ESTIMATION OF PARAMETERS IN BIOREACTOR DYNAMICS MODELS

J. P. Castro Júnior.[a]
E. C. Rodrigues[b]
M.B. Carmo[a]
D.S. Costa[c]
D. C. Estumano[a]

[a]University Federal of Pará
Faculty of Biotechnology
ZIP CODE. 66075-110, Belém, Pará, Brasil
dcestumano@ufpa.br

[b]University Federal of Pará
Faculty of Chemistry Engineering
ZIP CODE. 66075-110, Belém, Pará, Brasil
ecr@ufpa.br

[c]University Federal of Pará
Faculty of Materials Engineering
ZIP CODE. 66075-110, Belém, Pará, Brasil
deibsonsc@yahoo.com.br

Received: Nov 16, 2023
Reviewed: Dec 04, 2023
Accepted: Dec 12, 2023

ABSTRACT

Bioreactors are applied in the production of various products and the analysis of parameters that describe the production/consumption kinetics of the species becomes important to be able to design the bioreactors, as well as through mathematical models to be able to carry out simulations that make it possible to infer the concentration of species in scenarios where there is no experimental data. In this context, this article shows the application of Bayesian techniques (Monte Carlo Via Markov Chain-MCMC) to estimate both parameters and state variables in which there are no experimental measurements. The application was carried out using a model that has as state variables substrate (S), product (P) and biomass (X) using the Monod model as the kinetic model. The estimates obtained had good accuracy and precision in the evaluated scenario.

Keyword: biorreactors, MCMC, Monod

NOMENCLATURE

X	Biomass
S	Substrate
P	Product
F	feed rate
V	volume of reactor
K_S	Half of maximum Substrate concentration
J	reduced sensibility coefficient
w	search step

Greek symbols

ε_p	perturbação
ε	random variable N(0,1)
θ	vector of parameters
σ^2_{meas}	variânce of synthetic measurement
α	probability of acceptance of Metropolis-Hastings
θ^*	candidate vector vetor

Subscripts

t	time
j	counter of parameters parâmetros
N_P	total of parameter
T_{end}	final time
T	transpose

1. INTRODUCTION

The evaluation of the kinetics of biomass, product and substrate dynamics involves analyzing the concentration evolution of one or more components of a bioreactor. It comprises components: Microorganism (biomass), nutrients (substrate), metabolites (product). Which are generally represented by X, S and P in mathematical models (Andrews 1968; Miller and Block, 2020; Pradhan et al., 2016 Himmi et al., 2000; Stowers et al. 2014).

Mathematical models are useful for simulating scenarios in which experimental data is not available. In this sense, in addition to estimating in different experimental conditions, the model can also assist in research related to scaling up.

In the model explored in this article, Monod kinetics was used, which is widely used (Wang and Wan, 2009; Zhao et al., 2003; Monod, 1949). The

mathematical models are composed of the initial biomass value (X), substrate-to-product conversion factor (Yp/s), substrate-to-cell conversion factor (Yx/s and specific cell growth rate (μx). However, for Before having the complete model, it is still necessary to have a function for μx (kinetic model). Kinetic models are generally represented by a system of ordinary and coupled differential equations that describe the reactions and interactions between the elements of the reaction (Miller and Block, 2020).

One difficulty encountered is determining the parameters relating to the mathematical model. One way to make inferences about unknown parameters is to apply statistical techniques to estimates. In this article, the Bayesian Monte Carlo technique via Markov Chain was used with the Metropolis-Hastings acceptance/rejection algorithm.

2. DIRECT MODEL

Mathematical modeling in bioreactors depends on the configuration in which such modeling will be applied. In this article, a model that can be applied in both batch and continuous feeding (F) will be discussed. The direct model is represented by a system of four coupled differential equations (Equations 1.a-d) obtained from the mass balances applied to determine the biomass, substrate and product state variables (X, S and P). In batch fermentations, the must feed flow rate (F) was equal to zero (Marinho et al. 2018). E.D.O system from mass balance:

$$\frac{dX}{dt} = \left(\mu - \frac{F}{V}\right) X \quad (1.a)$$

$$\frac{dS}{dt} = (C_{SM} - S) * \frac{F}{V} - \frac{1}{Y_{VX\_S}} \mu X \quad (1.b)$$

$$\frac{dP}{dt} = \frac{Y_{E\_S}}{Y_{VX\_S}} \mu X - \frac{F}{V} P \quad (1.c)$$

$$\frac{dV}{dt} = F \quad (1.d)$$

Although there are several models for describing growth kinetics, this article adopted the classic Monod model presented in equation 1.e (Monod, 1949).

$$\mu = \mu_{max} \frac{S}{Ks + S} \quad (1.e)$$

where μ_{max} represents half speed and K_S corresponds when the variable S is equal to half of its maximum.

