

A era digital e suas implicações sociais: Desafios e contribuições

CALIBRAÇÃO MULTIVARIADA APLICADA A FLUORESCÊNCIA EEM PARA A DETERMINAÇÃO DE DBO E DQO EM EFLUENTES

Ellisson Henrique de Paulo^{1,2,3}(PG), Caroline Pereira Pimentel¹(TC), Guilherme Medeiros da Silva¹(TC), Márcia Helena Cassago Nascimento³(PQ), Otávio Arruda Heringer¹(PQ), Paulo Roberto Filgueiras³(PQ), Marco Flôres Ferrão²(PQ).

¹Tommasi Ambiental, Rua Arara Azul, 187, Novo Horizonte – 29163-306 – Serra-ES, Brasil, ellisson.hp@gmail.com, cpimentel1992@gmail.com, eng.guims@gmail.com, o.heringer@tommasiambiental.com.br.

²Universidade Federal do Rio Grande do Sul, Avenida Bento Gonçalves, 9500, Agronomia – 90650-001 – Porto Alegre-RS, Brasil, mfferrao@gmail.com.

³Universidade Federal do Espírito Santo, Avenida Fernando Ferrari, 514, Goiabeiras - 29075-910 - Vitória-ES, Brasil, marcia.cassago@gmail.com, paulo.filgueiras@ufes.br.

Resumo

Este estudo avaliou a aplicação de modelos de regressão em dados de fluorescência em matriz de excitação-emissão (EEM) para estimar a demanda bioquímica de oxigênio (DBO) e a demanda química de oxigênio (DQO) de efluente sanitário tratado (EST) e bruto (ESB). Os espectros de EEM foram modelados pela regressão por mínimos quadrados parciais (PLS), regressão por componentes principais (PCR) e regressão por vetores de suporte (SVR). Foram desenvolvidos modelos individuais para EST e ESB, além de modelos combinando as duas matrizes (EST+ESB). Os modelos PLS e SVR predominaram como melhores estratégias. Para DBO em EST, o melhor modelo foi o PLS com erro de predição (RMSEP) igual a 0,0364. Para estimar DBO em ESB, o modelo SVR teve RMSEP igual a 0,0211. Para DQO, o SVR prevaleceu tanto para EST quanto para ESB, com RMSEP igual a 0,0174 e 0,0175. Na combinação EST+ESB, ambos modelos PLS e SVR foram os melhores para estimar DBO com RMSEP iguais a 0,0318 para PLS e 0,0312 para SVR. Para DQO, o PLS apresentou RMSEP igual a 0,0285 e o SVR apresentou RMSEP igual a 0,0269.

Palavras-chave: fluorescência EEM. DBO. DQO. PLS. SVR.

Área do Conhecimento: Ciências Exatas e da Terra.

Introdução

A matéria orgânica dissolvida (DOM, do inglês *dissolved organic matter*) em corpos d'água é uma complexa mistura de compostos orgânicos de origem natural e de atividades humanas (FINDLAY, 2003). A determinação do nível de contaminação por DOM é realizada através da análise da demanda bioquímica de oxigênio de cinco dias (DBO₅) e da demanda química de oxigênio (DQO) (BAIRD, 1990). A DBO₅ mede indiretamente a matéria orgânica biodegradável, monitorando a diferença nas concentrações de oxigênio dissolvido (OD) consumido por microrganismos durante um período de incubação de 5 dias a 20°C. Já a DQO é um parâmetro que quantifica o oxigênio utilizado para oxidar a matéria orgânica e inorgânica presente na água por meio de uma reação química com um agente oxidante forte como o dicromato de potássio (K₂Cr₂O₇) ou permanganato de potássio (KMnO₄) (BAIRD, 1990). Esses parâmetros de análise são amplamente utilizados em sistemas de tratamento de efluentes, indústrias, entre outros, para avaliar a carga de matéria orgânica lançada em corpos hídricos e a qualidade geral da água. As limitações desses métodos de análise incluem o consumo de reagentes tóxicos, o longo tempo necessário para a análise, a presença de diversos poluentes nas amostras, como surfactantes, bem como a possibilidade de imprecisão devido a contaminação ou manipulação incorreta por parte do analista, entre outros (BAIRD, 1990).

