Incremental Model Identification of Bio-processes from Data: Application to Microbial Production of Hyaluronic Acid

Kamakshi C[a,b], Guhan Jayaraman[a], Nirav P Bhatt[a,b]

[a] *Department of Biotechnology, Bhupat and Jyoti Mehta school of Biosciencies,* [b] *Prospective CoE for Network Systems Learning, Control, and Evolution, Indian Institute of Technology Madras, Chennai, India (e-mail:bt18s007@smail.iitm.ac.in, guhanj@iitm.ac.in niravbhatt@iitm.ac.in).*

Abstract: The development of reliable kinetic models of bioprocesses from data is a challenging task. In this work, a systematic approach to develop a kinetic model of bioprocess involving single biomass from concentration data is proposed without imposing any kinetic model *a priori*. The proposed incremental model identification approach decomposes the model-building task into a set of sub-tasks such as determining the yield coefficients and maintenance coefficient, specific growth rate structure identification, and parameter estimation. It is shown that the proposed approach allows identifying the mechanism of product formation. The proposed approach is applied to the microbial production of Hyaluronic acid (HA), an important biopolymer, using a recombinant *Lactococcus lactis* MKG6. An unstructured kinetic model is developed for the HA production from data. It is shown that HA production is a growth-associated process. Further, the specific growth rate of HA production is identified from a set of rate candidates. It is revealed that the specific growth rate in the HA production follows the non-competitive HA inhibition model. The parameters obtained by the incremental identification are further refined to obtain statistically optimal estimates using the simultaneous model identification. Validation of the identified kinetic model of HA production on new experimental data shows that the proposed approach leads to a reliable kinetic model with the optimal parameter estimates.

Keywords: Kinetic models, Incremental identification, Hyaluronic acid, Bioprocesses, Unstructured kinetics

1. INTRODUCTION

Kinetic models of bioprocesses are essential for developing process with optimal operational conditions in production of desired biological products. Further, kinetic models allow us to monitoring and control of bioprocesses during the operation. These models of bioprocesses are formulated based on conserved mass balance equations and are written as either structured or unstructured kinetic model forms. In contrast to structured models, unstructured models consider a macroscopic reaction scheme and develop a simple but comprehensive model describing bioprocesses (Richelle and Bogaerts, 2015). Hence, unstructured models are suitable for model-based monitoring, control, and optimization of bioprocesses. However, developing unstructured models of bioprocesses is a challenging task due to complex bioprocesses involving biomass, intra- and extracellular metabolites.

The selection of an appropriate model describing growth kinetics, substrate uptake and product formation from a set of kinetic models becomes cumbersome when the number of model candidates is large(Wang et al., 2007). In general, the estimation of kinetic parameters are carried out by solving the model equations simultaneously. This simultaneous identification approach estimates the parameters by comparing the model prediction and measurements (Brendel et al., 2006; Saa and Nielsen, 2017). This needs to be repeated for all rate expression combinations, when several model candidates are proposed to describe a single reaction. The approach is computationally costly when there are several candidates for each rates. Furthermore, the issue of parameter identifiability (Asprey and Macchietto, 2000), accounting of errors due to single mismatch, and proper choice of initial guess values of parameters are some of the drawbacks of simultaneous identification approaches (Brendel et al., 2006). Further, practical unidentifiablity of some of the parameters of these kinetic models from the available data also poses additional challenges in estimating parameters and appropriate models (Villaverde et al., 2016; Gábor et al., 2017; Varghese et al., 2018).

The incremental model identification of reaction systems is an alternative approach for handling problems of model discrimination and parameter estimations from data (Bhatt et al., 2012; Bardow and Marquardt, 2004). In this approach, model identification is decomposed into several sub-tasks such as identification of reaction stoichiometry, model discrimination, and parameter estimation. In the literature, the incremental model identification approaches have been developed and demonstrated

for chemical and enzymatic reaction systems(Bardow and Marquardt, 2004; Brendel et al., 2006; Zavrel et al., 2008).

In this work, the incremental model identification approach is extended to identify model for microbial fermentation processes. The extended incremental identification approach includes steps such as computation of rates, identification of microbial growth kinetic mechanism, and yield and maintenance coefficients, structure of specific growth rate, identifiability analysis of selected and parameter estimation. Further, the proposed approach is demonstrated on an experimental data involving microbial production of hyaluronic acid in a batch reactor.

