Contents lists available at ScienceDirect

Chemical Engineering Research and Design

journal homepage: www.elsevier.com/locate/cherd

Machine learning-based adaptive model identification of systems: Application to a chemical process

Bhavana Bhadriraju[a,b], Abhinav Narasingam[a,b], Joseph Sang-Il Kwon[a,b,*]

[a] Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX 77845, USA
[b] Texas A&M Energy Institute, Texas A&M University, College Station, TX 77845, USA

ARTICLE INFO

Article history:
Received 3 August 2019
Received in revised form 2 September 2019
Accepted 4 September 2019
Available online 4 October 2019

Keywords:
Nonlinear dynamics
Adaptive identification
System identification
Sparse regression
Feature selection
Stepwise regression

ABSTRACT

Many of the existing offline system identification methods cannot completely comprehend the dynamics of an evolving complex process without relying on impractically large data sets. As a solution to this, a systematic procedure capable of identifying and predicting the nonlinear dynamics on the fly promises to provide a useful representation of the process model. Motivated by this, an adaptive model identification framework that relies on the methods of sparse regression and feature selection is presented in this work. The proposed method is a three-step procedure: (1) identifying potential functions from a candidate library using recently developed Sparse Identification of Nonlinear Dynamics (SINDy), (2) updating coefficients of the identified model using ordinary least-squares regression, (3) selecting the most important features using stepwise regression. The proposed algorithm is implemented as follows. Initially, a baseline model is identified offline using SINDy, and as a new data becomes available, the subsequent online steps are triggered based on a pre-specified tolerance to further update the model. Such an adaptive identification scheme facilitates in perceiving the model structure using a less amount of data than its offline counterpart, SINDy. To highlight its significance, the dynamics of a continuous stirred tank reactor is identified using the proposed adaptive method and is compared with a model identified using SINDy alone.

1. Introduction

Over the past decades, growing production technologies have contributed to a rise in complex processes across a major sector of industries. When dealing with such processes, first principles based modeling may be intractable and cannot be relied upon to discover the underlying governing equations, particularly when a strict constraint is imposed on the computational requirement (Hou and Wang, 2013). This has guided several promising developments in the field of data-based system identification. The focus of system identification is to determine the structure of the model based on the input–output relations and provide an accurate future prediction (Hong et al., 2008). In order to meet these goals, several data-driven methods have been established over the years based on equation-free modeling (Cisternas et al., 2004), artificial neural networks (Hunt et al., 1992; Wu et al., 2019b,c), empirical dynamic modeling (Ye et al., 2015), nonlinear Laplacian spectral analysis (Giannakis and Majda, 2012), automated inference from dynamics (Daniels and Nemenman, 2015), and surrogate modeling (Brigham and Aquino, 2007) to name a few. Another important class of data-driven methods that

* Corresponding author at: Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX 77845, USA.
E-mail address: kwonx075@tamu.edu (J.S.-I. Kwon).
https://doi.org/10.1016/j.cherd.2019.09.009
0263-8762/Published by Elsevier B.V. on behalf of Institution of Chemical Engineers.

have been in use for quite sometime, particularly in the areas of process control, is subspace identification. This includes methods like numerical algorithms for subspace state-space identification (N4SID) (Van Overschee and De Moor, 1994), multivariable output-error-state-space (MOESP) (Verhaegen and Dewilde, 1992), and canonical variate analysis (Larimore, 1990) which are competent in identifying simple state-space models for multivariable dynamical systems based on measured input-output data (Viberg, 1994). Owing to an easy accessibility to a vast amount of data and advancements in machine learning algorithms in recent times, these data-driven approaches are becoming more feasible and prominent.

Despite the proven success of the aforementioned "black-box" approaches in many applications, there has been an increasing shift in integrating data-driven methods with physics laws, especially in the case of dynamical systems. This is driven by a limitation of black-box models to extrapolate the dynamics to the entire state-space, beyond where they were sampled and constructed. Besides, it is necessary sometimes to develop a complete understanding of the physical mechanisms that govern the dynamical system of interest. Within this context, genetic programming-based symbolic regression has contributed in determining the governing dynamics from data, along with providing an actual sense of the process (Bongard and Lipson, 2007; Schmidt and Lipson, 2009). But in the case of large scale systems, symbolic regression can be prohibitively expensive and is prone to over-fitting. Some of the researchers have addressed these issues by developing system identification methods using sparse regression and compressive sensing. These techniques are based on the fact that only a few nonlinear terms are sufficient in determining the governing dynamics of a complex process. One such method that has recently received much attention is Sparse Identification of Nonlinear Dynamics (SINDy) algorithm developed by Brunton et al. (2016b). It has been extensively used for data-driven discovery of underlying dynamics by constraining the model structure based on a priori knowledge such as symmetries and conservation laws. The significance of SINDy has been widely researched in various fields. Some of the applications include rapid model recovery from abrupt system changes (Quade et al., 2018), simultaneous identification of both micro-scale and macro-scale dynamics (Champion et al., 2019), sparse learning of reaction kinetics (Hoffmann et al., 2019), model-predictive control (Kaiser et al., 2018), reduced-order modeling of high-fidelity systems (Narasingam and Kwon, 2018; Loiseau et al., 2018), understanding rational function nonlinearities (Mangan et al., 2016) and parameterized dynamics (Li et al., 2018), discovering partial differential equations (Rudy et al., 2017; Schaeffer, 2017), ranking the models based on Akaike Information Criterion (AIC) (Mangan et al., 2017), Koopman operator based control (Kaiser et al., 2017; Brunton et al., 2016a), and developing Galerkin regression models for fluid-flow (Loiseau and Brunton, 2018). Due to the ease of implementation and the ability to incorporate any known process knowledge, SINDy could be effectively applied for a large number of nonlinear dynamical systems. Additionally, several theoretical developments have also been established to show that the SINDy algorithm rapidly converges to a local minimizer under specific conditions (Zhang and Schaeffer, 2018).

Though SINDy proved to be successful in inferring the dynamics of various systems of interest, it holds the limitation of uncovering all of the underlying subtle dynamics for a complex process when only a small amount of data is available. Especially for the processes exhibiting complex nonlinear characteristics, the type regularly encountered in chemical sector, numerous samples may be required for obtaining an accurate model in the absence of enhanced sampling strategies. However, collecting such a large amount of measured data may be expensive and also, handling such massive data is computationally demanding. With this in mind, in this work, an adaptive identification method is proposed for identifying complex process dynamics with limited use of data. For nonlinear processes whose dynamics are time-varying and poorly understood, adaptive identification is a favorable approach. Most importantly, in the case of systems with parameter uncertainties and evolving dynamics, there is a need for adaptive modeling as a new data becomes available (Varshney et al., 2009; Åström and Eykhoff, 1971; Seborg et al., 1986). This is particularly useful because re-training the model may not be fast enough to cope with the real-time demands. Moreover, offline trained models can be significantly improved when they are simply updated using a new data. Therefore, for real-time applications, adaptive model identification helps in handling any plant-model mismatch that may occur during process operation. Recently, several methods contributing to the data-driven online model identification have been developed; in Alanqar et al. (2017), the authors proposed an error-triggered online identification approach and in Wu et al. (2019a), a combination of event-triggered and error-triggered online identification mechanism based on recurrent neural networks is discussed. Although these approaches are shown to be useful for model predictive control of real-time processes, they do not provide an interpretable model.

Apart from using a small amount of data, it is important to identify a model which is free of redundant and irrelevant variables (Jović et al., 2015). Specifically, the features that do not contribute to the optimal model performance are considered irrelevant and the ones that are weakly relevant but can be replaced by other distinctive features are redundant. The presence of such features reduces prediction speed and accuracy. To tackle this issue, one can use feature selection techniques to eliminate the unwanted variables which do not significantly contribute to the process dynamics. These methods usually result in improved model prediction accuracy and reduced computational burden. Within the class of feature selection methods, numerous techniques are available and can be broadly grouped as filter, wrapper and embedded methods (Chandrashekar and Sahin, 2014; Guyon and Elisseeff, 2003). Filter methods rank the features based on their correlation with the output without using any machine learning algorithm. Though these methods are computationally less expensive and evaluate the importance of each variable individually, they may fail in providing the best subset of variables as they do not actually train the model. On the other hand, wrapper and embedded methods select the best subset of features based on predictor performance. While wrapper methods use the combination of search strategies and modeling algorithm, embedded methods like LASSO (Tibshirani, 1996) and RIDGE (Hoerl and Kennard, 1970) have feature selection integrated within their algorithm. In this work, a wrapper-based stepwise regression is used as a part of the proposed framework (Thompson, 1978; Billings and Voon, 1986; Tsai, 2009).

