Grey wolf optimizer trained ANN technique for development of explainable model of commercial ethylene oxide reactor and multi-objective optimization to maximize profit

Sucharita Pal[1] , Sandip K Lahiri [2]

[1,2] *Chemical Engineering Department, National Institute of Technology Durgapur, Durgapur – 713209*

Abstract - *The current research focuses on the use of a grey wolf optimizer trained ANN for the building of an industrial reactor model, as well as its optimization for profit maximisation. As a case study, an ethylene oxide (EO) reactor was used. The EO reactor is a multiphase catalytic reactor with a complex nature that makes developing a reliable first principle-based model challenging, while data-driven models lack explanability. As a result, a general methodology has been devised in this study, in which a data-driven technique such as grey wolf optimizer trained ANN was utilised as a modelling tool, and the model was then post-processed to improve the model's explanability. The performance of the model in capturing the system's inherent physics was used to make the decision. Following the creation of an appropriate model, it was subjected to optimization. Maximization of catalyst selectivity and minimization of reactor temperature were examined here, both of which have a significant impact on the plant's profitability. Because the two aims are incompatible, a pareto solution was found by using a multi-objective evolutionary algorithm on the generated model.*

Key Words: Grey wolf optimizer trained ANN; Ethylene oxide reactor; Modelling; Optimization; Genetic algorithm

1.INTRODUCTION

Due to intense global competition, the business environment has changed dramatically since last decade. Chemical process industries are seeing their profit margins dwindle as a result of increased cutthroat competition in a volatile market, thanks to globalization. Furthermore, the developing environmental concern of global warming has prompted government authorities in several countries to establish rigorous environmental regulations, which add to the expense of doing business. The only way to successfully minimize these difficulties is through technological innovation to improve process and energy efficiency, yield maximization, and environmental impact reduction. Chemical firms all around the world are looking for new and imaginative methods to save costs and increase profits. One of the most appealing creative strategies to explore is the use of artificial intelligence-based techniques to extract value from massive amounts of plant operating data through data mining and knowledge discovery.

Chemical reactors are attracting the attention of researchers and plant managers who are looking for new ways to create more profit. The reactor is the only significant piece of equipment that adds real value to raw materials by converting them into finished products. Reactor optimization has a huge potential impact on total plant profitability in this regard [1]. Modeling complicated systems of industrial chemical processes is essential as a first step in optimizing the commercial reactor. Complex reaction kinetics and thermodynamics were involved in industrial chemical reactions. Building a reliable phenomenological model for commercial reactors is a time-consuming and challenging task that necessitates a thorough understanding of industrial heterogeneous catalytic behavior such as mass diffusion, catalyst deactivation, and so on. Furthermore, several parameters such as agitation speed, catalyst interfacial area, negative influence of poisons present in incoming gas, diffusional coefficients, and others affect reaction rate, yield, and selectivity in commercial plants, the mechanisms of which are poorly understood. As a result, a lack of understanding of chemical reaction kinetics obstructs the reactor's optimization. The reactors in most chemical plants remain a black box, and operation engineers do not tamper with them in a live plant for safety and reliability reasons. This leads to non-optimal operating of industrial reactors, which has a significant impact on the plant's profitability. As a result, reactors are regarded as uncharted area in the chemical industry. In large-scale operations, however, a slight increase in catalyst selectivity and reaction yield has a significant impact on raw material consumption and plant profitability [1].

Typically, laboratory scale kinetic analysis is performed in a very ideal and controlled setting to derive the kinetic equation of a particular commercial reaction. Because of the vast scale of industrial reactors, the existence of poisons and inert in input gases, and the varied heat and mass transfer environment, the use of such a method for kinetic modelling of industrial reactors is debatable. Furthermore, due to the limited market window for chemical goods, plant owners are hesitant to devote sufficient time and effort to a thorough examination of these complex reaction mechanisms. As a result, an alternative simple way is to

use a data-driven effective computational strategy to construct approximated reactor models for these complex reactions systems which can be subsequently used to optimize the reactor and increase profit.

Because most chemical plants gather and store massive amounts of reactor intake and exit historic plant operational data every minute, the true challenge is figuring out how to use this wealth of data to earn more money. Data is the new oil, and data-driven modelling approaches like artificial neural networks (ANN) and support vector machines (SVM) are unquestionably the new IC engines of our time. ANN and SVR have been highly prominent black box modelling approaches in the recent decade, with a number of applications developed for commercial reactors. Plant engineers, on the other hand, dislike ANN and SVM models because of their lack of explainability and the model's black box nature. ANN model does not provide any insights of underlying physics of chemical reactors.

Because ANN modelling gives a black box type equation consisting of a complex sigmoidal function with multiple tuning factors known as weights and biases, the model suffers from explainability limitations, despite its great prediction capacity. Process engineers typically desire comprehensible equations in differential/ algebraic form that relate output variables to input process parameters so they can gain more insight and benefit. It is preferable that the generated model include a closed form equation that can describe the effect of key process parameters on the output variable. SVM has the same problem as ANN in that none of them provides closed form explainable equations that are portable and easy to implement in a plant DCS system. For plant engineers, gaining insights and obtaining a closed form explainable model equation is critical to their acceptance of using the model in the plant.

Despite the ANN approach's impressive prediction capabilities, applications of this technology to chemical reaction engineering jobs in commercial reactors are conspicuously lacking. In this study, an attempt was made to apply ANN modelling to a commercial chemical ethylene oxide (EO) reactor. One of the key goals of this research is to turn the ANN model into an explainable closed form equation, which will reveal important information about the phenomenology of the reactor.

The second goal of this work is to use recently discovered nature-inspired grey wolf optimization approaches to train the ANN model. Learning is an important component of every neural network, and it is attracting the attention of many academics. The traditional [7] or enhanced [8– 10] Back-Propagation (BP) algorithm has been used to train feedforward neural networks (FNNs) in most applications. The BP method is a gradient-based algorithm with some flaws, including delayed convergence [11] and a possibility to get stuck in local minima [12]. During the learning process of FNNs, the goal is to identify the best combination of connection weights and biases to achieve the least amount of error. FNNs, on the other hand, frequently converge on points that are the best answer locally but not globally. In other words, rather than the global minimum, learning methods lead FNNs to local minima. According to [11], the BP algorithm's convergence is strongly dependent on the weights, biases, and parameters' initial values. Using unique heuristic optimization methods or evolutionary algorithms to improve the problems of BP-based learning algorithms is a common option in the literature.