2. MASS BALANCE OF BIOREACTORS

In this section, model equations for bioprocesses involving a single substrate, single biomass and multiple products are developed in a bioreactor. The bioreactor with one inlet [1] and one outlet is considered. It is assumed that the reactor is a well-mixed tank reactor. In microbial fermentation processes, the substrate S is predominately consumed for the growth and product synthesis during the cell growth. Whereas in the stationary phase, the available substrate is directed towards the maintenance of the cell. In some scenarios, products such as secondary metabolites, antibiotics are synthesised as a part of cellular maintenance. Therefore, the reaction scheme describing these scenarios for P_j, $j = 1, \ldots, p$ formation of products from the substrate S using a biomass X can be written as:

$$\begin{aligned} R1: \; & Y_{SX} S \xrightarrow{r_1} Y_{P_1X} P_1 + ... + Y_{P_pX} P_p + X \\ R2: \; & m_S S \xrightarrow{r_2} \beta_{P_1} P_1 + ... + \beta_{P_p} P_p \end{aligned} \tag{1}$$

where Y_{SX}, Y_{P_jX}, $j = 1, \ldots, p$ are the yield coefficients for the substrate and products in the reaction R1. m_S is the maintenance coefficient. β_{P_j}, $j = 1, \ldots, p$ are the yield coefficients of products in the reaction R2. r_1 and r_2 are the unknown rates of the reactions R1 and R2. With the constant density assumption, the overall mass balance equation can be written as the change in volume (V) during the fermentation as follows:

$$\dot{V} = q_{in} - q_{out} \tag{2}$$

where q_{in} and q_{out} are the inlet and outlet flow rates. For all the substrate, biomass and products, the mass balance equations are written as:

$$\dot{\mathbf{m}} = \frac{d(\mathbf{c}V)}{dt} = \mathbf{Kr(t)}V + \mathbf{c}_{in} q_{in} - \mathbf{c}\; q_{out} \tag{3}$$

where $\mathbf{m} = [m_X, m_S, m_{P_1}, \ldots, m_{P_p}]^T$ is the $(p+2)$-dimensional vector of mass of species in the bioreactor, $\mathbf{c} = [X, S, P_1, \ldots, P_p]^T$ is the $(p+2)$-dimensional vector containing the concentrations of species in the bioreactor, $\mathbf{K}$ is $(p+2) \times 2$-dimensional yield coefficient matrix, $\mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix}$ is the reaction rate vector, and $\mathbf{c}_{in}$ is the $(p+2) \times 1$-dimensional inlet composition matrix. Using Eq. (2), the mass balance equation can be simplified into

$$\dot{\mathbf{c}} = \mathbf{Kr}(t) + \frac{(\mathbf{c}_{in} - \mathbf{c}) q_{in}}{V} \tag{4}$$

[1] The approach can be extended to more than one inlet in a straightforward manner

Eq. (4) describes the change of concentrations of substrate, biomass, and products inside a bioreactor. Typically, for bioprocesses involving a single substrate, one inlet stream is considered. Note that the multiple inlets will affect the volume in the reactor. However, it will not change the number of reactions and parameters to be estimated in a bioprocess. For batch processes, $q_{in} = 0$. Then, Equation (4) can be further simplified as

$$\dot{\mathbf{c}} = \mathbf{Kr(t)} \tag{5}$$

For example, a bioprocess with two products carried out in a batch reactor with $\mathbf{c} = [X, S, P_1, P_2]^T$, Eq. (5) can be written as

$$\dot{\mathbf{c}} = \begin{bmatrix} \dot{X} \\ \dot{S} \\ \dot{P}_1 \\ \dot{P}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -Y_{XS} & -m_S \\ Y_{P_1X} & \beta_{P_1} \\ Y_{P_2X} & \beta_{P_2} \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix}. \tag{6}$$

3. INCREMENTAL IDENTIFICATION OF UNSTRUCTURED MODELS FROM DATA

In this section, the incremental unstructured model identification of bioprocesses is discussed. In microbial fermentation processes, the product formation can be characterized by one of the following mechanism: (i) growth associated, (ii) non-growth associated, or (iii) mixed-growth associated. The production formation mechanism will be determined by estimating coefficients of $\mathbf{K}$ from experimental data. Then, a specific growth rate structure and corresponding parameters will be identified.

