Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 59-6 (2025) 451–456

Robust Reaction Rate Estimation with Application to Mammalian Cell Cultures

Guilherme A. Pimentel * **Fernando N. Santos-Navarro** *
Laurent Dewasme * **Alain Vande Wouwer** *
* *Systems, Estimation, Control and Optimization (SECO) , University of Mons, 7000 Mons, Belgium*
(e-mail: <guilherme.araujopimentel, fernandonobel.santosnavarro, laurent.dewasme, alain.vandewouwer>@umons.ac.be).

Abstract: Macroscopic modeling of bioprocesses is a common approach to derive dynamic predictors suitable for process optimization and control. This study presents a data-driven methodology for inferring reaction rates without resorting to the classical numerical differentiation of experimental data, which is prone to errors in the presence of noise. The approach is based on the minimization of a nonlinear least square criterion, which parameterizes the rates in terms of the temporal values. The proposed method is appropriate for datasets with sparse measurements and few experimental replicates. A use case considering protein production by mammalian cells is used to validate the proposed approach.

Keywords: reaction rates, robust estimation, mechanistic modeling, Hermite basis functions, biotechnology

1. INTRODUCTION

Macroscopic models describe the mass balance between macroscopic species in biological systems for process prediction, media optimization, and feeding control. Developing biologically interpretable kinetics relies on enzyme-catalytic or modulation effects, such as activations, inhibitions, and limitations. When little prior knowledge of biological systems is available, constructing the kinetics as a product of individual modulation phenomena is common (Grosfils et al., 2007; Richelle and Bogaerts, 2015; Wang et al., 2020; Dewasme et al., 2023). However, due to the nonlinear model structure, model identification involving model selection and parameter estimation can be challenging.

The first challenge is defining the number of macroscopic reactions from the datasets, which could be subject to the test of several candidate macroscopic reaction schemes. A well-known data-driven method to identify macroscopic reaction schemes is the Principal Component Analysis (PCA) (Bernard and Bastin, 2005), which can determine the number of reactions and the stoichiometric relations between the compounds in a data-driven way. An extension of this method, Maximum Likelihood Principal Component Analysis (MLPCA), accounts for higher levels of measurement noise (Mailier et al., 2012). Furthermore, Pimentel et al. (2023b) used a data-driven methodology with a robust algorithm for parallel implicit sparse identification to deduce macroscopic reaction schemes, while in Pimentel et al. (2024a), a new method using a low-rank matrix approximation approach was developed, which infers the number of reactions and its stoichiometry.

Apart from the number of reactions and their stoichiometric relations, the next challenging aspect is determining the kinetic laws, which have been mainly tackled either by empirically defining and testing different candidate kinetic laws – such as Monod (Monod, 1949) for substrate activation and Jerusalimski (Jerusalimski and Engamberdiev, 1969) for inhibition – based on the minimization of the fitting error of the model. This approach is sensitive to parameter initialization as the optimization problem is multimodal and possesses several local minima (Wang et al., 2020; Dewasme et al., 2023). Another well-known method consists of obtaining the reaction rate signals from the process derivative measurements and using the logarithm transformation to rewrite the problem into a linear identification problem that reveals the process compound involved in the activation and inhibition of the reaction rates (Grosfils et al., 2007; Richelle and Bogaerts, 2015). One of the bottlenecks of this approach is the need for numerical differentiation of the measurement signals. It is well known that obtaining derivatives from measurement datasets is an ill-posed, ad-hoc problem, and the solutions are susceptible to noise in the measurements (van Breugel et al., 2020). In addition, numerical differentiation tools do not allow the consideration of constraints such as the positivity of the concentrations in biosystems (if these concentrations were reconstructed by integrating the calculated derivatives), or additional prior knowledge. Moreover, the computation of the derivatives is obtained solely by considering one reactant/product time series at a time. To overcome obtaining derivatives from measurement for sparse and noisy datasets, Hebing et al. (2020) presented a method for selecting small sets of elementary modes and estimating reaction rates from noisy concentration measurements. The approach uses the sums of squared residual function with a regularization term that increases the robustness against noisy measurements.

In the spirit of Hebing et al. (2020), this work presents a robust method to extract biological kinetic rate evolutions from noisy and sparse datasets, with the original feature of a multicriteria cost function, which includes basic knowledge about the process, such as biologically inspired information, noise characteristics, and occurrence (or exclusion) of different biochemical reactions. This procedure is illustrated with the case study of

10.1016/j.ifacol.2025.07.187

a protein production process from mammalian cell cultures, considering a metabolic shift to lactate consumption.

