Nonlinear Dyn (2024) 112:6475–6501
https://doi.org/10.1007/s11071-023-09258-0

ORIGINAL PAPER

Nonlinear system identification in coherence with nonlinearity measure for dynamic physical systems—case studies

Joanofarc Xavier · S. K. Patnaik · Rames C. Panda

Received: 27 September 2021 / Accepted: 23 December 2023 / Published online: 5 February 2024

Abstract With the recent success in using time series data, many nonlinear identification tools have emerged to learn the nonlinear dynamics of unknown physical systems. However, if the nonlinearity level is very high or too small, in many cases, the identified model fails to precisely learn the actual dynamics of the system, which in turn makes the closed-loop control more challenging. Finding out a suitable system identification routine for identifying a given nonlinear system based on the nonlinearity level is still cryptic. In this article, we propose an integrated framework 'System identification in coherence with nonlinearity measure' that involves three reliable nonlinear system identification methods and a 'Convergence area-based Nonlinear Metric' (CANM). The nonlinear identification methods in order are (a) An enhanced key term-based Sparse Identification of Nonlinear Dynamics with control (kSINDYc) (b) Standard Nonlinear Least Square method (NL2SQ) and (c) Neural Network-based Nonlinear Auto Regressive Exogenous input (N3ARX) schemes. This article revolves around the central idea of developing kSINDYc to capture the nonlinear dynamics of high nonlinear systems. Furthermore, the nonlinear metric CANM computes the process nonlinearity in the dynamic physical systems that classify the unknown process under mild, medium or highly nonlinear categories. Simulation studies are carried out on five industrial systems with divergent nonlinear dynamics. The user can make a flawless choice of a specific identification method suitable for a given process from CANM.

Keywords kSINDYc · Nonlinear least square · Nonlinear dynamics · System identification · Nonlinearity measure · CANM

J. Xavier
Dublin, Ohio 43016, United States
e-mail: joanjoean@gmail.com

S. K. Patnaik
Department of EEE, College of Engineering, Anna University, Chennai 600025, India
e-mail: skpatnaik@annauniv.edu

R. C. Panda (✉)
Department of Chemical Engineering, Chief Scientist and Faculty, CSIR-Central Leather Research Institute, Adyar, Chennai 600020, India
e-mail: joanjoean@gmail.com

1 Introduction

Most of the physical systems in real-life applications are nonlinear in nature. Safe operational practice of various industrial units needs mathematical modeling of the physical system, optimization and design of the control system. 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 (FP) concept was desirable for only

systems where adequate knowledge of it is available. Moreover, if the assumptions made in the derived models are inadequate or unrealistic, then the FP-based model might fail to capture the essential process dynamics. An alternative proposition to build a nonlinear model directly from the observation of measured data of the process is called system identification. 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 of the usage and complexity of many notable identification algorithms for nonlinear systems. To manifest a few, Schoukens and Ljung presented a review of 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 times. 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 was discussed in [8, 48]. 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 the 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 the 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 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]. Control of pH using adaptive nonlinear model-based control was implemented where process parameters were estimated [49]. Takagi–Sugeno (TS) fuzzy modeling with an unscented Kalman filter was carried over for a practical heat exchanger process [16]. Yet, the real challenge lies in the choice of fuzzy rule numbers on the output precision. In another research, the authors of [17] have put forth a Reliable Fuzzy Neural Network (ReFNN) which can handle reliability of nonlinear systems using an information reliability measure. The Stone-Weierstrass theorem was used to prove the universal approximation property of ReFNNs. Results showed that ReFNNs outperformed

traditional Feed forward Neural Networks in terms of error and sensitivity, especially in the presence of noise. Nevertheless, an integrated framework using data driven system identification tools and nonlinear metric has not executed by the researchers up to this point. In this article, we propose an integrated framework relating these two concepts. 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. It becomes difficult to select suitable identification techniques for a given system with unknown dynamics and unmeasured nonlinearity. In this article, we emphasize the importance of quantifying nonlinearity in order to choose an appropriate identification method. A control-relevant nonlinearity measure (CRNM) was proposed for measuring the nonlinear degree of a system when a linear control strategy is selected [18]. The CRNM method is an integrated multi-model control framework based on the gap metric and the gap metric stability margin. In spite of the investigations on two CSTR systems, the nonlinear metric failed to classify them based on nonlinearity level. Jiang et al. proposed a nonlinearity measure-based damage location method for beam-like structures [19]. Nonlinearity degree of the characteristic points in undamaged and damaged structures was compared to identify the location of damage by finding positions of maximal change. Nonlinear approaches are more sensitive in detecting breathing cracks in blades. A bicoherence-based nonlinearity measurement method was intended for identifying the location of breathing cracks in blades, which evaluated the extent of nonlinearity in the responses of blades under random excitations [20]. The Total Nonlinearity Index was used to establish the indicators of the cracks in blades, and the crack location was identified by finding the maximum components of the indicators. Generally, tools for analyzing nonlinear systems, like, describing function, phase portrait, perturbation, stability criteria (Lyapunov or Popov), and passivity are well-established. However, some of the existing units demand fault diagnosis and model order reduction of complex systems. Another salient aspect in the analysis of nonlinear physical systems, is the synthesis of closed loop systems with advanced control techniques. Zhaou [21], in his article focused on stability analysis and controller synthesis for stochastic networked control systems (NCS) under aperiodic denial-of-service (DoS) jamming attacks. An observer was constructed to estimate unmeasurable states, and a new adaptive event-triggered mechanism was proposed to reduce the transmission burden and mitigate the effects of DoS attacks. Furthermore, an observer-based controller was designed, and a switched system with time-varying delays was introduced in a mass-spring-damper mechanical system. In another work, a new control strategy for stochastic nonlinear systems (SNS) with state constraints and time-varying delays was presented [22]. The strategy used an event-triggered adaptive artificial neural network (AANN) and a barrier Lyapunov function (BLF) to handle the state constraints. The constructed AANN control scheme guaranteed stability and did not violate predefined constraints. The developed method also enriched the AANN control design of SNS. Alternatively, the design of the controller or achieving better closed loop-performance requires a pertinent identification scheme.

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 in separate attempts. However, there exists a break in the continuity between these two concepts for over years. There exist enumerable identification methods and methods for quantification of nonlinearity. However, well-designed directions or guidelines for the selection of identification algorithms (based on nonlinearity-measure) are rare and need to be established. The motivation for this research comes up with a bang by readdressing the issues of system identification and the concept of nonlinear metrics in a joint venture. The

research idea discussed in this article will overcome the existing disruption by formulating an integrated framework that relates the nonlinear metric (Δ_0) with three noteworthy system identification tools.

1.2 Major contributions

Research studies in the past have not attempted to employ the identification schemes (kSINDYc, NL2SQ and N3ARX) for mild, medium and highly nonlinear systems in the process engineering domain. The readers may refer to our earlier research contribution on the nonlinear metric CANM [23], to understand the detailed concepts of quantification of nonlinearity. Furthermore, in this article, we have made a reminiscent improvement from our earlier work on CANM [23], by introducing the Δ_0 criteria for stable, unstable and marginally stable systems. In this research, we have proposed a new framework for making an appropriate choice of system identification method by inspecting the degree of nonlinearity of the nonlinear system under investigation. The substructure of the proposed framework involves three steps.

1. To conduct nonlinear system identification of the physical system using (a) proposed data-driven kSINDYc identification (b) Neural network-based data-driven N3ARX method (c) parametric NL2SQ identification method.
2. To measure the degree of nonlinearity of the physical system at the specified operating region using the nonlinearity measure CANM (Δ_0).
3. To make a suitable choice of system identification by mapping the nonlinearity level Δ_0 with the identification method that dispenses the least RMSE.
4. Furthermore, the nonlinear metric, namely CANM method, is upgraded in this article by recommending certain directives, on the computation of Δ_0 for stable, unstable and marginally stable systems.

Besides, assimilating the proposed framework for five different physical systems from chemical engineering units with different nonlinear levels has not been carried out in the existing literature. The data driven SINDYc identification proposed in [24], does not provide descriptive library terms based on the nonlinearity, which in turn makes it ill-conditioned in the prediction of complex nonlinear processes. This major concern is drenched here, by choosing a fewer number of relevant key terms in the candidate library of the kSINDYc scheme. This paper also addresses this issue by providing the relevant choice of key terms based on the degree and type of nonlinearity of dynamic nonlinear systems. The paper is divided into six Sects. Section 1 has introduced the literature review of many system identification routines and nonlinear metric tools. Section 2 investigates the nonlinearity metric CANM for stable, unstable and marginally stable systems. Section 3 elucidates the concept of the proposed framework consisting of the three identification methods kSINDYc, NL2SQ and N3ARX and nonlinear metric CANM. It is followed by simulation results in Sect. 4 which show that the computed nonlinearity Δ_0 as well as evaluation index RMSE witness a major lively role in deciding the choice of nonlinear identification method for the five dynamic systems with contrasting nonlinearity. Besides, Sect. 4 also adds increased flavor to the current study by suggesting a suitable parsimonious model for every physical system under study and Sect. 5 concludes the article. Section 6 presents the future directions of research of this article.

2 Nonlinearity metric–CANM

The nonlinearity of the physical systems is an important characteristic to be inscribed 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 the nonlinearity of typical industrial processes and their impacts on popular system identification schemes. Nevertheless, there are several nonlinear indices to mark the value of nonlinearity in dynamic systems [25–27]. The concept of Convergence-area-based-nonlinearity measure (CANM) proposed in [23] has

been endorsed in the current study. Additionally, the calculation of the nonlinear metric Δ_0 for stable, unstable and marginally stable systems is, refurbished in this article by recommending some directives. Without loss of generality, consider a nonlinear dynamic system of the form

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

in which $x(t) \in \mathbb{R}^m$ denotes the state variables of a system at a time t.Eq. (1) also generalizes the first principle model of nonlinear systems. If $y_{\mathrm{True}}(t)$ and $y_{\mathrm{lin}}(t)$ represent the measured output (True output) of the nonlinear system and its linearized response at the jth operating point P_j, then the nonlinear metric Δ_{0j} quantifies the level of nonlinearity as given in Eq. (2).

$$\Delta_{0j} = \frac{\left|\left|\int_0^{t_f} y_{\mathrm{True}}\mathrm{d}t\right| - \left|\int_0^{t_f} y_{\mathrm{lin}}\mathrm{d}t\right|\right|}{\left|\int_0^{t_f} y_{\mathrm{True}}\mathrm{d}t\right|} \tag{2}$$

where t_f represents the settling time 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 $\Delta_{0\mathrm{nom}}$ is shown in Eq. (3)

$$\Delta_{0\mathrm{nom}} = \frac{\Delta_{0P_1} + \Delta_{0P_2} + \cdots \Delta_{0P_j} + \cdots \Delta_{0P_m}}{m} \tag{3}$$

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 $y_{\mathrm{lin}}(t)$ which thoroughly depends on analysis of an operating point. The stability of the operating point decides the current dynamic behavior of the plant. A nonlinear system has multiple operating points, unlike a linear system which has only one operating point with zero initial condition. In a nonlinear system, with multiple operating points, the initial condition by itself is an operating point, which may be stable or unstable. Another class of nonlinear systems are chaotic processes, which don't have initial conditions. The scope of study in this manuscript does not include any chaotic system. Moreover, CANM is an operating point-dependent nonlinear metric. So to maintain a standard consistency in the nonlinear metric, the examples explored in Sect. 4 of this article are subjected to initial conditions and excitation inputs at nominal operating points referring to the concerned literature. The effect of the initial condition and the type of excitation signal applied to a physical system will definitely affect Δ_0. Considering this characteristic, the nonlinear systems elaborated in Sect. 4 are subjected to step (u_{nom}) and PRBS (u_{prbs}) inputs, and the effect of Δ_0 over the excitation signals is also investigated.

