Chemical Engineering Journal Advances 18 (2024) 100605

Contents lists available at ScienceDirect

Chemical Engineering Journal Advances

journal homepage: www.sciencedirect.com/journal/chemical-engineering-journal-advances

Partial derivative-based dynamic sensitivity analysis expression for non-linear auto regressive with exogenous (NARX) model—case studies on distillation columns and model's interpretation investigation

Waqar Muhammad Ashraf, Vivek Dua [*]

The Sargent Centre for Process Systems Engineering, Department of Chemical Engineering, University College London, Torrington Place, London WC1E 7JE, UK

ARTICLE INFO

Keywords:
Time-series data
Explainable AI
NARX
Partial-derivative based sensitivity analysis
Interpretability

ABSTRACT

Constructing the reliable dynamic sensitivity profile for the output variable using the machine learning model is a challenging task; however, the dynamic sensitivity trends are helpful to understand the impact of the input variables on the system's performance. In this paper, we have derived the partial-derivative approach-based sensitivity analysis expression for the non-linear auto regressive with exogenous (NARX) model for the first time. The engineering systems-based case studies, i.e., two distillation columns with five and ten stages, respectively are taken which are commonly found in the chemical processing plants. Two output variables, i.e., liquid composition in tray 2 and tray 4 (Y_2 and Y_4) of a five-stage distillation column, and liquid composition in tray 7 (Y_7) of a ten-stage (higher) distillation column are modelled by NARX with respect to time, feed concentration (X_f) and feed flow rate (L_f). The dynamic sensitivity profiles of the output variables with respect to X_f and L_f for the two distillation columns are plotted by the derived partial derivative-based sensitivity expression on the NARX model. Furthermore, the forward difference method of sensitivity analysis (first principle method) is also applied on the ordinary differential equations of the distillation columns to compute the sensitivity values of the output variables. A good agreement in the dynamic sensitivity values of the output variables with respect to the input variables is found for the two sensitivity analysis techniques thereby demonstrating the effectiveness of the partial-derivative approach for the improved NARX's interpretability performance. This research presents the explicit partial-derivative based sensitivity analysis expression for the NARX model which can be utilised for time-series applications and can provide the insights about the model's interpretation performance.

1. Introduction

The rising energy demand and increased emissions discharge to the environment is expected on the face of increasing population size, consumption based economies and improved life style around the world [1]. However, we are experiencing industry 4.0 revolution in the twenty first century, and the productivity, quality and efficiency of industrial complexes and energy systems have been boosted many folds under the technological advancement [2]. The industrial systems store heaps of data in the data storage banks formally called supervisory information systems. This creates the situation to deploy the industrial data for intensive data-exploratory and value-creating analytics for the time dependent applications [3,4]. The data-driven analytics can be applied for the operational excellence, informed decision making and data-aware policy making for enhancing the performance of any system under investigation [5].

Machine learning (ML) algorithms are used for the data-driven modelling and optimisation analytics as they can detect the pattern and hidden features in the data with good accuracy [6,7], and can be computationally cheaper compared with the first-principle model based analyses [8,9]. Artificial neural network is one amongst the popular modelling algorithms of ML [10] and is deployed for the data-driven modelling applications due to their versatility, excellent ability to approximate the function and low memory requirements [11,12]. Multi layered perceptron (MLP) is the basic component of ANN and its working mimics how human brain processes information to make decisions [13]. The nonlinear autoregressive with exogenous (NARX) network is another variant of ANN and incorporates the MLP structure along with the past observations of the input and output variables that act as sliding windows. The sliding window passes over the training data

[*] Corresponding author.
E-mail address: v.dua@ucl.ac.uk (V. Dua).

https://doi.org/10.1016/j.ceja.2024.100605

Available online 5 April 2024

and contributes to model the nonlinear dynamics of timeseries datasets [14–17]. The sliding window is termed as delay in the algorithm of NARX and is a critical hyperparameter that is to be optimised to achieve the good modelling accuracy of NARX.

Neural networks offer predictive benefits as compared to other algorithms such as their ability to develop hard-to-find nonlinear and complex relationships and interactions between the input and output variables, and to handle volumes of data for building effective functional mapping with reasonable computational resource utilisation [18]. Neural networks are essentially black-box models and thus, it is quite difficult to explain how the model simulates the output variables' value for the given input conditions (low interpretability) [19]. Therefore, research community is actively engaged on developing the techniques to understand and explain the causal effects of input variables on the neural network's predictions [20]. Some examples are described as follows:

Neural interpretation diagram (NID) is a modification to the structure of neural network and it highlights the width and colour of the connections between the neurons depending upon the sign and width of weights thereby differentiating the significant input variables [21]. Garson's method computes the summation of the product of the absolute value of weights from the input to output variables via hidden layer, and scaled it relative to all other input variables to identify the significant variable [22]. Olden's method [22] also performs the similar computation except that real value of weights is used and the resultant value is not scaled. Input perturbation method [23] adds a noise to the value of the input variable under investigation whereas other input variables are kept at a certain value. The constructed experiments are simulated from the model and the change in the selected performance metrics presents the relative variable importance. Similar to the input perturbation method, profile method [24] for sensitivity analysis allows to vary the value of selected input variable while keeping the other input variables at different quantile values. Thus, different plots for the input variables are created. Beck M.W [25] presented a modification to this technique where central value of training data clusters is utilised instead of selecting the quantile values.

Partial dependence plot (PDP) approach [26–29] plots the individual conditional expectation (ICE) curves for the output variable against the selected input variable. Later, the average value of the ICE curves is taken in order to visualise the PDP curve for the input variable. Local interpretable model-agnostic explanations method [30] explains the complexity of ANN by locally approximating it with the interpretable models like linear regression or decision tree. Shapley values computed by SHAP (SHapley Additive exPlanations) method computes the marginal contribution of the input variable to explain the output of neural network and can be computed by conditional expectation function [31]. Stepwise addition and elimination methods [32] rebuild the neural network by adding or removing the input neuron respectively in a sequential approach, and the change in the chosen performance metrics at each step depicts the relative importance of input variable. Partial derivative method takes the partial derivative of neural network's output with respect to the input variable and evaluate the resulting expression on a dataset.

The techniques mentioned above are useful to explain the interpretability of neural networks. However, they have certain drawbacks as well thereby limiting their applicability. Explaining the results of NID method is difficult given the weight connections in the neural networks. Garson and Olden method only consider the weights connection from input to hidden layer and Lek's profile technique can present the analysis on the constructed scenarios not supported by the training data. PDP may present misleading results for correlated input variables. Local linearisation can provide information only in certain regions thus quantitative information on the entire dataset is missing. SHAP method is not an exact measure of causality, and can be computationally expensive for the large number of input variables. Forward and backward elimination is a computationally exhaustive method and may produce inconsistent results based on order of input variables to be added or removed. Whereas, computing the partial-derivative of function with respect to large number of input variables can be time-consuming.

It is important to mention here that the partial derivative method analytically estimates the variable significance by taking the derivative of the neural network with respect to the input variable from the explicit sensitivity analysis expression. The contributions of the weight connections and activation function are also accounted in addition to the value of the input variables. The partial derivative method provides more robust diagnostics for a well-trained neural network, and thus the derivative of the network will provide stable results thereby providing a competitive edge, in terms of model interpretability compared to the mentioned sensitivity analysis techniques and also outweighs the time-consuming aspect of the approach.

