Nonlinear System Identification in coherence with Nonlinearity measure for Dynamic Physical systems- Case Studies

Joanofarc Xavier
Anna University Chennai
Rames C Panda (panda@clri.res.in)
Central Leather Research Institute CSIR
SK Patnaik
Anna University Chennai

Research Article

Keywords: kSINDYc, nonlinear least square, nonlinear dynamics, system identification, nonlinearity measure, sensitivity analysis

Posted Date: September 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-945119/v1

Nonlinear System Identification in coherence with Nonlinearity measure for Dynamic Physical systems- Case Studies

Joanofarc Xavier[1], S.K Patnaik[1], Rames C Panda[2]
[1]Dept of EEE, College of Engineering, Anna University, Chennai-600025, India
Email: joanjoean@gmail.com, skpatnaik@annauniv.edu
[2]Sr Principal Scientist & Faculty, Dept of Chemical Engineering, CSIR-Central Leather Research Institute,
Adyar, Chennai, 600020, India, Email: panda@clri.res.in

Highlight of research:

1. The submission serves as a bridge to fill the leveraging gap between the computation of nonlinearity and the suitable choice of nonlinear system identification for nonlinear dynamic physical systems.
2. Given any nonlinear process, the CANM (Convergence Area based Nonlinear Measure) method categorizes it under a particular class of nonlinearity quantified within the range 0 to 1.
3. A notable development on the (Sparse Identification of Nonlinear Dynamics with control) SINDYc candidate library function, by introducing the *'key nonlinear term'* from the plant dynamics (kSINDYc), apart from the other higher order polynomials of the processes.
4. The research also extends to nonlinear system identification with the nonparametric N3ARX (Neural network based nonlinear auto regressive exogenous input) scheme and the most popular parametric NL2SQ (Nonlinear Least Square) method for five different dynamic processes with different levels of nonlinearity.
5. Finally, for all the case studies, a qualitative and quantitative relation is made between the nonlinear system identification methods and the nonlinearity measures.

Abstract:

With the recent success of using the time series to vast applications, one would expect its boundless adaptation to problems like nonlinear control and nonlinearity quantification. Though there exist many system identification methods, finding suitable method for identifying a given process is still cryptic. Moreover, to this notch, research on their usage to nonlinear system identification and classification of nonlinearity remains limited. This article hovers around the central idea of developing a 'kSINDYc' (key term based Sparse Identification of Nonlinear Dynamics with control) algorithm to capture the nonlinear dynamics of typical physical systems. Furthermore, existing two reliable identification methods namely NL2SQ (Nonlinear least square method) and N3ARX (Neural network based nonlinear auto regressive exogenous input scheme) are also considered for all the physical process-case studies. The primary spotlight of present research is to encapsulate the nonlinear dynamics identified for any process with its nonlinearity level through a mathematical measurement tool. The nonlinear metric Convergence Area based Nonlinear Measure (CANM) calculates the process nonlinearity in the dynamic physical systems and classifies them under mild, medium and highly nonlinear categories. Simulation studies are carried-out on five industrial systems with divergent nonlinear dynamics. The user can make a flawless choice of specific identification methods suitable for given process by finding the

Δ_0). Finally, parametric sensitivity on the measurement has been studied on CSTR and Bioreactor to evaluate the efficacy of kSINDYc on system identification.

Keywords: kSINDYc, nonlinear least square, nonlinear dynamics, system identification, nonlinearity measure, sensitivity analysis.

1 Introduction:

Most of the physical systems we come-across are non-linear in nature. Safe operational practice of various industrial units needs mathematical modelling of the physical system, optimization and design of control system. Generally, tools for analyzing nonlinear systems, like, describing function, phase portrait, perturbation, stability criteria (Lyapunov or Popov), passivity are well-established. However, some of the existing units demands for fault diagnosis and model order reduction of complex systems. The design of controller or achieving better closed loop-performance requires parametric identification/estimation. It becomes difficult to select suitable identification techniques for a given system with nonlinear/unknown dynamics. In the past, researchers contributed a lot of work in obtaining the mathematical model of a nonlinear system from its first principle concepts. The first principle concept was desirable for only processes where adequate knowledge of process is available. In many of the process industries, whose internal functions are complex to understand, formulate and compute, parametric and nonparametric nonlinear system identification are adopted, by providing the measured input and output data.

The main advantage of nonlinear-system-identification is the fact that, even, if an unknown system is given, just with the measured input and output data, the nonlinear dynamics of the process can be retrieved accurately. In recent years, many literatures have brought out the features, pros and cons on the usage and complexity of many notable identification algorithms for nonlinear system. To manifest a few, Schoukens and Ljung presented a review on identification methods of linear and nonlinear systems [1]. The article also indexed an exemplary summary of many parametric identification methods. Block oriented nonlinear models can be classified under (i) Hammerstein (ii) Wieners (iii) Voltera-series. The review article conferred by [2], not only portrayed the block-oriented identification methods, but also delivered a deep thought on the most prominent nonlinear control schemes of recent time. A non-stochastic subspace algorithm was considered for multi-dimensional nonlinear system identification based on measured output data. However, the procedure was not tested for systems with different structural nonlinearities [3].

The autoregressive models with exogenous inputs are employed in applications where state transitions are triggered by external events [4]. Stochastic gradient parametric estimation using moving window data was presented in [5] to estimate the system's response to discrete measured data. However, the effectiveness of the method was shown only by using numerical examples and not on physical systems. The identification of LPV time-delay systems with missing output data using multiple-model approach is framed in [6]. Output-error (OE) model representing the process dynamics of CSTR and continuous fermenter, are recovered using the expectation-maximization (EM) algorithm to obtain the final global model. Reference [7] is concerned with the parametric identification of a special class of nonlinear-systems called as bilinear state space systems. Parametric identification of time-delay-systems were discussed in [8]. Multi-innovation theory is put forward in stochastic gradient algorithm based on state observer and recursive least-squared identification algorithm to improve their accuracy and convergence rate. In another work by [9], a

generalized identification scheme for integral-order systems is utilized for identification of fractional-order nonlinear systems with both non-chaotic and chaotic behaviors. Being under the class of black-box modeling, Hammerstein-Wiener models can be employed for identification of complex nonlinear systems with static nonlinearity as well as dynamic linear regions [10,11].

Machine learning approaches are very powerful tools to identify a variety of highly nonlinear systems. The approaches come out with high fidelity models, that reflect the underlying physics of the nonlinear system. Many standard machine learning methods have shown spectacular performance in predicting dynamics of any interpolated system, but the resulting models usually lack generalizability and interpretability [12,13]. Recently, in one article by [14], the authors have reviewed system identification in context to powerful tools of computational intelligence methods which include genetic algorithm, particle swarm optimization and differential evolution. A variety of highly nonlinear occurrences are contemplated to assess the competence and the fast computing intelligence of genetic programming in [15].Takagi-Sugeno(TS) fuzzy modelling with unscented Kalman filter was carried over for a practical heat exchanger process in [16]. Yet, the real challenge lies on the choice of fuzzy rule numbers on the output precision.

Nonetheless there are several identification methods where the real challenge lies in developing a parsimonious model with the smallest possible number of parameters that can adequately describe the dynamics of the physical system. Also, the confrontation lies in determining the underlying dynamics of the process from the measured data.

1.1 Motivation:

The above discussion reveals that there are indigenous number of articles that discuss the concepts of system identification and measurement of nonlinear metric [21] in separate attempts. There exist enumerable identification methods and methods for quantification of nonlinearity. However, well-designed directions or guidelines for selection of identification algorithms (based on nonlinearity-measure) are rare and need to be established. Past literature explains that there exists a break in the continuity between these two concepts for (as mentioned above) over years. The motivation for this research comes up with a bang by readdressing the issues of system identification and the concept of nonlinear metric in a jointed venture. The research idea discussed in this article will overcome the existing disruption by relating the nonlinear metric (Δ_0) with the noteworthy system identification tools.

1.2 Contributions:

In this respect, we have established three significant system identification methods namely key term kSINDYc, N3ARX and NL2SQ methods to identify peculiar nonlinear systems from the process engineering glebe. SINDYc (Sparse Identification of Nonlinear Dynamics with control) algorithm is a symbolic sparse regression problem, which may be susceptible to over fitting problem if care is not taken to balance the model complexity [17]. This major concern is drenched here, by choosing a fewer number of relevant key terms in the candidate library of the SINDYc scheme. This paper also addresses this issue by providing the perfect choice of key terms based on the degree and type of nonlinearity of dynamic nonlinear systems. Moreover, the performance index of all the methods are correlated along with the degree of nonlinearity of each process and the outcomes are enumerated.

The paper is divided into six sections. Section 1 has introduced the literature review of many system identification methods. Section 2 examines about the CANM method to measure nonlinearity level Δ_0 .Section 3 gives a brief review of the kSINDYc, NL2SQ and N3ARX approaches. It is followed by simulation results in section 4 which show that Δ_0 as well as RMSE witness a major lively role in deciding the choice of nonlinear identification method for the five dynamic systems with contrasting nonlinearity. Besides, Section 4 also adds increased flavour to the current study by suggesting the suitable parsimonious model for every process. Section 5 narrates the significance of parametric sensitivity analysis of physical systems and section 6 concludes the article.

2 Proposed CANM Method Δ_0 :

The nonlinearity of the physical systems is an important issue to be addressed in controller design, bifurcation and uncertainty analysis. It varies with respect to the initial condition of state variables, excitation signals given, and input constraints associated with it. This research brings out the strength of nonlinearity of typical industrial process and its impact on the popular system identification schemes. Nevertheless, there are several nonlinear indices to mark the value of nonlinearity in the dynamic systems [18-20]. The concept of Convergence-area-based-nonlinearity measure (CANM) has been endorsed in the current study [21].

