Memorias del 2023 Congreso Nacional de Control Automático
25-27 de Octubre, 2023. Acapulco, Guerrero, México.


# Sparse Identification of Chemical Processes: A Feasibility Study ⋆

**Itzae Hernández-Fuentes** [*,**] **Daniela Vera-Martínez** [*,**]
**Roberto G. Ramírez-Chavarría** [*] **Lizeth Torres** [*]
**Daniel Martínez-Gutiérrez** [***]

[*] *Instituto de Ingeniería, Universidad Nacional Autónoma de México, Coyoacán 04510, CDMX, México (e-mail: RRamirezC@iingen.unam.mx*
[**] *Facultad de Química, Universidad Nacional Autónoma de México, Coyoacán 04510, CDMX, México*
[***] *Facultad de Ingeniería, Universidad Nacional Autónoma de México, Coyoacán 04510, CDMX, México*

**Abstract:** In the ongoing search for efficient techniques to identify and validate dynamic models in various fields of science and engineering, the sparse identification of nonlinear dynamics (SINDy) algorithm has emerged as a promising tool to carry out this task. In this study, SINDy is applied to identify two chemical processes: pH neutralization and temperature control in a bioreactor. To accomplish this, firstly, the dynamic behavior of each system was modeled, and data were collected for both processes. Then, the studied algorithm was employed to identify the dynamic models of each process. Then, the identified models by SINDy were validated by comparing them with first-principles models. Several tests were conducted to assess the capability of SINDy-based models in predicting the dynamic behavior of the pH and temperature processes. The results revealed a satisfactory agreement between the models obtained by SINDy and the first-principles models. Hence, we show the SINDy algorithm could be reliable tool for identifying dynamic models for chemical processes.

*Keywords:* System identification, nonlinear systems, chemical processes, data-driven models

## 1. INTRODUCTION

The fundamental problem in control engineering is to determine an action at the input of the system in such a way that the output exhibits a desired response. This is achieved through a controller, which assumes knowledge of the dynamics of the system to be controlled. This assumption makes system identification studies significantly important because, without an appropriate estimation stage, the basic control problem cannot be autonomously achieved.

Thus, the objective of identification is to obtain a representation of the system based on the available data. Generally, this task carried out in two stages. The first stage, based on prior knowledge of the process, is to select the type of model where the predicted output depends on previous inputs and the set of parameters to be determined. In the second step, an optimization process minimizes the difference between the process outputs and those estimated by the model. A realistic situation is that there is some uncertainty associated with the obtained nominal model. Successful identification is highly system-specific, as it depends in an appropriate choice of model type, data set, and algorithm. Sometimes these techniques are difficult to interpret and cannot be easily applied to known or partially known first principles.

Recently, the sparse identification of nonlinear dynamics (SINDy) algorithm has emerged as an attractive data-driven modeling framework. SINDy works around a sparse regression to find the minimal number of terms from a catalog of candidate functions necessary to model the dynamics. Due to its reliance on a linear regression that promotes sparsity, this approach allows for the incorporation of partial knowledge of physics, such as symmetries, constraints, and conservation laws (e.g., mass, momentum, and energy conservation).

The SINDy algorithm produces interpretable and generalizable models of dynamic systems from limited data. Furthermore, SINDy has been widely applied to identify models for optical systems, fluid flows, quantum mechanics, and model predictive control (MPC) (Abdullah and Christofides, 2023). However, there is scarce information on SINDy for chemical dynamical processes, which may

⋆ This work was supported by the grant UNAM-PAPIME PE100523.

result as a powerful tool to deal with their identification and modeling for efficiency promoting.

In this work, we consider two interesting biochemical processes, specifically in the fermentation industry: i) pH and ii) temperature modeling. In the former case, a neutralization process was simulated, in a continuous stirred-tank reactor (CSTR) (Morales et al., 2022). On the other hand, for the temperature modeling, a fermentation process for ethanol production was simulated. Interestingly, both processes exhibit nonlinear dynamics, time invariance, and time delay, making it challenging to achieve effective control of these parameters (Michael, 1994).