Simulated Annealing (SA) [13,14], Genetic Algorithms (GAs) [15], Particle Swarm Optimization (PSO) algorithms [16–20], Magnetic Optimization Algorithm (MOA) [21], and Differential Evolution (DE) [22] are some of the heuristic optimization methods that have been utilized to train FNNs. According to [11], some algorithms, such as SA and GA, can lessen the likelihood of local minima trapping, although they still have poor convergence rates.

Despite the fact that numerous meta-heuristics have been used widely in ANN learning, none of them has performed well in all applications. Furthermore, existing metaheuristics have a number of drawbacks [23-27], including slow convergence speed, trapping in local minima, long computational time, tuning many parameters, and a difficult encoding scheme. As a result, it appears that improving the performance of prior meta-heuristics or proposing new ones is required to improve the efficiency of ANN learning in various domains. GWO is a recent innovative stochastic and metaheuristic optimization approach, which is introduced by [18]. In this paper, the efficiencies of GWO technique are investigated for training FNNs.

The current study opted to model a commercial EO reactor because of its significant financial significance in worldwide petrochemical markets. Because it is used to make polyester (which is used to make textile) and PET bottles, the ethylene glycol industry is regarded the second most significant petrochemical product in terms of volume, profit margins, and global market demands. Recently, the glycol sector has seen intense competition in the global market, with players vying for higher reactor yield and selectivity, which has a significant impact on production costs.

The next goal of this study is to use the developed model to raise the profit of the ethylene oxide plant while reducing its environmental impact. This is accomplished by using a model-based, nature-inspired metaheuristic optimization method to optimize the input process parameters in order to improve selectivity (i.e., reactor performance) while lowering reactor temperature (which maximize catalyst longevity). The two main operational objectives in commercial EO reactors, namely,

maximize selectivity (i.e., maximize profit) and minimize reactor temperature (corresponds to greater catalyst longevity), are mutually exclusive and conflicting, and one cannot achieve the highest selectivity without sacrificing temperature, and vice versa. The multi objective genetic algorithm (GA) is used to improve the input space of the EO reactor ANN model in order to provide pareto optimal solutions that meet both objectives in the most efficient way feasible.

2. Case study of ethylene oxide reactor

2.1 Background

Because of the lack of understanding and complexity of multiphase catalytic processes in commercial ethylene oxide reactors, trustworthy first principle-based models are hard to come by. Data driven modelling is a feasible alternative strategy due to the vast amount of reactor operating parameter data available from commercial reactors. The current project aims to make use of a huge amount of process data to create a framework for converting the data's information to knowledge and subsequently into profit.

2.2 Reactions:

The manufacture of ethylene oxide (EO) is the first step in the ethylene glycol manufacturing process. Ethylene oxide is produced in commercial glycol plants by reacting ethylene with oxygen (Equation 1) at 230–270°C and 20 bar pressures. In addition to ethylene oxide, an unfavorable side reaction (Equation 2) produces carbon di-oxide, which is released into the atmosphere and has a severe negative influence on the environment. Catalyst selectivity is computed using equation 3 and measures the extent of the first reaction (desired) relative to the second reaction (undesired). Selectivity maximization is a key production goal for increasing profits while minimizing environmental damage.

$$C_2H_4 + \frac{1}{2}O_2 \rightarrow C_2H_4O \quad (1)$$

$$C_2H_4 + 3O_2 \rightarrow 2CO_2 + 2H_2O \quad (2)$$

$$Catalyst\ selectivity = \frac{Moles\ of\ EO\ produced}{Moles\ of\ ethylene\ consumed} \quad (3)$$

2.3 Process flow diagram

Figure 1 [17] shows a simplified process flow diagram for the synthesis of EO. The EO reactor is a shell and tube reactor in which the catalyst is put in the tube side as packed bed and the reactions described in Equations 1 and 2 occur in this part. Water is cycled through the shell side to recover heat because the reactions are exothermic. Because the conversion of EO from ethylene and oxygen is very low par pass, the unreacted reactants are recycled back through the cycle gas compressor to boost the conversion. In industrial, unreacted ethylene, unreacted oxygen, carbon dioxide, methane, EO, nitrogen, (usually called cycle gas) are combined with fresh ethylene and oxygen streams. The reactor output gas was routed through EO absorbers and CO_2 absorbers, as shown in Figure 1. The EO is selectively absorbed by water in an EO absorber, whilst the CO_2 is absorbed by a carbonate solution in a CO_2 absorber. Plant's overall profitability is determined by the catalyst's selectivity, and plant aims to decreases the second undesired reaction while maximizing the first favorable reaction.