It is assumed that the following experimental data for L_N time instants are available: Concentrations of metabolite and biomass $\mathbf{c}$ (g/L), inlet and outlet flow rates, q_{in}, q_{out}, and feed concentration c_{in} (g/L). By integrating Eq. (2), the volume at different time can be obtained. The steps involved in the proposed incremental model identification approach are as follows:

(1) Estimation of metabolite fluxes from measurements
(2) Estimating elements of $\mathbf{K}$
(3) Identifying an appropriate specific growth rate structure from a set of rate structures
(4) Checking for identifiability of the identified specific growth rate structure
(5) Refining parameter estimates of the identified model in Step 3 using the simultaneous parameter estimation

3.1 Estimation of metabolite fluxes from measurements

Estimating metabolite fluxes, via the computation of derivatives from noisy measurements is ill-posed problem (Mhamdi and Marquardt, 2004). It often leads to poor estimation of rates. Hence, the noise need to be reduced using smoothing or filtering operation before computing rates. Various methods exist in the literature for filtering noisy data such as Tikhonov–Arsenin filtering, cubic smoothing spline method(Mhamdi and Marquardt, 2004), L-curve criterion (Brendel et al., 2006), spline and general cross validation (GCV) (Craven and Wahba, 1978). In this work, the cubic smoothing spline method is used for smoothness of concentration data denoted as $\tilde{\mathbf{c}}(t)$. Then, the metabolite (or species) flux vector $\mathbf{f}$ at the sampled

time t_h is computed using the smoothed concentrations, $\tilde{\mathbf{c}}$, and its derivatives, and $V(t_h)$

$$\mathbf{f}(t_h) = \mathbf{K}\mathbf{r}(t_h) = \frac{d\tilde{c}}{dt}(t_h) - \frac{(\mathbf{c}_{in} - \tilde{\mathbf{c}}(t_h))q_{in}(t_h)}{V(t_h)} \tag{7}$$

To compute the reaction rate $\mathbf{r}$ from the species flux vector $\mathbf{f}$, $\mathbf{K}$ has to be known. In the next section, the elements of $\mathbf{K}$ are estimated from the fluxes and metabolite concentrations.

3.2 Estimating elements of $\mathbf{K}$

The elements of $\mathbf{K}$ are estimated by exploiting the structure of $\mathbf{K}$. Note that there are two reactions R1 and R2 taking place in the system. For the reactions R1 and R2, the coefficients of biomass X is known to be one, and zero. Hence, the rate of reaction R1 can be computed as:

$$r_1^m(t_h) = f_X(t_h) \tag{8}$$

where f_X is the biomass flux, and the superscript m indicates the computed from measurements. Further, the reaction R2 is associated with the cell maintenance and product formation. In bioprocesses, the rate of reaction R2 is equal to the biomass concentration and is given by:

$$r_2^m(t_h) = \tilde{X}(t_h) \tag{9}$$

where $\tilde{X}$ is the smoothed biomass concentration. Using Eq. (7), the species flux vector for the substrate and the products can be written as:

$$\begin{aligned} f_S(t_h) &= -Y_{SX} f_X(t_h) - m_S \tilde{X}(t_h) \\ f_{P_j}(t_h) &= Y_{P_jX} f_X(t_h) + \beta_{P_j} \tilde{X}(t_h),\ j = 1, \ldots, p \end{aligned} \tag{10}$$

where f_S and f_{P_i} are the substrate and the j product fluxes, respectively. Eq. (10) establishes $(p+1)$ relationships between the computed species fluxes, the computed biomass flux and the biomass concentration at the time instant t_h and the unknown yield coefficients. These relationships can be formulated as multiple linear regression models in the form of $y = m_1 x_1 + m_2 x_2$ with $y = f_S$ or f_{P_j} and $x_1 = f_X$ and $x_2 = X$. By fitting these multiple linear regressions in Eq. (10), the yield coefficients and maintenance coefficient are estimated.