3 INVERSE PROBLEM

3.1 SENSITIVITY ANALYSIS

Before solving the inverse problem, it is necessary to analyze the reduced sensitivity coefficients of the model parameters. This analysis allows evaluating which parameters can be successfully estimated since the low magnitude of sensitivity or the presence of linear dependence between the parameters points to the existence of several solutions for the same problem (Orlande et al, 2011, Naveira-Cotta, 2009). This would directly interfere with the estimate, as the problem is classified as ill-conditioned. Therefore, it is desirable that the parameters present high sensitivity and are not linearly dependent on each other in order to obtain an accurate estimate of them.

The partial derivatives of the state variable that have measures $\mathbf{Y^T}$= (S,X,P), measured over time $\mathbf{t} = \{1, \ldots, t_{end}\}$, relative to known parameters $\boldsymbol{\theta}_\mathbf{j}$ for $\mathbf{j} = \{1, \ldots, N_P\}$ calculated by centered finite differences can be used to determine the reduced sensitivity coefficients through the following equation with NP parameters and perturbation ε_p (Orlande et al, 2011, Estumano, 2016).

$$J_{\theta_j} = \frac{Y\left(\theta_1,..\theta_j+\varepsilon_P\theta_j,..\theta_{N_P}\right) - Y\left(\theta_1,...\theta_j-\varepsilon_P\theta_j,...\theta_{N_P}\right)}{2\varepsilon_P} \quad (2)$$

3.3 MARKOV CHAIN MONTE CARLO

In this work, to obtain an approximation of the posterior distribution, the Monte Carlo method with Markov Chain (MCMC) was used, simulating samples of $\pi_{posterior}(\boldsymbol{\theta}|\mathbf{Y})$. The idea is to obtain a sample from the posterior distribution and calculate sample estimates of characteristics of this distribution.

To this end, the Metropolis-Hastings algorithm will be used. This algorithm is based on the acceptance-rejection method, where candidate values are generated $\boldsymbol{\theta}^*$ belonging to a proposal distribution $p\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(t-1)}\right)$. In this work, Gaussian distributions were used. The Metropolis-Hastings algorithm is described below (Metropolis et al, 1953; Hastings, 1970; Kaipio e Somersalo, 2004; Gamerman e Lopes, 2006; Orlande et al, 2011; Ehlers, 2018 ; Oliveira et al., 2018; Van Ravenzwaaij, 2018):

1. The chain iteration counter is initialized $i = 1$ e arbitra-se um valor inicial $\theta^{(0)}$;

2. Generate a candidate θ* from distribution $p\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(i-1)}\right)$:

$$\boldsymbol{\theta}^* = \boldsymbol{\theta}^{(i-1)}\left(1 + w\varepsilon\right) \quad (3)$$

where ε a random number coming from a normal distribution, N(0,1) and w is the search step.

3. Calculate the probability of acceptance $\alpha\left(\boldsymbol{\theta}^{(i-1)} \mid \boldsymbol{\theta}^*\right)$ of the candidate value in the form:

$$\alpha\left(\boldsymbol{\theta}^{(i-1)} \mid \boldsymbol{\theta}^*\right) = \min\left[1, \frac{f\left(\boldsymbol{\theta}^* \mid \mathbf{C}/\mathbf{C_0}\right)}{f\left(\boldsymbol{\theta}^{(i-1)} \mid \mathbf{C}/\mathbf{C_0}\right)}\right] \quad (4)$$

4. A random number is generated u from uniform distribution, like $u \sim U(0,1)$.

5. If $u \leq \alpha\left(\boldsymbol{\theta}^{(i-1)} \mid \boldsymbol{\theta}^{*}\right)$, the new parameter vector is accepted and $\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{*}$. Otherwise, do $\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)}$.

6. The counter is increased from i to i+1 and returns to step 2.

As the experiments were not carried out, synthetic measurements obtained were used, adding uncertainty to the solution obtained with the reference parameters according to equation 5.

$$\mathbf{Y}_{\text{meas}} = \mathbf{Y}_{\text{exact}} + \sigma_{\text{med}}\epsilon \qquad (5)$$

where $\mathbf{Y}_{\text{meas}}$ represents the synthetic measure generated and this can be (X,P, S), σ_{med} is the standard deviation of the measurements and is a random variable N(0,1)

4. RESULTS AND DISCUSSION

The application of the Bayesian approach to the problem of estimating parameters of the model applied in bioreactors has as reference data the parameters presented in Table 1 (Marinho et al., 2018). Subject to the following initial conditions: X(0) = 180; S(0)= 12 ;P(0) = 0. The solution of the system of equations was carried out using the 4th order Runge-kutta method.