A era digital e suas implicações sociais: Desafios e contribuições

Todavia, as técnicas espectroscópicas oferecem vantagens em relação aos métodos tradicionais. Essas técnicas reduzem consideravelmente o tempo necessário para a análise, eliminando etapas de incubação ou digestão da amostra. Além disso, requerem pouco ou nenhum preparo da amostra, diminuindo a necessidade de uso de reagentes e resultando em um baixo consumo amostral (MESQUITA *et al*, 2017). Dentre as principais técnicas utilizadas na análise de água está a espectroscopia de fluorescência. A fluorescência é o fenômeno em que uma molécula absorve energia da luz em um comprimento de onda e a emite quase instantaneamente em outro. Essas moléculas são chamadas de fluoróforos e possuem um espectro de excitação *versus* a intensidade da luz absorvida e um espectro de emissão *versus* a intensidade da luz emitida. A fluorescência é mais sensível do que a absorção no ultravioleta-visível (UV-Vis) porque não é afetada pela turbidez do meio. A matriz de excitação-emissão (EEM) é uma extensão da espectroscopia de fluorescência, na qual os comprimentos de onda de emissão de fluorescência são registrados em função de vários comprimentos de onda de excitação, resultando num gráfico bidimensional para cada amostra.

Quando combinada com métodos de calibração multivariada, a EEM pode ser utilizada para construir modelos de regressão capazes de estimar os valores de DBO₅ e DQO (DAI *et al*. 2022; MESQUITA *et al*, 2017). Neste estudo, os espectros de EEM foram utilizados para estimar DBO₅ e DQO em amostras de efluente sanitário tratado (EST) e efluente sanitário bruto (ESB) através da regressão por mínimos quadrados parciais (PLS, do inglês, *partial least squares*), regressão por componentes principais (PCR, do inglês, *principal component regression*) e regressão por vetores de suporte (SVR, do inglês, *support vector regression*).

Metodologia

Neste estudo foram empregadas 366 amostras de EST e 239 amostras de ESB. Os espectros de EEM foram adquiridos no espectrômetro Aqualog da Horiba. O Aqualog usa um detector de dispositivo de carga acoplada (CCD), um detector de matriz multicanal baseado em silício de luz UV, visível e infravermelha. As amostras foram filtradas usando um filtro de seringa de 0,45 µm para um volume de 4 mL de uma cubeta de quartzo de 10 x 10 mm em temperatura ambiente seguindo o método ASTM D8431 (ASTM, 2022). Os espectros EEM foram coletados em comprimentos de onda variando de 200 nm a 800 nm para excitação e de 250 nm a 800 nm para emissão usando 3,28 nm para emissão, um ganho médio, 5 nm de passa-banda, integração de 2 s para detecção de emissão e com lâmpada de arco de xenônio de 150 W montada verticalmente (GILMORE, 2013). Os espectros EEM foram corrigidos para a excitação instrumental e resposta de emissão, detector de correntes escuras, emissão de amostra em branco, efeitos de filtro interno e mascaramento da dispersão de Rayleigh de primeira e segunda ordem para cada espectro (BARHAM, 2006).