The estimated yield coefficients and maintenance coefficient value are useful to discriminate the product formation mechanisms. In order to discriminate the product formation mechanisms, the significance of the estimated parameters in Eq. (10) is determined using the hypothesis tests. The t-statistic for each estimated parameter is tested against a null hypothesis of zero. The corresponding p-value for each parameter is checked for the level of significance, $\alpha = 0.05$ and the $(L_N - p)$ degree of freedom. The t-statistics is determined by the ratio of estimated coefficient $\hat{\theta}$ and standard error of the estimate, $SE(\theta)$ as follows:

$$T_{stat} = \frac{\hat{\theta}}{SE(\hat{\theta})} \tag{11}$$

The p-value is computed using the T_{stat} in Eq. (11). Then, the null hypothesis is accepted for p-value>0.05. Note that the decision to accepted the null hypothesis can be done by comparing T_{stat} in Eq. (11) with the theoretical t-statistic with $\alpha = 0.05$ and the $(L_N - p)$ degree of freedom. In this manner, the elements of $\mathbf{K}$ can be determined. If the hypothesis tests' results in the yield coefficients associated with the reaction rate r_2 to be zero, then, the bioprocess can be modelled using the reaction R1.

3.3 Identification of specific growth rate structure

In the previous step, $\mathbf{K}$ is estimated by exploiting the structure of the systems and the product formation mechanism is determined. Then, the next is to identify specific growth rate structure using the computed reaction rate r_1. In bioprocesses, the reaction rate r_1 is expressed as

$$r_1 = \mu(S, X, P_j, \boldsymbol{\gamma})\, X \tag{12}$$

where $\mu(S, X, P_j, \boldsymbol{\gamma})$ is the unknown specific growth rate of bioprocess. The structure of μ and corresponding parameters are determined from a set of the proposed structures using the estimated profile, r_1^m and concentration measurements.

The model discrimination among specific growth rate models is performed by ranking the models based on the selection criteria. A nonlinear least-square estimation problem is formulated by minimizing the sum square of errors as follows

$$\begin{aligned} f_i = \min_{\boldsymbol{\gamma}} \sum_{t_h=0}^{L_N} (r_1^m(t_h) - r_{1,i}^s(t_h)))^2 \\ s.t.\ \boldsymbol{\gamma}\ \in [\boldsymbol{\gamma}^{lb}, \boldsymbol{\gamma}^{ub}] \geq 0 \end{aligned} \tag{13}$$

where r_1^m is the computed rate using the measured concentrations, $r_{1,i}^s = \mu_i X$ is the predicted rate using the ith proposed structure of specific growth rate, and $\boldsymbol{\gamma}$ parameters. The optimization problem is solved for all proposed specific growth rate structures. The models with minimum objective function values are selected for carrying out further model discrimination. These models are labelled as the screened models. The F-test is carried out to identify the best model out of the screened models. The F-test for two models m_1 and m_2 with different parameters p_1 and p_2, where $p_2 > p_1$ is computed for L_N data points with sum of squares of error (SSE) as follows:

$$F = \frac{(SSE_1 - SSE_2)(L_N - p_1)}{SSE_2(p_2 - p_1)} \tag{14}$$

If the calculated F-value for the model is greater than the reference F_{ref}, the model with more parameters is selected. The reference value is determined from the F-distribution table using degree of freedom $(p_1 - p_2,\ L_N - p_1)$.

3.4 Check for model Identifiability

Model identifiability allows us to identify whether the parameters in the given specific rate structure can be uniquely identified from the data. The identified specific growth rate structure is used to determine identifiability of parameters in the model. Typically, *a priori model identifiability* is carried out for all the proposed model candidates before estimating the parameters. However, it leads to combinatorial complexity due to the number of model candidates proposed for specific growth rate. In order to avoid this issue, the model identifiability will be carried out only for the selected model after the model discrimination step. The differential algebraic approach is used to check identifiability of the model

(Audoly et al., 2001). In this work, Differential Algebra for Identifiability of SYstems (DAISY) software tool is used to perform identifiability of parameters (Bellu et al., 2007). The estimated **K** and the identified specific growth rate structure in **r** are used in Eq. (4) for identifiability check using the DAISY. In case of non-identifiable model, either reparameterization of the identified model or re-select a new model from the literature will be considered.