The methodology can generally be used to first define the time evolution of the kinetic rates and, in a second step, to estimate the parameters of kinetic laws based on the extracted information on the rate evolutions. The methodology considers the prior knowledge of the stoichiometric matrix, which, as mentioned above, can be achieved in various ways, such as PCA. Moreover, in addition to the estimation of the reaction rate profiles, the approach provides the estimation of the derivatives of the measurement aftermath.

This paper is organized as follows. Section 2 presents the model of the protein production by mammalian cells considering lactate shift. Section 3 presents the proposed approach, as well as some numerical results considering different measurement noise levels, and a possible approach to scalability to larger datasets. Section 4 presents the conclusion and future work.

2. PROTEIN PRODUCTION CONSIDERING LACTATE SHIFT

The proposed methodology is developed and illustrated in the context of a protein production process based on the cultivation of mammalian cells, which predominantly use glucose as a primary carbon source and produce lactate as a byproduct. However, lactate accumulation is usually not recommended since it inhibits cell growth and protein (the product of interest) production. Moreover, at high lactate levels combined with low glycolysis (the main substrate consumption pathway) rates, cells are likely to adjust their metabolism, shifting to lactate consumption. This process was modeled based on experimental data in (Pimentel et al., 2023a) and can be represented by the following three reactions:

$$k_{31}G + Gn \xrightarrow{\varphi_1} k_{11}X_v + k_{51}L + k_{61}P \tag{1a}$$

$$k_{52}L \xrightarrow{\varphi_2} X_v \tag{1b}$$

$$k_{13}X_v \xrightarrow{\varphi_3} X_d + k_{63}P \tag{1c}$$

where X_v, X_d, G, Gn, L and P are the concentrations of viable biomass, dead biomass, glucose, glutamine, lactate, and proteins, respectively. The first reaction considers substrate consumption (glycolysis) to produce biomass and byproducts at a rate φ_1. The second reaction describes lactate consumption, producing viable biomass, governed by a rate φ_2. The third reaction represents viable biomass decay, leading to the production of dead biomass and the release of proteins into the medium. Applying mass balance to (1) yields the following ordinary differential equation system:

$$\frac{dX_v}{dt} = k_{11}\varphi_1 + \varphi_2 - k_{13}\varphi_3, \tag{2a}$$

$$\frac{dX_d}{dt} = \varphi_3, \tag{2b}$$

$$\frac{dG}{dt} = -k_{31}\varphi_1, \tag{2c}$$

$$\frac{dGn}{dt} = -\varphi_1, \tag{2d}$$

$$\frac{dL}{dt} = k_{51}\varphi_1 - k_{52}\varphi_2, \tag{2e}$$

$$\frac{dP}{dt} = k_{61}\varphi_1 + k_{63}\varphi_3, \tag{2f}$$

where the reaction rates are defined as:

$$\varphi_1 = \mu_{max,1}\frac{Gn}{(K_{Gn}+Gn)}\frac{G}{(K_G+G)}X_v, \tag{3a}$$

$$\varphi_2 = \mu_{max,2}\frac{L}{(K_L+L)}\frac{K_{GI}}{(K_{GI}+G)}X_v, \tag{3b}$$

$$\varphi_3 = \mu_{dmax}\frac{K_{Gnd}}{(K_{Gnd}+Gn)}X_v, \tag{3c}$$

K_{Gn}, K_G, K_L are the half-saturation parameters, $\mu_{max,1}$, $\mu_{max,2}$, and μ_{dmax} the maximum reaction rate parameters, and K_{GI} and K_{Gnd} the inhibition parameters. The reaction rate φ_1 is driven by two Monod factors activated by glucose and glutamine. Likewise, φ_2 stands for the selective consumption of lactate activated by lactate and inhibited by glucose (considered as the primary and preferred substrate). φ_3 models the biomass death rate inhibited by the presence of glutamine, which is the primary nitrogen source of the cell, ensuring its viability.

The model described in this section is used as a process emulator to generate synthetic data with various sampling rates and noise levels. The specific parameter values are listed in Table 1.

3. ROBUST ESTIMATION OF REACTION RATE TIME EVOLUTIONS

As discussed in the introduction section, the method considers the knowledge of the stoichiometric matrix $K \in \mathbb{R}^{N\times M}$, which can be determined in various ways, including a classical PCA. The main objective of the procedure is to estimate the time evolution of the reaction rates using the measurements of the species concentrations without resorting to explicit differentiation but rather by minimizing a multicriteria cost function.