The rest of the paper is organized as follows. Section 2 introduces the basis of SINDy. Afterwards, the pH and temperature systems are briefly explained in Section 3. In Section 4 the results and a thorough discussion are presented. Finally, Section 5 is devoted to the conclusions.

## 2. SPARSE IDENTIFICATION OF NONLINEAR DYNAMICS (SINDY)

The SINDy algorithm (Brunton et al., 2016), (Kaheman et al., 2020), (Champion et al., 2020), identifies fully nonlinear dynamics systems from measurement data. It considers a system of the form:

$$\frac{d}{dt}x(t) = f(x(t)), \tag{1}$$

where $x(t) \in \mathbb{R}^n$ represents the measurements of the system's state variables at the time $t$, and the non-linear function $f(x(t))$ represents the interactions between these state variables.

To determine the function, we need a dataset of measurements collected at different time instances (with one row per time measurement and one column per variable) and the temporal derivatives of these measurements, which will be grouped into two different matrices, $X = [x_1(t_m)\, x_2(t_m)\, \cdots\, x_n(t_m)]^\top$, and $\dot{X} = [\dot{x}_1(t_m)\, \dot{x}_2(t_m)\, \cdots\, \dot{x}_n(t_m)]^\top$, respectively, with $^\top$ denoting the transpose operator.

Next, a $\Theta(X)$ library is constructed from trajectory data $X$ as:

$$\Theta(X) = \begin{bmatrix} | & | & | & & | & | \\ 1 & X & X^{P2} & \cdots & \sin(X) & \cos(X) \\ | & | & | & & | & | \end{bmatrix}, \tag{2}$$

where each column represents a set of basis functions that are candidates to interpret the nonlinear dynamics of the system. The selection of these functions is arbitrary and may consist of linear $X$, polynomial $X^{P2}$, or trigonometric functions, for instance.

Afterwards, a sparse regression problem is set up to retrieve the sparse coefficient vectors $\xi_k$ that determine which nonlinearities are active in characterizing the dynamics. Hence, the regression problem is given by:

$$\xi_k = \arg\min_{\xi_k} \| \dot{X} - \xi_k \Theta(X) \|_2 + \lambda \| \xi_k \|_1, \tag{3}$$

where $\| \cdot \|_2$ and $\| \cdot \|_1$ are the norm 2 and 1, respectively, $\lambda > 0$ is a hyperparameter determining the strength of the regularization term $\lambda \| \xi_k \|_1$, which promotes sparsity in $\xi_k$.

Once all sparse vectors $\xi_k$ have been estimated, they can be collected in the sparse matrix $\Xi = [\xi_1\ \xi_2\ \cdots\ \xi_n]$, so that the system dynamics can be computed as:

$$\dot{X} = \Theta(X)\Xi, \tag{4}$$

which corresponds to the matrix notation of the system in (1), where the number of vectors equals the dimension of the state vector, $n$, so $k = 1, \ldots, n$. As a result, SINDy retrieves the model of each row of the governing equations as

$$\dot{x}_k = f_k(x) = \Theta(x^\top)\xi_k. \tag{5}$$

Note that each column of (2) requires a distinct optimization to find the sparse vector of coefficients $\xi_k$. Thus, the final form of the system governing equations is indicated by

$$\dot{x} = f(x) = \Xi^\top(\Theta(x^\top))^\top, \tag{6}$$

where it is evident that SINDy is modeling the nonlinear function $f(\cdot)$ in (1).

## 3. SIMULATION EXAMPLES

### *3.1 pH neutralization system*

The operation of the process takes place in a continuous stirred tank of constant volume V, where a basic flow rate $q_3(t)$ (titrant) is manipulated to mix it with an acidic flow rate $q_1(t)$ that acts as a disturbance to the process in order to obtain an output flow rate $q_4(t)$ with the required pH value. Meanwhile, a constant flow of buffering current $q_2(t)$ helps maintain the pH value insensitive to small additions of acids or bases. $pH_4$ represents the process output (Elameen et al., 2019).