As concentrações de DBO₅ e DQO foram determinadas experimentalmente pelos métodos normatizados pela Associação Americana de Saúde Pública (APHA, do inglês *American Public Health Association*) (BAIRD, 1990). Os valores de DQO foram determinados medindo a absorção espectral em 420 nm e 620 nm usando um Fotômetro Espectral VIS Hach DR3900. A partir dos resultados de DQO foram estimados os volumes das amostras e os fatores de diluição necessários para a análise de DBO₅. Os valores de DBO₅ foram determinados medindo a concentração inicial de OD, a concentração final de OD, a concentração de OD da semente e o fator de diluição usando um oxímetro (BAIRD, 1990). A análise dos dados foi realizada por meio da versão 9.1 do software Eigenvector Solo® e do MATLAB® R2013a (The Mathworks, Natick, EUA). Os espectros EEM foram desdobrados para construir uma única matriz **X** com todas as amostras. Após isso, as amostras foram divididas em 70% para o conjunto de calibração e 30% para a predição definida pelo algoritmo de Kennard-Stone (KENNARD & STONE, 1969). A matriz **X** foi pré-processada por normalização, centragem na média, variação normal padrão (SNV, do inglês *standard normal variate*) e autoescalamento antes da modelagem de regressão. A Equação 01 foi testada nos valores de DBO₅ e DQO devido à sua grande variação, onde y_i são os valores iniciais de DQO e DBO₅ obtidos pela análise interna, $\log$ é o operador logarítmico e y_f é o valor padronizado de DQO e DBO₅ (DIAS, 2011).

$$y_f = \log(\log(y_i + 1.1)) \quad (01)$$

Os modelos foram otimizados com o método *k-fold* no procedimento de validação cruzada para seleção de variável latente (LV) (ALLEGRINI, 2016). Para avaliar qual modelo PLS, PCR e SVR apresentou melhor capacidade de ajuste e predição, foi utilizado cálculo do RMSE de calibração

A era digital e suas implicações sociais: Desafios e contribuições

(RMSEC) e predição (RMSEP). Os parâmetros foram calculados de acordo com as Equações 02 e 03, onde y_i é o valor de referência da propriedade, $\hat{y}_i$ é o valor da propriedade previsto pelo modelo, $\bar{y}_i$ é o valor médio de referência, $ncal$ é o número de amostras usadas para calibração e $npred$ é o número de amostras usadas para previsão. Além disso, coeficientes de determinação (R^2) foram calculados usando a Equação 04 para calibração (R^2_c) e predição (R^2_p) (OLIVIERI, 2014).

$$RMSEC = \sqrt{\sum_{i=1}^{ncal} \frac{(y_i - \hat{y}_i)^2}{ncal - LV - 1}} \quad (02)$$

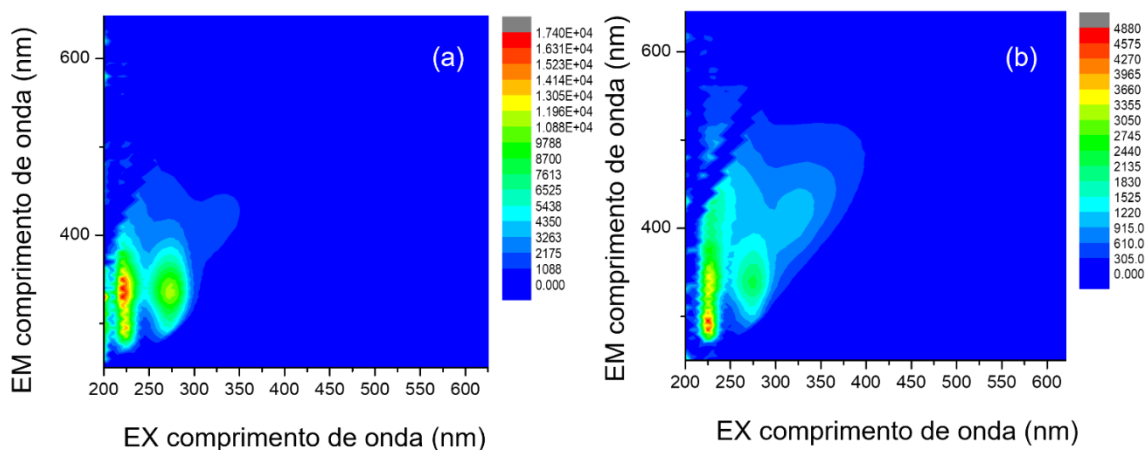
$$RMSEP = \sqrt{\sum_{i=1}^{npred} \frac{(y_i - \hat{y}_i)^2}{npred}} \quad (03)$$

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2} \quad (04)$$

Resultados

Na Figura 1 são apresentados espectros de EEM para uma amostra EST (a) e uma amostra ESB (b). Devido à enorme carga de matéria orgânica, os sinais da amostra ESB são muito mais pronunciados se comparados aos da EST. Contudo, a amostra EST usada como exemplo possui composição química muito similar a ESB. A maior intensidade de fluorescência ocorre na excitação entre 200-250 nm para ambos efluentes, mas com emissões em 350 nm para EST e 250 nm para ESB.