The simulations for the computation of nonlinearity CANM are restricted only to SISO systems. CANM method proposed in [23], is upgraded in this article by recommending the following directives, on computation of Δ_0 for stable, unstable and marginally stable systems.

Case (i): Δ_0 for stable systems.

For any stable system, the eigen values of the Jacobian linearized model will have their eigen values on L.H.S of 's' plane. If t_f represents the settling time of the nonlinear system around the vicinity of the stable steady state operating point P, then Δ_{0P} is operating point dependent and is given as

$$\Delta_{0\_\mathrm{stable}} = \frac{\left|\left|\int_0^{t_f} y_{\mathrm{True}}\mathrm{d}t\right| - \left|\int_0^{t_f} y_{\mathrm{lin}}\mathrm{d}t\right|\right|}{\left|\int_0^{t_f} y_{\mathrm{True}}\mathrm{d}t\right|} \tag{4}$$

Case (ii): Δ_0 for marginally stable systems.

A crucial point in CANM is finding the nonlinearity for systems with transient states (marginally stable system and unstable systems). In a marginally stable system, the eigen values of the linearized model (y_{lin}) are located on the imaginary axis. The response y will display sustained oscillations and there is no steady state t_f. While finding Δ_0,instead of choosing t_f, it is suggested to use t_{cycle} as the sustained oscillations repeat with the same time period after every cycle. Then Δ_0 becomes

Table 1 Computation of Δ_0 for nonlinear systems

System	Behavior of the system	Δ_0 at operating point P
Stable	t_f is available, attains steady state	$\Delta_{0\_\text{stable}} = \frac{\left\|\left\|\int_0^{t_f} y_{\text{True}}\mathrm{d}t\right\| - \left\|\int_0^{t_f} y_{\text{lin}}\mathrm{d}t\right\|\right\|}{\left\|\int_0^{t_f} y_{\text{True}}\mathrm{d}t\right\|}$
Marginally stable	Oscillatory response. Obtain the time period for one cycle of sustained oscillations t_{cycle}	$\Delta_{0\_\text{marg.stable}} = \frac{\left\|\left\|\int_0^{t_{\text{cycle}}} y_{\text{True}}\mathrm{d}t\right\| - \left\|\int_0^{t_{\text{cycle}}} y_{\text{lin}}\mathrm{d}t\right\|\right\|}{\left\|\int_0^{t_{\text{cycle}}} y_{\text{True}}\mathrm{d}t\right\|}$
Unstable	No steady state. t_f cannot be selected as infinity; Instead consider small time intervals $(t_1, t_2, \ldots t_n)$ in regions $(R_1, \ldots R_n)$ $(t_1, t_2, \ldots t_x \ldots t_n)$.Measure Δ_0 from t_1 to t_n	$\Delta_{0R_x} = \frac{\left\|\left\|\int_{t_{x-1}}^{t_x} y_{\text{True}\_x}\mathrm{d}t\right\| - \left\|\int_{t_{x-1}}^{t_x} y_{\text{lin}\_x}\mathrm{d}t\right\|\right\|}{\left\|\int_{t_{x-1}}^{t_x} y_{\text{True}\_x}\mathrm{d}t\right\|}$ $\Delta_{0\_unstable} = \frac{\Delta_{0R_1}+\Delta_{0R_2}+\cdots\Delta_{0R_x}+\cdots\Delta_{0R_n}}{n}$

$$\Delta_{0\_\text{marg.stable}} = \frac{\left\|\left\|\int_0^{t_{\text{cycle}}} y_{\text{True}}\mathrm{d}t\right\| - \left\|\int_0^{t_{\text{cycle}}} y_{\text{lin}}\mathrm{d}t\right\|\right\|}{\left\|\int_0^{t_{\text{cycle}}} y_{\text{True}}\mathrm{d}t\right\|} \tag{5}$$

Case (iii): Δ_0 for unstable systems:

In an unstable system, the eigen values of the Jacobian linearized model will occur on the R.H.S of the '*s*' plane. But there is no steady state t_f for unstable system. (Ex: Batch and transient processes in Chemical Reactors). The unstable response shows a transient behavior. Moreover t_f cannot be chosen as infinite. In such cases, local nonlinearity analysis will be an alternative solution. The nonlinear metric Δ_0 for unstable system can be obtained by making a trajectory dependent analysis of the measured output. To attain this feature, the whole sequence of output y_{True} is considered a trajectory which can be broken into many short time intervals $(t_1, t_2, \ldots t_x \ldots t_n)$ with n number of regions such that we can obtain piece-wise models. t_n corresponds to the time instant applied by the user to sort out the dynamic transient response. Then the nonlinearity metric Δ_0 becomes trajectory dependent and is computed by taking a cumulative mean from all regions $(R_1, R_2, \ldots R_x \ldots R_n)$ as follows.

$$\Delta_{0R_x} = \frac{\left\|\left\|\int_{t_{x-1}}^{t_x} y_{\text{True}\_x}\mathrm{d}t\right\| - \left\|\int_{t_{x-1}}^{t_x} y_{\text{lin}\_x}\mathrm{d}t\right\|\right\|}{\left\|\int_{t_{x-1}}^{t_x} y_{\text{True}\_x}\mathrm{d}t\right\|} \tag{6}$$

$$\Delta_{0\_\text{unstable}} = \frac{\Delta_{0R_1} + \Delta_{0R_2} + \cdots \Delta_{0R_x} + \cdots \Delta_{0R_n}}{n} \tag{7}$$

The main difference between Δ_{0_stable},$\Delta_{0\_\text{marg.stable}}$ and $\Delta_{0\_\text{unstable}}$ lies in the time interval limit t_f,t_{cycle} and t_n. This sort of analysis can also be applied to batch processes in many chemical reactor units. In many batch processes, the eigen values are stable only at the beginning, but as the batch process continues, the eigen values become unstable.

$$\text{Nonlinearity level} = \begin{cases} \Delta_0 \leq 0.3, \text{ mild nonlinear} \\ 0.3 < \Delta_0 \leq 0.7, \text{ medium nonlinear} \\ \Delta_0 > 0.7, \text{ highly nonlinear} \end{cases} \tag{8}$$

Equation (8), implies the classification of nonlinearity as mild, medium or highly nonlinear using the CANM metric where the value of Δ_0 for any nonlinear system is consigned between 0 and 1. Table 1 gives an outright summary on the computation of Δ_0 for stable,

unstable and marginally stable systems, working at a single operating point P.

Over and above, that the nonlinearity level of any dynamic system will have a serious impact on identifying the dynamics of the complex 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 System identification in coherence with nonlinearity measure

So far, literatures have discussed the idea of nonlinear system identification and nonlinearity measurement in individual research studies. Moreover, to this notch, research on the usage of nonlinear system identification based on classification of nonlinearity remains very limited. This implication necessitates the requirement of a mathematical tool to bridge the gap between nonlinearity measurement and nonlinear system identification. 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. Viewed in this way, we have proposed a single framework 'System identification in coherence with nonlinear measure' with the assorted combination of (a) nonlinear identification schemes (kSINDYc, NL2SQ, N3ARX) and (b) CANM nonlinear metric. This combined framework will ensure an appropriate choice of system identification by inspecting the degree of nonlinearity of the nonlinear system under investigation. Among the three identification methods, kSINDYc identification is proposed in this article and its accuracy is compared with NL2SQ and N3ARX methods available in existing literature.

3.1 System identification using kSINDYc

The recent impeccable SINDYc (Sparse Identification of Nonlinear Dynamics with control) algorithm is a celebrated parsimonious system identification technique introduced by Brunton [24]. 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 [29–32]. The SINDy (Sparse Identification of Nonlinear Dynamics) algorithm is a symbolic sparse regression problem, to identify nonlinear systems. It uses a candidate library with higher order polynomials, trigonometric terms, logarithmic functions etc., in Eq. (9) to identify any unknown process. The term 'candidate library' refers to a set of diversity of many functions to determine the learned SINDYc models.

$$\Theta(X,U) = \begin{bmatrix} | & | & | & | & | & | & | & | \\ 1 & X & U & X\otimes X & X\otimes U & \ldots \sin(X) & \ldots e^{X} & \ldots \log(X) \\ | & | & | & | & | & | & | & | \end{bmatrix} \tag{9}$$

where $X \otimes U$ denotes the vector of all product combinations in X and U. The use of higher order polynomials or trigonometric nonlinearities or other mathematical functions in Eq. (9) without observing the system nonlinearity might cause numerical problems, which in turn engenders unnecessary oscillations in the predicted model outputs. Moreover, without descriptive library terms, the size of the $\Theta(X,\ U)$ grows rapidly, which sequentially drives the SINDYc library to be ill-conditioned. Recent literature by [28] have substantiated that SINDYc may be susceptible to over fitting problem if care is not taken to balance the model complexity and polynomial order in its candidate library. In this article, this major concern is drenched, by the introduction of kSINDYc identification scheme for nonlinear systems. In kSINDYc method, we have refined the selection of the candidate library $\Theta(X,\ U)$ with the inclusion of *key nonlinear terms* (k_{nl}) that describe the system dynamics. The term k_{nl} is chosen as a basic nonlinear function that plays a vital role in deciding the nonlinear dynamics of the physical system. Our research concentrates on establishing a streamlined library function $\Theta(X,\ U)$ for kSINDYc, where library

terms are chosen based on the nonlinearity that contemplates the system dynamics rather than choosing the library elements in trial and error criteria as in SINDYc. Even though kSINDYc is a data-driven approach, dynamic equations describing the identified model can be obtained from Eq. (11) after the inclusion of k_{nl} in the library Θ(X, U). Consequently, the learned model using the data-driven kSINDYc approach predicts the nonlinear dynamics of the physical system, with a lesser number of parametric terms in $\Theta(X,\ U)$ without overfitting issues. 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. Consider a nonlinear dynamic system of the form given in Eq. (10).

$$\dot{x}(t) = f(x) + g(x,u) \tag{10}$$

where $x \in \mathbb{R}^m$ denotes the state of the system at time t and $u \in \mathbb{R}^p$ gives the manipulated input vector. The function $f(\cdot)$ and $g(\cdot)$ represent the system parameters that capture the physics-based dynamics of the system. Inspired by its application to many physical systems, kSINDYc is formulated to determine the nonlinear dynamics of Eq. (10) using the measured input and output data.$\dot{X}$ is a data matrix that gives the time derivatives of state variables in the sparse regression problem in Eq. (11). Θ is the augmented library matrix 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 matrix $U = \left[u_1(t), u_2(t) \ldots u_p(t)\right]^T \in \mathbb{R}^p$, the state vector (data matrix) $X = [x_1(t), x_2(t) \cdots x_m(t)]^T \in \mathbb{R}^m$, then the sparse regression becomes

$$\dot{X} = \Theta(X,U)\xi \tag{11}$$

$$\xi = [\xi_1\ \xi_2 \ldots\ \xi_m] \tag{12}$$

Equation (12) is vector that has the sparse coefficient $\xi_1 \xi_2 \ldots \xi_n$ corresponding to $\Theta(X,U)$.