3.1 Multicriteria optimization

The proposed method is based on the minimization of a cost function J

$$\theta = \arg\min_{\theta} J(\theta) : \bar{\theta} < \theta < \underline{\theta}, \tag{4}$$

to estimate the values of unknown parameters θ under upper and lower bound constraints.

The cost function J can contain several criteria relative to the distance between model predictions and the corresponding measurements, the smoothness of the time trajectory of the reaction rates, and biologically inspired penalties such as, for instance, penalties on the occurrence of some specific rates. In our case study, the cost function could take the following form:

$$\begin{aligned} J(\theta) = &\sum_{k=0}^{n} \underbrace{(\xi_{m,k} - \hat{\xi}_k)^\top \cdot Q \cdot (\xi_{m,k} - \hat{\xi}_k)}_{\text{Maximum Likelihood Cost Function}} + \\ &\sum_{k=0}^{n-1} \underbrace{(\hat{\varphi}_{k+1} - \hat{\varphi}_k)^\top \cdot R \cdot (\hat{\varphi}_{k+1} - \hat{\varphi}_k)}_{\text{Smoothness Cost}} + \\ &\sum_{k=0}^{n} \underbrace{(\hat{\varphi}_{1,k} \cdot \hat{\varphi}_{2,k})^\top \cdot W \cdot (\hat{\varphi}_{1,k} \cdot \hat{\varphi}_{2,k})}_{\text{Biologically Inspired Cost}}, \end{aligned} \tag{5}$$

where $\xi_{m,k}$ is the process measurement vector at time k, Q is the covariance matrix of the measurement errors and R and W are penalty matrices. In our case study, the last term in the cost function rejects the cooccurrence of φ_1 and φ_2, i.e., it expresses

Table 1. Parameters to generate simulation data.

Parameter	Velues	Parameter	Values	Parameter	Values
$\mu_{max,1}$ $[g/(10^9 Cells\ d))]$	0.46	K_{gnd} $[g/L]$	0.002	K_L $[g/L]$	1.20
$\mu_{max,2}$ $[g/(10^9 Cells\ d))]$	0.40	k_{11} $[(10^9 Cells/g)]$	6.80	K_{GI} $[g/L]$	0.80
μ_{dmax} $[d^{-1}]$	0.03	k_{31} $[g/g]$	18.0	k_{61} $[mg/g]$	107.80
K_G $[g/L]$	1.10	k_{51} $[g/g]$	10.70	k_{63} $[mg/g]$	2.90
K_{Gn} $[g/L]$	0.25	k_{52} $[g/g]$	1.20		

that the production and consumption of lactate are unlikely to coincide.

On a more practical aspect, the Matlab fmincon optimizer is selected to solve the constrained minimization problem (4) where θ is lower-bounded by zero to ensure the parameter positiveness and no upper bound is defined. The optimizer uses the solution obtained by an explicit Euler method of the discretized macroscopic process model

$$\hat{\xi}_{k+1} = \hat{\xi}_k + K\varphi(\hat{\xi}_k, \vartheta)dt, \tag{6}$$

where $\hat{\xi}_k$ is the vector of discretized state (compound) estimates, $\varphi(\hat{\xi}_k, \vartheta)$ is the reaction rate vector with ϑ the parameters of the rate kinetics, and dt is the time step interval.

3.2 Rare measurements

When only a small dataset is available, it is proposed to estimate the values of the reaction rates at each instant, yielding $n \times M$ values to estimate, where M is the number of reactions and n is the number of measurements. More precisely, the unknown parameter vector is defined as

$$\theta = \begin{bmatrix} \hat{\xi}_0^\top & \hat{\varphi}_1^\top & \hat{\varphi}_2^\top & \cdots & \hat{\varphi}_M^\top \end{bmatrix}, \tag{7}$$

where the vector $\hat{\xi}_0 \in \mathbb{R}_+^{N\times 1}$ contains the initial concentration values (even if these concentrations are initially measured, they are corrupted by noise and a good practice consists in estimating their most likely values), and the vectors $\hat{\varphi}_i \in \mathbb{R}_+^{n\times 1}$ contains the time evolution of each reaction rate.