The objective of this work is to derive the partial-derivative based sensitivity analysis expression for the NARX model and utilise the expression to obtain the dynamic sensitivity information of the input variables of the trained NARX model. The explicit partial-derivative based sensitivity analysis expression for NARX is not available in the open literature and thus, this research bridges this identified gap by providing the explicit partial-derivative based sensitivity analysis expression for the NARX model. The partial-derivative approach explains the complexity of neural network through the explicit expression that offers the insights about the model's interpretation performance. In this research, two case studies on time-series based engineering system's applications, i.e., a distillation column and a higher-order distillation column are taken, and NARX models are trained on the data obtained after solving the ordinary differential equations (ODEs) of the distillation columns corresponding to different initial conditions. Later, the dynamic sensitivity trend of the output variables against the input variables of the distillation columns is plotted by the derived NARX based partial-derivative method. Furthermore, the forward difference method on the ODEs of the distillation columns is applied to compute the variable's sensitivity and is compared with that of the partial-derivative approach to confirm the accuracy of the dynamic sensitivity trends plotted by partial derivative-based sensitivity analysis carried out using the NARX model. The comparison also allows to investigate the interpretability performance of NARX model in terms of the significance order of the input variables towards predicting the output variables [33, 34]. Thus, the derived partial derivative based sensitivity analysis expression can be utilised in various real-life applications to plot the dynamic trends of the system's performance complemented with the improved interpretability of NARX algorithm which is helpful to make informed and knowledgeable decisions.

This paper is structured as: The working of the NARX model is described in Section 2. The partial derivative of the NARX model is calculated and presented in Section 3. The dynamic sensitivity trends of the output variables against the input variables are plotted for two examples, distillation and higher-order distillation column, and the details are provided in Sections 4.1 and 4.2 respectively. Finally, conclusion of this research is mentioned in Section 5.

2. Development of non-linear auto regressive with exogenous (NARX) model

A nonlinear autoregressive network with exogenous (NARX) model is a time-series based function approximation algorithm for modelling the dynamic profile of a system. A NARX model is basically a MLP network and can incorporate the past input and output observations to predict the current output value. It can include the delay terms of the input as well as output time-series to map their causal relationships. Mathematically, the working of NARX model can be expressed as [35, 36]:

$$y(t) = f\left(u(t-n_u), \ldots, u(t-1), u(t), y\left(t-n_y\right), \ldots, y(t-1)\right) \tag{1}$$

where, $u(t-n_u),\ldots,u(t-1),u(t)$ is the input time series and $y(t-n_y),\ldots,y(t-1)$ is the output time-series. n_u and n_y are the lag terms introduced in the time-series of input and output variables respectively. f represents the non-linear MLP function mapping the two exogenous series to predict the current value of the output variable ($y(t)$). The states of the NARX model are specified with respect to n_u and n_y which are the tapped delays for input and output time-series respectively. The states of the NARX model are updated as:

$$x_i(t+1) = \begin{cases} u(t)\ i = n_u \\ y(t)\ i = \ n_u + \ n_y \\ x_{i+1}(t)\ 1 \le i\ <\ n_u\ and\ n_u <\ i\ <\ n_u + \ n_y \end{cases} \tag{2}$$

so that, at time 't', the taps correspond to the values:

$$x(t) = \left[u(t-n_u)\ldots u(t-1),\ y\left(t-n_y\right)\ldots y(t-1)\right] \tag{3}$$

According to the working of feedforward MLP, the output produced at the i^{th} neuron of the hidden layer at time 't' is given as ($H_i(t)$):

$$H_i(t) = f_1\left[\sum_{r=0}^{n_u} w_{ir}u(t-r) + \ \sum_{l=1}^{n_y} w_{il}y(t-l) + \ a_i\right] \tag{4}$$

here, w_{ir} is the weight connecting the input time series neuron $u(t-r)$ with the i^{th} hidden layer neuron. Similarly, w_{lr} corresponds to the connection weight between the feedback neuron $y(t-l)$ and the i^{th} hidden layer neuron. a_i is the bias value applied at the i^{th} hidden layer neuron and f_1 is the activation function applied at the hidden layer.

The final output produced at the output neuron of the NARX network is computed as:

$$\hat{y}_j(t) = \ f_2\left[\sum_{i=1}^{n_h} w_{ji}H_i(t) + \ b_j\right] \tag{5}$$

where, w_{ji} is the weight of the link connecting the j^{th} and i^{th} neuron of the output and hidden layer neuron respectively. b_j is the bias value applied at the j^{th} neuron of the output layer; $n_h$is the number of hidden layer neurons; f_2 is the activation function applied at the output layer. Finally, $\hat{y}_j(t)$ is the output value predicted by the NARX network for the given exogenous input and output time series. A simple architecture demonstrating the working of NARX model is presented in Fig. 1. Two input and feedback delays as well as three hidden layer neurons are considered in the NARX network, and the model predicted response for the output is represented by $\hat{y}(t)$ for the input series $u(t)$ and delay states $x_1(t)$ to $x_4(t)$.

The modelling performance of the developed NARX network is measured by two statistical terms namely co-efficient of determination (R^2) and root-mean-squared-error (RMSE). The performance matrix built on these two terms are utilised in research studies for evaluating the prediction efficiency of the machine learning model [37,38]. Mathematically, R^2 and RMSE are written as:

$$R^2 = 1 - \ \frac{\sum_i^N \left(y_i - \ \hat{y}_i\right)^2}{\sum_i^N \left(y_i - \ \bar{y}_i\right)^2} \tag{6}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2} \tag{7}$$

R^2 is regarded as the accuracy of the developed model to predict the value of the output variable corresponding to the input vector. The value of R^2 varies from zero (poor predictability) to one (perfect prediction). Whereas, RMSE indicates the difference between the model-simulated responses and the true observations, and should be minimum thereby the model has excellent prediction performance.

3. Partial derivative-based sensitivity expression of NARX model

The sensitivity of NARX model can be expressed as first-order partial derivative between the input and output variables. NARX consists of MLP structure along with the feedback loop to map the output variable with respect to the input as well as the delayed responses of the variables. The information received at any neuron of the hidden layer at time 't' can be expressed as:

$$S_{h(t)} = \ N_{p(t)}\ w_{hp} + \sum_{i\ \neq p} N_{i(t)}\ w_{hi} + N_{p(t-d)}\ w_{hp}^{'} + \sum_{i\ \neq p} N_{i(t-d)}\ w_{hi}^{'} \\ + \sum_{i=1}^{d} w_{hj}\ y_{i(t-d)}\ + b_h \tag{8}$$

where, $N_{p(t)}$ is the dynamic input variable whose sensitivity on the output is to be evaluated. w_{hp} is the weight connection from $N_{p(t)}$ to a hidden layer neuron; $N_{i(t)}$ is the set of other dynamic input variables having connection weights w_{hi} with the hidden layer neuron; $N_{p(t\ -\ d)}$ and $N_{i(t\ -\ d)}$ are the delayed connections of $N_{p(t)}$ and $N_{i(t)}$ in the NARX having the weights connections w'_{hp} and w'_{hi} respectively; $y_{i(t\ -\ d)}$ is the delayed feedback response of the output variable and w_{hj} is the weight connection of $y_{i(t\ -\ d)}$ with the hidden layer neuron; b_h is the bias introduced to the hidden layer; and $S_{h(t)}$ represents the information collected at the hidden layer of NARX. The activation function (ϕ) is applied on $S_{h(t)}$ which is expressed as:

$$N_{h(t)} = \ \phi_h\left(S_{h(t)}\right) \tag{9}$$

here, $N_{h(t)}$ represents the information signal forwarded to the output layer from the hidden layer of NARX. The information processing at the

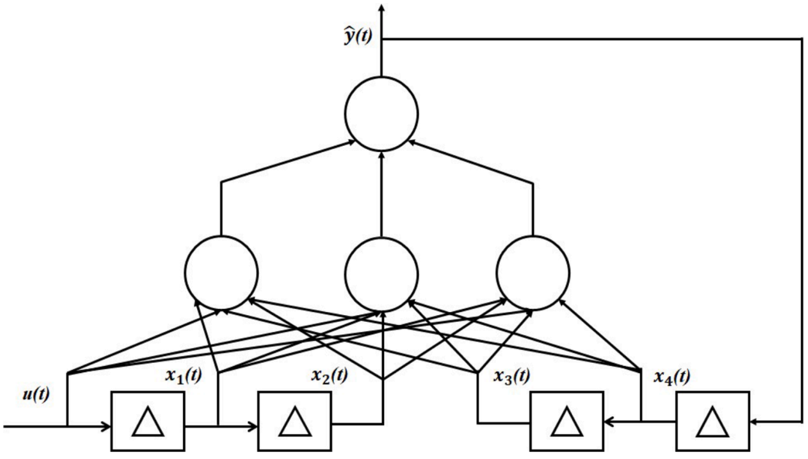

Fig. 1. The multilayer perceptron-based architecture of NARX network. Here, $n_u = n_y = 2$ and $H = 3$.