Without loss of generality, consider a nonlinear dynamic system of the form

$$\frac{dx(t)}{dt} = f\left(x(t), u\right) \tag{1}$$

in which $x(t) \in \mathbb{R}^n$ denotes the state variables of a system at time t .Eq. (1) also generalizes the first principle model of nonlinear systems. If $y_{act}(t)$ and $y_{lin}(t)$ represent the actual output of the nonlinear system and its linearized response at the j^{th} operating point P_j, then the nonlinear metric Δ_0 quantifies the level of nonlinearity as given in Eq. (2).

$$\Delta_0 = \frac{\left\| \left| \int_0^{t_f} y_{act} dt \right| - \left| \int_0^{t_f} y_{lin} dt \right| \right\|}{\left| \int_0^{t_f} y_{act} dt \right|} \tag{2}$$

Where t_f represents the final time period of the nonlinear system. For a nonlinear process with m number of operating sectors such as P_1, $P_2 \ldots P_j \ldots P_m$, the overall nonlinearity Δ_{0nom} is shown in Eq. (3)

$$\Delta_{0nom} = \frac{\Delta_{0P_1} + \Delta_{0P_1} + \ldots \Delta_{0P_j} + \ldots \Delta_{0P_m}}{m} \tag{3}$$

The CANM method conferred in this work stands distinct for its amenability in dealing with wide range of nonlinear dynamic systems. The method uses Jacobian linearization to find out linear approximation $y_{lin}(t)$ which thoroughly depends on analysis of an operating point. The stability of the operating point decides the current dynamic behavior of the process. As many chemical, biomedical and biological processes often operate on a predesigned operating region with multiple operating points, this CANM method will be most beneficial to them.

3 Compendium on Data driven Identification Schemes:

[Insert Fig. 1 here]

Fig. 1 presents a schematic view of the approach to show the steps involved in the proposed view of nonlinear-system- identification and nonlinearity-quantification in a single framework. In this figure, the terms k_{para} and x_{init} denote the nominal input parameters and initial states of the nonlinear system respectively. The excitation signal u and the nonlinear output $y_{act}(t)$ from the nonlinear process are treated as measured input and output data. Region I contain the proposed kSINDYC algorithm to learn the dynamics of $y_{act}(t)$. The most essential term in the kSINDYc library, which plays a critical role in determining the nonlinear dynamics, is weighed from the governing equations of the process. The predicted output $y_{pred}(t)$ of Region I and measured output $y_{act}(t)$ are used to calculate performance using *RMSE* criteria. On the flip side, the nonlinear system under study is linearized about its operating point P_j using Jacobian linearization and is expressed as $y_{lin}(t)$. The Δ_0 is computed using CANM method as given in Eq. (2). On completion of the learned dynamics using kSINDYc, Region I of Fig. 1 is replaced by N2LSQ and N3ARX identification methods. A graphical plot is made between Δ_0 and RMSE to fill the leveraging gap between computation of nonlinearity and the suitable choice of identification for nonlinear dynamic physical systems.

3.1 Key term based SINDYc (kSINDYc):

The recent impeccable SINDYc algorithm is a celebrated parsimonious system identification technique introduced by Brunton [22]. Abundant collection of technical records is garnered with widespread curiosity on the remarkable progress made in sparse dynamics in many disciplines ranging from biology to control engineering [23-25]. Backdrop in this section, we provide a brief retrospect to the SINDYc algorithm, which forms the bottom line of the proposed 'kSINDYc' system identification methodology. Inspired by its application to many physical systems, regression problem using SINDYc is formulated as follows:

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

$\dot{X}$ gives the derivatives of state variables. Θ is the library function with all the candidate terms in kSINDYc library. For a time period $t = \left[t_1, t_2 \ldots t_f\right]^T$, consider the input vector $u = \left[u_1(t), u_2(t) \ldots u_p(t)\right]^T \in R^p$, the state vector $x(t) = [x_1(t) x_2(t) \ldots x_n(t)]^T \in R^n$, then

$$\Xi = \left[\xi_1 \ \xi_2 \ \cdots \ \xi_n\right] \tag{5}$$

Eq. (5) is vector that has the sparse co-efficients $\xi_1 \ \xi_2 \ \cdots \ \xi_n$ corresponding to $\Theta(x,u)$.

$$X = \begin{bmatrix} x^T(t_1) \\ x^T(t_2) \\ \vdots \\ x^T(t_f) \end{bmatrix} = \begin{bmatrix} x_1(t_1) & x_2(t_1) & \ldots & x_n(t_1) \\ x_1(t_2) & x_2(t_2) & \cdots & x_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(t_f) & x_2(t_f) & \cdots & x_n(t_f) \end{bmatrix} \tag{6}$$

$$U = \begin{bmatrix} u^T(t_1) \\ u^T(t_2) \\ \vdots \\ u^T(t_f) \end{bmatrix} = \begin{bmatrix} u_1(t_1) & u_2(t_1) & \ldots & u_p(t_1) \\ u_1(t_2) & u_2(t_2) & \cdots & u_p(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ u_1(t_f) & u_2(t_f) & \cdots & u_p(t_f) \end{bmatrix} \tag{7}$$

$$\dot{X} = \begin{bmatrix} \dot{x}^T(t_1) \\ \dot{x}^T(t_2) \\ \vdots \\ \dot{x}^T(t_f) \end{bmatrix} = \begin{bmatrix} \dot{x}_1(t_1) & \dot{x}_2(t_1) & \ldots & \dot{x}_n(t_1) \\ \dot{x}_1(t_2) & \dot{x}_2(t_2) & \cdots & \dot{x}_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \dot{x}_1(t_f) & \dot{x}_2(t_f) & \cdots & \dot{x}_n(t_f) \end{bmatrix} \tag{8}$$

The key terms in the library function $\Theta(x,u)$ are given in Eq. (9)

$$\Theta(x,u) = \left[1 \ x \ u \quad x \otimes x \quad x \otimes u \cdots k_{nl}\right]. \tag{9}$$

$x \otimes u$ denotes the vector of all product combinations in x and u .It also indicates the quadratic nonlinearities in the unknown system.

$$\xi_k = \underset{\hat{\xi}_k}{\operatorname{argmin}} \frac{1}{2} \left\| \dot{X}_k - \hat{\xi}_k \Theta^T(X,U) \right\|_2^2 + \lambda \left\| \hat{\xi}_k \right\|_1 \tag{10}$$

The optimization problem given in Eq. (10) can be evaluated using sparsity promoting scheme called *STLS* (Sequential Threshold least square method). The second part of Eq. (10) has the penalty term with the tunable weighing parameter $\lambda \geq 0$ to establish model parsimony. ξ_k represents k^{th} row of Ξ and $\dot{X}_k$ represents k^{th} row of $\dot{X}$

The algorithmic pseudocode for the proposed kSINDYc identification is given below:

Algorithm A1. key term SINDYc (kSINDYc)

Require: The time series data u and y for n number of sets

Input : Candidate function Θ ;time derivatives $\dot{X}$;sparcification tuner λ ; threshold ξ ; no of *iters*

Output : Sparse coefficient vectors Ξ , RMSE

1: Procedure 'kSINDYc'

2: *for* $i = 1, ..., iters$ do

3: Evaluate Θ , the library function using data x and $u = u_{nom}$ given in Eq. (6) and (7)

4: Include the key nonlinearity term k_{nl} in library Θ

5: **Build** the matrix $\dot{X} = \Theta(x, u)$

6: Compute $\dot{X}$ as given in Eq. (4)

7: **Solve** the sparse regression problem ξ as given in Eq. (10) using *STLS* method

8: **Procedure** ' *STLS* ' to evaluate ξ

a. Start the least square solution for Ξ

b. Fix the cut off value for λ as $\lambda_{\min} < \lambda \leq \lambda_{\max}$

c. Threshold all the coefficients $\xi_k < \lambda$ (*i.e*) $\xi_k = 0$

e. Identify non-zero coefficients (*i.e*) $\xi_k \neq 0$ and obtain another least square solution for Ξ

f. Threshold the new cut-off using λ

g. Repeat steps *a to f* until the non-zero coefficients converge.

end procedure *STLS*

9: **Calculate** the coefficients of all states Ξ

10: end *for*

11: Repeat the above steps for PRBS excitation u_{prbs}

12: Compute the fitness function RMSE

13: end procedure 'kSINDYc'

A coarse sweep of the tunable parameter λ in the range $\lambda_{\min} < \lambda \leq \lambda_{\max}$ is considered in Algorithm A1 to attain an optimal solution with minimum error and maximum convergence rate. The specific convergence criteria for *STLS* algorithm in SINDYc framework is provided in [26]. kSINDYc can handle large candidate library with the regularizing tuner λ. Not limitingly, the choice of the sparsity knob λ is made in such a way that there is a tradeoff between accuracy and complexity of the kSINDYc algorithm.

3.2 NL2SQ method:

Many works on parametric system identification have used least square method to estimate the numerical values of parameters [27-29]. Other techniques to estimate the parameters of the physical systems include Subspace identification methods and Hammerstein-wiener models, [4,12]. NL2SQ with Levenberg–Marquardt algorithm is another established method used for optimizing the process parameters in the field nonlinear system identification [16, 28, 30]. The

$y_{act}(t)$ and $y_{pred}(t)$ given by

$$F(\theta) = \frac{1}{2}\sum_{i=1}^{m}\left\|f_i(\theta)\right\|_2^2 = \frac{1}{2} f(\theta)^T f(\theta) \tag{11}$$

The model parameters θ are estimated with the basic requirement of minimizing the objective function in Eq. (11).

$$\nabla F(\theta) = f(\theta)^T J(\theta) \tag{12}$$

where $J(\theta)$ is the $m \times p$ Jacobian matrix with the transformation $f : \mathbb{R}^m \times \mathbb{R}^p$

The second derivative of Eq. (11) is

$$\nabla^2 F(\theta) = J(\theta)^T J(\theta) + \sum_{i=1}^{m} f_i(\theta) H_i(\theta) \tag{13}$$

Where $H_i(\theta) = f_i''(\theta)$ is $p \times p$ Hessian Matrix with the function $f_i(\theta)$

The Jacobian matrix $J(\theta)$ of $H(\theta)$ has to be found out to optimize θ for m number of samples. Using LM algorithm, the objective function for NL2SQ method is modified as h_{LM} [29].

$$(J^T J + \mu I) h_{LM} = -J^T f \tag{14}$$

$$h_{LM} = -(J^T J + \mu I)^{-1} J^T f \tag{15}$$

The damping factor is always $\mu \geq 0$,for which the following effects are observed. When $\mu > 0$, the co-efficient $(J^T J + \mu I)$ is positive definite and so h_{LM} is in descent direction. If μ is very large $h_{LM} = -\frac{1}{\mu} J^T f$ and goes into the steepest descent direction. On the other hand, if μ is very small $h_{LM} = h_{GN}$, the LM algorithm converges with the Gauss Newton method. In the gradient descent method, the h_{LM} is minimized by refreshing the parameters in the steepest-descent direction. The gradients of the process are calculated using automatic differentiation. On the other hand, in the Gauss-Newton method h_{LM} is reduced by considering the least square module to be locally quadratic to its parameters and sorting out the minimum value from this quadratic term. The LM algorithm operates similar to gradient-descent method when the parameters are away from their optimal value, and behaves more like a Gauss-Newton scheme when the parameters are very near to the optimal point. It can be concluded that the LM algorithm involves the cross combination of gradient descent and Gauss-Newton methods.