3.5 Simultaneous parameter optimization

To obtain statistically optimal values of the parameters, the simultaneous parameter estimation is carried out using the estimated parameters and model structures in the previous section as follows:

$$\begin{aligned} \hat{\boldsymbol{\gamma}} = \underset{\boldsymbol{\gamma}}{\operatorname{argmin}} & \sum_{j=1}^{p+2} \sum_{l=1}^{L_N} (c_{jl}^m - c_{jl}(\boldsymbol{\gamma}))^2 \\ & s.t. \text{ model equations (4)} \\ & \boldsymbol{\gamma} \in [\boldsymbol{\gamma}^{lb}, \boldsymbol{\gamma}^{ub}] \geq 0 \end{aligned} \tag{15}$$

where for c_{jl}^m is the jth species concentration at the lth time instants, $c_{jl}(\boldsymbol{\gamma})$ is the predicted jth species concentration at the lth time instants by the model equations. The $\hat{\boldsymbol{\gamma}}$ is the optimally refined estimated parameters for the reaction system. The parameter values obtained in the model discrimination step is used as the initial guess values with the appropriate bounds.

4. MODEL IDENTIFICATION FOR HYALURONIC ACID PRODUCTION

In this section, the incremental model identification proposed in the previous section is used to identify the best unstructured model that can describe the fermentation of Hyaluronic acid (HA) using the recombinant *Lactococcus lactis* MKG6 (transformed from *Lactococcus lactis* NZ9020) (Kaur and Jayaraman, 2016).

4.1 Experimental method

Experiments were carried out with the initial glucose concentration of 30 g/L. The fermentation was carried out in 2.4L KLF 2000 bioreactor (Bioengineering AG, Switzerland) with 1.2 Litre working volume in a batch mode of operation. The fermentation was carried out in an unaerated condition with agitation, pH and temperature maintained at 200 rpm, 7, and 30^{o}C, respectively. The cells were induced with Nisin (2 ng/mL) when it reached OD_{600}=0.6 for the production of HA. All the experiments were carried out until glucose was utilized completely. The fermentation of the recombinant *Lactococcus lactis* MKG6 yields biomass (X), Hyaluroinc acid (HA), acetic acid (A), formate (F), and ethanol (E) from glucose (S). The samples collected at the regular intervals were analyzed using a HPLC system (Shimadzu Prominence, Japan) to measure the concentrations of glucose, lactic acid, acetic acid, formate and ethanol in the fermentation broth. Two experiments were carried out in the identical conditions. The experimental data from one batch are used to identify the model and then, the other batch are used to validate the model.

Table 1. Product formation mechanism

Specific rate	Product formation
q_p=$\alpha\mu$	Growth associated
q_p=β	Non-growth associated
q_p=$\alpha\mu + \beta$	Mixed-growth associated

In this process, the substrate consumed are taken up for the product formation in addition to growth and maintenance of cell. The rate of substrate consumption is a function of growth rate, product formation rate and cellular maintenance. The specific rate of substrate uptake is derived as,

$$\begin{aligned} q_S &= \frac{\mu}{Y_{XS}} + \frac{q_{P_i}}{Y_{P_iS}} + m_s \\ q_S &= G_a\mu + m_T \end{aligned} \tag{16}$$

where G_a (g/g) and m_T (g/g h) are lumped parameters describing the parameters associated with growth rate and maintenance, respectively. q_{P_i} is the specific rate of product formation (h^{-1}), described by the Leudeking-piret equation. It has combination of both growth associated (α) and non-growth associated (β) terms.

$$q_{P_i} = \alpha_{P_i}\mu + \beta_{P_i} \tag{17}$$

where $P_i = [A, HA, F, E]^T$. α_{P_i} is the yield coefficient for the ith product. The product formations are classified based on the value of α_{P_i} and β_{P_i} (Table 1). For a growth associated product formulation, α_{P_i} is be equal to Y_{P_iX}, yield coefficient.

The products produced during the fermentation of *Lactococcus lactis* can either growth associated or mixed growth associated product formation. The objective here is to identify a kinetic model for the production of HA. Based on the description of the production of HA, the reactions can be written as

$$\begin{aligned} G_aS &\xrightarrow{r_1} \alpha_AA + \alpha_{HA}HA + \alpha_FF + \alpha_EE + X \\ m_SS &\xrightarrow{r_2} \beta_AA + \beta_{HA}HA + \beta_FF + \beta_EE \end{aligned} \tag{18}$$