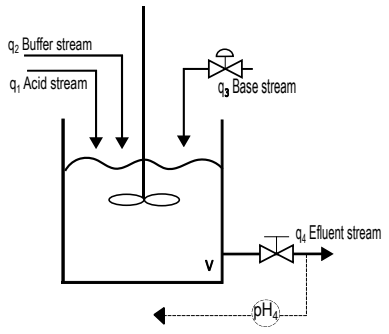


Fig. 1. The pH neutralization process

The chemical reactions in the system are as follows

$$H_2CO_3 \leftrightarrow HCO_3^- + H^+,$$
$$HCO_3^- \leftrightarrow CO_3^2 + H^+,$$
$$H_2O \leftrightarrow OH^- + H^+.$$

The equilibrium constants for the previous chemical reactions are defined as follows

$$K_{a1} = \frac{[\mathrm{H^+}][\mathrm{HCO_3^-}]}{[\mathrm{H_2CO_3}]}, \qquad K_{a2} = \frac{[\mathrm{H^+}][\mathrm{CO_3^{2}}]}{[\mathrm{HCO_3^-}]},$$
$$K_w = [H^+][OH^-].$$

In this document, the reaction invariants method is used, which is one of the most comprehensive descriptions for pH systems, focused on modeling certain acid-base processes Castelán and Gonzalez (2003).

From the concept of chemical equilibrium, the static mathematical model is developed by introducing the reaction invariant moments for each incoming flow

$$W_{ai} = [H^+]_i - [OH^-]_i - [HCO_3^-]_i - 2[CO_3^{2-}]_i \quad (7)$$
$$W_{bi} = [H_2CO_3]_i + [HCO_3^-]_i + [CO_3^{2-}], \quad (8)$$

where $i = 1, 2, 3, 4$ represent acid, buffer, base, and output flow, respectively. $W_{ai}$ represents the invariant charge, and $W_{bi}$ represents the concentration of carbonate ions, which are independent of the reaction extent.

The hydrogen ion concentration is computed using the equation

$$W_{ai} + W_{bi}\frac{\frac{[\mathrm{K_{a1}}]}{[\mathrm{H^+}]_i} + \frac{2\mathrm{K_{a1}K_{a2}}}{[\mathrm{H^+}]_i^2}}{1 + \frac{[\mathrm{K_{a1}}]}{[\mathrm{H^+}]_i} + \frac{\mathrm{K_{a1}K_{a2}}}{[\mathrm{H^+}]_i^2}} + \frac{\mathrm{K_w}}{[\mathrm{H^+}]_i} - [H^+]_i = 0. \quad (9)$$

Hence, the pH of the output stream is calculated through the hydrogen ion concentration $[H^+]_4$ using the following relationship

$$pH_4 = -\log_{10}([H^+]_4). \quad (10)$$

On the other hand, the linear dynamic model of the system is developed based on the material balance of the reactor as

$$V\frac{\mathrm{dW_{a4}}}{\mathrm{d}t} = q_1(W_{a1} - W_{a4}) + q_2(W_{a2} - W_{a4}) + q_3(W_{a3} - W_{a4}), \quad (11)$$

$$V\frac{\mathrm{dW_{b4}}}{\mathrm{d}t} = q_1(W_{b1} - W_{b4}) + q_2(W_{b2} - W_{b4}) + q_3(W_{b3} - W_{b4}). \quad (12)$$

Therefore, the mathematical model can be represented by the following nonlinear state-space model Kyu et al. (2004)

$$\dot{X} = f(x) + g(x)u + p(x)d, \quad (13)$$
$$c(x, y) = 0, \quad (14)$$

where $x = [W_{a4}\ W_{b4}]$; $\quad d = q_2$; $\quad u = q_3$;

$$f(x) = \begin{bmatrix} \frac{\mathrm{q_1}}{\mathrm{V}}(W_{a1} - x_1) \\ \frac{\mathrm{q_1}}{\mathrm{V}}(W_{a1} - x_2) \end{bmatrix}; \quad g(x) = \begin{bmatrix} \frac{1}{\mathrm{V}}(W_{a3} - x_1) \\ \frac{1}{\mathrm{V}}(W_{a3} - x_2) \end{bmatrix}$$

$$p(x) = \begin{bmatrix} \frac{1}{\mathrm{V}}(W_{a2} - x_1) \\ \frac{1}{\mathrm{V}}(W_{a2} - x_2) \end{bmatrix};$$

