

II-277 - MODELANDO A DEMANDA BIOLÓGICA DE OXIGÊNIO EFLUENTE EM UM SISTEMA DE TRATAMENTO DE ESGOTO POR LODOS ATIVADOS EMPREGANDO MACHINE LEARNING

Saulo Nonato de Souza ⁽¹⁾

Engenheiro Civil pela UFMG. Especialista em Gestão de Pessoas pelo IBMEC. Mestre em Saneamento, Meio Ambiente e Recursos Hídricos pela UFMG. Engenheiro de Operação na Unidade de Serviços de Tratamento de Esgoto Metropolitana na COPASA MG.

Gustavo de Souza Groppo ⁽²⁾

Graduado em Economia pela Universidade Federal de Viçosa (UFV), mestrado em Economia Aplicada pela Universidade de São Paulo (ESALQ/USP) e doutorado em Saneamento pela Universidade Federal de Minas Gerais (DESA/UFMG). Analista de Desenvolvimento Tecnológico da COPASA-MG

Endereço(1): Rua Mar de Espanha, 453 – Santo Antônio - Belo Horizonte - MG - CEP: 30.330-900 - Brasil - Tel: (31) 99819-7952 - e-mail: saulo.souza@copasa.com.br.

RESUMO

A capacidade de remoção de DBO em uma estação de tratamento de esgoto está relacionada à tecnologia de tratamento existente na planta e pode ser influenciada por uma série de variáveis, como as propriedades do esgoto bruto, tais como DBO afluente, pH, alcalinidade total, Demanda Química de Oxigênio afluente (DQOa), Nitrogênio Total afluente (NTa), assim como questões operacionais relacionadas à operação da planta, como o controle de oxigênio dissolvido nos tanques de aeração, ou a quantidade de lodo excedente removido do processo, entre outras. Diante desta complexidade, é fundamental que existam ferramentas que auxiliem no controle operacional de uma estação, pois a salvaguarda da saúde pública e ambiental dependem do adequado funcionamento desta. Este trabalho teve o objetivo de avaliar a influência de um conjunto de variáveis monitoradas ao longo do processo de tratamento de esgoto de uma ETE implantada com sistema de reatores UASB seguidos de lodos ativados sobre a predição de remoção do DBOe, empregando os modelos Feed Forward Neural Network (FFNN), Rede Neural Bayesiana (BRNN), Cubist (CUBIST) e Random Forest (RF). O modelo com a melhor performance foi o BRNN e as variáveis que possuem os maiores pesos na predição do valor da DBOe foram o sst_e e a dqo_e . Tal constatação é importante no acompanhamento operacional de uma estação de tratamento pois o tempo de duração para a determinação em laboratório destes parâmetros é bem inferior ao observado para a DBO, o que permite que ações corretivas sejam implementadas de forma mais ágil caso seja identificada a possibilidade de aumento das concentrações de DBO no efluente tratado da planta, contribuindo assim para uma melhor avaliação do desempenho de estações de tratamento de esgoto.

PALAVRAS-CHAVE: Tratamento de Esgoto, Controle Operacional, Modelagem matemática, Predição de DBO efluente.

INTRODUÇÃO

As Estações de Tratamento de Esgoto – ETEs são sistemas implantados e utilizados para remover os poluentes presentes nos esgotos brutos que são coletados para tratamento e disposição final. A eficiência dos processos de tratamento é estabelecida e regulada pela legislação ambiental e regulatória estabelecida em cada estado da federação, sendo que a legislação federal brasileira também deve ser considerada. Entre as questões que são analisadas na avaliação de desempenho de uma ETE está a capacidade de uma unidade remover a Demanda Biológica de Oxigênio (DBO) presente nos esgotos brutos, que é um parâmetro que indica a quantidade de oxigênio consumida por microrganismos presentes em uma certa amostra de efluente.

Adicionalmente, a capacidade de remoção de DBO está relacionada à tecnologia de tratamento existente na ETE e pode ser influenciada por uma série de variáveis, como as propriedades do esgoto bruto, tais como DBO afluente, pH, alcalinidade total, Demanda Química de Oxigênio afluente (DQOa), Nitrogênio Total afluente (NTa), assim como questões operacionais relacionadas à operação da planta, como o controle de oxigênio

dissolvido nos tanques de aeração, ou a quantidade de lodo excedente removido do processo, entre outras. Diante desta complexidade, é fundamental que existam ferramentas que auxiliem no controle operacional de uma estação, pois a salvaguarda da saúde pública e ambiental dependem do adequado funcionamento desta.