$$X = \begin{bmatrix} x^T(t_1) \\ x^T(t_2) \\ \vdots \\ x^T(t_f) \end{bmatrix} = \begin{bmatrix} x_1(t_1) & x_2(t_1) & \ldots & x_m(t_1) \\ x_1(t_2) & x_2(t_2) & \cdots & x_m(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(t_f) & x_2(t_f) & \cdots & x_m(t_f) \end{bmatrix} \tag{13}$$

$$\boldsymbol{U} = \begin{bmatrix} u^T(t_1) \\ u^T(t_2) \\ \vdots \\ u^T(t_f) \end{bmatrix} = \begin{bmatrix} u_1(t_1) & u_2(t_1) & \ldots & u_p(t_1) \\ u_1(t_2) & u_2(t_2) & \cdots & u_p(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ u_1(t_f) & u_2(t_f) & \cdots & u_p(t_f) \end{bmatrix} \tag{14}$$

$$\dot{X} = \begin{bmatrix} \dot{x}^T(t_1) \\ \dot{x}^T(t_2) \\ \vdots \\ \dot{x}^T(t_f) \end{bmatrix} = \begin{bmatrix} \dot{x}_1(t_1) & \dot{x}_2(t_1) & \ldots & \dot{x}_m(t_1) \\ \dot{x}_1(t_2) & \dot{x}_2(t_2) & \cdots & \dot{x}_m(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \dot{x}_1(t_f) & \dot{x}_2(t_f) & \cdots & \dot{x}_m(t_f) \end{bmatrix} \tag{15}$$

The key terms in the kSINDYc library function matrix $\Theta(X,U)$ are given in Eq. (16)

$$\Theta(X,U) = \begin{bmatrix} | & | & | & | & | & | \\ 1 & X & U & X \otimes X & X \otimes U & \ldots k_{nl} \\ | & | & | & | & | & | \end{bmatrix} \tag{16}$$

$X \otimes U$ in Eq. (16) denotes the vector of all product combinations in X and U.It also indicates the quadratic nonlinearities in the unknown system.

$$\xi_k = \underset{\hat{\xi}_k}{\operatorname{argmin}} \frac{1}{2}\left\|\dot{X}_k - \Theta(X,U)\hat{\xi}_k\right\|_2^2 + \lambda\left\|\hat{\xi}_k\right\|_1 \tag{17}$$

The l_1 regularized optimization problem given in Eq. (17) can be evaluated using sparsity promoting scheme called STLS (Sequential Threshold least square method). Equation (17) penalizes the number of active terms in the candidate library $\Theta(X,U)$. The second part of Eq. (17) has the penalty term with the tunable weighing parameter $\lambda \geq 0$ to establish model parsimony.ξ_k represents kth row of ξ and $\dot{X}_k$ represents kth row of $\dot{X}$. Equation (17) is solved iteratively till the coefficients converge. A notable development is made on the SINDYc candidate library function, by introducing the *'key nonlinear term'* from the plant dynamics, apart from the other higher-order polynomials of the processes. The algorithmic pseudocode for the proposed kSINDYc identification is given below:

Algorithm 1 Key term SINDYc (kSINDYc)

Require: The time series data U and X for n number of sample sets
Input : Candidate function Θ ;time derivatives $\dot{\mathrm{X}}$;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 X and U matrices given in Eq. (13) and (14)
4: Include the key nonlinearity term k_{nl} in library Θ
5: **Build** the matrix $\dot{\mathrm{X}} = \Theta(\mathrm{X},\mathrm{U})$
6: Compute $\dot{\mathrm{X}}$ as given in Eq. (11)
7: **Solve** the sparse regression problem ξ as given in Eq. (12) 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*) $\xi_k = 0$
d. Identify non-zero coefficients (*i.e*) $\xi_k \neq 0$ and obtain another least square solution for ξ
e. Threshold the new cut-off using λ
f. Repeat steps *a to e* until the non-zero coefficients converge
end procedure *STLS*
9: **Calculate** the sparse coefficients ξ for all elements of X
10: end *for*
11: Compute the evaluation index RMSE
12: end procedure 'kSINDYc'

A coarse sweep of the tunable parameter λ in the range $\lambda_{\min} < \lambda \leq \lambda_{\max}$ is considered in Algorithm 1 to attain an optimal solution with minimum error and maximum convergence rate. The specific convergence criteria for STLS algorithm in sparse regression framework is provided in [33]. 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.

A comparison between the existing SINDy method and the proposed kSINDYc is summarized in Table 2. The descriptive library $\Theta(X, U)$ built for each nonlinear system is unique and the same is discussed elaborately in Sect. 4 of this article.

3.2 System identification using NL2SQ

Many works on parametric system identification have used least square method to estimate the numerical values of parameters [33, 34]. NL2SQ with Levenberg–Marquardt algorithm is another established method used for optimizing the process parameters in the field nonlinear system identification [35, 36]. An optimization problem on any nonlinear system targets on maximizing profit or minimizing the overall cost. The optimization problem so formed will have a set of independent variables with some restrictions called constraints. The solution to any optimal problem depends on the allows set of variables where the objective function $f(x)$ is minimized and attains an optimal value. For physical systems with many nonlinear functions, the objective function $f(x)$ is framed as the sum of squares of the nonlinear function $r(x)$ as follows

$$f(x) = \frac{1}{2}\sum_{j=1}^{m}\left(r_j(x)\right)^2 = \frac{1}{2}\|r(x)\|_2^2 \tag{18}$$

In Eq. (17) the objective function $f(x)$ has to be minimized. Then Eq. (18) becomes

Table 2 Overview of SINDy and kSINDYc

Feature	SINDy	kSINDYc
Accuracy and model fit	Noisy or inconsistent data can lead to inaccurate models. Sufficient data is essential for robust modeling	Works well for systems with limited data samples also. The identified model is robust to uncertainties
Data requirements	Requires quality data less corrupted by noise	Can handle noisy or sparse data easily
Sparsity	Finding a compact and interpretable model that captures the essential physics of the nonlinear system with minimal library terms is challenging	A subset of key nonlinear terms is selected in the candidate library $\Theta(X, U)$ based on the nonlinearity of the system to promote sparsity
Model complexity	Complexity increases for highly nonlinear systems	Comparatively simpler than SINDy for highly nonlinear systems
Nonlinearity handling	Works well for mild and medium nonlinear systems with subset of trigonometric and polynomial terms in the library	Suitable for highly nonlinear system with a descriptive key term library
Incorporation of prior knowledge	Prior knowledge of nonlinear system is not mandatory	Prior knowledge of nonlinear system is important for choice of candidate library
Computational efficiency	Comparatively less	Fast

$$\min_x f(x) = \sum_{j=1}^{m} \left(r_j(x)\right)^2 \tag{19}$$

In Eq. (18), the sum of least squares of the nonlinear function $r(x)$ is minimized, and hence the optimization problem with this methodology is called Nonlinear Least Square method (NL2SQ). If a nonlinear system has a model function $\phi(x)$ and the measured output be y_{True} then

$$r_j(x) = \phi\left(x, t_j\right) - y_{True(j)} \tag{20}$$

$$r(x) = (r_1(x), r_2(x), \cdots r_m(x))^T \tag{21}$$

The residual vector $r(x)$ has m number of components. To solve the least square problem, the most common algorithms used are Gauss Newton method and Levenberg Marquardt (LM) methods. Gradient of the objective function $\nabla f(x)$ is expressed from Eq. (22) as

$$\nabla f(x) = \sum_{j=1}^{m} r_j(x)\nabla r_j(x) = J(x)^T r(x) \tag{22}$$

where the Jacobian term $J(x)$ is given as

$$J(x) = \left(\frac{\partial r_j}{\partial x_i}\right)_{j=1,\ldots m;\ i=1,\ldots n} = \begin{pmatrix} \nabla r_1(x)^T \\ \nabla r_2(x)^T \\ \vdots \\ \nabla r_m(x)^T \end{pmatrix} \tag{23}$$

Hessian matrix $H(f(x))$ consists of second-order partial derivatives of $r(x)$ and is given by $\nabla^2 f(x)$

$$\begin{aligned} \nabla^2 f(x) &= \sum_{j=1}^{m} \nabla r_j(x)\nabla r_j(x)^T + \sum_{j=1}^{m} r_j(x)\nabla^2 r_j(x) \\ &= J(x)^T J(x) + \sum_{j=1}^{m} r_j(x)\nabla^2 r_j(x) \end{aligned} \tag{24}$$

The Hessian matrix observed in Eq. (24) must be positive definite in all the least square solutions. Equation (23) can be approximated to have the first term of Jacobian function alone, eliminating the second term $\nabla^2 r(x)$, when the residual is very close to the actual solution. This approximation is followed in LM method adopted in this article. The Jacobian matrix $J(x)$ of $H(x)$ has to be found out to optimize x for m number of samples. Using LM algorithm in [35], the objective function for NL2SQ method is modified as $h_{\text{LM}}(x)$.

$$(J(x)^T J(x) + \mu I) h_{\mathrm{LM}}(x) = -J(x)^T f(x) \tag{25}$$

$$h_{\mathrm{LM}}(x) = -(J(x)^T J(x) + \mu I)^{-1} J(x)^T f(x) \tag{26}$$

The damping factor is always $\mu \geq 0$,for which the following effects are observed. When $\mu > 0$, the coefficient $(J(x)^T J(x) + \mu I)$ is positive definite and so $h_{LM}(x)$ is in descent direction. If μ is very large $h_{\mathrm{LM}}(x) = -\frac{1}{\mu} J(x)^T f(x)$ and goes into the steepest descent direction. On the other hand, if μ is very small $h_{\mathrm{LM}}(x) = h_{\mathrm{GN}}(x)$, the LM algorithm converges with the Gauss Newton method. In the gradient descent method, the $h_{\mathrm{LM}}(x)$ is minimized by updating 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 $h_{\mathrm{LM}}(x)$ 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 System identification using N3ARX

Neural Networks is another computational intelligence approach for identifying nonlinear systems in real world scenario with accurate estimations [37]. Neural networks are well-suited for nonlinear modeling tasks because they can learn complex patterns and relationships in the data, without knowing any prior knowledge about the dynamics of the system. N3ARX combines exogenous inputs (X) using feed forward neural networks to capture complex nonlinear relationships in time-series data and make predictions based on both past values of the series and external inputs. The exogenous inputs (X) represent additional factors or variables that may influence the time series but are not directly part of the series itself are fed into the neural network along with the autoregressive inputs (AR). The N3ARX method is a standard identification technique and is found in enormous literatures [38–41]. 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 [42]. 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. The N3ARX model employs neural networks to capture nonlinear relationships between the autoregressive and exogenous inputs and the target variable. The hidden layers of the neural network enable the model to capture and represent these nonlinear relationships. 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 is chosen iteratively. A simple Neural network structure is taken with 1 hidden layer, and nonlinear Rectified linear Activation function (RELU) for the simulation studies carried out in this article. The regressor equation for the N3ARX model is given by

$$y(t) = F(y(t-1), y(t-2), \ldots, y(t-n), u(t), u(t-1), \ldots, u(t-m)) \tag{27}$$

where $y(t)$ refers the target variable at time t. $y(t-1), y(t-2), \ldots, y(t-n)$ are the past value of the target variable also called AR inputs. $u(t), u(t-1), \ldots, u(t-m)$ are the exogenous inputs (X) at time t and their past values. The function F is the feedforward neural network architecture with the weights and bias terms learned during the training process. It is evident from Eq. (27) that F maps the inputs to the target variable $y(t)$. The N3ARX model is trained using the historical time-series data with corresponding exogenous inputs. Followingly, the predicted model is optimized by adjusting the weights and biases of the neural network to minimize the prediction error between the model's output and the actual target values. This is typically done using gradient descent optimization algorithm. Given the previous values of the time series and the corresponding exogenous inputs, the N3ARX model can generate predictions for the future values of the target variable.