Table 1: Reference Parameters.

Parameter	Value
μ_{max} (h^{-1})	0.157
Ks ($g.L^{-1}$)	19.98
$Y_{E/S}$	0.446
Y_{VXS}	0.0622
F (Batch)	0

Initially, the analysis of the reduced sensitivity coefficients is evaluated to verify which parameters can be estimated and which state variable (X, S and P) the reduced sensitivity coefficients have the greatest magnitude. These analyzes are presented in Figures 1-3.

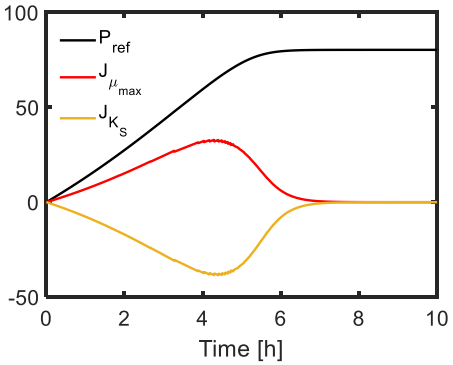

Figure 1: Reduced sensitivity coefficients in relation to P (product).

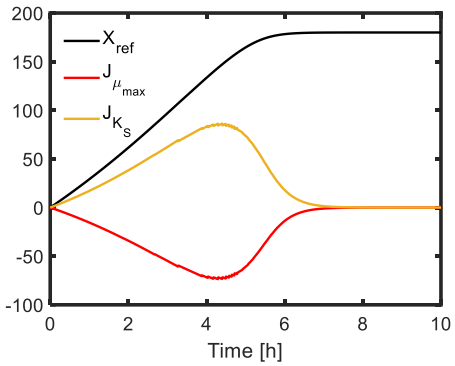

Figure 2: Reduced sensitivity coefficients in relation to X (Biomass).

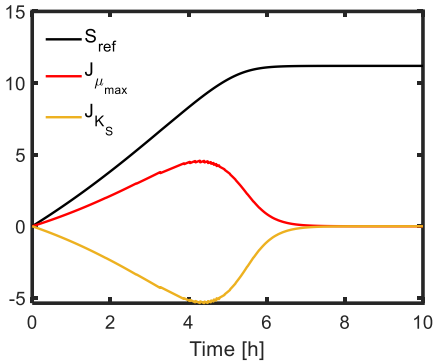

Figure 3: Reduced sensitivity coefficients in relation to S (substrate).

Analysis of the sensitivity coefficients reveals that the parameters μ_{max} and Ks have considerable magnitude in relation to all the model state variables (X,S and P). Regarding linear dependence, it is observed that these parameters are linearly dependent. The sensitivity analysis reveals that one can choose one of the state variables to consider the measurements. In this article, only measurements of the substrate S were considered with the objective of estimating the parameters and inferring the other state variables P and X. The Markov chains in Figures 4-5 are presented below for estimating the parameters.

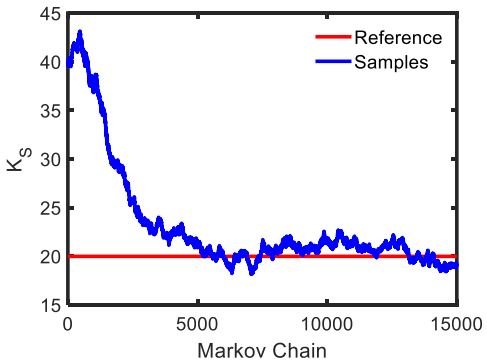

Figure 4: Markov Chain for parameter K_S.