Neste contexto, a utilização de ferramentas de machine learning na predição da eficiência de ETEs pode ser fundamental para o bom desempenho das estações. Inúmeros modelos têm sido aplicados com esse objetivo. Guo et al. (2017) utilizaram a Feed Forward Neural Network (FFNN) e o Support Vector Machine (SVM) para prever a concentração do Nitrogênio Total efluente (N_{Te}). Nourani et al. (2018), empregaram as abordagens FFNN, SVM, Adaptative Neuro Fuzzy Inference System (ANFIS), Multi-Linear Regression (MLR) e três modelos de ensemble para prever o a DBO_e, o DQO_e e o N_{Te}. Mais recentemente, Wang et al. (2021) aplicaram uma Deep Neural Network (DNN) e uma Random Forest (RF) para avaliar controle de qualidade do efluente no tocante aos Sólidos Suspensos Totais (SST_e) e do Fosfato (PO_{4e}).

Assim, o objetivo deste trabalho será avaliar a influência de um conjunto de variáveis monitoradas ao longo do processo de tratamento de esgoto em sistemas UASB com lodos ativados sobre a predição de remoção do DBO_e, empregando os modelos Feed Forward Neural Network (FFNN), Rede Neural Baysiana (BRNN), Cubist (CUBIST) e Random Forest (RF).

MATERIAIS E MÉTODOS

O estudo foi realizado para a estação de tratamento de esgoto por lodos ativados Betim Central na Região Metropolitana de Belo Horizonte (RMBH). Para a análise e predição do DBO efluente utilizaram um conjunto de variáveis chaves que são monitoradas ao longo do processo de tratamento. Também foi empregada a variável pluviometria, visto que tem um peso muito grande no processo, dado que muita água de chuva é canalizada de forma indevida para as redes coletoras de esgoto, ocasionando o aumento de vazão afluente e diluição do esgoto bruto, o que impacta o processo de tratamento. A Figura 1 a seguir apresenta uma foto aérea da estação com a indicação dos pontos onde foram realizadas as coletas para as análises das variáveis consideradas no estudo, assim como na Tabela 1 são indicadas as variáveis, sua descrição e os respectivos pontos de coleta.



Figura 1 – Foto aérea ETE Betim Central com indicação dos pontos de monitoramento. Fonte: arquivo dos autores

Tabela 1: Indicação dos pontos de monitoramento onde foi coletado o esgoto para a realização das análises das variáveis consideradas no estudo

Variável	Descrição	Ponto de coleta
vaz_a	Vazão Afluente (L/s)	Medidor de vazão final
dbo_a	DBO (mg/L)	Esgoto Bruto
dqo_a	DQO (mg/L)	Esgoto Bruto
sst_a	SST (mg/L)	Esgoto Bruto
alc_a	Alcalinidade (mg/L)	Esgoto Bruto
pH_a	pH	Esgoto Bruto
ssed_a	Sólidos Sedimentáveis (mL/L)	Esgoto Bruto
temp_a	Temperatura (°C)	Esgoto Bruto
ntk_a	Nitrogênio Total (mg/L)	Esgoto Bruto
naa_a	Nitrogênio Total Amoniacal (mg/L)	Esgoto Bruto
pluv	Pluviosidade (mm chuva/mês)	Estação Pluviométrica
rel_dqo_dbo	Adimensional	-
dbo_e	DBO (mg/L)	Esgoto Tratado
dqo_e	DQO (mg/L)	Esgoto Tratado
sst_e	SST (mg/L)	Esgoto Tratado
ntk_e	Nitrogênio Total (mg/L)	Esgoto Tratado
naa_e	Nitrogênio Total Amoniacal (mg/L)	Esgoto Tratado
od_ta	Oxigênio Dissolvido (mg/L)	Tanques de aeração
sssta	Sólidos suspensos totais (mg/L)	Tanques de aeração
ssvta	Sólidos suspensos voláteis (mg/L)	Tanques de aeração

O conjunto de dados é composto por 84 observações (01/2015 a 12/2021), que representam as médias mensais para cada uma das variáveis empregadas no presente estudo. Os dados passaram por um pré processamento onde foram normalizados usando a técnica de mínimo – máximo [$x_{scaled} = x - \min(x) / \max(x) - \min(x)$], onde x_{scaled} é o valor normalizado, x é o valor atual, $\min(x)$ e $\max(x)$ compreendem o menor e o maior valor presente em cada uma das colunas, respectivamente. Este conjunto de dados foram divididos em 70% para treinar os modelos e 30% para testar os modelos e garantir uma melhor avaliação dos modelos finais.

Na predição do dbo_e foi utilizado o conjunto de covariáveis descritas acima. Todos os modelos foram implementados usando o pacote Caret do R (The R Foundation 2022).