3.4 Proposed framework

In Fig. 1, 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

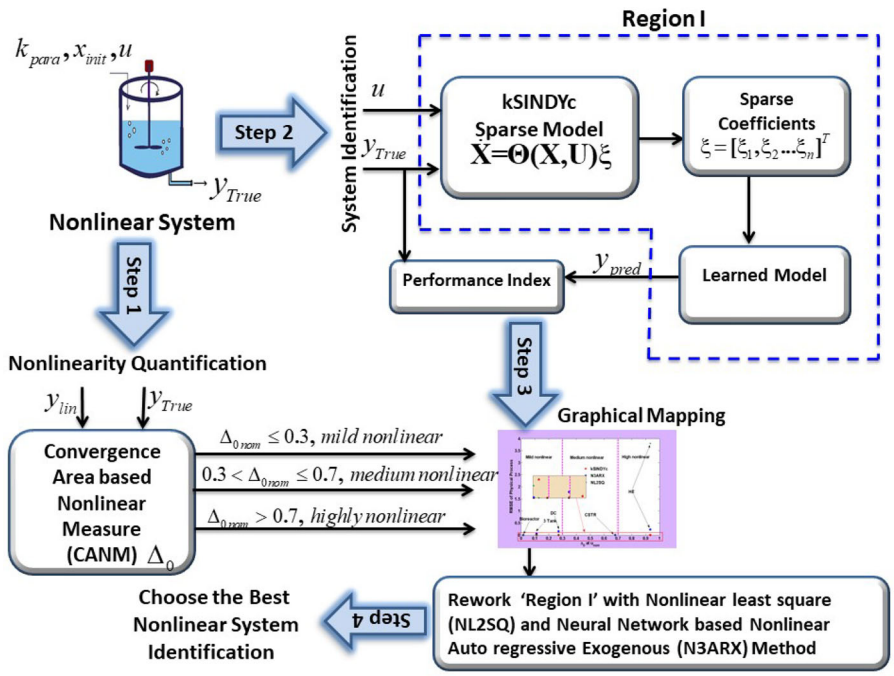

Fig.1 Proposed framework- System identification in coherence with nonlinear measure

nonlinear output $y_{\text{True}}(t)$ from the nonlinear process are treated as measured input and output data. Region I contain the proposed kSINDYc algorithm to learn the dynamics of $y_{\text{True}}(t)$.

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 physical system. The predicted output $y_{\text{pred}}(t)$ of Region I and measured output $y_{\text{True}}(t)$ 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 $y_{\text{lin}}(t)$. The metric 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, that makes $y_{\text{pred}}(t) = \{y_{\text{kSINDYc}}, y_{\text{NL2SQ}}, y_{\text{N3ARX}}\}$. The steps involved in the proposed view of nonlinear system identification and nonlinearity quantification are as follows:

1. Measure the nonlinearity Δ_0 of the nonlinear dynamic system (Plant -P) under study using the CANM metric.
2. Identify (Predict) the dynamics of the plant $y_{\text{pred}}(t)$ using (a)kSINDYc (b)NL2SQ and (c) N3ARX methods.
3. Create a graphical mapping between Δ_0 obtained in step 1 and $y_{\text{pred}}(t)$ acquired in step 2 using a performance index.
4. Sort out the suitable choice of system identification from kSINDYc, NL2SQ and N3ARX, by mapping the nonlinearity level Δ_0 with the identification method which dispenses least evaluation index.

The crucial step in the proposed framework is the selection of an applicable identification from kSINDYc, NL2SQ and N3ARX 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. This article differs from the existing literature by providing guidelines for suitable nonlinear identification from the three methods based on Δ_0. The proposed substructure also 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.

4 Simulation study

A reliable quantitative analysis is exemplified to cohere the nonlinear metric Δ_0 with the above said identification methods for five nonlinear systems with divergent nonlinear strengths. The simulation study is carried out on five Industrial physical systems with different nonlinearity levels, ranging from chemical to biological domain from the process engineering field. The nonlinear metric Δ_0 subjected to excitations u_{nom} and u_{prbs} are figured out for all the examples of Sect. 4. The computed values of Δ_0 can be inspected in Tables 4 and 5, to check whether the nonlinear system falls under mild, medium or highly nonlinear. The measured data (y_{True}) for all the five case studies are obtained from the first principle equations (also called true models). The measured outputs (y_{True}) are utilized in the data driven identification of kSINDYc and N3ARX methods. The models developed using kSINDYc, N3ARX and NL2SQ identification techniques are called predicted models (learned models) whose outputs are indicated as (y_{pred}). 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 input output data Z^N obtained using first principle equations.

$$Z^N = [u(1), y(1), u(2), y(2)...u(N), y(N)] \tag{28}$$

u and y corresponds to the excitation signal and the response of the SISO system, t_f denotes the final time for the Nth sample. The data set Z^N is subjected to preprocessing before proceeding with the prediction, by splitting it into training set $(Z^{N_{\text{train}}})$ and testing set $(Z^{N_{\text{test}}})$

$$Z^{N_{\text{train}}} = 0.7Z^N \tag{29}$$

$$Z^{N_{\text{test}}} = 0.3Z^N \tag{30}$$

where

$$N = N^{\text{train}} + N^{\text{test}}; \tag{31}$$

N^{train} and N^{test} conform to the training and testing data samples. As observed from Eq. (29) and (30), 70% of Z^N is taken randomly for training and remaining data (30%) for testing purpose. All the case studies are simulated with a sampling time of $T_s = 0.01\ s$ in N number of sample space.$Z^{N_{\text{train}}}$ intends the input–output data taken for training kSINDYc, N3ARX and NL2SQ algorithms and remaining $Z^{N_{\text{test}}}$ corresponds to testing dataset. The following Section exemplifies five industrial systems 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 3rd tank by measuring the level in each tank. The dynamic equations and the associated process parameters c_{12}, c_{23}, c_3, A_1 are acquired from [43]. The terms h_1, h_2 and h_3 stand for the individual level of each tank in the cascaded-arrangement of three-tank process, respectively. The state variable h_3 is the only measurable output when the input flow rate to Tank 1 is $q_1\ m^3\ s^{-1}$. The nonlinear differential equations of the three-tank system are given by Eq. (32–34). 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 $u = q_1\ m^3\ s^{-1} \quad q_1 \in [0 - 1e^{-5}]m^3s^{-1}$ and $u_{\text{nom}} = 0.5e^{-5}m^3s^{-1}$.

$$\dot{h_1} = \frac{q_1}{A_1} - c_{12}\sqrt{h_1 - h_2} \tag{32}$$

$$\dot{h_2} = c_{12}\sqrt{h_1 - h_2} - c_{23}\sqrt{h_2 - h_3} \tag{33}$$

$$\dot{h_3} = c_{23}\sqrt{h_2 - h_3} - c_3\sqrt{h_3} \tag{34}$$

$$y_{\text{True}} = h_3 \tag{35}$$

The step response for the tank level $y_{\text{True}}(t)$ from Eq. (34) is compared with that $y_{\text{pred}}(t)$ obtained from the predicted models identified by kSINDYc, N3ARX and NL2SQ where the training dataset $Z^{N\,\text{train}}$ is presented in Fig. 2.The nonlinearity of the three-tank using CANM adheres to The three-tank process has a large settling time of around 5000 s with a weak nonlinear behavior.

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.The height of h_3 swirls over a level band between $(0 - 1.5)$m when subjected to u_{prbs} unlike the

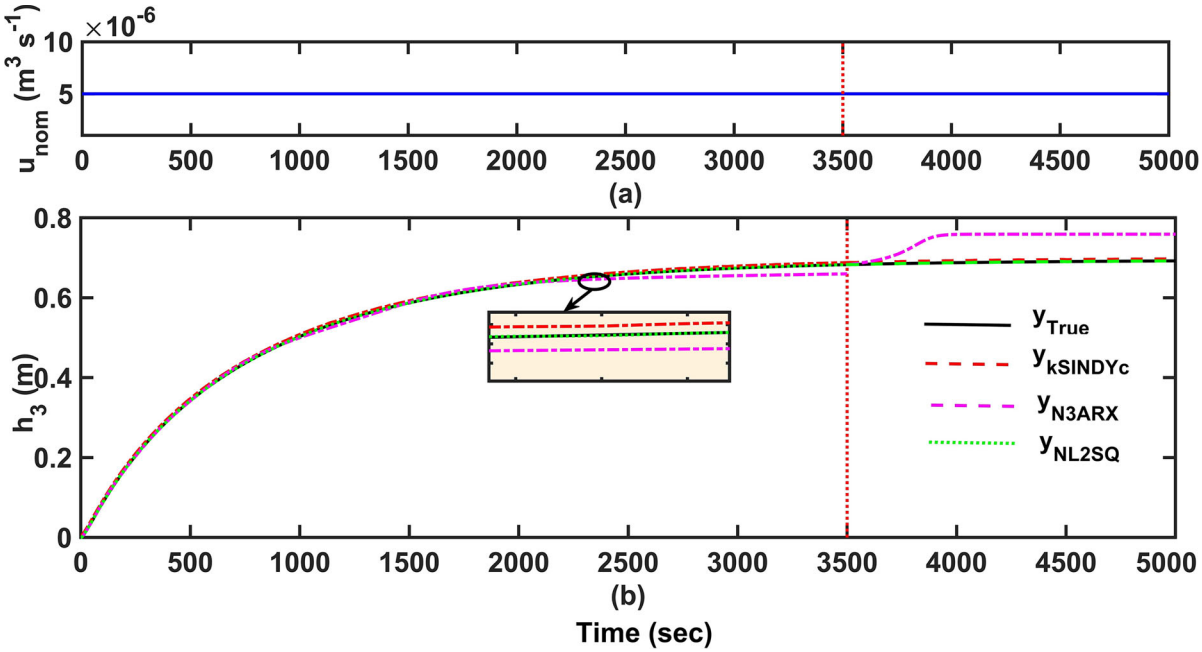

Fig. 2 True and learned model response h_3 of three tank process for step input at $u_{\text{nom}} = 0.5e^{-5}\text{m}^3\text{s}^{-1}$