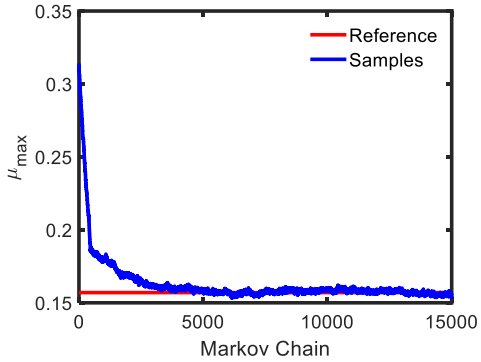

Figure 5: Markov Chain for the parameter $\mu_{max.}$

When observing Figures 4-5, it can be seen that 5000 states were needed for the Markov chains to reach equilibrium. Therefore, the heating was considered 5000 and the states from the heating states are samples of the posterior probability distribution of the parameters, that is, the estimates. Below, Figure 6 shows samples of estimates for both parameters. It can be seen in Figure 6 that such samples are correlated (ρ = 0.8) and the mean and 99% credibility interval are presented in Table 2.

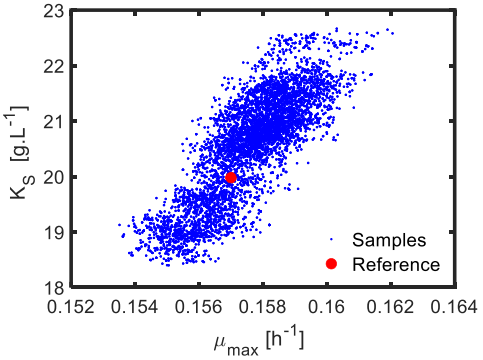

Figure 6: Samples from thr posterior distribution of both parameters.

Table 2: Mean and credible interval of 99% for the parameters estimations.

Parameter	Reference	Initial Estimation	Mean (C.I 99%)
μ_{max} (h^{-1})	0.157	0.34	0.157 (0.15;0.16)
Ks (g.L^{-1})	19.98	39.96	20.57 (18.5;22.5)

After determining the parameter estimates, the direct model was solved for each posterior sample of the parameters and the average of the solutions obtained was evaluated to be able to compare with the state variables in which the experimental measurements were not considered. Such comparisons between estimates and exact values (solution obtained with reference parameter) are presented in Figures 7-9.

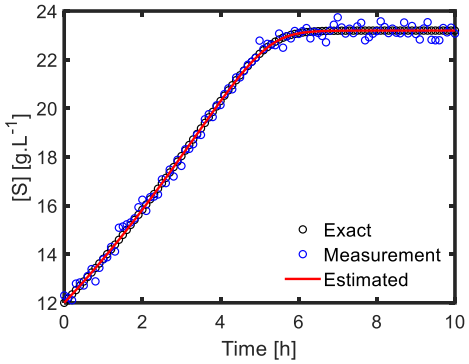

Figure 7: Comparison between simulated and exact measurements for substrate concentration.

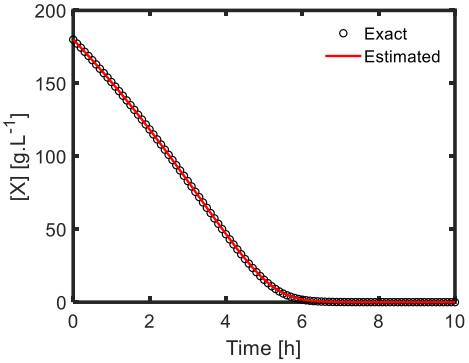

Figure 8: Comparison between simulated and exact measurements for biomass concentration.

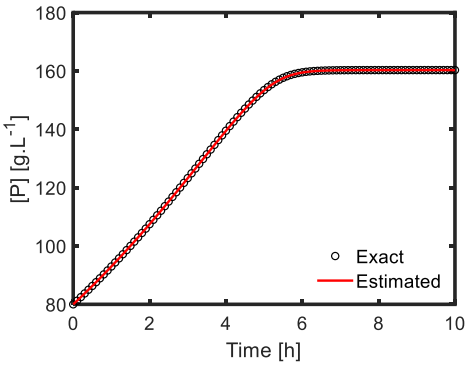

Figure 9: Comparison between simulated and exact measurements for Product concentration.

When comparing the simulated and exact measurements in Figures 7-9, it was observed that there was excellent agreement. Therefore, the strategy used to consider only substrate measurements for parameter estimation and inference of substrate, product and biomass is valid. Therefore, the technique applied to estimate parameters appears to be promising and useful so that it is not necessary to spend investments to obtain measurements of biomass and products.

5. CONCLUSION

This article showed the application of Bayesian techniques in mathematical models applied to bioreactors.

The results presented show that the parameters Ks and μ_{max} are correlated when evaluating the reduced sensitivity coefficient and subsequently proven with the analysis of samples of the posterior probability distribution obtained from the parameters, in which a correlation of 0.8 ($\rho = 0.8$) was verified.