3.3 N3ARX Method:

Neural Networks is another computational intelligence approach for identifying nonlinear systems in real world scenario with accurate estimations [31, 33, 34]. Neural Networks can be employed

without knowing any prior knowledge about the dynamics of the system The NARX method is a standard identification technique and is found in enormous literatures [32, 35]. A novel optimal identification algorithm is presented for NMPC based on the Neural network model for different operating regions of highly nonlinear dynamic processes in [36]. Hybrid combination of Neural network algorithm with NARX method is investigated in this research to make a strong comparison with the kSINDYc method of identification. A simple Neural network structure is taken with 1 hidden layer, and linear activation function in Matlab. The number of neurons required to identify each process will differ depending upon the nonlinearity and operating region. The number of hidden layer nodes in N3ARX method can be chosen iteratively.

4 Simulation study:

The simulation study is carried out here for typical Industrial processes from chemical to biological field. The learned models are developed using kSINDYc, N3ARX and NL2SQ identification techniques. The nominal operating data of all the simulation examples discussed in this section can be referred from the relevant literatures cited inside the article. In order to acquire an accurate estimate of the learned models, two test signals namely the step (u_{nom}) and PRBS (u_{prbs}) signals are input excited on all the examples. Consider the set of first principle input output data Z^N as

$$Z^N = [u(1), y(1), u(2), y(2) \ldots u(N), y(N)] \tag{16}$$

u and y corresponds to the excitation signal and the response of the process, t_f denotes the final time for the N^{th} sample.

$$Z^{N_{train}} = [u(1), y(1), u(2), y(2) \ldots u(N^{train}), y(N^{train})] = 0.7Z^N \tag{17}$$

$$Z^{N_{valid}} = [u(N^{train+1}), y(N^{train+1}), u(N^{train+2}), y(N^{train+2}) \ldots u(N), y(N)] = 0.3Z^N \tag{18}$$

where $N = N^{train} + N^{valid}$; (19)

N^{train} and N^{valid} conform to the training and validation samples set. All the case studies are simulated with a sampling time of $T_s = 0.01s$ for N number of sample space. $Z^{N_{train}}$ intends the input-output data taken for training kSINDYc, N3ARX and NL2SQ algorithms and remaining $Z^{N_{valid}}$ corresponds to validation dataset .The following section exemplifies five industrial processes with divergent nonlinear dynamics and their time response to nonlinear system identification methods at u_{nom} and u_{prbs} .

Example 1: Three Tank Process

A three-tank hydraulic process with the configuration of first pump supplying a liquid to first tank is considered in the present work. The objective is to control liquid levels in each tank by measuring the level of 3rd tank. The dynamic equations and the associated process parameters c_{12}, c_{23}, c_3, A_1 h_1, h_2 *and* h_3

$q_1\ m^3\ s^{-1}$. The nonlinear differential equations of the three-tank system are given by Eq. (20) to (22). The initial level of the all the tanks is assumed to be zero. q_1 denotes the inflow rate of the liquid in the first tank with the constraint $u = q_1\ m^3\ s^{-1}$ $q_1 \in [0 - 1e^{-5}]m^3 s^{-1}$ and $u_{nom} = 0.5e^{-5}\ m^3\ s^{-1}$.

$$\dot{h}_1 = \frac{q_1}{A_1} - c_{12}\sqrt{h_1 - h_2} \tag{20}$$

$$\dot{h}_2 = c_{12}\sqrt{h_1 - h_2} - c_{23}\sqrt{h_2 - h_3} \tag{21}$$

$$\dot{h}_3 = c_{23}\sqrt{h_2 - h_3} - c_3\sqrt{h_3} \tag{22}$$

$$y = h_3 \tag{23}$$

The step response for the tank level $y_{act}(t)$ from Eq. (23) is compared with that $y_{pred}(t)$ obtained from the learned models identified by kSINDYc, N3ARX and NL2SQ where the training dataset Z^{Ntrain} is presented in Fig.2. The three-tank process has a large settling time of around 5000 sec with a weak nonlinear behavior.

[Insert Fig. 2 here]

[Insert Fig. 3 here]

A $\pm 10\%$ variation in feed flow rate from u_{nom} also termed as u_{prbs} is adopted (through PRBS mode) to check the open loop response of tank level h_3 in Fig. 3. It has been observed that NL2SQ identification approach outperforms kSINDYc in the case of a sluggish nonlinear system like three-tank process.

Example 2: CSTR

An exothermal, continuous stirred tank reactor (CSTR) is widely used to convert reactants to products $(A \rightarrow P)$

$q_c = u_{nom}$ is considered as manipulated input and temperature of the reactor T is the output variable. The states are concentration of reactants C_a and temperature of reactor T. The nominal operating data for the reaction is available in [38]. The initial states and steady state points of the Concentration gradient C_a of the species A and the effluent temperature of the reactor T are assumed to be the same where $(Ca_{nom}, T_{nom}) = (0.08235\, mol\, l^{-1}, 441.81\, K)$. The open loop study obtained for a nominal input $u_{nom} = 102\ l\ min^{-1}$ can be viewed from Fig. 4.

$$\dot{C}_a = \frac{q}{V}\left(C_{af} - C_a\right) - k_0 C_a \exp\left(-\frac{E}{RT}\right) \tag{24}$$

$$\dot{T} = \frac{q}{V}\left(T_f - T\right) + \frac{(-\Delta H) k_0 C_a}{\rho C_p} \exp\left(-\frac{E}{RT}\right) + \frac{\rho_c C_{pc}}{\rho C_p V} q_c \left[1 - \exp\left(-\frac{hA}{q_c \rho_c C_{pc}}\right)\right](T_c - T)(6) \tag{25}$$

$$y = T \tag{26}$$

By carefully observing Eqs. (24) and (25), we can clearly understand that the activation energy level E has an effect on rate-constant of reaction which further influences the outputs of the CSTR, and depends upon the operating conditions and mechanism of species A undergoing the reaction $(A \rightarrow P)$. Therefore, the key term for the kSINDYc identification method in an exothermal CSTR appears in $\exp\left(-\frac{E}{RT}\right)$.

[Insert Fig. 4 here]

The open loop (temperature) responses, with the jacketed-coolant flowrate at $u_{nom} = 102\ l\ min^{-1}$ of the CSTR studied in Fig.4, ensure that both kSINDYc and NL2SQ expedite the process dynamics more accurately. Moreover, the N3ARX method shows notable deviations in predicted temperature ($y_{pred}(t)$) from the true reactor temperature $y_{act}(t)$ where accuracy falls down with a value of RMSE=3.2382.

[Insert Fig. 5 here]

The input PRBS region, $u_{prbs} \in [90, 110] l\, min^{-1}$ is near the vicinity of the steady state point u_{nom}. The results are obtained with respect to state variable temperature, $T(K)$. The open loop simulations for $\pm 10\%$ PRBS type of changes on the coolant flow rate at u_{nom} are presented in Fig. 5. These graphs, showing outlet temperatures, in Fig. 4 and Fig. 5 prove that the dynamic characteristics of CSTR undergo wide variations when it is operated at input regions u_{nom} and u_{prbs}. The learned dynamics (outputs from sparse space) from kSINDYc identification method outperforms the other two methods for a nonlinear CSTR process.

Example 3: Heat Exchanger

A Heat Exchanger (HE) is a device where a cold fluid is heated by another hot stream mostly by convection principle. Recently, first principle modelling of a heat exchanger for a high temperature milk pasteurization unit was enumerated using log mean temperature difference approach [39]. In our research, a nonlinear physical model of a fluid-fluid HE process is detailed in this section, adopted from [40]. Here, we consider the outlet temperature of the process fluid T_{po} as the controlled variable and flow rate F_c of the heating fluid as the manipulated variable. The operating conditions and parameter of the heat exchanger are acquired from [40]. The steady state values and the initial states (T_{co_nom}, T_{po_nom}) are found to be $(115,150)^\circ F$. The nonlinear material balance equations of the process are given in Eqs. (27-28).

$$\dot{T}_{co} = \frac{2}{M_c}\left[F_c\left(T_{ci} - T_{co}\right) + \left(UA\Delta T_{lm} / C_{pc}\right)\right] \tag{27}$$

$$\dot{T}_{po} = \frac{2}{M_p}\left[F_p\left(T_{pi} - T_{po}\right) - \left(UA\Delta T_{lm} / C_{pp}\right)\right] \tag{28}$$

$$\Delta T_{lm} = \frac{\left(T_{po} - T_{ci}\right) - \left(T_{pi} - T_{co}\right)}{\log\left(T_{po} - T_{ci}\right) - \log\left(T_{pi} - T_{co}\right)} \tag{29}$$

ΔT_{lm} is the logarithmic mean temperature difference of the process

$$y = T_{po} \tag{30}$$

[Insert Fig. 6 here]

The open-loop temperature from nonlinear physical process, $y_{act}(t)$ namely, outlet fluid temperature T_{po} depicted in Fig. 6, reveals that HE model is highly nonlinear where T_{po} values drop drastically from $150^\circ F$ to a new steady state at $T_{po} = 45.11^\circ F$. The predicted response of the fluid temperature $y_{pred}(t)$ of kSINDYc method reached the steady state at $45.02^\circ F$ and surpassed other identification schemes for a highly nonlinear heat exchanger at $u_{nom} = 40\ lbm\ min^{-1}$.

[Insert Fig. 7 here]

The output variable T_{po} is plotted for all the three methods kSINDYc, NL2SQ and N3ARX in Fig.7 when the excitation signal is $u_{prbs} \in [36, 44]\ lbm\, min^{-1}$. The response of kSINDYc and NL2SQ are relatively closer and better than N3ARX identification.