Redes Neurais

O neurônio artificial é composto por três elementos básicos, quais sejam, os pesos das conexões ou pesos sinápticos, a junção somadora ou função aditiva e a função de ativação (Haykin, 2001). Cada sinapse é definida como sendo um estímulo de entrada multiplicado pelo seu peso sináptico. Em termos matemáticos, um neurônio k poderá ser escrito como:

$$u_k = \sum_{j=1}^m w_{kj} x_j + b_k$$

A função será calculada através da combinação linear dos pesos sinápticos e das entradas ou, equivalentemente, ao produto escalar do vetor de pesos $w^T = [w_{k1}, w_{k2}, \dots, w_{km}]$ e do vetor de entradas $x = [x_1, x_2, \dots, x_m]$.

$$u_k = w^T x + b_k$$

A saída será dada aplicando-se uma função de ativação

$$y_k = f(u_k) = f\left(\sum_{j=1}^m w_{kj} x_j + b_k\right)$$

A função de ativação do neurônio, também conhecida como função restritiva, tem o propósito proporcionar uma função contínua derivável em todo domínio em um intervalo de $[0,1]$ ou alternativamente $[-1,1]$. As funções de ativação podem ser lineares ou não-lineares, tais como função limiar ou degrau, função linear, função gaussiana, função sigmoide logística e tangente hiperbólica e função unidade linear retificadora mais conhecida como ReLU.

No presente estudo serão estudadas a RNA gaussiana (BRNN) e a Multilayer Perceptron (MLP) empregando a função sigmoide.

A rede MLP é uma das principais arquiteturas de redes neurais, a qual utiliza diversos neurônios de McCulloch-Pitts dispostos em camadas. São amplamente usadas na resolução de problemas complexos, especialmente no tocante à previsão, devido a sua capacidade de mapeamento de entrada-saída desejada. A MLP é uma rede neural, do tipo feedforward. Um dos algoritmos mais utilizados para o treinamento desta rede neural é conhecido como algoritmo de retropropagação de erro (error backpropagation), desenvolvido por Rumelhart et al. (1986). As funções radiais são uma classe de funções, que em princípio, podem ser empregadas em qualquer tipo de modelo e qualquer tipo de rede. A Rede Neural ajusta-se a uma rede neural de duas camadas [MacKay, 1992 e Foresee e Hagan, 1997] e emprega o algoritmo de Nguyen e Widrow (1990) para atribuir pesos iniciais e o algoritmo de Gauss-Newton para realizar a otimização.

As funções radiais são uma classe de funções, que em princípio, podem ser empregadas em qualquer tipo de modelo e qualquer tipo de rede. Contudo, desde o artigo seminal de Broomhead e Lowe (1988), as Redes Neurais Bayesianas (RBNN) foram tradicionalmente associadas a funções radiais em uma única rede de camadas (Orr, 1996).

Na camada de entradaenters information without any other processing, and the implicit layer is usually composed of a series of RBF são inseridas as informações sem qualquer outro processamento, e a camada implícita é geralmente composta por uma série de funções RBF. functions.The most commonly used RBF function is the Gaussian function. A função RBF mais comumente empregada é a função Gaussiana.

$$h_j(x) = \exp\left(-\|x - c_j\|^2 / r_j^2\right)$$

Na fórmula, x é o vetor de entrada, $h_j(x)$ é a saída da primeira função RBF, c_j é o centro do j nós escondidos, e r_j é o raio da RBF de ordem j . Cada nó da RBF representa uma função RBF, e cada RBF tem o centro e o raio correspondentes. A camada de saída BRNN geralmente é uma função linear. O processo de treinamento e otimização da BRNN é determinar os seguintes parâmetros da rede: O número de funções RBF, o centro c_j e raio r_j de RBF, e o peso ligação w_{kj} entre o j nó da camada oculta e o nó da saída k . Quando todos os parâmetros da rede são determinados, uma rede com a capacidade de generalização é formada, de modo que a rede possa ser usada para predição. O número de RBF afeta a capacidade de predição da BRNN em grande medida. Se o número for muito pequeno, a rede é muito simples para refletir a complexidade do sistema de pesquisa. Por outro lado, se o número de RBF for muito grande, pode causar excesso de ajuste, resultando em deterioração da capacidade de generalização do modelo.

Em geral, no processo BRNN, a primeira fase é a aprendizagem sem supervisão entre a camada de entrada e a camada oculta. A segunda fase é a aprendizagem supervisionada entre a camada oculta e a camada de saída linear.

$$R(x^q - C_i) = \exp[-(\|w1_i - X^q\|xb1_i)^2]$$

em que: $\|w1_i - X^q\|$ é a Distância Euclidiana (DE), C é o centro da função Gaussiana.

