção primária e floração de algas e cianobactérias tóxicas, promovem também a elevação da turbidez, a mortandade dos peixes, perda do habitat natural e riscos à saúde humana (XU, 2014).

O fósforo (P) é um nutriente importante para a manutenção da vida, visto que organismos fotossintéticos o utilizam para sintetizar seus tecidos (RUTTENBERG, 2014). Isso explica a importância de se medir o elemento, visto que quantidades elevadas deste nutriente colabora com o exagerado crescimento de produtores primários, contribuindo com o acúmulo de fósforo em forma particulada, sedimentada ou dissolvida no meio aquático e conseqüentemente, na eutrofização de ecossistema.

De acordo com CARVALHO, “o nitrogênio é um elemento importante na natureza e essencial à produção das culturas. Embora a fonte principal deste nutriente se encontre na atmosfera, as plantas utilizam este nutriente a partir do solo”. Se ocorrer o aumento da concentração de nutrientes na água, automaticamente acontece um estímulo natural de crescimento bacteriano causando interferências na estrutura física e química da água, redução de transparência, exagero de biomassas, e florações de cianobactérias na qual não permitem a salubridade de um recurso hídrico.

Especificamente no Brasil, o Conselho Nacional de Meio Ambiente, por meio da Resolução do CONAMA 357, de 17 de Março de 2005, estabeleceu condições de qualidade para o enquadramento dos sistemas hídricos no território nacional, segundo os seus usos dominantes, e para o lançamento de águas residuais. Como instrumento legal, esta resolução estabeleceu limites superiores e inferiores para diversas variáveis em sistemas de água doce, salobra e salgada. Para o fósforo, o limite é de 0,20 mg /L de água doce e para o nitrogênio de 20 mg /L de água doce. Estes valores servem como base para as análises monitoramento da qualidade das águas superficiais, de forma homogênea de acordo com meio analisado,

OBJETIVO

O objetivo deste trabalho é propor dois modelos baseados em Redes Neurais Artificiais (RNA), sendo o Modelo RAP para prever o teor de nitrogênio e Modelo RAN para prever o teor de nitrogênio ao final do processo de reatores do tipo UASB de uma Estação de Tratamento de Esgoto (ETE), utilizando dados do processo e suas diferenças de eficiência em cima de diferentes intensidades de correlação das variáveis.

METODOLOGIA UTILIZADA

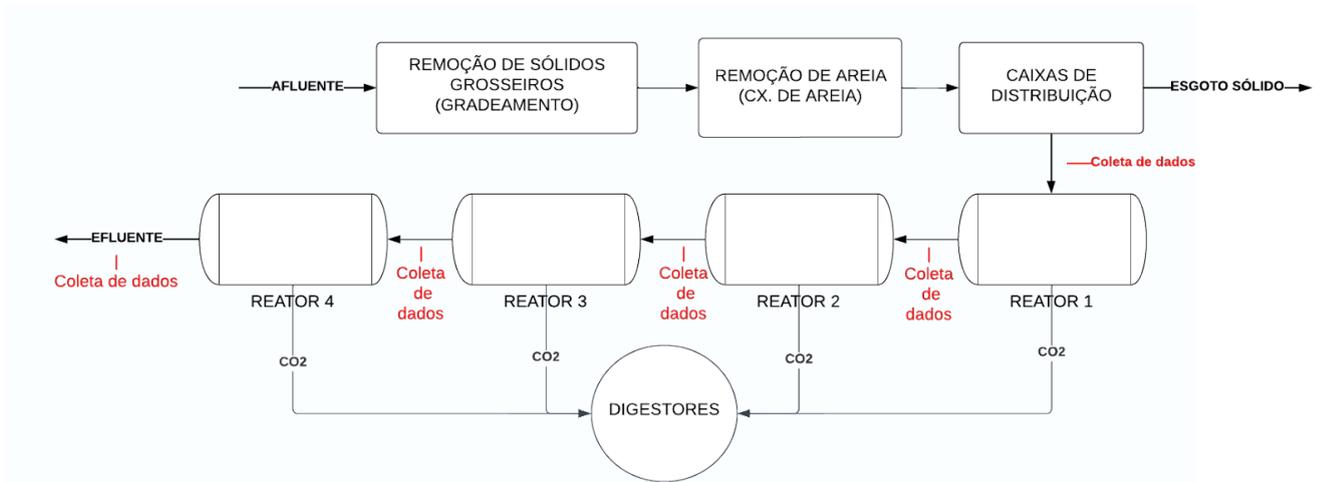
3.1 ÁREA DE ESTUDO E TRATAMENTO DE DADOS

O estudo foi desenvolvido em uma Estação de Tratamento de Esgoto (ETE) da Região Metropolitana do Vale do Aço, localizada no leste de Minas Gerais onde pertence a Bacia Hidrográfica do Rio Piracicaba (BHRP) que integra a Bacia Hidrográfica do Rio Doce (BHRD).

De acordo com SILVA, 2016 “o tratamento de esgoto é realizado na Estação de Tratamento de Efluentes (ETE), que segundo a ABNT/NBR 12209/1992, é um conjunto de unidades de tratamento, equipamentos, órgãos auxiliares, acessórios e sistemas de utilidades cuja finalidade é a redução das cargas poluidoras do esgoto sanitário”.

As primeiras etapas da ETE consistem na remoção de sólidos grosseiros, utilizando os processos de gradeamento (grades metálicas que funcionam como uma barreira para os sólidos) e pelas caixas de areia (ocorre por meio da sedimentação de matéria orgânica) visualizado na Figura 1. Todo esgoto sólido retirado nessas etapas é transferido para um aterro sanitário.

Figura 1) Fluxograma de etapas da ETE estudada.



