



## C A P Í T U L O   1

# Modelagem Cinética e Estimação de Parâmetros na Remoção de Poluentes por Crescimento Microbiano

 <https://doi.org/10.22533/at.ed.892112621011>

### **Matheus Bastos do Carmo**

Faculdade de Biotecnologia, Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém – PA, Brasil

### **João Victor Tovany Soares da Silva**

Faculdade de Biotecnologia, Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém – PA, Brasil

### **Waldecléia Queiroz da Costa**

Faculdade de Biotecnologia, Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém – PA, Brasil

### **Miguel Fernando Saraiva Maia**

Faculdade de Engenharia Química, Instituto de Tecnologia, Universidade Federal do Pará, Belém – PA, Brasil

### **Ana Paula Souza de Sousa**

Programa de Pós-Graduação em Engenharia Química, Universidade Federal do Pará, Belém – PA, Brasil.

### **Haianny Beatriz Saraiva Lima**

Programa de Pós-Graduação em Engenharia Química, Universidade Federal do Pará, Belém – PA, Brasil.

### **Diego Cardoso Estumano**

Laboratório de Biologia Computacional, Centro de Computação de Alto Desempenho, Universidade Federal do Pará, Belém – PA, Brasil

**RESUMO:** Este capítulo apresenta a modelagem matemática e a estimativa bayesiana de parâmetros cinéticos aplicadas a um processo de remoção de poluentes por crescimento microbiano, com base no modelo proposto por Mohiuddin et al. (2022). O sistema foi descrito por equações diferenciais ordinárias que representam a dinâmica da biomassa microbiana e o consumo dos substratos carbono, nitrogênio e fósforo. O problema direto foi resolvido numericamente no MATLAB®, utilizando o solver *ode15s*, permitindo reproduzir de forma coerente o comportamento qualitativo das variáveis experimentais. A análise de sensibilidade indicou que o parâmetro associado à velocidade máxima de crescimento microbiano ( $\mu_{\text{Max}}$ ) e a constante de meia saturação do carbono ( $K_c$ ) exercem maior influência sobre as dinâmicas, enquanto os demais parâmetros apresentaram sensibilidade reduzida nas condições analisadas. Observou-se ainda dependência linear entre  $\mu_{\text{Max}}$  e  $K_c$ , inviabilizando a estimativa simultânea desses parâmetros. Com base nesses resultados, realizou-se a estimativa bayesiana de  $\mu_{\text{Max}}$  por meio do método de Monte Carlo via Cadeias de Markov, utilizando o algoritmo Metropolis–Hastings, resultando em um valor estimado de  $\mu_{\text{Max}} = 0,5879 \text{ h}^{-1}$ . A utilização do parâmetro estimado trouxe bons ajustes para a biomassa e o carbono, com coeficientes de determinação de  $R^2 = 0,9080$  e  $R^2 = 0,9446$ , respectivamente. Já para o nitrogênio e fósforo os ajustes não foram tão adequados, com coeficientes de determinação ( $R^2$ ) de 0,8513 para o nitrogênio e 0,4420 para o fósforo. Os resultados demonstram a eficácia da abordagem bayesiana na estimativa de parâmetros cinéticos e na quantificação das incertezas associadas ao modelo.

**PALAVRAS-CHAVE:** modelagem matemática; estimativa bayesiana; Monte Carlo via Cadeias de Markov; análise de sensibilidade; remoção de poluentes.

## Kinetic Modeling and Parameter Estimation in Pollutant Removal via Microbial Growth

**ABSTRACT:** This chapter presents the mathematical modeling and Bayesian parameter estimation applied to a pollutant removal process based on microbial growth, using the model proposed by Mohiuddin et al. (2022). The system is described by a set of ordinary differential equations representing the dynamics of microbial biomass and the consumption of carbon, nitrogen, and phosphorus substrates. The forward problem was numerically solved in MATLAB® using the *ode15s* solver, allowing a coherent qualitative reproduction of the experimental behavior of the system variables. Sensitivity analysis indicated that the parameter associated with the maximum microbial growth rate ( $\mu_{\text{Max}}$ ) and the carbon half-saturation constant ( $K_c$ ) exerts the greatest influence on biomass growth and carbon consumption, while the remaining parameters exhibited reduced sensitivity under the analyzed conditions.

A linear dependence between  $\mu_{\text{Max}}$  and  $K_c$  preventing their simultaneous estimation. Based on these results, Bayesian estimation of  $\mu_{\text{Max}}$  was performed using the Markov Chain Monte Carlo method with the Metropolis–Hastings algorithm, resulting in an estimated value of  $\mu_{\text{Max}} = 0,5879 \text{ h}^{-1}$ . The use of the estimated parameter led to a good agreement between experimental data and simulated curves, particularly for biomass and carbon, with coefficients of determination of  $R^2 = 0,9080$  and  $R^2 = 0,9446$ , respectively. For nitrogen and phosphorus, the obtained  $R^2$  values were 0,8513 and 0,4420, respectively, reflecting poorer adjustment for those dynamics. The results demonstrate the effectiveness of the Bayesian approach for estimating relevant kinetic parameters and for quantifying the uncertainties associated with the proposed model.

**KEYWORDS:** mathematical modeling; Bayesian estimation; Markov Chain Monte Carlo; sensitivity analysis; pollutant removal.

## INTRODUÇÃO

A preocupação com os recursos hídricos tem sido cada vez mais enfatizada ao longo do tempo devido à sua ampla utilização nos setores industrial, agropecuário e doméstico. Além do uso direto, os corpos d'água constituem ambientes que abrigam diversas comunidades biológicas e desempenham papel essencial na manutenção da vida na Terra. (LOUCKS; JIA, 2011)

Segundo o relatório *Water for Prosperity and Peace* (UNESCO, 2024), a agricultura é responsável por aproximadamente 70% das retiradas globais de água doce, seguida pela indústria (20%) e pelo uso doméstico (12%). Além disso, desde a década de 1980, a demanda mundial por água doce cresce cerca de 1% ao ano, tendência intensificada pela urbanização e pelo aumento populacional.

Nesse sentido, a poluição da água surge como resultado da industrialização, das práticas agrícolas e da insuficiência de tratamentos adequados de efluentes. Destilarias, curtumes, indústrias têxteis, de celulose, alimentícias e metalúrgicas liberam substâncias orgânicas e inorgânicas, metais, solventes e agentes tóxicos que, quando descartados sem tratamento, intensificam a contaminação dos corpos hídricos (Chaudhry; Malik, 2017; Schwarzenbach *et al.*, 2010).

Para mitigar esses impactos, diversas abordagens têm sido estudadas, entre elas processos físicos, químicos e biológicos de tratamento. A biorremediação destaca-se como alternativa promissora, utilizando microrganismos capazes de degradar compostos contaminantes por meio de sua atividade metabólica (Madigan; Brock; Martinko, 2016). Bactérias, microalgas e fungos são amplamente aplicados nesses processos (Harms; Schlosser; Wick, 2011; Khalid *et al.*, 2021)

A utilização de modelos matemáticos permite prever e quantificar a dinâmica de consumo de substrato, crescimento microbiano e formação de produtos, além de possibilitar a estimativa de parâmetros essenciais para a compreensão e otimização do processo. Modelos podem possuir parâmetros difíceis ou impossíveis de estimar utilizando observações diretas ou medidas experimentais. Devido a essa dificuldade de obter os valores dos parâmetros, utilizou-se a técnica Bayesiana de Monte Carlo via Cadeia de Markov para inferir os parâmetros do modelo. (Helcio R.B. Orlande *et al.*, 2011; Helcio R.B. Orlande; Necati Özisik, 2021; Saltelli *et al.*, 2020)

Métodos estocásticos, como o algoritmo de Monte Carlo via Cadeias de Markov (MCMC), permitem inferir parâmetros de difícil determinação analítica (Helcio R.B. Orlande *et al.*, 2011; Helcio R.B. Orlande; Necati Özisik, 2021). Este trabalho tem como objetivo simular e estimar os parâmetros relevantes do modelo cinético proposto por Mohiuddin *et al.* (2022), avaliando seu ajuste aos dados experimentais.