Figura 1 - Espectros de EEM de uma amostra de EST (a) e uma amostra de ESB (b).

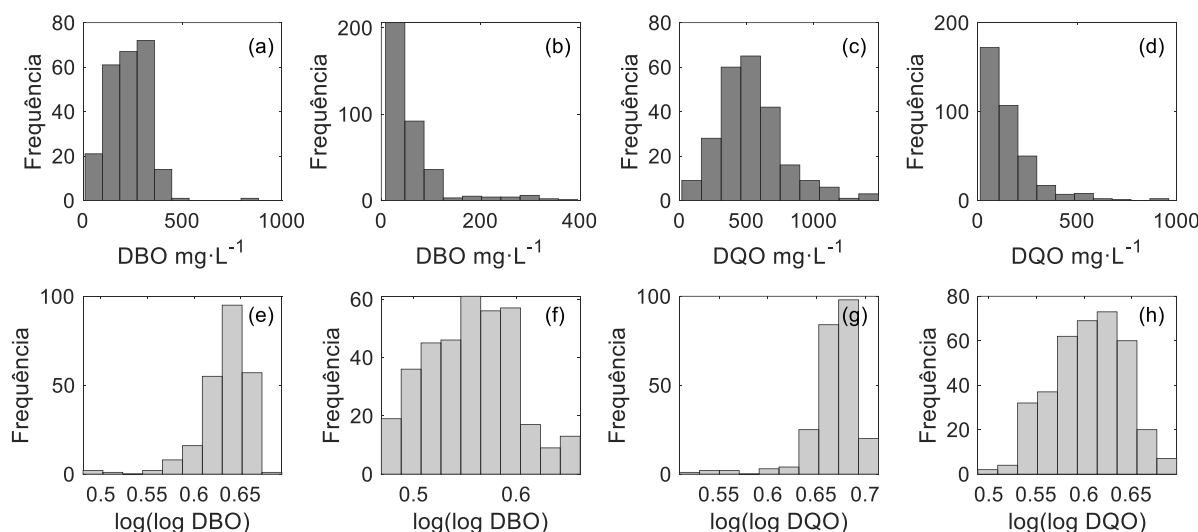


Fonte: Os autores.

Na Figura 2 tem-se a distribuição dos valores de DBO para ESB (a) e EST (b) com a maioria das amostras de ESB abaixo de 500 mg·L⁻¹ (Figura 2a) e a maioria das amostras de EST abaixo de 200 mg·L⁻¹ (Figura 2b). Devido à grande amplitude dos dados de DBO e DQO das amostras, os dados *in-house* foram tratados com logaritmo, tornando os erros dos modelos (RMSEP e RMSEC) adimensionais. O tratamento com logaritmo está nas Figuras 2(e) e 2(f) para ESB e EST, respectivamente. As concentrações adquiridas para os resultados de DQO de ESB e EST também seguem o mesmo padrão apresentado pela DBO, como está apresentado nas Figuras 2(c), 2(d), 2(g) e 2(h). A Tabela 1 traz os resultados dos modelos de regressão para EST e ESB. Individualmente, o melhor modelo para estimar DBO em EST foi o PLS com RMSEP igual a 0,0364 e R^2_p igual a 0,5478. Para estimar DBO em ESB, o modelo SVR teve RMSEP igual a 0,0211 e R^2_p igual a 0,6524. Para DQO, o SVR prevaleceu tanto para EST quanto para ESB, com RMSEP e R^2_p iguais a 0,0174 e 0,5589 para o primeiro, e 0,0175 e 0,5433 para o segundo. A Figura 3 mostra o comportamento das amostras para os melhores modelos de regressão. Na Tabela 2 estão os resultados para a combinação EST+ESB, onde ambos modelos PLS e SVR foram os melhores para estimar DBO com RMSEP e R^2_p iguais a 0,0318 e 0,7879 para PLS e RMSEP e R^2_p iguais a 0,0312 e 0,7991 para SVR. Para DQO, o PLS apresentou RMSEP e R^2_p iguais a 0,0285 e 0,7494 e o SVR apresentou RMSEP e R^2_p iguais a 0,0269 e 0,7805. A Figura 4 (a) tem-se o modelo SVR para DBO e na Figura 4 (b), o modelo SVR para DQO.