output layer is given as:

$$S_{o(t)} = N_{h(t)}\, w_{oh} + \sum_{j\neq h} N_{j(t)}\, w_{oj} + b_o \tag{10}$$

$$N_{o(t)} = \phi_o\left(S_{o(t)}\right) \tag{11}$$

here, b_o is the bias term applied at the output layer; ϕ_o refers to the activation function applied on the output layer of NARX; 'j' denotes the neuron in the hidden layer; and $N_{o(t)}$ is the value simulated by NARX for the given input values of the variables. Also, we have:

$$\frac{\partial S_{h(t)}}{\partial N_{p(t)}} = w_{hp} \tag{12}$$

$$\frac{\partial S_{o(t)}}{\partial N_{h(t)}} = w_{oh} \tag{13}$$

The first-order partial derivative of output variable with respect to the input variable X_p ($N_i = X_p$) is:

$$\frac{\partial N_{o(t)}}{\partial X_{p(t)}} = \frac{\partial N_{o(t)}}{\partial N_{p(t)}} = \frac{\partial N_{o(t)}}{\partial N_{h(t)}} \frac{\partial N_{h(t)}}{\partial N_{p(t)}} = \left(\frac{dN_{o(t)}}{dS_{o(t)}} \frac{\partial S_{o(t)}}{\partial N_{h(t)}}\right)\left(\frac{dN_{h(t)}}{dS_{h(t)}} \frac{\partial S_{h(t)}}{\partial N_{p(t)}}\right) \tag{14}$$

Considering Eqs. (9) and (11):

$$\frac{dN_{o(t)}}{dS_{o(t)}} = \phi_o'\left(S_{o(t)}\right) \tag{15}$$

$$\frac{dN_{h(t)}}{dS_{h(t)}} = \phi_h'\left(S_{h(t)}\right) \tag{16}$$

Eq. (14) can be expressed as:

$$\frac{\partial N_{o(t)}}{\partial X_{p(t)}} = \phi_o'\left(S_{o(t)}\right) w_{oh} \phi_h'\left(S_{h(t)}\right)\, w_{hp} \tag{17}$$

Since, the hidden layer consists of more than one neuron, the general form of partial derivative-based input sensitivity of three-layer MLP based NARX model for 'nh' hidden layer neurons is expressed as:

$$\frac{\partial N_{o(t)}}{\partial X_{p(t)}} = \sum_{h=1}^{nh} \phi_o'\left(S_{o(t)}\right) w_{oh} \phi_h'\left(S_{h(t)}\right)\, w_{hp} \tag{18}$$

In this work, ($\phi_h(S_h) = (\exp(2S) - 1)/(\exp(2S) + 1)$) is the tangent hyperbolic based activation function deployed on the hidden layer while linear activation function ($\phi_h(S_o) = S_o$)is implemented on the output layer. Therefore, first-derivative of tangent hyperbolic function is given as:

$$\phi'(S) = 1 - \phi^2(S) = 1 - N^2 \tag{19}$$

Thus, Eq. (18) can be expressed as:

$$\frac{\partial N_{o(t)}}{\partial X_{p(t)}} = \sum_{h=1}^{nh} w_{oh}\left(1 - N^2\right) w_{hp} \tag{20}$$

The Eq. (20) describes the absolute dynamic sensitivity of output variable $N_{o(t)}$ for per unit change in input variable $X_{p(t)}$ which can be deployed to identify the significant input variables for the system under investigation.

4. Results and discussion

4.1. Development of NARX model and its partial-derivative based sensitivity analysis for distillation column

A distillation column is a commonly used industrial component in the material separation techniques and is used for the range of applications including water desalination systems [39,40], crude oil and mixture separations in the process industries [41–43]. The distillation column is a non-linear dynamic system and is considered to investigate the comparison between the sensitivity analysis made by partial derivative method on the NARX model and forward difference method on ODEs of distillation column (first-principle method). The distillation tower consists of a total condenser, five trays and a reboiler. Feed in liquid phase is maintained at its boiling point and enters the tower at tray 3. It is assumed that constant molar flow rate and accurate control of the levels in the reboiler and condenser are maintained. The disturbances are introduced in feed concentration X_f and feed flow rate L_f. The dynamic operation of the distillation column is represented by the following ordinary differential equations (ODEs):

$$\text{Reboiler}: \; H_r\frac{dX_1}{dt} = \left(L + L_f\right)(X_2 - X_1) + V(X_1 - y_1) \tag{21}$$

$$\text{Tray 2}: \; H_t\frac{dX_2}{dt} = \left(L + L_f\right)(X_3 - X_2) + V(y_1 - y_2) \tag{22}$$

$$\text{Feed tray}: \; H_t\frac{dX_3}{dt} = L_f X_f + LX_4 - \left(L + L_f\right)X_3 + V(y_2 - y_3) \tag{23}$$

$$\text{Tray 4}: \; H_t\frac{dX_4}{dt} = L(X_5 - X_4) + V(y_3 - y_4) \tag{24}$$

$$\text{Condenser}: \; H_c\frac{dX_5}{dt} = V(y_4 - X_5) \tag{25}$$

The equilibrium in vapour–liquid state in the distillation column is expressed as:

$$y_i = \frac{\alpha X_i}{1 + (\alpha - 1)X_i} \tag{26}$$

where, X_i represents the light component's liquid mole fraction at tray i ($i = 1,2,\ldots,5$) and y_i denotes the light component's vapor mole fraction above the tray i. V and L are vapor and liquid molar flow rates. The two liquid compositions measured at the tray 2 (X_2 now represented as Y_2) and tray 4 (X_4 now represented as Y_4) are taken as the output variables to be modelled by NARX, and are depicted on five stage distillation column diagram on Fig. 2. The values of the parameters are found from the literature [44,45] and are taken as: $L = 27.3755$ mol (min)$^{-1}$, $H_c = 30$ mol, $H_r = 30$ mol, $H_t = 20$ mol, $\alpha = 5$, and V=32.3755 mol (min)$^{-1}$ and are utilised to numerically solve the first principle equations of the distillation column.

The two input variables, X_f and L_f, are varied in the operating range, i.e., 0.5–0.7 and 6 to 10 as reported in literature [44,45]. The step-size of 0.01 and 0.2 is taken for X_f and L_f respectively, and 20 experiments are constructed for the input variables. Subsequently, the developed ODEs of the distillation column are numerically solved in MATLAB 2021b version using ode23 solver. The dynamic profiles of the two output variables, Y_2 and Y_4 are retained for two time-step values, i.e., $t = 13, 26$. Thus, the simulated datasets consisting of the causal inputs and the output variables and having 40 observations are normalised into −1 to 1 scale and are deployed to develop the NARX model for Y_2 and Y_4.