Consequentemente, a arquitetura RBNN é dada pela equação que segue abaixo:

$$Y = \sum_{j=1}^m W_{jm} \cdot \varphi$$

em que: Y valores output; φ = número de neurônios escondidos; e W = pesos (0-1).

Floresta Aleatória

As Florestas Aleatórias [Breiman, 2001] são classificadores muito flexíveis e poderosos que tem como base as árvores de decisão do tipo CART (Classification and Regression Trees), inicialmente propostas por Breiman et al. (1984). RF constrói inúmeras árvores binárias com base em amostras aleatórias bootstrapped do conjunto de dados de treinamento. O bootstrapping garante que cada árvore de decisão tenha um subconjunto diferente de dados de treinamento, fornecendo árvores de decisão exclusivas. Embora cada árvore de decisão na RF seja fraca, o resultado final classificador é poderoso, dado que esse algoritmo emprega o conceito de ensemble (comitê). Métodos de ensemble como o RF são práticos para mitigar tanto underfitting quanto overfitting [Altman e Krzywinski, 2017].

Matematicamente, o RF é regido de acordo com as equação abaixo. Define-se $\hat{C}_b(\mathbf{x})$ como a predição da b-ésima árvore, ou seja, o seu voto; e $C_1^B = \{\hat{C}_1(\mathbf{x}), \hat{C}_2(\mathbf{x}), \dots, \hat{C}_B(\mathbf{x})\}$

Portanto, tem-se que

$$\hat{C}(\mathbf{x}) = \arg \max_c \sum_{i=1}^B \delta_{c, \hat{C}_i(\mathbf{x})}$$

Sujeito à $c \in \{1, 2, \dots, m\}$,

Em que: $\delta_{i,j}$ é o delta de Kronecker, dado por:

$$\delta_{i,j} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{caso contrário} \end{cases}$$

A RF funciona bem em conjuntos de dados de grande escala, sendo capaz de lidar com dados de alta dimensão e de não precisar selecionar os recursos. Após o treinamento, pode-se dizer quais recursos são mais importantes. Durante o processo de treinamento, é possível detectar a influência mútua entre recursos, é fácil implementar a paralelização do processo de classificação, e a velocidade de treinamento é rápida.

Na modelagem de uma RF três parâmetros precisam ser especificados: o ntree, o mtry e o nodesize. O ntree define a quantidade de árvores de decisão a ser gerada, sendo este relevante pois impacta o custo computacional do modelo, dado que, quanto maior o número de árvores, maior o custo de execução. Para cada árvore, e a cada divisão dos nodos, a RF escolhe aleatoriamente o número de variáveis (mtry), que corresponde a um subconjunto das variáveis do conjunto original de dados. Como padrão de classificação [Breiman 2001] o mtry é dado por $mtry = \sqrt{p}$ para classificações de categoria binária, onde p é o número de variáveis preditoras. Já o (nodesize) diz respeito ao número mínimo de observações nos nós terminais das árvores, controlando a complexidade das árvores. O tamanho do nó menor permite árvores mais profundas e complexas, o que significa que mais divisões são realizadas até os nós terminais [Probst et al. 2019], enquanto um nó maior resulta em árvores mais rasas. Este é outro trade-off de viés-variância [Geman et al. 1992] onde as árvores mais profundas introduzem mais variância (risco de overfitting) e arvores mais rasas introduzem mais bias (risco de underfitting).

Cubist

O Cubist é um modelo de regressão orientado a predição que combina as ideias de Quinlan (1992) e Quinlan (1993), que é uma extensão do modelo M5 model tree de Quinlan (1992). Este algoritmo divide automaticamente o conjunto de dados de entrada em subconjuntos conforme um critério de decisão e constrói uma estrutura de árvore de decisão com funções de regressão linear nas pontas, ao invés de atributos numéricos contínuos. Regressões lineares são criadas assumindo que existe uma relação linear entre entradas e saídas do

modelo, no entanto a construção do M5 model tree vai na direção de não-linearidade, pois o resultado é um modelo localmente linear, porém globalmente não-linear (Solomatine et al., 2008).

O atributo utilizado para particionar os subconjuntos no M5 model tree é chamado de splitting criterion, e é baseado no desvio padrão do subconjunto em cada nó, no qual é escolhido o atributo que maximiza a redução do erro esperado, isto é, que o atributo ofereça a maior probabilidade de ter um desvio padrão mínimo. Essa função de redução do erro esperado SDR (Standart Deviation Reduction) é dada por:

$$SDR = dp(T) - \sum_{i \in \{D, E\}} \frac{|T_i|}{T} \times dp(T_i)$$

em que: T é o conjunto de exemplos cobertos pelo nó de uma árvore com o atributo escolhido; T_D e T_E são os exemplos que resultam do particionamento e dp é o desvio padrão dos valores do atributo.

