Nonlinear-Control-Oriented Modelling of the Multi-Variable UCG process for Underground Coal Gasification Project Thar: A Machine Learning Perspective

Afaq Ahmed[a,], Ali Arshad Uppal[a], Syed Bilal Javed[b]

[a]*Department of Electrical and Computer Engineering, COMSATS University Islamabad, Pakistan*

[b] *Centers of Excellence in Science & Applied Technologies, Islamabad, Pakistan*

Abstract

The Underground Coal Gasification (UCG) process is a complex and multi-physics phenomenon, thus making it difficult to develop a mathematical model that encapsulates all dynamical aspects. In this regard, data-driven modeling techniques offer a reliable alternative for prediction, control, and optimization of dynamical systems, but their application in UCG is still in the early stages. This work aims to bridge this gap by implementing three cutting-edge nonlinear identification structures: Non-linear Autoregressive with Exogenous Inputs (NARX), Hammerstein-Wiener (HW), and State-Space Neural Networks (SSNN) on the UCG process to obtain a multivariable control-oriented model. The contributions include synthesizing an excitation signal for data acquisition, outlining the non-linear system identification procedure, and comparing predictive capabilities using statistical tools. The simulation results demonstrate a rigorous comparison of various

*Corresponding author.*E-mail address:* ali_arshad@comsats.edu.pk.

techniques for the heating value and flowrate of the syngas, which are the outputs of the UCG process. The results of the analysis show that NARX outperforms other structures in statistical metrics, with MAE, RMSE, and Best fit values of $1.51, 1.9$, and 0.9, respectively, for the heating value; and $0.25, 0.31$, and 0.94, respectively, for the flowrate. Consequently, the outputs of the NARX model are compared with the experimental data obtained from the UCG project Thar, which show a good match for both outputs.
Keywords: Underground Coal Gasification (UCG), Non-linear sytem identification, Data-driven modeling, Energy conversion process, Machine Learning, Thar coal field

1. Introduction

Underground Coal Gasification (UCG) is an alternative method to conventional coal mining that uses drilled wells to access coal beds. Its working principle involves the injection of input gas mixture – see figure 1–, the inputs of the process are composition (steam to Oxygen ratio) and flowrate of the injected gases. Upon its interaction with already ignited coal bed, the inlet gas mixture initiates a series of chemical reactions, which mainly include pyrolysis, coal combustion and gasification. As a result syngas (hydrogen, carbon monoxide, methane, carbon dioxide, and nitrogen etc) is produced which is recovered from the production well. The output of the process are heating value and flowrate of the syngas. The recovered gas can be further processed for power generation and other industrial applications [1, 2].

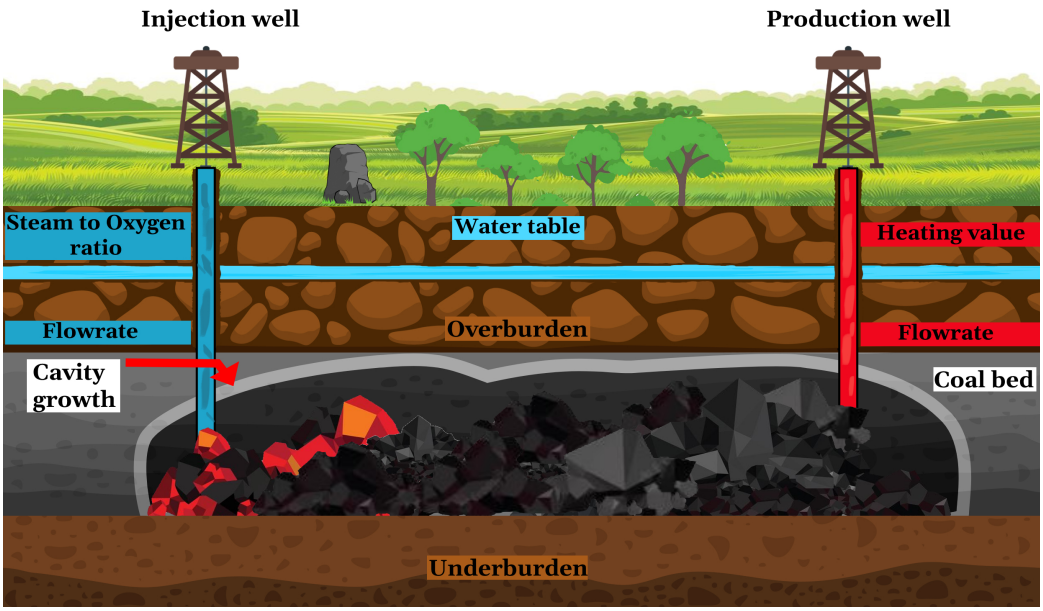


Figure 1: UCG process schematic

Because there are intricate chemical interactions and physical phenomena involved, accurate modelling is essential for safe and effective operation. In this regard, two modelling approaches are available: first-principle-based and data-driven modelling. First-principle-based models incorporate laws of nature, whereas data-driven models adopt the entirely different strategy: formulating a mathematical map between input and output data of dynamical system. However, this process is complex due to the lack of a-priori knowledge of certain parameters and the structure of the model.

1.1. Motivation

In our previous works, we have mainly explored the process models based on the first principle and control-oriented models based on data-driven techniques. For instance, in [3], a non-linear, time-dependent UCG model was

formulated by assuming certain assumptions in the existing models of [4] and [5]. Moreover, in [6], the analytical model considered was a set of partial differential equations; despite giving accurate predictions of UCG output, its 1D formulation constrained its ability to handle other process variables, such as cavity growth and its shape. In [7], we parametrized and validated the Cavity Simulation Model (CAVSIM), which was developed by [8], with Underground Coal Gasification Project Thar (UPT) field data. CAVSIM is a high-fidelity, 3D UCG process simulator that can predict the UCG plant's outputs and other process variables, such as cavity growth and its interaction with environment, to a high degree of accuracy.

However, CAVSIM is represented as the bulk of computational routines in FORTRAN, making it unsuitable for synthesising a model-based multi-variable controller for the UPT field. Thus, a data-driven linear model of the UPT gasifier was derived using the subspace N4SID technique with data from CAVSIM [7]. While this model yielded statistically acceptable results for the UCG process, it could only handle a limited range of inputs due to its linear nature. However, in practical scenarios, the UCG reactor is driven across a wide range of inputs to maximize output gas yield. Therefore, a linear model-based controller does not ensure desired UCG system operation. Hence, selecting a control-oriented, non-linear model is necessary to overcome these constraints.

.

1.2. Related Work

Conforming to the theme of this work, we are going to thoroughly investigate the different modelling strategies undertaken to model the UCG system

with first-principles models in subsection 1.2.1, and data-driven models in subsection 1.2.2

1.2.1. First-Principle based models

In comparison to data-driven approach, there has been strenuous investigations undertaken to model UCG process using first-principle based techniques. Considering the multiphysics nature of this process- chemical kinetics, Computational Fluid Dynamics (CFD), stress and structural analysis, and thermodynamics etc- researchers have been coming up with models that are severely bounded in terms of their scope, assumptions, predictive capabilities, and the relevant physics. The underlying simplifications whilst deriving the model chiefly constitute dimensions and time response of the model; incorporation of multi-physics modules: heat transfer and CFD; predictive capability range: cavity evolution and interaction with the environment. These parameters qualitatively scour the following aspects of a UCG model: the model's dimensions assumed in the problem formulation, i.e., 1D, 2D, or 3D; the model's time response—steady state, transient response, pseudo-steady state, or semi-steady state; the inclusion of various heat transfer mechanisms—conduction, convection, or radiation—by which heat is exchanged between the plant and its surroundings; the physical model, if any, used to incorporate fluid-flow into the plant's dynamics, i.e., Naiver-Stokes equations, Darcy flow, or mixed approach; the capacity to simulate certain physical phenomena such as chemical reactions, thermo-mechanical failure, and bulk collapse; and the provisions to simulate water influx, heat, and mass loss. Based on these criterion the UCG models can be categorized into following groups: packed bed models, channel models, coal block models,

and resource recovery models. Table 1 gives an extensive overview of the first-principle UCG models based on the metrics as identified above.

Packed bed, channel, and coal block models are simpler and more suitable for laboratory setups but cannot predict cavity geometry due to the lack of a thermo-mechanical failure module. Resource-recovery models, on the other hand, are more comprehensive and can be used in actual field setups to predict cavity growth, but at the expense of greater mathematical and computational complexity. Moreover, of the same model's category, CAVSIM is a widely used benchmark simulator in UCG sites worldwide.

1.2.2. Data-Driven based models

Recently, researchers have used data-driven techniques to model the UCG process. Various Machine Learning (ML) algorithms have been employed to capture the underlying dynamics of the system [34–38] in different scenarios, including laboratory-scaled, field-test, and computer simulations. Some researchers have also considered the synthesis of the excitation signal, which is an essential step in system identification. Table 2 summarizes the key components of the system identification process for the UCG system found in the literature.

Table 2 highlights that different ML algorithms can be incorporated to model the UCG process, and there is no all-encompassing model structure that works equally well on all dynamical systems. Similarly, data acquired through field tests tends to yield accurate outputs as important process variables can be captured. However, which process variables get excited depends entirely on the information content of the excitation signal.

Authors	Year	Model Formulation		Heat transfer			Mass Diffusion	Fluid flow	Cavity Evolution			Interaction with environment		
		Dimension	Time dependence	Conduction	Convection	Radiation			Chemical reactions	Thermomechanical failure	Bulk collapse	Water influx	Heat loss	Mass loss
Packed bed models														
Gunn & Whitman [9]	1976	1D	PS	✓					✓					
Winslow [10]	1977	1D	T	✓	✓			D	✓					
Thorsness et al. [11]	1978	1D	T	✓	✓		✓	D	✓					
Thorsness & Kang [12]	1985	2D	T	✓	✓		✓	D	✓	✓		✓		
Abdel Hadi & Hsu [13]	1987	1D	T	✓	✓		✓	D	✓					
Khadse et al. [14]	2006	1D	PS	✓	✓				✓					
Uppal et al. [6]	2014	1D	PS	✓	✓				✓					
Channel models														
Magnani & Ali [15]	1975	1D		✓	✓				✓					
Pasha et al. [16]	1978	2D	T	✓	✓			P	✓	✓				
Dinsmoor et al. [17]	1978	1D	T	✓	✓	✓	✓	M	✓			✓	✓	
Eddy et al. [18]	1983	1D	T	✓	✓	✓	✓	NS	✓			✓		
Kuyper & Van [19]	1994	2D	T	✓	✓	✓	✓	D	✓			✓	✓	
Batenburg [20]	1995	1D	SS	✓	✓	✓	✓	D	✓			✓		
Pirlot et al. [21]	1998	2D	S	✓	✓		✓	M	✓			✓	✓	
Perkins & Saha [22]	2008	2D	T	✓	✓	✓	✓	P	✓					
Luo et al. [23]	2009	2D	T	✓	✓	✓			✓					
Seifi et al. [24]	2013	1D	S	✓	✓									
Coal bed models														
Tsang [25]	1980	1D	T	✓	✓	✓		D	✓					
Massaquoi & Riggs [26]	1983	1D	S	✓			✓	D	✓				✓	
Park & Edgar [27]	1987	1D	T	✓	✓		✓	NS	✓			✓	✓	
Perkins & Sahajwalla [28]	2005	1D	PS	✓	✓	✓			✓	✓				
Resource recovery model														
Britten & Thorsness [29]	1989	2D	T	✓	✓	✓		M	✓	✓		✓	✓	✓
Biezen et al. [30]	1996	3D	PS	✓	✓	✓	✓	D	✓	✓	✓	✓	✓	✓
Nitao et al. [31]	2011	3D	T	✓	✓	✓	✓	M	✓	✓		✓	✓	
Samdani et al. [32]	2016	2D	T	✓	✓	✓		M	✓	✓		✓		
Akbarzadeh et al. [33]	2016	3D	T	✓	✓			D						