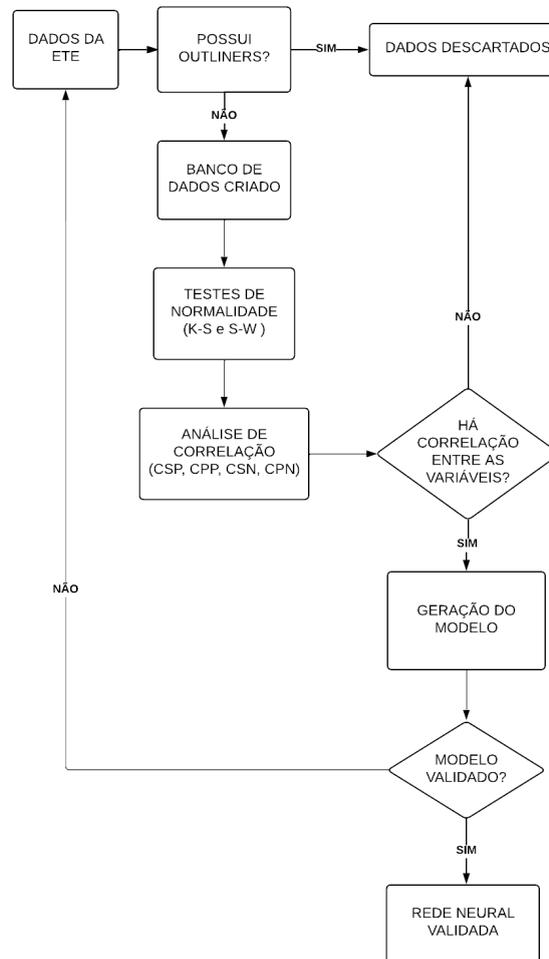
Fonte: Autoras, 2022.

Iniciado o tratamento, utilizam-se quatro reatores do tipo UASB que operam em série. Esta é uma tecnologia de tratamento biológico de esgotos baseada na decomposição anaeróbia da matéria orgânica. “Consiste em uma coluna de escoamento ascendente, composta de uma zona de digestão, uma zona de sedimentação, e o dispositivo separador de fases gás-sólido-líquido” (UTE, 2016). “O processo de digestão anaeróbia possibilita a decomposição da matéria orgânica, evitando que esse material contamine rios e lagos, evitando, portanto, a poluição ambiental” (AFONSO, 2016). Esse processo também garante a produção de subprodutos como o dióxido de carbono e metano, chamado de biogás, que pode ser utilizado para geração de energia, fonte de calor ou combustível veicular. Segundo PRADO (2010), “outros subprodutos gerados são o lodo, que é resultado da produção de sólidos biológicos no sistema e o efluente líquido clarificado”.

Para as coletas de dados, baseou-se em todas as entradas dos reatores e a saída do último reator, como ilustrado na Figura 1.

A metodologia proposta seguirá segundo o fluxograma descrito na Figura 2.

Figura 2) Fluxograma da metodologia proposta.



Fonte: Autoras, 2022.

A primeira etapa metodológica é o recebimento do banco de dados da ETE e a remoção dos outliers (sendo valores atípicos ou inconsistentes). A existência de outliers implica em prejuízos para a geração do modelo e validação dos resultados de testes estatísticos aplicados às amostras, sendo estes dados descartados do processo.

Após a retirada dos outliers, criou-se um banco de dados para a realização dos testes de normalidade com o objetivo de identificar se estes se comportavam como “normal” ou “não normal”, ou seja, se os dados estavam tendendo a uma regressão linear ou não linear. Com este conhecimento, pode-se realizar as análises de correlação a fim de mensurar o grau de relacionamento das variáveis com os parâmetros preditores: fósforo e nitrogênio.

As variáveis que apresentaram correlações significativas permaneceram, sendo logo após, filtradas em diferentes níveis de correlação variando de 0 a 1 de acordo com o modelo (0,1 a 0,3 para fósforo e 0,2 a 0,6 para nitrogênio), e enviadas separadamente para a geração do modelo de RNA's. Após a verificação dos erros e valores de r^2 , os modelos foram validados.

A Tabela 1 descreve as siglas das metodologias desenvolvidas em todas as etapas do trabalho, bem como as análises estatísticas realizadas para modelagens, interpretação e análise dos dados utilizados no processo.



Tabela 1) Descrição dos modelos.

METODOLOGIA	SIGLA
Correlação de Spearman fósforo	CSP
Correlação de Tabela Pearson fósforo	CPP
Correlação de Spearman nitrogênio	CSN
Correlação de Pearson nitrogênio	CPN
Rede Neural Artificial de fósforo	Modelo RAP _k
Rede Neural Artificial de nitrogênio	Modelo RAN _j

*k = correlação variando de 0,1 a 0,3

*j = correlação variando de 0,2 a 0,6

Fonte: Autoras, 2022.

Utiliza-se como controle independente variáveis das análises físico-químicas, apresentadas na Tabela 2, em todos os pontos de coleta. O período de análise compreende de maio de 2021 até maio de 2022, totalizando 12 meses de observação. Os dados das análises físico-químicas utilizados foram obtidos mediante análises dos afluentes e efluentes utilizando a metodologia descrita nos Standard Methods (APHA-WEF, 2016).

Tabela 2) Descrição das variáveis das análises físico-químicas.

Variável	Descrição	Unidade
pHi	Potencial hidrogeniônico	Adimensional
CONDi	Condutividade	Siemens por metro (S/m)
SSTi	Sólidos suspensos totais	Miligramas por litro (mg/L)
Pi	Fósforo	Miligramas por litro (mg/L)
Ni	Nitrogênio	Miligramas por litro (mg/L)
DBOi	Demanda bioquímica de oxigênio	Miligramas por litro (mg/L)
DQOi	Demanda química de oxigênio	Miligramas por litro (mg/L)
SSEDi	Sólidos sedimentáveis totais	Mililitros por litro (mL/L)

*i = Índice para reator (variando de 1 a 4 reatores)

Fonte: Autoras, 2022.

Com os dados obtidos da ETE, montou-se um banco de dados utilizando o software SPSS Statistic, e em seguida retirou-se os outliers de todas as variáveis selecionadas, sendo possível a diminuição das interferências de dados não conformes.

3.2 Análises de normalidade

O banco de dados foi submetido por um teste de normalidade por meio dos testes de Kolmogorov-Smirnov (K-S) e de Shapiro-Wilk (S-W), a análise da normalidade é uma importante ferramenta analítica utilizada para a verificação de uma determinada amostra de dados, visando classificar se estes se apresentavam como “normal” ou “não normal”, ou seja, linear ou não linear, colaborando diretamente com a retirada de qualquer ponto fora da curva que afete negativamente o algoritmo.