## METODOLOGIA

O presente trabalho tem como objetivo estimar os parâmetros presentes no modelo cinético proposto por Mohiuddin *et al.* (2022), utilizando o método de Monte Carlo via Cadeia de Markov, mais especificamente o algoritmo Metropolis-Hastings.

A primeira etapa é resolver o problema direto, calcular os resultados das equações diferenciais, dadas as condições iniciais. O problema direto foi resolvido utilizando o solver *ode15s* presente no software MATLAB.

Posteriormente, a análise de sensibilidade é resolvida. A análise de sensibilidade é a etapa que analisa quais parâmetros exercem maior influência sobre o modelo e quais são linearmente dependentes.

A estimativa de parâmetros é feita após a análise de sensibilidade, utilizando o método de Monte Carlo via Cadeia de Markov, resolvido via algoritmo de Metropolis-Hastings. A estimativa vai permitir o cálculo de valores de parâmetros presentes no modelo que seriam de difícil obtenção experimental ou cálculo analítico.

## PROBLEMA DIRETO

A modelagem do sistema de remoção de poluentes foi realizada com base no modelo cinético proposto por Mohiuddin *et al.* (2022), composto por quatro equações diferenciais ordinárias que descrevem a dinâmica da levedura e dos substratos carbono, nitrogênio e fósforo. O conjunto completo dessas equações é apresentado nas Equações 1.a-d.

$$\frac{d[Yeast_i]}{dt} = \mu_{max} \times Yeast_i \times \frac{[C]}{[C] + K_C} \times \frac{[N]}{[N] + K_N} \times \frac{[P]}{[P] + K_P} \#(1.a)$$

$$\frac{dC}{dt} = -Y_C \times \mu_{max} \times Yeast_i \times \frac{[C]}{[C] + K_C} \times \frac{[N]}{[N] + K_N} \times \frac{[P]}{[P] + K_P} \#(1.b)$$

$$\frac{dN}{dt} = -Y_N \times \mu_{max} \times Yeast_i \times \frac{[C]}{[C] + K_C} \times \frac{[N]}{[N] + K_N} \times \frac{[P]}{[P] + K_P} \#(1.c)$$

$$\frac{dP}{dt} = -Y_P \times \mu_{max} \times Yeast_i \times \frac{[C]}{[C] + K_C} \times \frac{[N]}{[N] + K_N} \times \frac{[P]}{[P] + K_P} \#(1.d)$$

Os parâmetros contidos no sistema são  $\theta = [\mu_{Max}, Yeast_i, K_C, K_N, K_P, Y_C, Y_N \text{ e } Y_P]$ . Onde  $\mu_{Max}$  representa a velocidade máxima de crescimento celular;  $Yeast_i$  representa a concentração inicial de levedura;  $K_C$ ,  $K_N$ , e  $K_P$  representam os constantes de meia saturação dos substratos: carbono, nitrogênio e fósforo, respectivamente; e  $Y_C$ ,  $Y_N$  e  $Y_P$  representam os coeficientes de rendimento do consumo de cada substrato: carbono, nitrogênio e fósforo. Ou seja, o quanto de substrato é consumido pelas células.

A solução preliminar do problema direto utilizou-se os valores de parâmetros disponibilizados no estudo original, apresentados na Tabela 1. Esses valores também foram adotados como estimativas iniciais para a etapa posterior de estimativa de parâmetros, uma vez que fornecem um ponto de partida fisicamente consistente para o método estocástico empregado.

Parâmetros	Valor	Unidade de medida
$\mu_{Max}$	0.59	1/h
$Yeast_i$	10	mg/L
$K_C$	33,8	mg/L
$K_N$	1,0021	mg/L
$K_P$	1,0061	mg/L
$Y_C$	0,897	mg[C]/mg[Levedura]
$Y_N$	0,086	mg[N]/mg[Levedura]
$Y_P$	0,013	mg[P]/mg[Levedura]

Tabela 1 – Tabela de Parâmetros

Os coeficientes de rendimento podem ser calculados utilizando as medidas finais e iniciais dos substratos e concentração celular, conforme descrito na Equação (2).

$$\frac{Y_s}{X} = \frac{(S_0 - S_f)}{(X_f - X_0)} \quad (2)$$

O sistema de equações diferenciais ordinárias foi resolvido numericamente no software MATLAB®, utilizando-se o solver *ode15s*, apropriado para sistemas considerados rígidos (*stiff*). A escolha desse método deve-se à sua estabilidade numérica em cenários onde diferentes variáveis apresentam escalas de variação distintas, características de sistemas *stiff*. (Shampine; Reichelt, 1997)

## ANÁLISE DE SENSIBILIDADE

A análise de sensibilidade foi conduzida com o objetivo de identificar quais parâmetros exercem maior influência sobre as variáveis de estado do modelo, bem como verificar a existência de possíveis dependências lineares entre eles. Essa etapa é fundamental para determinar quais parâmetros apresentam impacto significativo sobre a dinâmica do sistema e, portanto, são candidatos adequados à estimativa no problema inverso.

A sensibilidade de uma variável de estado  $Y_j$  em relação a um parâmetro  $\theta_i$  é expressa pelo coeficiente de sensibilidade apresentado na Equação (3). Esse coeficiente permite avaliar como pequenas perturbações no valor de um parâmetro afetam a resposta do modelo. Coeficientes de baixa magnitude indicam que o parâmetro exerce pouca influência sobre a variável analisada, enquanto valores elevados sugerem forte impacto sobre a dinâmica do sistema. (Helcio R.B. Orlande *et al.*, 2011; Helcio R.B. Orlande; Necati Özisik, 2021; James Vere Beck; Arnold, 1977)

$$J_{ij} = \frac{\partial Y_j}{\partial \theta_i} \quad (3)$$

Além disso, a análise dos coeficientes de sensibilidade ao longo do tempo auxilia na identificação de dependência linear entre parâmetros. A presença desse fenômeno compromete a estimativa simultânea, pois diferentes combinações paramétricas passam a produzir respostas semelhantes, tornando o problema inverso mal condicionado.