Table 1: Summary of First principle modeling approach (T: Transient, S: Steady state, P: Pseudo-steady state, SS: Semi-steady state, D: Darcy flow, M: Mixed and NS: Naiver Stokes)

Authors	Year	Machine Learning Algorithm	System Dimensions	Data Acquisition	Excitation Signal
Jan Kacur et al. [34]	2017	Support Vector Machine (SVM)	MISO	Laboratory scaled experiment	-
Alicja Krzemien [35]	2018	Multivariate Adaptive Regression Splines (MARS)	SIMO	Pilot-scaled experiment	-
Yuteng Xiao et al. [36]	2021	Long Short-Term Memory (LSTM)	SISO	Field test	-
Yuteng Xiao et al. [37]	2020	Convulution Neural Network (CNN) & Long Short-Term Memory (LSTM)	SISO	Field test	-
S.B Javed et al. [38]	2021	State-space model(N4SID)	MIMO	Simulation	PRBS

Table 2: Summary of Data-driven modeling approach (MISO: Multiple-inputs single-output, SIMO: Single-input multiple-outputs, SISO: Single-input single-output, MIMO: Multiple-inputs multiple-outputs and PRBS: Pseudo Random Binary Sequence)

1.3. Gap analysis

High fidelity, first-principle models like CAVSIM are not well suited for model-based controller design precisely due to their complex structure. On the contrary, the application of data-driven approach to UCG is in seminal state. As the landscape of non-linear identification is vast and enriched with the fusion of diversified fields, any lapse in any step can lead to erroneous results; table 2 signifies that in almost all case studies a crucial step in system identification for UCG is amiss: synthesis of the excitation signal. Thus, training a ML algorithm without appropriate data acquisition will perceive the problem as time-series prediction, rather than capturing the dynamics of the process. Despite the fact that the data-driven techniques have demonstrated exceptional performance in modelling a wide range of dynamical systems such as aircraft [39], turbo-engines [40], and gas turbines [41], it has not been widely adopted for a UCG plant. This approach, as applicable to a UCG system, is still in a nascent state.

1.4. Major contributions

Keeping in mind, any backsliding in the non-linear system identification process and its subsequent applications in controller design for a UCG system as inspired from the related work, the contributions of this work are iterated as follows.

- Design of excitation signal: Amplitude Modulated Pseudo Random Binary Signal (APRBS) for non-linear system identification of the Thar coal UCG process.
- Derivation, identification, and validation of three different model structures- Nonlinear Autoregressive Network with Exogenous Inputs (NARX), Hammerstein-Wiener (HW), and State Space Neural Networks (SSNN).
- Explanation of the working principle of system identification process applied to UCG, ensuring reusability and scalability.
- Control theory applications are discussed along with design limitations and how easily model structures may be created.

The rest of the paper is arranged in the following manner. The nonlinear system identification of the UPT gasifier is discussed in section 2. The results and discussions are presented in section 3, a comprehensive qualitative comparison of the identified models is given in 4, and the article is concluded in section 5.

2. Non-linear System Identification of the UCG Process

System identification is an important modelling paradigm that attempts to capture the important dynamics of the physical system by setting up a mathematical map. Though much involved, a holistic overview of the system identification process is illustrated in figure 2. It is pertinent to mention here that the inputs to the UCG process (CAVSIM in this case) are the flowrate and the composition (steam to Oxygen ratio) of the inlet gas mixture, whereas, the flowrate and the heating value of the syngas are the outputs of the process.

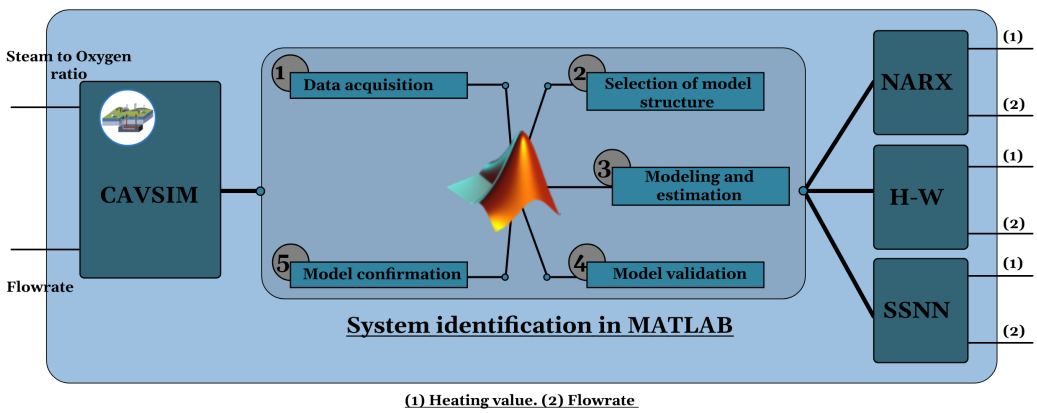


Figure 2: System identification pipeline in MATLAB

2.1. Design of Excitation Signal

The Design of Experiment (DoE) is a critical task in controlling the behavior of a system's states, as it involves synthesizing an excitation signal. For non-linear identification, this task is more complex than for linear identification. Excitation signals can be broadly categorized as impulse, step, ramp,

sinusoidal, PRBS, and random for linear system identification, and chirp, sweep, multisine, burst, and Amplitude Modulated Pseudo Random Binary Sequence (APRBS) for non-linear systems. The selection of signals depends on the context of the design scenario, and each signal has its strengths and weaknesses. Therefore, multiple signals can be used to gain a comprehensive understanding of the system. In this work, an APRBS signal is used due to its ease of design, rich frequency content, and experimental viability [42].

PRBS is commonly used for linear system identification, but it is not appropriate for non-linear problems because it only alternates between two extreme values and cannot provide additional system information. To overcome this limitation, APRBS was proposed by [43], where each step of the PRBS signal is given a different amplitude value. Thus, APRBS is based on the design of PRBS. The PRBS signal is parametrized by several parameters, i.e., signal bandwidth, clock frequency, switching time, sampling time, etc., whose appropriate selection is critical in determining the efficacy of PRBS in exciting the modes of a dynamical system. Though there has been some discussion in the literature guiding the reader in synthesizing PRBS [43, 44], the discussion is mostly scattered; here we attempt to present the guidelines in a unified approach.

2.1.1. Signal Bandwidth

The signal bandwidth is a critical parameter that determines both the frequency range of system excitation and the level of noise present in the signal. Therefore, the bandwidth must be selected based on the system's characteristics. If the system parameters are already known, the bandwidth

can be determined using the following inequality

$$\omega_l = \frac{1}{\beta T_H} \leq \omega \leq \omega_H = \frac{\xi}{T_l}, \tag{1}$$

where, ω_l and ω_H in (rad/s) are the lowest and highest frequencies, T_L (s) and T_H (s) represent the lowest and highest time constants of the dynamical system, while β and ξ represent the settling time parameter and the ratio of open-to-closed-loop time constants, respectively. β s a measure of how fast the system reaches its steady state after being excited by an input signal, whereas ξ is a measure of how much feedback is present in the system. It is pertinent to mention that some prior tests such as staircase experiment, step response need to be conducted to determine the values of β and ξ [45].

2.1.2. Clock Frequency

This parameter dictates how frequently the new values of the PRBS signal are generated by the shift register. Consequently, it plays a vital role in determining some statistical properties, such as frequency content, correlation, noise levels, etc., of the signal. To excite the dynamical systems that have a wide bandwidth, a higher clock frequency is recommended, but this comes at the cost of increased noise level and sampling time, thus increasing the computational cost. On the contrary, if the system's bandwidth is limited, a lower clock frequency may be selected, thus reducing the computational burden. Therefore, keeping the conflicting objectives in mind, a trade-off between sampling rate, bandwidth, and computational cost has to be reached; however, the general rule of thumb in selecting the clock frequency is:

$$f_{ck} = 2.5 f_{int}, \tag{2}$$

where f_{int} (Hz) is maximum frequency of interest.

2.1.3. Switching Time

Switching time $T_{sw}(s)$ reflects the duration for which the signal holds a particular binary value. Consequently, it directly impacts the frequency characteristics of the signal, with shorter switching time resulting in higher frequency components and vice versa. The system's dynamical properties, such as how fast it responds to the input and the frequency range of interest, play a key role in the proper selection of $T_{sw}(s)$. Therefore, $T_{sw}(s)$ should be small enough to be able to excite the system for a range of frequency interest while long enough to reduce the aliasing and noise levels. To ensure that the PRBS signal is sufficiently random and does not exhibit any predictable behaviour, T_{sw} is particularly recommended to be much shorter than the duration of the sequence itself; therefore, it can be selected as follows:

$$\frac{2 \cdot 8\tau_l}{\xi} \leq T_{sw} \leq \frac{2\pi\beta\tau_H}{2^n - 1}, \tag{3}$$

where n is the number of shift registers.

2.1.4. Number of Shift Registers

They are an integral component in generating a PRBS sequence; by creating a feedback loop and repeatingly shifting their input, they generate a sequence of binary values with the desired properties. The number of shift registers (n) used determines the signal's length and period and can affect other parameters such as correlation properties. In a system identification context, the PRBS must desirably have an increasing number of (n) if the range of interested frequencies spans a wide bandwidth. A larger number of (n) implies a long period and a flatter spectrum, thus covering more frequencies. By the same token, fewer (n) are sufficient and computationally

efficient for the system spanning lower bandwidths. Moreover, the length of the signal is $2^n - 1$, implying that it wholly depends on n. The number of shift registers for system identification purposes can be selected as:

$$2^n - 1 = \max\left(\frac{2\pi\beta\tau_H}{T_{sw}}, pt_d\right), \tag{4}$$

where p represents the number of plant's inputs and t_d (s) is the time-delay, to be defined shortly.