The estimates were reported with good accuracy and precision, since when comparing the estimates of the parameters and state variables with the reference values in all evaluations, excellent agreements were obtained. Therefore, with the results obtained, it can be seen that the monta carlo technique via Markov chain is robust enough to estimate parameters of the model studied considering only measurements of the substrate (S) and estimate the dynamics of the variables P and X (product and biomass).

ACKNNOWLEDGMENT

We would like to thank FAPESPA for the financial support through the project entitled "Adsorção de Gases em Leito Fixo: Uso de Adsorventes Produzidos a Partir de Resíduos de Mineração em Sistema com Escala Semi Piloto" by agreement N°. 013/2022

REFERENCE

Andrews, J. F. (1968). A mathematical model for the continuous culture of micro-organisms utilizing inhibitory substrates. Biotechnol Bioeng

Andrews, J. F. (1968). A mathematical model for the continuous culture of micro-organisms utilizing inhibitory substrates. Biotechnol Bioeng

D. Gamerman, H. Lopes, H., Markov Chain Monte Carlo: Stochastic simulation for Bayesian inference, London: Chapman & Hall, 2006.

Ehlers, R. S., 2011, Inferência bayesiana. Departamento de Matemática Aplicada e Estatıstica, ICMC-USP, Vol. 64.

ESTUMANO, D. C. (2016), Estimativa de parâmetros e variáveis de estado de modelos aplicados à neurônios citomegálicos utilizando dados experimentais do protocolo de tensão fixa. UFRJ/COPPE – Rio de Janeiro – RJ, 233 p. (dissertação de doutorado).

H. Orlande, O. Fudym, D. Maillet, R. Cotta, R., Thermal Measurements and Inverse Techniques, CRC Press, Boca Raton, 2011.

Himmi, E. H., Bories, A., Boussaid, A., & Hassani, L. (2000). Propionic acid fermentation of glycerol and glucose by Propionibacterium acidipropionici and Propionibacterium freudenreichii ssp. shermanii. Applied Microbiology and Biotechnology, 53, 435-440

J. Kaipio, E. Somersalo, *Computational and Statistical Methods for Inverse Problems*, Springer, Berlin, 2004.

Marinho, C., Santos, A., Barreto, L., Saraiva, S., Carvalho, F., & Coêlho, d.(2018) avaliação cinética e modelagem matemática na fermentação propiônica. Xxii congresso brasileiro de engenharia química, são paulo - sp

Miller, K. V., & Block, D. E. (2020). A review of wine fermentation process modeling. Journal of Food Engineering, 273, 109783.

Monod, J. (1949). The growth of bacterial cultures. Annual review of microbiology, 3(1), 371-394.

N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller, Equation of State Calculation by Fast Computing Machines, J. Chemical Phys., vol. 21,1087-1092, 1953.

NAVEIRA-COTTA, C. P. Problemas inversos de condução de calor em meios heterogêneos: analise teórico-experimental via transformação integral, inferência bayesiana e termografia por infravermelho. 2009. Tese de Doutorado. Tese, Rio de janeiro, UFRJ/COPPE.

Oliveira, C., Junior, J. L., Knupp, D. C., Neto, A. S., Prieto-Moreno, A., & Llanes-Santiago, O. (2018). Estimation of kinetic parameters in a chromatographic separation model via Bayesian inference. Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería, 34(1).

Pradhan, N., Dipasquale, L., d'Ippolito, G., Fontana, A., Panico, A., Lens, P. N.,& Esposito, G. (2016). Kinetic modeling of fermentative hydrogen production by Thermotoga neapolitana. International Journal of Hydrogen Energy, 41(9), 4931-4940.

Stowers, C. C., Cox, B. M., & Rodriguez, B. A. (2014). Development of an industrializable fermentation process for propionic acid production. Journal of Industrial Microbiology and Biotechnology, 41(5), 837-852.

Van Ravenzwaaij, D., Cassey, P., & Brown, S. D. (2018). A simple introduction to Markov Chain Monte–Carlo sampling. Psychonomic bulletin & review, 25(1), 143-154.

W. Hastings, Monte Carlo Sampling Methods using Markov Chains and their Applications, Biometrika, vol.57, pp. 97-109, 1970.

Wang, J., & Wan, W. (2009). Kinetic models for fermentative hydrogen production: a review. International journal of hydrogen energy, 34(8), 3313-3323.

Zhao, M., Zhao, S., & Liu, F. (2023). Semi−Supervised Hybrid Modeling of the Yeast Fermentation Process. Machines, 11(1), 63.