Taking the above mentioned considerations into account, this work proposes an adaptive sparse identification method that identifies the emerging process dynamics of a complex system through a sequence of steps. First, a sparse model

based on SINDy is identified offline using the initial data. In the next step, when the previous model fails to predict accurately, the coefficients of the identified functions are updated using ordinary least-squares regression. Finally, the identified model is updated by retaining only the essential features via stepwise feature selection. The main advantage of this sequential approach is that it requires a less amount of data for identifying a complex nonlinear dynamical system compared to SINDy.

The outline of this paper is summarized as follows: In Section 2, a detailed description of the proposed methodology is presented, and in Section 3, application of the proposed method in identifying a highly nonlinear Continuous Stirred Tank Reactor (CSTR) model is described. In Section 4, numerical simulations carried out to identify the CSTR dynamics using the proposed method are discussed. In the following subsections, the performance of the model identified by the proposed algorithm is analyzed, validated and then compared with the model identified by SINDy offline. A few conclusions are presented in the final section.

2. Proposed methodology

In this section, the proposed adaptive data-based model identification method is detailed. The governing process dynamics of the system under study is represented as

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

where the vector $\mathbf{x}(t)$ denotes the states of the system at time t, $\mathbf{u}(t)$ is the vector containing the inputs applied to the system at time t, and $\mathbf{f}(\mathbf{x}, \mathbf{u})$ represents the governing equations describing the process dynamics. In some cases when the underlying dynamics are unknown and cannot be determined using physics laws, the function $\mathbf{f}$ has to be identified from measurement data. To achieve this, an adaptive sparse identification method is proposed in this work. The method is executed according to the following three steps:

1. **Sparse model identification**: With the initial data available, a parsimonious model is obtained offline from a large set of candidate functions using SINDy.
2. **Re-estimation of regression coefficients**: As a new data becomes available, the coefficients of the identified functions are updated by performing ordinary least-squares regression.
3. **Feature selection**: Using stepwise regression, the best subset of functions is selected that represents the structure of the actual dynamics.

A flowchart representing each step of the proposed method is illustrated in Fig. 1. Instead of the conventional way of using SINDy for overall process model identification, the idea is to apply SINDy for identifying potential functions from a large library matrix. Note that, the first model is identified offline with a data available initially and is further improved online as a new data is available. At a point where the SINDy model fails, a new data is regressed onto the identified function library to update the values of the function coefficients. Furthermore, stepwise feature selection is implemented to develop a more accurate and computationally efficient model by selecting only the essential functions. As a preliminary step, it is recommended to pre-process the data for efficient regression analysis using standard techniques such as normalization or filtering depending on the nature of the data. In the following subsections, each step of the adaptive identification method is discussed in detail.

2.1. *Step 1: Sparse model identification*

This is the first step in the proposed framework. In this step, a parsimonious model is obtained from some initial data (that may be collected at different operating conditions) using SINDy offline. For example, this initial data can be obtained from process history. This paper presents a brief overview of the SINDy method and for more information, the readers can refer the original work by Brunton et al. (2016b). The SINDy algorithm is developed based on an assumption that out of all possible functions considered, only few of them govern the system dynamics. In accordance with that, a sparse regression problem is solved by balancing sparsity with accuracy. This eliminates the intractable brute-force approach of searching for the right model among all the possible models in the given function-space. For finding the function $\mathbf{f}$, time-series data of state variables and applied inputs are collected. The time series data can be obtained either from physical sensor measurements or from numerical simulation of Eq. (1). The collected m snapshots of n state variables and their corresponding inputs $u(t)$ are arranged into matrices $\mathbf{X}$ and $\mathbf{U}$ as shown below.

$$\mathbf{X} = \begin{pmatrix} x_1(t_1) & x_2(t_1) & \cdots & x_n(t_1) \\ x_1(t_2) & x_2(t_2) & \cdots & x_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(t_m) & x_2(t_m) & \cdots & x_n(t_m) \end{pmatrix}$$

$$\mathbf{U} = \begin{pmatrix} u_1(t_1) & u_2(t_1) & \cdots & u_n(t_1) \\ u_1(t_2) & u_2(t_2) & \cdots & u_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ u_1(t_m) & u_2(t_m) & \cdots & u_n(t_m) \end{pmatrix} \tag{2}$$

Next, the derivatives of state variables are either measured or numerically computed. When the time-series derivatives cannot be measured directly, they must be determined carefully for efficient working of SINDy. In general, this can be done by finite difference method. But in the presence of noise, it is suggested to use rigorous methods such as total variation regularized differentiation (Chartrand, 2011) and Knowles and Wallace variational method (Knowles and Renka, 2014). Using these computed derivatives, a matrix is constructed at different time points as follows:

$$\dot{\mathbf{X}} = \begin{pmatrix} \dot{x}_1(t_1) & \dot{x}_2(t_1) & \cdots & \dot{x}_n(t_1) \\ \dot{x}_1(t_2) & \dot{x}_2(t_2) & \cdots & \dot{x}_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \dot{x}_1(t_m) & \dot{x}_2(t_m) & \cdots & \dot{x}_n(t_m) \end{pmatrix} \tag{3}$$

After obtaining the derivatives of state variables, the collected time-series data of $\mathbf{X}$ and $\mathbf{U}$ are utilized to build a candidate function library containing all possible potential

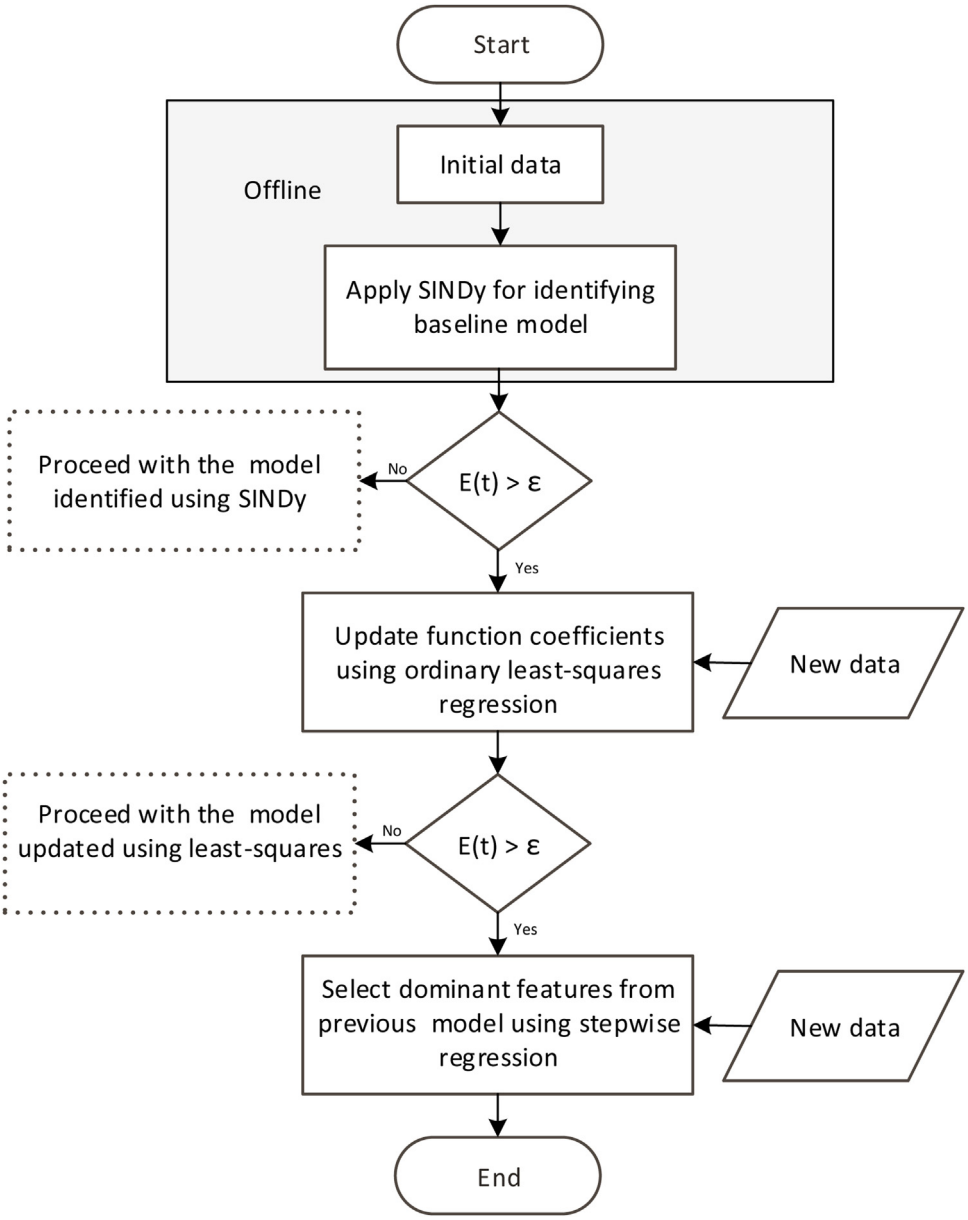


Fig. 1 – Flow chart of the proposed methodology.