2.1.5. Sampling Time

The sampling time t_s is the time interval between successive samples of the PRBS signal, and it is typically chosen to be faster than the switching time of the signal in order to accurately capture its dynamics.

$$t_s = \frac{T_{sw}}{4}. \tag{5}$$

2.1.6. Delay Time

For MIMO systems, multiple input channels can be simultaneously excited with APRBS. To ensure that each channel receives a distinct contribution from excitation, the input signals must be uncorrelated. This can be accomplished by adding a delay t_d to the signal, as shown below:

$$t_d = \frac{5\tau_H}{T_{sw}}. \tag{6}$$

2.1.7. Signal Amplitude

When defining the range of a PRBS signal for non-linear system identification, the amplitude is constrained by economic viability and the saturation points of the actuators to avoid unrealistic operating regions. Once a standard PRBS is generated, steps are randomly counted and are assigned levels

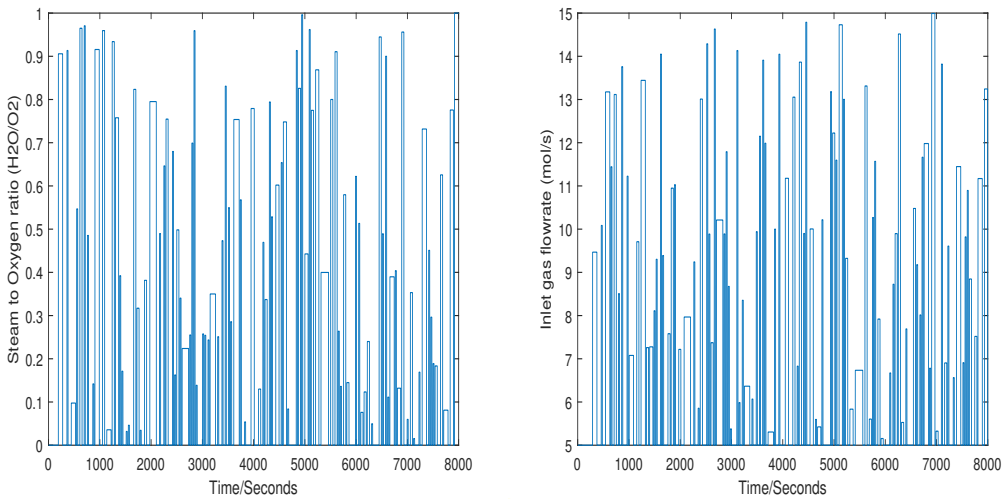
within the interval from maximum to minimum. Increasing the length of the experiment can decrease the number of gaps in the input space to obtain an APRBS signal that fully covers the input space.

In this work, for data acquisition, we have considered CAVSIM as the benchmark. Moreover, due to the limitations of the FORTRAN programming language in which CAVSIM's computational routines are written, CAVSIM had limitations in terms of graphical capabilities and a lack of toolboxes for data acquisition and controller design; therefore, we integrated CAVSIM with MATLAB [46]—the integrated package is available at [47]. This integration helped streamline the system identification process so that its all steps—data collection, data pre-processing, model identification, validation, and analysis—could be carried out in one place—see figure 2.

Based on above guidelines, important parameters derived for APRBS are given in table 3, and the subsequent inputs are given in figures 3a, and 3b.

ω_H	ω_l	f_{ck}	T_{sw}	n	t_s	t_d	β	ξ
0.1 rad/s	0.015 rad/s	0.25 Hz	27 s	4	6.75 s	4 s	3	2

Table 3: APRBS parameters



(a) Excitation signal for Input 1

(b) Excitation signal for Input 2

Figure 3: Plots of excitation signals for UCG system

In figures 3a and 3b, the amplitudes for u_1 (the steam-to-oxygen ratio) and u_2 (flowrate) are selected as per actuator constraints at UPT as $0 \leq u1 \leq 1$, and $5 \leq u2 \leq 15$. Likewise, the signal length is selected so as to encompass the range of input space, as discussed by [44, 48], as follows: signal length $> 2^n - 1 \times T_{sw} = 405$ s.

2.2. Model Estimation using Machine Learning Algorithms

The models considered in this work belong to a certain class of ML algorithms that follow the same structural procedure of training, validation, and testing as outlined in figure 2. The general formulation of the ML algorithm is as follows:

$$D = \{(x_1, y_1), (x_2, y_2) \ldots, (x_n, y_n)\} \subseteq X \times Y, \tag{7}$$

$$(x_i, y_i) \sim P, \tag{8}$$

where (x_n, y_n) are labelled dataset taken from unknow probability distribution P. And the task of the algorithm is to make a prediction y_i against the given input x_i. The output, or prediction, y_i is a function of the structure of the given ML algorithms and varies from algorithm to algorithm. The problem is formulated by defining and finding the minimum of the cost function. In system identification context, where the algorithm predicts a continous numerical value of the y_i, Mean Squared Error (MSE) is the apt choice for the cost function and has the following form

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{9}$$

where n is the total number of samples, y_i is the actual value for the i-th sample, and $\hat{y}_i$ is the predicted value for the i-th sample.

To solve the optimization problem, one of the ubiquitous algorithms used is Levenberg-Marquardt (LM) [49]. It is a second-order training algorithm that is used to solve optimization problems where the cost functions are non-linear. Its approach involves iteratively improving the initial guess of the model's parameters by using a blend of the gradient descent and Newton methods.

In this work, the model structures considered for system identification of the UCG process are the NARX, HW, and SSNN. In the subsequent subsections, we are going to briefly describe their structure and subsequent training and identification.

2.2.1. NARX

It is well known that Neural Networks (NN) are a universal function approximator [50], and exploiting this property of NN, researchers are actively applying them in system identification paradigm. Generally, a dynamical system can be represented either in state-space form or input-output relationship; the latter assumes the following representation

$$yp(k) = \psi(yp(k-1), ..., yp(k-n), u(k-1), ..., u(k-m)) + \xi(k), \quad (10)$$

where $\psi(\cdot)$ is a nonlinear function, n is the order of the model, m and p are positive integer constants, $u(k)$ is a vector of input control signals of the dynamical system, and $\xi(k)$ is the disturbance vector. Our aim in system identification is to estimate the map $\psi(\cdot)$ between inputs and outputs, and if $\psi(\cdot)$ is approximated using NN, equation (10) is then called NARX.

Let $U \in \mathscr{R}^{a\times 1}$ and $Y \in \mathscr{R}^{b\times 1}$ be the input and output vector of the UCG process, respectively, then the NARX model that identifies the UCG process is given by

$$Y(t) = \mathscr{F}(Y(t-1), Y(t-2), ..., Y(t-n), U(t-1), U(t-2), ..., U(t-m)), \quad (11)$$

where n is the number of past outputs used as inputs to the model (lag order), and m is the number of past inputs used as inputs to the model (input delay).

The function $\mathscr{F}$ is typically modeled using a feedforward neural network– see figure 4 for the schematic of NARX– with one or more hidden layers, and can be written as

$$\mathscr{F} = \mathscr{G}^m \left(\sum_{i=1}^{n} W y_i y(t-i) + \sum_{j=1}^{m} W x_j x(t-j) + b \right), \quad (12)$$

where $Wy, Wy_2, \ldots, Wy_n$ and $Wx, Wx_2, \ldots, Wx_m$ are the weights for the past outputs and inputs, respectively, b is the bias term, and $\mathscr{G}$ is the activation function.

The weights and bias terms are estimated during the training phase using a dataset of input-output pairs $(u(t), y(t))$. The goal is to minimize the mean squared error between the model predictions and the true outputs.

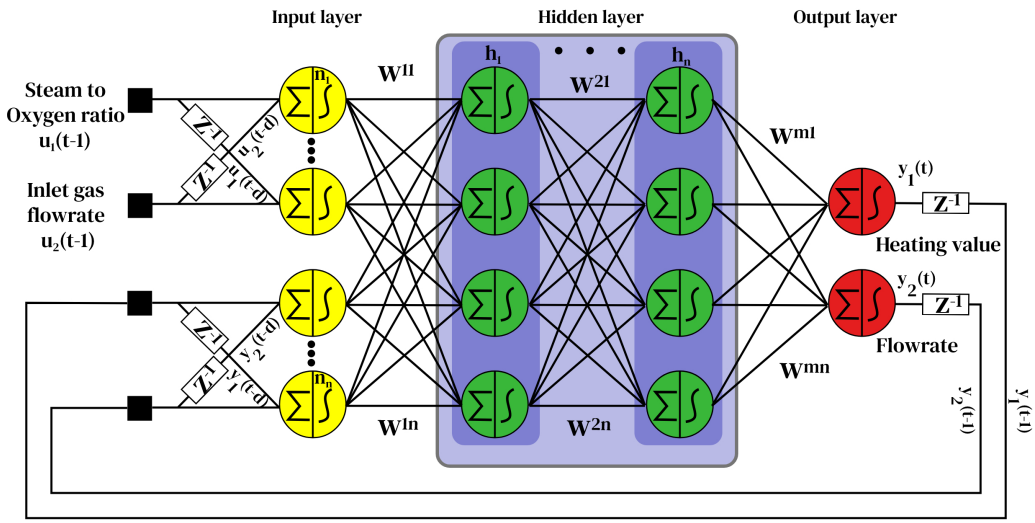


Figure 4: NARX general structure

The NARX model's nonlinear part can be represented as a multiple-input multiple-output or multiple (2)-input single output models, with the former being called NARX MIMO, and latter NARX SISO. The trained parameters of the NARX model are summarized in table 4.

2.2.2. Hammerstein-Wiener

The Wiener, Hammerstein, and Hammerstein-Wiener (HW) modeling techniques utilize block structures to represent dynamical systems. These

a	b	h_n	n_n	$\mathscr{G}^1$	$\mathscr{G}^2$	n	m
2	2	1	10	$tanh(.)$	$ReLU$	2	2

Table 4: NARX parameters

structures consist of interconnected blocks with both static non-linearity and dynamic linearity functionalities, making them suitable for modeling various types of systems. When there is prior knowledge of a plant's dynamics, these structures can be used for grey-box modeling. Among these configurations, the HW model is widely used. In the following paragraphs, we will briefly explain the HW model

The HW model's unified structure expands the identification range of dynamical systems, specifically when non-linear actuators and sensors are present. The HW model consists of a linear block—modelled as transfer function—flanked at both ends by two inputs and outputs static non-linear blocks. Figure 5 gives the structure of the HW model for UPT gasifier.