O teste pode ser aplicado em diversas análises estatísticas e inferência. Testes de normalidade estatística realizam a comparação dos valores de média e desvio padrão de uma amostra testada, com outro conjunto que possui uma distribuição normal. A hipótese nula explica que a amostra é normal, então se o teste é significativo ($p < 0,05$) a amostra é não normal.



O trabalho de LOPES explica que:

“Os testes K-S e S-W fornecem o parâmetro valor de prova α (valor-p, p-value ou significância), que pode ser interpretado como a medida do grau de concordância entre os dados e a hipótese nula (H_0), sendo H_0 correspondente à distribuição Normal. Quanto menor for o valor-p, menor é a consistência entre os dados e a hipótese nula. Então, a regra de decisão adotada para saber se a distribuição é Normal ou não é rejeitar H_0 : (i) se $\text{valor-p} \leq \alpha$, rejeita-se H_0 , ou seja, não se pode admitir que o conjunto de dados em questão tenha distribuição Normal; (ii) se $\text{valor-p} > \alpha$, não se rejeita H_0 , ou seja, a distribuição Normal é uma distribuição possível para o conjunto de dados em questão.” (v. 21, n. 1 p. 59–66, 2013).

3.3 Análises de correlação

A análise de correlação permite interpretar o quanto uma variável afeta o desempenho de um sistema de dados, podendo compreender o grau de relacionamento entre eles e se este comportamento está sendo forte, fraco, positivo ou negativo.

A finalidade de se realizar análises de correlação provém da investigação e medir graus de relação existentes, utilizando-se usualmente o coeficiente de correlação de Pearson, que segundo VIEIRA (2018) “Existe uma medida para o grau de relação linear entre duas variáveis numéricas”. Contudo, se as amostras analisadas não apresentaram comportamento linear é “utilizado o coeficiente de correlação de Spearman, onde se avalia a relação entre duas variáveis; sejam elas lineares ou não” (GUIMARAES, 2022).

Foram realizados testes de correlação, sendo utilizados os métodos de Pearson e Spearman, a fim de comparar o desempenho de ambos nos valores de correlação mediante os testes de normalidade.

3.4 Redes Neurais Artificiais

HAYKIN diz que:

As redes neurais artificiais (RNAs) são um modelo computacional que foi projetado inspirando-se no funcionamento do cérebro humano e sua capacidade de aprender tarefas. As RNAs foram inspiradas pelo funcionamento do cérebro humano com as redes neurais biológicas, composta de neurônios que se comunicam por meio de sinapses neurotransmissores através das quais os impulsos nervosos são transmitidos, propiciando a capacidade de aprender. O neurotransmissor pode excitar ou inibir a geração de pulsos nesse outro neurônio (2ª Ed., 696p).

“Assim, essa técnica não é baseada em regras comuns; ela se difere de outras formas de inteligência computacional pelo fato de ser treinada para reconhecer e generalizar a relação entre um conjunto de entradas e saídas” (CRISTALDO, 2020).

Utilizou-se a do tipo perceptrons de múltiplas camadas, conhecida como Multilayer Perceptron (MLP), sendo o algoritmo de treinamento escolhido o Resilient Propagation.

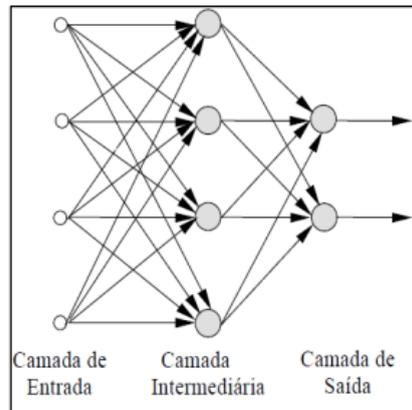
De acordo com SOUSA:

As RNA MLP, são tipos de redes que são aplicadas a diferentes problemas combinatórios e na solução de diversas tarefas, possuindo uma estrutura composta basicamente por três ou mais camadas (Figura 3), sendo i) a camada de entrada de cada rede, na qual não faz o processamento, apenas envia os dados de entrada; ii) única camada escondida em cada rede neural, composta por neurônios com função de ativação; podem ter várias linhas de pesos; iii) camada de saída, formada por neurônios com função linear (2010, v. 14, n. 2, 3p).

“Na camada de entrada são inseridos as variáveis que serão preditoras no processo, ou seja, as variáveis independentes, na camada intermediária ocorrem o processamento dos neurônios e ajustes das funções e pesos sinápticos e, na camada de saída, estão os parâmetros a serem previstos, denominados de variáveis

dependentes” (CARNEIRO et al, 2019). Como apresentado na Figura 3, a arquitetura de uma rede neural artificial.

Figura 3) Arquitetura da Rede Neural Artificial MLP.



Fonte: Haykin (2001)

A partir da Tabela 2 é possível visualizar as variáveis predictoras que são inseridas nas camadas de entrada, ocorrendo o processamento dos neurônios nas camadas intermediárias, onde são ajustadas de acordo com os pesos e funções para que na camada de saída, obtenha-se os valores das variáveis de nitrogênio e de fósforo a serem previstas.

Os pesos então foram criados para proceder ao treinamento da rede, na qual foi treinada considerando 70% do conjunto de dados obtidos pelas correlações realizadas e os demais 30% utilizados para validação do treinamento, sendo fósforo para o Modelo RAP e nitrogênio para o Modelo RAN. A escolha dos dados de cada grupo foi feita aleatoriamente pelo mesmo software utilizado anteriormente.

RESULTADOS OBTIDOS E DISCUSSÃO DOS RESULTADOS

4.1 Tratamento de dados

Inicialmente, criou-se o banco de dados utilizando o software SPSS Statistic, tendo como base 3.200 dados e em seguida analisou-se todas as variáveis por visualização em diagramas de dispersão, a fim de identificar se há uma variabilidade gradual entre os conjuntos e se essa variação é positiva ou negativa. Descartou-se todos os outliers encontrados no processo.