functions as

$$\boldsymbol{\Theta}(\mathbf{X},\mathbf{U}) = \begin{bmatrix} \mathbf{1} & \mathbf{X} & \mathbf{X}^2 & \cdots & \mathbf{U} & \cdots & \exp(\mathbf{X}) & \sin(\mathbf{X}) & \cos(\mathbf{X}) & \cdots \end{bmatrix} \quad (4)$$

The choice of selecting the potential functions can be based on the knowledge of physics and prior information about the process. For example, since a majority of process models usually contain polynomial terms, it is useful to include them in the function library. In addition, it is recommended to populate the library with as many functions as possible like constant terms, trigonometric, and any other nonlinear functions so that the true process dynamics are well represented by the library with a higher probability. After evaluating time-derivatives of states and building the candidate function library, a regression problem is formulated as

$$\dot{\mathbf{X}} = \boldsymbol{\Theta}(\mathbf{X},\mathbf{U})\boldsymbol{\Sigma} \quad (5)$$

where the vector $\dot{\mathbf{X}}$ denotes the time-series derivatives of state variables, $\boldsymbol{\Theta}(\mathbf{X},\mathbf{U})$ is the library of possible potential functions representing the system dynamics, and $\boldsymbol{\Sigma}$ is the vector containing the function coefficients. However, solving the above problem directly using ordinary regression does not provide a parsimonious model. In order to promote sparsity in $\boldsymbol{\Sigma}$, Eq. (5) should be expressed in the form of a convex l_1-regularized regression as

$$\boldsymbol{\Sigma} = \underset{\boldsymbol{\Sigma}'}{\operatorname{argmin}} \|\dot{\mathbf{X}} - \boldsymbol{\Theta}(\mathbf{X},\mathbf{U})\boldsymbol{\Sigma}'\|_2 + \lambda\|\boldsymbol{\Sigma}'\|_1 \quad (6)$$

The above problem is solved using Sequential Thresholded Least-Squares (STLS) algorithm (Brunton et al., 2016b), which is similar to ordinary least-squares regression with an additional step of hard thresholding. The variable coefficients having values less than the thresholding parameter are rendered zeros and the regression problem is iteratively solved until convergence of parameter coefficients is attained. In this step, the parameter λ is crucial in eliminating the unwanted functions and its value can be evaluated using several hyper-

parameter tuning strategies such as grid-search (Abraham et al., 2014), random search (Bergstra and Bengio, 2012; Bergstra et al., 2011), and Bayesian Optimization (Eggensperger et al., 2013). Identifying the correct functions along with their exact coefficients requires a large number of samples. Since Eq. (6) is solved using limited data samples available initially, it is unlikely to realize an accurate model in this step. Therefore, the identified model will only be used to approximate the system until it diverges from the actual process dynamic behavior. To quantify the accuracy of the identified model, the relative error based on Frobenius norm can be used, which is calculated as

$$E(t) = \frac{\|x(t) - \hat{x}(t)\|_{fro}}{\|x(t)\|_{fro}} \tag{7}$$

where $\|.\|_{fro}$ denotes the Frobenius norm, $x(t)$ is the actual process value, and $\hat{x}(t)$ is the model predicted value. When the error evaluated between model prediction and process measurement exceeds a pre-specified tolerance, ϵ, i.e., $E(t) > \epsilon$, Step 2 becomes functional.

2.2. Step 2: Re-estimation of regression coefficients

The objective of this step is to update the coefficients of previously determined functions. As a new data becomes available, this is done using ordinary least-squares regression. Specifically, a new library matrix is constructed considering only the functions identified in Step 1, and the new data is regressed onto this function library without any thresholding. As there is no thresholding, the regression process is computationally more attractive. Again, when the model obtained in this step performs poorly, Step 3 of the method is initiated.

2.3. Step 3: Feature selection

This step is aimed to further enhance the prediction accuracy by selecting only the essential features from the previously identified functions. For this purpose, a statistics-based approach of feature selection helps in selecting the best subset of variables without altering their representation, thus achieving a balance between model simplicity and goodness of fit. The model obtained in Step 2 is tested using stepwise forward and backward regression, by formulating a null hypothesis as

$$\gamma_j = 0 \tag{8}$$

where γ_j represents the estimated coefficient of a feature, z_j, considered for selection. In this work, the terms present in the model obtained from Step 2 represent the feature candidates available in this step. Features are added or removed from the model based on a statistical criterion like p-value (the probability of a null hypothesis to be true), obtained from F-test. The order of adding features to the model is decided by measuring the correlation between the dependent variable, y, and the independent feature, z, as follows:

$$r_{zy} = \frac{\sum_{i=1}^{m}(z_i - \bar{z})(y_i - \bar{y})}{\left(\sum_{i=1}^{m}(z_i - \bar{z})^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^{m}(y_i - \bar{y})^2\right)^{\frac{1}{2}}} \tag{9}$$

For m time-series samples considered, $\bar{z}$ and $\bar{y}$ are the mean values of the considered feature and the dependent variable, respectively. The most promising feature with the highest correlation coefficient is first added to the model and its statistical significance is examined using F-test. If this feature is significant, then the next features are added one at a time based on their partial correlation coefficient (Draper and Smith, 2014; Klein et al., 1981). At every step, the significance of all the previously selected features and the new feature to be added is evaluated. At any time during the evaluation, a previously added feature can be removed if it becomes insignificant in contributing to the desired prediction accuracy. If the p-value for a feature is less than the pre-specified significance level (i.e., α-value), then the null hypothesis is rejected indicating that the feature is valuable and is added to the model. The selection procedure stops when further addition or removal of features cannot improve goodness of fit. With this heuristic approach, only the most informative features are retained in the model, thus reducing the run time and complexity associated with more parameters. Please note that, for the cases with very few samples and more predictor variables, this particular selection technique may not deliver expected results always (Lewis, 2007). Fortunately, for most of the dynamic processes, the number of samples available is more than that of the predictor variables considered, including the application demonstrated in this work.

Remark 1. For the cases where a large amount of data is available, often times it is possible to fully identify the original process model using Step 1 (SINDy) alone with a suitable value of λ (please refer the case studies presented in Brunton et al. (2016b)).

Remark 2. Please note that measurement noise is not considered in this case study. However, in many practical applications, the measurement data is often contaminated by noise that may affect the performance of the proposed method. This can be addressed by denoising the data using the available nonlinear noise reduction techniques such as filtering and smoothing (Schreiber, 1993; Aguirre and Billings, 1995; Lalley et al., 1999; Kostelich and Schreiber, 1993). Furthermore, the differentiation of data in Step 1 has to be carried out using robust methods (Chartrand, 2011; Knowles and Renka, 2014) that can compensate for noise. Additionally, it is important to implement a feature selection method which is robust to noisy data; for example, one can use a hybrid feature selection method combining both filter and wrapper methods (Kumar et al., 2005). In this work, stepwise regression was used as it was shown to perform well in the presence of measurement noise (Burkholder and Lieber, 1996).

3. Application to CSTR

This section demonstrates the application of the proposed method for a perfectly mixed, non-isothermal CSTR. An exothermic, irreversible reaction A → B with the second order kinetics is considered whose reaction rate is given by

$$r = K C_A^2 \tag{10}$$

where K is the temperature dependent rate constant, and C_A is the time-varying concentration of reactant A. The reaction rate constant is determined by Arrhenius law as

$$K = K_0 \exp\left(\frac{-E}{RT(t)}\right) \tag{11}$$

Table 1 – Parameter values for simulation.