Para mitigar os efeitos de variações de escala entre parâmetros, utilizou-se o coeficiente de sensibilidade reduzido, apresentado na Equação (4). Esse coeficiente

é obtido ao multiplicar  $J_{ij}$  pelo próprio parâmetro  $\theta_j$ , permitindo comparações mais adequadas entre parâmetros com ordens de grandeza distintas. (Helcio R.B. Orlande et al., 2011; Helcio R.B. Orlande; Necati Özisik, 2021; James Vere Beck; Arnold, 1977)

$$X_{ij} = \theta_j \frac{\partial Y_j}{\partial \theta_i} \# (4) \quad (4)$$

Os coeficientes de sensibilidade foram calculados numericamente por meio do método de diferenças finitas centrais, conforme a Equação (5). Esse método foi escolhido por apresentar maior precisão em relação à diferença para frente, uma vez que utiliza perturbações simétricas em torno do ponto de avaliação. (Helcio R.B. Orlande et al., 2011; Helcio R.B. Orlande; Necati Özisik, 2021; James Vere Beck; Arnold, 1977)

$$J_{ij} \equiv \frac{Y_i(\theta_1, \theta_2, \dots, \theta_j + \varepsilon \theta_j, \dots, \theta_N) - Y_i(\theta_1, \theta_2, \dots, \theta_j - \varepsilon \theta_j, \dots, \theta_N)}{2 \varepsilon P_j} \quad (5)$$

A partir desses cálculos, obtiveram-se os coeficientes de sensibilidade reduzidos apresentados na Tabela 2, os quais serviram de base para selecionar os parâmetros estimados na etapa de problema inverso.

$X_{ij}$	Yeast	C	N	P
$\mu_{Max}$	$X_{ij} = \mu_{Maxj} \frac{\partial (\text{Yeast})_j}{\partial \mu_{Maxi}}$	$X_{ij} = \mu_{Maxj} \frac{\partial (C)_j}{\partial \mu_{Maxi}}$	$X_{ij} = \mu_{Maxj} \frac{\partial (N)_j}{\partial \mu_{Maxi}}$	$X_{ij} = \mu_{Maxj} \frac{\partial (P)_j}{\partial \mu_{Maxi}}$
$K_c$	$X_{ij} = K_{cj} \frac{\partial (\text{Yeast})_j}{\partial K_{ci}}$	$X_{ij} = K_{cj} \frac{\partial (C)_j}{\partial K_{ci}}$	$X_{ij} = K_{cj} \frac{\partial (N)_j}{\partial K_{ci}}$	$X_{ij} = K_{cj} \frac{\partial (P)_j}{\partial K_{ci}}$
$K_n$	$X_{ij} = K_{nj} \frac{\partial (\text{Yeast})_j}{\partial K_{ni}}$	$X_{ij} = K_{nj} \frac{\partial (C)_j}{\partial K_n}$	$X_{ij} = K_{nj} \frac{\partial (N)_j}{\partial K_{ni}}$	$X_{ij} = K_{nj} \frac{\partial (P)_j}{\partial K_{ni}}$

$$K_p \quad X_{ij} = K_{pj} \frac{\partial (Yeast)_j}{\partial K_{pi}} \quad X_{ij} = K_{pj} \frac{\partial (C)_j}{\partial K_{pi}} \quad X_{ij} = K_{pj} \frac{\partial (N)_j}{\partial K_{pi}} \quad X_{ij} = K_{pj} \frac{\partial (P)_j}{\partial K_{pi}}$$

Onde  $X_{ij}$  é o coeficiente de sensibilidade reduzido

Tabela 2 – Coeficientes de sensibilidade reduzidos de cada parâmetro para cada variável

## PROBLEMA INVERSO

O modelo estudado é definido por oito parâmetros,  $\theta = [\mu_{Max}, Yeast_i, K_c, K_N, K_p, Y_c, Y_N \text{ e } Y_p]$ . Embora o modelo tenha oito parâmetros, um subconjunto pode ser calculado ou obtido usando medidas experimentais das concentrações iniciais e finais de levedura, carbono, nitrogênio e fósforo,  $\theta = [Yeast_i, Y_c, Y_N \text{ e } Y_p]$ . Dessa forma, trataremos apenas o subconjunto de parâmetros que não pode ser determinado por observações diretas como desconhecido para o problema inverso,  $\theta = [\mu_{Max}, K_c, K_N, K_p]$ , aplicando o método Monte Carlo via Cadeia de Markov no mesmo.

Os algoritmos de Monte Carlo via Cadeia de Markov são empregados quando é inviável ou muito trabalhoso determinar uma solução analítica para a distribuição de probabilidade posterior, seja pela dificuldade de calcular a constante de normalização, pela forma da distribuição a priori ou a forma que a verossimilhança interage com a distribuição a priori. As Equações 6.a-d representam as distribuições posteriores do sistema.

$$\pi(\theta | Yeast_{exp}) \propto \pi(Yeast_{exp} | \theta) \pi(\theta) \quad (6.a)$$

$$\pi(\theta | C_{exp}) \propto \pi(C_{exp} | \theta) \pi(\theta) \quad (6.b)$$

$$\pi(\theta | N_{exp}) \propto \pi(N_{exp} | \theta) \pi(\theta) \quad (6.c)$$

$$\pi(\theta | P_{exp}) \propto \pi(P_{exp} | \theta) \pi(\theta) \quad (6.d)$$

Onde  $Yeast_{exp}$ ,  $C_{exp}$ ,  $N_{exp}$  e  $P_{exp}$  representam o conjunto de observações experimentais disponível e  $\theta$  representa o vetor de parâmetros já mencionado.

As distribuições a priori representam o conhecimento prévio adquirido acerca dos parâmetros do modelo, antes de incorporar os dados experimentais. Nesse trabalho adotou-se que cada parâmetro é representado por uma distribuição normal, com média e variância. A distribuição a priori é representada pela Equação 7.

$$p(\theta_j) \sim N(\mu_{0,j}, \sigma_{0,j}^2) \quad (7)$$

A função de verossimilhança simboliza a probabilidade de observar os dados experimentais dado um vetor de parâmetros específico. Neste trabalho assume-se que as medidas são independentes e que a verossimilhança segue uma distribuição gaussiana para os dados experimentais. As Equações 8.a, 8.b, 8.c e 8.d representam as verossimilhanças do sistema.

$$p(Yeast_{\text{exp}}|\theta) = \frac{1}{(2\pi\sigma_T^2)^{\frac{n_{\text{obs}}}{2}}} \exp\left(-\frac{(Yeast_{\text{exp}} - Yeast_{\text{sim}}(\theta))^T (Yeast_{\text{exp}} - Yeast_{\text{sim}}(\theta))}{2\sigma_T^2}\right) \quad (8 \text{ a})$$

$$p(C_{\text{exp}}|\theta) = \frac{1}{(2\pi\sigma_T^2)^{\frac{n_{\text{obs}}}{2}}} \exp\left(-\frac{(C_{\text{exp}} - C_{\text{sim}}(\theta))^T (C_{\text{exp}} - C_{\text{sim}}(\theta))}{2\sigma_T^2}\right) \quad (8 \text{ b})$$

$$p(N_{\text{exp}}|\theta) = \frac{1}{(2\pi\sigma_T^2)^{\frac{n_{\text{obs}}}{2}}} \exp\left(-\frac{(N_{\text{exp}} - N_{\text{sim}}(\theta))^T (N_{\text{exp}} - N_{\text{sim}}(\theta))}{2\sigma_T^2}\right) \quad (8 \text{ c})$$

$$p(P_{\text{exp}}|\theta) = \frac{1}{(2\pi\sigma_T^2)^{\frac{n_{\text{obs}}}{2}}} \exp\left(-\frac{(P_{\text{exp}} - P_{\text{sim}}(\theta))^T (P_{\text{exp}} - P_{\text{sim}}(\theta))}{2\sigma_T^2}\right) \quad (8 \text{ d})$$

Onde  $Yeast_{\text{exp}}$ ,  $C_{\text{exp}}$ ,  $N_{\text{exp}}$  e  $P_{\text{exp}}$  representam as observações experimentais,  $Yeast_{\text{sim}}$ ,  $C_{\text{sim}}$ ,  $N_{\text{sim}}$  e  $P_{\text{sim}}$  representam as curvas simuladas, geradas por MCMC,  $n_{\text{obs}}$  é o número de observações e  $\sigma_T^2$  é a variância associada à incerteza da medida.