In order to illustrate the approach, a batch culture simulation is performed considering independent and identically distributed (IID) Gaussian noise $\varepsilon \sim (0, \sigma^2)$ corrupting all the measurements. The imposed standard deviations σ are 0.1×10^6cells/ml, 0.0167×10^6cells/ml, 0.2 g/l, 0.01 g/l, 0.1 g/l, and 1.0 μg/ml for viable biomass X_v, dead biomass X_d, glucose G, glutamine Gn, lactate L, and proteins P, respectively. Four measurements are collected per day ($dt = 0.25$ days) for each process compound in a batch process of 7 days. The estimations of the reaction rates are obtained by solving (4) with the following decision vector

$$\theta = [X_{v,0} \quad X_{d,0} \quad G_0 \quad G_{n,0} \quad L_0 \quad P_0 \quad \hat{\varphi}_{1,1\times n} \quad \hat{\varphi}_{2,1\times n} \quad \hat{\varphi}_{3,1\times n}]. \tag{8}$$

The penalty matrices are set as $Q = diag(\sigma_i^2)^{-1}$, $R = diag([1000 \ \ 100 \ \ 100])$, and $W = 1000$. The rate-selection penalty weighted by W considers that the first biochemical reaction (1a) can not co-occur with reaction (1b), i.e., the consumption of lactate only occurs when glucose (and, in turn, glycolysis) is low. The tuning of W, therefore, depends on the biological a priori knowledge of the user and the confidence in the metabolic shift occurrence.

Figure 1 presents the concentration fitting where error bars (corresponding to 95 % confidence intervals) are displayed for each measurement. The dark gray dashed lines represent the estimation of the process compounds. Figure 2 shows the three reaction rate estimate trajectories. Even though scarce data is available, the method provides estimates with a fair accuracy level for the three reaction rate profiles since the root mean squared errors (RMSE) are 0.012, 0.026, and 0.005 for reaction rates 1, 2, and 3, respectively.

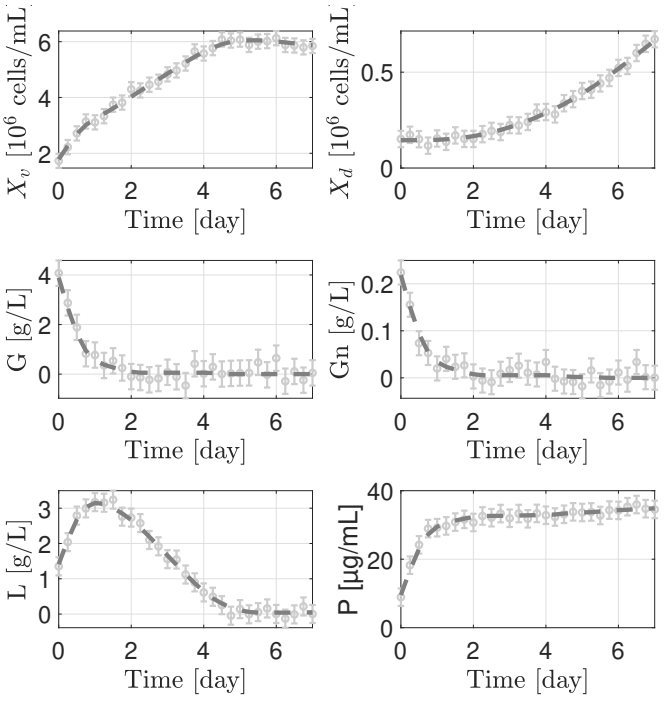

Fig. 1. Simulation restults. Error bars are the batch measurements, and dark gray dashed lines are the state estimations.

3.3 Robustness analysis

To test the robustness of the approach when facing data scarcity, data collection with a larger sampling interval is considered, e.g., one measurement per day ($dt = 1$ day). Three experimental replicates ($n_r = 3$) with different noise realizations are generated. The same metric as the one considered in (Hebing et al., 2020) is selected, where the average reconstruction error

$$\bar{\varphi}_{i_e} = \frac{1}{n_r} \sum |\hat{\varphi}_i - \varphi_i|, \quad i = 1,2,3 \tag{9}$$

of reaction rates considering different levels of noise is computed.

Table 2 presents the average reconstruction errors for different noise levels. The same penalty matrices $R = diag\,([10\ 3\ 1000])$ and $Q = diag\left(\sigma_i^2\right)^{-1}$ were set for all cases (σ_i^2 is the corresponding noise variance) except for the noise-free case where $R = diag([0.1\ 0.02\ 1.0])$ and $Q = diag\left(max(\xi_{m,i})^2\right)^{-1}$. Despite

Table 2. Robustness analysis: Average reconstruction error $\bar{\varphi}_{i_e}$ at different levels of simulated measurement noise.