Example 4: Bio reactor

A bioreactor otherwise called a fermenter, a special type of heterogeneous reactor, is an essential automated system used in food processing and pharmaceutical industries. A fed-batch reactor with the manipulated input of dilution rate D and the process output, biomass concentration X is adopted from [41]. The mass balance equations representing the kinetic model of the bioreactor are given in Eqs (31-33). At high substrate concentration, S, rate of product formation is independent of S due to limited amount of enzyme; at low substrate concentration, the rate of product formation becomes proportional to S and follows first-order kinetics. Fermenters generally produce heat respiration and maintenance of bio-chemical pathways by microbes. Control becomes essential in large scale installations. However lack in proper knowledge behind kinetic pathways, calculation of cooling, aeration, pH, and agitations need attention. Here growth rate (μ_{max}), yield factor (Y_{XS}), rate constant for conversion of substrate to product (K_m) and rate of inhibition (K_1) are the vital process parameters. The manipulated input D occupies the region $[0, 0.6] hr^{-1}$. The initial values and nominal (operating) points of the state variables are $(X_{nom}, S_{nom}) = (1.530, 0.174)\, g\, l^{-1}$. The density of microbial cells also called biomass concentration X of any microorganism grows by consuming the substrate S fed to it.

$$\dot{S} = \frac{-1}{Y_{XS}} \mu(S) X + D(S_{in} - S) \tag{31}$$

$$\dot{X} = (\mu(S) - D) X \tag{32}$$

where a Haldane type of specific growth rate is given by

$$\mu(S) = \mu_{\max} \frac{S}{\left(S + K_m + K_1 S^2\right)} \tag{33}$$

$$y = X \tag{34}$$

The nonlinearity of the bioreactor varies w.r.t the specific growth rate $\mu(S)$, the type of excitation given, initial states of S, X and the operating region of dilution rate u. Therefore a bioreactor can be contemplated as a very sensitive nonlinear system, subjected to the above factors.

[Insert Fig. 8 here]

Fig. 8 represents the true response $y_{act}(t)$ D is operated at $u_{nom} = 0.27\, hr^{-1}$

u_{nom}

response of biomass concentration $y_{act}(t) = 1.547\, g\, l^{-1}$ to reach $y_{pred}(t) = 1.55\, g\, l^{-1}$. On the contrary, kSINDYc responds well and captures the true nonlinear dynamics $y_{act}(t)$ of the bioreactor, when operated in the safe region at $u_{nom} = 0.27\, hr^{-1}$.

[Insert Fig. 9 here]

The response of $X(g/litre)$ due to PRBS input in the dilution rate of the feed flow $u_{prbs} = \pm 10\%\ u_{nom}$ is portrayed in Fig. 9. It can be noticed that the three methods of identification showed excellent tracking of the biomass-concentration (nonlinear dynamics) with very sharp variations.

Example 5: Distillation Column

A 9 stage ($n_s = 9$) binary Distillation Column (DC), to separate methanol-water mixture, operated in the LV (liquid-vapour) configuration with the manipulated variable as reflux rate to the column $u_{nom} = L\, kmol\, min^{-1}$ is taken for the study from [41]. The distillate composition x_D **(***kmol***)** which is the top most product is the output variable y. The feed mixture containing 50% Methanol has to be rectified continuously to 98% purity. The common problems are vapor cross-flow channeling, foaming and unaccounted interactions. The presence of many state-variables and process parameters make the simulation of DC model more complex. Accordingly, certain process assumptions are made as follows for an easier analysis: A perfect binary mixture with constant pressure, no vapor holdup, and constant relative volatility, on all stages are considered. The ordinary differential equations governing the DC are Eqs. (35-39).

$$\dot{x}_1 = \frac{1}{M_D}\left[V_R\left(y_2 - y_1\right)\right] \tag{35}$$

$$\dot{x}_i = \frac{1}{M_T}\left[L_R x_{i-1} + V_R y_{i+1} - L_R x_i - V_R y_i\right] \tag{36}$$

$$\dot{x}_{NF} = \frac{1}{M_T}\left[L_R x_{NF-1} + V_R y_{NF+1} + F z_F - L_S x_{NF} - V_R y_{NF}\right] \tag{37}$$

$$\dot{x}_i = \frac{1}{M_T}\left[L_S x_{i-1} + V_S y_{i+1} - L_S x_i - V_S y_i\right] \tag{38}$$

$$\dot{x}_{NS} = \frac{1}{M_B}\left[L_S x_{NS-1} - B x_{NS} - V_S y_{NS}\right] \tag{39}$$

$$y = x_D \tag{40}$$

[Insert Fig. 10 here]

The DC model when operated in a wider operating region instead of a fixed input at u_{nom} imparts a massive nonlinear phenomenon and does not suit the normal operation. Therefore, the initial conditions of distillate x_D and bottoms composition x_B are carefully chosen to be $(0.005, 1.05)$. The nominal operating values so obtained are $(x_{Dnom}, x_{Bnom}) = (0.775, 0.2225)$. The time response of x_D in Fig. 10 is noticed when the DC is operated at reflux rate $u_{nom} = 2.706\, kmol\, min^{-1}$. Except for N3ARX method, the learned models $y_{pred}(t)$ from kSINDYc and NL2SQ approaches bear a very close follow up to the true output of the distillate composition $y_{act}(t) = 0.775\,(kmol)$.

[Insert Table 1 here]

Table 1, summarizes the major key terms selected in the candidate library function of each process. The nonlinear function of each process is decided by the key term k_{nl}. The candidate library Θ has relatively a fewer functional terms. k_{nl} term is chosen carefully by checking the influencing terms from the dynamic equations of every process. Moreover, introducing the key nonlinear terms in the candidate library function of kSINDYc is intended to build models of dynamic physical systems with diverse nonlinear behavior.

[Insert Fig. 11 here]

Fig. 11 is an important graphical representation that relates the nonlinearity of each process with the three nonlinear system identification methods in terms of the objective function. The criteria to be accounted for is the trade-off between Goodness of fit and the complexity of the learned model $y_{pred}(t)$.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_{act}(t) - y_{pred}(t))^2} \tag{41}$$

The RMSE from Eq. (41) is found for all the processes described in the previous sections for N samples. Table 2 will provide the inference made on the choice of the appropriate system identification method, on the basis of nonlinearity level and the RMSE of every process. As seen from Fig.11, the crucial factor in determining the choice of system identification rests on the minimum RMSE between the actual and learned dynamics of the identified model among all the three identified models.

[Insert Table 2 here]

By looking at Table 2, a major conclusion can be brought over in the concept of data driven modelling. The usage of NL2SQ method is preferred only for mild systems like three tank process. Diversely, kSINDYc is the best opted non parametric model for highly nonlinear processes like

HE. To organize all the nonlinear dynamic systems into a wider category based on the nonlinearity level, significant variables which include the excitation signal, the initial values of state variables and the operating region for each process are deliberately chosen such that they fall under one class of nonlinearity as shown in Eq. (42).

$$Nonlinearity\ level = \begin{cases} \Delta_{\circ} \le 0.3, mild\ nonlinear \\ 0.3 > \Delta_{\circ} \ge 0.7, medium\ nonlinear \\ \Delta_{\circ} > 0.7, highly\ nonlinear \end{cases} \tag{42}$$

For all the case studies discussed in this section, a qualitative comparison is made between the RMSE of all the nonlinear system identification methods and Δ_0 in Fig.11. The identification method which gives least RMSE under each class of Δ_0 imply a better estimate on all basis and is exclusively picked up for the perfect choice of system identification. The physical quantities of each process addressed in Table 3, have different orders of magnitude. To sustain a uniform scale in measuring the nonlinearity, the time period t, input u and output variable $y_{act}(t)$ of all nonlinear physical process are normalized between 0 and 1, and thereafter the CANM method Δ_0 is intended from Eq.(2)

[Insert Table 3 here]

Table 3, affirms the better choice for system identification of medium and high nonlinear process to be kSINDYc approach whereas the mild nonlinear systems can follow NL2SQ method to learn the process dynamics. The effectiveness of N3ARX approach trails behind kSINDYc and NL2SQ in terms of RMSE and execution time. Table 3 also presents assertive conclusive remarks on the selection of identification method based on strength of nonlinearity Δ_0.

Remark 1: N3ARX method requires large number of training data set to provide accurate solution. The learned dynamics using this approach did not meet the acceptable limit at u_{nom} .Consequently, it can be used for systems with broader range of excitation signals u_{prbs} and u_{chirp} (CHIRP excitation). The major setback of N3ARX method compared to other methods is its long computation time in MATLAB.

Remark 2: The learned dynamics using NL2SQ is satisfactory for mild and medium nonlinear systems. As it is a parametric identification scheme it requires a healthy knowledge of the process parameters and the nominal operating regions. This method fails to identify the process model whose measured data is within the low data limit.

Remark 3: kSINDYc is computationally attractive, requires less data, assumes a few number of candidate terms in Θ to make an interpretable efficient model at u_{nom} and u_{prbs} . The method outstrips NL2SQ and N3ARX by accurately following the plant dynamics of medium and highly nonlinear systems.

5 Parametric Sensitivity Analysis (PSA):

Parametric variations occurring in a unit in an industrial process is a decisive topic to be separately discussed. The uncertainities and the parametric fluctuations will definitely affect the dynamic stability of the process which in turn will be a matter of concern for nonlinear identification of dynamic systems. Local and global sensitivity analysis were carried out in semi-batch reactors to find out the intensity of interactions between the input-output parameters and to understand the variations in output variables for infinitesimal disturbances in input parameters [42]. This section gives a precise view of global parametric sensitivity analysis and the parametric influence on output variables. The model variance V_i is used to compute the sensitivity index Si . The effect of each individual parameter on the output can be measured from Si [42]. The parametric sensitivity of each process is evaluated considering uniform sampling for $N_P = 500$ samples. The simulation study in Section 4 confirms that the CSTR and Bioreactor processes are immensely sensitive to nonlinearity and excitation signal u . Therefore the metric Δ_0 changes with respect to operating point P , inputs u_{nom} , u_{prbs} and Sensitivity index Si .