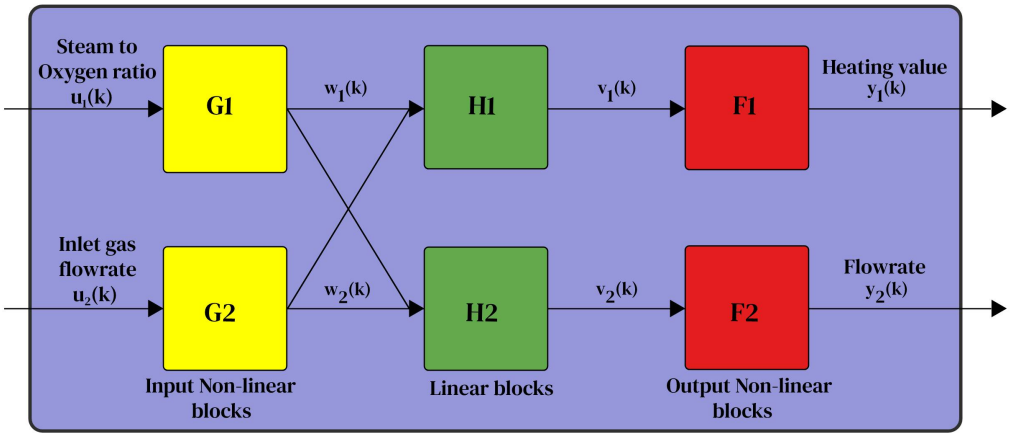


Figure 5: General srtucture of HW

The HW approach seeks to represent a system as a combination of parallel linear and nonlinear components. Various nonlinear functions can be utilized with the HW method, including piecewise linear functions, wavelet networks, one-layer sigmoid networks, dead zones, saturation, and one-dimensional polynomials. Thus, the pool of known non-linear estimators and transfer functions allows the HW structure to identify diverse set of dynamical system, and are particularly useful in grey-box modeling paradigm. In figure 5, it can be seen that for the HW model, inputs to the plants are first passed through the non-linear block G, taking on the following form

$$x(k) = G(u(k)), \tag{13}$$

where $G(.)$ can be any non-linear function identified above.

Similarly, the outputs from $G(.)$ get mapped by the linear block H, which are typically given by a transfer function. The input/output relation of linear block is given as follows

$$v(k) = H(z)x(k), \tag{14}$$

where $H(z)$ is defined as follows

$$H(z) = \begin{bmatrix} H_{11}(z) & H_{12}(z) \\ H_{21}(z) & H_{22}(z) \end{bmatrix} = \frac{B(z)}{F(z)},$$

where $H(z)$ is a transfer function matrix with elements $H_{ij}(z)$, $B(z)$ is a polynomial of degree n in z^{-1} representing the numerator of the transfer function, and $F(z)$ is a polynomial of degree m in z^{-1} representing the denominator of the transfer function.

Finally, the output of the linear block is fed to to the static non-linear

output, thus the output of the model takes the form

$$Y(k) = F(V(k)), \tag{15}$$

where F can be any non-linear function identified above.

To obtain the HW parameters, it is necessary to conduct a training phase that minimizes the gap between the measured and estimated outputs. The parameters of the HW model's linear and nonlinear components can be calculated using iterative methods such as the gradient descent scheme. In this study, the "nlhw" subroutine available in MATLAB is utilized for this purpose. Table 5 presents the key parameters of the HW model that has been identified for the UCG process.

G_1	G_2	F_1	F_2	n	m
Deadzone	Deadzone	Piecewise linear	Piecewise linear	2	3

Table 5: HW parameters

2.2.3. State Space Neural Network

One of the most ubiquitous ways to model or represent dynamical systems is state-space representation that also preserves the system's physical interpretability in its first-principle canonical form, thus offering opportunities to analyse and control the plants. Despite having distinct structures and ways to model information, state-space representation and Recurrent Neural Networks (RNN) are often considered prime candidates to model dynamical systems. Borrowing bits of structural representation and learning algorithms

from both gives rise to the notion of SSNN. Though SSNN are a subset of RNN, their architecture bears little resemblance to that of traditional NNs, with the former being more akin to state-space representation.

2.2.3.1 SSNN Architecture

The mathematical formulation of a genereal non-linear discrete time process is as follows

$$\begin{cases} x\left(k+1\right) = \mathscr{F}\left(x\left(k\right), u\left(k\right)\right), \\ y\left(k\right) = \mathscr{G}\left(x\left(k\right)\right) + v\left(k\right), \end{cases} \tag{16}$$

where $x \in \mathbb{R}^s$, $u \in \mathbb{R}^n$, and $y \in \mathbb{R}^m$, are system's states, inputs, and outputs respectively. Similarly, $\mathscr{F}$ and $\mathscr{G}$ are the non-linear mappings approximated using NN.

SSNN is a subset of RNN such that its structure has the flexibility of representing a dyanmical system in non-linear state-space form. Figure 6 represents the general structure of SSNN.

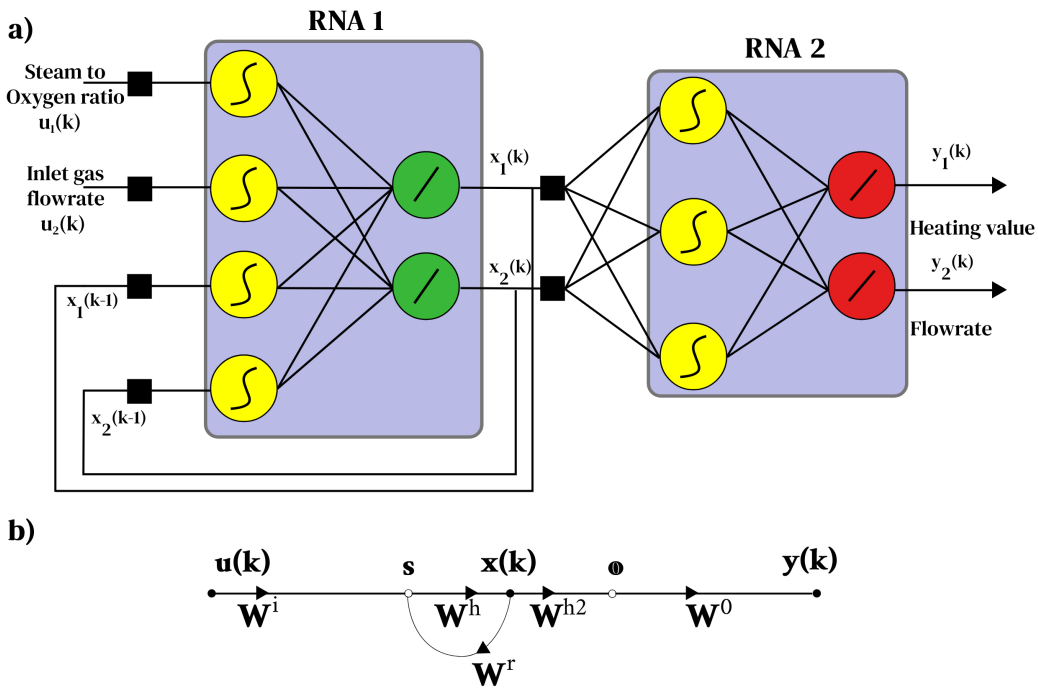


Figure 6: General srtucture of SSNN

Furthermore, the SSNN's architecture can also be construed as two blocks of NN with an estimated state-space between them. The structure of SSNN lends it a mathematical form, see equation (17), that bears resemblance to state-space representation of a non-liear dynamical system

$$\begin{cases} \hat{x}\left(k+1\right) = W^{h}\mathscr{F}\left(W^{r}\hat{x}\left(k\right) + W^{i}u\left(k\right) + B^{h}\right) + B^{l}, \\ \hat{y}\left(k\right) = W^{0}\mathscr{G}\left(W^{h2}\hat{x}\left(k\right) + B^{h2}\right) + B^{l2}, \end{cases} \tag{17}$$

where $\hat{x}\left(k\right) \in \mathbb{R}^{s}$ is the estimated state vector, $\hat{y}\left(k\right) \in \mathbb{R}^{m}$ is the estimated output vector, $u\left(k\right) \in \mathbb{R}^{n}$ is the input vector, $W_i, W_h, W_r, W_0, W_{h2}$ are weight matrices that determine the strength of the connections between the nodes of the neural network, B_h, B_l, B_{h2}, B_{l2} are bias vectors that shift the activation functions of the nodes, $\mathscr{F}$ and $\mathscr{G}$ are non-linear activation functions

that introduce non-linearity into the network. The parameters h and $h2$ are the number of hidden nodes in the first and second layer of the network, respectively.

The estimation of dynamical system by SSNN requires finding the right values of weights (W), and biases (B); whereas, the parameters $\mathscr{F}$, $\mathscr{G}$, s— number of states—, h and h_2 are updatable parameters. Moreover, prior to training, SSNN needs to know the value of x_0— initial conditions—, and total number of inputs and outputs of plant—n and m respectively. Given the optimal values of weights and biases, SSNN can replicate the behavior of the actual process. Moreover, unlike other neural structures such as Multi-Layered Perceptron (MLP) and Feed-Forward Neural-Network (FFNN), SSNN is also capable of estimating the values of system's states $\hat{x}(k)\ \forall\ k$, owing to its distinct structure.

2.2.3.2 SSNN Training

The SSNN learning process involves determining the appropriate weights and biases values that minimize the difference between its predicted and actual values. To accomplish this, learning is framed as a non-linear optimization problem, where the objective function or cost function is minimized to determine the optimal values of W and B. Generally, Mean Squared Error (MSE) is taken as the cost function

$$E = \frac{1}{2}\sum_{k=1}^{N} \|e(k)\|^2 = \frac{1}{2}\sum_{k=1}^{N} \|y(k) - \hat{y}(k)\|^2 . \tag{18}$$

The MSE is minimized by the gradient descent algorithm to find the values of weights and biases such that difference between model's prediction ($\hat{y}(k)$)

and actual plant ($y(k)$) is minimal. For training SSNN in this work, we used MATLAB's "nlssest" subroutine. The functions $\mathscr{F}$ and $\mathscr{G}$ are neural networks with one hidden layer and $tanh(.)$ activation function, and the other important parameters of the SSNN after training are given in table 6.

n	m	s	h	h_2
2	2	2	10	5

Table 6: SSNN parameters

Before we discuss results and analysis, let's consider how ML model structures and data preprocessing affect I/O data quality for system identification. We used NARX, SSNN, and H-W models with different tuning parameters requiring optimization. Increasing model complexity, i.e., increasing the number of tuning parameters, improves performance but demands more computational resources; thus, balancing models' complexity and computation is crucial [51]. Moreover, in this work, we used trial and error method iteratively to determine the ML model's tuning parameters that yielded the best performance. Similarly, prior to models' training, data quality is vital, as real-world sensor data may contain noise and errors. Data preprocessing tasks, which include outlier detection, missing value handling, feature selection, and dimensionality reduction, enhance data quality. It also helps identify key inputs using techniques like correlation analysis and principal component analysis [52]. However, in this study, we attempted to model and simulate the actual scenario of the Thar field, which has only two measurable inputs: the steam-to-oxygen ratio and its flowrate. Moreover, since the training data was generated from the CAVSIM simulator, there was less need

for data preprocessing [45].

3. Results and Discussion

In this work, all the models are trained to make single-step ahead prediciton (SSP), wherein the model makes the next prediction of outputs based on the current values of inputs and outputs unlike multi-step ahead prediction (MSP). SSP is more suitable for situations where information about the immediate future is of greater importance. This is the case with UCG, which is a slow process and does not require long-term forecasting. Moreover, from an online control perspective, SSP is preferable as the control actions require information about the current values of the system's outputs. SSP is simpler to train as the models need to learn one step ahead map, whereas in MSP, models have to learn multiple mapping functions for different time horizons, thus making them more difficult to train.