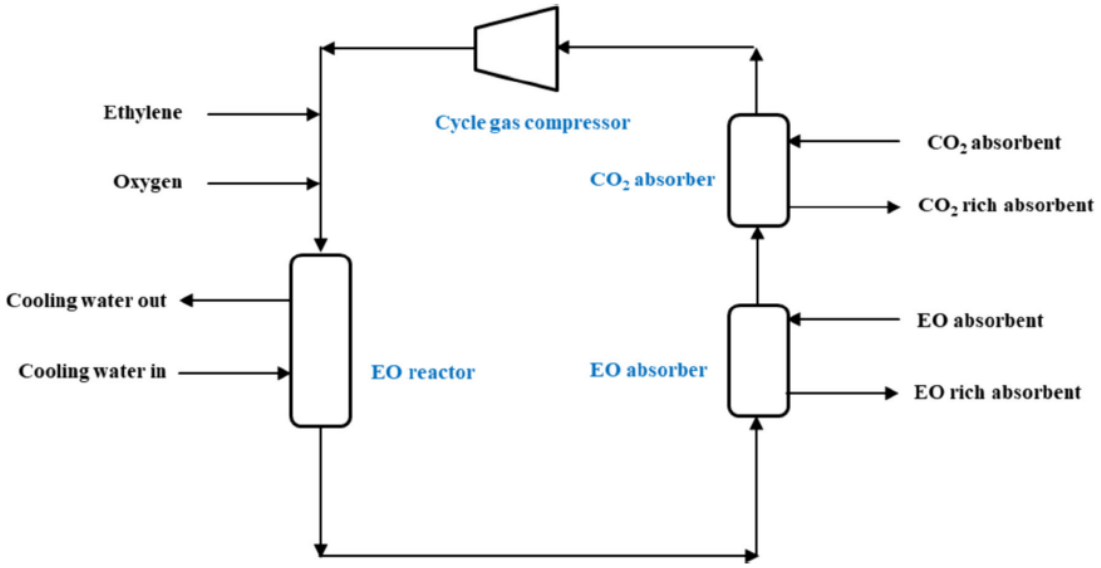

FIGURE 1: Process flow diagram of EO production process

2.4 Production Objectives:

The catalyst selectivity declines gradually with the age of the catalyst due to catalyst poisoning and sintering effects, and the catalyst selectivity maintains about 90% at the start of operation (SOR) and drops to 80% at the end of operation (EOR) (usually after 2-4 years). The average catalytic life of today's catalysts is between 24 and 48 months. The temperature of the reactor is gradually increased over the life of the catalyst [say from 225 ˚C (SOR) to 275˚C (EOR)] to enhance the reaction rate and keep the plant production on target while selectivity is dropping. The selectivity of the catalyst impacts the plant's profitability, as better selectivity means less consumption of expensive ethylene and oxygen to produce same amount of EO i.e., less production cost. Reactor temperature, on the other hand, determines catalyst longevity. A faster rise in reactor temperature causes a loss of catalyst age and a higher fixed cost of catalyst replacement, which necessitates a plant downtime. Optimizing reactor operating parameters in such a way that selectivity maximization and reactor temperature minimization (i.e., maximization of catalyst longevity) are achieved simultaneously while maintaining plant production targets are two primary production objectives in commercial plants.

3. Hybrid FNN- GWO methodology

This section explains the basic background materials needed for a thorough comprehension of the suggested method. The Multi-layer perceptron (MLP) as a Feed-forward neural network (FNN) is first presented, followed by the GWO methods, which are then compared to the suggested technique for MLP learning in industrial reactors modelling. A quick explanation of the hybrid ANN and GWO training process follows MLP network.

Because of its capability to map any function to an indefinite degree of precision, the feed-forward neural network (FNN) is one of the most prevalent ANNs used to solve real life problems and attracts a lot of academic attention. The multi-layer perceptron has been extensively used in finance, medicine, engineering, geology, physics, and biology, among other fields. It has been widely used in the chemical industry for nonlinear process modelling, fault diagnostics, and process control.

MLP has one input layer that receives external inputs, one or more hidden layers, and one output layer that displays the output, as shown in Figure 2. Except for the input layer, all levels are built up of processing nodes and activation functions. Data is supplied at the input layer, and network nodes perform calculations in successive layers until each output node obtains an output value.

3.1 Training of ANN by GWO

In general, there are three ways to train FNNs with a heuristic algorithm. First, heuristic techniques are used to discover the best combination of weights and biases for a FNN with the least amount of error. Second, heuristic algorithms are used to determine the best structure for a FNN in each problem. The final way is to tune the parameters of a gradient-based learning system, such as the learning rate and momentum, using an evolutionary algorithm.

Before training FNNs, the structure is fixed in the first scenario. A training algorithm's job is to discover a good value for all connection weights and biases in order to reduce the overall error of the FNNs. FNN structures differ in the second scenario. To discover the appropriate structure for a given problem, a training procedure is applied to an FNN. Manipulate the connections between neurons, the number of hidden layers, and the number of hidden nodes in each layer of the FNN to change the structure.

GWO is applied to a FNN using the first technique in this paper; these processes are referred to as hybrid FNN-GWO. This means that the FNN's structure is fixed, and the GWO algorithm finds a set of weights and biases that yields the least amount of error for the FNN. The following essential elements must be defined in order to design FNN-GWO. To evaluate fitness in FNN-GWO, first develop a fitness function based on the FNN's error. Second, for the agents of FNN-GWO, an encoding approach should be established to encode the FNN's weights and biases. These components are detailed in further detail below.

3.2 Fitness function

The following is the formula for the fitness function used in this article [10]:

The number of input nodes is equal to n, the number of hidden nodes is equal to h, and the number of output nodes is equal to m in Figure 2, which has two levels (one input, one hidden, and one output layer). Each hidden node's output is calculated as follows at the end of each learning epoch:

$$f(s_j) = 1/\left(1 + \exp\left(-\left(\sum_{i=1}^{n} w_{ij}.x_i - \Theta_j\right)\right)\right), \quad j = 1, 2, \ldots, h \qquad (4)$$

In $\sum_{i=1}^{n} w_{ij}.x_i - \Theta_j$, n is the number of the input nodes, w_{ij} is the connection weight from the i^{th} node in the input layer to the j^{th} node in the hidden layer, Θ_j is the bias (threshold) of the j^{th} hidden node, and x_i is the i^{th} input.

After calculating outputs of the hidden nodes, the final output can be defined as follows:

$$o_k = \sum_{i=1}^{n} w_{kj}.f(s_j) - \Theta_k, \quad k = 1,2, \ldots, m, \qquad (5)$$

where w_{kj} is the connection weight from the j^{th} hidden node to the kth output node and Θ_k is the bias (threshold) of the k^{th} output node.

Finally, the learning error E (fitness function) is calculated as follows:

$$E_k = \sum_{i=1}^{m} \left(o_i^k - d_i^k\right)^2 \qquad (6)$$

$$E = \sum_{k=1}^{q} \frac{E_k}{q} \qquad (7)$$

where q is the number of training samples, d_i^k is the desired output of the i^{th} input unit when the k^{th} training sample is used, and o_i^k is the actual output of the i^{th} input unit when the kth training sample is used. Therefore, the fitness function of the i^{th} training sample can be defined as follows:

Fitness $(X_i) = E(X_i)$

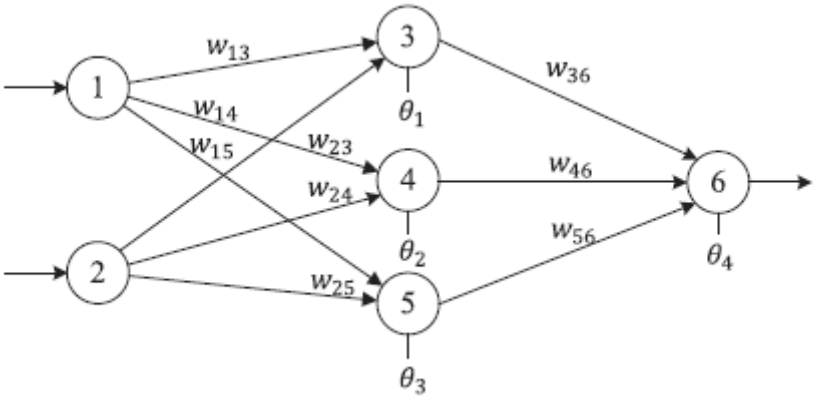

FIGURE 2: FNN with a 2-3-1 structure

3.3 Encoding strategy

Following the specification of the fitness function for FNN-GWO, the next step is to choose an encoding strategy for each agent in FNN-GWO to encode the FNN's weights and biases. In matrix encoding strategy, each agent is encoded as a matrix. An example of this encoding strategy for the FNN of Figure 2 is provided as follows:

$$\text{particle}(:,:,\text{i}) = [W_1, B_1, W_2', B_2] \quad (8)$$

$$W_1 = \begin{bmatrix} w_{13} & w_{23} \\ w_{14} & w_{24} \\ w_{15} & w_{25} \end{bmatrix}, \quad B_1 = \begin{bmatrix} \Theta_1 \\ \Theta_2 \\ \Theta_3 \end{bmatrix}, \quad W_2' = \begin{bmatrix} w_{36} \\ w_{46} \\ w_{56} \end{bmatrix}, \quad B_2 = [\Theta_4] \quad (9)$$

where W_1 is the hidden layer weight matrix, B_1 is the hidden layer bias matrix, W_2 is the output layer weight matrix, W_2' is the transpose of W_2, and B_2 is the hidden layer bias matrix.

3.4 Grey Wolf Optimization (GWO): at a glance

Mirjalili et. Al. [28] introduced GWO, a new stochastic and metaheuristic optimization algorithm. Two important characteristics of meta heuristic optimization are exploration and exploitation. Exploration means how quickly an optimization algorithm able to identify the potential area where the global minima may exist from all feasible large search area. Once this potential area is found, exploitation means how quickly and accurately detect the exact point where the global minima exist by detail search. In literature, it is found that GWO algorithm able to make a delicate balance between exploration and exploitation by emulating the cooperative hunting behavior of grey wolves in nature. This bionic optimization method promotes the grey wolf pack's hierarchical system and predation behavior, in which the wolves take prey by surrounding, haunting, and attacking it under the command of the head grey wolf [31]. In this large-scale search methodology, all wolves form a rigid hierarchical pyramid as depicted in Figure 3. but decision is based on the top three grey wolves siting at the top of pyramid.

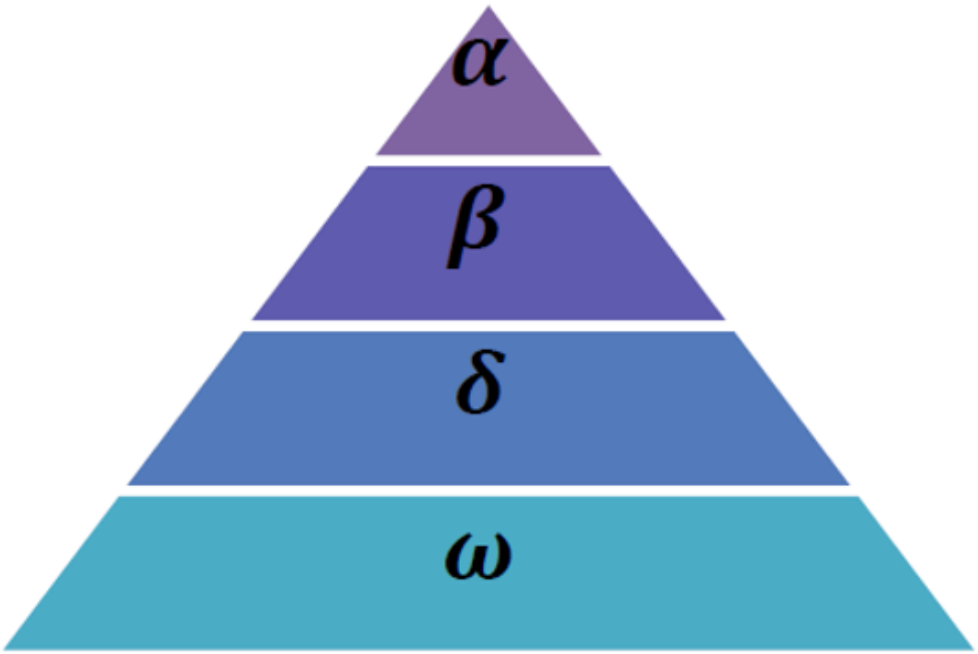

FIGURE 3: Hierarchy of grey wolf (dominance decreases from top down)

The average group size is 5-12 wolves. The top layer (α layer), which consists of a male and female leader, is the team's most powerful and capable member for making decisions concerning predation and other activities. The second and third layers in the hierarchy, β and δ layer respectively, are responsible for assisting in the behavior of group organizations. Most of the overall populace resides in the bottom of the pyramid, which is referred to as ω layer. They are primarily in charge of satisfying the entire pack by balancing the population's internal relationships, caring for the young, and sustaining the dominance hierarchy [28].

3.5. Mathematical modelling of GWO:

The main key point of the GWO model is the social hierarchy, encircling, hunting, attacking and searching prey.

A) Social hierarchy

In the model of GWO, α is considered as fittest solution. β and δ are considered as second and third best solution, respectively. The rest of the solutions are assumed to be ω.

B) Encircling

At first, the location of the prey is determined and during the hunting process grey wolves encircled the prey. The following equations are proposed for mathematically modelling [28].

$$\overrightarrow{D_p} = \left|\vec{C}.\overrightarrow{X_p}(t) - \vec{X}(t)\right| \quad (10)$$

$$\vec{X}(t+1) = \overrightarrow{X_p}(t) - \vec{A}.\left(\overrightarrow{D_p}\right) \quad (11)$$

Where t is the number of current iterations, $\vec{X}(t)$ is the position vector of one grey wolf, $\vec{X}(t+1)$ is the next position vector it arrives, $\overrightarrow{X_p}(t)$ is the position vector of the prey, $\vec{A}$ and $\vec{C}$ are coefficient vectors which are evaluated as follows:

$$\vec{A} = 2\vec{a}.\overrightarrow{r_1} - \vec{a} \quad (12)$$

$$\vec{C} = 2.\vec{r_2} \quad (13)$$

r_1 and r_2 are random vectors in [0,1], a is decreasing value during the iteration in [0,2], typically $a = 2 - 2t/I$ (I is the maximum number of iterations).

In this concept, grey wolves move around the best solution in hyper-cubes within an n dimensional space and able to detect the position of the prey and encircle it.