[Insert Fig. 12 here]

Fig.12 shadows the sensitivity of the bioreactor with respect to variations in each parameter for around 10% variations. The yield coefficient Y_{XS} gives a maximum sensitivity index Si as seen from Fig.12 (a) for a particular time instant $t = 10$ *hours* .A scatter plot was obtained between the most influencing parameter Y_{XS} and the output variable X in Fig. 12(b).A gradual increase in Y_{XS} from 0.38 to 0.42,resulted in increase of biomass concentration from $X = 1.45$ *to* $1.64\, mol/l$. The sensitivity index of Y_{XS} plotted for a time period $t = 0$ *to* $10\, hrs$ confirms minimum Si occurs at $t = 0.2\, hours$ and maximum Si at $t = 10\, hrs$.

[Insert Fig. 13 here]

Referring from Eqs. (24, 25) in the exothermal CSTR, the variations in input parameter E/R has a significant effect on the output as viewed from Fig.13 (a). For a parametric variation $8000 < E/R \le 10000$, the reactor temperature $T(K)$, swirls around 460 K in Fig. 13(b). However, for $E/R \ge 10000$, the temperature makes a sudden fall to 350 K .Also in Fig.13(c), we can see a nonlinear relationship between the sensitivity index E/R_{Si} and the output variable $T(K)$. It is evident from Figs. 12 and Fig.13, that the PSA has a major implication on identifying the dynamics of nonlinear systems, which in turn will affect the identification of the nonlinear systems.

6 Conclusion:

In the present work, we have devised kSINDYc, N3ARX and NL2SQ system identification methods which are popular for identification of many nonlinear dynamic physical systems. Input-

$y_{pred}(t)$ from the three identification methods kSINDYc, N3ARX and NL2SQ are validated with the true response $y_{act}(t)$ through RMSE for all the case studies. In particular, the kSINDYc identification method is formulated by tuning the sparcification knob λ set in the ratio between 0 and 10. Additionally, the CANM Δ_0 targets to find out the degree of nonlinearity of those nonlinear examples. A fresh quantitative analysis is exemplified to cohere the nonlinear metric Δ_0 with the above said identification methods. Finally, a global parametric sensitivity analysis was investigated on CSTR and Bioreactor to perceive the changes in the nonlinear dynamics caused by the varying the sensitive process parameters. The presented adaptation of these advisable system identification methods is therefore considered important for all users who are interested in finding an interpretable, identification method for complex and diverse nonlinear systems.

Scope:

This article exemplifies a fresh quantitative analysis that correlates the nonlinear metric Δ_0 with the most admired non-parametric and parametric system identification schemes. Initially, the nonlinear dynamic processes are assorted in a single framework, as mild, medium and highly nonlinear using CNLM method. However, if the process is unstable, the index Δ_0 fails to measure the nonlinearity, due to the inadequacy in finding steady state operating point. Consequently, the measure may be deficient, when the process is operated in a region far beyond the operating point. Such deprivation issues have to be addressed in the sequel while measuring nonlinearity. Conclusively, this research study will be definitely instrumental for the researchers and academicians of nonlinear dynamics community, but needs to be further tested in real-world physical systems.

Acknowledgement: The first author acknowledges the financial support 'Anna Centenary Research Fellowship' (ACRF) of Anna University, Chennai to carry-out this study. Authors also state that there is no conflict of interests in this publication.

It is declared that data, code and material will be available on request.

The research does not include any study on human/ animal.

All the authors have consented for publication.

References

1. Schoukens, J., Ljung, L.: Nonlinear system identification: A user-oriented road map. IEEE Control Syst. Mag. 39(6), 28-99 (2019). https://doi.org/10.1109/MCS.2019.2938121
2. Xavier, J., Patnaik, S.K., Panda, R.C.: Process Modeling, Identification Methods, and Control Schemes for Nonlinear Physical Systems–A Comprehensive Review. ChemBioEng Rev. 8(4), 1-21, (2021). https://doi.org/10.1002/cben.202000017

3. .Sadeqi, A., Moradi, S., Shirazi, K.H.: Nonlinear subspace system identification based on output-only measurements. J. Franklin Inst. 357(17), 12904-12937 (2020). https://doi.org/10.1016/j.jfranklin.2020.08.008
4. Xu, H., Ding, F., Yang, E.: Modeling a nonlinear process using the exponential autoregressive time series model. Nonlinear Dyn. 95(3), 2079-2092 (2019) https://doi.org/10.1007/s11071-018-4677-0
5. Xu, L.: The parameter estimation algorithms based on the dynamical response measurement data. Adv. Mech. Eng. 9(11), 1687814017730003 (2017). https://doi.org/10.1177/1687814017730003
6. Xiong, W., Yang, X., Huang, B., Xu, B.: Multiple-model based linear parameter varying time-delay system identification with missing output data using an expectation-maximization algorithm. Ind. Eng. Chem. Res. 53(27), pp.11074-11083 (2014). https://doi.org/10.1021/ie500175r
7. Zhang, X., Ding, F., Alsaadi, F.E. and Hayat, T., 2017. Recursive parameter identification of the dynamical models for bilinear state space systems. Nonlinear Dyn. 89(4), 2415-2429. https://doi.org/10.1007/s11071-017-3594-y
8. Zhang, T., Lu, Z.R., Liu, J.K., Liu, G.: Parameter identification of nonlinear systems with time-delay from time-domain data. Nonlinear Dyn. 1-17 (2021).
9. Mani, A.K., Narayanan, M.D. and Sen, M., 2018. Parametric identification of fractional-order nonlinear systems. Nonlinear Dyn. 93(2), 945-960. https://doi.org/10.1007/s11071-018-4238-6
10. Kazemi, M., Arefi, M.M.: A fast iterative recursive least squares algorithm for Wiener model identification of highly nonlinear systems. ISA Trans. 67, 382-388 (2017). https://doi.org/10.1016/j.isatra.2016.12.002
11. Yu, F., Mao, Z., Jia, M.: Recursive identification for Hammerstein–Wiener systems with dead-zone input nonlinearity. J. Process Control. 23(8), 1108-1115 (2013). https://doi.org/10.1016/j.jprocont.2013.06.014
12. Mauroy, A., Goncalves, J.: Koopman-based lifting techniques for nonlinear systems identification. IEEE Trans. Autom. Control. 65(6), 2550-2565(2019). https://doi.org/10.1109/TAC.2019.2941433
13. Pillonetto, G., Dinuzzo, F., Chen, T., De Nicolao, G., Ljung, L.: Kernel methods in system identification, machine learning and function estimation: A survey. Automatica. 50(3), 657-682 (2014). https://doi.org/10.1016/j.automatica.2014.01.001
14. Quaranta, G., Lacarbonara, W., Masri, S.F.: A review on computational intelligence for identification of nonlinear dynamical systems. Nonlinear Dyn. 99(2), 1709-1761 (2020). https://doi.org/10.1007/s11071-019-05430-7
15. Bolourchi, A., Masri, S.F., Aldraihem, O.J.: Development and application of computational intelligence approaches for the identification of complex nonlinear systems. Nonlinear Dyn. 79(2), 765-786 (2015). https://doi.org/10.1007/s11071-014-1702-9
16. Vafamand, N., Arefi, M.M., Khayatian, A.: Nonlinear system identification based on Takagi-Sugeno fuzzy modeling and unscented Kalman filter. ISA Trans. 74, 134-143 (2018). https://doi.org/10.1016/j.isatra.2018.02.005
17. Kaiser, E., Kutz, J.N., Brunton, S.L.: Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. Proc. R. Soc. London, Ser. A, 474(2219), 20180335 (2018). https://doi.org/10.1098/rspa.2018.0335

18. Du, J. and Johansen, T.A.: Integrated multimodel control of nonlinear systems based on gap metric and stability margin. Ind. Eng. Chem. Res. 53(24), 10206-10215 (2014). https://doi.org/10.1021/ie500035p
19. Hahn, J. and Edgar, T.F.: A gramian based approach to nonlinearity quantification and model classification. Ind. Eng. Chem. Res. 40(24), 5724-5731 (2001). https://doi.org/10.1021/ie010155v
20. Alanqar, A., Durand, H., Christofides, P.D.: On identification of well-conditioned nonlinear systems: Application to economic model predictive control of nonlinear processes. AIChE J. 61(10), 3353-3373 (2015). https://doi.org/10.1002/aic.14942
21. Xavier, J., Patnaik, S.K., Panda, R.C.: Nonlinear Measure for Nonlinear Dynamic Processes Using Convergence Area: Typical Case Studies. J. Comput. Nonlinear Dyn. 16(5), p.051002 (2021). https://doi.org/10.1115/1.4050553
22. Brunton, S.L., Proctor, J.L., Kutz, J.N.: Sparse identification of nonlinear dynamics with control (SINDYc). IFAC-Papers OnLine, 49(18), 710-715 (2016). https://doi.org/10.1016/j.ifacol.2016.10.249
23. Mangan, N.M., Brunton, S.L., Proctor, J.L., Kutz, J.N.: Inferring biological networks by sparse identification of nonlinear dynamics. IEEE Trans. Mol. Biol. Multi-Scale Commun. 2(1), 52-63 (2016). https://doi.org/10.1109/TMBMC.2016.2633265
24. Tang, X., Dong, Y.: Expectation maximization based sparse identification of cyberphysical system. Int. J. Robust Nonlinear Control. 31(6), 2044-2060 (2021). https://doi.org/10.1002/rnc.5325
25. Champion, K., Zheng, P., Aravkin, A.Y., Brunton, S.L. and Kutz, J.N., 2020. A unified sparse optimization framework to learn parsimonious physics-informed models from data. IEEE Access. 8, 169259-169271. https://doi.org/10.1109/ACCESS.2020.3023625
26. Zhang, L., Schaeffer, H.: On the convergence of the SINDy algorithm. Multiscale Model. Simul. 17(3), 948-972 (2019). https://doi.org/10.1137/18M1189828
27. Lin, M., Cheng, C., Peng, Z., Dong, X., Qu, Y., Meng, G.: Nonlinear dynamical system identification using the sparse regression and separable least squares methods. J. Sound Vib. 505, 116141 (2021). https://doi.org/10.1016/j.jsv.2021.116141
28. Yin, Q., Zhou, L., Wang, X.: Parameter identification of hysteretic model of rubber-bearing based on sequential nonlinear least-square estimation. Earthquake Eng. Eng. Vibr. 9(3), 375-383(2010). https://doi.org/10.1007/s11803-010-0022-4
29. Transtrum, M.K. and Sethna, J.P., 2012. Improvements to the Levenberg-Marquardt algorithm for nonlinear least-squares minimization. arXiv preprint arXiv:1201.5885.
30. Kommenda, M., Burlacu, B., Kronberger, G., Affenzeller, M.: Parameter identification for symbolic regression using nonlinear least squares. Genet. Program. Evolv. Mach. 21(3), 471-501(2020). https://doi.org/10.1007/s10710-019-09371-3
31. Cheng, L., Liu, W., Hou, Z.G., Yu, J., Tan, M.: Neural-network-based nonlinear model predictive control for piezoelectric actuators. IEEE Trans. Ind. Electron. 62(12), 7717-7727 (2015). https://doi.org/10.1109/TIE.2015.2455026
32. Liu, H. and Song, X.: Nonlinear system identification based on NARX network. In 2015 10th Asian Control Conference (ASCC) 1-6. IEEE (May 2015). https://doi.org/10.1109/ASCC.2015.7244449
33. Han, Y., Ding, N., Geng, Z., Wang, Z., Chu, C.: An optimized long short-term memory network based fault diagnosis model for chemical processes. J. Process Control. 92, pp.161-168(2020). https://doi.org/10.1016/j.jprocont.2020.06.005