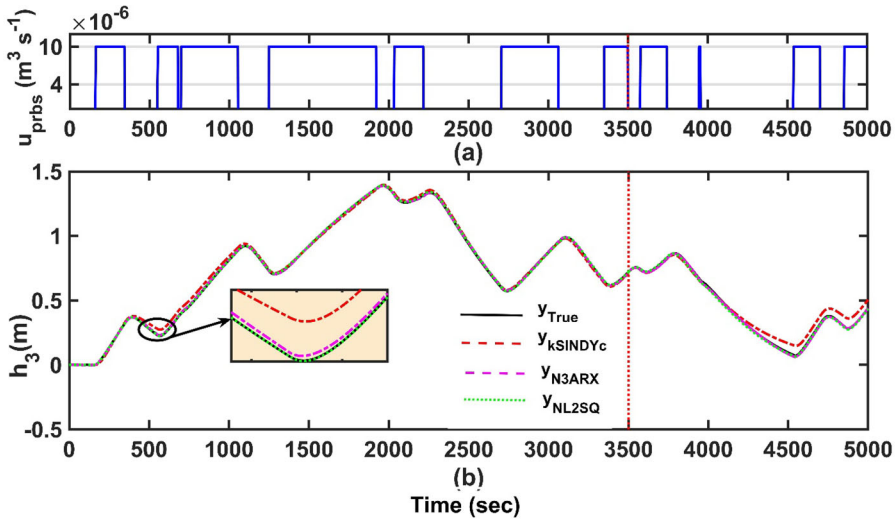

Fig. 3 True and learned model response of three-tank process for input at u_{prbs}

response to u_{nom} that attains the steady state at $h_3 = 0.7$m. The system suffers from mild nonlinearity where the value of $\Delta_0 < 0.3$ using CANM. It has been observed that NL2SQ identification approach outperforms kSINDYc in predicting the dynamics of $y_{\text{True}} = h_3$ for 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 $(A \rightarrow P)$. The reactor suffers from operational difficulties like complex behavior, output multiplicity (as it shows multiple steady states), oscillations and chaos due to its nonlinear dynamics. Here we consider a uniformly mixed CSTR which undergoes a single irreversible, exothermic reaction. Rate of reaction and heat transfer from heating media to reactor wall impose nonlinearity to the system. Sometimes, due to economic reason, the reactor is

preferred to be operated at an unstable steady state. The rate of heat generation and rate of heat removal should be balanced by rate of cooling for efficient control so that dynamic disturbances can be safely handled. The standard state variable representation of the reactor is given in Eqs. (36, 37). The coolant flow rate $q_c = u_{\text{nom}}$ 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 [44]. 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 $(C_{\text{anom}}, T_{\text{nom}}) = (0.08235 \text{ mol l}^{-1}, 441.81\text{K})$.The open loop study obtained for a nominal input $u_{\text{nom}} = 102\ l\,\text{min}^{-1}$ can be viewed from Fig. 4.

$$\dot{C}_a = \frac{q}{V}\left(C_{af} - C_a\right) - k_0 C_a \exp\left(-\frac{E}{\text{RT}}\right) \tag{36}$$

$$\dot{T} = \frac{q}{V}\left(T_f - T\right) + \frac{(-\Delta H)k_0 C_a}{\rho C_p}\exp\left(-\frac{E}{\text{RT}}\right) + \frac{\rho_c C_{pc}}{\rho C_p V} q_c \left[1 - \exp\left(-\frac{hA}{q_c \rho_c C_{pc}}\right)\right](T_c - T) \tag{37}$$

$$y_{\text{True}} = T \tag{38}$$

By carefully observing Eqs. (36) and (37), 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 the term $\exp\left(-\frac{E}{\text{RT}}\right)$.

The open loop (temperature) responses, with the jacketed-coolant flowrate at $u_{\text{nom}} = 102\ l\,\text{min}^{-1}$ of the CSTR studied in Fig. 4, ensures that both kSINDYc and NL2SQ expedite the process dynamics more accurately. Moreover, the N3ARX method shows notable deviations in predicted temperature $(y_{\text{pred}}(t))$ from the true reactor temperature $y_{act}(t)$ where accuracy falls down with a value of RMSE = 3.2382. The input PRBS region, $u_{\text{prbs}} \in [90, 110]\ 1\ \text{min}^{-1} [90, 110]\ 1\ \text{min}^{-1}$ is near the vicinity of the steady state point u_{nom}. The results are obtained with respect to reactor temperature, $T(K)$. The open loop simulations for $\pm 10\%\ u_{\text{nom}}$ type of changes on the coolant flow rate at $u_{\text{prbs}} = (90 - 110)\,1\,\text{min}^{-1}$ 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}. Identification using N3ARX at u_{nom} resulted in a small offset error from the steady state at $y_{\text{True}} = 440.31$ K with $y_{\text{N3ARX}} = 440.22K$. The learned dynamics from kSINDYc (y_{kSINDYc}) and NL2SQ (y_{NL2SQ}) identification method outperformed

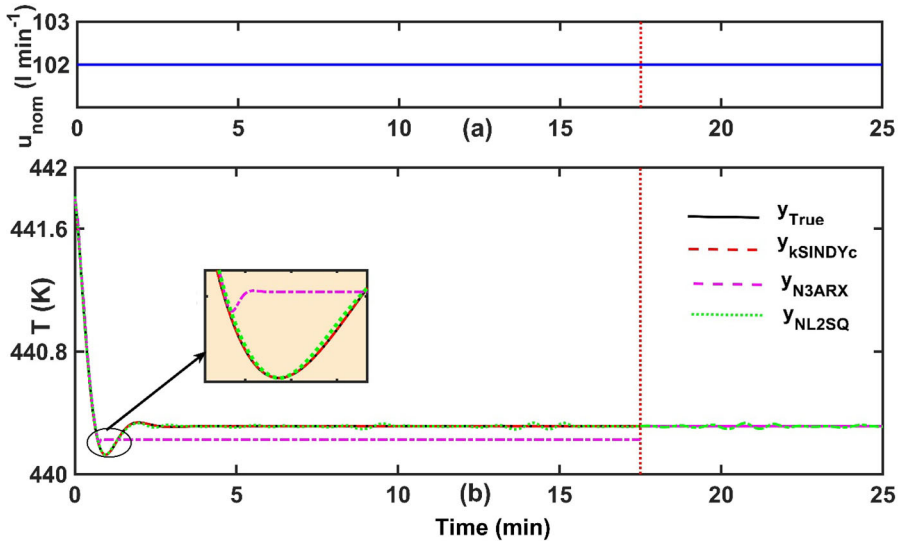

Fig. 4 True and learned model response of CSTR process for step input u_{nom}

N3ARX for the medium nonlinear CSTR process $(0.3 < \Delta_0 \leq 0.7)$ by tracking the steady state at y_{True}.

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 modeling of a heat exchanger for a high temperature milk pasteurization unit was enumerated using log mean temperature difference approach [45]. In our research, a nonlinear physical model of a fluid–fluid HE processes is detailed is adopted from [46]. 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 [46]. The steady state values and the initial states $(T_{\text{co\_nom}}, T_{\text{po\_nom}})$ are found to be $(115, 150)$ °F. The nonlinear material balance equations of the process are given in Eqs. (39, 40).

$$\dot{T}_{\text{co}} = \frac{2}{M_c}\left[F_c(T_{\text{ci}} - T_{\text{co}}) - \left(UA\Delta T_{\text{lm}}/C_{\text{pp}}\right)\right] \tag{39}$$

$$\dot{T}_{\text{po}} = \frac{2}{M_p}\left[F_p(T_{\text{pi}} - T_{\text{po}}) - \left(UA\Delta T_{\text{lm}}/C_{\text{pp}}\right)\right] \tag{40}$$

$$\Delta T_{\text{lm}} = \frac{(T_{\text{po}} - T_{\text{ci}}) - (T_{\text{pi}} - T_{\text{co}})}{\log(T_{\text{po}} - T_{\text{ci}}) - \log(T_{\text{pi}} - T_{\text{co}})} \tag{41}$$

ΔT_{lm} is the logarithmic mean temperature difference of the Heat exchanger system.

$$y_{\text{True}} = T_{\text{po}} \tag{42}$$

The process fluid temperature from the Heat Exchanger $y_{\text{True}}(t) = T_{\text{po}}(t)$ also called, outlet fluid temperature depicted in Fig. 6, reveals that HE model is highly nonlinear where T_{po} values drop drastically from 150°F to a new steady state at $T_{\text{po}} = 44.62$°F at u_{nom}. The predicted response of the fluid temperature $y_{\text{pred}}(t)$ of kSINDYc method reached the steady state surpassing other identification schemes for a highly nonlinear heat exchanger at $u_{\text{nom}} = 40$ lbm min^{-1}. The output variable T_{po} is plotted for all the three methods kSINDYc, NL2SQ and N3ARX in Fig. 7 when the excitation signal is $u_{\text{prbs}} \in [36, 44]$ lbm min^{-1}. The measured temperature y_{True} for PRBS excitation displays a steady state at $y_{\text{True}} = 44.7$°F which is relatively closer to the response at u_{nom}. The predicted response of y_{kSINDYc} and y_{NL2SQ} follows y_{True} accurately compared to y_{N3ARX} when excited at u_{prbs}. The high level of nonlinearity $\Delta_0 > 0.7$ for step and PRBS inputs of the Heat Exchanger can be noticed from Tables 5 and 6.

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 [47]. The mass balance equations representing the kinetic model of the bioreactor are

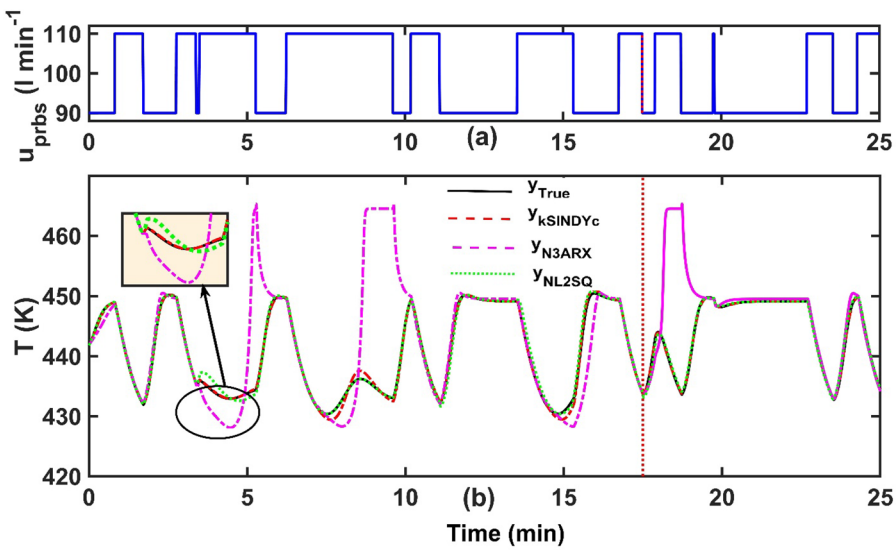

Fig. 5 True and learned model response of CSTR process for input u_{prbs}

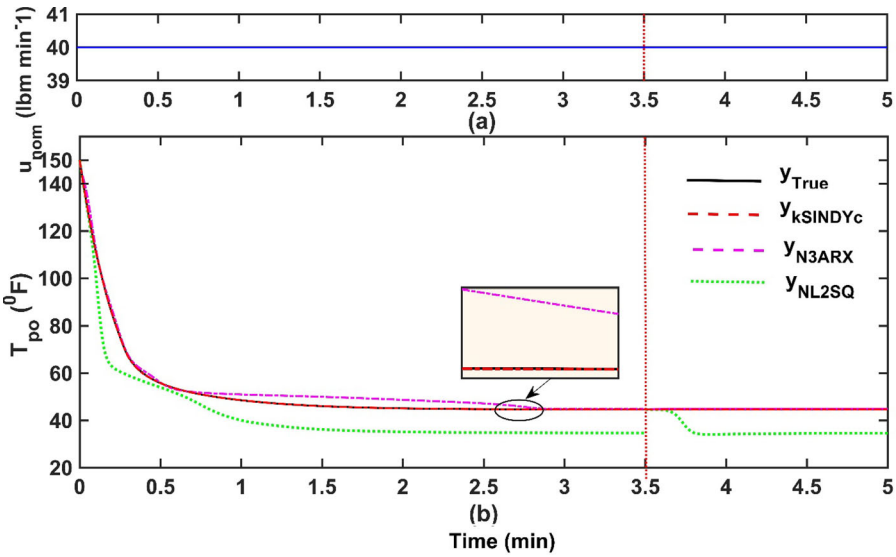

Fig. 6 True and learned model temperature responses T_{po} of heat exchanger process for step input at u_{nom}

given in Eqs. (43–45). 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 $(\mu_{\max})$, 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 $[0, 0.6]\,\text{hr}^{-1}$. The initial values and nominal (operating) points of the state variables are $(X_{\text{nom}}, S_{\text{nom}}) = (1.530, 0.174)\ \text{g l}^{-1}$. The density of microbial cells also called biomass concentration X of any microorganism grows by consuming the substrate S fed to it.