Após a construção da árvore, o modelo avalia se existem muitas sub-árvores, o que poderia gerar overfitting e consequentemente baixa capacidade de generalização. A aplicação de pruning deixa o modelo mais robusto. Por fim, o processo de smoothing é aplicado para evitar descontinuidades entre as folhas das árvores, ou seja, entre os modelos de regressão linear.

Este modelo também pode usar um esquema de reforço chamado de comitê, onde o modelo iterativo cria as árvores em sequência. Diferentemente do reforço tradicional, os pesos de estágio para cada comitê não são usados para calcular a média das previsões de cada árvore modelo e a previsão final será a média simples das previsões de cada árvore modelo, onde cada árvore é cultivada onde as folhas terminais contêm modelos de regressão linear, que são baseados nos preditores usados nas divisões anteriores.

RESULTADOS E DISCUSSÕES

As correlações estatísticas entre os recursos (x) e o alvo (y) são apresentadas na Figura 2. Foi identificada uma correlação positiva bem fraca da DBOe em relação aos parâmetros na entrada da estação como o nitrogênio total e amoniacal, sólidos sedimentáveis, pH, DBO e vazão afluente a ETE. Em relação aos parâmetros pluviometria, temperatura do esgoto bruto, SST e DQO no afluente foi identificada uma correlação negativa fraca.

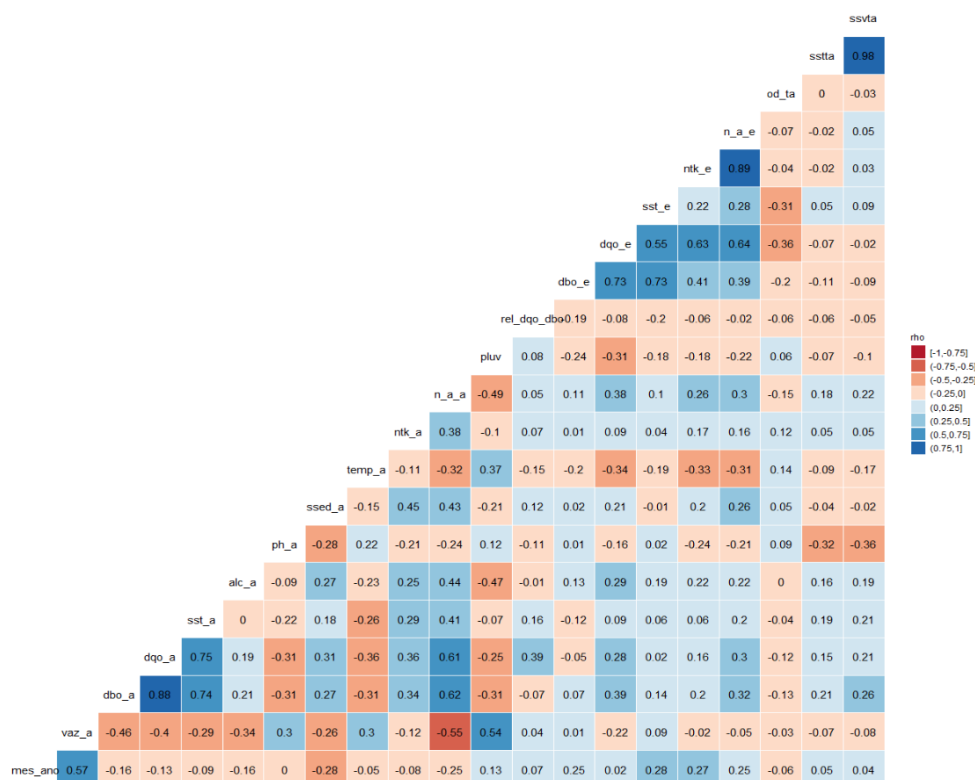


Figura 2 - Matriz de Correlação entre as variáveis empregadas na modelagem

Em relação aos parâmetros monitorados no efluente final da estação foi identificada uma correlação fraca positiva com o nitrogênio amoniacal e moderada positiva com o nitrogênio total. Enquanto que foi observada uma correlação forte positiva com os parâmetros DQO e SST. Quando analisada a correlação entre a DBOe e os parâmetros monitorados nos tanques de aeração da estação foi verificada uma correlação fraca negativa com o oxigênio dissolvido no sistema de lodos ativados, além de uma correlação bem fraca com os parâmetros sólidos suspensos totais e voláteis dessa etapa do processo.