34. Subudhi, B., Jena, D.: A differential evolution based neural network approach to nonlinear system identification. Appl. Soft Comput. 11(1), 861-871 (2011). https://doi.org/10.1016/j.asoc.2010.01.006
35. Sahoo, H.K., Dash, P.K., Rath, N.P.: NARX model based nonlinear dynamic system identification using low complexity neural networks and robust H∞ filter. Appl. Soft Comput. 13(7), pp.3324-3334 (2013). https://doi.org/10.1016/j.asoc.2013.02.007
36. Saki, S., Fatehi, A.: Neural network identification in nonlinear model predictive control for frequent and infrequent operating points using nonlinearity measure. ISA Trans. 97, pp.216-229 (2020). https://doi.org/10.1016/j.isatra.2019.08.001
37. Bistak, P., Huba, M.: Three-tank laboratory for input saturation control based on matlab. IFAC- Papers online. 49(6), 207-212 (2016). https://doi.ord/10.1016/j.ifacol.2016.07.178
38. Pottman, M., Seborg, D.E.: Identification of non-linear processes using reciprocal multiquadric functions. J. Process Control. 2(4), 189-203 (1992). https://doi.org/10.1016/0959-1524(92)80008-L
39. Indumathy, M., Sobana, S., Panda, R.C.: Modelling of fouling in a plate heat exchanger with high temperature pasteurisation process. Appl. Therm. Eng. 189, 116674(2021). https://doi.org/10.1016/j.applthermaleng.2021.116674
40. Alvarez-Ramirez, J., Cervantes, I., Femat, R.: Robust controllers for a heat exchanger. Ind. Eng. Chem. Res. 36(2), 382-388(1997). https://doi.org/10.1109/PC.2015.7169947
41. Bequette, B.W.: Process control: modeling, design, and simulation. Prentice Hall Professional (2003).
42. Zang, N., Qian, X.M., Shu, C.M., Wu, D.: Parametric sensitivity analysis for thermal runaway in semi-batch reactors: Application to cyclohexanone peroxide reactions. J. Loss Prev. Process Ind. 70, 104436 (2021). https://doi.org/10.1016/j.jlp.2021.104436

Table 1: Key term based SINDYc (kSINDYc):

Process	Key nonlinear term 'k_{nl}'	Candidate Library
Three Tank	$\sqrt{h_1-h_2}$, $\sqrt{h_2-h_3}$, $\sqrt{h_3}$	$\left[k_{nl} \;\; u \;\; h_1 \;\; h_2 \;\; h_3 \;\; h_1h_2 \;\; h_2h_3 \;\; h_3h_1 \;\; h_1u \;\; h_2u \;\; h_3u\right]$
CSTR	$e^{\left(\frac{-E}{RT}\right)}$	$\left[k_{nl} \;\; u \;\; C_a \;\; T \;\; C_a^{\,2} \;\; T^2 \;\; C_aT \;\; C_au \;\; Tu\right]$
HE	$\log\left(\frac{T_{po}-T_{ci}}{T_{pi}-T_{co}}\right)$	$\left[k_{nl} \;\; u \;\; T_{co} \;\; T_{po} \;\; T_{co}^{\,2} \;\; T_{po}^{\,2} \;\; T_{co}T_{po} \;\; T_{co}u \;\; T_{po}u\right]$
Bioreactor	$\frac{\mu_{max}}{K_1S^2}$	$\left[k_{nl} \;\; u \;\; X \;\; S \;\; X^2 \;\; S^2 \;\; SX \;\; Su \;\; Xu\right]$
DC	$\frac{\alpha x}{1+(\alpha-1)x}$	$\left[k_{nl} \;\; u \;\; x_{i=1\ldots n_s} \;\; x^2_{i=1\ldots n_s} \;\; x_{i=1..(n_s-1)}x_{j=1..n_s} \;\; x_{i=1\ldots n_s}u\right]$

Table 2: Comparison of RMSE for System identification using kSINDYc, N3ARX and NL2SQ methods. Note that the RMSE value is found for different excitation signals u_{nom} and u_{prbs}

Process	RMSE for u_{nom}			RMSE for u_{prbs}		
	kSINDYc	**N3ARX**	**NL2SQ**	**kSINDYc**	**N3ARX**	**NL2SQ**
Three Tank	0.005	0.0569	1.632e-6	0.0282	0.0043	0.007
CSTR	1.2954e-6	0.0016	2.493e-4	0.01	3.2382	0.0521
HE	4.311e-4	0.2282	8.6746	0.0322	0.3154	0.0403
Bioreactor	1.098e-7	1.1912e-4	0.0033	4.8147e-5	1.15e-4	8.9196e-4
DC	7.4743e-12	0.1559	8.5535e-8	4.8689e-4	0.0054	2.6418e-5

Table 3: System Identification from Δ_0

Process	Δ_0 at u_{nom}	Class of NL	Best Choice for System Identification
CSTR	0.68	Medium	kSINDYc
HE	0.9328	High	kSINDYc
Three Tank	0.1123	Mild	NL2SQ
Bioreactor	0.0157	Mild	N3ARX
DC	0.2710	Mild	NL2SQ/kSINDYc

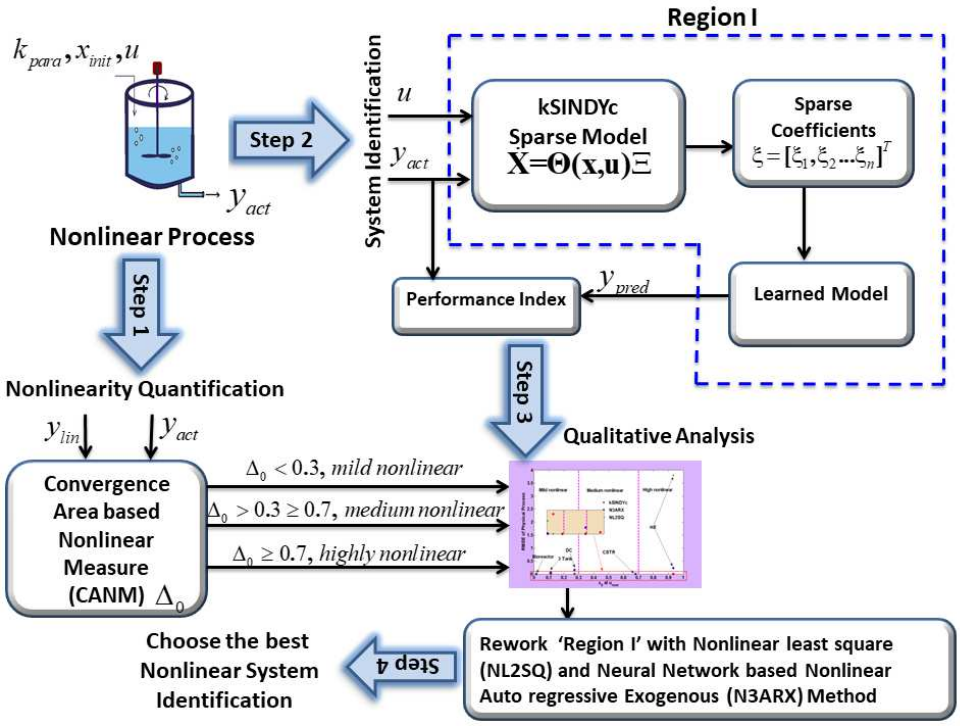

Fig.1 Schematic diagram of the CANM based System Identification method

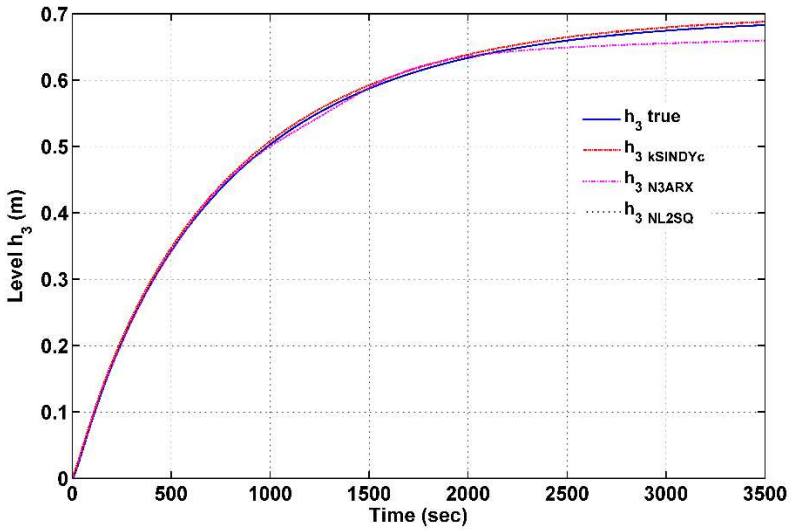

Fig. 2 True and learned model response h_3 of three tank process for step input at $u_{nom} = 0.5e^{-5}\ m^3\ s^{-1}$