4.2 Análise de normalidade

Para a realização dos testes de Normalidade, foi utilizado todos os valores de entrada dos reatores como as variáveis independente como base os modelos de Kolmogorov-Smirnov (K-S) e de Shapiro-Wilk (S-W), sendo esses realizados com grau de confiança de 95% (nível de significância (α) de 5%). Ao lançar os dados dos ensaios, foram obtidos os valores apresentados na Tabela 3.

Tabela 3) Resultado dos testes de K-W e S-W pelo software.

Variável	Kolmogorov-Smirnov (K-S)			Shapiro-Wilk (S-W)		
	Statistic	sig**	Conclusão	Statistic	sig**	Conclusão
1 SSTi	0,122	0,2	Não rejeita a H0	0,915	0,09	Não rejeita a H0



2 DQO _i	0,189	0,074	Não rejeita a H ₀	0,884	0,025	Rejeita a H ₀
3 DBO _i	0,114	0,2	Não rejeita a H ₀	0,956	0,495	Não rejeita a H ₀
4 SSED _i	0,215	0,019	Rejeita a H ₀	0,863	0,011	Rejeita a H ₀
5 FÓSFORO _i	0,133	0,2	Não rejeita a H ₀	0,937	0,236	Não rejeita a H ₀
6 NITROGÊNIO _i	0,217	0,042	Rejeita a H ₀	0,874	0,017	Rejeita a H ₀
7 pH _i	0,151	0,2	Não rejeita a H ₀	0,927	0,15	Rejeita a H ₀
8 CONDi	0,201	0,021	Rejeita a H ₀	0,876	0,019	Rejeita a H ₀

*i = Índice para reator (variando de 1 a 4 reatores)

** sig = significância

*** H₀ = Normalidade

Fonte: Autoras, 2022.

Considerou-se que baixos valores de Significância (Sig < 0,05) indica que a distribuição dos dados difere significativamente de uma distribuição normal, sendo assim, os ensaios demonstraram que para o teste de S-W os parâmetros DQO, SSED, Nitrogênio, pH e COND estavam se comportando de forma *não* normal, demonstrando comportamento não linear. Enquanto SST, DBO e Fósforo estavam tendendo a linearidade.

Para os ensaios de K-S, os parâmetros de SST, DQO, DBO, Fósforo e pH se comportam de forma normal, mas para os parâmetros de SSED e Nitrogênio deve-se rejeitar a normalidade.

Os testes concordaram entre si na *não* normalidade nos parâmetros de SSED, Nitrogênio e COND; enquanto em SST, DBO e Fósforo se comportam de forma normal; mas ambos discordam nos parâmetros de: DQO e pH.

“A avaliação da normalidade da distribuição dos dados é primordial para a adequada descrição da amostra e sua análise inferencial” (NORMAN, 2014). “Muitos dados apresentam distribuição *não* normal, especialmente em eventos de grande variabilidade, com desvio padrão maior que a metade do valor médio, contraindicado o uso de técnicas estatísticas destinadas a amostras normais, sob pena de enviesamento dos parâmetros e da inferência dos testes” (SCOTTON, 2014). Sendo assim, a importância da realização da análise de normalidade, convém em garantir melhores escolhas em técnicas estatísticas para interpretação dos dados.

Os dados, em sua maioria, se adequaram ao teste de Kolmogorov-Smirnov, tendendo a *não* normalidade. O teste deve ser dedicado apenas à verificação de aderência da amostra a distribuições com outros parâmetros, visto que é superado por outros testes de normalidade. Sendo assim, foram realizados testes de correlação.

4.3 Análises de correlação

“A magnitude do efeito da correlação entre duas ou mais variáveis é representada pelos coeficientes de correlação, que assumem valores de -1 até +1, passando pelo zero (ausência de correlação). Coeficientes positivos ($r > 0$) indicam correlação direta entre as variáveis; já coeficientes negativos ($r < 0$) significam uma correlação inversa” (MIOT, 2018).

De forma geral, foram considerados os valores de MUKAKA, 2012 “sendo as correlações entre 0 e 0,3 são biologicamente desprezíveis; entre 0,31 e 0,5 são correlações fracas; entre 0,51 e 0,7 são moderadas; entre 0,71 e 0,9 são correlações fortes”. Para uma classificação menos minuciosa, utilizou-se os valores de COHEN

(1988) que tem estabelecido uma classificação baixa de: “grande” 0,50 a 1,00; “moderada” de 0,30 a 0,19 e “pequena” de 0,01 a 0,19.

Os resultados das análises de correlação dos testes de Spearman e Pearson podem ser visualizados na Tabela 4:

Tabela 4) Resultado das análises de correlação pelo software.

Variável	CSP	CPP	CSN	CPN
SST _i	-0,184	-0,041	0,655	0,619
DQO _i	0,184	0,029	-0,15	-0,381
DBO _i	-0,113	-0,188	0,556	-0,532
SSE _i	-0,129	-0,382	0,226	0,19
FÓSFORO _i	0,938	0,861	-0,304	-0,203
NITROGÊNIO _i	-0,321	-0,245	1	1
pH _i	-0,176	-0,066	-0,392	-0,292
COND _i	-0,259	-0,033	-0,753	0,217

*i = Índice para reator (variando de 1 a 4 reatores)

Fonte: Autoras, 2022.

Foi observado por meio de testes que os dados de correlação se comportaram de forma parecida em todas as entradas descritas na Figura 1 em relação às variáveis dependentes propostas. Sendo assim, foram utilizados os valores de entrada dos 4 reatores descritos na Tabela 4.

As análises de correlação mostraram-se alinhadas com os testes de normalidade, mostrando que se comportavam de forma não linear. Isso explica por que mais parâmetros foram considerados desprezíveis no teste de Pearson, uma vez que essa correlação trabalha com dados lineares.

“A escolha de Redes Neurais Artificiais para a predição de uma variável, apresenta como vantagem a não necessidade de linearidade entre as variáveis, devido a sua capacidade de absorver a não linearidade das mesmas” (BAPTISTELLA, 2005).

Visto que os dados se comportaram de forma não linear, a mesma sendo comprovada pelas correlações de Spearman as Redes Neurais se tornam uma ótima opção de modelagem para a predição de fósforo e nitrogênio.