A era digital e suas implicações sociais: Desafios e contribuições

Figura 2 - Distribuição dos valores brutos de DBO em ESB (a), valores brutos de DBO em EST (b), valores brutos de DQO em ESB (c), valores brutos de DQO em EST (d), valores de DBO transformados com logaritmo para ESB (e), valores de DBO transformados com logaritmo para EST (f), valores de DQO transformados com logaritmo para ESB (g) e valores de DQO transformados com logaritmo para EST (h).



Fonte: Os autores.

Tabela 1 - Resultados dos modelos para EST e ESB.

Modelo	Prop	Pretrat	RMSEC	RMSEP	R ² _c	R ² _p
EST						
PLS(10)	DBO	SNV	0,0313	0,0364	0,7248	0,5478
PCR(4)	DBO	SNV	0,0428	0,0471	0,4611	0,2706
SVR	DBO	SNV	0,0156	0,0398	0,9289	0,4800
PLS(12)	DQO	SNV	0,0261	0,0335	0,7317	0,5189
PCR(4)	DQO	SNV/No	0,0401	0,0405	0,3322	0,2942
SVR	DQO	SNV	0,0106	0,0174	0,9131	0,5589
ESB						
PLS(10)	DBO	SNV	0,0186	0,0235	0,7112	0,5496
PCR(3)	DBO	SNV	0,0263	0,0262	0,3774	0,4475
SVR	DBO	SNV	0,0156	0,0211	0,7953	0,6524
PLS(9)	DQO	SNV	0,0174	0,0207	0,7771	0,5100
PCR(3)	DQO	SNV	0,0274	0,0238	0,4133	0,2699
SVR	DQO	SNV	0,0110	0,0175	0,9055	0,5433

Prop = Propriedade; Pretrat = Pré-tratamento, SNV = variação normal padrão, No = normalização.

Fonte: Os autores.

Tabela 2 - Resultados dos modelos para EST+ESB.