At this point, it is worth reiterating that input/output measurement data for the UPT process has been taken from CAVSIM; with this in mind, we can now turn our attention to the main discussion. The models used in this study belong to a specific class of ML algorithms. The input-output data, see figures 3 and 7-10, were split into 70-30 segments randomly for training and validation; this is because APRBS signal has a pseudo-random nature, and the system modes are not excited in a deterministic order [53]. Moreover, after training and validating models, the performance of individual models is compared with each other. Finally, the best-trained model is selected, and its performance is further tested on actual data taken from an experimental setup that was previously unseen by the trained model. The statistical tools

used to evaluate the models' performances are Root Mean Squared Error (RMSE), Mean Average Error (MAE) and Best fit, which are presented in table 7.

Tool	**Formulas**	**Criterion**
RMSE	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$	Low values indicate good fit
MAE	$\frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$	Low values indicate good fit
BEST FIT	$\left(1 - \frac{\sqrt{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}}{\sqrt{\sum_{i=1}^{n}\left(y_i - \frac{\sum_{i=1}^{n} y_i}{n}\right)^2}}\right)$	High values indicate good fit

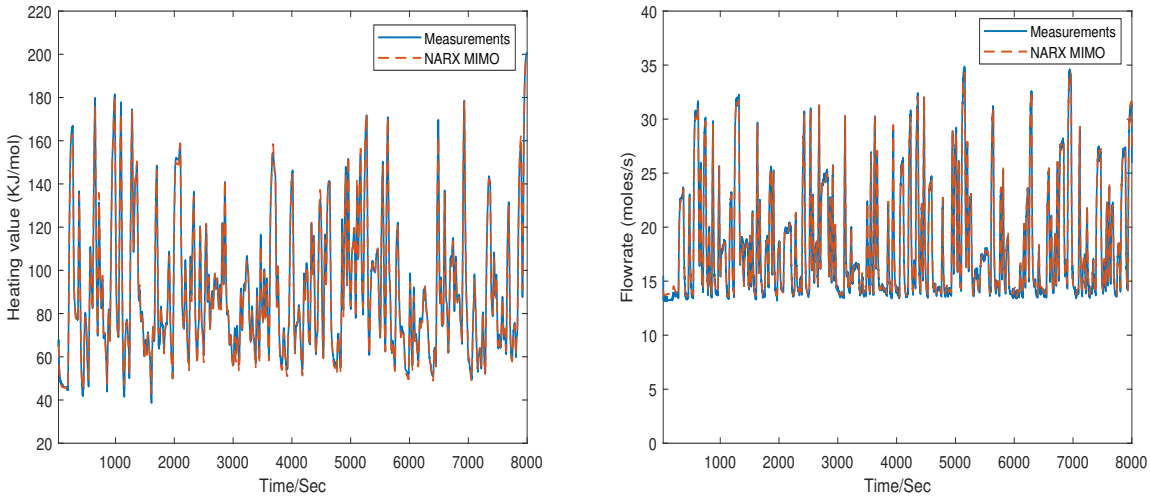
Table 7: Performance measures and criteria

Figures 7–10 are the predictions of the heating values and flowrate of all model structures after being trained on 70 % of the training-validation data-set. It is important to reiterate that figures 3 and 7-10 contain a 70–30 training–validation data-set; however, due to the random splitting of this data-set, as discussed in the previous paragraph, the distinction between training and validation data-sets is not realizable. Similary, figures 11a and 11b represents the residuals of heating values and flowrate respectively, of all model structures. A closer look at the residuals of output 1 (figure 11a) reveals that NARX MIMO gave the minimum error between the actual and predicted output, with a maximum value of 3.96 KJ/mol in the positive direction, while SSNN yielded the maximum error of 17.84 KJ/mol in the positive direction. Similarly, the residuals of output 2 (figure 11b) show that

HW and SSNN gave the minimum and maximum errors of 2.28 and 4.92 mol/s, respectively, in the negative direction; table 8 quantifies the residuals of both outputs of all models. However, the results show that there is no clear winner in terms of performance, as different models outperformed each other at different time instances. Additionally, there was a high frequency of lower error terms that were bounded under a certain threshold with occasional outliers.

Model	Heating value (KJ/mol)			Flowrate (moles/s)		
	Average	Max	Min	Average	Max	Min
NARX-MIMO	0.80	3.96	3.63×10^{-5}	0.25	2.40	2.27×10^{-5}
NARX-SISO	1.23	9.81	1.85×10^{-5}	0.25	2.47	2.90×10^{-5}
H-W	1.10	7.62	4.83×10^{-4}	0.45	2.28	2.24×10^{-5}
SSNN	2.28	17.84	8.03×10^{-5}	0.79	4.92	4.32×10^{-4}

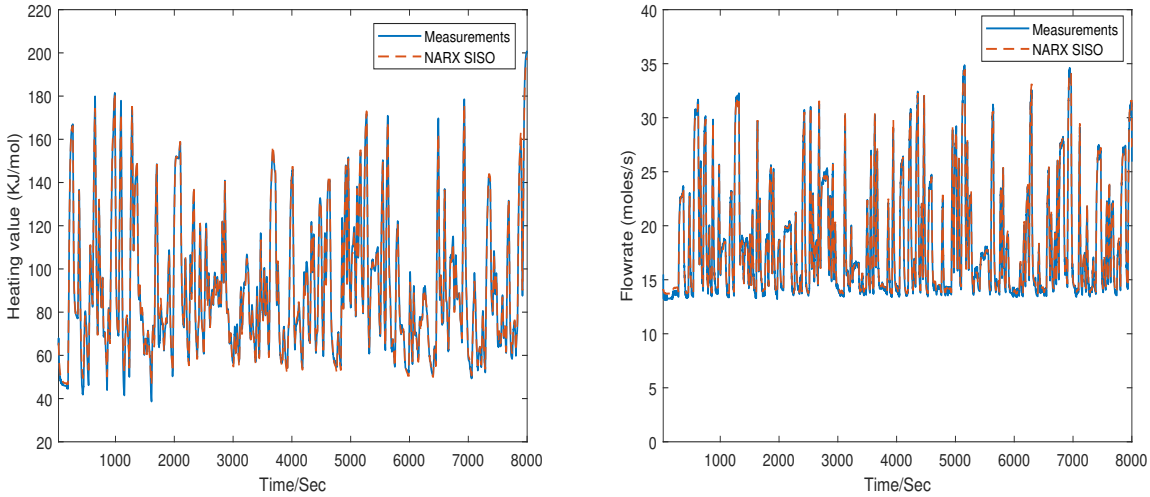
Table 8: Residuals of all models' predictions



(a) Heating value

(b) Flowrate

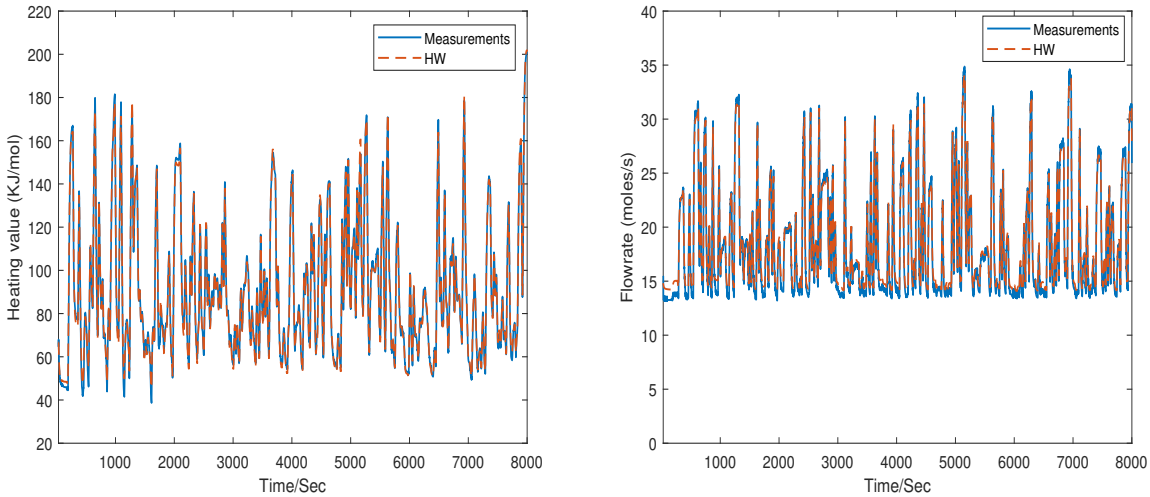
Figure 7: Training and validation of NARX MIMO



(a) Heating value

(b) Flowrate

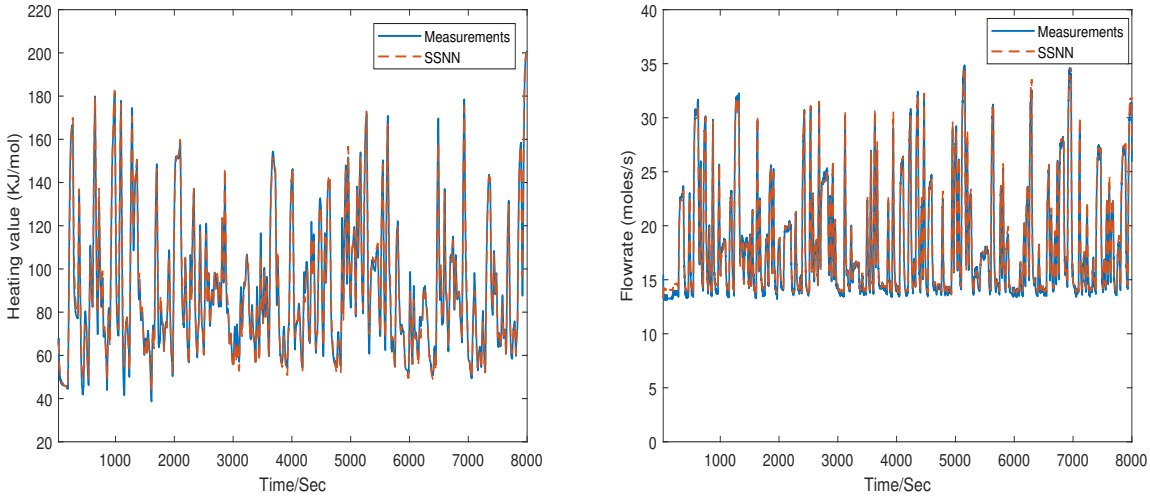
Figure 8: Training and validation of NARX SISO