Parameter	Units	Value
Flowrate, F	m^3/h	5
Arrhenius pre-exponential factor, K_0	1/h	8.46×10^6
Reactor volume, V_r	m^3	1
Gas constant, R	kJ/Kmol-K	8.314
Inlet temperatue, T_0	K	300
Initial concentration, C_0	$Kmol/m^3$	4
Activation energy, E	kJ/Kmol	5×10^4
Enthalpy change, ΔH	kJ/Kmol	-1.15×10^4
Fluid density, ρ	kg/m^3	1000
Specific heat, c_p	kJ/kg-K	0.231

where K_0 is the pre-exponential factor, E is the activation energy of the reaction, R is the universal gas constant, and T is the time-varying reactor temperature in Kelvin. The temperature is maintained by adjusting the amount of heat transferred through the reactor jacket. The following equations obtained from mass and energy balance of the reactor are the mathematical representation of concentration and temperature dynamics in a CSTR. These equations are used to generate simulation data for identifying the governing dynamics of the process.

$$\frac{dC_A(t)}{dt} = \frac{F}{V_r}(C_0 - C_A(t)) - K_0 \exp\left(\frac{-E}{RT(t)}\right) C_A(t)^2 \tag{12}$$

$$\frac{dT(t)}{dt} = \frac{F}{V_r}(T_0 - T(t)) - \frac{\Delta H}{\rho c_p} K_0 \exp\left(\frac{-E}{RT(t)}\right) C_A(t)^2 + \frac{Q(t)}{\rho c_p V_r} \tag{13}$$

In the above equations, F is the feed flow rate to the reactor, V_r is the reactor volume, ΔH is the heat of reaction, Q is the manipulated rate of heat input, and ρ and c_p are the density and specific heat capacity of the fluid in the reactor, respectively. The temperature-dependent rate constant and the coupled dynamics between temperature and concentration contribute to the complex nonlinear dynamics, making the process of system identification challenging. The objective of this case study is to develop an adaptive model that captures the evolving dynamics of concentration and temperature with a higher prediction accuracy. In the following section, the performance of models identified using the adaptive method and its offline counterpart, SINDy, is evaluated.

4. Simulation results

This section presents the results obtained from the numerical experiments carried out for model identification of the CSTR dynamics. The characteristics of the models developed using the proposed method and the original SINDy method are compared on the basis of prediction accuracy and the total number of data samples required. In this work, all the simulations were performed using MATLAB R2018b programming platform.

The input–output data required for training the models is generated by solving open-loop simulations of the mathematical models, Eqs. (12) and (13), using the `ode45` solver. The process is subjected to a random heat input profile with signals varying between -6×10^4 kJ/h to 10×10^4 kJ/h. A simulation time step of 1×10^{-6} h is considered within the solver, and the data is collected with a sampling time step of 1×10^{-4} h. Assuming full state measurements are available, the process outputs are C and T. The parameter values considered for numerical simulation are shown in Table 1. It is expected that the function coefficients of the identified model must be approximately in the same range as the true values shown in Table 2. In the following subsections, the adaptive sparse identification and SINDy based models are identified, validated and then compared.

Table 2 – True coefficients.

Functions	$\frac{dC_A}{dt}$	$\frac{dT}{dt}$
1	20	1500
C	−5	0
T	0	−5
$\exp\left(\frac{-E}{RT}\right) C^2$	-8.46×10^6	4.21×10^8
Q	0	4.33×10^{-3}

4.1. Adaptive model identification

As mentioned earlier, the proposed algorithm is a three-step method having different goals in each step as:

Step 1) Identify an initial set of potential functions.

Step 2) Update the identified function coefficients.

Step 3) Select the best combination of essential functions.

In Step 1, the original SINDy algorithm is applied to identify the governing functions of the concentration and temperature dynamics. To this end, a candidate library matrix is built with 22 functions as represented in Eq. (14). The advantage of SINDy, which is to incorporate a priori knowledge such as the temperature dependence of the rate constant via Arrhenius law, is utilized by including a temperature-dependent exponential term in the function library. The library developed in this step is a $m \times 22$ matrix, where m indicates the size of the time series data. In the subsequent steps, as the model gets updated, the column size of the library may vary. The simulated outputs of concentration, temperature, and the manipulated heat input are represented as x_1, x_2 and u, respectively.

$$\boldsymbol{\Theta}(\mathbf{x}, \mathbf{u}) = \left[1 \quad x_i^n \quad \exp\left(\frac{-E}{Rx_2}\right) x_1^2 \quad u \quad u^2 \quad x_i u \quad \sin(x_i) \quad \cos(x_i) \right] \tag{14}$$

In the above equation, the subscript $i = 1, 2$ corresponds to the concentration and temperature variables, respectively, and $n = 1, \ldots, 6$ indicates the degree of the polynomial. This step is performed offline using the available historical data of $m = 5 \times 10^3$ samples. This historical data is obtained by simulating the process starting at initial conditions $C = 1.9$ kmol/m^3 and $T = 400$ K for a total duration of $t = 0.5$ h. Please note that, in this specific application, the concentration and temperature values are different by two orders of magnitude. This disparity in the scales can prompt poor sparsity, especially in the scenario of dealing with many functions. This issue is handled by normalizing the concentration data. Specifically, the concentration values are multiplied with the ratio of the mean values of temperature and concentration. Also, the magnitude of the exponential term in the candidate library is relatively higher than that of the other functions present, and this leads to a scaling issue. To solve this problem, each library element is divided by the mean of the corresponding library column. After pre-processing the data, the samples are differentiated by finite difference method and are used to solve the sparse regression problem as presented in Eq. (6). For this purpose, the STLS method is used with 100 iter-

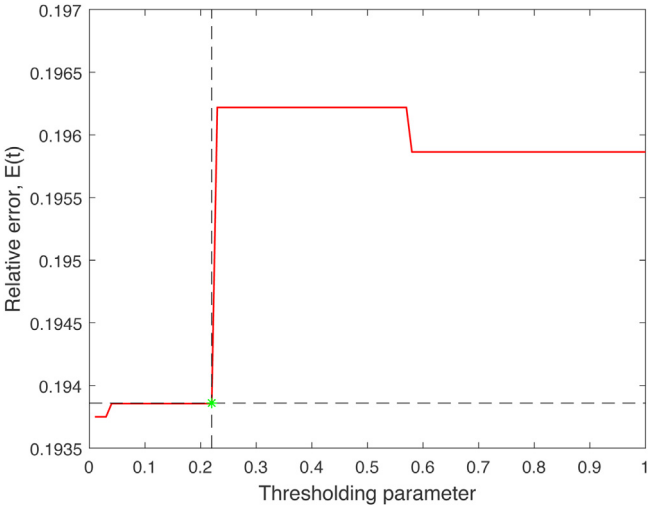


Fig. 2 – Relative error vs. thresholding parameter.

Table 3 – Sparse coefficients estimated in Step 1.

Functions	$\frac{dC_A}{dt}$	$\frac{dT}{dt}$
1	−85.937	-2.95×10^7
C	−0.30311	-9.99×10^4
T	1.781	4.95×10^5
C^2	−4.510	1.12×10^5
T^2	−0.012	−3458.71
C^3	2.294	-6.55×10^4
T^3	0	12.854
$\exp\left(\frac{-E}{RT}\right) C^2$	-8.46×10^6	8.98×10^8
Q	0	4.14×10^{-3}
C^4	0	0
T^4	0	−0.0268
C^5	0	0
T^5	0	0
C^6	0	0
T^6	0	0
CQ	0	0
TQ	0	0
Q^2	0	0
sin(C)	0	0
cos(C)	0	0
sin(T)	0	0
cos(T)	0	0

ations to ensure optimum convergence of coefficients. The value of the thresholding parameter, λ, affects the degree of sparsity observed (Goharoodi et al., 2018). Different models are obtained for different values of λ and increasing λ results in a more sparse model. But as the degree of sparsity increases, many functions are disregarded and because of this, the error computed between the predicted value and the measured value increases. Therefore, the λ value is selected by balancing sparsity and accuracy. In this case, the relative error given by Eq. (7) is considered as a measure for quantifying model accuracy. As shown in Fig. 2, the Pareto front analysis gives the optimum value of thresholding parameter as 0.22. For this value of λ, the model identified in this step performs well with respect to training data as presented in Fig. 3. In reference to the model structure, Step 1 identifies only 10 out of 22 functions as the potential candidates and the results are shown in Table 3. From the table it can be observed that by solving Eq. (6), all the original functions present in Eqs. (12) and (13) are correctly identified along with some additional functions which are not part of the original system. However, the values of the identified function coefficients are not close to the