For the batch operation, Eq. (5), the rates of change in concentrations of biomass, and metabolites, is written as:

$$\dot{\mathbf{c}} = \mathbf{K}\mathbf{r}, \text{ with } \mathbf{c} = \begin{bmatrix} X \\ S \\ A \\ HA \\ F \\ E \end{bmatrix}, \mathbf{K} = \begin{bmatrix} 1 & 0 \\ -G_a & -m_T \\ \alpha_A & \beta_A \\ \alpha_{HA} & \beta_{HA} \\ \alpha_F & \beta_F \\ \alpha_E & \beta_E \end{bmatrix}, \mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} s \tag{19}$$

The rate expressions for r_1 and r_2 can be written as $r_1 = \mu X$ (identical to Equation (8)) and $r_2 = X$. The proposed models for specific growth rate are given in Table 2. Eq. (19) is written in the form of Eq. (5).

4.2 Results & Discussion

The concentrations for biomass, substrate, and products are available as the discrete points (not shown here). The concentration profile is obtained from average of duplicate measurements. The proposed incremental model identification approach is applied to the concentration data to identify a kinetic model of HA production.

The noisy concentration measurements are filtered using the cubic smoothing spline here. The rate profiles ($\frac{d\tilde{\mathbf{c}}}{dt}$)

Table 2. Dynamic model candidates for specific growth rate μ

Model	Specific growth rate (μ)
Logistic model	$\mu^{(1)} = \mu_0\left(1 - \frac{X}{X_m}\right)$
Monod model	$\mu^{(2)} = \frac{\mu_m S}{k_s + S}$
Competitive HA inhibition model	$\mu^{(3)} = \frac{\mu_m S}{k_S\left(1 + \frac{HA}{k_{i,HA}}\right) + S}$
Competitive HA and acetate inhibition model	$\mu^{(4)} = \frac{\mu_m S}{k_S\left(1 + \frac{HA}{k_{i,HA}}\right)\left(1 + \frac{A}{k_{i,A}}\right) + S}$
Competitive substrate inhibition model	$\mu^{(5)} = \frac{\mu_m S}{k_S\left(1 + \frac{S}{k_{i,S}}\right) + S}$
Non-competitive substrate inhibition model	$\mu^{(6)} = \frac{\mu_m S}{\left(k_S + S + \frac{S^2}{k_{i,S}}\right)}$
Linear HA inhibition model	$\mu^{(7)} = \frac{\mu_m S}{k_S + S}\left(1 - \frac{HA}{k_{i,HA}}\right)$
Non-competitive HA inhibition model	$\mu^{(8)} = \frac{\mu_m S}{(k_S + S)\left(1 + \frac{HA}{k_{i,HA}}\right)}$

Table 3. Yield coefficients estimated using the metabolite fluxes

Kinetics	Parameter	value	p-value
Substrate utilization	G_a	8.49	2.12×10^{-5}
	m_T	0.155	0.431
Acetate Production	α_A	0.529	0.0059
	β_A	0.026	0.258
HA Production	α_{HA}	0.392	0.01
	β_{HA}	-0.012	0.53
Formate Production	α_F	1.79	6.31×10^{-5}
	β_F	0.044	0.34
Ethanol Production	α_E	2.14	9.79×10^{-6}
	β_E	-0.007	0.869

computed using three-point numerical differentiation to compute the metabolite fluxes used to estimate the yield coefficients associated with the substrate consumption and product formation rates as mentioned in Section 3.2. The results of the regression analysis are given in Table 3. It can be observed that the p-values from the t-statistic rejects null hypothesis for some parameters. For the substrate consumption rate, the p-value of cellular maintenance coefficient m_T is greater than the significance level, $\alpha = 0.05$. Therefore, the parameter m_T is rejected from the model. Similarly, the non-growth associated β in all the products production rate are rejected due to larger p-values. Hence, it can be concluded that the production of HA follows the growth associated product formation. In other words, the HA production can be modelled as a single reaction scheme with no maintenance.

With the estimated elements of $\mathbf{K}$, the specific growth rate model structure is identified from the set of models proposed in Table 2 by fitting the least-squares problem in Eq. (13). The objective function values for the eight models are compared in Fig. 1. It can be observed that the models $\mu^{(1)}$, $\mu^{(3)}$, $\mu^{(4)}$, $\mu^{(7)}$ and $\mu^{(8)}$ have the least values among all the models. Then, the F- statistics for these five screened models are calculated and the model $\mu^{(8)}$ is found to be statistically better than other models.