(a) Heating value

(b) Flowrate

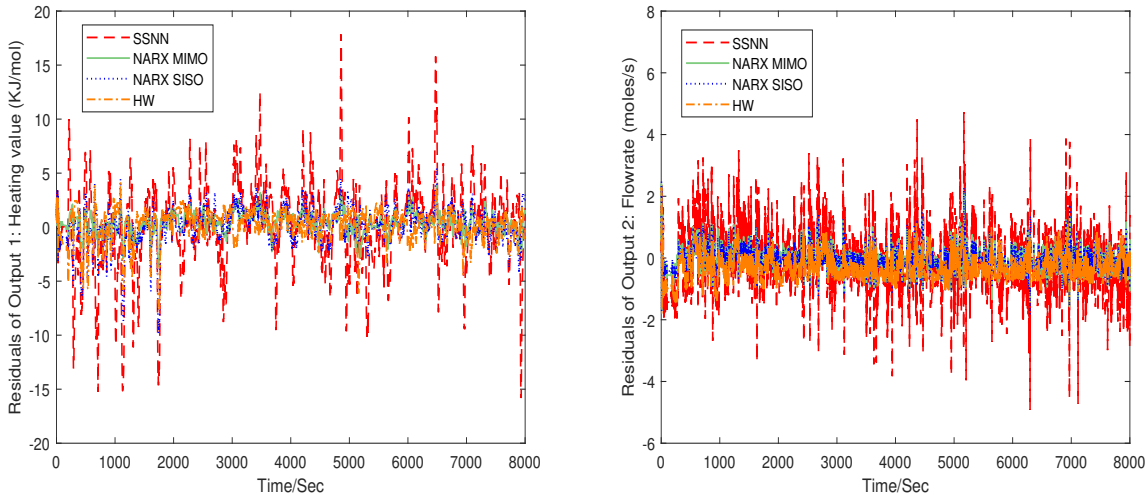
Figure 9: Training and validation of HW



(a) Heating value

(b) Flowrate

Figure 10: Training and validation of SSNN



(a) Heating value

(b) Flowrate

Figure 11: Residuals of the outputs

The models' performance relies heavily on parameter fine-tuning, which was achieved through a trial-and-error strategy. MAE and RMSE metrics were used to track the absolute difference between the predicted and actual output, providing linear and quadratic scores, respectively. As demonstrated in figure 12, NARX MIMO exhibited the lowest MAE and RMSE scores in both training and testing phases, outperforming HW, NARX SISO, and SSNN. For output 2- see figure 13- NARX MIMO and NARX SISO showed similar metric values, followed by HW and SSNN, respectively. The average values for both outputs- figures 14 and 15- displayed a consistent pattern, with NARX MIMO outperforming the other models.

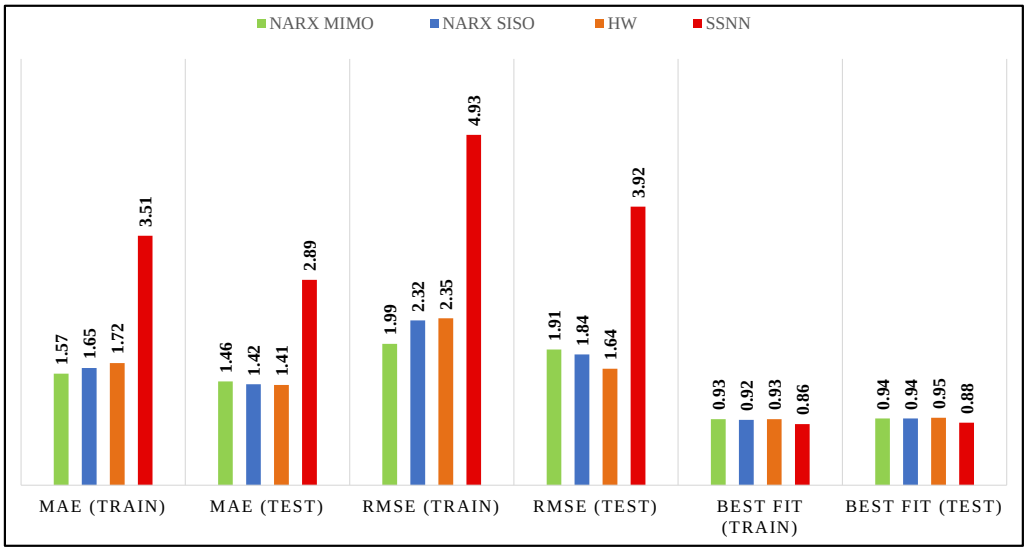


Figure 12: Statistical measures for training and validation of heating value.

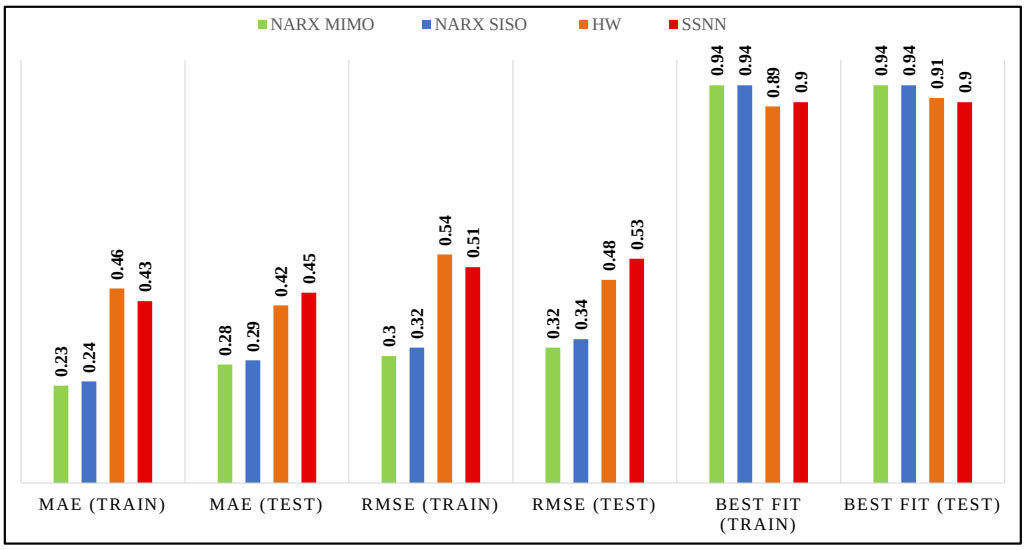


Figure 13: Statistical measures for training and validation of flowrate.

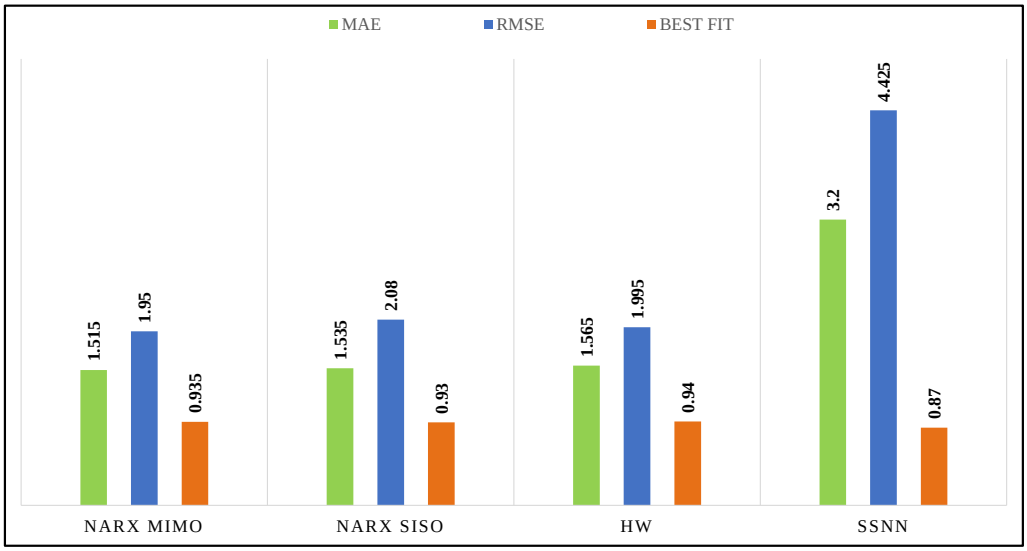


Figure 14: Average statistical measures for training and validation of heating value.

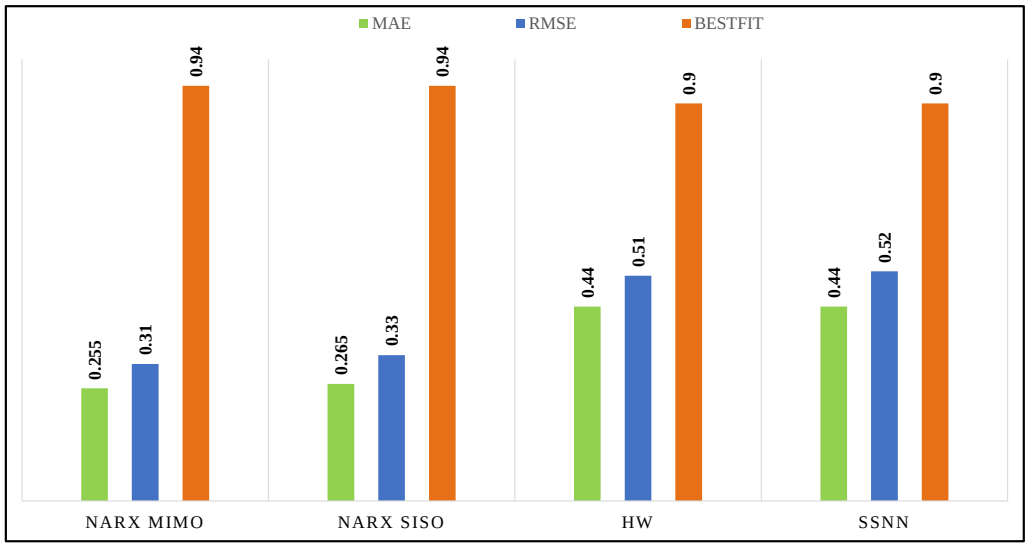
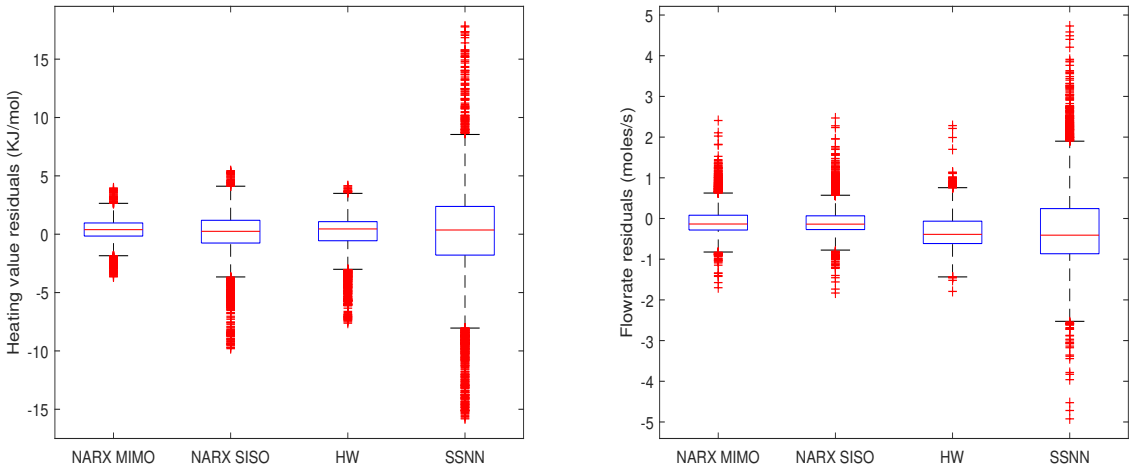


Figure 15: Average statistical measures for training and validation of flowrate.