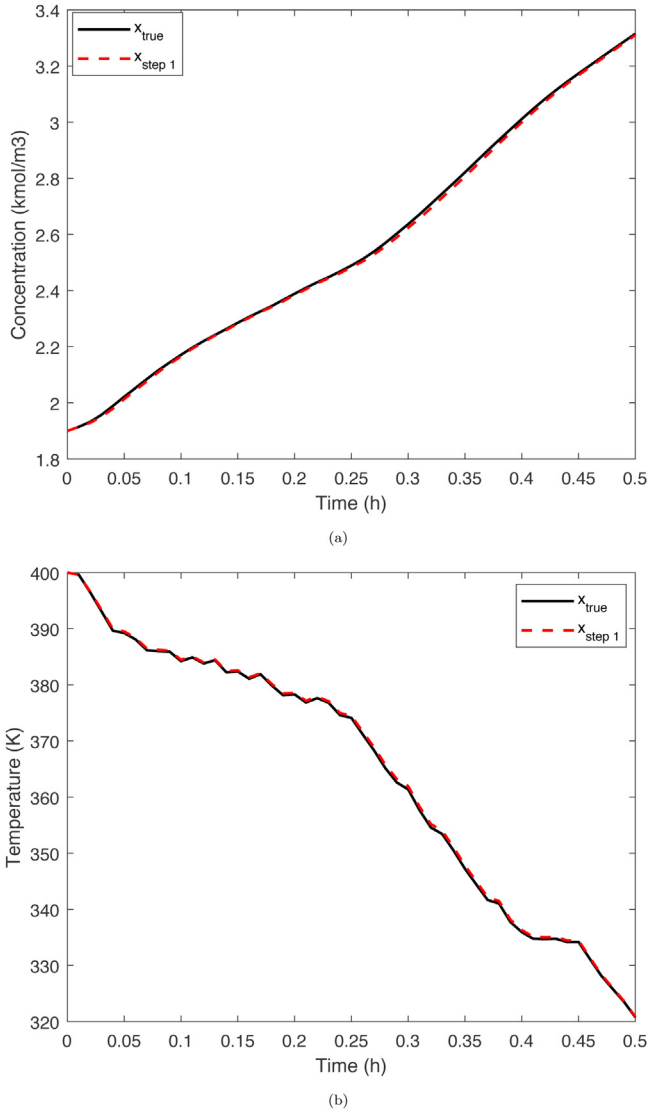


Fig. 3 – Results obtained in Step 1 using the training data for (a) concentration and (b) temperature profiles.

actual model and thus, the model needs to be updated in the subsequent steps.

Once the model is obtained offline through Step 1, it is used as a baseline model to predict the dynamics of a process starting at initial conditions $C = 3.31\,\text{kmol/m}^3$ and $T = 320.75\,\text{K}$. A relative error tolerance value of $\epsilon = 5 \times 10^{-3}$ is considered for both temperature and concentration and any model having an error exceeding this value is deemed poor. The divergence point where the obtained model prediction deviates from the actual measurement serves as an indication to start Step 2. In Fig. 4(a), the model obtained from Step 1 predicts well from $t = 0\,\text{h}$ to $t = 0.18\,\text{h}$, and after that it begins to deviate and can no longer be used for predicting the future states. The relative error computed between the predicted output and the measured data is illustrated in Fig. 4(b). It can be observed that the relative error exceeds the tolerance at $t = 0.18\,\text{h}$ and at this point, Step 2 of the proposed framework is initiated.

In Step 2, ordinary least-squares regression is performed for updating the coefficients of the previously identified functions. The amount of data utilized in this step is 5×10^3 samples, which are collected from the process between $t = 0$ to $t = 0.18\,\text{h}$. The library matrix is re-constructed using only the 10 functions identified in Step 1 and their coefficients values are determined by solving Eq. (5), without any thresholding.

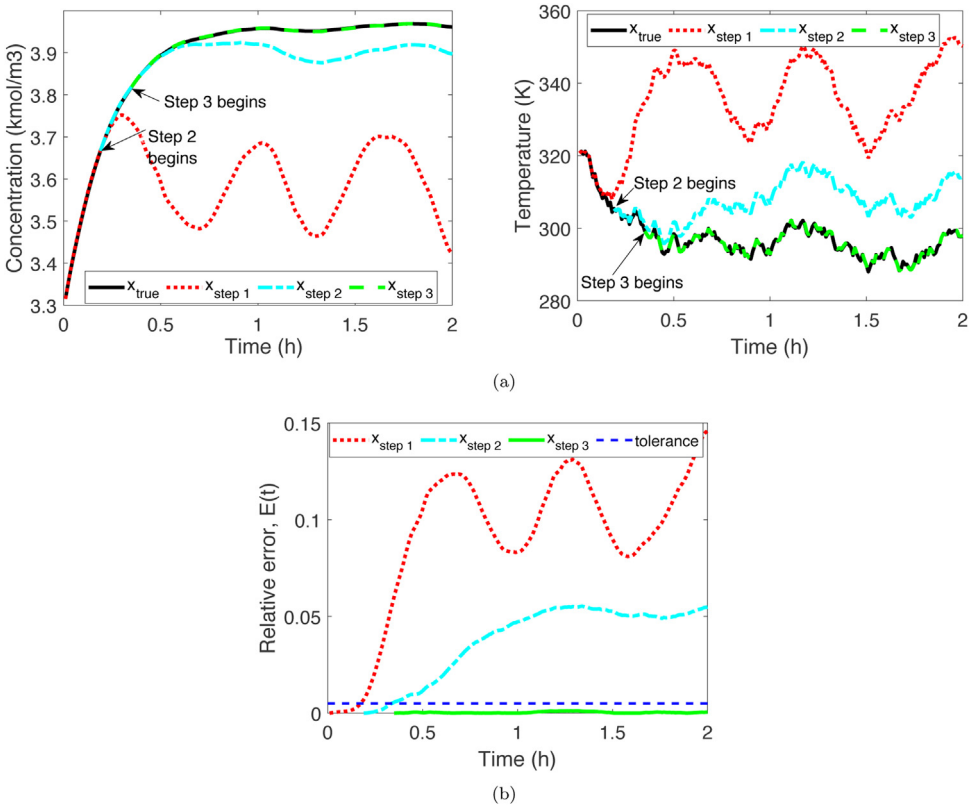


Fig. 4 – (a) Comparing the future behavior of individual models obtained using the proposed method for concentration and temperature profiles. (b) Relative errors of the models identified using the proposed method with respect to time.

Table 4 – Regression coefficients estimated in (a) Step 2 and (b) Step 3.

Functions	$\frac{dC_A}{dt}$	$\frac{dT}{dt}$
(a) Model obtained in Step 2		
1	19.896	-1.05×10^5
C	−4.982	7.81×10^3
T	7.10×10^{-4}	1.02×10^3
C^2	-5.31×10^{-3}	−2345.24
T^2	0	−3.881
C^3	5.34×10^{-4}	237.284
T^3	0	6.05×10^{-3}
$\exp\left(\frac{-E}{RT}\right) C^2$	-8.46×10^6	-6.41×10^8
Q	0	4.3×10^{-3}
T^4	0	0
(b) Model obtained in Step 3		
1	19.896	1311.4
C	−4.982	16.993
T	7.10×10^{-4}	−4.587
C^2	-5.31×10^{-3}	0
T^2	0	0
C^3	5.34×10^{-4}	0
T^3	0	0
$\exp\left(\frac{-E}{RT}\right) C^2$	-8.46×10^6	4.06×10^8
Q	0	4.3×10^{-3}

The results obtained from Step 2 are presented in Table 4(a). From the table, it can be seen that the coefficients for concentration are nearly identical to the original model and T^4 term is observed to play no role in the reactor dynamics as can be seen from its zero coefficient value. Therefore, only the remaining 9 functions are taken into account for improving the model further. From Fig. 4(a) it can be observed that at $t = 0.34$ h the temperature profile deviates from the actual model, i.e., the relative error exceeds the pre-specified tolerance, ϵ, and this triggers Step 3. Additionally, the overall performance of the model updated through Step 2 is better than the model identified in Step 1, as can be seen from the plots depicted in Fig. 4(b). Note that, as the concentration prediction fits well with the actual behavior (Table 4(a)), Step 3 is performed to update the temperature model only.