NARX network is trained on the data simulated by the ODEs of the distillation column. Levenberg Marquardt algorithm is deployed for the parametric optimisation of the network and sum-of-square-error is used as loss function [46]. The activation function applied at the hidden and output layer of NARX is tangent sigmoidal and linear respectively [47]. Various combinations of the delays (input and feedback) and hidden layer neurons are tried for the NARX model development. The performance metrics constructed on R^2 and RMSE are measured corresponding to the developed NARX network. Fig. 2 shows the performance metrics computed for the NARX network of Y_2 and Y_4 under various architectural configuration (hidden layer neurons, input delay, feedback delay). The performance metrics of the developed networks are compared. It is found that NARX network with five hidden layer neurons and one feedback delay has the comparatively improved values of R^2 and RMSE, i.e., 1 and 0.00076 respectively for Y_2. Similarly, the optimal architectural configuration developed for Y_4 has four hidden layer neurons and

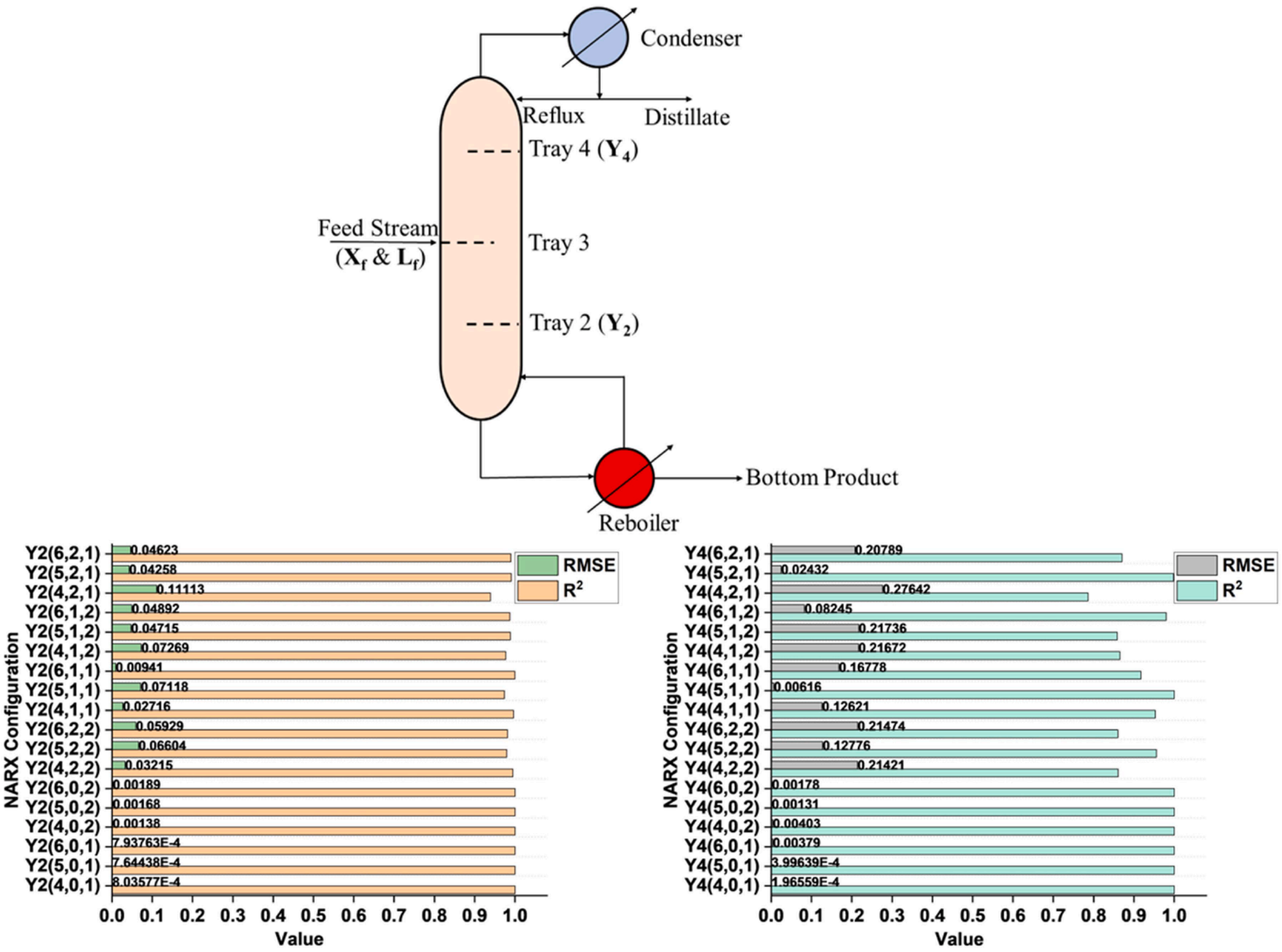

Fig. 2. Graphical visualisation of five stage distillation column where X_f and L_f are deployed to model Y_2 and Y_4 by NARX. Development of NARX network for Y_2 and Y_4 under different configurations (hidden layer neurons, input delay, feedback delay) is presented. Y_2 (5,0,1) and Y$_4$ (4,0,1) achieved comparatively higher R^2 and lower RMSE values than those of other NARX networks having different structural configuration.

one feedback delay with R^2 value of 1 and RMSE is 0.00020. The two NARX models have comparatively higher merit of performance in modelling the two output variables and are deployed for conducting the partial-derivative based sensitivity analysis.

The partial derivative-based sensitivity of the input over the output variable is computed using Eq. (20) for the NARX model. The weight vector of the input variable from input layer to the hidden layer of NARX is compiled. Moreover, the weight vector from the hidden layer neurons to the output neuron is constructed. Similarly, the term $(1 - N^2)$ is computed using the Eq. (19). The dynamic sensitivity profile of the input variables, i.e., X_f and L_f is evaluated over the output variables (Y_2 and Y_4) for two time step values. Similarly, the dynamic sensitivity profile of the input variables over the two output variables is also developed by the forward difference method applied on the ODEs (Eqs. (21)–(26)).

Fig. 3 compares the sensitivity of two output variables, i.e., Y_2 and Y_4 towards the input variables (X_f and L_f) at $t = 13,26$. The dynamic sensitivity profiles of the two output variables are constructed at the four operating points of X_f and L_f taken as 0.54–0.57 and 7.6–8.8 with the step size of 0.01 and 0.4 respectively. During the evaluation of dynamic sensitivity of Y_2 and Y_4 with respect to X_f, L_f is kept at 8. Similarly, X_f is set at 0.55 to investigate the dynamic sensitivity of two output variables towards L_f.

A general increasing trend in the dynamic sensitivity of the two output variables is observed with respect to X_f and L_f computed by partial derivative method on NARX model and forward difference method on ODEs (first principle). Y_2 appears to be more sensitive to X_f since the dynamic sensitivity values are comparatively bigger than that of L_f. However, Y_4 has higher dynamic sensitivity to L_f compared with that of X_f. It is important to note that the two sensitivity analysis methods present the comparable and similar dynamic sensitivity trends for the output variables which are computed with respect to the input variables. Furthermore, the similar significance order of the input variables is established by the two techniques confirming the accurate interpretability performance of the NARX model to predict the values of the output variables as investigated by the derived partial-based sensitivity expression.

Initially, the dynamic sensitivity profiles of Y_2 and Y_4 with respect to X_f and L_f are constructed relative to the time scale for $t = 13$, 26. Fig. 4 presents sensitivity trend of Y_4 plotted with respect to X_f and L_f at $t = 13$, 26 computed by first principle and partial-derivative approach applied on NARX. An increasing dynamic sensitivity trend is observed for X_f and L_f when time is kept at 13 and 26. The sensitivity values at the two-time steps are in good agreement for two sensitivity analysis techniques indicating the good functional mapping created in the trained NARX model for the predictive analysis.