The implicit output equation between pH and reaction invariants can be written by the following polynomial approximation

$$c(x, y) = x_1 + 10^{y-14} - 10^{-y} + x_2 c_{x2} = 0, \quad (15)$$

with

$y = pH_4$; $\quad c_{x2} = \frac{1+2*10^{\mathrm{y-pk2}}}{1+10^{\mathrm{pk1-y}}+10^{\mathrm{y-pk2}}}$;

$p_{k1} = -\log_{10} k_{a1}$; $\quad p_{k2} = -\log_{10} k_{a2}$.

Table 1 shows the typical operating conditions of the process.

Table 1. Nominal operating conditions for the pH neutralization system.

| Parameter | Nominal Value | Parameter | Nominal value |
|---|---|---|---|
| V | 2900 mL | $q_3$ | 15.8 ml/s |
| $K_{a1}$ | $4.47 * 10^{-7}$ | $pH_4$ | 7 |
| $K_{a2}$ | $5.62 * 10^{-11}$ | $W_{a1}$ | 0.003 M |
| $[q_1]$ | 0.003 M HNO3 | $W_{b1}$ | $5 * 10^{-}5$ M |
| | $5 * 10^{-5} MNaHCO3$ | $W_{a2}$ | -0.03 M |
| $[q_2]$ | 0.03 M NaHCO3 | $W_{b2}$ | 0.03 M |
| $[q_3]$ | 0.003 M NaHCO3 | $W_{a3}$ | $-3.05 * 10^{-3}$ M |
| | $5 * 10^{-5} MNaHCO3$ | $W_{b3}$ | $5 * 10^{-5}$ M |
| $q_1$ | 16.6 ml/s | $W_{a4}$ | $-4.5 * 10^{-4}$ M |
| $q_2$ | 0.55 ml/s | $W_{b4}$ | $5.5 * 10^{-4}$ M |

### *3.2 Temperature modeling in a bioreactor*

We work with alcoholic distillation for ethanol production, where a batch reactor is used to carry out the process, various considerations are taken into account for the reactor design, such as achieving perfect mixing, maintaining a constant agitation speed, constant pH, and constant inflow and outflow rates. At the reactor inlet, there is an initial substrate concentration ($c_{s0}$) and biomass concentration ($c_{x0}$). Within the reactor, there are concentrations of biomass ($c_x$), substrate ($c_s$), which serves as the feed for microorganisms, and our product, ethanol ($c_p$), at the outlet, the concentrations of these components remain the same, and the reactor volume ($V$) is constant. For yeast cells to have good growth during fermentation, the presence of dissolved oxygen ($c_{O_2}$) is necessary.

In Fig. (2), we can observe the components present at the inlet and outlet of the bioreactor, as well as the flows within the bioreactor jacket

Since the reactor operates continuously, the global mass balance of the reactor is given by

$$\frac{\mathrm{d}v}{\mathrm{d}t} = F_i - F_e, \quad (16)$$

where $F_i$ is the inflow rate into the reactor and $F_e$ is the outflow rate from the reactor.