In Step 3, only the functions that contribute most significantly towards improved prediction accuracy are selected using feature selection. To do this, a feature matrix is constructed using the available 5×10^3 time samples till $t = 0.34$ h. In this case, the products of the 9 identified functions and their updated coefficients from Step 2 represent the features to be considered for selection. Within the feature selection algorithm, the standard significance level of $\alpha = 0.05$ is specified as the criterion to reject the null hypothesis, i.e., a feature is added to the model if its p-value is less than 0.05. Among the 9 features considered, only 5 of them are selected in this step as the dominating ones (Table 4(b)), resulting in a more concise and efficient model. It can be observed from Table 4(b) that the identified model is very close to the true model both in terms of the governing functions as well as their coefficients. As shown in Fig. 4(a), both the temperature and concentration profiles obtained from Step 3 are nearly identical to the actual process, and the relative error of the updated model is within the tolerance limits (Fig. 4(b)). Thus, in terms of predictive performance, the model identified in Step 3 is superior to the previously obtained models as it contains essential features only.

In real-time applications, a major challenge of model adaptation is to rapidly update the model to capture all of the changing dynamics. Therefore, analyzing the computational time of each step is important. For the case study presented in this work, the model is updated from a point where it diverges from the actual behavior of the system. The computational

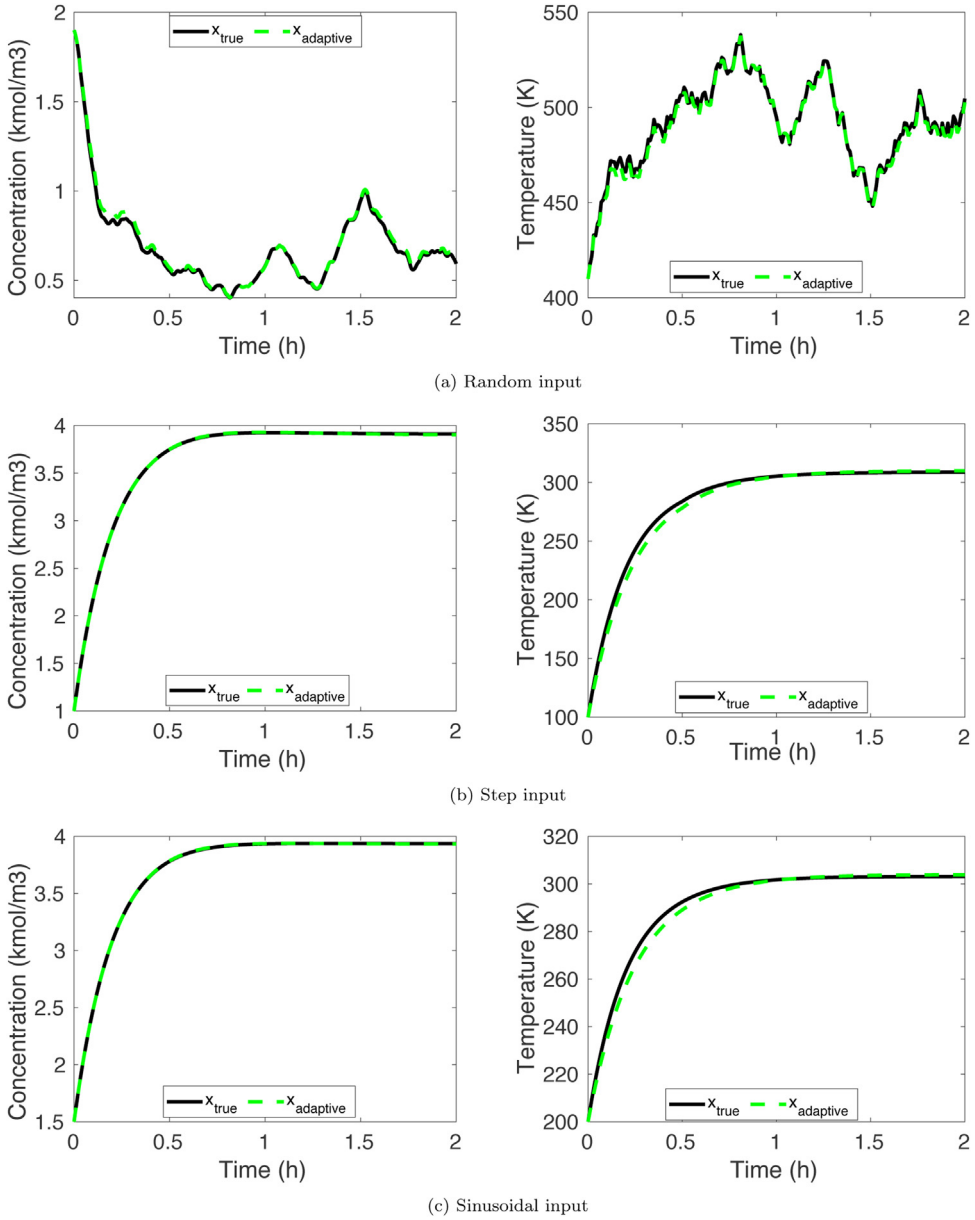


Fig. 5 – Open-loop validation of concentration and temperature profiles described by the adaptive model for three different input settings.

times taken for updating the models in Step 2 and Step 3 of the proposed method are 0.137 s and 1.862 s, respectively. From the results it can be interpreted that in these steps the model structure is improved almost instantaneously since only a limited amount of data is used in both the steps. Note that, during this time of model update, the model identified in the previous steps continues to be in use until a new model is identified.

Remark 3. The number of samples used for training the model is one of the most important factors for any data-driven model identification methods. In this work, multiple numerical simulations were performed considering varying number of data samples in order to train the model. However, not all of the results obtained are reported in this work. Only the results of the best model identified and the corresponding data size is discussed in the manuscript. The proposed algorithm is observed to perform well for any dataset having the size larger than this optimum value.

Remark 4. Please note that the presence of noise may lead to frequent model updates if not handled appropriately. Once the data is cleaned using the previously mentioned techniques, the effects of measurement noise can be mitigated preventing frequent model updates.

4.2. *Validation of the adaptive method*

To evaluate the quality of the final model derived using the proposed method, it is validated against various sets of input profiles at different operating conditions. The validation datasets are generated using three kinds of heat input profiles; a random input with signals varying between -1.2×10^5 kJ/h and 2×10^5 kJ/h, an input with a step of 5×10^3 kJ/h, and a sinusoidal input with an amplitude and frequency of 6 kJ/h and 2 h^{-1}. The sampling and simulation time steps are the same as that of the training data, and the validation results are presented in Fig. 5. The results observed from the figure show that in all the three cases, the adaptive model predicts the process dynamics accurately.