A implementação do método MCMC foi feita através do uso do algoritmo de Metropolis-Hastings (Hastings, 1970; Metropolis *et al.*, 1953). O algoritmo usa critério de aceitação-rejeição, que segue os seguintes passos (Davila *et al.*, 2023a; Davila *et al.*, 2023b; Nunes *et al.*, 2021; Oliveira *et al.*, 2020; Sousa *et al.*, 2025; Toffoli *et al.*, 2023):

- O número de estados da cadeia de Markov (n) é selecionado, indicando o número de simulações;
- O contador de iterações é iniciado ( $i=0$ ) e um valor inicial  $\theta^0$  é selecionado;
- Um valor candidato  $\theta^*$  é gerado a partir da distribuição  $p(\theta^*|\theta^{(i)})$ , obtido a partir da Equação 9, mostrada a seguir;

$$\theta^* = \theta^{(i)}(1 + w\epsilon) \quad (9)$$

Onde  $w$  é o passo de procura e é uma variável aleatória proveniente de uma distribuição normal  $N(0,1)$

- A probabilidade de aceitação  $\alpha(\theta^*|\theta^{(i)})$  do valor candidato é obtida usando a equação de Hastings, representada pela Equação 10;

$$\alpha(\theta^*|\theta^{(i)}) = \min \left[ 1, \frac{\pi(\theta^*|y_{\text{sim}}) p(\theta^{(i)}|\theta^*)}{\pi(\theta^{(i)}|y_{\text{sim}}) p(\theta^*|\theta^{(i)})} \right] \quad (10)$$

Onde  $\pi(\theta^*|y_{\text{sim}})$  é a distribuição de probabilidade a posteriori

- Uma amostra auxiliar aleatória é gerada a partir de uma distribuição uniforme  $u \sim U(0,1)$ .
- Se  $u \leq \alpha(\theta^{(i)}|\theta^*)$ , o novo valor é aceito e  $P^{(i+1)} = P^*$ . Caso  $u \geq \alpha(\theta^{(i)}|\theta^*)$ , o valor é rejeitado e  $P^{(i+1)} = P^i$ .
- O contador é acrescido de  $i$  até  $i+1$ , e retorna ao passo 3.

A definição dos parâmetros a serem estimados depende dos resultados obtidos na análise de sensibilidade, onde serão estimados apenas os parâmetros que possuem magnitude para influenciar o modelo e parâmetros que não sejam linearmente dependentes, evitando um problema mal posto.

## MÉTRICA DE COMPARAÇÃO DE MEDIDA SIMULADA E EXPERIMENTAL

O coeficiente de determinação, também conhecido como  $R^2$  é utilizado para avaliar a qualidade do ajuste entre os dados experimentais e as curvas simuladas pelo modelo. É das métricas mais utilizadas na análise de desempenho de modelos

aplicados a dados discretos. Essa métrica expressa o quanto da variabilidade observada é explicada pelo modelo.

O cálculo de R<sup>2</sup> depende inicialmente da determinação da soma total dos quadrados ( $SQ_{Total}$ ) e da soma dos quadrados dos resíduos ( $SQ_{Resíduos}$ ), definidas pelas Equações (11) e (12). A primeira quantifica a variabilidade total dos dados experimentais em torno de sua média, enquanto a segunda representa a variabilidade não explicada pelo modelo, isto é, o erro entre valores medidos e valores simulados. (Gujarati, 2000; Mann, 2021; Montgomery; Peck; G Geoffrey Vining, 2013)

$$SQ_{Total} = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (11)$$

$$SQ_{Resíduos} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (12)$$

Onde  $y_i$  é o valor observado, experimental;  $\hat{y}_i$  é o valor simulado, previsto;  $\bar{y}$  é a média das observações e  $n$  é o número de observações.

A partir dessas quantidades, o coeficiente de determinação é obtido pela Equação (13), cuja formulação varia entre 0 e 1. Valores próximos de 1 indicam forte concordância entre modelo e dados experimentais, enquanto valores mais baixos sinalizam inconsistências ou limitações na representação do sistema (Gujarati, 2000; Mann, 2021; Montgomery; Peck; G Geoffrey Vining, 2013).

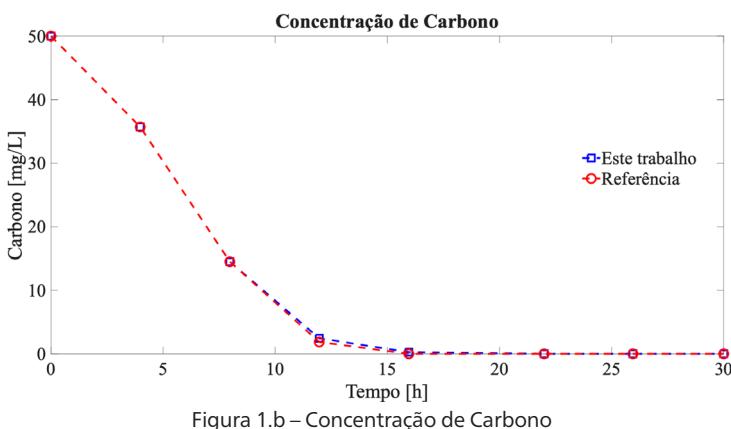
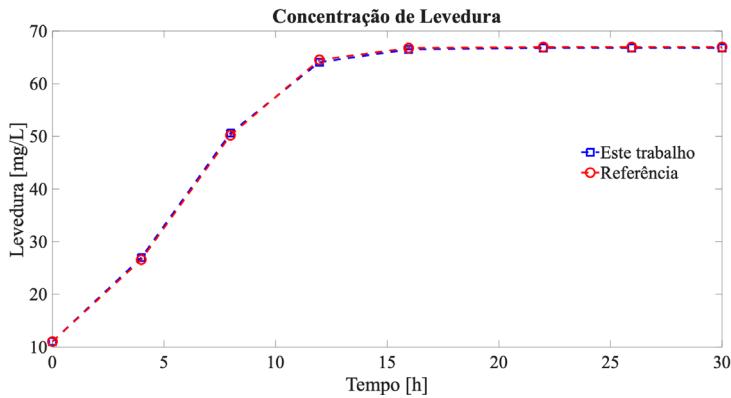
$$R^2 = 1 - \frac{SQ_{Resíduos}}{SQ_{Total}} \quad (13)$$

As métricas de ajuste foram calculadas para cada uma das variáveis do modelo (levedura, carbono, nitrogênio e fósforo), permitindo identificar quais dinâmicas foram mais adequadamente representadas pelo modelo ajustado.