$$\dot{S} = \frac{-1}{Y_{XS}}\mu(S)X + D(S_{\text{in}} - S) \tag{43}$$

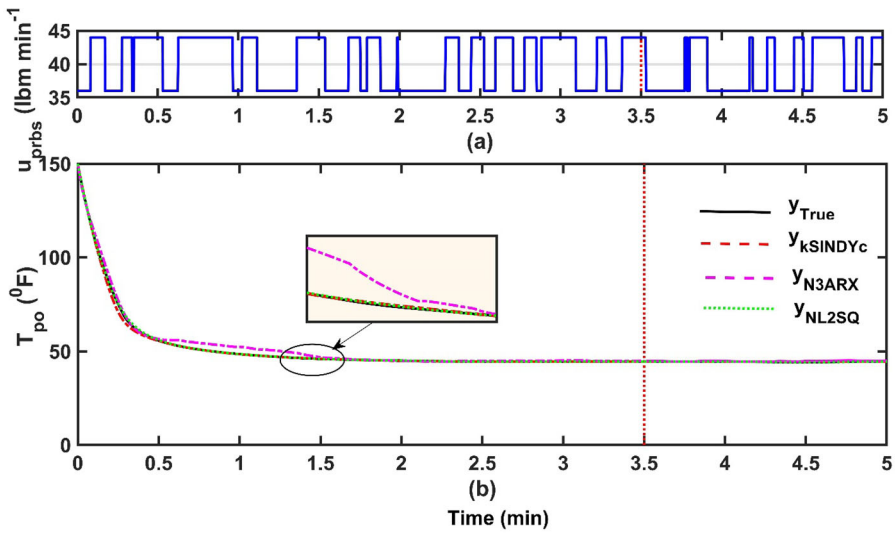

Fig. 7 True and learned model response of heat exchanger process for step input at u_{prbs}

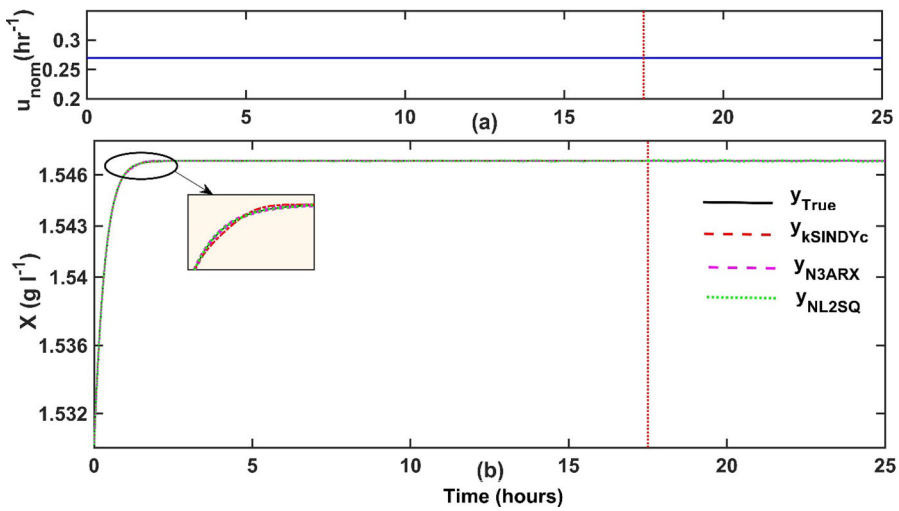

Fig. 8 True and learned model response of bioreactor process for input u_{nom}

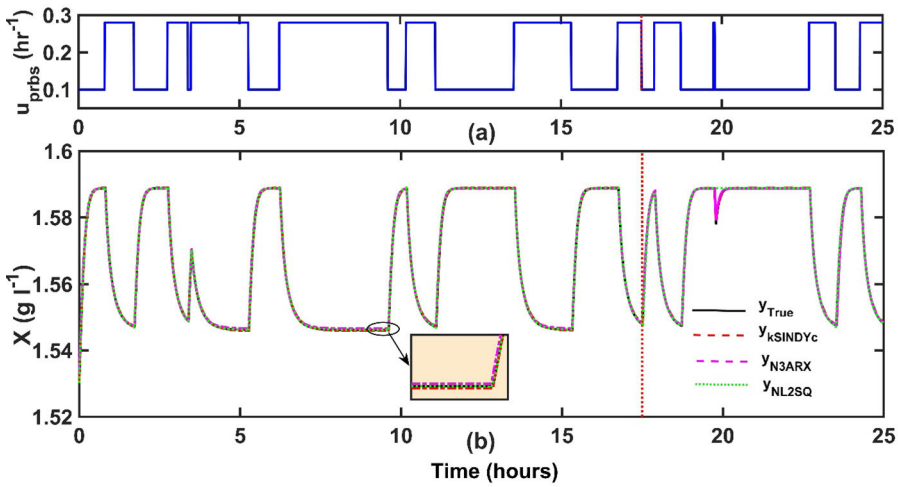

Fig. 9 True and learned model response X(g/litre) of bioreactor process for input u_{prbs}

$$\dot{X} = (\mu(S) - D)X \tag{44}$$

where a Haldane type of specific growth rate is given by

$$\mu(S) = \mu_{\max} \frac{S}{(S + K_m + K_1 S^2)} \tag{45}$$

$$y_{\text{True}} = X \tag{46}$$

The nonlinearity of the bioreactor varies w.r.t the specific growth rate $\mu(S)$, the type of excitation given (u), initial states of (S, X) and the operating region of dilution rate $(u = D)$.Therefore a bioreactor can be contemplated as a very sensitive nonlinear system, subjected to the above actors. The Bioreactor is highly sensitive, whose nonlinearity may switch from mild to medium based on the excitation input $u = D$.The input constraints of the Bioreactor lies in the range $u = (0.1 - 0.3)\text{hr}^{-1}$. For dilution rate $D < 0.3$, the Bioreactor system remains in the mild nonlinear category. The step input at $D_c = 0.3$ is called critical dilution rate where the biomass concentration (X) disappears, and the system becomes unstable [46]. In this example, we restrict our analysis with the operating point P at $D = 0.27$, as the microbial cell growth gets affected beyond D_c.

Figure 8 represents the response $y_{\text{True}} = X(\text{g l}^{-1})$ at steady state, when the dilution rate is operated at $u_{\text{nom}} = 0.27\,\text{hr}^{-1}$. The bioreactor is operated at the stable operating point u_{nom}. It can be observed that the predicted response generated by NL2SQ,N3ARX and

kSINDYc methods, show a pattern that follows the nonlinear dynamics of the bioreactor very accurately for $y_{\text{pred}} = \{y_{\text{kSINDYc}}, y_{\text{NL2SQ}}, y_{\text{N3ARX}}\}$ with the steady state of biomass concentration (X) at $u_{\text{nom}} = 0.27\ \text{hr}^{-1}$ to settle at $y_{\text{True}}(t) = 1.547\ \text{g l}^{-1}$. The response of X (g/litre) due to PRBS input which has a feed flow $u_{\text{prbs}} = \pm 10\%\ u_{\text{nom}}$ is portrayed in Fig. 9. It can be noticed that the three methods of identification $(y_{\text{kSINDYc}}, y_{\text{NL2SQ}}, y_{\text{N3ARX}})$ showed excellent tracking of the biomass-concentration with very sharp variations when excited with PRBS signal also. However, the evaluation criteria RMSE, is very small for y_{N3ARX} than kSINDYc and NL2SQ identification both for u_{nom} and u_{prbs} excitations. The N3ARX identification is also computationally faster compared to the other two methods for the bioreactor, which exhibits mild nonlineatity of $\Delta_0 < 0.3$ when operated below the critical dilution rate $u < 0.3\ \text{hr}^{-1}$.

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 $u_{\text{nom}} = 2.704\ \text{kmol min}^{-1}$ is taken for the study from [47]. The distillate composition x_D (mole fraction) 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 uniform 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. (47–51).

Condenser Stage:

$$\dot{x}_1 = \frac{1}{M_D}[V_R(y_2 - y_1)] \tag{47}$$

Rectifying section above feed tray (i=2 to NF-1)

$$\dot{x}_i = \frac{1}{M_T}[L_R x_{i-1} + V_R y_{i+1} - L_R x_i - V_R y_i] \tag{48}$$

Feed stage (NF)

$$\dot{x}_{NF} = \frac{1}{M_T}[L_R x_{NF-1} + V_R y_{NF+1} + F z_F - L_S x_{NF} - V_R y_{NF}] \tag{49}$$

Rectifying section below feed tray (i=NF+1 to NS-1)

$$\dot{x}_i = \frac{1}{M_T}[L_S x_{i-1} + V_S y_{i+1} - L_S x_i - V_S y_i] \tag{50}$$

Re-boiler Stage

$$\dot{x}_{NS} = \frac{1}{M_B}[L_S x_{NS-1} - B x_{NS} - V_S y_{NS}] \tag{51}$$

$$y_{\text{True}} = x_D \tag{52}$$

The DC model when operated in a wider operating region instead of a fixed input at u_{nom} imparts a

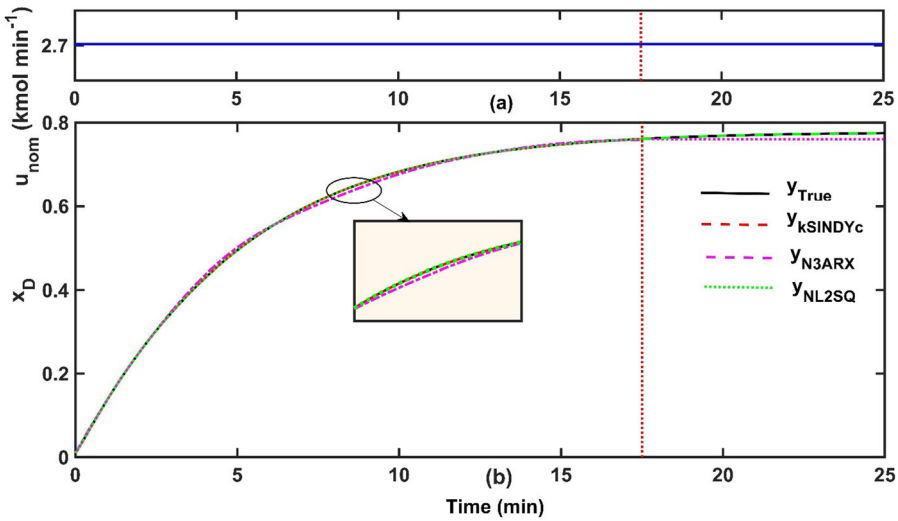

Fig. 10 True and learned model response x_D of distillation column process for step input at u_{nom}

Table 3 Key term-based SINDYc (kSINDYc)

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

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 $(x_{\text{Dnom}}, x_{\text{Bnom}}) = (0.775, 0.2225)$. The time response of x_D in Fig. 10 is noticed when the DC is operated at reflux rate $u_{\text{nom}} = 2.704\,\text{kmol min}^{-1}$.Fig. 11 flaunts the time response of x_{D} when excited in the PRBS range $u_{\text{prbs}} \in (2.43, 3)\,\text{kmol min}^{-1}$. Except for N3ARX method, the learned models $y_{\text{pred}}(t)$ from kSINDYc and NL2SQ approaches bear a very close follow up to the true output of the distillate composition $y_{\text{True}} = x_D = 0.775$. The DC exhibits medium nonlinearity using CANM method with Δ_0 in the range $0.3 < \Delta_0 \leq 0.7$ when excited by both u_{nom} and u_{prbs}.