	0%*	2.5%	5%	7.5%	10%	12.5%	15%	20%	30%
$\bar{\varphi}_{1_e}$	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
$\bar{\varphi}_{2_e}$	0.04	0.09	0.1	0.1	0.1	0.1	0.1	0.1	0.2
$\bar{\varphi}_{3_e}$	0.003	0.005	0.005	0.005	0.004	0.004	0.004	0.005	0.007

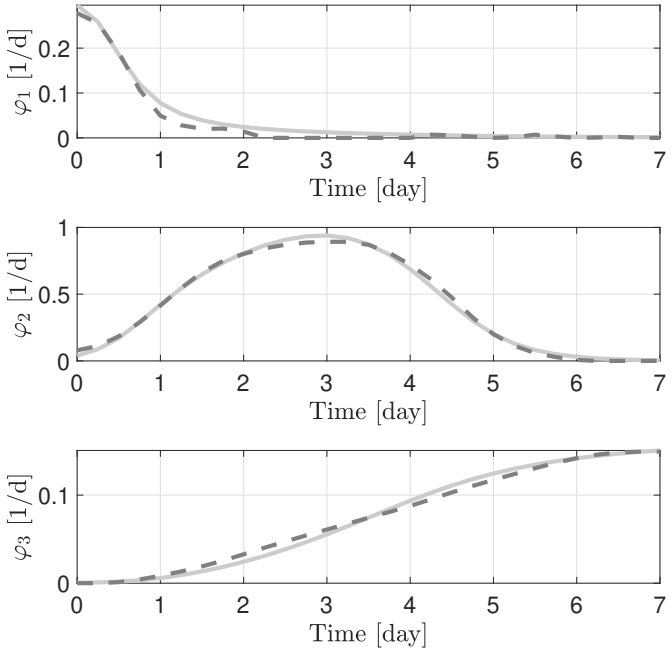

Fig. 2. Simulation restults. Light gray solid lines are the ground truth, and dark gray dashed lines are the reaction rate estimations.

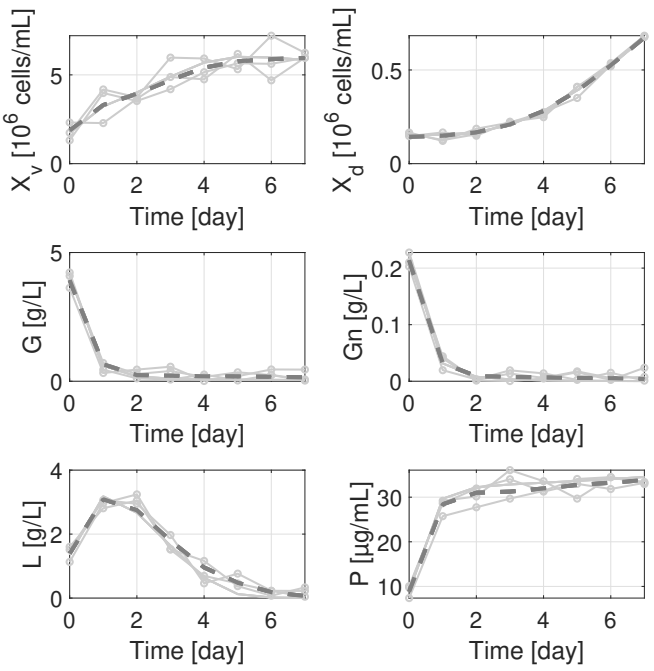

Fig. 3. Example of measurements with a 15% noise level. The light gray solid lines represent the three replicates, while the dark gray dashed lines show the direct validation results based on the average estimates.

the increasing noise levels, the average reconstruction errors remain at the same orders of magnitude. It should be noticed that in (Hebing et al., 2020), the values obtained for the average reconstruction error are around 0.06.

Figure 3 shows the reconstruction of the concentrations for the case with a noise variance of 15%, while Figure 4 shows the fitting of the reaction rate estimates. The proposed method also delivers the confidence intervals of the reaction rate estimation, contributing to better assessing the accuracy of the estimated rates.

3.4 Scalability for larger datasets

Increasing the dataset size improves the accuracy of the solution of problem (4). However, it also increases the computational burden as the number of decision variables increases. This observation suggests using an alternative strategy where the size of θ is not directly proportional to the amount of data under consideration. To this end, a piecewise cubic Hermite polynomial is chosen, which approaches the rates by finite elements with a local basis of low-order polynomials.