4.4 Redes Neurais Artificiais

Para a modelagem das Redes neurais, a função de ativação utilizada pelo programa é a do tipo função tangente hiperbólica, descrita pela Equação 1:

$$\varphi_i(t + 1) = (1 - e^{-v_i(t)}) \quad (\text{Eq.1})$$

A função tangente hiperbólica preserva a forma sigmoideal mas apresenta valores no intervalo de [-1,1]. A função de ativação φ , também conhecida como função restritiva, é aplicada para restringir a amplitude da saída de um neurônio a um valor finito; Onde t é a taxa de aprendizado e $e^{-v_i(t)}$ é o sinal de entrada no neurônio. Neste trabalho a função de tangente hiperbólica foi utilizada nos neurônios da camada escondida

4.4.1 Modelo RAP

Foi realizada uma RNA para cada nível de correlação visando comparar sua eficiência em função da intensidade de correlação. Utilizou-se os seguintes níveis de correlação para o modelo de fósforo observado na Tabela 5:

Tabela 5) Descrição das RNA's para fósforo realizadas.

Níveis de correlação	Modelos
0,1	RAP1
0,2	RAP2
0,3	RAP3

Fonte: Autoras, 2022.

Modelou-se cada RNA para cada nível de correlação obtida na Tabela 5, sendo assim, utilizou-se dados de todos os reatores em função da variável dependente: Fósforo, de acordo com o seu nível de correlação. Deste modo, as variáveis excluídas exibidos na Tabela 6, retratam as variáveis descartadas em todos os reatores analisados, utilizando apenas os que possuíam correlações iguais ou superiores às analisadas. Os resultados das RNA's obtida para o Modelo RAP são apresentadas:

Tabela 6) Variáveis descartadas nos 4 reatores analisados

Modelo Rede Neural Artificial		RAP1	RAP2	RAP3
Variáveis Excluídas		DQO	SST	SST
			DQO	DQO
			DBO	DBO
			SSED	SSED
			pH	pH
			COND	COND
Somadas dos erros quadráticos	Treinamento	9,60%	4,90%	7,50%
	Validação	0,50%	1,30%	1,30%
Eficiência		90,40%	95,10 %	98,50%

Fonte: Autoras, 2022.

O algoritmo de retropropagação do erro é baseado na regra delta proposto por WINDROW E HOFF (1960), sendo por isto também chamada de regra delta generalizada. "Este algoritmo propõe uma forma de definir o erro dos nodos das camadas intermediárias, possibilitando o ajuste de seus pesos". Os ajustes dos pesos são realizados utilizando-se o método do gradiente.

Segundo GONÇALVES, 2010, "a derivação da regra delta generalizada é semelhante à derivação da regra delta". A função custo a ser minimizada é uma função de erro, definida pela soma dos erros quadráticos e representada pela Equação 2":

$$E = \frac{1}{2} \sum_p \sum_{i=1}^k (d_i^p - y_i^p)^2 \quad (\text{Eq. 2})$$

Onde: E é a medida de erro total, p é o número de padrões, k é o número de unidades de saída, di é a i-ésima saída desejada e yi é a i-ésima saída gerada pela rede. Esta equação define o erro total cometido pela rede, ou a quantidade em que, para todos os padrões de um dado conjunto, as saídas geradas pela rede diferem das saídas desejadas.

A eficiência foi calculada em cima dos erros obtidos na fase de treinamento, sendo assim, nota-se que o modelo RAP1 já se comportava positivamente obtendo um comportamento de 90,4% com a exclusão de somente uma

variável, mas ao deixar a correlação mais elevada utilizando os níveis acima de 0,3 obteve-se uma eficiência de 98,5%.

De acordo com o trabalho de NETO (2015), que possuía como variáveis de entrada no que se refere ao processo de simulação do reator UASB, DQO bruta; DQO filtrada; SST; Sólidos Suspensos Voláteis (SSV); Sólidos Suspensos Fixos (SSF); Alcalinidade total; Alcalinidade a bicarbonatos; Ácidos voláteis; pH; Fósforo; Nitrogênio Kjeldahl Total (NTK); N-Orgânico; N-amoniaco. no estudo para tratamento de esgotos utilizando RNA do tipo MLP. Foi realizada a correlação para simular um reator UASB, a fim de avaliar e estabelecer parâmetros de controle, através de Redes Neurais Artificiais (RNA), onde obtiveram uma eficiência de 91% utilizando variáveis com correlação acima de 0,9. Esses resultados podem ser observados na Tabela 7:

Tabela 7) Comparação das intensidades de correlação com a eficiência da RNA

Modelos	Correlação	Eficiência RNA
Modelo RAP1	Acima de 0,1	90,40%
Modelo RAP2	Acima de 0,2	95,10%
Modelo RAP3	Acima de 0,3	98,50%
Neto (2015)	Acima de 0,9	Superior a 91%
Junior (2000)	0,2	Média de 35,5
	0,6	80,07%

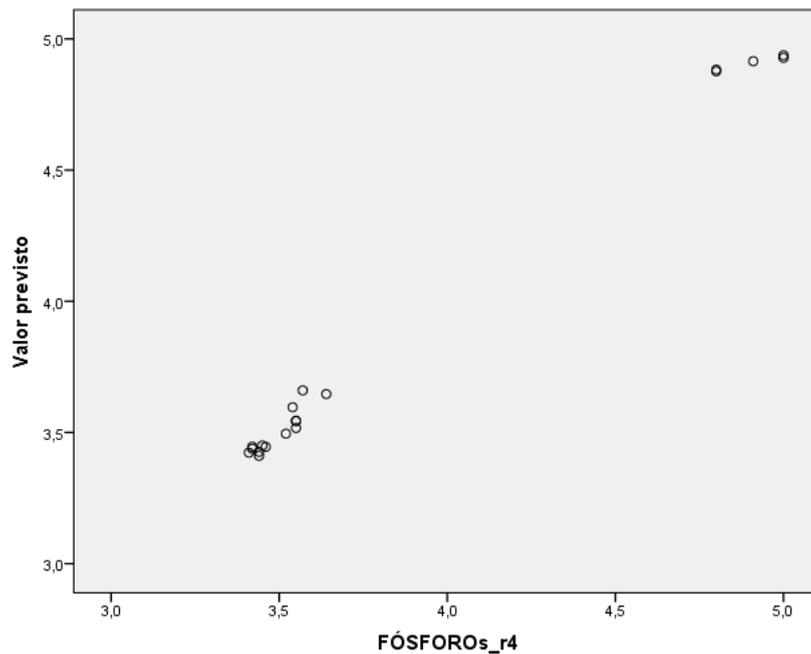
Fonte: Autoras, 2022.