4.1 Selection of key nonlinear terms k_{nl}

The key nonlinear term k_{nl} in every system taken for the study is the predominant nonlinear function that determines the nonlinear dynamics of the physical system. In each Example, the dynamic equations have some mathematical terms that will have the most significant influence on the system's behavior. This could be found out by careful examination of the Ordinary Differential Equation (ODE) of every Example from (Eqs. 32–52). Nonlinear terms typically include products, powers, trigonometric functions, exponential functions, and other nonlinear operations. In the examples considered for the present study, the significant terms in each ODE equation that involve nonlinear functions of the dependent variable(s) or their derivatives are carefully observed based on prior knowledge of the system. These terms often have larger coefficients or play a central role in the dynamics. Alternatively, numerical simulation tools can also be adopted (e.g., differential equation solvers) to simulate the dynamic response of the nonlinear system to different inputs and initial conditions. In such cases, the transient behavior of all examples has to be carefully analyzed for any observation of any nonlinear effects, oscillations, or stability issues. Moreover, interest readers may refer Global Sensitivity Analysis (GSA) methods reported in [50] for sorting out k_{nl} of complex nonlinear systems. Applying GSA will definitely assess how changes in various parameters (e.g., reaction rate constants, flow rates, temperature) affect the system's response. Furthermore, this method will also reveal the predominant terms that have the most significant impact on the system's nonlinear behavior. Table 3 shows the kSINDYc library terms obtained for all the five examples elaborated in Sect. 4. kSINDYc is an empirical data driven method, which needs only the measured input and output data to identify any nonlinear system. Even if kSINDYc is data driven identification, using the sparse regression and key nonlinearity terms (k_{nl}) in its library function, the governing equation can be found out along with the parameters.

The parametric coefficients of each example can be found out after solving Algorithm A1 in iterative steps. Table 3 summarizes the basis terms to be carefully placed inside the candidate library for all the five case studies. The nonlinear function of each process is decided by the rudimentary 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. The active terms in the dynamics from the library $\Theta(X, U)$ are identified using the sparse regression algorithm defined in Eq. (17). It is evident that compared to SINDYc identification, the proposed kSINDYc scheme requires only a smaller number of relevant key terms in the candidate library, thereby

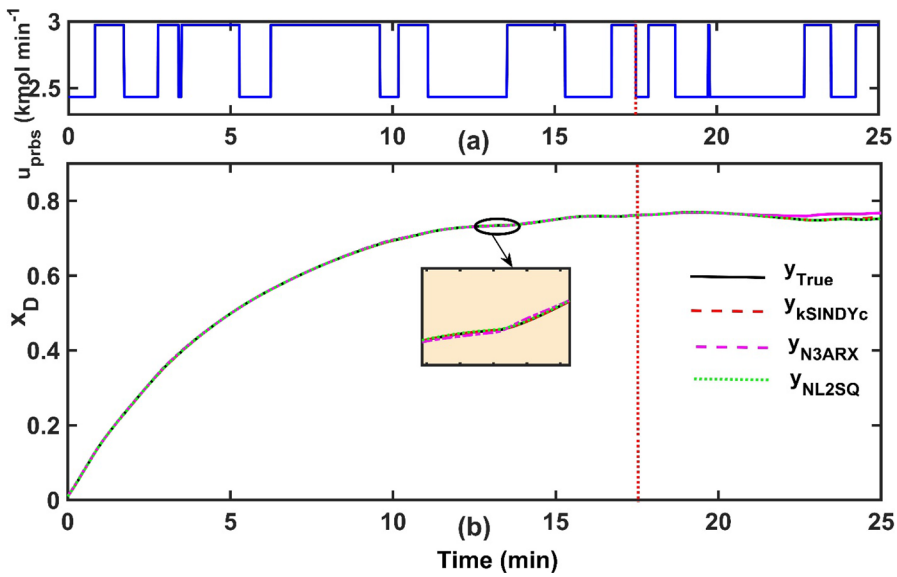

Fig. 11 True and learned model response x_D of distillation column process for step input at u_{prbs}

Table 4 Comparison of RMSE for System identification using kSINDYc, N3ARX and NL2SQ methods

Process	RMSE for u_{nom}			RMSE for u_{prbs}		
	kSINDYc	N3ARX	NL2SQ	kSINDYc	N3ARX	NL2SQ
Three Tank	0.0050	0.0569	1.632e−6	0.0282	0.0043	0.0070
CSTR	1.2954e−6	0.0160	2.493e−4	0.0184	3.2382	0.0521
HE	4.311e−4	0.2282	4.2615	0.0322	0.3154	0.0403
Bioreactor	0.00832	0.0015	0.00561	0.0102	0.0049	0.0065
DC	3.314e−4	0.0098	5.212e−6	6.273e−3	0.0054	2.641e−5

reducing the number of parameters required to identify the nonlinear system. Using the candidate library terms of kSINDYc in Table 3, every nonlinear process can be identified using higher-order polynomial equations. 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. It is always interesting to observe how the dynamics of nonlinear system changes in response to different excitations/test signals.

4.2 Performance evaluation

In our proposed framework we intend to validate the performance of kSINDYc, NL2SQ and N3ARX for all examples subjected to step and PRBS excitation. By looking at Table 4, a major conclusion can be brought over in the concept of data driven modeling. A careful observation of the y_{pred} at u_{nom} and u_{prbs} in all the examples signify that N3ARX does not mimic the system dynamics accurately for CSTR and Heat exchanger process. This inspection ensures that N3ARX identification can be adopted for systems with mild nonlinearity. However, it predicts the PRBS response of all the examples with an acceptable RMSE. Furthermore, the identified model using kSINDYc and NL2SQ under u_{nom} and u_{prbs}, predicted the nonlinear dynamics of all the examples with high accuracy. The predicted model accuracy of each identification method (kSINDYc, NL2SQ, N3ARX) are validated using the performance Index RMSE. By inspecting the level of nonlinearity using CANM, a user can flexibly choose the appropriate identification method among the three methods investigated in the study. To emphasize this point, in all the five case studies, the RMSE at u_{nom} and u_{prbs} are summarized in Table 3 and Δ_0 computed using CANM are graphically related

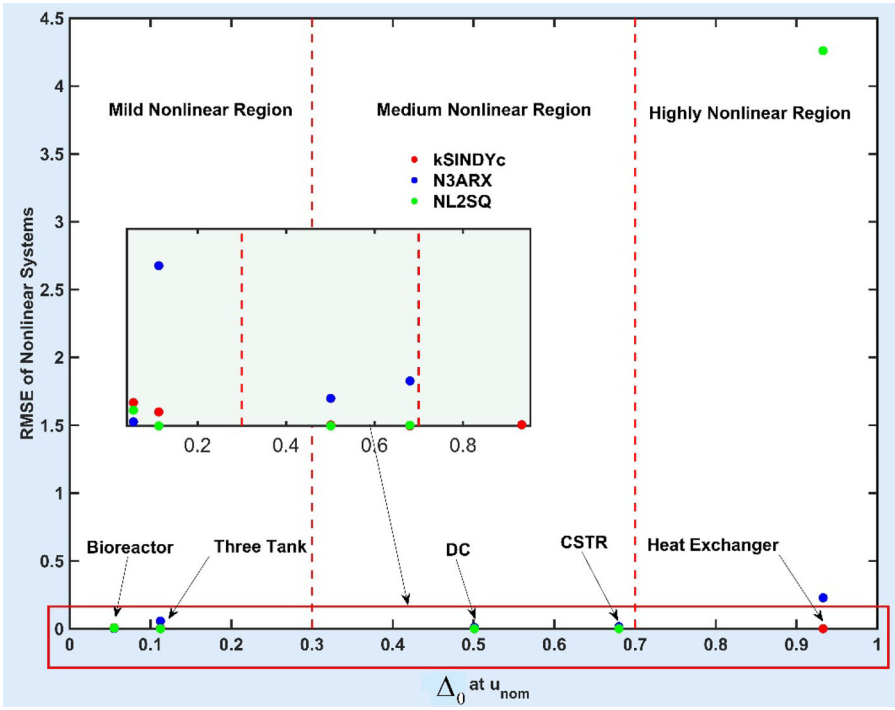

Fig. 12 RMSE of learned models of all systems with its metric Δ_0 at input u_{nom}

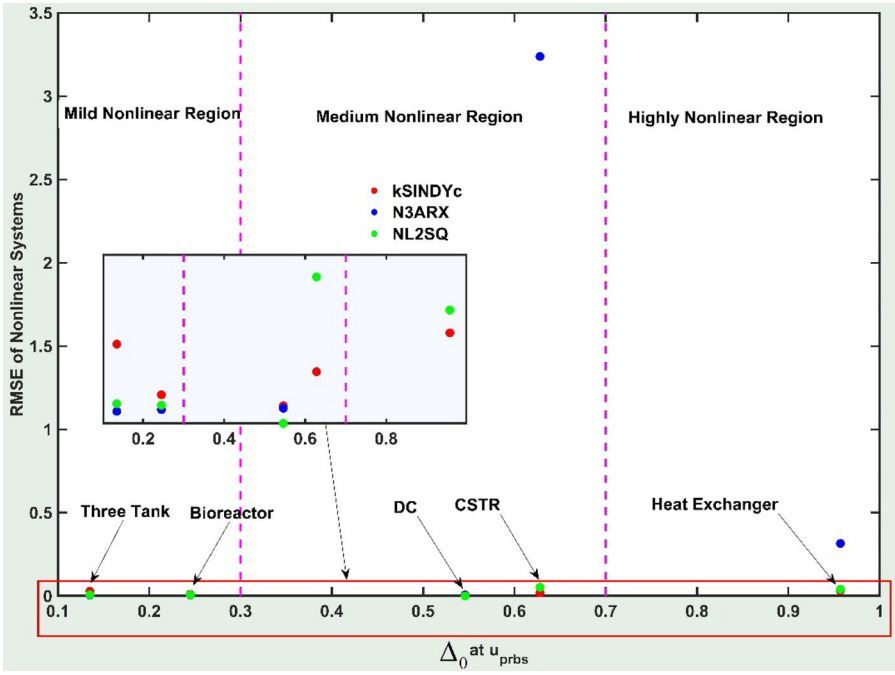

Fig. 13 RMSE of learned models of all systems with its metric Δ_0 at input u_{prbs}

using scatter plot in Figs. 12 and 13.The quantitative analysis in Figs. 12 and 13 are very important graphical representations that relates the nonlinearity of each system with the three nonlinear system identification methods in terms of performance evaluation criteria.