Figure 16 shows box and whisker plots to select the best models for pre-

diction and control. These plots offer a visual representation of data distributions and can reveal information about skewness, outliers, and percentiles. They are often used in statistical analysis to compare multiple datasets and identify abnormal values.



(a) Heating value

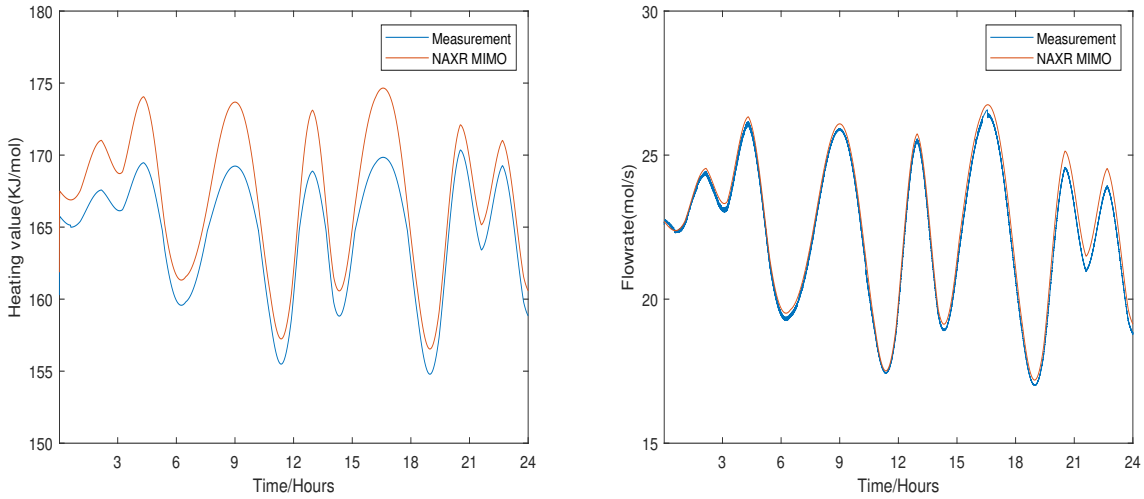
(b) Flow rate

Figure 16: Box and Whisker plots for model selection

Regarding the central tendency of heating value and flowrate –refer to figures 16a and 16b respectively–, all models except SSNN exhibit a proximity to zero. Further, for heating value, the models exhibit error spread in an increasing order of NARX MIMO, HW, NARX SISO, and SSNN, respectively. In contrast, for flowrate, the NARX MIMO, NARX SISO, and HW models indicate nearly similar error spreads, with the SSNN being an outlier.

Although it is common practice to validate a model using training-validation data-set, it is crucial to also evaluate the model's performance using exper-

imental data that was not used for training. This is because a model may perform well on the training-validation data-set, but may not generalize well to new experimental data, thereby affecting the model's ability to accurately predict a system's behavior under real-world conditions [54]. Furthermore, this evaluation can provide insights into the model's stability, robustness, and generalizability to new, untested data. To this end, an unseen input signal obtained from a field-setup [7] was fed into the NARX MIMO, which was selected based on its superior relative performance as evident from the statistical analysis. The output of the NARX MIMO was then compared with CAVSIM, the results of which are given in figures 17a and 17b



(a) Heating value prediction by NARX MIMO for experimental data

(b) Flowrate prediction by NARX MIMO for experimental data

Figure 17: Comparison with experimental data

Observations show that maximum deviations in predictions from actual

values are 7 KJ/mol and 0.7 mol/s for heating value and flowrate, respectively. Additionally, statistical analysis demonstrates similar performance of NARX MIMO in predicting flowrate on experimental data, compared to the training and validation phases, with MAE, RMSE, and Best fit values of 0.13, 0.16, and 0.9 respectively; similarly, the metrics of heating value are 3.84, 4.43, and 0.7 respectively. Overall, NARX MIMO performance is acceptable against experimental data from a practical standpoint.

4. Qualitative Assessment of Models' Performance And Their Application Scope

This section aims to serve two purposes: to draw a qualitative comparison between the trained models and their potential applications in the field of UCG. The analysis showed that NARX MIMO exhibited the most accuracy in both training and experimental phases. To present our findings in a comprehensive manner, we have included a comparison table 9 that highlights the features of each of these structures.

The table shows that NARX MIMO and NARX SISO structures have a simple and fast training process with few tunable parameters, but NARX SISO is not suitable for control applications. Overall, the results presented in this table provide valuable insights into the strengths and weaknesses of each of the system identification structures, which can be useful for researchers and practitioners working in the field of UCG process identification and control. It is important to note that while the results presented in the table may be specific to UCG processes, the methodology and approach used in this study can be applied to other complex and nonlinear systems as well.

Models	**Features**		
	Prediction accuracy	Ease of implementation	Application in Control system
NARX MIMO	✓✓✓✓	Few tunable parameters. Fastest to train. Simple structure.	Neural and intelligent based control methods.
NARX SISO	✓✓✓	Few tunable parameters. Fastest to train. Simple structure.	Not suitable for control applicatons.
HW	✓✓	Relatively complicated structure. Suitable for grey box modeling. Faster to train	Neural and intelligent based control. methods.
SSNN	✓✓	Simple structure. Large number of tuning parameters. Slowest to train.	All classical, optimal and intelligent based control methods.

Table 9: Comparison of different system identification structures

Finally, to wrap up, we are going to briefly reiterate the contribution and highlight the Ml models' application in a wider context. The aim of this work was to rigorously follow the SI pipeline (see figure 2) to the UCG Thar process. One of the distinguishing features of current work is the methodological synthesis of the excitation signal, APRBS, to capture the full non-linear operating range of the UCG process' dynamics. The information richness of our APRBS is further validated by NARX-MIMO predictions on experimental data, which was previously unseen by ML models (see figure 17). The experimental validation of ML models is amiss in the context of UCG literature (table 2), where conventional validation on 70-30 datasets is usually performed. Moreover, these ML models are control-oriented, a demonstration of which is given in our earlier work [46], wherein we designed Model Predictive Control (MPC) for the UCG plant with NARX-MIMO as the mathematical model employed by MPC for prediction purposes. Furthermore, apart from prediction and control purposes, these ML models can be utilized for open-loop trajectory optimization of the UCG process, wherein the optimal trajectories of the UCG's inputs can be found to maximize the heating value and flowrate of syngas. This can be achieved by using ML models as surrogate models for trajectory optimization of UCG plants using black box optimization (BBO)[55]; where surrogate models provide the initial conditions for the search algorithms, consequently restricting the search space and making BBO algorithms faster.

5. Conclusion and Future Work

The system identification of Underground Coal Gasification (UCG) processes is a challenging task due to the complex and nonlinear nature of the process. In this research article, we have explored and compared three different system identification structures, namely NARX, HW, and SSNN. In order to evaluate the performance of these structures, we have used statistical metrics–MAE, RMSE, and Best-fit-and compared their prediction accuracy. The metrics showed that NARX MIMO had the superior performance than all other models in terms of prediction on both training and experiental phases. Moreover, this study also provided a quantitave comparsion of models in terms of predicitve capabilities, ease of implementation, and applications in control systems.

For future work, we have twofold agenda: considering NARX MIMO as a surrogate model for trajectory optimization of UCG system using Black-Box optimization techniques—with CAVSIM being the Black-Box—, and considering SSNN for designing Gain-Scheduled Model Predictive Control for the same. Similarly, an exciting avenue of research that could precede this work is the online training of ML models, wherein the tuning parameters are udpated in real-time operation as the input/output data become available through sensors.

Funding

No funding associated with this work.

Credit author contribution statement

Afaq Ahmed: Methodology, Simulation, Validation, and Writing—Original draft. **Ali Arshad Uppal**: Conceptualization, Methodolgy, Supervision, Validation, and Writing—Review & editing. **Syed Bilal Javed**: Conceptualization, Supervision, Data curation, and Writing—Review and editing.

References

[1] S. B. Javed, A. A. Uppal, R. Samar, A. I. Bhatti, Design and implementation of multi-variable H_∞ robust control for the underground coal gasification project thar, Energy 216 (2021) 119000. doi:https://doi.org/10.1016/j.energy.2020.119000.
URL https://www.sciencedirect.com/science/article/pii/S0360544220321071

[2] P. J. Megía, A. J. Vizcaíno, J. A. Calles, A. Carrero, Hydrogen Production Technologies: From Fossil Fuels toward Renewable Sources. A Mini Review, Energy Fuels 35 (20) (2021) 16403–16415. doi:10.1021/acs.energyfuels.1c02501.

[3] A. Arshad, A. I. Bhatti, R. Samar, Q. Ahmed, E. Aamir, Model development of UCG and calorific value maintenance via sliding mode control, in: 2012 International Conference on Emerging Technologies, IEEE, 2012, pp. 1–6. doi:10.1109/ICET.2012.6375477.

[4] A. M. Winslow, Numerical model of coal gasification in a packed bed, Symp. Combust. 16 (1) (1977) 503–513. doi:10.1016/S0082-0784(77)80347-0.

[5] C. B. Thorsness, R. B. Rozsa, In-Situ Coal Gasification: Model Calculations and Laboratory Experiments, SPE J. 18 (02) (1978) 105–116. doi:10.2118/6182-PA.

[6] A. A. Uppal, A. I. Bhatti, E. Aamir, R. Samar, S. A. Khan, Control oriented modeling and optimization of one dimensional packed bed model

of underground coal gasification, J. Process Control 24 (1) (2014) 269–277. doi:10.1016/j.jprocont.2013.12.001.

[7] S. B. Javed, A. A. Uppal, A. I. Bhatti, R. Samar, Prediction and parametric analysis of cavity growth for the underground coal gasification project Thar, Energy 172 (2019) 1277–1290. doi:10.1016/j.energy.2019.02.005.

[8] C. B. Thorsness, J. A. Britten, CAVISM user manual, [Online; accessed 2. Sep. 2022] (Mar. 1989). doi:10.2172/6307133.

[9] R. D. Gunn, D. L. Whitman, In situ coal gasification model (forward mode) for feasibility studies and design, [Online; accessed 2. Sep. 2022] (Feb. 1976).
URL https://www.osti.gov/biblio/7354280-situ-coal-gasification-model-forward-mode-feasibility-studies-design

[10] A. M. Winslow, Numerical model of coal gasification in a packed bed, Symp. Combust. 16 (1) (1977) 503–513. doi:10.1016/S0082-0784(77)80347-0.

[11] C. B. Thorsness, R. B. Rozsa, In-Situ Coal Gasification: Model Calculations and Laboratory Experiments, SPE J. 18 (02) (1978) 105–116. doi:10.2118/6182-PA.