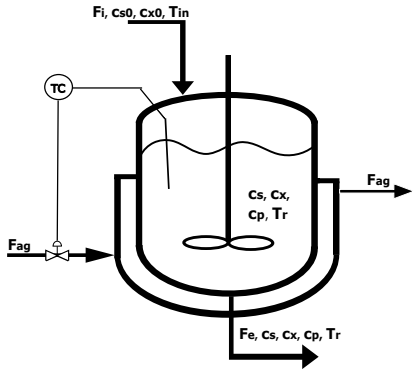


Fig. 2. Fermentation process for ethanol production

For simulation, both mass balance equations and energy balance equations are considered. The mass balance equations consist of the biomass balance

$$\frac{\mathrm{d}c_x}{\mathrm{d}t} = \mu_x c_x \frac{c_s}{K_s + c_s} e^{-K_p c_p} - \frac{F_e}{V} c_x, \tag{17}$$

where $\mu_x$ is the maximum specific growth rate, $K_s$ is the substrate constant for growth and $k_p$ is the growth inhibition constant by ethanol.

On the other hand, the ethanol balance is given by

$$\frac{\mathrm{d}c_p}{\mathrm{d}t} = \mu_p c_x \frac{c_s}{K_{s1} + c_s} e^{-K_{p1} c_p} - \frac{F_e}{V} c_p, \tag{18}$$

where $\mu_p$ is the maximum specific fermentation rate, $k_{s1}$ is the substrate constant for ethanol production and $k_{p1}$ is the fermentation inhibition constant by ethanol.

The substrate balance is written as

$$\begin{aligned} \frac{\mathrm{d}c_s}{\mathrm{d}t} &= -\frac{1}{R_{sx}} \mu_x c_x \frac{c_s}{K_s + c_s} e^{-K_p c_p} - \frac{1}{R_{sp}} \\ &\frac{c_s}{k_{s1} + c_s} e^{-K_{p1} c_p} + \frac{F_i}{V} c_{sin} - -\frac{F_e}{V} c_s, \end{aligned} \tag{19}$$

where $R_{sx}$ is the ratio of cell produced per glucose consumed for growth, and $R_{sp}$ is the ratio of ethanol produced per glucose consumed for fermentation.

Furthermore, the dissolved oxygen balance follows that

$$\frac{\mathrm{d}c_{O_2}}{\mathrm{d}t} = (K_L a)(c^*_{O_2} - c_{O_2}) - r_{O_2}, \tag{20}$$

where $c_{O_2}$ is the oxygen concentration in the liquid phase, $c_{O_2}$ is the equilibrium concentration of oxygen in the liquid phase, $K_{La}$ is the product of mass-transfer coefficient for oxygen, and gas-phase specific area and $r_{O_2}$ is the rate of oxygen consumption.

The energy balance for the reactor is given by the equation (20).

$$\begin{aligned} \frac{\mathrm{d}T_r}{\mathrm{d}t} &= \frac{F_i}{V}(T_{in} + 273) - \frac{F_e}{V}(T_r + 273) + \frac{r_{O_2} \Delta H_r}{32 \rho_r C_{heat,r}} \\ &+ \frac{K_T A_T (T_r - T_{ag})}{V \rho_r C_{heat,r}}, \end{aligned} \tag{21}$$

Table 2. Nominal operating conditions for the temperature system.

| Parameter | Initial values | Parameter | Initial values |
|---|---|---|---|
| $m_{NaCl}$ | 500 g | $F_{ag}$ | 18 L h$^{-}$1 |
| $m_{CaCO_3}$ | 100 g | $T_{in}$ | 25° C |
| $m_{MgCl_2}$ | 100 g | $T_{in,ag}$ | 15° |
| $pH$ | 6 | $c_{sin}$ | 60 g $L^1$ |
| $F_i = F_e$ | 51 L h$^{-}$1 | - | - |

where $T_r$ is the volume of the jacket, $T_{in}$ is the temperature of the substrate flow entering to the reactor, $\Delta H_r$ is the reaction heat of fermentation, $\rho_r$ is the density of the mass of reaction, $C_{heat,r}$ is the heat capacity of mass of reaction, $K_T$ is the heat transfer coefficient and $A_T$ is the heat transfer area.