Uma vez apresentadas as análises descritivas, utilizou-se os modelos a fim de inferir acerca da importância de cada uma das variáveis na predição do DBO efluente. A Figura 3 apresenta a importância das variáveis para cada um dos modelos avaliados.

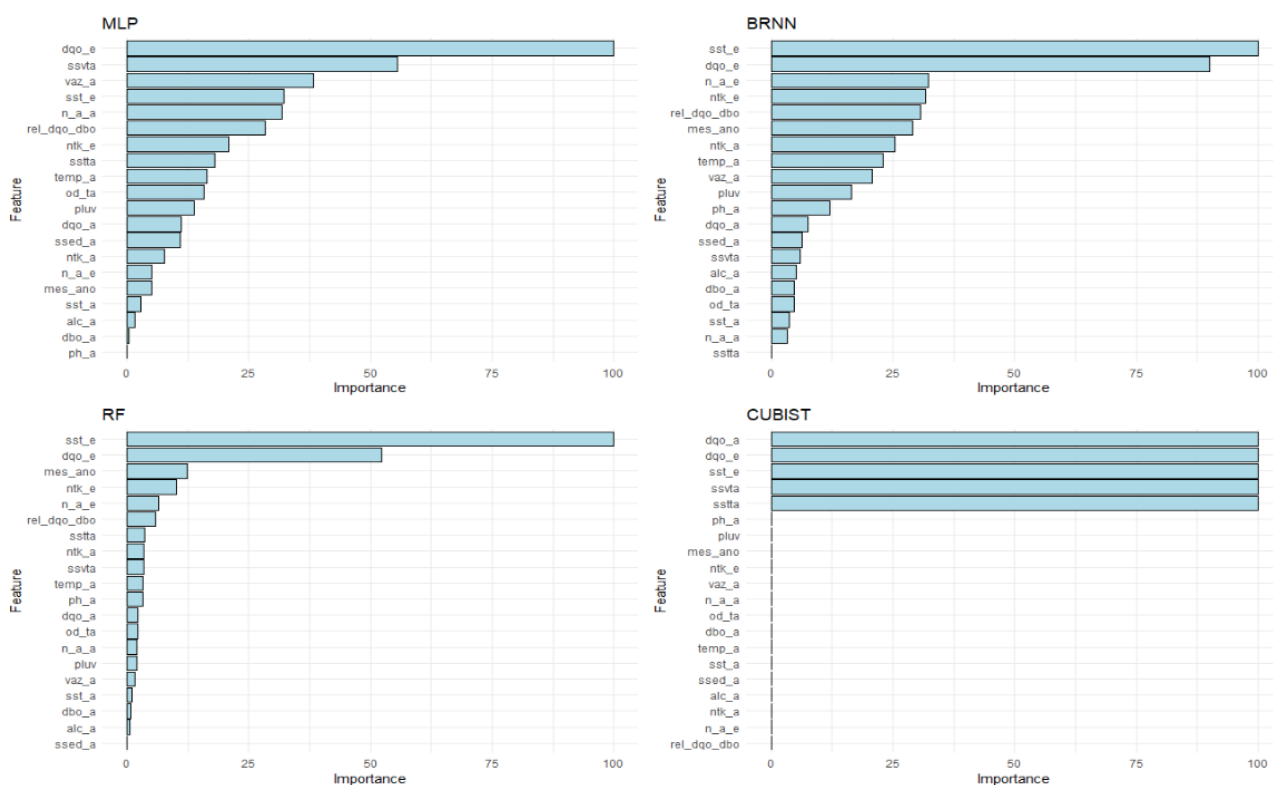


Figura 3 – Importância das variáveis

O modelo com a melhor performance foi o BRNN e as variáveis que possuem os maiores pesos na predição do valor da DBOe foram o sst_e e a dco_e. Tal constatação é importante pois indica a probabilidade dos valores de DBOe serem previstos em função dos resultados dos parâmetros SST e DQO que são monitorados no efluente da estação e que tem um tempo de realização dos ensaios em laboratório bem inferior ao de determinação da DBO, cuja duração é de cinco dias entre o início e o término do ensaio, permitindo assim maior agilidade na tomada de decisão sobre a necessidade de intervenção operacional no processo de tratamento.

Para avaliar o desempenho dos modelos em prever a DBOe, três métricas foram empregadas: coeficiente de determinação (R^2), Raiz Quadrada do Erro Médio (RMSE) e Erro Absoluto Médio (MAE). Os resultados estão apresentados na Tabela 2.

Tabela 2 – Avaliação de desempenho dos modelos empregados no estudo.