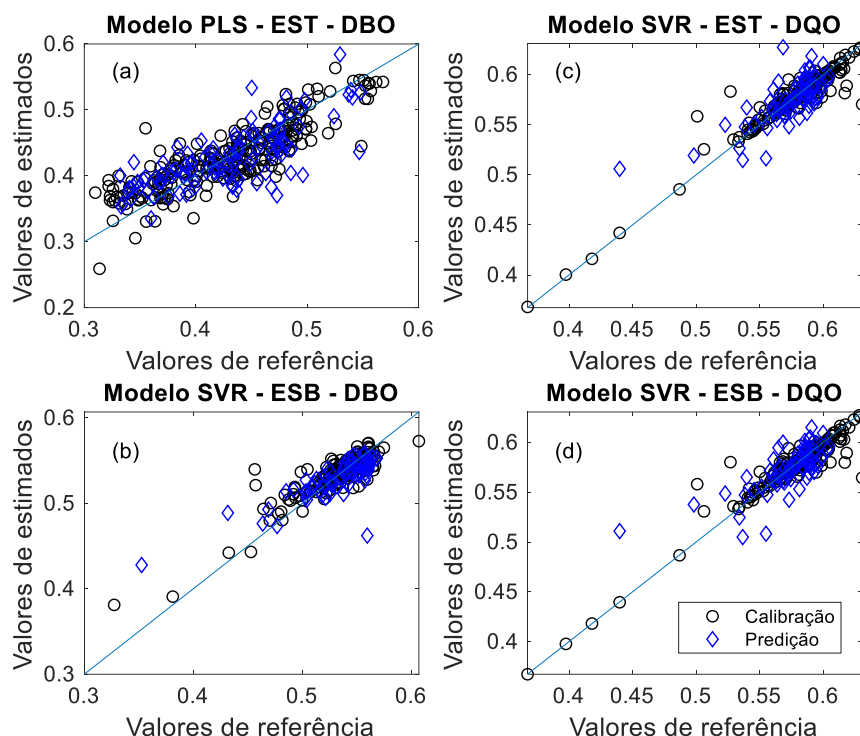
Modelo	Prop	Pretrat	RMSEC	RMSEP	R ² _c	R ² _p
PLS(13)	DBO	SNV	0,0293	0,0318	0,8343	0,7879
PCR(4)	DBO	SNV/No	0,0432	0,0391	0,6267	0,6819
SVR	DBO	SNV	0,0196	0,0312	0,9234	0,7991
PLS(12)	DQO	SNV/No	0,0259	0,0285	0,8158	0,7494
PCR(4)	DQO	SNV/No	0,0383	0,0349	0,5834	0,6251
SVR	DQO	SNV/No	0,0167	0,0269	0,9222	0,7805

Prop = Propriedade; Pretrat = Pré-tratamento, SNV = variação normal padrão, No = normalização.

Fonte: Os autores.

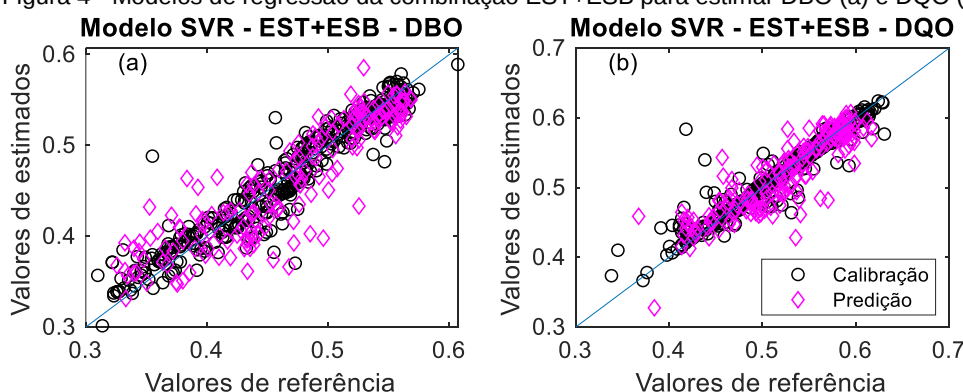
A era digital e suas implicações sociais: Desafios e contribuições

Figura 3 - Modelos de regressão para DBO em EST (a), DBO em ESB (b), DQO em EST (c) e DQO em ESB (d).



Fonte: Os autores.

Figura 4 - Modelos de regressão da combinação EST+ESB para estimar DBO (a) e DQO (b).



Fonte: Os autores.

Discussão

Apesar de EST apresentar valores próximos ao adquirido pelo bruto, sua maior distribuição amostral se restringe a uma faixa menos concentrada. Tais tendências de concentração, podem ser explicadas com as etapas que o ESB passa para que seja removida a matéria orgânica presente. De outro modo, ao analisarmos a DQO e a DBO do ESB, que não foram submetidas a nenhum tratamento, é habitual encontrarmos valores concentração de compostos orgânicos mais elevados. Além disso, a reação química na DQO pode oxidar uma variedade maior de compostos, o que resulta em valores de DQO sempre mais altos do que os da DBO₅. Apesar do elevado número de amostras tanto para EST quanto para ESB, as faixas de valores de DBO e DQO variaram muito. Isso pode justificar o porquê da

A era digital e suas implicações sociais: Desafios e contribuições

combinação EST+ESB ter gerado modelos com melhores métricas que os individuais, na qual a distribuição dos valores das propriedades se complementou. Os modelos PLS e SVR apresentaram os menores erros e os coeficientes de determinação mais elevados em relação aos modelos PCR. Tanto DBO quanto DQO demonstraram um bom ajuste linear para as amostras de calibração e de teste.