Finally, the energy balance for the jacket is given by

$$\frac{\mathrm{d}T_{ag}}{\mathrm{d}t} = \frac{F_{ag}}{V_j}(T_{in,ag} - T_{ag}) + \frac{K_T A_T (T_r - T_{ag})}{V_j \rho_{ag} C_{heat,ag}}, \tag{22}$$

where $T_{ag}$ is the temperature of cooling agent in the jacket, $T_{in,ag}$ is the temperature of cooling agent entering to the jacket, $C_{heat,ag}$ is the heat capacity of cooling agent, $\rho K u_{ag}$ is the density of cooling agent, $F_{ag}$ is the flow of cooling agent, and $V_j$ is volume of the jacket.

For the simulation, there are some additional equations (such as considerations of ionic forces) as well as required values for the model to work, as indicated in the work by Nagy (2007). However, the initial input values that must be known primarily are provided in Table 2.

## 4. RESULTS AND DISCUSSIONS

The methodology for system identification consists of 4 steps: data acquisition, model selection, estimation, and model validation. To study the performance of SINDy, we computed the root-mean-squared error (RMSE) for both processes. The RMSE is given by:

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{(\hat{y} - y)^2}{N}}, \tag{23}$$

where $\hat{y}$ is the estimated output, $y$ is the actual output and $N$ is the length of the signal.

### *4.1 pH neutralization*

*Data generation* This stage of the work involves conducting experiments in the system to collect input and output data of the variables of interest using a planned excitation signal. The staircase test is chosen in this work, which consists of exciting the process with ascending and descending step-like signals. The advantage of this type of signal is that it allows observing the dynamics of the process in each area.

In the case of the conducted test, the range of the base flow rate is: 11.1 to 16.8 mL/s, and the corresponding output range is: 5.11 to 8.05 High nonlinearity is observed in the process, where the system's output values are not

uniform. In the face of unit changes in the input variable, there is high sensitivity and abrupt changes in pH. The I/O data is shown in Fig. 3.

The SINDy algorithm identified the system in the following third-degree polynomial space that establishes the relationship between the input u and the system's output y:

$$f(u,y) = -0.563216 - 0.002341u + 0.277789y \\ -0.045512y^2 + 0.002278y^3 + 0.000948uy \quad (24)$$

During the simulation, model degenerations were observed when including higher-order polynomial terms. The optimal lambda parameter results from the best candidate adapted to the estimation and validation data, with a value of $\lambda = 5 \cdot 10^{-4}$.

In Fig. 4(a), it can be observed that the model obtained with the SINDy algorithm well fits the estimation data using a stair caise input. On the other hand, for the validation of the SINDy-based model structure, a pseudo-random binary sequence (PRBS) signal was generated to excite the system. Fig. 4(b) shows that SINDy fulfills the identification task. Thus, with this test, we demonstrate that the model retrieved by SINDy captures the dynamics as it adapts to any type of variation in the base flow input.

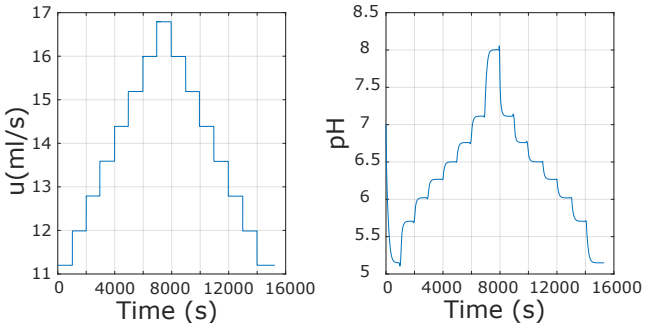


Fig. 3. Input-output data for identifying the pH process.

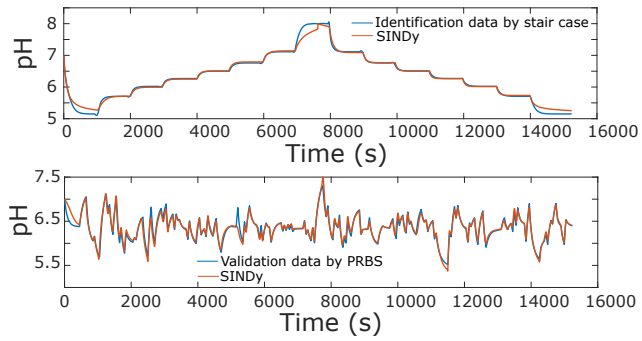


Fig. 4. Estimation and validation data for the pH process.