C) Hunting

After encircling the prey, grey wolves have the ability to hunt the prey with the guidance of alpha. The beta and delta also take part occasionally in hunting process. In the mathematical stimulation of hunting behaviour, the first three best solutions update the position of other search agents (including the omegas). The following equations are proposed in this regard.

$$\overrightarrow{D_\alpha} = \left|\overrightarrow{C_1}.\overrightarrow{X_\alpha} - \vec{X}(t)\right|;\ \overrightarrow{D_\beta} = \left|\overrightarrow{C_2}.\overrightarrow{X_\beta} - \vec{X}(t)\right|;\ \overrightarrow{D_\gamma} = \left|\overrightarrow{C_3}.\overrightarrow{X_\gamma} - \vec{X}(t)\right| \quad (14)$$

$$\overrightarrow{X_1} = \overrightarrow{X_\alpha}(t) - A_1\left(\overrightarrow{D_\alpha}\right);\ \overrightarrow{X_2} = \overrightarrow{X_\beta}(t) - A_2.\left(\overrightarrow{D_\beta}\right);\ \overrightarrow{X_3} = \overrightarrow{X_\gamma}(t) - A_3.\left(\overrightarrow{D_\gamma}\right) \quad (15)$$

$$\overrightarrow{X_p}(t+1) = \frac{X_1+X_2+X_3}{3} \quad (16)$$

D) Attacking prey (exploitation)

Grey wolves complete the hunting phase and ready to capture the prey. For the purpose of mathematically modelling, the value of $\vec{a}$ gradually decrease. Therefore, the fluctuation rate of $\vec{A}$ is also decreased by $\vec{a}$. $\vec{A} \epsilon [-2a, 2a]$ where a is decremented from 2 to 0 over the course of iterations. When the random values of $\vec{A} \epsilon [-1,1]$, the next position of search agent can be in any position between its current position and the position of the prey. When $|A| \geq 1$, the grey wolves would diverge from the prey to achieve global search and when $|A| \leq 1$, the grey wolves would converge towards the prey and complete it.

E) Searching prey (exploration)

GWO algorithm uses an efficient exploration methodology by allowing its search agents to update their position on the basis of alpha, beta, delta and attacks towards the prey. This mechanism creates a good diversity in the problem search space. Grey wolves are stay away from each other for global search of prey and close to each other for attacking the prey. For $\vec{A}$, values between 1 and -1 are taken. $\vec{C}$ vector is also favoured the exploration technique and random values $[0,2]$ are used. After generating the random population, alpha, beta and delta determined the position of the best prey. For selection of exploration and exploitation, the value of $\vec{a}$ is decreased from 2 to 0 respectively.

The GWO algorithm terminates when the criterion is satisfied. This metaheuristic approach is applied in various real-world problems because of its efficient and simple performance ability by tuning the fewest operators [28-32].

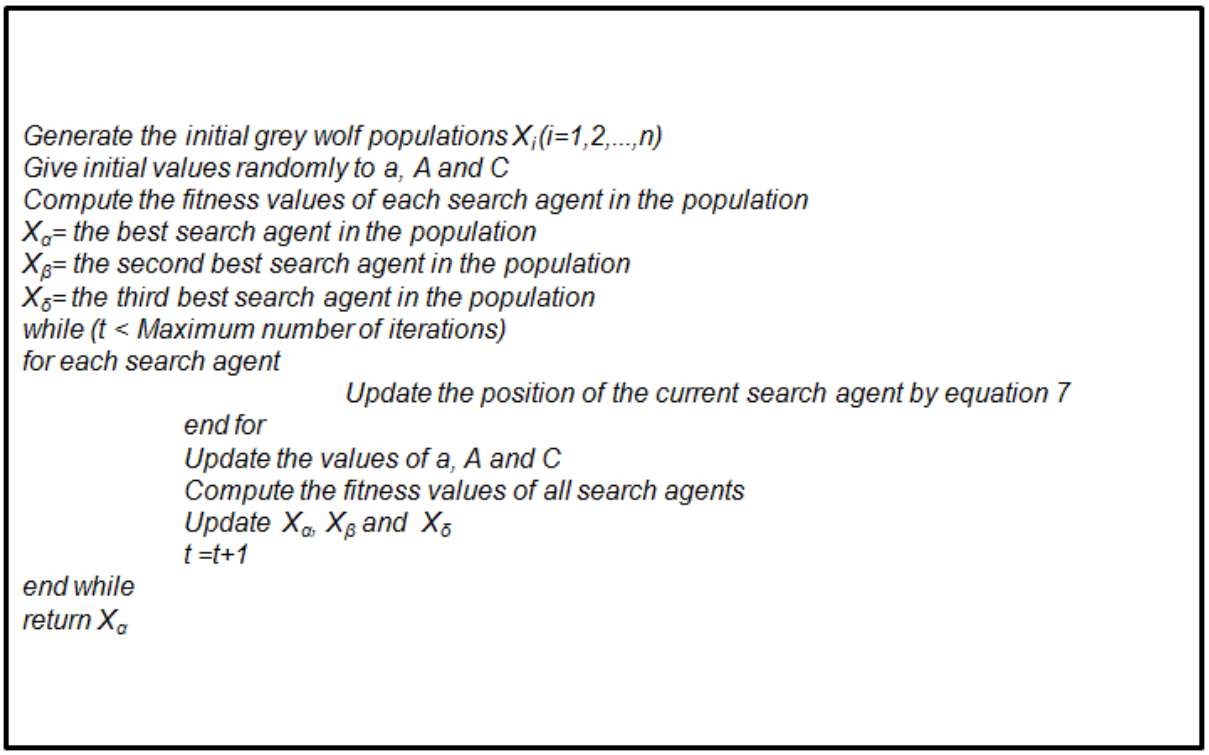

FIGURE 4: Pseudo code of Grey Wolf Optimisation

3.6. Hybrid learning FNN network by GWO

Essentially the training algorithm of FNN is an optimization problem where optimized set of weigths and biases are to be determined which minimizes the error (MSE) beween actual output and predicted output. The optimization capability of GWO is utilised for FNN learning in this part. To improve the accuracy of the network, a hybrid learning of GWO and FNN networks (GWO-FNN) is used. By training the network, the algorithm will concurrently determine the set of weights and their corresponding accuracy. The network weights and biases of the FNN network can be represented as a D-dimensional vector. Equation 17 defines the vector for FNN (4). The dimension of each particle is treated as the vector D as given below:

D = (Input × Hidden) + (Hidden × Output) +Hidden bias + Output bias, (17)

where Input, Hidden and Output are referred the number of input, hidden and output neurons of FNN network respectively. The number of biases in the hidden and output layers is also known as Hidden bias and Output bias. Figure 5 depicts the flowchart of GWO and FNN hybrid leaning (GWO-FNN). A dataset is collected, normalised, and read to begin the GWO-FNN. Following that, the appropriate number of input, output, and hidden neurons are specified to establish the particle dimension as Equation (7). The population is initialised, and the training error is determined as a fitness function given in equation 17 following FNN training. Every particle (wolf) modifies its velocity and position based on training error. The new places represent the FNN network's new weights, which are supposed to minimise the fitness function. The fitness function is computed based on test set error. These steps will go on until meeting stop conditions.