## RESULTADOS

### PROBLEMA DIRETO

A solução do problema direto permitiu simular a dinâmica da levedura e dos substratos durante o processo experimental. As Figuras 1.a, 1.b e 1.c apresentam as curvas obtidas a partir do conjunto de parâmetros iniciais fornecido por Mohiuddin et al. (2022) reproduzido na Tabela 1 e a verificação da solução do modelo.



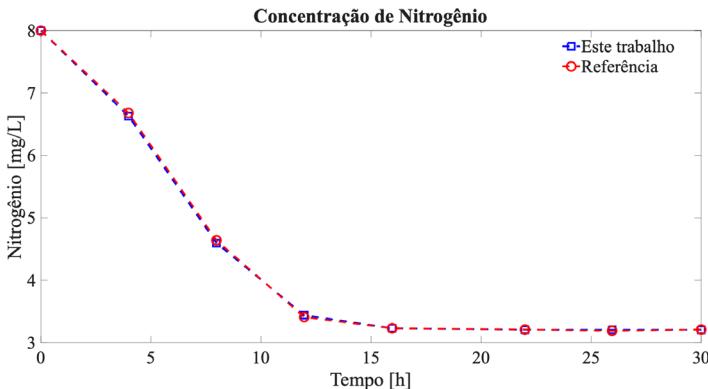


Figura 1.c – Concentração de Nitrogênio

A comparação dos resultados é referente a etapa de verificação do modelo, onde é comparada a solução deste trabalho com a solução da referência. Dessa forma, as curvas dos dois trabalhos tiveram dinâmicas similares. Em ambas as resoluções, o regime permanente foi atingido por volta das 15 horas, quando a concentração de levedura e dos substratos permaneceu aproximadamente constante.

A concentração de levedura apresenta crescimento ao longo do tempo, conforme esperado para sistemas em que há oferta de nutrientes essenciais. O aumento da biomassa está associado ao consumo progressivo do carbono, nitrogênio e fósforo, observado pela redução contínua das concentrações desses substratos ao longo da simulação. Esse comportamento é compatível com a fisiologia microbiana e indica que o modelo foi capaz de reproduzir adequadamente a tendência geral dos dados experimentais (Mcneil; Harvey, 2008, P. 103–108; Mohiuddin *et al.*, 2024; Zhang; Elser, 2017).

A dinâmica nas simulações diretas forneceu uma base importante para as etapas subsequentes, permitindo avaliar a sensibilidade do modelo e justificar a seleção dos parâmetros a serem ajustados pela abordagem bayesiana. Além disso, essas simulações foram essenciais para definir bons valores iniciais para o algoritmo MCMC.

## ANÁLISE DE SENSIBILIDADE

As Figuras 2.a, 2.b, 2.c e 2.d apresentam os coeficientes de sensibilidade reduzidos calculados para cada parâmetro do modelo. Esses coeficientes permitem avaliar a influência relativa de cada parâmetro sobre as variáveis de estado, considerando adequadamente suas diferentes escalas.

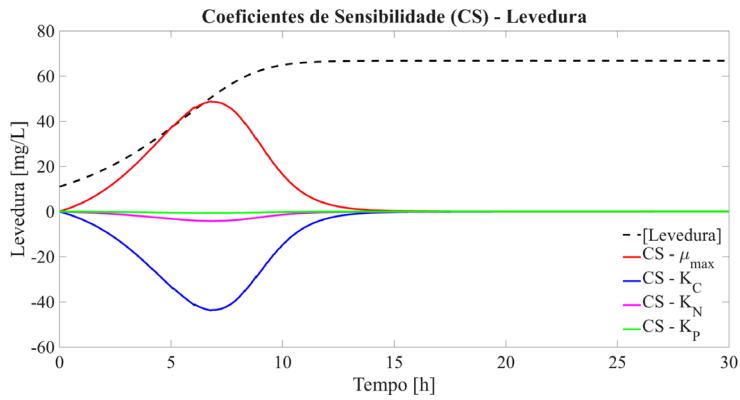


Figura 2.a – Coeficiente de Sensibilidade (Levedura)

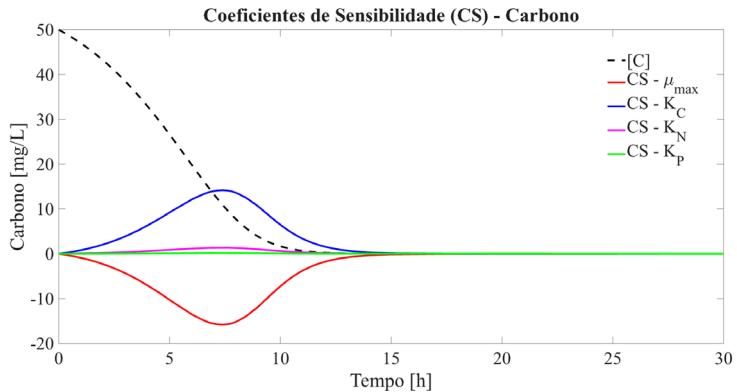


Figura 2.b – Coeficiente de Sensibilidade (Carbono)

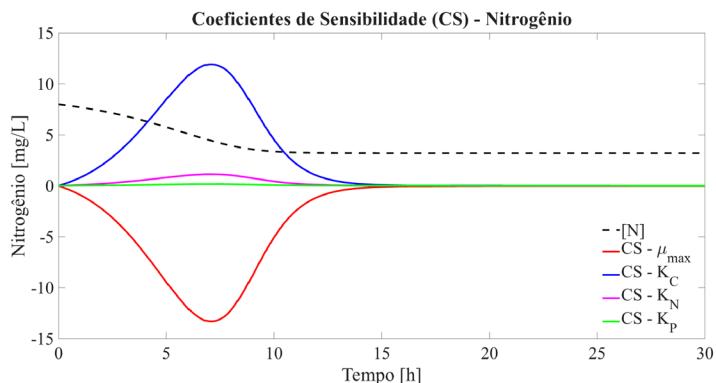


Figura 2.c – Coeficiente de Sensibilidade (Nitrogênio)

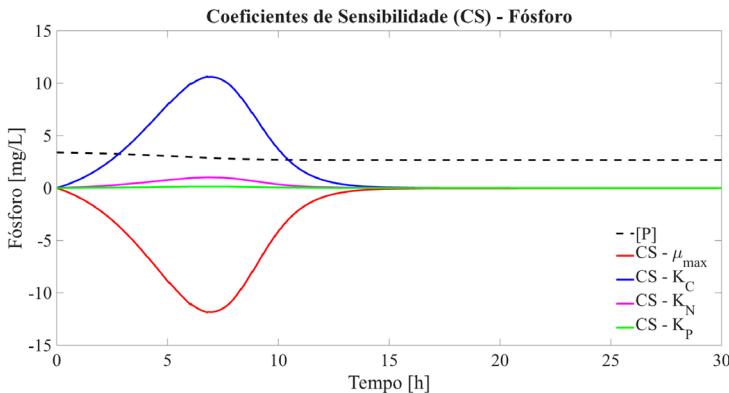


Figura 2.d – Coeficiente de Sensibilidade (Fósforo)

Os resultados indicam que apenas os parâmetros  $\mu_{\max}$  e  $K_c$  apresentam sensibilidade expressiva sobre as curvas de biomassa e de carbono, especialmente nos primeiros instantes da simulação, quando a taxa de crescimento celular é mais pronunciada. Tal dinâmica é coerente com a formulação do modelo, uma vez que ambos os parâmetros estão diretamente associados à cinética de crescimento microbiano (Arnaldos et al., 2015; Monod, 1949).