The interpolating function is defined as

$$g(x) = \sum_{i=0}^{(N_{nd}+1)(p+1)-1} a_i x^i, \tag{10}$$

where p is the desired derivative order, up to p^{th} at node N_{nd}. Considering this representation, it is possible to piecewise the trajectories of $\hat{\varphi}_j$ in e-elements, considering a two-data point cubic Hermite interpolation function and its first derivative. Thus, $p = 1$ and $N_{nd} + 1 = 2 \Rightarrow (p+1)(N_{nd}+1) = 3$, resulting in a cubic interpolating function as:

$$g^{(e_i)}(x) = a_0^{(e_i)}\phi_0(x) + a_1^{(e_i)}\phi_1(x) + a_{p,0}^{(e_i)}\psi_0(x) + a_{p,1}^{(e_i)}\psi_1(x), \tag{11}$$

where (e_i) is the number of i elements of $\hat{\varphi}$ defined by the user. The Hermite basis functions are defined as:

$$\phi_0(x) = 2x^3 - 3x^2 + 1, \tag{12}$$
$$\phi_1(x) = -2x^3 + 3x^2, \tag{13}$$
$$\psi_0(x) = x^3 - 2x^2 + x, \tag{14}$$
$$\psi_1(x) = x^3 - x^2. \tag{15}$$

Consequently, for the estimate considering only the first derivative approximation, each node in the element e_i has two Hermite basis functions: the first associated with the function value of the node zero (12), and the second with the first derivative of node zero, (14). Equivalently, node one has two Hermite basis functions (13) and (15). The parametrization considers the interpolated function's value and slope (possibly constrained, Vande Wouwer et al. (2014)), ensuring C1 continuity.

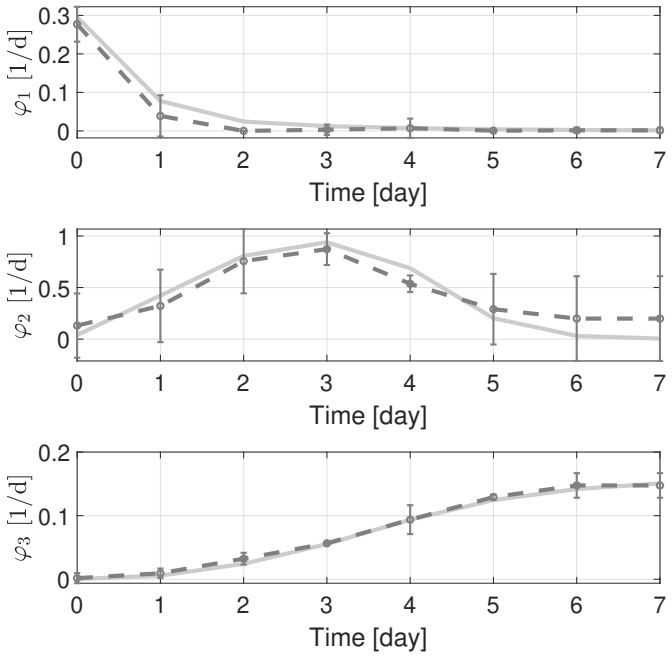

Fig. 4. Example of the $\varphi_{1,2,3}$ considering a 15% noise level. The light gray lines are the true reaction rate profiles, while the dark gray dashed lines are the average values of the estimation of the reaction rates with its error bar with 95% of the confidence intervals.

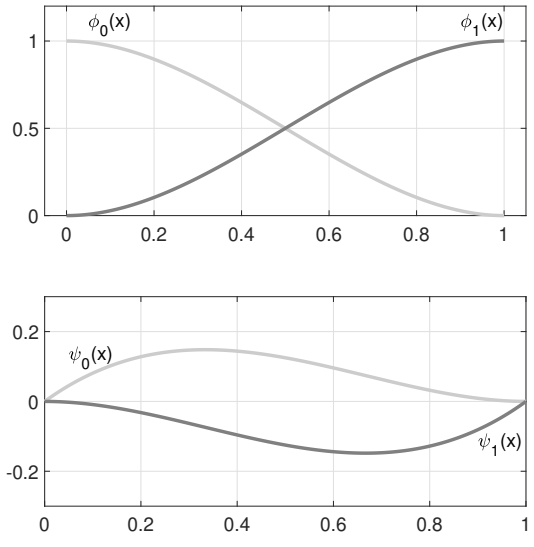

Fig. 5. Representation of the Hermite basis functions.