Finally, to quantify the performance of the identification, the RMSE was 0.0723 for the pH neutralization system. Hence, as this value is as close to zero, the goodness of the fit indicates agreement between the output of the SINDy model, and the experimental data generated by simulating the system.

### *4.2 Temperature modeling in a bioreactor*

As show in Fig. 5, there are six different states of the dynamical system, including concentrations of yeast, substrate, oxygen, and ethanol, as well as the final temperatures of the reactor and its jacket. With five input signals to the system, we can identify it as a multiple-input, multiple-output (MIMO) system. However, for practical purposes, we worked with a single-input single-output (SISO) system. For this purpose, the reactor temperature was considered the output, i.e. as the variable controlled, while the feed flow to the reactor jacket was taken as the manipulated variable.

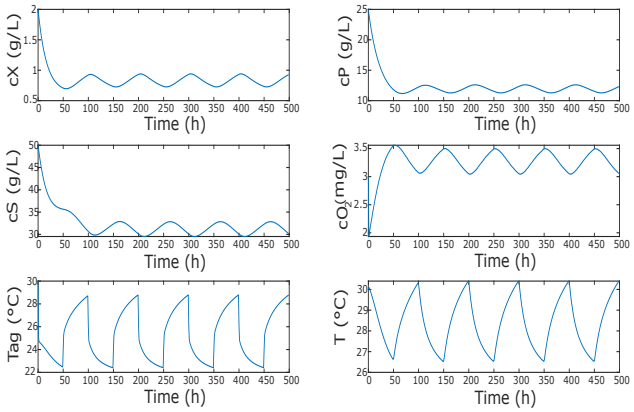


Fig. 5. Dynamic response of the reactor for temperature modeling.

Taking as a reference Rahman (2013) and Lorenzo (2020), it was concluded that an optimal temperature for the reactor to operate correctly would range between 25 and 30 °C. At these temperatures, the system is favored with a higher ethanol concentration at the outlet. Additionally, this is supported owing that at higher temperatures the yeast would deactivate, and the fermentation reaction could not be completed.

A stair case signal was applied as the system input, ranging from 10 to 48 L/h. Meanwhile, the response signal shows a change in the reactor temperature, in the range from 31 to 26 °C, which is an optimal temperature for the fermentation process.

In line with our research problem, the main objective was to obtain a dynamic model that describes the variations in reactor temperature as a function of the coolant flow through the reactor jacket. For this purpose, three alternatives of model classes were considered:

- State-space (SS) model
- Output-error (OE) model

- SINDy model

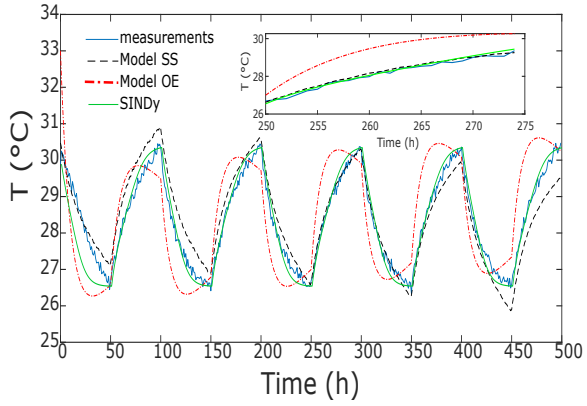


Fig. 6. Temperature response of the reactor.

The results of the bioreactor identification process are shown in Fig. 6, which compares the output of the selected model classes and the data generated by the first-principles model. To determine which model adapted best to the data, the RMSE was computed. As a result, the SS model exhibited an RMSE of 0.36, whereas the SINDy-based model retrieved RMSE of 0.32. These were chosen as the best models for identifying the dynamic model of the process. However, through various tests conducted, it was found that SINDy outperforms the other model classes. That is, When changing $F_{ag}$ to obtain a different response in $T_r$, the SS model did not adapt to these changes, whereas SINDy was able to adapt in a wide range of operating points.