Os resultados obtidos pela RAP, ressalta que o nível de correlação está associado diretamente com a eficiência da RNA, sendo que a qualidade dos conjuntos de dados impactam na eficiência final do modelo. Onde, dados consistentes e bem tratados resultam em uma rede com baixo erro e alta eficiência. Dessa forma, foi possível se destacar pois obteve-se um conjunto de dados maior e mais eficiente, mesmo com uma correlação menor do que a obtida por NETO (2015).

JUNIOR (2000) possuía entre as variáveis de entrada renda per capita das famílias, distância de viagem e número de pessoas, com o objetivo de propor desenvolvimento de modelos destinados à estimar potenciais de viagens urbanas, aspectos da acessibilidade e o impacto da mobilidade fazendo uso das Redes Neurais Artificiais. O autor obteve resultados de correlações de 0,2, ocasionando em uma eficiência média de 35,4% em suas RNA's. Mas ao utilizar correlações mais fortes de 0,6, melhorou seus modelos para uma eficiência de 80,07%. Isto indica que correlações mais fortes apontam modelos mais eficientes.

Na Figura 4, observa-se o gráfico de predição do Modelo RAP3 (na fase de testes) na qual apresentou a maior correlação, sendo o melhor modelo proposto para a predição de fósforo.

Figura 4) Gráfico de Valor Previsto de Fósforo mg/L x Valor Predito de Fósforo mg/L (RAP3).



Fonte: Autoras, 2022.

4.4.2 Modelo RAN

Para as RNA de nitrogênio, os níveis de correlação utilizados para comparação da eficiência estão descritos na Tabela 8:

Tabela 8) Descrição das RNA's para nitrogênio realizadas.

Níveis de correlação	Modelos
0,2	RAN2
0,3	RAN3
0,4	RAN4
0,5	RAN5
0,6	RAN6

Fonte: Autoras, 2022.

Nos modelos RAN foram utilizados os dados de todos os reatores em função da variável dependente: nitrogênio. Deste modo, as variáveis excluídas, exibidas na Tabela 9 retratam as variáveis descartadas em todos os reatores analisados, utilizando apenas os que possuíam correlações iguais ou superiores às analisadas.

Tabela 9) Variáveis descartadas nos 4 reatores analisados.



Modelo Rede Neural Artificial		RAN2	RAN3	RAN4	RAN5	RAN6
Variáveis Excluídas		DQO	DQO SSED	DQO SSED FÓSFORO	DQO SSED FÓSFORO pH	DQO DBO SSED FÓSFORO pH
Somadas dos erros quadráticos	Treinamento	7,30%	8,20%	1,90%	3,80%	0,80%
	Validação	11,50%	1,60%	0,60%	7,10%	1,40%
Eficiência		92,70%	91,80%	98,10%	96,20%	99,20%

Fonte: Autoras, 2022.

Para os modelos de RAN, realizou-se mais Redes Neurais pois os níveis de correlação tiveram melhor desempenho, visualizados no tópico 4.3.

As variações na eficiência das RNA's apresentadas entre as RAN3 e RAN4 e conseqüentemente entre RAN5 e RAN6, pode ser explicada analisando a soma dos erros quadráticos, na qual, conforme aumentam os erros na etapa de treinamento implica na diminuição da eficiência da RNA. Sendo desta forma, o modelo RAN6 apresenta menor erro, desta maneira possui uma alta eficiência em relação aos outros modelos.

De acordo com os trabalhos apresentados por ROCHA (2014), este utilizou a RNA para predição da turbidez da água no processo de coagulação de uma Estação de Tratamento de Água (ETA), obtendo 90% de eficiência. Na Tabela 10, é relacionado, a eficiência de todos os modelos RAN em comparação aos autores citados.

Tabela 10) Comparação da eficiência das RNA's com outros autores.

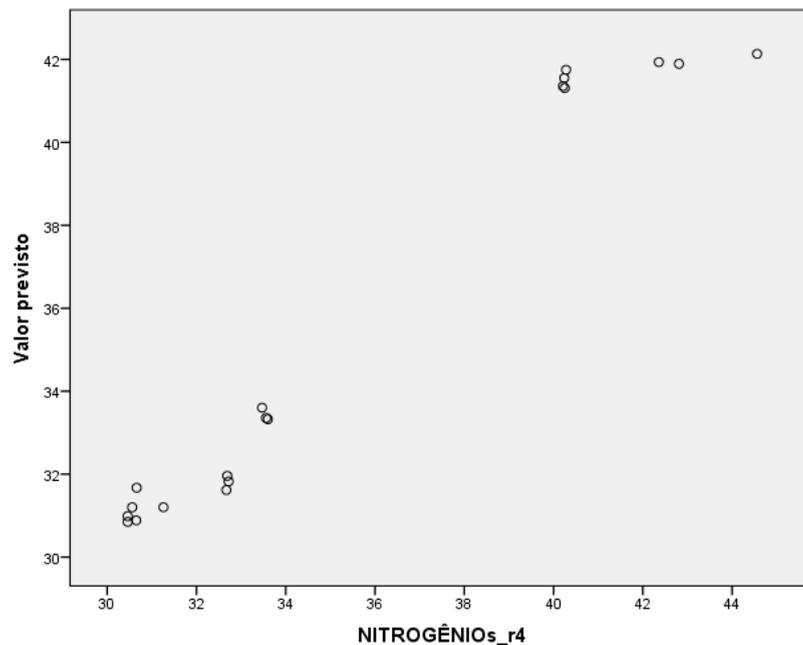
Modelo	Correlação	Eficiência RNA
RAN2	0,2	92,70%
RAN3	0,3	91,80%
RAN4	0,4	98,10%
RAN5	0,5	96,20%
RAN6	0,6	99,20%
Rocha (2014)	-	90%
Henriques et al (2018)	0,96	92%
Henriques et al (2018)	0,907	82,30%

(-) Não especificado pelo autor

Fonte: Autoras, 2022.