Entretanto, observa-se que as curvas de sensibilidade reduzida desses dois parâmetros apresentam formas muito semelhantes ao longo do tempo. Essa similaridade indica a existência de dependência linear entre eles: mudanças simultâneas em  $\mu_{\max}$  e  $K_c$  podem produzir efeitos quase indistinguíveis nas saídas do modelo. Arnaldos et al. (2015) reforçam que a estrutura matemática da função de Monod impede a diferenciação clara entre a afinidade pelo substrato e a velocidade máxima de crescimento a partir de dados experimentais convencionais.

A presença dessa dependência compromete a estimativa simultânea dos dois parâmetros, pois diferentes combinações podem gerar respostas praticamente equivalentes, tornando o problema inverso mal condicionado.

Os demais parâmetros exibem coeficientes de sensibilidade reduzidos próximos de zero para todas as variáveis analisadas, o que implica influência limitada sobre a dinâmica do sistema nas condições estudadas. Dessa forma, estimativas nesses parâmetros não resultariam em melhorias significativas na capacidade preditiva do modelo. (Helcio R.B. Orlande et al., 2011; Helcio R.B. Orlande; Necati Özisik, 2021; Moura et al., 2021)

Com base nesses resultados, conclui-se que somente  $\mu_{\text{Max}}$  apresenta sensibilidade suficiente e independência estrutural para ser estimado de forma confiável. Esse resultado orientou a escolha desse parâmetro como foco da etapa de estimação bayesiana apresentada na Seção 3.3.

## PROBLEMAS INVERSOS

A Figura 3 apresenta o *traceplot* da cadeia gerada para o parâmetro  $\mu_{\text{Max}}$  ao longo das 10.000 iterações do algoritmo Metropolis–Hastings. Observa-se que, após aproximadamente 600 iterações, a cadeia atinge uma região de comportamento estacionário, caracterizada pela flutuação em torno de um nível médio estável. Esse comportamento indica que o período de aquecimento (*burn-in*) foi adequadamente removido e que a cadeia passou a amostrar a distribuição posterior do parâmetro.

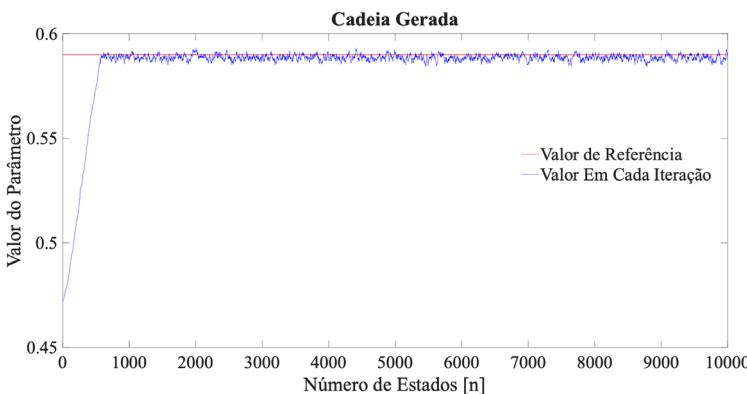


Figura 3 – Traceplot da Cadeia de  $\mu_{\text{Max}}$

A convergência observada não decorre de propriedade inerente à cadeia de Markov, mas do fato de que, após número suficiente de iterações, a distribuição dos estados converge para a distribuição posterior. Essa estabilidade temporal é um indicativo de que as amostras restantes representam apropriadamente a distribuição de interesse (Helcio R.B. Orlande *et al.*, 2011; Helcio R.B. Orlande; Necati Özisik, 2021; James Vere Beck; Arnold, 1977; Moura *et al.*, 2021).

A Figura 4 apresenta o histograma das amostras da distribuição de probabilidade a posteriori, juntamente com a estimativa da densidade posterior. A distribuição obtida apresenta forma unimodal e concentração elevada em torno de valores próximos ao reportado por Mohiuddin *et al.* (2022).

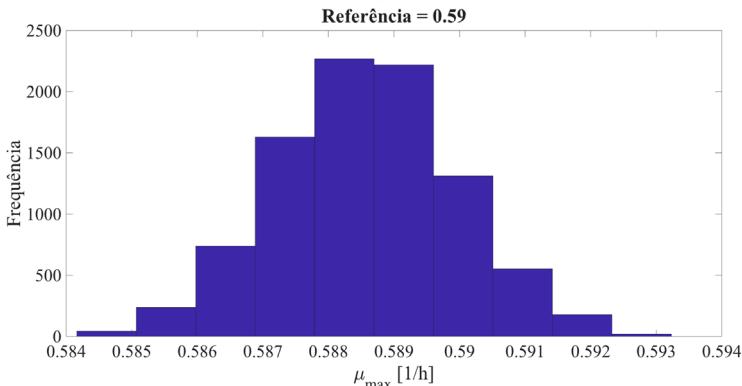


Figura 4 – Histograma da distribuição a posteriori de  $\mu_{\max}$

A média de obtida para o parâmetro, dentro de um intervalo de credibilidade de 99%, foi de:  $\mu_{\max} = 0,5879 \text{ h}^{-1}$ .

O intervalo de credibilidade de 99% foi obtido diretamente a partir dos quantis da distribuição posterior, evitando suposições adicionais de normalidade. Esse intervalo fornece uma estimativa da incerteza associada ao parâmetro e permite avaliar a confiabilidade da inferência.

As Figuras 5.a-d apresentam as comparações entre os dados experimentais, que estão legendados como “Medida” nos gráficos, e as curvas simuladas utilizando o valor estimado. Observa-se melhora no ajuste para a variável carbono, especialmente na fase de consumo mais acentuado. Para a biomassa, a curva ajustada aproxima-se mais dos pontos experimentais nas fases intermediárias da simulação. Apesar disso, alguns desvios permanecem, indicando que ajustes adicionais — envolvendo outros parâmetros ou modelos alternativos — poderiam ser investigados em estudos futuros.

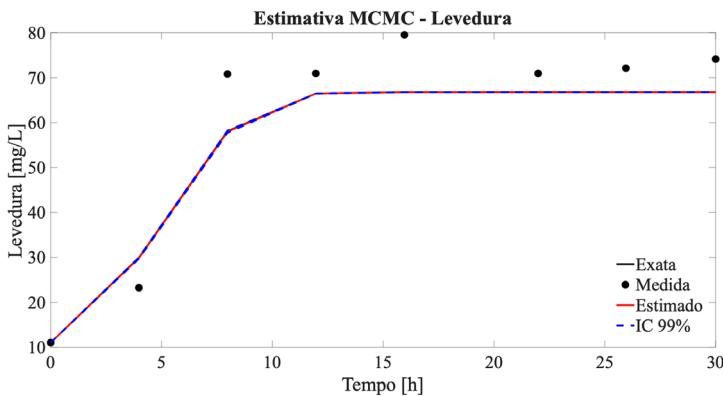


Figura 5.a – Estimativa MCMC (Levedura)