4.2. Development of NARX model and its partial-derivative based sensitivity analysis for higher distillation column

A higher order distillation column comprising on ten stages including reboiler, condenser and feed entering at stage five is considered to implement the partial-derivative method for the sensitivity analysis of NARX model. The liquid composition at stage seven is taken as output variable (X_7 now taken as Y_7) to be modelled by *t*, X_f and L_f and

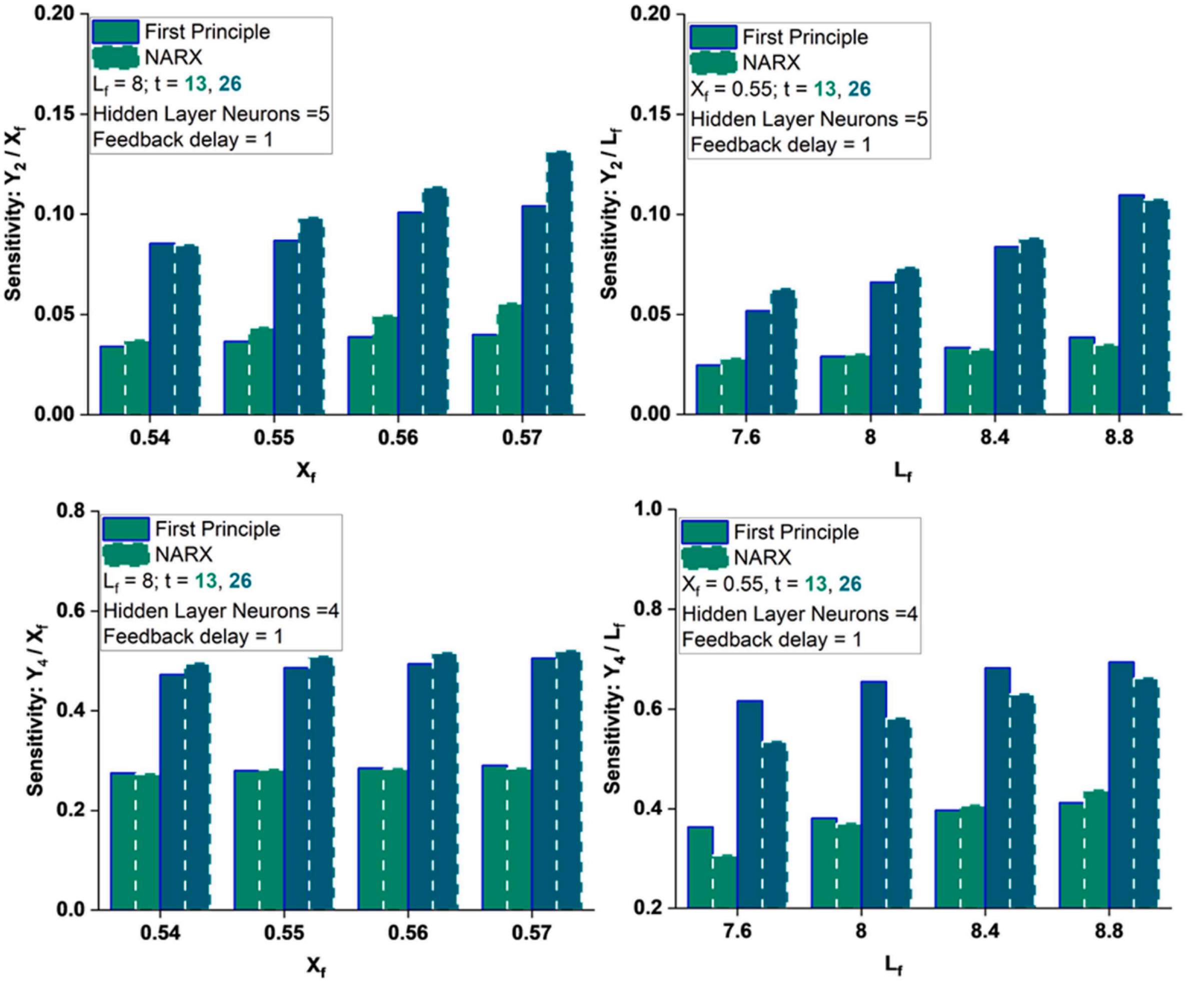

Fig. 3. Comparison of partial-derivative based sensitivity analysis of developed NARX models and first principle method for Y_2 and Y_4 with respect to X_f and L_f. A good match is observable for the dynamic sensitivity values of Y_2 and Y_4 against the input variables.

also shown on Fig. 5. The mathematical model of the considered higher distillation column is given as follows:

$$\text{Reboiler}: \; H_r\frac{dX_1}{dt} = (L+L_f)(X_2 - X_1) + V(X_1 - y_1) \tag{27}$$

$$\text{Tray 2}: \; H_t\frac{dX_2}{dt} = (L+L_f)(X_3 - X_2) + V(y_1 - y_2) \tag{28}$$

$$\text{Tray 3}: \; H_t\frac{dX_3}{dt} = (L+L_f)(X_4 - X_3) + V(y_2 - y_3) \tag{29}$$

$$\text{Tray 4}: \; H_t\frac{dX_4}{dt} = (L+L_f)(X_5 - X_4) + V(y_3 - y_4) \tag{30}$$

$$\text{Feed tray}: \; H_t\frac{dX_5}{dt} = L_fX_f + LX_6 - (L+L_f)X_5 + V(y_4 - y_5 \tag{31}$$

$$\text{Tray 6}: \; H_t\frac{dX_6}{dt} = L(X_7 - X_6) + V(y_5 - y_6) \tag{32}$$

$$\text{Tray 7}: \; H_t\frac{dX_7}{dt} = L(X_8 - X_7) + V(y_6 - y_7) \tag{33}$$

$$\text{Tray 8}: \; H_t\frac{dX_8}{dt} = L(X_9 - X_8) + V(y_7 - y_8) \tag{34}$$

$$\text{Tray 9}: \; H_t\frac{dX_9}{dt} = L(X_{10} - X_9) + V(y_8 - y_9) \tag{35}$$

$$\text{Condenser}: \; H_c\frac{dX_{10}}{dt} = V(y_9 - X_{10}) \tag{36}$$

The vapour–liquid equilibrium maintained in the distillation column is expressed as:

$$y_i = \frac{\alpha X_i}{1 + (\alpha - 1)X_i} \tag{37}$$

The values of parameters are same as considered for distillation column in the previous Section 4.1 and are deployed for solving the ODEs of higher distillation columns in MATLAB 2021b version using ode23 solver. X_f and L_f are varied from 0.5 to 0.7 and 6 to 10 respectively with the step size of 0.01 and 0.4 thereby making 20 input conditions. The simulated values of Y_7 are retained and taken corresponding to $t = 0.39,\ 1.0,\ 1.44,\ 2.07$ thereby making 80 observations for the input–output dataset.

The simulated dataset is normalised into −1 to 1 scale and is deployed for modelling Y7 on the input variables by NARX algorithm. Tangent sigmoidal and linear activation function are implemented at the hidden and output layer of NARX respectively. Various combinations of the delays (input and feedback) and hidden layer neurons are tried for the NARX model development. R^2 and RMSE are calculated corresponding to the architecture of the network. Fig. 5 shows the modelling performance of the NARX networks developed under different combination of hidden layer neurons, input delay and feedback delay. R^2 value is observed from 0.62 to 1.0 whereas RMSE is varied from 0.0168 to 0.312. Closely comparing the performance metrics of the developed