Modelo	R^2	RMSE	MAE
MLP	0,69578	0,14897	0,11063
BRNN	0,71260	0,14476	0,10549
CUBIST	0,71179	0,14850	0,10384
RF	0,51143	0,19506	0,13486

Os resultados obtidos demonstram que a BRNN obteve os melhores resultados para as métricas R^2 e RMSE com 0,71260 e 0,14476, respectivamente. O CUBIST obteve o menor valor para a métrica MAE com 0,10384, seguido pela BRNN com 0,10549.

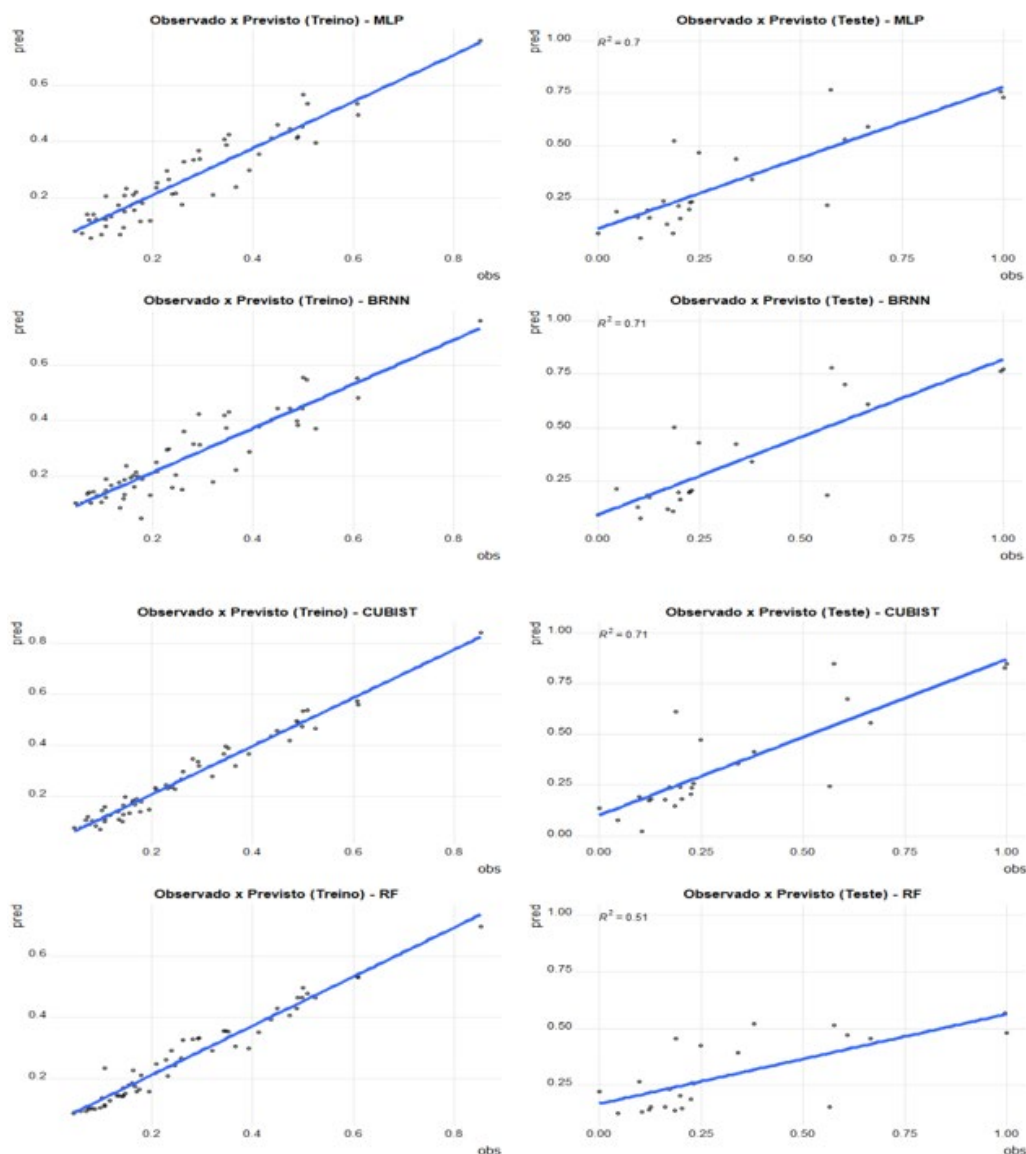


Figura 4 – DBOe previsto versus observado para o conjunto de treino e de teste para todos os modelos