Figure 5 shows the properties of the Hermite basis functions. The plot on the top shows that $\phi_0(x)$ starts with a unitary value $\phi_0(0) = 1$ and ends with zero $\phi_0(1) = 0$. Likewise, $\phi_1(x)$ exhibits a mirror behavior, resulting in a smooth transition when more than one element is interconnected to interpolate a desired function. As for the derivative, $\psi_0(x)$ is zero at the beginning and the end of the interval, i.e., $\psi_0(0) = \psi_0(1) = 0$. As expected, $\psi_1(x)$ has a complementary behavior, guaranteeing a smooth transition between sections.

The proposed approach solves the minimization problem (4) to find the coefficients of the Hermite interpolating function (11) for a predefined number of elements e. The size of vector θ

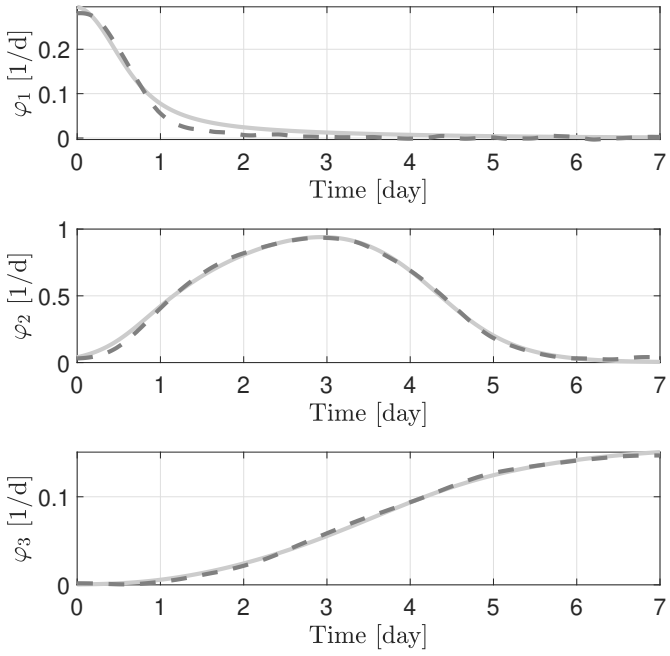

Fig. 6. Proposed method with the Hermite basis functions results. Error bars are the measurements, and the dark gray dashed line is the state estimation

is drastically reduced since it has four parameters for the first element and two additional parameters for each subsequent e-element. The bounding conditions between elements i and $i+1$ ensure function continuity, i.e., for $i = 1$, $a_0^{(e_2)} = a_1^{(e_1)}$ and $a_{p,0}^{(e_2)} = a_{p,1}^{(e_1)}$. The decision vector θ for the indirect method, therefore reads

$$\theta = \Big[\hat{\xi}_0^\top \quad \underbrace{a_0^{(e_1)} \quad a_{p,0}^{(e_1)} \quad a_1^{(e_1)} \quad a_{p,1}^{(e_1)}}_{element\ 1} \quad \underbrace{a_1^{(e_2)} \quad a_{p,1}^{(e_2)}}_{element\ 2} \quad \cdots \quad \cdots \quad \underbrace{a_1^{(e_i)} \quad a_{p,1}^{(e_i)}}_{element\ i}\Big] \tag{16}$$

To illustrate the application of this new parametrization, the measurements are now considered to be collected every 15 minutes ($t_s \approx 0.01$ days). The penalty matrices are set as $Q = diag(\sigma_i^2)^{-1}$, $R = diag([10\ \ 2\ \ 100]) \times 10^3$, and $W = 1000$. If the first approach was considered, the minimization problem would involve a θ vector of 2022 parameters, which is a very hard problem. The second approach considers interpolating Hermite polynomials over 30 elements, resulting in 62 parameters per reaction rate, or a total of 192 parameters, plus the identification of the state's initial conditions. Using fmincon and restricting the solutions to be positive, the results displayed in Figure 6 are obtained, with RMSEs of 0.011, 0.018, and 0.002 for the respective estimations of reaction rates 1, 2, and 3.

While the RMSE value of φ_1 almost remains equivalent, with a slight improvement of 4.5%, φ_2 and φ_3 estimations are improved by 40%, and 160 %, respectively. Since more measurements are available, these improvements were expected. Furthermore, both approaches show reaction rate trajectories with minimal noise amplitudes, as the noise in the measurements is effectively filtered by the smoothness term of (5).