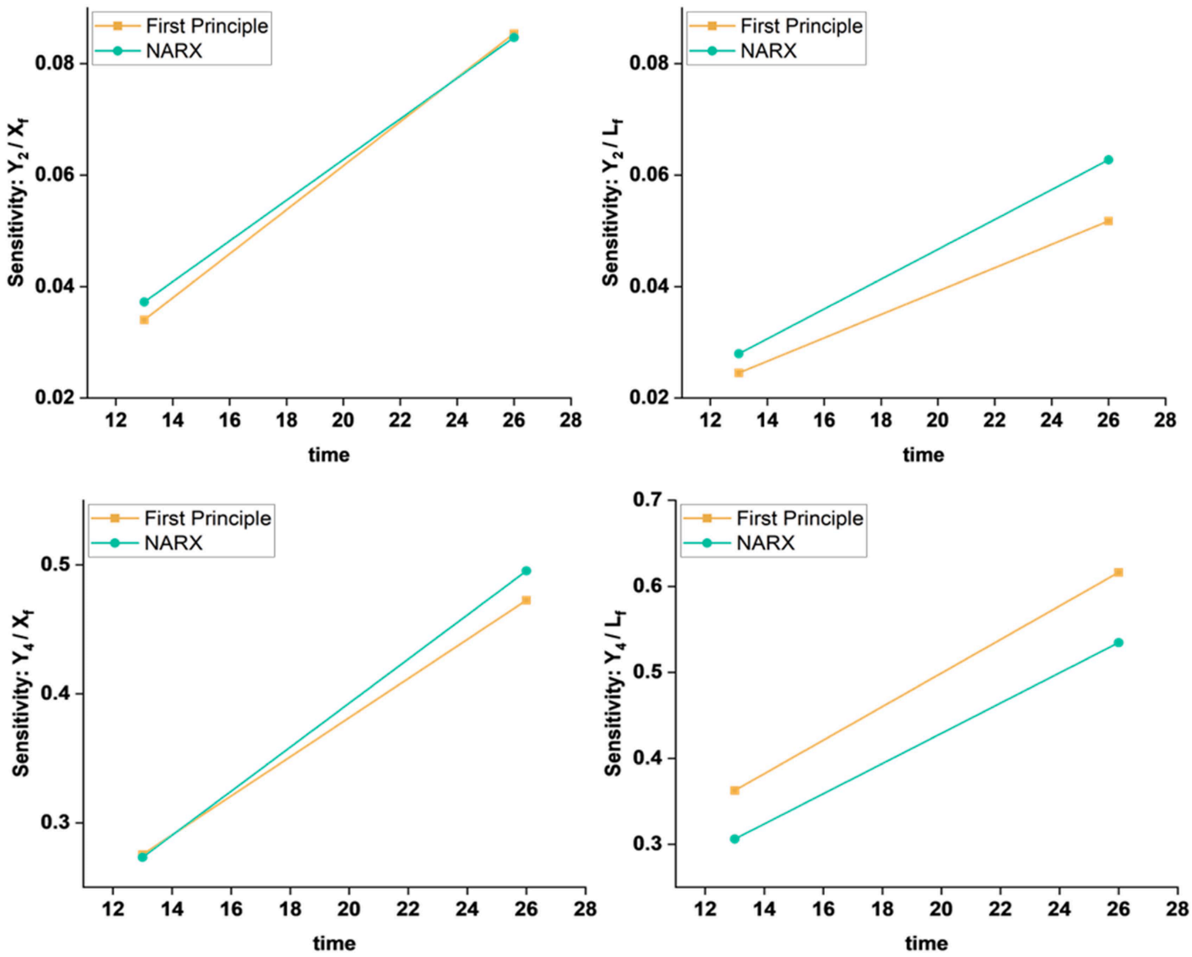

Fig. 4. Comparison of the dynamic sensitivity trend of Y_2 and Y_4 against X_f and L_f as based upon the first principle and NARX method. A reasonably good match among the dynamic sensitivity values and pointing of the trend is observable for the two approaches.

models, it is found that NARX network with four hidden layer neurons and one feedback delay has comparatively better values of performance metrics, i.e., $R^2 = 1$ and RMSE = 0.0168. Thus, the developed NARX model is deployed to evaluate the sensitivity of Y_7 towards the input variables by partial-derivative method.

The developed NARX network for modelling Y_7 is deployed for conducting the partial derivative-based sensitivity analysis. The weight matrices w_{oh} and w_{hp} are compiled and the term $(1 - N^2)$ is computed on the data deployed for conducting the sensitivity analysis. Fig. 6 shows the sensitivity of Y_7 towards the input variables at four-time step values computed by partial derivative approach on NARX network and forward difference method on the ODEs (first principle) of higher distillation column. The dynamic sensitivity of output variable is evaluated for $X_f =$ 0.54 to 0.7 and L_f = 7.6 to 8.8 with the step size of 0.01 and 0.4 respectively at $t =$ 0.39, 1.0, 1.44, 2.07. Similarly, L_f and X_f is taken as 8 and 0.6 during dynamic sensitivity evaluation with respect to X_f and $L_f$respectively. A non-linear dynamic sensitivity trend is observed for Y_7 with respect to X_f and L_f for four-time step values as shown on Fig. 6. There exists a good agreement between the NARX based partial derivative approach and forward difference method on the ODEs for the sensitivity analysis indicating the accuracy of the derived partial-derivative expression for the NARX model. Another important aspect to note here is the significance order of the input variables towards the sensitivity of Y_7. It is apparent from Fig. 6 that Y_7 is more sensitive to L_f as compared to X_f since the sensitivity values computed by the two techniques for the output variable are higher with respect to L_f than those of X_f. The correct significance order of input variables is established by the partial-derivative based approach offering the accurate interpretability performance of NARX model. This further confirms the accurate interpretability performance of the NARX model as computed through leveraging the mathematical rigor of partial-derivative approach.

The dynamic sensitivity of Y_7 to X_f and L_f is visualised with respect to time scale and presented in Fig. 7. The dynamic sensitivity trend is plotted corresponding to $X_f = 0.54$ and $L_f = 8.8$ for time scale: $t = 0.39$, 1, 1.44, 2.07. A nonlinear sensitivity trend of Y_7 initially increased from t = 0.39 to t = 1.0 and subsequently decreased until t = 1.44 and finally increased up to t = 2.07 for both input variables, i.e., X_f and L_f. The two sensitivity analysis approaches, i.e., partial derivative method of NARX model and first principle method on ODEs present the closer sensitivity values and follows the trend in good agreement.

5. Conclusion

In this paper, we have derived the partial derivative based explicit dynamic sensitivity analysis expression for the non-linear auto regressive with exogenous (NARX) model. The derived expression for the dynamic sensitivity analysis is applied on two engineering system-based case studies, i.e., distillation column and higher distillation column. The input variables of the distillation column, i.e., t, X_f and L_f are deployed to model liquid mole fraction corresponding to second & fourth stage of distillation column (Y_2 & Y_4), and liquid mole fraction corresponding to seventh stage of higher distillation column (Y_7).

- Y_2 is appeared to be relatively more sensitive to X_f than L_f, whereas L_f is relatively more significant towards the dynamic sensitivity of Y_4 as