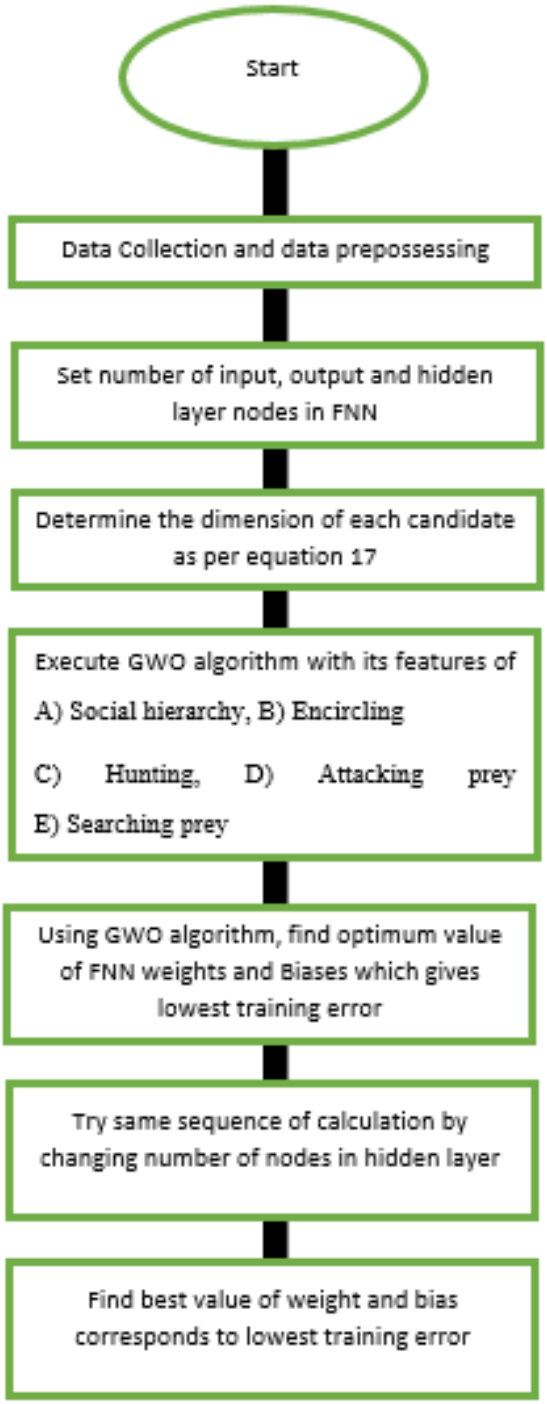

FIGURE 5: Flowchart of Hybrid FNN-GWO methodology

4. Mathematical modelling and optimization of EO reaction process

4.1 Selection of input and output variables for modelling

Because selectivity and reactor temperature have such a significant impact on plant profitability, they are preserved as output variables. Selectivity was chosen because it directly depicts plant production costs and has a significant impact on plant profitability. Reactor temperature, on the other hand, impacts catalyst longevity and thus adds a fixed cost to plant production. Faster temperature rises imply a shorter catalyst life and a higher catalyst cost. As a "wish list" of input variables, all reactor operational parameters that potentially effect selectivity and reactor temperature are kept. For the first three years, all EO reactor operational data (daily average) was gathered. Then, after consulting with a plant expert, all of the input variables that could affect the output variables were recorded. Following that, a cross-correlation study was performed. The correlation coefficients of each of the input variables with each of the output variables, as well as the inter input cross-correlation coefficients, were determined using this procedure.

The input variables are shortlisted using the criteria below.

(i) There should be a strong cross-correlation coefficient between input and output variables for a given input variable.

(ii) The values of inter input variable cross-correlation coefficients should be low.

(iii) To keep the model as simple as feasible, the input set of variables was kept to a minimum.

Eight input variables are finally shortlisted and tabulated in Table 1 based on the aforesaid criteria.

TABLE 1 Input Output variables for model building and their range

Variables used in modeling	**Data Range**
Input Variables	
Oxygen inlet concentration, mole% (x_1)	4.38 - 7.75
C_2H_4 inlet concentration, mole% (x_2)	17.69 - 34.97
CO_2 inlet concentration, mole% (x_3)	0.21 - 0.45
CG Pressure, bar (x_4)	21.84 – 22.64
CG Flow (MT/h) (x_5)	687.78 – 803.39
Work Rate (Kg/hr. /m3) (x_6)	104.38 – 202.47
Cumulative EOE (MT/m^3) (x_7)	275.28 – 4469.32
Total chloride concentration, PPM (x_8)	2.51 – 5.86
Output Variables	
Catalyst selectivity, % (Y_1)	81.00 – 89.44
Reactor Temperature, °C, (Y_2)	225.83– 250.75

4.2. Data collection, Data cleaning and removal of outliers

Data historian software is used in commercial facilities to gather and store every operating parameter at every minute. We acquired EO reactor operating data (daily average) for all parameters listed in Table 1 from a commercial plant that had been operational for nearly three years. Following that, the raw data is subjected to data cleaning and outlier removal tools.

Because industry data contains noise, spikes, outliers, and error as a result of malfunctioning sensors, transmitters, analyzers, control systems, and data historian software, data cleansing is one of the most important tasks in industrial process modelling. It is commonly established that the quality of data utilized to construct a data-driven model determines the model's quality. Data quality is an important issue to consider when using data driven modelling because noisy and faulty data can have a significant

impact on model performance. Due to the large volume of process data, this study created an automatic effective data cleaning technique to prevent manual cleaning. From the plant operating dataset, an automated MATLAB-based algorithm was constructed to generate a multivariate statistical vector called t-squared. The corresponding rows in the t-squared vector with values over the 95th percentile was therefore considered outliers and were eliminated from the dataset.

4.3. Modelling through FNN-GWO algorithm

The cleansed data was used to create an ANN-based model. The dataset, which included eight input variables and two output variables (Table 1), was randomly partitioned into a training set (80% of the total data) and a test set (20% of the total data). The training data set was used to compute the model by maximizing the fitness value, while the test data set was utilized to cross-validate the produced expression. The basic goal of cross-validation is to improve the model's generalizability.