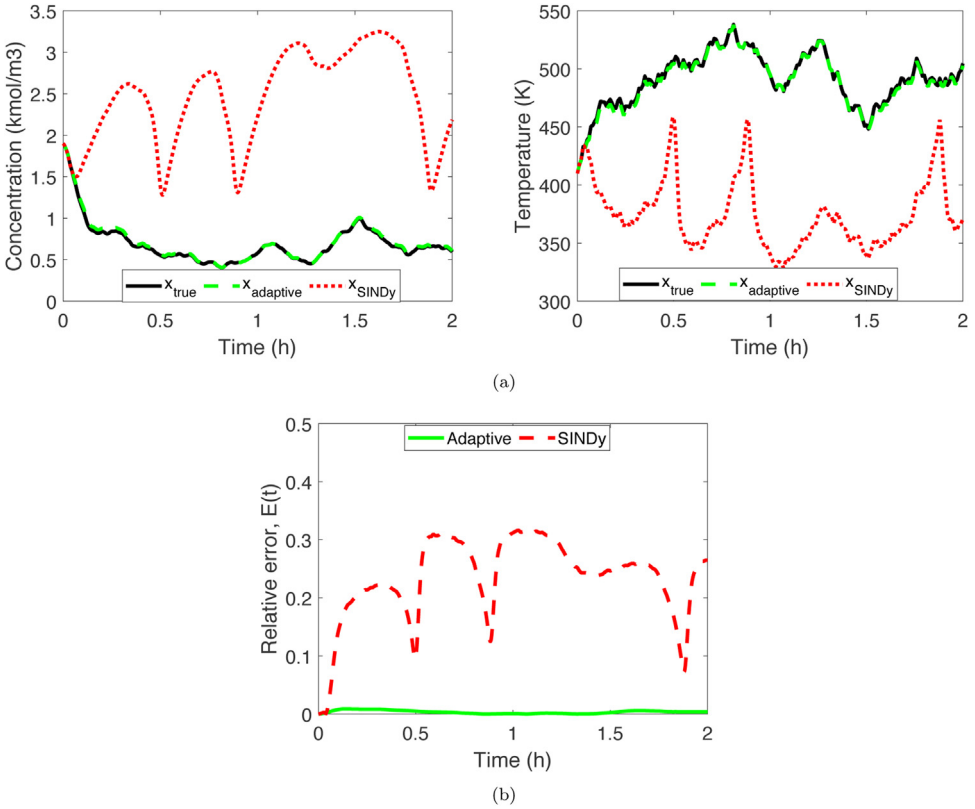


Fig. 6 – (a) Open-loop validation of the adaptive and SINDy models identified using the same number of samples for concentration and temperature profiles. (b) Comparison of the relative errors for the adaptive and SINDy models identified using the same amount of data.

4.3. *Comparison with SINDy*

In this subsection, the proposed method is compared with SINDy in terms of future state prediction. In order to do this, the input–output dataset used in obtaining the adaptive model is considered. For the comparison to be consistent, the number of samples used to perform SINDy is taken to be the same as the total number of samples used for the proposed method (i.e., all three steps combined). The resulting sparse regression problem is solved using STLS with a thresholding parameter of 0.22, and 100 iterations are employed for proper convergence. Once the models are discovered by the proposed method and SINDy, their prediction performance is compared using validation data generated with a random input profile. From the results presented in Fig. 6(a), it can be observed that the model identified by SINDy fails to interpret the actual dynamics while the adaptive model (final model obtained after implementing Step 3) accurately represents both the concentration and temperature dynamics. Furthermore, the final model identified by the proposed method has a very low relative error at each time point (Fig. 6(b)). For the same number of samples, the proposed adaptive method could approximately predict the real model (Table 4), while an additional amount of data may be required for SINDy to fully identify the system. To generalize the validation results across different input settings, both the models are compared using 100 different random input profiles. The relative errors based on Frobenius norm are calculated for each of the input profiles and their average values are shown in Table 5, highlighting the superior performance of the proposed method.

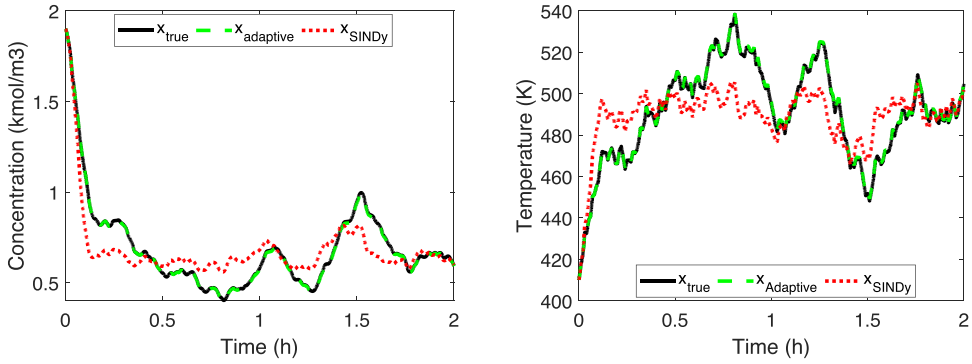
Additionally, the performance of both the methods is tested for the case when the SINDy model is trained using a large amount of training data. Specifically, 6 times the total number of samples used in identifying the adaptive model is used in training the SINDy model. In Fig. 7(a), the response of the final model previously identified by the proposed method is compared with the model identified by SINDy with respect to the concentration and temperature variables. Also, the relative errors estimated at each time point for both the models are compared in Fig. 7(b). Although the performance of the model identified by SINDy using a high number of samples is improved when compared to the previous case (using a less number of samples), the accuracy of the adaptive model still outperformed SINDy. Overall, the results ascertain the usefulness of the proposed approach in obtaining a reasonable model using a less amount of data.

Table 5 – Average relative error computed for 100 different datasets.

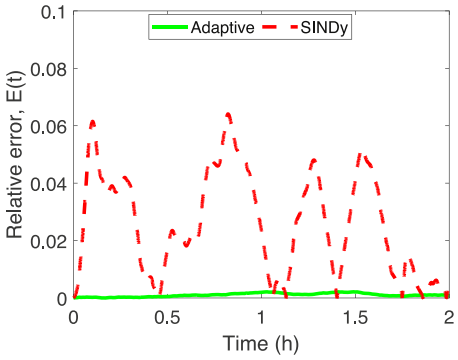
Model	Relative error
Adaptive	3.60×10^{-3}
SINDy	2.51×10^{-1}

5. Conclusions

In this work, an adaptive model identification method is proposed that seeks to identify and improve the model following the three steps. First, a set of potential candidate functions is identified using sequentially thresholded sparse regression. In the following step, the coefficients of these identified functions are updated using least-squares regression, and lastly, stepwise regression is implemented for selecting the best combination of the most important features. The choice of candidate library functions, the thresholding parameter



(a)



(b)

Fig. 7 – (a) Open-loop validation of the adaptive model and SINDy model identified using a large amount of data for concentration and temperature profiles. (b) Comparison of the relative errors for the adaptive model and SINDy model identified using a large amount of data.

value considered, the feature selection criterion for stepwise regression, and the number of samples used at every step significantly affect the performance of this adaptive approach. The effectiveness of the proposed methodology was demonstrated on a CSTR system with second-order kinetics. For a less amount of data, the adaptive model successfully identified the coupled dynamics between concentration and temperature variables. In all the cases tested, the prediction accuracy of the model identified by the proposed algorithm was much higher than that of its offline counterpart, SINDy. Furthermore, the final model was observed to perform well when it is validated with different operating conditions, making it a viable representation of the actual dynamics. In conclusion, the method presented is suitable for adaptive model identification and prediction of complex process dynamics. Such an adaptive model identification framework will be of great interest in adaptive control applications which will be considered in the future.

Acknowledgements

The authors gratefully acknowledge financial support from the National Science Foundation (CBET-1804407), the Texas A&M Energy Institute, and the Artie McFerrin Department of Chemical Engineering.

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., Varoquaux, G., 2014. Machine learning for neuroimaging with scikit-learn. Front. Neuroinform. 8, 14.

Aguirre, L.A., Billings, S., 1995. Identification of models for chaotic systems from noisy data: implications for performance and nonlinear filtering. Physica D: Nonlinear Phenomena 85, 239–258.

Alanqar, A., Durand, H., Christofides, P.D., 2017. Error-triggered on-line model identification for model-based feedback control. AIChE J. 63, 949–966.

Åström, K.J., Eykhoff, P., 1971. System identification-a survey. Automatica 7, 123–162.

Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. J. Mach. Learn. Res. 13, 281–305.

Bergstra, J.S., Bardenet, R., Bengio, Y., Kégl, B., 2011. Algorithms for hyper-parameter optimization. In: Advances in Neural Information Processing Systems, New York, USA, pp. 2546–2554.

Billings, S., Voon, W., 1986. A prediction-error and stepwise-regression estimation algorithm for non-linear systems. Int. J. Control 44, 803–822.

Bongard, J., Lipson, H., 2007. Automated reverse engineering of nonlinear dynamical systems. Proc. Natl. Acad. Sci. USA 104, 9943–9948.

Brigham, J.C., Aquino, W., 2007. Surrogate-model accelerated random search algorithm for global optimization with applications to inverse material identification. Comput. Methods Appl. Mech. Eng. 196, 4561–4576.

Brunton, S.L., Brunton, B.W., Proctor, J.L., Kutz, J.N., 2016a. Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control. PLOS ONE 11, e0150171.

Brunton, S.L., Proctor, J.L., Kutz, J.N., 2016b. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proc. Natl. Acad. Sci. USA 113, 3932–3937.

Burkholder, T.J., Lieber, R.L., 1996. Stepwise regression is an alternative to splines for fitting noisy data. J. Biomech. 29, 235–238.