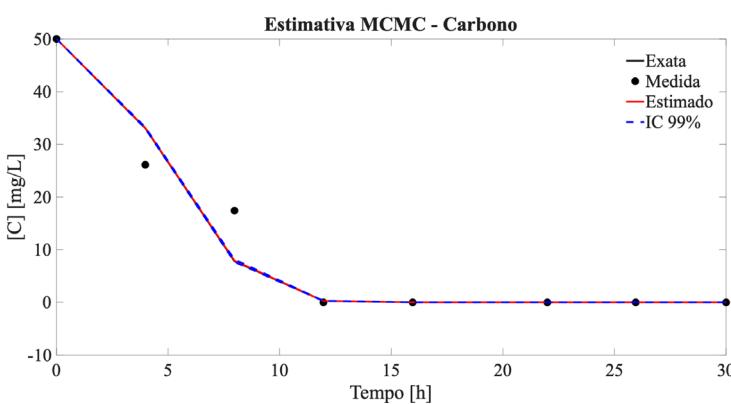


Figura 5.b – Estimativa MCMC (Carbono)

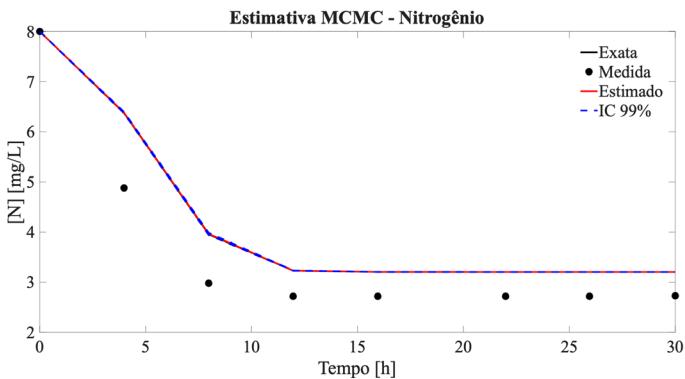


Figura 5.c – Estimativa MCMC (Nitrogênio)

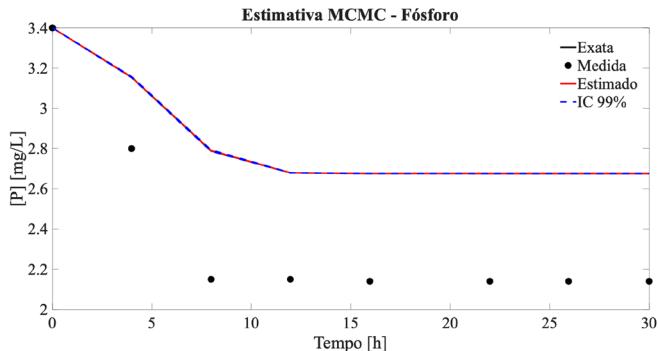


Figura 5.d – Estimativa MCMC (Fósforo)

A Tabela 3 apresenta os valores de  $R^2$  calculados para cada variável do modelo utilizando o valor estimado de  $\mu_{Max}$ . De modo geral, os resultados confirmam o desempenho satisfatório do modelo, embora com variações entre as diferentes variáveis analisadas.

Dinâmica	$R^2$
Levedura	0,9080
Carbono	0,9446
Nitrogênio	0,8513
Fósforo	0,4420

Tabela 3 – Coeficientes de determinação ( $R^2$ )

A estimativa bayesiana de parâmetros, conduzida via algoritmo Metropolis-Hastings, resultou em um valor médio para a velocidade máxima de crescimento de  $\mu_{Max} = 0,5879 \text{ h}^{-1}$ . A aplicação deste parâmetro ao modelo cinético permitiu a simulação das dinâmicas de consumo de substratos e crescimento da biomassa, cujos desempenhos foram avaliados por meio do coeficiente de determinação ( $R^2$ ).

Os resultados indicaram um ajuste satisfatório para a biomassa da levedura ( $R^2 = 0,9080$ ) e excelente para o carbono orgânico ( $R^2 = 0,9446$ ). A literatura aponta que o modelo de Monod, base desta modelagem, apresenta alta robustez para descrever a remoção de substratos carbonáceos, uma vez que a relação entre o consumo de energia e o incremento de biomassa é mais direta nesses sistemas (Muloiwa; Nyende-Byakika; Dinka, 2020).

Contudo, observou-se uma redução na precisão do ajuste para o nitrogênio ( $R^2 = 0,8513$ ) e, de forma mais acentuada, para o fósforo ( $R^2 = 0,4420$ ). No caso do nitrogênio, a dispersão dos dados em relação ao modelo pode ser atribuída à complexidade metabólica. Diferente de processos puramente assimilativos, leveduras como a *Candida tropicalis* possuem vias para nitrificação e desnitrificação simultâneas, o que introduz mecanismos de remoção não capturados integralmente por uma cinética de crescimento simples (He et al., 2021).

O baixo índice de ajuste para o fósforo pode estar relacionado a acúmulo ou inadequação do modelo. Conforme discutido por Dong et al. (2022), sistemas baseados em leveduras frequentemente apresentam o fenômeno de “consumo de luxo” (*luxury uptake*), onde o fósforo é acumulado intracelularmente na forma de polifosfato de maneira desacoplada da taxa de crescimento instantânea. Essa dinâmica sugere que a utilização de modelos de cota celular, como o de Droop, poderia ser mais adequada para descrever a remoção de nutrientes específicos que sofrem estocagem intracelular (Lee; Jalalizadeh; Zhang, 2015).

## CONCLUSÕES

O presente estudo analisou a dinâmica de crescimento microbiano e o consumo de substratos em um sistema de remoção de poluentes por meio do modelo cinético proposto por Mohiuddin et al. (2022), empregando técnicas de modelagem matemática e inferência bayesiana para a estimativa de parâmetros.

A solução do problema direto permitiu reproduzir de forma coerente o comportamento qualitativo das variáveis experimentais, evidenciando crescimento da biomassa e consumo progressivo de carbono, nitrogênio e fósforo. No entanto, discrepâncias quantitativas entre as curvas simuladas e os dados experimentais indicaram a necessidade de ajuste paramétrico para aprimorar a capacidade preditiva do modelo.

A análise de sensibilidade demonstrou que o parâmetro  $\mu_{\text{Max}}$  apresenta a maior influência sobre a dinâmica do sistema, especialmente sobre a biomassa e o carbono, enquanto os demais parâmetros exibiram sensibilidade reduzida nas condições estudadas. Além disso, a dependência linear observada entre  $\mu_{\text{Max}}$  e  $K_c$  inviabilizou a estimativa simultânea desses parâmetros, justificando a escolha de  $\mu_{\text{Max}}$  como único parâmetro a ser estimado no problema inverso.

A estimação bayesiana por meio do método de Monte Carlo via Cadeias de Markov, utilizando o algoritmo Metropolis–Hastings, resultou em um valor estimado de  $\mu_{\text{Max}} = 0,5879 \text{ h}^{-1}$ , obtido a partir da média da distribuição posterior, com intervalo de credibilidade de 99% determinado diretamente a partir dos quantis da posterior. A análise do *traceplot* indicou comportamento estacionário da cadeia após o período de aquecimento, evidenciando a convergência adequada do algoritmo e a confiabilidade das amostras geradas.

A utilização do valor estimado de  $\mu_{\text{Max}}$  proporcionou melhora no ajuste das curvas simuladas em relação aos dados experimentais, especialmente para a biomassa e o carbono. Os valores do coeficiente de determinação confirmaram esse comportamento, resultando em  $R^2 = 0,9080$  para a levedura e  $R^2 = 0,9446$  para o carbono. Para o nitrogênio e fósforo os valores de  $R^2$  foram de 0,8513 e 0,4420, respectivamente, refletindo um menor ajuste dessas dinâmicas aos dados experimentais.