The FNN-GWO-based model was created using code written in MATLAB 2019a. The fitness function used in this work was the mean-squared error (MSE) between actual and predicted outputs, and the program was run in such a way that the MSE value of test set data was minimized. The software was run 100 times to construct the model due to the stochastic nature of the ANN.

4.4. Optimization through Multi objective grey wolf optimization (MOGWO)Algorithm

The model is then used to optimize the reactor process parameter once a reliable and accurate EO reactor model has been built. The goal is to establish the ideal process condition that maximizes the profitability of the EO plant, and to do so, the following conflicting criteria must be met:

(i) The selectivity of the catalyst must be maximized.

(ii) The reactor temperature must be kept as low as possible.

In the present work, multi-objective grey wolf optimization (MOGWO) algorithm is used to strike a balance between two conflicting objectives. MOGWO has proven itself to be an excellent multi objective optimization tool and has been used extensively in diverse field of engineering and medicine.

As there are two conflicting objectives in this particular case study, MOGWO is used to generate pareto optimal solutions. For implementation of MOGWO algorithm, two objective functions were developed which are as follows (Equation 18 and 19):

$$F_1(x) = 1/(1+Y_1(x)) \quad (18)$$

$$F_2(x) = Y_2(x) \quad (19)$$

Where $Y_1(x)$ and $Y_2(x)$ are functions of the models corresponding to catalyst selectivity and reactor temperature respectively. Therefore, in MOGWO both $F_1(x)$ *and* $F_2(x)$ have to be minimized in order to minimize reactor temperature and maximize catalyst selectivity.

5. Results and discussions

5.1. Performance of FNN-GWO model

The main goal of this research is to develop a closed model equation for an EO reactor that is accurate, simple, portable, and easy to understand.

The three-year daily average dataset was cleaned, and 1220 data sets were qualified for model construction after outliers were removed. A trial-and-error approach and a literature study were used to determine the values of ANN parameters required for modelling. The number of nodes in the hidden layer in the current study varies systematically from 5 to 25, and the FNN GWO method is employed each time to discover the best weights and bias that generate the lowest MSE between actual and predicted output. These procedures give 21 [each model for each number of nodes] different ANN model with varying accuracy.

Shortlisting the models: The following criteria were used to pick a viable model from a pool of probable candidates or representative model equations with varied degrees of complexity and accuracy:

(i) Simplicity: The model should have as little complexity as feasible. The number of nodes in the hidden layer was used to determine model complexity.

(ii) Accuracy of prediction: The difference between expected and actual selectivity.

(iii) The model equation should capture the process's underlying physics. In other words, model equations should contain physical understanding of the system under study, not just a predictive association. This is a crucial factor to consider while creating realistic reactor models. Their plant operating knowledge, experience and observations of reactor behavior are summarized in Table 2.

TABLE 2 Rules to select best model from real Plant operation observations.

Sl. no	Parameters changed keeping all other parameters constant	What happen to selectivity?	What happen to temperature?
1	If oxygen inlet concentration increase	Increase	Decrease
2	If ethylene inlet concentration increase	Increase	Decrease
3	If CO_2 inlet concentration increase	Decrease	Increase
4	If CG Pressure increase	Increase	Decrease
5	If CG flow increase	Increase	Decrease
6	If Work rate increase	Decrease	Increase
7	If Cum EOE/m^3 of catalyst increase	Decrease	Increase

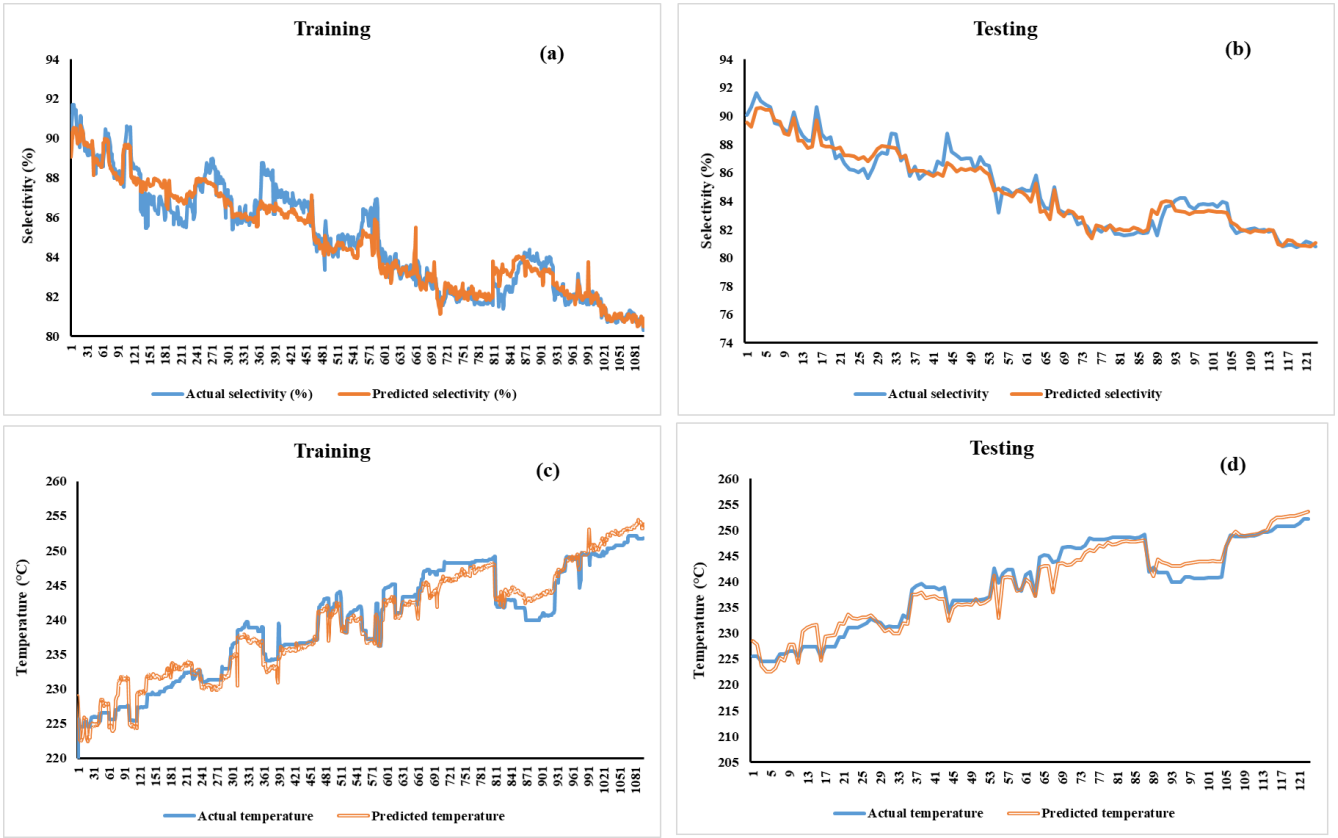

FIGURE 6: Actual vs. predicted plots of (a) catalyst selectivity with training data (b) catalyst selectivity with testing data (c) reactor temperature with training data (d) reactor temperature with testing data