Table 5 System identification from Δ_0 for excitation input u_{nom}

Process	u_{nom}	Operating point	Δ_0	Class of NL	Best choice for system identification
Three Tank	$0.5e^{-5}m^3s^{-1}$	(0.914, 0.80, 0.692)	0.1123	Mild	NL2SQ
CSTR	102 l min^{-1}	(0.0823, 441.81)	0.68	Medium	kSINDYc
HE	40 lbm min^{-1}	(47.018, 44.58)	0.9328	High	kSINDYc
Bioreactor	0.27 hr^{-1}	(1.547, 0.138)	0.055	Mild	N3ARX
DC	2.704 kmol/min	(0.775, 0.225)	0.501	Medium	NL2SQ

Table 6 System identification from Δ_0 for excitation input u_{prbs}

Process	u_{prbs}	Operating point	Δ_0	Class of NL	Best choice for system identification
Three Tank	$(0-1)e^{-5}m^3\ s^{-1}$	(0.914, 0.80, 0.692)	0.135	Mild	N3ARX
CSTR	$(90-110)$ l min^{-1}	(0.0823, 441.81)	0.628	Medium	kSINDYc
HE	$(36-44)$ lbm min^{-1}	(47.018, 44.58)	0.957	High	kSINDYc
Bioreactor	$(0.1-0.3)$ hr^{-1}	(1.547, 0.138)	0.245	Mild	N3ARX
DC	$(2.435-2.972)$ kmol/min	(0.775, 0.225)	0.546	Medium	NL2SQ

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_{\text{True}}(t) - y_{\text{pred}}(t)\right)^2} \tag{53}$$

The RMSE for N samples are found from Eq. (53) for all the dynamic physical systems described in this Sect. As seen from Figs.12 and 13, the crucial factor in determining the choice of system identification rests on the minimum RMSE between the actual and predicted dynamics of the identified model among all the three identified models $\left\{y_{\text{pred}} = y_{\text{kSINDYc}}, y_{\text{NL2SQ}}, y_{\text{N3ARX}}\right\}$. A graphical comparison is made between the RMSE of all the nonlinear system identification methods for u_{nom} is given in Fig. 12. The scatter plot in Fig. 13 maps the RMSE with Δ_0 for nonlinear systems excited at u_{prbs}. 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 accurate choice of system identification.

CANM is an operating point dependent nonlinear metric. All the five examples examined in this manuscript, are continuous processes working at a stable steady state operating point (x_{as}, x_{bs}). The effect of nonlinearity Δ_0 around the vicinity of the fixed point (x_{as}, x_{bs}) are carefully investigated in this section. The usage of NL2SQ method is preferred only for mild systems like three tank process and Bioreactor. Diversely, kSINDYc is the best opted non-parametric model for highly nonlinear processes like Heat Exchanger. A low value of RMSE ensures that y_{pred} is very close to $y_{\text{True}}(t)$ capturing the underlying patterns and the nonlinear dynamics with high precision. In an application like DC, RMSE is very low in the order of 10^{-4} or lesser in the predicted models kSINDYc and NL2SQ. The RMSE precision of all the learned models in the Bioreactor implies that even though the predicted data has high fidelity, the model is highly sensitive to noise that leads to overfitting of $y_{\text{True}}(t)$. By using this combined framework, any user can find out the conducive system identification method based on the nonlinearity in the desired operating region. CANM is an operating point dependent nonlinear metric.

The physical quantities of each process addressed in Tables 5 and 6, have different orders of magnitude. To sustain a uniform scale in measuring the nonlinearity, the time period t, input u and output variable $y_{\text{True}}(t)$ of all nonlinear physical process are normalized between 0 and 1, and thereafter the CANM method Δ_0 is intended from Eq. (2). The performance index (RMSE) of all the system identification methods are correlated along with the degree of nonlinearity Δ_0 of each nonlinear system and the outcomes are enumerated. To verify the impact of excitation inputs on Δ_0, u_{nom} and u_{prbs} are applied to all the examples.

Even though the Δ_0 values were different for u_{nom} and u_{prbs} signals operated at operating point P, the class of nonlinearity (mild, medium, high) will remained the same. This perception reveals an important remark that irrespective of the excitation inputs, the class of nonlinearity will not change but remains the same using CANM. Tables 5 and 6 will provide the inference made on the choice of the appropriate system identification method, on the basis of Δ_0 and the RMSE of every Example. The special features of kSINDYc method lies in its concrete algorithm to develop an exact model even with limited number of samples, without involving in under fitting or over fitting of highly nonlinear systems. kSINDYc proves to be a better choice for system identification in high nonlinear process whereas the mild nonlinear systems can follow N3ARX method and medium nonlinear units can adopt NL2SQ to learn the process dynamics as observed from Tables 5 and 6. The effectiveness of N3ARX approach trails behind kSINDYc and NL2SQ approaches in terms of RMSE and execution time. From the detailed system identification analysis carried out on all case studies, some assertive conclusive remarks are presented below over the selection of appropriate identification method based on degree 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}. The prime hindrance of N3ARX method compared to NL2SQ and kSINDYc schemes 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. However, this approach crashes to identify the process models with less measured I-O data.

Remark 3 kSINDYc is computationally attractive, requires less data, assumes a few numbers 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 highly nonlinear systems.

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. The use of the proposed kSINDYc identification necessitates the knowledge of the system dynamics to acquire the key term k_{nl} which appear to be a hindrance in nonlinear systems whose governing equations are completely unknown and unpredictable. In such cases, SINDy algorithm performs the same task without involving the system dynamics in the form of ODEs.

5 Conclusion

Selection of an appropriate identification method is very decisive for any complex nonlinear system. The proposed framework 'System identification in coherence with nonlinearity measure' indisputably accomplishes this task by mathematically relating the nonlinearity level with the applicable identification method. An integrated framework comprising three identification methods (kSINDYc, N3ARX and NL2SQ) and a nonlinearity measure called CANM is devised in this research study. In particular, the proposed data driven kSINDYc scheme, identifies nonlinear processes under the sparse dimensional space using key nonlinear terms in its candidate library. A notable development is made in the 'kSINDYc candidate library', by introducing the *'key nonlinear terms'* from the plant dynamics, along with the polynomial terms. The kSINDYc identification uses a sparcification knob λ set between 0 and 10, to identify nonlinear dynamics of five physical systems with divergent nonlinear strengths. The method surmounts NL2SQ and N3ARX by meticulously adopting the process dynamics of highly nonlinear processes. Additionally, this research comprehends the nonlinear metric CANM that targets to find out the degree of nonlinearity between 0 and 1 and subclassifies the five examples to fit in mild, medium or highly nonlinear category. This article exemplifies a contemporary quantitative analysis that correlates the nonlinear metric with the system identification schemes. The proposed framework is tested for five nonlinear systems with diverse nonlinear strengths. This article differs from the existing literature by providing a configuration for suitable system

identification from the three methods based on the computed Δ_0 using CANM metric.

6 Scope

The nonlinear systems considered in this study, are classified as mild, medium and highly nonlinear using CANM method. However, 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. Extending the proposed framework to more complex MIMO process structures should be carried out without losing the dynamic behavior of the system. The choice of nonlinear control schemes based on the computation of Δ_0 is another decisive study that has to be devised in the future research. 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.

Acknowledgements The first author acknowledges the financial support 'Anna Centenary Research Fellowship' (ACRF) funded by Centre for Research, Anna University, Chennai, India to carry-out this study.

Funding The authors have not disclosed any funding.

Data availability The data analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest. 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), 11074–11083 (2014). https://doi.org/10.1021/ie500175r
7. Zhang, X., Ding, F., Alsaadi, F.E., Hayat, T.: Recursive parameter identification of the dynamical models for bilinear state space systems. Nonlinear Dyn. **89**(4), 2415–2429 (2017). 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. (2021). https://doi.org/10.1007/s11071-021-06454-8
9. Mani, A.K., Narayanan, M.D., Sen, M.: Parametric identification of fractional-order nonlinear systems. Nonlinear Dyn. **93**(2), 945–960 (2018). 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. Rafiei, H., Akbarzadeh-T, M.R.: Reliable fuzzy neural networks for systems identification and control. IEEE Trans. Fuzzy Syst. (2022). https://doi.org/10.1109/TFUZZ.2022.3222036

18. Du, J., Johansen, T.A.: Control-relevant nonlinearity measure and integrated multi-model control. J. Process. Control. **57**, 127–139 (2017). https://doi.org/10.1016/j.jprocont.2017.07.001
19. Jiang, M., Zhang, W., Lu, Q.: A nonlinearity measure-based damage location method for beam-like structures. Measurement **146**, 571–581 (2019). https://doi.org/10.1016/j.measurement.2019.06.049
20. Jiang, M., Wang, D., Kuang, Y., Mo, X.: A bicoherence-based nonlinearity measurement method for identifying the location of breathing cracks in blades. Int. J. Non-Linear Mech. **135**, 103751 (2021). https://doi.org/10.1016/j.ijnonlinmec.2021.103751
21. Zhao, N., Shi, P., Xing, W., Chambers, J.: Observer-based event-triggered approach for stochastic networked control systems under denial of service attacks. IEEE Trans. Control Netw. Syst. **8**(1), 158–167 (2020). https://doi.org/10.1109/TCNS.2020.3035760
22. Liu, Y., Zhu, Q.: Event-triggered adaptive neural network control for stochastic nonlinear systems with state constraints and time-varying delays. IEEE Trans. Neural Netw. Learn. Syst. **34**(4), 1932–1944 (2021). https://doi.org/10.1109/TNNLS.2021.3105681
23. 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), 051002 (2021). https://doi.org/10.1115/1.4050553
24. Brunton, S.L., Proctor, J.L., Kutz, J.N.: Sparse identification of nonlinear dynamics with control (SINDYc). IFAC-PapersOnLine **49**(18), 710–715 (2016). https://doi.org/10.1016/j.ifacol.2016.10.249
25. Du, J., 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
26. Hahn, J., 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
27. 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
28. 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**(22), 20180335 (2018)
29. 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
30. 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
31. Champion, K., Zheng, P., Aravkin, A.Y., Brunton, S.L., Kutz, J.N.: A unified sparse optimization framework to learn parsimonious physics-informed models from data. IEEE Access. **8**, 169259–169271 (2020). https://doi.org/10.1109/ACCESS.2020.3023625
32. 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
33. 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
34. 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
35. Transtrum, M.K., Sethna, J.P.: Improvements to the Levenberg-Marquardt algorithm for nonlinear least-squares minimization (2012) arXiv preprint
36. 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
37. Liu, H., Song, X.: Nonlinear system identification based on NARX network. In 2015 10th Asian Control Conference (ASCC) 1-6. IEEE (2015). https://doi.org/10.1109/ASCC.2015.7244449
38. 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
39. Taghavifar, H.: Neural network autoregressive with exogenous input assisted multi-constraint nonlinear predictive control of autonomous vehicles. IEEE Trans. Veh. Technol. **68**(7), 6293–6304 (2019). https://doi.org/10.1109/TVT.2019.2914027
40. 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
41. 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), 3324–3334 (2013). https://doi.org/10.1016/j.asoc.2013.02.007
42. Saki, S., Fatehi, A.: Neural network identification in nonlinear model predictive control for frequent and infrequent operating points using nonlinearity measure. ISA Trans. **97**, 216–229 (2020). https://doi.org/10.1016/j.isatra.2019.08.001
43. Bistak, P., Huba, M.: Three-tank laboratory for input saturation control based on matlab. IFAC-Papers Online **49**(6), 207–212 (2016). https://doi.org/10.1016/j.ifacol.2016.07.178
44. 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
45. Indumathy, M., Sobana, S., Panda, R.C.: Modeling 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

46. 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
47. Bequette, B.W.: Process control: modeling, design, and simulation. Prentice Hall Professional, New Jersey (2003)
48. Sobana, S., Panda, R.C.: Control of pH process using double-control scheme. Nonlinear Dyn. **67**, 2266–2277 (2012). https://doi.org/10.1007/s11071-011-0144-x
49. Atanu, P., Panda, R.C.: Adaptive nonlinear model-based control scheme implemented on the nonlinear processes. Nonlinear Dyn. **91**, 2735–2753 (2018). https://doi.org/10.1007/s11071-017-4043-7
50. 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