Os resultados obtidos demonstram que a abordagem bayesiana adotada foi eficaz na estimativa de parâmetros cinéticos relevantes, permitindo não apenas a obtenção de um valor consistente com os dados experimentais, mas também a quantificação explícita da incerteza associada à estimativa. Apesar das limitações impostas pela dependência linear entre parâmetros, o estudo evidencia o potencial do uso de métodos estocásticos, como o MCMC, na análise e interpretação de sistemas biológicos complexos.

## REFERÊNCIAS

ARNALDOS, M. et al. From the Affinity Constant to the half-saturation index: Understanding Conventional Modeling Concepts in Novel Wastewater Treatment Processes. **Water Research**, v. 70, p. 458–470, mar. 2015.

CHAUDHRY, F.; MALIK, M. Factors Affecting Water pollution: a Review. **Journal of Ecosystem & Ecography**, v. 07, n. 01, 2017.

DAVILA, I. et al. Phosphate Removal Using Dolomite Modified with ultrasound: Mathematical and Experimental Analysis. **Journal of Environmental Science and Health Part A**, v. 58, n. 5, p. 469–482, 2 abr. 2023a.

DAVILA, V. et al. Fixed-bed Column for Phosphate Adsorption Combining Experimental observation, Mathematical simulation, and statistics: Classical and Bayesian. **Separation and purification technology**, v. 317, p. 123914–123914, 1 jul. 2023b.

DONG, X. et al. Advanced Simultaneous Nitrogen and Phosphorus Removal for Non-Sterile Wastewater through a Novel Coupled Yeast-Sludge System: Performance, Microbial Interaction, and Mechanism. [S. l.: s. n.], **SSRN Electronic Journal**, 2021.

GUJARATI, D. N. **Econometria Básica**. São Paulo: Makron Books, 2000.

HARMS, H.; SCHLOSSER, D.; WICK, L. Y. Untapped potential: Exploiting Fungi in Bioremediation of Hazardous Chemicals. **Nature Reviews Microbiology**, v. 9, n. 3, p. 177–192, 7 fev. 2011.

HASTINGS, W. K. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. **Biometrika**, v. 57, n. 1, p. 97–109, 1 abr. 1970.

HE, Y. et al. Candida Tropicalis Prompted Effectively Simultaneous Removal of Carbon, Nitrogen and Phosphorus in Activated Sludge Reactor: Microbial Community Succession and Functional Characteristics. [S. l.: s. n.], 2021. **SSRN Electronic Journal**.

HELCIO R.B. ORLANDE et al. **Thermal Measurements and Inverse Techniques**. [s.l.] CRC Press, 2011.

HELCIO R.B. ORLANDE; NECATI ÖZSIK, M. **Inverse Heat Transfer**. 2. ed. 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742: CRC Press, 2021.

JAMES VERE BECK; ARNOLD, K. J. **Parameter Estimation in Engineering and Science**. [s.l.] James Beck, 1977.

KHALID, F. E. et al. Bioremediation of Diesel Contaminated Marine Water by Bacteria: a Review and Bibliometric Analysis. **Journal of Marine Science and Engineering**, v. 9, n. 2, p. 155, 3 fev. 2021.

LEE, E.; JALALIZADEH, M.; ZHANG, Q. Growth Kinetic Models for Microalgae cultivation: a Review. **Algal Research**, v. 12, p. 497–512, nov. 2015.

LOUCKS, D. P.; JIA, H. Managing Water for Life. **Frontiers of Environmental Science & Engineering**, v. 6, n. 2, p. 255–264, 29 set. 2011.

MADIGAN, M. T.; BROCK, T. D.; MARTINKO, J. M. **Brock Biology of Microorganisms**. 14. ed. Boston: Pearson, 2016. p. 653–657

MANN, P. S. **Introductory statistics**. Hoboken, Nj: Wiley, 2021.

MCNEIL, B.; HARVEY, L. **Practical Fermentation Technology**. [s.l.] John Wiley & Sons, 2008. p. 103–108

METROPOLIS, N. et al. Equation of State Calculations by Fast Computing Machines. **The Journal of Chemical Physics**, v. 21, n. 6, p. 1087–1092, jun. 1953.

MOHIUDDIN, O. et al. Kinetic Modelling of Yeast Growth and Pollutant Removal in Secondary Effluent. **Journal of Water Process Engineering**, v. 50, p. 103244, dez. 2022.

MOHIUDDIN, O. et al. Bioremediation of Waste by Yeast Strains. **Electronic Journal of Biotechnology**, v. 69, p. 30–42, maio 2024.

MONOD, J. The Growth of Bacterial Cultures. **Annual Review of Microbiology**, v. 3, n. 1, p. 371–394, out. 1949.

MONTGOMERY, D. C.; PECK, E. A.; G GEOFFREY VINING. **Introduction to Linear Regression analysis**. Oxford: Wiley-Blackwell, 2013.

MOURA, C. et al. Parameter Estimation in Population Balance through Bayesian Technique Markov Chain Monte Carlo. **J. Appl. Comput. Mech**, v. 7, n. 2, p. 890–901, 2021.

MULOIWA, M.; NYENDE-BYAKIKA, S.; DINKA, M. **Comparison of Unstructured Kinetic Bacterial Growth Models**. [s.l.: s.n.]. Acesso em: 19 jan. 2026.

NUNES, K. et al. Evaluation of Zinc Adsorption through Batch and Continuous Scale Applying Bayesian Technique for Estimate Parameters and Select Model. **Journal of Environmental Science and Health Part A**, v. 56, n. 11, p. 1228–1242, 19 set. 2021.

OLIVEIRA, R. et al. Cr (VI) Adsorption in Batch and Continuous scale: a Mathematical and Experimental Approach for Operational Parameters Prediction. **Environmental technology & innovation**, v. 20, p. 101092–101092, 1 nov. 2020.

SALTELLI, A. et al. Five Ways to Ensure That Models Serve society: a Manifesto. **Nature**, v. 582, n. 7813, p. 482–484, 1 jun. 2020.

SHAMPINE, L. F.; REICHELT, M. W. The MATLAB ODE Suite. **SIAM Journal on Scientific Computing**, v. 18, n. 1, p. 1–22, jan. 1997.

SOUZA, A. P. et al. A Proposal for the Incorporation of Temperature into the Langmuir Model in Adsorption Isotherms. **Processes**, v. 13, n. 12, p. 4036–4036, 13 dez. 2025.

TOFFOLI, J. et al. Bayesian Computational Technique for Modeling Caffeine Adsorption in a Fixed-Bed Column: Use of the Maximum Adsorption Capacity Deterministically and Experimental Design. **Industrial & Engineering Chemistry Research**, v. 62, n. 18, p. 7127–7137, 28 abr. 2023.

UNESCO. **Water for Prosperity and Peace Water for Prosperity and Peace Facts, Figures and Action Examples the United Nations World Water Development Report 2024**. [s.l.: s.n.]. Acesso em: 13 jan. 2026.

ZHANG, J.; ELSEY, J. J. Carbon:Nitrogen:Phosphorus Stoichiometry in Fungi: A Meta-Analysis. **Frontiers in Microbiology**, v. 8, 14 jul. 2017.