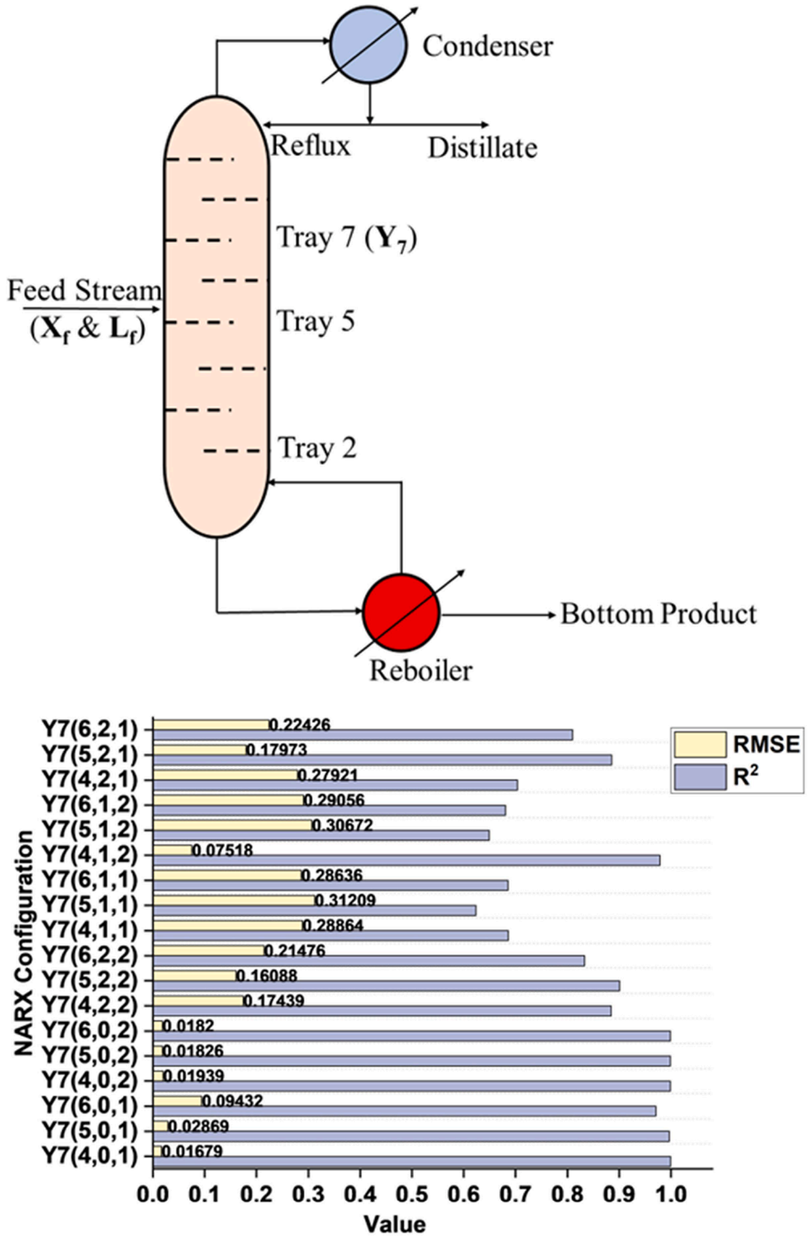

Fig. 5. Graphical visualisation of ten stage distillation column where X_f and L_f are deployed to model Y_7 by NARX development of NARX network to model Y_7 of higher-order distillation column under various architectural configurations (hidden layer neurons, input delay, feedback delay) is presented. Y_7 (4,0,1) achieved R^2 value of 1 and RMSE value of 0.0168 representing improved performance metrics in comparison with those of other NARX networks having different configurations.

evaluated on partial-derivative approach on NARX model and first principle method. Moreover, higher relative sensitivity of Y_7 towards L_f as compared with that of X_f is observed as analysed by partial-derivative and first-principle approach.

- The comparison of the sensitivity values computed from the NARX model by partial-derivative and first-principle approach allows to investigate the interpretation performance of the NARX model. We observe the similar order of significance of the input variables towards the output variables of distillation columns as established by partial-derivative and first-principle approach that confirms the good interpretation performance analysed by partial-derivative approach on the NARX model.
- This research presents the explicit expression derived by the partial derivative approach to plot the dynamic sensitivity trends for the NARX model that is helpful to explain the interpretability performance of the NARX model. Furthermore, the partial-derivative based sensitivity expression can help get the insight of the impact of causal variables on the dynamic operations of system under investigation. The current work presents the dynamic sensitivity information corresponding to the time-steps upon which the NARX model is trained. In the future work, the dynamic sensitivity values can be computed at different values of the time and the information can be used for the control and decision-making for the dynamic operations of real-life applications.

Funding

The Punjab Education Endowment Fund (PEEF) provides the funding (funding number: PEEF/CMMS/21/142) to Waqar Muhammad Ashraf to pursue his PhD from University College London.

CRediT authorship contribution statement

Waqar Muhammad Ashraf: Writing – original draft, Software,

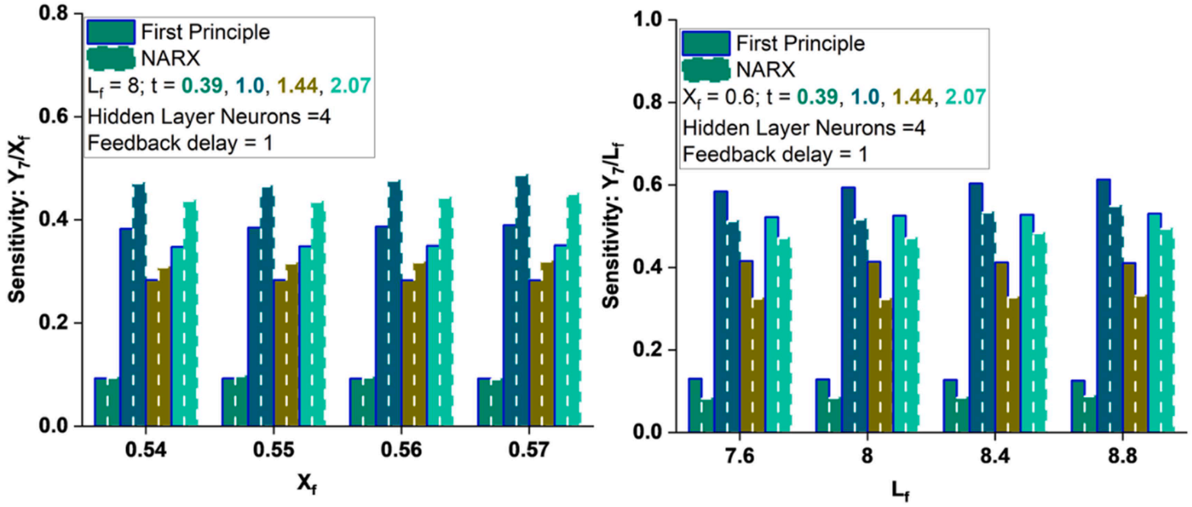

Fig. 6. Comparison of partial-derivative based sensitivity analysis of developed NARX model by partial-derivative and first principle method for Y_7. The dynamic sensitivity values are computed for four time-step values. A good comparison is observable for the computed sensitivity values of Y_7 with respect to X_f and L_f by the two approaches.

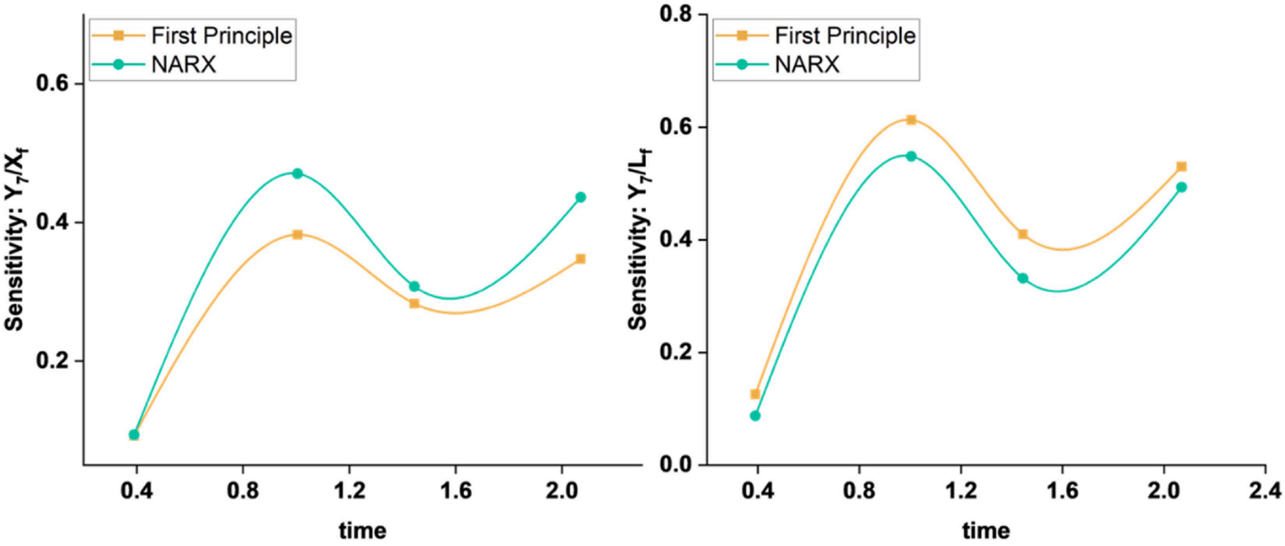

Fig. 7. Comparison of the dynamic sensitivity trend plotted by partial-derivative approach on NARX model and first-principle method for Y_7 against X_f and L_f. A reasonably good match in the computed sensitivity values as well as dynamic sensitivity trend lines is observed.

Methodology, Investigation, Data curation. **Vivek Dua:** Writing – review & editing, Supervision, Project administration, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] M.W. Shahzad, et al., Energy-water-environment nexus underpinning future desalination sustainability, Desalination 413 (2017) 52–64.

[2] M. Ghobakhloo, Industry 4.0, digitization, and opportunities for sustainability, J. Clean. Prod. 252 (2020) 119869.

[3] C.P. Chen, C.-Y. Zhang, Data-intensive applications, challenges, techniques and technologies: a survey on Big Data, Inf. Sci. (Ny) 275 (2014) 314–347.