The model involving non-competitive HA inhibition is the best model describing the HA production. The identified model is checked for parameter identifiability using the

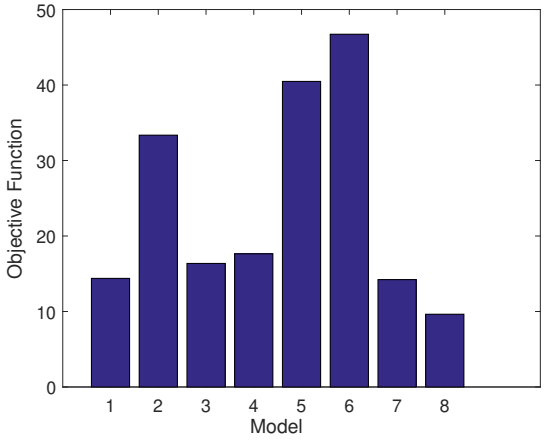

Fig. 1. Objective function values for discrimination of specific growth rate models

Table 4. Optimal parameter ($\boldsymbol{\gamma}$) values for the kinetic model of HA production

$\boldsymbol{\gamma}$	value	units	$\boldsymbol{\gamma}$	value	units
μ_m	0.562	h^{-1}	α_A	0.761	$\frac{g}{g\ biomass}$
k_S	2.892	$\frac{g}{L}$	α_{HA}	0.459	$\frac{g}{g\ biomass}$
$k_{i,HA}$	0.636	$\frac{g}{L}$	α_F	2.214	$\frac{g}{g\ biomass}$
G_a	9.899	$\frac{g}{g\ biomass}$	α_E	2.19	$\frac{g}{g\ biomass}$

DAISY software. The output of the DAISY indicates that the selected model is all algebraically observable and globally identifiable. Next, The model parameters are refined using the simultaneous non-linear least-squares optimization. Table 4 contains the statically optimal values for all the parameters in the selected model. The predictive ability of the identified model is validated on the new experimental data as shown in Fig. 2. It can be observed that the model predicts all the metabolites and biomass during the transient period. This demonstrates that the high fidelity model is built using the proposed incremental identification approach. The model can be used to perform online state and parameter estimation (Vargas et al., 2014; Nuñez et al., 2013)

5. CONCLUSIONS

In this work, the incremental model identification approach has been extended to bioprocesses involving a single substrate and single biomass and multiple products. The proposed approach involves steps such as (i) estimation of $\mathbf{K}$, (ii) identification of specific growth rate structure, (iii) model identifiability. The application of the proposed approach has been successfully demonstrated on the experimental data obtained from HA production in a batch bioreactor. A high fidelity kinetic model explaining HA production in a batch reactor has been developed. The developed kinetic model has been validated on the experimental data for a new batch. Incremental model identification can be applied to complex networks, provided the measurement data are available. In the future, the proposed approach will be extended to more complex bioprocesses involving multiple substrates and biomass.

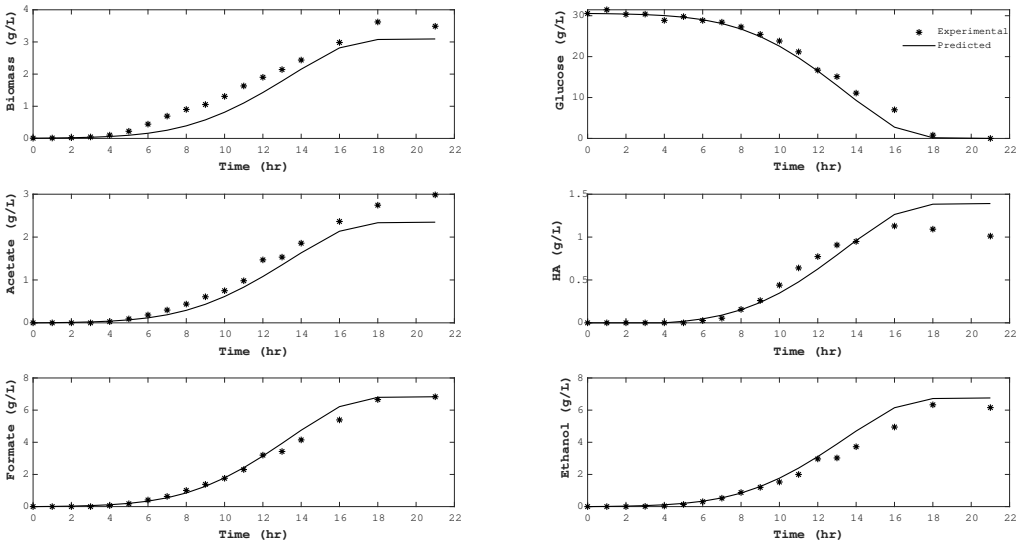

Fig. 2. Model prediction of metabolite concentrations compared to the experimental data