With the above clear, the sparse identification algorithm identified the system in a second-degree polynomial space. Which is an approximation of the energy balance for the reactor temperature shown in the main model (20).

$$f(u,x) = -31.232705 * x + 3.727858 * u^2 \\ +1.092737 * x^2 - 0.074973 * u^3 - 0.012750 * x^3 \quad (25)$$

with an optimal parameter $\lambda = 5 \cdot 10^{-4}$.

## 5. CONCLUSION

The study demonstrated that SINDy is an effective tool for the identification and validation of chemical processes in the industry. Real data was generated through the modeling of nonlinear process dynamics, which SINDy uses to search for candidate functions of the state variables. One of the advantages of SINDy is its ability to generate parsimonious models, meaning models that use a low number of parameters and functions. This is particularly useful when the true dynamics of the system are not fully known. Compared to other identification techniques simulated in this study, SINDy stands out in its ability to produce models that retain the physical meaning of the parameters used in the regression. As demonstrated, the generated models are easily interpretable and understandable.

## REFERENCES

Abdullah, F. and Christofides, P.D. (2023). Data-based modeling and control of nonlinear process systems using sparse identification: An overview of recent results. *Computers and Chemical Engineering*, 174. doi: 10.1016/j.compchemeng.2023.108247.

Brunton, S.L., Proctor, J.L., and Kutz, J.N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *The Proceedings of the National Academy of Sciences*, 113, 3932–3937. doi:10.1073/pnas.1517384113.

Castelán, R. and Gonzalez, B. (2003). Modelado de un sistema de control de ph. *Memorias del Congreso Nacional de Control Automático*, 3, 67–72.

Champion, K., Zheng, P., Aravkin, A.Y., and Brunton., S.L. (2020). A unified sparse optimization framework to learn parsimonious physics-informed models from data. *IEEE Access*, 8, 169259–169271. doi: 10.1109/access.2020.3023625.

Elameen, M.H.E., Hago, M.B.M., and Tan, W. (2019). ph neutralization process control based on active disturbance rejection control. *Advances in Science, Technology and Engineering Systems Journal*, 4, 104–108. doi:10.25046/aj040513.

Kaheman, K., Kutz, J.N., and Brunton, S.L. (2020). Sindy-pi: a robust algorithm for parallel implicit sparse identification of nonlinear dynamics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 476, 202–279. doi:0.1098/rspa.2020.0279.

Kyu, K.D., Soon, L.K., and Ryook, Y.D. (2004). Control of ph neutralization process using simulation based dynamic. *Korean Journal of Chemical Engineering*, 21, 942–949. doi:10.1007/BF02705575.

Lorenzo, M.H. (2020). *Determinación de los parámetros de etanol en el proceso de fermentación y destilación para la producción artesanal de POX con sabor*. Ph.D. thesis, Instituto Tecnológico de Tuxtla.

Michael, A.H. (1994). Adaptive nonlinear control of a ph neutralization process. *IEE Transactions of Control Systems Technology*, 2, 169–182. doi: 10.1109/87.317975.

Morales, L., Estrada, J.S., and Herrera, M. (2022). Hybrid approaches-based sliding-mode control for ph process control. *ACS Omega*, 7, 45301–45313. doi: 10.1021/acsomega.2c05756.

Nagy, Z.K. (2007). Model based control of a yeast fermentation bioreactor using optimally designed artificial neural networks. *Chemical Engineering Journal*, 127(1), 95–109. doi:10.1016/j.cej.2006.10.015.

Rahman, S.S. (2013). Effect of various parameters on the growth and ethanol production by yeasts insolated from natural sources. *Bangladesh Journal of Microbiology*, 30(1y2), 49–54. doi:10.3329/bjm.v30i1-2.28453.