[4] P. Valduriez, et al., Scientific data analysis using data-intensive scalable computing: the scidisc project, in: LADaS: Latin America Data Science Workshop, 2018. CEUR-WS. org.

[5] Z. Sun, L. Sun, K. Strang, Big data analytics services for enhancing business intelligence, J. Comput. Inf. Syst. 58 (2) (2018) 162–169.

[6] J. Krzywanski, et al., Modelling of SO2 and NOx emissions from coal and biomass combustion in air-firing, oxyfuel, iG-CLC, and CLOU conditions by fuzzy logic approach, Energies (Basel) 15 (21) (2022) 8095.

[7] J. Krzywanski, et al., Towards enhanced heat and mass exchange in adsorption systems: the role of AutoML and fluidized bed innovations, Int. Commun. Heat Mass Transf. 152 (2024) 107262.

[8] K.T. Butler, et al., Machine learning for molecular and materials science, Nature 559 (7715) (2018) 547–555.

[9] W. Quaghebeur, I. Nopens, B. De Baets, Incorporating unmodeled dynamics into first-principles models through machine learning, IEEE Access 9 (2021) 22014–22022.

[10] J. Krzywanski, W. Nowak, Artificial intelligence treatment of SO 2 emissions from CFBC in air and oxygen-enriched conditions, J. Energy Eng. 142 (1) (2016) 04015017.

[11] J. Bourquin, et al., Advantages of Artificial Neural Networks (ANNs) as alternative modelling technique for data sets showing non-linear relationships using data from a galenical study on a solid dosage form, Eur. J. Pharm. Sci. 7 (1) (1998) 5–16.

[12] F. Hajabdollahi, Z. Hajabdollahi, H. Hajabdollahi, Soft computing based multi-objective optimization of steam cycle power plant using NSGA-II and ANN, Appl. Soft Comput. 12 (11) (2012) 3648–3655.

[13] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, Nature 323 (6088) (1986) 533–536.

[14] Y.P. Lin, R. Dhib, M. Mehrvar, ARX/NARX modeling and PID controller in a UV/H2O2 tubular photoreactor for aqueous PVA degradation, Chem. Eng. Res. Des. 195 (2023) 286–302.

[15] E. Heidari, et al., Prediction of the droplet spreading dynamics on a solid substrate at irregular sampling intervals: nonlinear auto-regressive eXogenous artificial neural network approach (NARX-ANN), Chem. Eng. Res. Des. 156 (2020) 263–272.

[16] C.E. de Araújo Padilha, et al., Recurrent neural network modeling applied to expanded bed adsorption chromatography of chitosanases produced by Paenibacillus ehimensis, Chem. Eng. Res. Des. 117 (2017) 24–33.

[17] P. Azadi, et al., A hybrid dynamic model for the prediction of molten iron and slag quality indices of a large-scale blast furnace, Comput. Chem. Eng. 156 (2022) 107573.

[18] S. Haykin, Neural Networks and Learning Machines, 3/E, Pearson Education India, 2009.

[19] J.M. Benítez, J.L. Castro, I. Requena, Are artificial neural networks black boxes? IEEE Trans. Neural Netw. 8 (5) (1997) 1156–1164.

[20] Z. Zhang, et al., Opening the black box of neural networks: methods for interpreting neural network models in clinical applications, Ann. Transl. Med. 6 (11) (2018).

[21] S.L. Özesmi, U. Özesmi, An artificial neural network approach to spatial habitat modelling with interspecific interaction, Ecol. Modell. 116 (1) (1999) 15–31.

[22] Garson, D.G., Interpreting neural network connection weights. (1991).

[23] M. Scardi, L.W. Harding Jr, Developing an empirical model of phytoplankton primary production: a neural network case study, Ecol. Modell. 120 (2–3) (1999) 213–223.

[24] S. Lek, et al., Application of neural networks to modelling nonlinear relationships in ecology, Ecol. Modell. 90 (1) (1996) 39–52.

[25] M.W. Beck, NeuralNetTools: visualization and analysis tools for neural networks, J. Stat. Softw. 85 (11) (2018) 1.

[26] Y. Dimopoulos, P. Bourret, S. Lek, Use of some sensitivity criteria for choosing networks with good generalization ability, Neural Process. Lett. 2 (6) (1995) 1–4.

[27] I. Dimopoulos, et al., Neural network models to study relationships between lead concentration in grasses and permanent urban descriptors in Athens city (Greece), Ecol. Modell. 120 (2–3) (1999) 157–165.

[28] A. Muñoz, T. Czernichow, Variable selection using feedforward and recurrent neural networks, Eng. Intell. Syst. Electr. Eng. Commun. 6 (2) (1998) 91–102.

[29] H. White, J. Racine, Statistical inference, the bootstrap, and neural-network modeling with application to foreign exchange rates, IEEE Trans. Neural Netw. 12 (4) (2001) 657–673.

[30] M.T. Ribeiro, S. Singh, C. Guestrin, "Why should i trust you?" Explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.

[31] K. Futagami, et al., Pairwise acquisition prediction with SHAP value interpretation, J. Finance Data Sci. 7 (2021) 22–44.

[32] M. Gevrey, I. Dimopoulos, S. Lek, Review and comparison of methods to study the contribution of variables in artificial neural network models, Ecol. Modell. 160 (3) (2003) 249–264.

[33] R. Kumar, A.K. Singh, Chemical hardness-driven interpretable machine learning approach for rapid search of photocatalysts, NPJ Comput. Mater. 7 (1) (2021) 197.

[34] S. Zhao, et al., Interpretable machine learning for predicting and evaluating hydrogen production via supercritical water gasification of biomass, J. Clean. Prod. 316 (2021) 128244.

[35] I.J. Leontaritis, S.A. Billings, Input-output parametric models for non-linear systems part I: deterministic non-linear systems, Int. J. Control 41 (2) (1985) 303–328.

[36] I. Leontaritis, S.A. Billings, Input-output parametric models for non-linear systems part II: stochastic non-linear systems, Int. J. Control 41 (2) (1985) 329–344.

[37] W.M. Ashraf, V. Dua, Machine learning based modelling and optimization of post-combustion carbon capture process using MEA supporting carbon neutrality, Digit. Chem. Eng. 8 (2023) 100115.

[38] W.M. Ashraf, V. Dua, Artificial intelligence driven smart operation of large industrial complexes supporting the net-zero goal: coal power plants, Digit. Chem. Eng. 8 (2023) 100119.

[39] M.W. Shahzad, et al., Multi effect desalination and adsorption desalination (MEDAD): a hybrid desalination method, Appl. Therm. Eng. 72 (2) (2014) 289–297.

[40] K.C. Ng, et al., Recent developments in thermally-driven seawater desalination: energy efficiency improvement by hybridization of the MED and AD cycles, Desalination 356 (2015) 255–270.

[41] D. Ibrahim, M. Jobson, G. Guillén-Gosálbez, Optimization-based design of crude oil distillation units using rigorous simulation models, Ind. Eng. Chem. Res. 56 (23) (2017) 6728–6740.

[42] S. Fraser, Distillation in refining, Distillation (2014) 155–190.

[43] M. Waheed, A. Oni, Performance improvement of a crude oil distillation unit, Appl. Therm. Eng. 75 (2015) 315–324.

[44] K.H. Rasmussen, S.B. Jørgensen, Parametric uncertainty modeling for robust control: a link to identification, Comput. Chem. Eng. 23 (8) (1999) 987–1003.

[45] V. Prasad, B.W. Bequette, Nonlinear system identification and model reduction using artificial neural networks, Comput. Chem. Eng. 27 (12) (2003) 1741–1754.

[46] H. Yu, B.M. Wilamowski, Levenberg–Marquardt Training, in Intelligent Systems, CRC Press, 2018, p. 12-1-12-16.

[47] W.M. Ashraf, et al., Artificial Intelligence Modeling-Based Optimization of an Industrial-Scale Steam Turbine for Moving toward Net-Zero in the Energy Sector, ACS Omega, 2023.