Basically, the proposed method provides the reaction rate time evolutions with their associated confidence intervals considering datasets characterized by possibly scarce and noisy measurements. This scenario is common in the pharmaceutical industry, where small-scale wet laboratory experiments are conducted to identify specific manufacturing conditions supported by a model. The proposed method is a complementary tool that can help to further define a kinetic model structure by assimilating the estimates to well-known Monod, Haldane, or Contois Laws, to mention a few. Another possibility is to use sparse identification tools as proposed in Pimentel et al. (2024b) or the general rate parametrization approach of Richelle and Bogaerts (2015) to unveil the underlying kinetic structures.

4. CONCLUSION

This paper presents an approach for estimating the time trajectories of the reaction rates without resorting to a differentiation of the time trajectories of the extracellular component concentrations. To this end, a nonlinear minimization problem is solved based on a multicriteria cost function that considers the prediction error, smoothness constraints on the rate evolutions, as well as biological constraints inspired by prior knowledge about the reaction mechanisms. This tool is a brick in the modeling pipeline that can ease the further derivation of the kinetic rate structure, using one of the many published results in this area.

ACKNOWLEDGMENT

The authors acknowledge the support of the NEMO project (convention 8722) of the Biowin Walloon cluster (Belgium) achieved in collaboration with the CER Groupe, DNAlytics, and UCB. The scientific responsibility rests with its authors.

REFERENCES

Bernard, O. and Bastin, G. (2005). On the estimation of the pseudo-stoichiometric matrix for macroscopic mass balance modelling of biotechnological processes. *Mathematical Biosciences*, 193(1), 51–77.

Dewasme, L., Mäkinen, M., and Chotteau, V. (2023). Practical data-driven modeling and robust predictive control of mammalian cell fed-batch process. *Computers and Chemical Engineering*, 171(108164), 1–16.

Grosfils, A., Vande Wouwer, A., and Bogaerts, P. (2007). On a general model structure for macroscopic biological reaction rates. *Journal of biotechnology*, 130(3), 253–264.

Hebing, L., Neymann, T., and Engell, S. (2020). Application of dynamic metabolic flux analysis for process modeling: Robust flux estimation with regularization, confidence bounds, and selection of elementary modes. *Biotechnology and bioengineering*, 117(7), 2058–2073.

Jerusalimski, N. and Engamberdiev, N. (1969). *Continuous cultivation of microorganisms*, volume 517. Academic Press, New York.

Mailier, J., Remy, M., and Vande Wouwer, A. (2012). Stoichiometric identification with maximum likelihood principal component analysis. *Journal of Mathematical Biology*, 67(4), 739–765.

Monod, J. (1949). The growth of bacterial cultures. *Annual Review of Microbiology*, 3(1), 371–394.

Pimentel, G.A., Dewasme, L., Santos-Navarro, F.N., Boes, A., Côte, F., Filée, P., and Vande Wouwer, A. (2023a). Macroscopic dynamic modeling of metabolic shift to lactate consumption of mammalian cell batch cultures. In *9th International Conference on Control, Decision and Information Technologies*.

Pimentel, G.A., Dewasme, L., and Vande Wouwer, A. (2023b). On the number of reactions and stoichiometry of bioprocess macroscopic models: an implicit sparse identification approach. In *22nd World Congress of the International Federation of Automatic Control*.

Pimentel, G.A., Dewasme, L., and Vande Wouwer, A. (2024a). Data-driven inference of bioprocess models: A low-rank matrix approximation approach. *Journal of Process Control*, 134(103148), 1–10.

Pimentel, G.A., Santos-Navarro, F.N., Dewasme, L., and Vande Wouwer, A. (2024b). Sparse identification of stoichiometry and nonlinear kinetics of bioprocess macroscopic models. In *12th IFAC Symposium on Advanced Control of Chemical Processes*.

Richelle, A. and Bogaerts, P. (2015). Systematic methodology for bioprocess model identification based on generalized kinetic functions. *Biochemical Engineering Journal*, 100, 41–49.

van Breugel, F., Kutz, J.N., and Brunton, B.W. (2020). Numerical differentiation of noisy data: A unifying multi-objective optimization framework. *IEEE Access*, 8, 196865–196877.

Vande Wouwer, A., Saucez, P., and Vilas, C. (2014). *Simulation of ODE/PDE Models with MATLAB, OCTAVE and SCILAB*. Springer International Publishing Switzerland.

Wang, M., Risuleo, R.S., Jacobsen, E.W., Chotteau, V., and Hjalmarsson, H. (2020). Identification of nonlinear kinetics of macroscopic bio-reactions using multilinear Gaussian processes. *Computers & Chemical Engineering*, 133, 106671.