Champion, K.P., Brunton, S.L., Kutz, J.N., 2019. Discovery of nonlinear multiscale systems: sampling strategies and embeddings. SIAM J. Appl. Dyn. Syst. 18, 312–333.

Chandrashekar, G., Sahin, F., 2014. A survey on feature selection methods. Comput. Electr. Eng. 40, 16–28.

Chartrand, R., 2011. Numerical differentiation of noisy, nonsmooth data. ISRN Appl. Math. 2011.
Cisternas, J., Gear, C.W., Levin, S., Kevrekidis, I.G., 2004. Equation-free modelling of evolving diseases: coarse-grained computations with individual-based models. Proc. R. Soc. Lond. Ser. A: Math. Phys. Eng. Sci. 460, 2761–3277.
Daniels, B.C., Nemenman, I., 2015. Automated adaptive inference of phenomenological dynamical models. Nat. Commun. 6, 8133.
Draper, N.R., Smith, H., 2014. Applied Regression Analysis, vol. 326. John Wiley & Sons, New York, USA.
Eggensperger, K., Feurer, M., Hutter, F., Bergstra, J., Snoek, J., Hoos, H., Leyton-Brown, K., 2013. Towards an empirical foundation for assessing bayesian optimization of hyperparameters. NIPS workshop on Bayesian Optimization in Theory and Practice, 10, 3.
Giannakis, D., Majda, A.J., 2012. Nonlinear laplacian spectral analysis for time series with intermittency and low-frequency variability. Proc. Natl. Acad. Sci. USA 109, 2222–2227.
Goharoodi, S.K., Dekemele, K., Dupre, L., Loccufier, M., Crevecoeur, G., 2018. Sparse identification of nonlinear duffing oscillator from measurement data. IFAC-PapersOnLine 51, 162–167.
Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182.
Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12, 55–67.
Hoffmann, M., Fröhner, C., Noé, F., 2019. Reactive sindy: discovering governing reactions from concentration data. J. Chem. Phys. 150, 025101.
Hong, X., Mitchell, R.J., Chen, S., Harris, C.J., Li, K., Irwin, G.W., 2008. Model selection approaches for non-linear system identification: a review. Int. J. Syst. Sci. 39, 925–946.
Hou, Z.S., Wang, Z., 2013. From model-based control to data-driven control: survey, classification and perspective. Inf. Sci. 235, 3–35.
Hunt, K.J., Sbarbaro, D., Żbikowski, R., Gawthrop, P.J., 1992. Neural networks for control systems – a survey. Automatica 28, 1083–1112.
Jović, A., Brkić, K., Bogunović, N., 2015. A review of feature selection methods with applications. In: 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), IEEE, Opatija, Croatia, pp. 1200–1205.
Kaiser, E., Kutz, J.N., Brunton, S.L., 2017. Data-driven discovery of koopman eigenfunctions for control. Bull. Am. Phys. Soc., 62.
Kaiser, E., Kutz, J.N., Brunton, S.L., 2018. Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. Proc. R. Soc. A 474, 20180335.
Klein, V., Batterson, J.G., Murphy, P.C., 1981. Determination of Airplane Model Structure From Flight Data by Using Modified Stepwise Regression. NASA TP-1916.
Knowles, I., Renka, R.J., 2014. Methods for numerical differentiation of noisy data. Electron. J. Differ. Equ. 21, 235–246.
Kostelich, E.J., Schreiber, T., 1993. Noise reduction in chaotic time-series data: a survey of common methods. Phys. Rev. E 48, 1752.
Kumar, R., Jayaraman, V.K., Kulkarni, B.D., 2005. An svm classifier incorporating simultaneous noise reduction and feature selection: illustrative case examples. Pattern Recognit. 38, 41–49.
Lalley, S.P., et al., 1999. Beneath the noise, chaos. Ann. Stat. 27, 461–479.
Larimore, W.E., 1990. Canonical variate analysis in identification, filtering, and adaptive control. In: 29th IEEE Conference on Decision and Control. Piscataway, NJ, pp. 596–604.
Lewis, M., 2007. Stepwise versus hierarchical regression: Pros and cons. In: Paper presented at the Annual Meeting of the Southwest Educational Research Association, San Antonio, TX.
Li, S., Kaiser, E., Laima, S., Li, H., Brunton, S.L., Kutz, J.N., 2018. Discovering Time-Varying Aeroelastic Models of a Long-Span Suspension Bridge From Field Measurements by Sparse Identification of Nonlinear Dynamical Systems. arXiv preprint arXiv:1809.05707.
Loiseau, J.C., Brunton, S.L., 2018. Constrained sparse galerkin regression. J. Fluid Mech. 838, 42–67.
Loiseau, J.C., Noack, B.R., Brunton, S.L., 2018. Sparse reduced-order modelling: sensor-based dynamics to full-state estimation. J. Fluid Mech. 844, 459–490.
Mangan, N.M., Brunton, S.L., Proctor, J.L., Kutz, J.N., 2016. Inferring biological networks by sparse identification of nonlinear dynamics. IEEE Trans. Mol. Biol. Multi-Scale Commun. 2, 52–63.
Mangan, N.M., Kutz, J.N., Brunton, S.L., Proctor, J.L., 2017. Model selection for dynamical systems via sparse regression and information criteria. Proc. R. Soc. A: Math. Phys. Eng. Sci. 473, 20170009.
Narasingam, A., Kwon, J.S.I., 2018. Data-driven identification of interpretable reduced-order models using sparse regression. Comput. Chem. Eng. 119, 101–111.
Quade, M., Abel, M., Nathan Kutz, J., Brunton, S.L., 2018. Sparse identification of nonlinear dynamics for rapid model recovery. Chaos 28, 063116.
Rudy, S.H., Brunton, S.L., Proctor, J.L., Kutz, J.N., 2017. Data-driven discovery of partial differential equations. Sci. Adv. 3, e1602614.
Schaeffer, H., 2017. Learning partial differential equations via data discovery and sparse optimization. Proc. R. Soc. A: Math. Phys. Eng. Sci. 473, 20160446.
Schmidt, M., Lipson, H., 2009. Distilling free-form natural laws from experimental data. Science 324, 81–85.
Schreiber, T., 1993. Extremely simple nonlinear noise-reduction method. Phys. Rev. E 47, 2401.
Seborg, D., Edgar, T.F., Shah, S., 1986. Adaptive control strategies for process control: a survey. AIChE J. 32, 881–913.
Thompson, M.L., 1978. Selection of variables in multiple regression: Part I. A review and evaluation. Int. Stat. Rev.[Revue Internationale de Statistique] 46, 1–19.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc.: Ser. B (Methodol.) 58, 267–288.
Tsai, C.F., 2009. Feature selection in bankruptcy prediction. Knowl.-Based Syst. 22, 120–127.
Van Overschee, P., De Moor, B., 1994. N4sid: subspace algorithms for the identification of combined deterministic-stochastic systems. Automatica 30, 75–93.
Varshney, A., Pitchaiah, S., Armaou, A., 2009. Feedback control of dissipative pde systems using adaptive model reduction. AIChE J. 55, 906–918.
Verhaegen, M., Dewilde, P., 1992. Subspace model identification. Part 2. Analysis of the elementary output-error state-space model identification algorithm. Int. J. Control 56, 1211–1241.
Viberg, M., 1994. Subspace methods in system identification. IFAC Proc. Vol. 27, 1–12.
Wu, Z., Rincón, F.D., Christofides, P.D., 2019a. Real-time adaptive machine-learning-based predictive control of nonlinear processes. Ind. Eng. Chem. Res.
Wu, Z., Tran, A., Rincon, D., Christofides, P.D., 2019b. Machine learning-based predictive control of nonlinear processes. Part I: Theory. AIChE J. 65, e16729.
Wu, Z., Tran, A., Rincon, D., Christofides, P.D., 2019c. Machine learning-based predictive control of nonlinear processes. Part II: Computational implementation. AIChE J. 65, e16734.
Ye, H., Beamish, R.J., Glaser, S.M., Grant, S.C., Hsieh, C.H., Richards, L.J., Schnute, J.T., Sugihara, G., 2015. Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. Proc. Natl. Acad. Sci. USA 112, E1569–E1576.
Zhang, L., Schaeffer, H., 2018. On the Convergence of the Sindy Algorithm. arXiv preprint arXiv:1805.06445.