Conclusão

A metodologia utilizando fluorescência EEM e técnicas de regressão multivariadas para estimar DBO e DQO em efluentes se mostrou promissora. O tratamento com logaritmo melhorou a distribuição dos valores de DBO e DQO, o que resultou em resultados melhores para os modelos de regressão. Os métodos PLS e SVR predominaram na estimativa de DBO e DQO. Essa abordagem revelou-se como uma maneira rápida, prática e ambientalmente amigável para estimar as concentrações de DBO e DQO.

Referências

- ALLEGRI, F.; OLIVIERI, A.C. **Multi-way figures of merit in the presence of heteroscedastic and correlated instrumental noise: Unfolded partial least-squares with residual multi-linearization**, Chemometrics and Intelligent Laboratory Systems. 158 (2016) 200–209. <https://doi.org/10.1016/j.chemolab.2016.09.001>.
- ASTM International. **Standard Test Method for Detection of Water-soluble Petroleum Oils by A-TEEM Optical Spectroscopy and Multivariate Analysis 1**, (2022). <https://doi.org/10.1520/D8431-22>.
- BAHRAM, M.; BRO, R.; STEDMON, C.; AFKHAMI, A. **Handling of Rayleigh and Raman scatter for PARAFAC modeling of fluorescence data using interpolation**, J Chemom. 20 (2006) 99–105. <https://doi.org/10.1002/cem.978>.
- BAIRD, R.B.; EATON, A.D.; RICE, E.W., **Standard methods: For the examination of water and wastewater**, 1990. [https://doi.org/10.1016/0003-2697\(90\)90598-4](https://doi.org/10.1016/0003-2697(90)90598-4).
- DAI, Y.; WANG, H.; WANG, J.; WANG, X.; WANG, Z.; GE X. **Prediction of water quality based on SVR by fluorescence excitation-emission matrix and UV–Vis absorption spectrum**. Spectrochim. Acta - Part A Mol. Biomol. Spectrosc. 2022, 273.
- DIAS, J.C.M.; AGUIAR, P.F. **a Statistical Method for Acceptance of Crude Oil Viscosity-Temperature Curves**, Brazilian Journal of Petroleum and Gas. 5 (2011) 019–024. <https://doi.org/10.5419/bjpg2011-0003>.
- FINDLAY, S.E.G., SINSABAUGH, R.L. **Aquatic Ecosystems: Interactivity of Dissolved Organic Matter**, 1st Edition, Academic Press, San Diego, CA, 2003.
- GILMORE, A.M.; COHEN, S.M., **Analysis of the Chromophoric Dissolved Organic Matter in Water by EEMs with HOBIBA-Jobin Yvon Fluorescence Instrument called Aqualog**. English Edition. (2013) 19–24.
- KENNARD, RW; STONE, LA. **Computer Aided Design of Experiments**. Technometric. 1969;11:137–48. <https://doi.org/10.1080/00401706.1969.10490666>.
- MESQUITA, D.P.; QUINTELAS, C.; AMARA, A.L.; FERREIRA, EC. **Monitoring biological wastewater treatment processes: recent advances in spectroscopy applications**. Rev. Environ. Sci. Biotechnol. 2017, 16, 395–424.
- OLIVIERI, A.C. **Analytical figures of merit: From univariate to multiway calibration**, Chem Rev. 114 (2014) 5358–5378. <https://doi.org/10.1021/cr400455s>.

Agradecimentos

Tommasi Ambiental, UFRGS, UFES, CAPES, CNPq, FAPES.