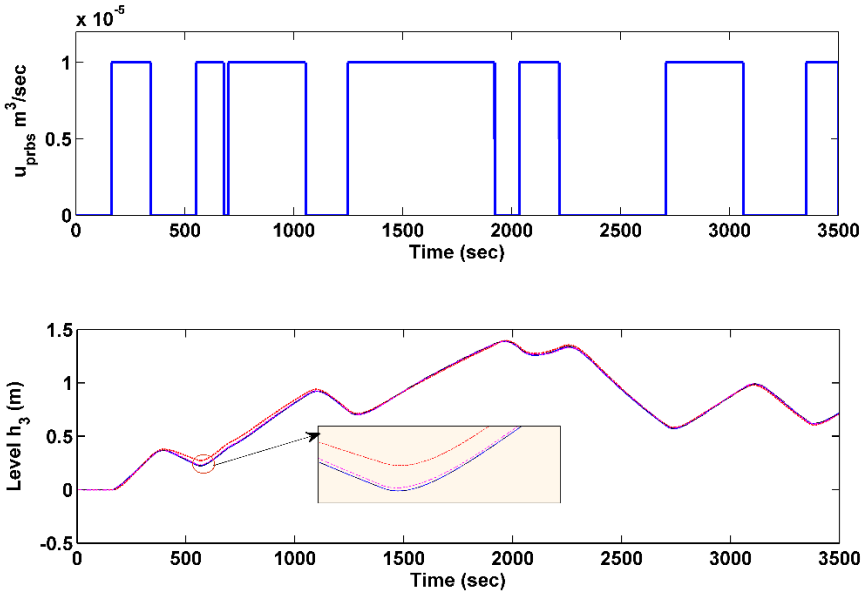

Fig. 3 True and learned model response of three-tank process for input u_{prbs} . Blue solid line: true process output; red solid line: learned model using kSINDYc; magenta dash dotted line: learned model using N3ARX; black dotted line: learned model using NL2SQ.

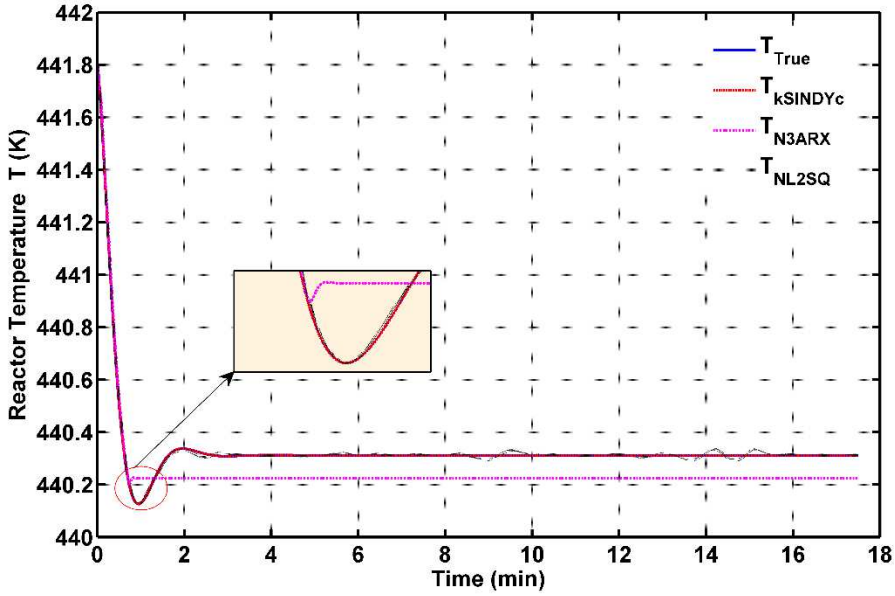

Fig. 4 True and learned model response of CSTR process for step input u_{nom}

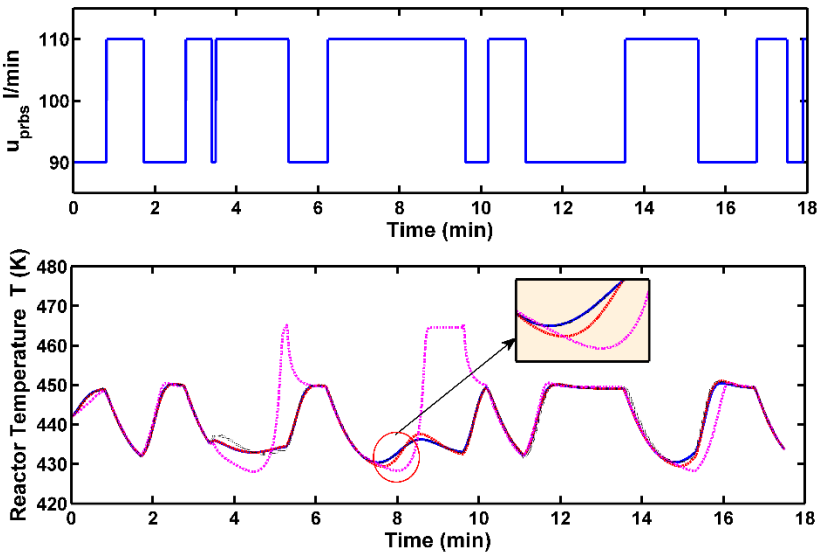

Fig. 5 True and learned model response of CSTR process for input u_{prbs} . Blue solid line: true process output; red solid line: learned model using kSINDYc; magenta dash dotted line: learned model using N3ARX ; black dotted line: learned model using NL2SQ.

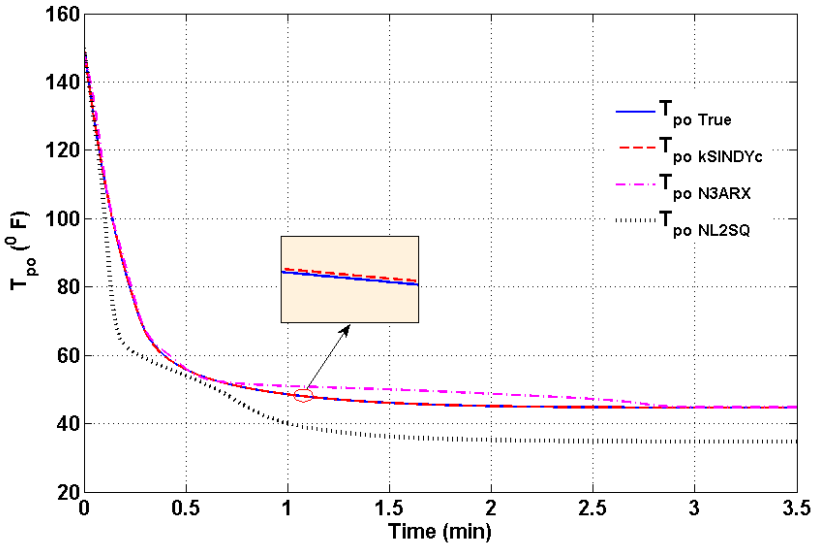

Fig. 6 True and learned model temperature responses T_{po} of heat exchanger process for step input at u_{nom}

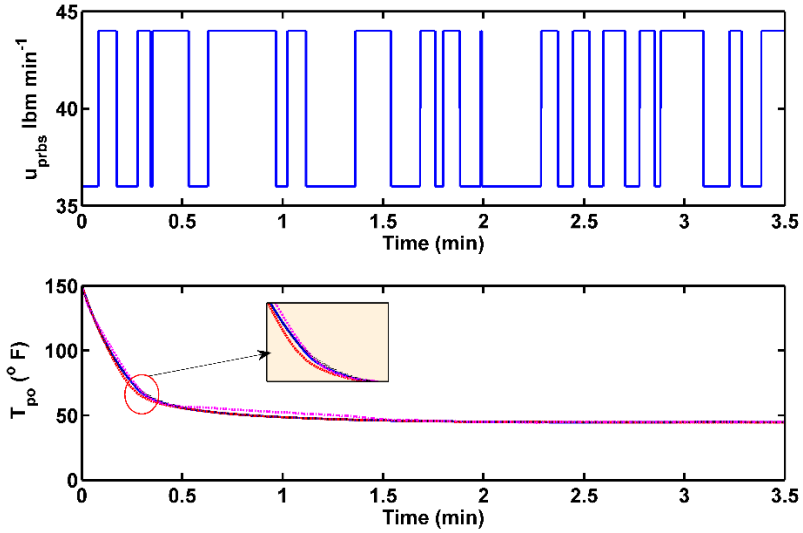

Fig. 7 True and learned model response of heat exchanger process for step input at u_{prbs} . Blue solid line: true process output; red solid line: learned model using kSINDYc; magenta dash dotted line: learned model using N3ARX ; black dotted line: learned model using NL2SQ.