[12] C. B. Thorsness, S. W. Kang, Further development of a general-purpose, packed-bed model for analysis of underground coal gasification processes, [Online; accessed 2. Sep. 2022] (Aug. 1985).
URL https://www.osti.gov/biblio/5212559-further-development

-general-purpose-packed-bed-model-analysis-underground-coal\
-gasification-processes

[13] E. A. A. Abdel-Hadi, T. R. Hsu, Computer Modeling of Fixed Bed Underground Coal Gasification Using the Permeation Method, J. Energy Res. Technol. 109 (1) (1987) 11–20. doi:10.1115/1.3231316.

[14] A. N. Khadse, M. Qayyumi, S. M. Mahajani, P. Aghalayam, Reactor Model for the Underground Coal Gasification (UCG) Channel, Int. J. Chem. Reactor Eng. 4 (1) (Nov. 2006). doi:10.2202/1542-6580.1351.

[15] C. F. Magnani, S. M. F. Ali, Mathematical Modeling of the Stream Method of Underground Coal Gasification, SPE J. 15 (05) (1975) 425–436. doi:10.2118/4996-PA.

[16] M. L. Pasha, An Advanced Numerical Model Of Underground Coal Gasification By The Stream Method, Using Simultaneous Solution, OnePetro (Oct. 1978). doi:10.2118/7416-MS.

[17] B. Dinsmoor, J. M. Galland, T. F. Edgar, The Modeling of Cavity Formation During Underground Coal Gasification, J. Pet. Technol. 30 (05) (1978) 695–704. doi:10.2118/6185-PA.

[18] T. L. Eddy, S. H. Schwartz, A Side Wall Burn Model for Cavity Growth in Underground Coal Gasification, J. Energy Res. Technol. 105 (2) (1983) 145–155. doi:10.1115/1.3230894.

[19] R. A. Kuyper, Th. H. Van Der Meer, C. J. Hoogendoorn, Turbulent natural convection flow due to combined buoyancy forces during under-

ground gasification of thin coal layers, Chem. Eng. Sci. 49 (6) (1994) 851–861. doi:10.1016/0009-2509(94)80022-7.

[20] 95/03476 A new channel model for underground gasification of thin, deep coal seams, [Online; accessed 2. Sep. 2022] (Jul. 1995). doi:10.1016/0140-6701(95)95130-W.

[21] 99/01679 A coupling of chemical processes and flow in view of the cavity growth simulation of an underground coal gasifier at great depth: Pirlot, P. et al. In Situ, 1998, 22, (2), 141–156, [Online; accessed 2. Sep. 2022] (Mar. 1999). doi:10.1016/S0140-6701(99)96860-1.

[22] G. Perkins, V. Sahajwalla, Steady-State Model for Estimating Gas Production from Underground Coal Gasification, Energy Fuels 22 (6) (2008) 3902–3914. doi:10.1021/ef8001444.

[23] [Online; accessed 2. Sep. 2022] (Oct. 2009). [link].
URL https://www.cfd.com.au/cfd_conf09/PDFs/196LUO.pdf

[24] M. Seifi, J. Abedi, Z. Chen, The Analytical Modeling of Underground Coal Gasification through the Application of a Channel Method, Energy Sources Part A 35 (18) (2013) 1717–1727. doi:10.1080/15567036.2010.531501.

[25] T. H. T. Tsang, Modeling of heat and mass transfer during coal block gasification, The University of Texas at Austin, 1980.

[26] J. G. M. Massaquoi, J. B. Riggs, Mathematical modeling of combustion and gasification of a wet coal slab—I: Model development and ver-

ification, Chem. Eng. Sci. 38 (10) (1983) 1747–1756. doi:10.1016/0009-2509(83)85031-3.

[27] K. Y. Park, T. F. Edgar, Modeling of early cavity growth for underground coal gasification, Ind. Eng. Chem. Res. 26 (2) (1987) 237–246. doi:10.1021/ie00062a011.

[28] G. Perkins, V. Sahajwalla, A Mathematical Model for the Chemical Reaction of a Semi-infinite Block of Coal in Underground Coal Gasification, Energy Fuels 19 (4) (2005) 1679–1692. doi:10.1021/ef0496808.

[29] C. B. Thorsness, J. A. Britten, Lawrence Livermore National Laboratory Underground Coal Gasification project, [Online; accessed 16. Feb. 2023] (Oct. 1989).

[30] E. N. J. Biezen, J. Bruining, J. Molenaar, An Integrated 3D Model for Underground Coal Gasification, OnePetro (Oct. 1995). doi:10.2118/30790-MS.

[31] J. J. Nitao, T. A. Buscheck, S. M. Ezzedine, S. J. Friedmann, D. W. Camp, An Integrated 3-D UCG Model for Predicting Cavity Growth, Product Gas, and Interactions with the Host Environment, [Online; accessed 16. Feb. 2023] (Sep. 2017).

[32] G. Samdani, P. Aghalayam, A. Ganesh, R. K. Sapru, B. L. Lohar, S. Mahajani, A process model for underground coal gasification – Part-II growth of outflow channel, Fuel 181 (2016) 587–599. doi:10.1016/j.fuel.2016.05.017.

[33] H. Akbarzadeh, R. J. Chalaturnyk, Sequentially coupled flow-geomechanical modeling of underground coal gasification for a three-dimensional problem, Mitigation Adapt. Strategies Global Change 21 (4) (2016) 577–594. doi:10.1007/s11027-014-9583-2.

[34] J. Kačur, M. Laciak, M. Durdán, P. Flegner, Utilization of Machine Learning method in prediction of UCG data, in: 2017 18th International Carpathian Control Conference (ICCC), IEEE, 2017, pp. 278–283. doi:10.1109/CarpathianCC.2017.7970411.

[35] A. Krzemień, Fire risk prevention in underground coal gasification (UCG) within active mines: Temperature forecast by means of MARS models, Energy 170 (2019) 777–790. doi:10.1016/j.energy.2018.12.179.

[36] Y. Xiao, H. Yin, T. Duan, H. Qi, Y. Zhang, A. Jolfaei, K. Xia, An Intelligent prediction model for UCG state based on dual-source LSTM, Int. J. Mach. Learn. Cybern. 12 (11) (2021) 3169–3178. doi:10.1007/s13042-020-01210-7.

[37] Y. Xiao, H. Yin, K. Xia, Y. Zhang, H. Qi, Utilization of CNN-LSTM Model in Prediction of Multivariate Time Series for UCG, in: Machine Learning for Cyber Security, Springer, Cham, Switzerland, 2020, pp. 429–440. doi:10.1007/978-3-030-62463-7_40.

[38] S. B. Javed, V. I. Utkin, A. A. Uppal, R. Samar, A. I. Bhatti, Data-Driven Modeling and Design of Multivariable Dynamic Sliding Mode Control for the Underground Coal Gasification Project Thar, IEEE

Trans. Control Syst. Technol. 30 (1) (2021) 153–165. doi:10.1109/TCST.2021.3057633.

[39] A. Dorobantu, A. Murch, B. Mettler, G. Balas, System Identification for Small, Low-Cost, Fixed-Wing Unmanned Aircraft, Journal of Aircraft (Jul. 2013). doi:10.2514/1.C032065.

[40] B. Yu, W. Shu, C. Cao, A Novel Modeling Method for Aircraft Engine Using Nonlinear Autoregressive Exogenous (NARX) Models Based on Wavelet Neural Networks, International Journal of Turbo & Jet-Engines 35 (2) (2018) 161–169. doi:10.1515/tjj-2017-0005.

[41] H. Asgari, E. Ory, Prediction of Dynamic Behavior of a Single Shaft Gas Turbine Using NARX Models, ASME Digital Collection (Sep. 2021). doi:10.1115/GT2021-58960.

[42] M. Deflorian, S. Zaglauer, Design of Experiments for nonlinear dynamic system identification, IFAC Proceedings Volumes 44 (1) (2011) 13179–13184. doi:10.3182/20110828-6-IT-1002.01502.

[43] O. Nelles, R. Isermann, A Comparison Between RBF Networks and Classical Methods for Identification of Nonlinear Dynamic Systems, IFAC Proceedings Volumes 28 (13) (1995) 233–238. doi:10.1016/S1474-6670(17)45355-9.

[44] D. E. Rivera, S. V. Gaikwad, X. Chen, CONTROL-ID: a demonstration prototype for control-relevant identification, in: Proceedings of 1994 American Control Conference - ACC '94, Vol. 2, IEEE, 1994, pp. 2055–2059vol.2. doi:10.1109/ACC.1994.752438.

[45] S. B. Javed, Cavity prediction and multi-variable control of underground coal gasification process, Ph.D. thesis, Capital University of Science and Technology (2023).
URL https://cust.edu.pk/static/uploads/2021/09/PhD-EE-Thesis-Syed-Bilal-Javed.pdf

[46] A. Ahmed, S. B. Javed, A. A. Uppal, J. Iqbal, Development of CAVLAB—A Control-Oriented MATLAB Based Simulator for an Underground Coal Gasification Process, Mathematics 11 (11) (2023) 2493. doi:10.3390/math11112493.

[47] A. Ahmed, S. B. Javed, A. A. Uppal, CAVLAB–UCG-process-simulator, [Online; accessed 23. Apr. 2023] (Apr. 2023).
URL https://github.com/Hilberto-inf/CAVLAB--UCG-process-simulator

[48] O. Nelles, Nonlinear Dynamic System Identification, in: Nonlinear System Identification, Springer, Berlin, Germany, 2001, pp. 547–577. doi:10.1007/978-3-662-04323-3_15.

[49] H. Liu, On the Levenberg-Marquardt training method for feed-forward neural networks, in: 2010 Sixth International Conference on Natural Computation, Vol. 1, IEEE, 2010, pp. 456–460. doi:10.1109/ICNC.2010.5583151.

[50] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, Neural Networks 2 (5) (1989) 359–366. doi:10.1016/0893-6080(89)90020-8.

[51] P. Goyal, P. Dollár, R. Girshick, P. Noordhuis, L. Wesolowski, A. Kyrola, A. Tulloch, Y. Jia, K. He, Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour, ArXiv (Jun. 2017). arXiv:1706.02677, doi:10.48550/arXiv.1706.02677.

[52] M. E. Sánchez-Gutiérrez, P. P. González-Pérez, Multi-Class Classification of Medical Data Based on Neural Network Pruning and Information-Entropy Measures, Entropy 24 (2) (2022) 196. doi:10.3390/e24020196.

[53] A. Salehi, M. Montazeri-Gh, Black box modeling of a turboshaft gas turbine engine fuel control unit based on neural NARX, Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment 233 (3) (2018) 949–956. doi:10.1177/1475090218797496.

[54] T. Hastie, J. Friedman, R. Tibshirani, Model Assessment and Selection, in: The Elements of Statistical Learning, Springer, New York, NY, New York, NY, USA, 2001, pp. 193–224. doi:10.1007/978-0-387-21606-5_7.

[55] M. Herraz, J.-M. Redonnet, M. Sbihi, M. Mongeau, Blackbox optimization and surrogate models for machining free-form surfaces, Comput. Ind. Eng. 177 (2023) 109029. doi:10.1016/j.cie.2023.109029.