CONCLUSÕES

Com o objetivo de avaliar a influência de um conjunto de variáveis monitoradas ao longo do processo de tratamento da ETE Betim Central foram utilizados os modelos Feed Forward Neural Network (FFNN), Rede Neural Baysiana (BRNN), Cubist (CUBIST) e Random Forest (RF). O modelo BRNN apresentou os melhores resultados para as métricas R2 e RMSE, assim como indicou que as variáveis *sst_e* e *dqo_e* tiveram maior peso na predição dos valores de *dbo_e*. Tal constatação é importante no acompanhamento operacional de uma estação de tratamento pois o tempo de duração para a determinação em laboratório destes parâmetros é bem inferior ao observado para a DBO, o que permite que ações corretivas sejam implementadas de forma mais ágil caso seja identificada a possibilidade de aumento das concentrações de DBO no efluente da planta, contribuindo assim para uma melhor avaliação do desempenho de estações de tratamento de esgoto.

O estudo realizado poderá contribuir para o desenvolvimento de estratégias de controle operacional mais sofisticadas em estações de tratamento de esgoto, o que poderá permitir maior confiabilidade do processo de

tratamento e menor tempo de resposta a problemas operacionais existentes nas ETEs. Trabalhos futuros serão processados considerando o incremento de mais variáveis operacionais e verificação dos modelos que apresentam melhor resultado na predição das variáveis de interesse.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Altman, N.; Krzywinski, M. Ensemble methods: bagging and random forests. *Nature Methods* 2017, 14, 933-934.
2. Breiman, L. Random Forests. *Machine Learning* 2001, 45, 5-32.
3. Breiman, L.; Friedman, J.; Olshen, R.; Stone, C. *Classification and Regression Trees*. Chapman and Hall/CRC: Wadsworth, New York, 1984.
4. Broomhead, D. S.; Lowe, D. Multivariable Functional Interpolation and Adaptive Networks. *Complex Systems*, 1988, 2, 321- 355.
5. Foresee, F. D., and M. T. Hagan. Gauss-Newton approximation to Bayesian regularization. In: *Proceedings of the 1997 International Joint Conference on Neural Networks*, 1997.
6. Geman, S.; Bienenstock, E.; Dourast, R. Neural network and the Bias/Variance Dilemma. *Neural Computation* 1992, 4, 1-58.
7. Guo, H.; Jeong, K.; Lim, J.; Jo, J.; Kim, Y. M.; Park, J. P.; Kim, J. H.; Cho, K. H. Prediction of effluent concentration in a wastewater treatment plant using machine learning models. *Journal of Environmental Sciences*, 2015, 32, 90-101.
8. Haykin, S. *Redes Neurais: princípios e prática*. Porto Alegre: Bookman, 2ª edição, 2001. 898 p.
9. MacKay, D. J. C. Bayesian interpolation, *Neural Computation*, 1992, 4, 3, 415-447.
10. Nguyen, D. and Widrow, B. Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights, *Proceedings of the 1990 International Joint Conference on Neural Networks*, 1990, 3, 21-26.
11. Nourani, V.; Elkirn, G. Abba, S. I. Wastewater treatment plant performance analysis using artificial intelligence – an ensemble approach. *Water Science & Technology*, 2018, 78, 10, 2064-2076.
12. Orr, M. J. L. Introduction to Radial Basis Function Networks. 1996, 96 p. Disponível em: < <https://www.cc.gatech.edu/~isbell/tutorials/rbf-intro.pdf> >. Consultado em: 22/11/17.
13. Probst, P.; Wright, M.N.; Boulesteix, A-L. Hyperparameters and tuning strategies for random forest. *WIREs Data Mining Knowledge Discovery* 2019, 9, 3.
14. Rumelhart, D. E.; Hinton, G. E.; Williams R. J. Learning representations by back-propagating errors. *Nature*, 1986, 323, 533-536.
15. Solomatine, D. P.; See, L. M.; Abrahart, R. J. Data-Driven Modelling: Concepts, Approaches and Experiences, in *Practical Hydroinformatics In: Computational Intelligence and Technological Developments in Water Applications*. Org. Abrahart, R. J.; See, L. M.; Solomatine, D. P., Springer, 2008, 17-32.
16. Quinlan, J. R. Learning with continuous classes. In: *Proceedings of the 5th Australian Joint Conference on Artificial Intelligence*, Hobart 16-18 November 1992, 343-348.
17. Quinlan, J. R. Combining instance-based and model-based learning. In: *Proceedings of the Tenth International Conference on Machine Learning*, 1993, 236-243.
18. Wang, D.; Thunéll, S.; Lindberg, U.; Jiang, L.; Trygg, J.; Tysklind, M.; Souihi, N. A machine learning framework to improve effluent quality control in wastewater treatment plants. *Science of the Total Environment*, 2021, 784, 147138.