Observa-se a melhor eficiência da RAN6 em relação aos outros modelos realizados anteriormente, e aos trabalhos apresentados por HENRIQUES et al (2018) e ROCHA (2014). O gráfico de predição na fase de testes pode ser observado na Figura 5, ressaltando o Modelo RAN6 como o melhor modelo desenvolvido neste trabalho para a predição de nitrogênio.

Figura 5) Gráfico de Valor Previsto de Nitrogênio mg/L x Valor Predito de Nitrogênio mg/L (RAN6).



Fonte: Autoras, 2022.

CONCLUSÃO

Esse trabalho teve como objetivo comparar os modelos de Redes Neurais Artificiais (RNA) em função da intensidade de correlação para determinação do valor de predição de fósforo e nitrogênio ao final do tratamento por reatores do tipo UASB. Utilizou-se a técnica com 1 camada oculta da rede realizada no software SPSS Statistic.

Inicialmente, foi recorrido as técnicas de Normalidade, utilizando os modelos de Kolmogorov-Smirnov (K-S) e Shapiro-Wilk (S-W) visando entender o comportamento dos dados obtidos, concluindo que se comportavam de forma não normal, utilizando os cálculos de K-S.

Para analisar a intensidade de correlação dos parâmetros com as saídas estudadas, aplicou-se testes de Pearson e Spearman, visto que as correlações de Spearman se comportavam melhor por não serem lineares, comprovado anteriormente com o teste de Normalidade.

A partir dos resultados obtidos pelas análises de correlação aplicado às Redes Neurais Artificiais, observou-se que quanto mais minuciosa a correlação considerada, maior vai ser a eficiência da RNA. Sendo assim, para o Modelo RAP obteve-se uma eficiência de 90,4% quando utilizada correlações a partir de 0,1 e quando considerado somente níveis acima de 0,3, evoluiu para 98,5% de eficiência. Enquanto o Modelo RAN obteve uma correlação de 0,6 resultando em uma RNA de 99,2% de eficiência.

Analisando o Modelo RAP, conclui-se que a melhor correlação para a modelagem proposta é o modelo de RAP3, por possuir o resultado de eficiência alto e um baixo erro na soma dos erros quadráticos. Para o Modelo RAN destaca-se o modelo de RAN6, sendo ideal para predição do teor de nitrogênio presentes em efluentes, por possuir uma alta eficiência atingindo ainda um baixo erro na etapa de treinamento e testes.

Conclui-se que o modelo estatístico de Redes Neurais Artificiais pode ser utilizado como importante ferramenta na predição, configurando assim, a possibilidade de ser aplicado para predição de fósforo e nitrogênio em uma estação de tratamento de efluentes.



REFERÊNCIAS BIBLIOGRÁFICAS

1. ABNT – ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 9.648: Estudo de concepção de sistemas de esgoto sanitário. [S.n] Rio de Janeiro: ABNT, 1992
2. APHA; AWWA; WPCF. *Standard methods for the examination of water and wastewater*. 19th ed. Washington D.C.: American Public Health Association, 2016. 953p.
3. ANA. Agência Nacional de Águas. Indicadores de qualidade-índice do estado trófico. 2018. Disponível em: <http://pnqa.ana.gov.br/indicadores-estado-trofico.aspx>. Acesso em: 13 mai. 2022.
4. AFONSO, Marcela da Silva et al. Reator anaeróbio de manta de lodo (UASB): características e funcionalidade. -RS. Disponível em: https://cti.ufpel.edu.br/siepe/arquivos/2016/EN_01769.pdf. Acesso em 21 de maio de 2023.
5. BAPTISTELLA, Marisa. O uso de Redes Neurais e Regressão Linear Múltipla na engenharia de avaliações: determinação dos valores venais de imóveis urbanos. 2005. Dissertação (Mestrado em Método numéricos em Engenharia). Universidade Federal do Paraná, Curitiba, 2005.
6. CARNEIRO JÚNIOR, J. B. A. SOUZA, C. C. Aplicação de redes neurais artificiais na previsão do produto interno bruto do Mato Grosso do Sul em função da produção de cana-de-açúcar, açúcar e etanol. *Revista Ibero Americana de Ciências Ambientais*, v.10, n.5, p.218-230, 2019. DOI: <http://doi.org/10.6008/CBPC2179-6858.2019.005.0019>.
7. CARVALHO, N. L. Zabet, V. (2012). Nitrogênio: nutriente ou poluente? *Revista Eletrônica em Gestão, Educação e Tecnologia Ambiental*, 6(6), 960–974.
8. COHEN, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
9. CRISTALDO, Marcia Ferreira et al. Redes neurais artificiais aplicadas à previsão de enchentes para região do pantanal no Mato Grosso do Sul. São Paulo, UNESP, Geociências, v. 39, n. 1, p. 191 - 201, 2020. Disponível em: <file:///C:/Users/reisa/Downloads/13645-18448-1-SM.pdf>. Acesso em 21 de maio de 2023.
10. GONÇALVES, Rodrigo Mikosz et al. Modelagem preditiva de linha de costa utilizando redes neurais artificiais. *Boletim de Ciências Geodésicas* [online]. 2010, v. 16, n. 3 Acesso em 26 Novembro 2022] , pp. 420-444. Disponível em: <https://doi.org/10.1590/S1982-21702010000300004>. Epub 15 Feb 2012. ISSN 1982-2170. <https://doi.org/10.1590/S1982-21702010000300004>.
11. GUIMARÃES, Paulo Ricardo B. Análise de Correlação e medidas de associação. Disponível em: <https://docs.ufpr.br/~jomarc/correlacao.pdf>. Acesso em: 9 de maio de 2022.
12. HAYKIN, S., *Redes neurais: princípios e prática*, 2ª Ed., Porto Alegre: BOOKMAN, 2013. 696p
13. HENRIQUES, Juscelino Alves; OLIVEIRA, Mariângela Dutra de. Modelo de predição de desempenho de estações de tratamento de água de pequeno porte usando redes neurais artificiais. *Revista DAE*. v. 68, n 221 / pp 87-100 , jan/mar. 2020. Disponível em: http://revistadae.com.br/artigos/artigo_edicao_221_n_1833.pdf. Acesso em: 9 de maio de 2022.
14. JUNIOR, Archimedes Azevedo Raia. *Acessibilidade E Mobilidade Na Estimativa De Um Índice De Potencial De Viagens Utilizando Redes Neurais Artificiais E Sistemas De Informações geográficas*. 2000. Disponível em: <https://www.teses.usp.br/teses/disponiveis/18/18137/tde-10112001-160812/publico/RaiaJrT.PDF> . Acesso em: 26 de novembro de 2022.
15. LOPES, Manuela de Mesquita; CASTELO BRANCO, Verônica Teixeira Franco; SOARES, Jorge Barbosa. Utilização dos testes estatísticos de Kolmogorov-Smirnov e Shapiro-Wilk para verificação da normalidade para materiais de pavimentação. *Transportes*, São Paulo-SP, v. 21, n. 1 p. 59–66, 2013.
16. Miot, Hélio Amante. Análise de correlação em estudos clínicos e experimentais. *Jornal Vascular Brasileiro* [online]. 2018, v. 17, n. 4 [Acessado 25 Outubro 2022] , pp. 275-279. Disponível em:



- <<https://doi.org/10.1590/1677-5449.174118>>. Epub 29 Nov 2018. ISSN 1677-7301. <https://doi.org/10.1590/1677-5449.174118>.
17. Miot HA. Avaliação da normalidade dos dados em estudos clínicos e experimentais. *J Vasc Bras.* 2017 Apr-Jun;16(2):88-91. Portuguese. doi: 10.1590/1677-5449.041117. PMID: 29930631; PMCID: PMC5915855.
 18. Mukaka MM. *Statistics corner: A guide to appropriate use of correlation coefficient in medical research.* *Malawi Med J.* 2012;24(3):69-71. PMID:23638278.
 19. NETO, Pedro Nogueira Da Silva. Modelagem e Simulação de um Sistema de Tratamento de Esgoto Sanitário Usando Rede Neural Artificial. Disponível em:<<http://tede.bc.uepb.edu.br/jspui/bitstream/tede/2472/2/PDF%20-%20Pedro%20Nogueira%20da%20Silva%20Neto.pdf>>. Acesso em: 27 de outubro de 2022.
 20. Norman G, Streiner D, editores. *Biostatistics: the bare essentials*. 3. ed. Hamilton: B.C. Decker; 2014.
 21. PRADO, Marco Antônio Calil; CAMPOS, Cláudio Milton Montenegro; SILVA, Julia Ferreira da. Estudo da variação da concentração de metano no biogás produzido a partir das águas residuárias do café. *Ciênc. agrotec.* Lavras, v. 34, n. 2, p. 475- 484, Abr. 2010. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S141370542010000200029&lng=en&nrm=iso>. Acesso em 20 out. 2022.
 22. ROCHA, Valder Matos et al. 2014. Uso das redes neurais artificiais na predição da turbidez da água no processo de coagulação de uma ETA. Disponível em:<https://abrh.s3.sa-east-1.amazonaws.com/Sumarios/175/a04972368265945c530bce077315755f_863dc95399926392523bfbdccecd0d321.pdf>. Acesso em: 27 de outubro de 2022.
 23. RUTTENBERG, K. C.; CANFIELD, D. E. *Chemical distribution of phosphorus in suspended particulate matter from twelve North American Rivers: evidence for bioavailability of particulate-P.* *EOS, Trans., AGU*, v. 75, n. 3, p. 110, 2014.
 24. SILVA, Valdemir Fonseca Da. 2016. Avaliação Da Eficiência De Tanque Séptico No Tratamento De Esgoto Sanitário E Proposição De Tecnologia Complementar Por Meio De Modelo De Apoio Multicritério À Decisão. Trabalho de Conclusão de Curso (Bacharelado em Engenharia Ambiental) Universidade Federal da Fronteira Sul, Cerro Largo RS. 2016.
 25. SOUSA, Wanderson dos S. e Sousa, Francisco de A. S. de Rede neural artificial aplicada à previsão de vazão da Bacia Hidrográfica do Rio Piancó. *Revista Brasileira de Engenharia Agrícola e Ambiental [online]*. 2010, v. 14, n. 2 [Acessado 24 outubro 2022], pp.
 26. SCHINDLER, D.W. 2012. *Recent Advances in the Understanding and Management of Eutrophication.* *Limnology and Oceanography*, 51: 356-363.
 27. Scotton MF, Miot HA, Abbade LP. *Factors that influence healing of chronic venous leg ulcers: a retrospective cohort.* *An Bras Dermatol.* 2014;89(3):414-22. PMID:24937814. <http://dx.doi.org/10.1590/abd1806-4841.20142687> »
 28. UTE. Reator UASB. UFRJ. Centro Experimental de Saneamento Ambiental, Rio de Janeiro, 2016. Acessado em 20 out. 2022. Online. Disponível em: <http://www.saneamento.poli.ufrj.br/site/pt-br/reator-uasb/>
 29. VIEIRA, Sonia. Estatística básica – 2ª edição revista e ampliada: Cengage Learning Brasil, 2018. 9788522128082. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9788522128082/>. Acesso em: 09 mai. 2022.
 30. Windrow, B.; HOFF M. E. Adaptive switching circuits. Institute of Radio Engineers, Western Electronic Show and Convention, 1960 p. 96-104
 31. YU, X. et al. *Restoration in northern Lake Gehu, a eutrophic lake in China.* *Chinese Journal of Oceanology and Limnology*, v. 35, n. 6, p. 1417-1431, 2017.
 32. XU, K.H.; WANG, Z., 2014. *Two-step methodology for retro t design of cooling water networks.* *Ind. Eng. Chem. Res.* 53, 274e286. <http://dx.doi.org/10.1021/ie400906r>