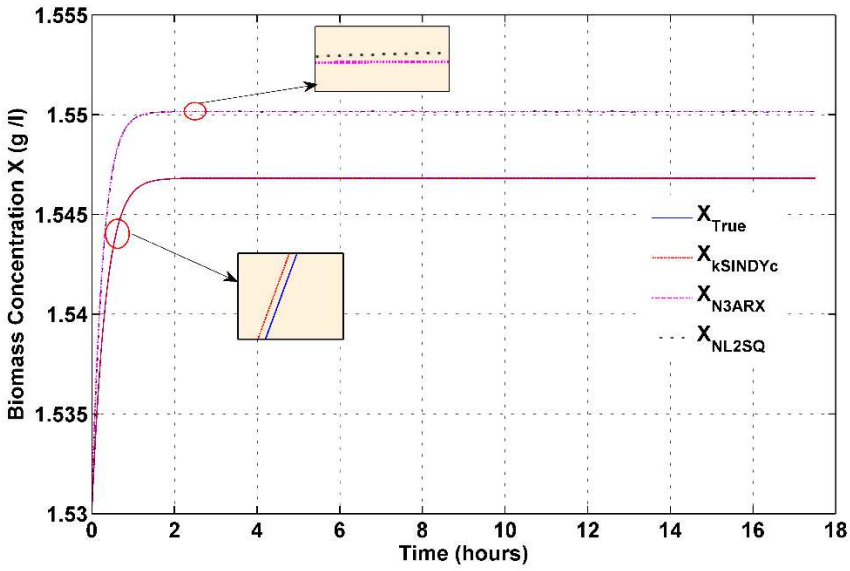

Fig. 8 True and learned model response of bioreactor process for input u_{nom}

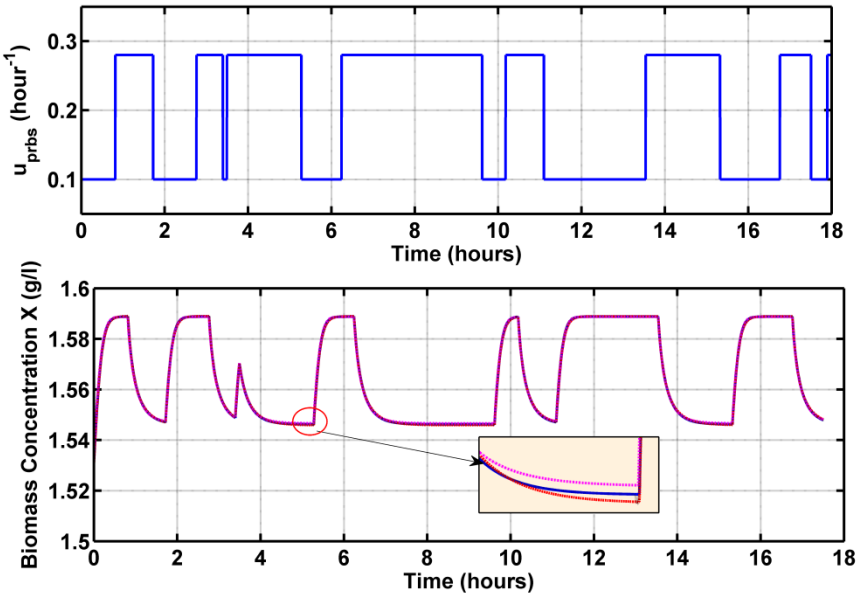

Fig. 9 True and learned model response $X\,(g/litre)$ of bioreactor process for input u_{prbs} . Blue solid line: true process output; red solid line: learned model using kSINDYc; magenta dash dotted line: learned model using N3ARX ; black dotted line: learned model using NL2SQ.

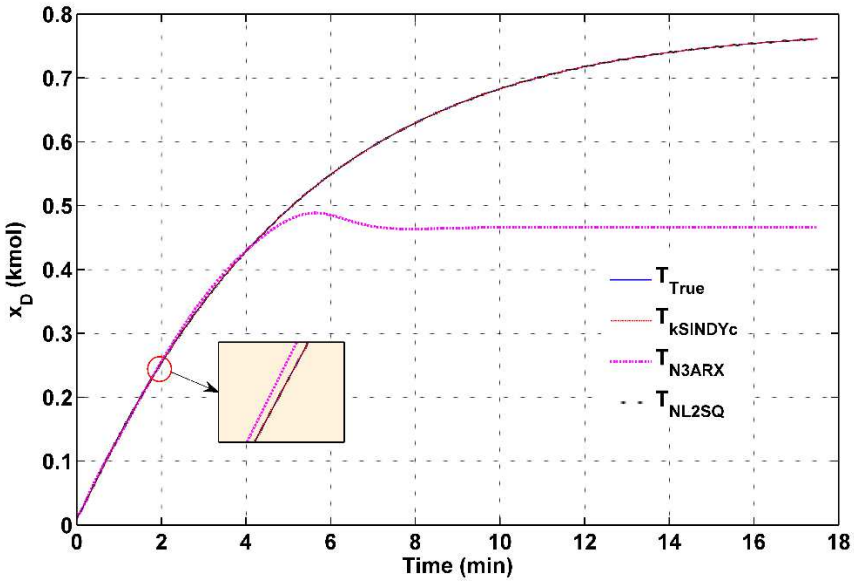

Fig. 10 True and learned model response x_D of distillation column process for step input at u_{nom}

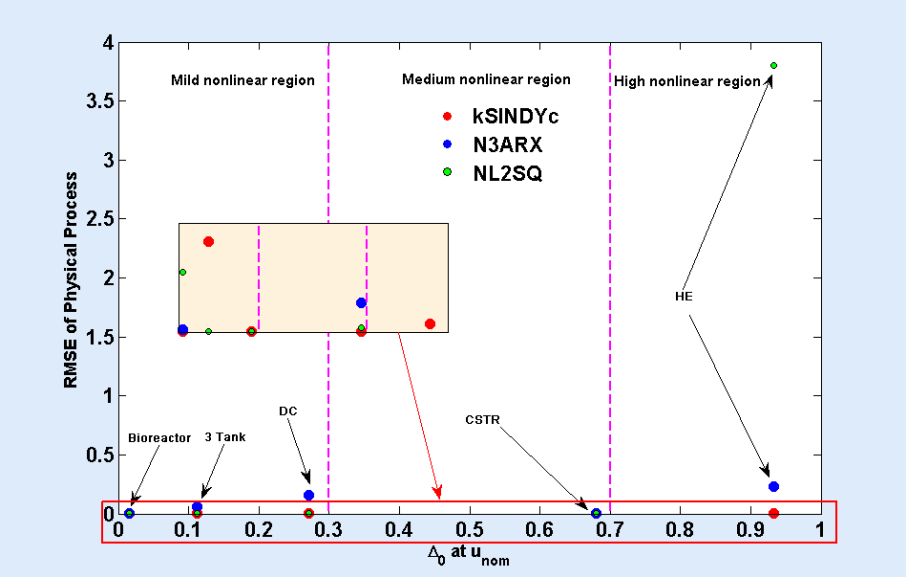

Fig. 11 RMSE of learned models of all process with its metric Δ_0 at input u_{nom}

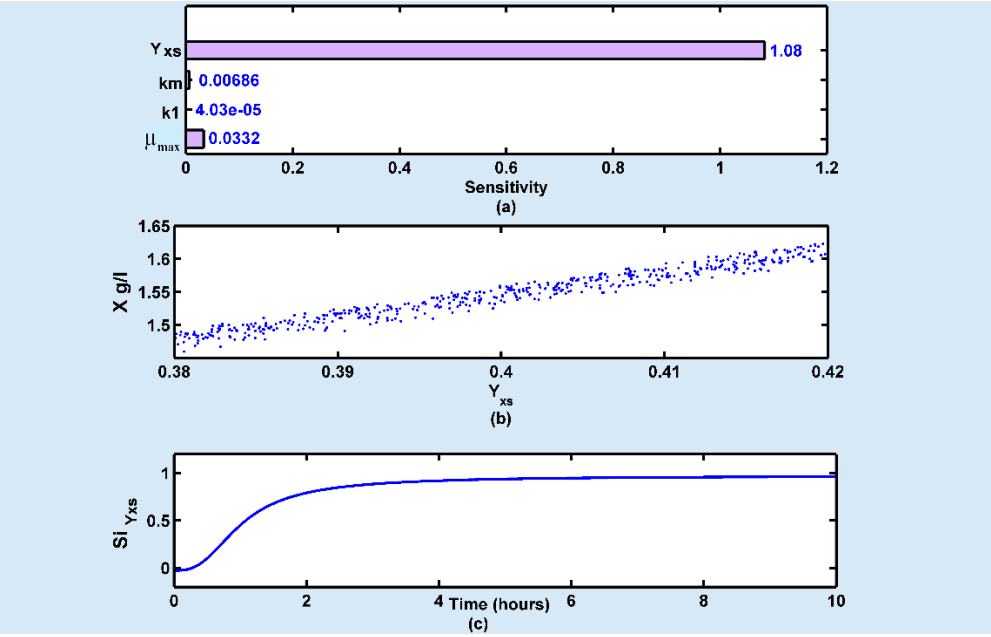

Fig. 12 (a) Influence of each parameter in bioreactor w.r.t senitivity (b) Scatter plot for ±10% variation in Y_{XS} and X. (c) Sensitivity index Si of Y_{XS} w.r.t time.

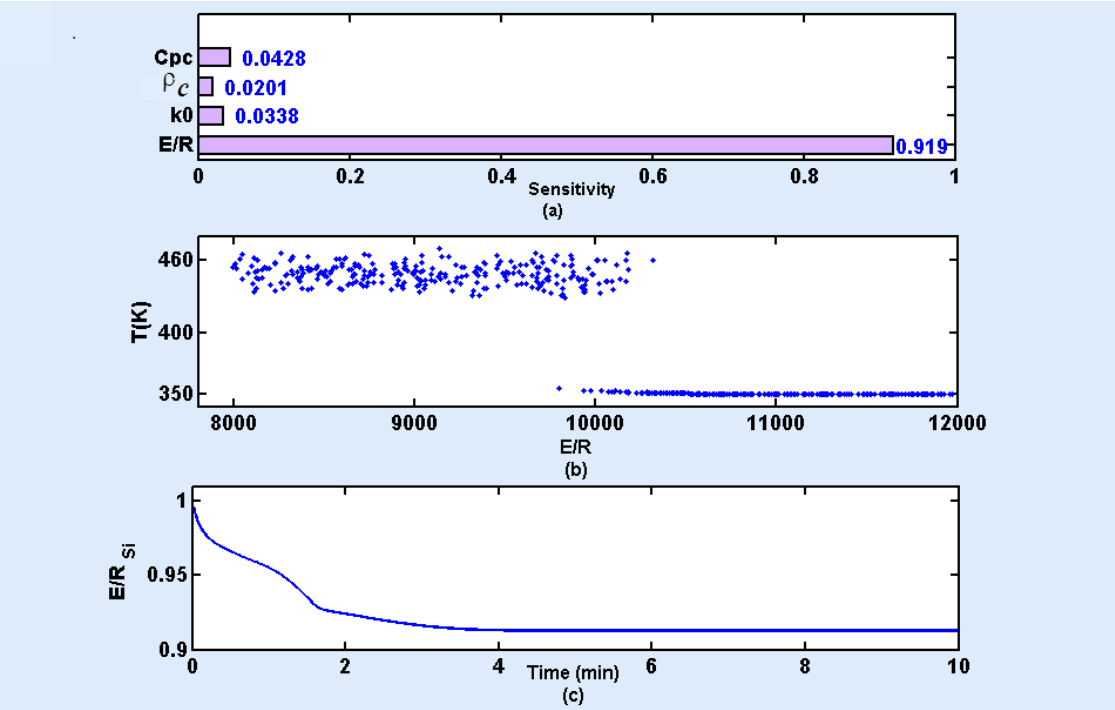

Fig. 13 (a) Influence of each parameter in CSTR w.r.t sensitivity (b). Scatter plot for ±10**%** variation in E/R and T. (c) Sensitivity index of E/R_{Si} w.r.t time