ACKNOWLEDGEMENTS

Ms. Kamakshi C is acknowledged financial support from the Ministry of Human Resource Development, GoI with the grant no. SB20210859BTMHRD008947.

REFERENCES

Asprey, S. and Macchietto, S. (2000). Statistical tools for optimal dynamic model building. *Computers & Chemical Engineering*, 24(2-7), 1261–1267.

Audoly, S., Bellu, G., D'Angio, L., Saccomani, M.P., and Cobelli, C. (2001). Global identifiability of nonlinear models of biological systems. *IEEE Transactions on Biomedical Engineering*, 48(1), 55–65.

Bardow, A. and Marquardt, W. (2004). Incremental and simultaneous identification of reaction kinetics: methods and comparison. *Chemical Engineering Science*, 59(13), 2673–2684.

Bellu, G., Saccomani, M.P., Audoly, S., and D'Angiò, L. (2007). DAISY: A new software tool to test global identifiability of biological and physiological systems. *Computer Methods and Programs in Biomedicine*, 88(1), 52–61.

Bhatt, N., Kerimoglu, N., Amrhein, M., Marquardt, W., and Bonvin, D. (2012). Incremental identification of reaction systems—a comparison between rate-based and extent-based approaches. *Chemical Engineering Science*, 83, 24–38.

Brendel, M., Bonvin, D., and Marquardt, W. (2006). Incremental identification of kinetic models for homogeneous reaction systems. *Chemical Engineering Science*, 61(16), 5404–5420.

Craven, P. and Wahba, G. (1978). Smoothing noisy data with spline functions. *Numerische Mathematik*, 31(4), 377–403.

Gábor, A., Villaverde, A.F., and Banga, J.R. (2017). Parameter identifiability analysis and visualization in large-scale kinetic models of biosystems. *BMC Systems Biology*, 11(1), 1–16.

Kaur, M. and Jayaraman, G. (2016). Hyaluronan production and molecular weight is enhanced in pathway-engineered strains of lactate dehydrogenase-deficient Lactococcus lactis. *Metabolic Engineering Communications*, 3, 15–23.

Mhamdi, A. and Marquardt, W. (2004). Estimation of reaction rates by nonlinear system inversion. *IFAC Proceedings Volumes*, 37(1), 167–172.

Nuñez, S., De Battista, H., Garelli, F., Vignoni, A., and Picó, J. (2013). Second-order sliding mode observer for multiple kinetic rates estimation in bioprocesses. *Control Engineering Practice*, 21(9), 1259–1265.

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

Saa, P.A. and Nielsen, L.K. (2017). Formulation, construction and analysis of kinetic models of metabolism: A review of modelling frameworks. *Biotechnology Advances*, 35(8), 981–1003.

Vargas, A., Moreno, J., and Wouwer, A.V. (2014). A weighted variable gain super-twisting observer for the estimation of kinetic rates in biological systems. *Journal of Process Control*, 24(6), 957–965.

Varghese, A., Narasimhan, S., and Bhatt, N. (2018). A priori parameter identifiability in complex reaction networks. *IFAC-PapersOnLine*, 51(15), 760–765.

Villaverde, A.F., Barreiro, A., and Papachristodoulou, A. (2016). Structural identifiability of dynamic systems biology models. *PLoS Computational Biology*, 12(10), e1005153.

Wang, F.S., Ko, C.L., and Voit, E.O. (2007). Kinetic modeling using s-systems and lin-log approaches. *Biochemical Engineering Journal*, 33(3), 238–247.

Zavrel, M., Schmidt, T., Michalik, C., Ansorge-Schumacher, M., Marquardt, W., Büchs, J., and Spiess, A.C. (2008). Mechanistic kinetic model for symmetric carboligations using benzaldehyde lyase. *Biotechnology and Bioengineering*, 101(1), 27–38.