All 21 models acquired using ANN are subjected to the aforementioned inspection to determine whether the created model is in agreement with plant data. All of the produced models have been thoroughly tested. Assume a ten-test data set was constructed with all variables set to their 50th percentile value, with the exception of oxygen inlet concentration, which was altered in 10-point intervals from its minimum to maximum value. The charts of oxygen inlet concentration vs. selectivity and oxygen inlet concentration vs. temperature were created after that. Table 2 observation number 1 was confirmed using these charts. Models that do not match the findings in Table 2 are discarded because they do not represent the basic physics of the EO reactor and do not correlate with plant observations. They are essentially representations of a complicated data fitting model that makes no sense.

Only one model for catalyst selectivity and one model for reactor temperature were chosen from the shortlisted models because they are exceedingly precise, obey the Table 2 observations, and capture the internal physics of the reactor, such two models are considered the representative model equations for selectivity and temperature.

The corresponding Coefficient of Determination (R^2) and Average Percentage Error (APE) of the above models for training and test data have been mentioned in Table 3.

TABLE 3. Performance of ANN model

	Training		**Testing**	
Model	**R^2**	**APE**	**R^2**	**APE**
Catalyst selectivity model	0.99	0.56	0.99	0.56
Reactor temperature model	0.99	0.24	0.99	0.23

With high R^2 and low APE for both the catalyst selectivity and reactor temperature models (Table 3), it can be concluded that the predicted output values are comparable to the actual output values and that the models developed are reliable, fairly accurate, and capture the inherent physics of the EO reactor. The model's generalizability and accurate learning on nonlinear input and output relationships are further indicated by the strong R^2 value on unseen test data and low APE.

Figure 6 depicts the models' prediction performance on training and testing data. The fact that the real and forecast curves almost coincide implies that the model has strong prediction accuracy.

From Table 3 and Figure 6 it is concluded that developed model is highly accurate and reliable as it also performs well with unseen test data.

Generation of explainable model equations: Though ANN generates a closed form of equation which has very high predictive capability, the developed equation is large and complex and sometimes difficult to directly interpret. In present study, a methodology is developed to enhance the interpretability of the developed equations. Figure 7 and Figure 8 summarize the developed methodology. Those figures are generated by changing one variable at a time from its minimum to maximum value (10 steps) while keeping all other 7 input variables at their 50-percentile value. Selectivity equations developed by ANN is used to predict the selectivity value in each case of these simulated test data. After plotting was done, a trendline was drawn through each data whose equation and R^2 value is shown in figure. Based on visual inspection and R^2 value trend line curve was selected (like straight line, or polynomial with degree 2 or 3 or more) so that generated trendline almost matches with the data. As seen from the figure 10, the developed trend lines are very decisive and monotonically increasing and decreasing. As mentioned earlier, they all match the plant actual observations and obey the Table 2. In short, developed models captures the nonlinear relationship between selectivity and reactor operating parameters. These trend lines can be used by plant operating engineers to get the insights how a particular input parameter affects the catalyst selectivity. For example, from figure 7, it is quantitively clear that increasing inlet oxygen and ethylene concentration, actually enhance the catalyst selectivity whereas increasing CO_2 reduces the selectivity.

The selectivity increases linearly with oxygen concentration with positive slope of 0.6772, whereas the relation of selectivity with ethylene and CO_2 are nonlinear and represented by second order polynomial. Now these trend line equations are used to develop the following explainable equation (Equation 20 and 21).

$$Selectivity = 0.6772\,(x_1 - x_{avg}) - 0.0004(x_2{}^2 - x_{avg}{}^2) + 0.1376\,(x_2 - x_{avg}) - 7.4768(x_3 - x_{avg}) - 0.1628\,(x_4 - x_{avg}) - 0.0076\,(x_5 - x_{avg}) - 0.0268\,(x_6 - x_{avg}) - 0.0017\,(x_7 - x_{avg}) - 0.0014(x_8{}^2 - x_{avg}{}^2) - 0.5158\,(x_8 - x_{avg}) + 84.86 \quad (20)$$

$$Temperature = -0.8435\,(x_1 - x_{avg}) + 0.0446\,(x_2 - x_{avg}) + 1.0246\,(x_3 - x_{avg}) - 2.2509\,(x_4 - x_{avg}) + 0.0001\,(x_6{}^2 - x_{avg}{}^2) + 0.0808\,(x_6 - x_{avg}) - 0.0066\,(x_7 - x_{avg}) + 0.269\,(x_8 - x_{avg}) + 240.61 \quad (21)$$

Where $x_1, x_2, \ldots x_8$ are the actual value of the 8 input variables and $x_{1,avg}, x_{2,avg}, \ldots x_{8,avg}$ are the average (50 percentile) value of input variables, respectively.

Each term in the equation 20 represents the change in selectivity if a particular parameter deviates from its average value. For example, the term, $(x_1 - x_{1,avg})$ represents the deviation ox inlet oxygen concentration from its average value and when it multiplied by co-efficient 0.6772, represents the selectivity gain (or penalty) due to oxygen. In this way, all 8 parameters contribution is calculated in equation 8 and it is added with 84.86% (average selectivity) to get the actual selectivity.

Main advantage of this equation (equation 20 and 21) over ANN model is that this equation is interpretable and easily explainable to plant engineer. Equation is simple and contains terms or parametric co-efficient which throw lights relative importance of each parameter on the overall selectivity if they deviate from this base value. Also, it indicates whether the effect of each parameter is linear or non-linear.

Equation 20 is then used to predict selectivity of three-year actual data and predicted, and actual selectivity is compared. Prediction error is 0.8% and R^2 is 0.97. This low value of prediction error and high value of R^2 signifies that developed equation (equation 20) is highly accurate and reliable.

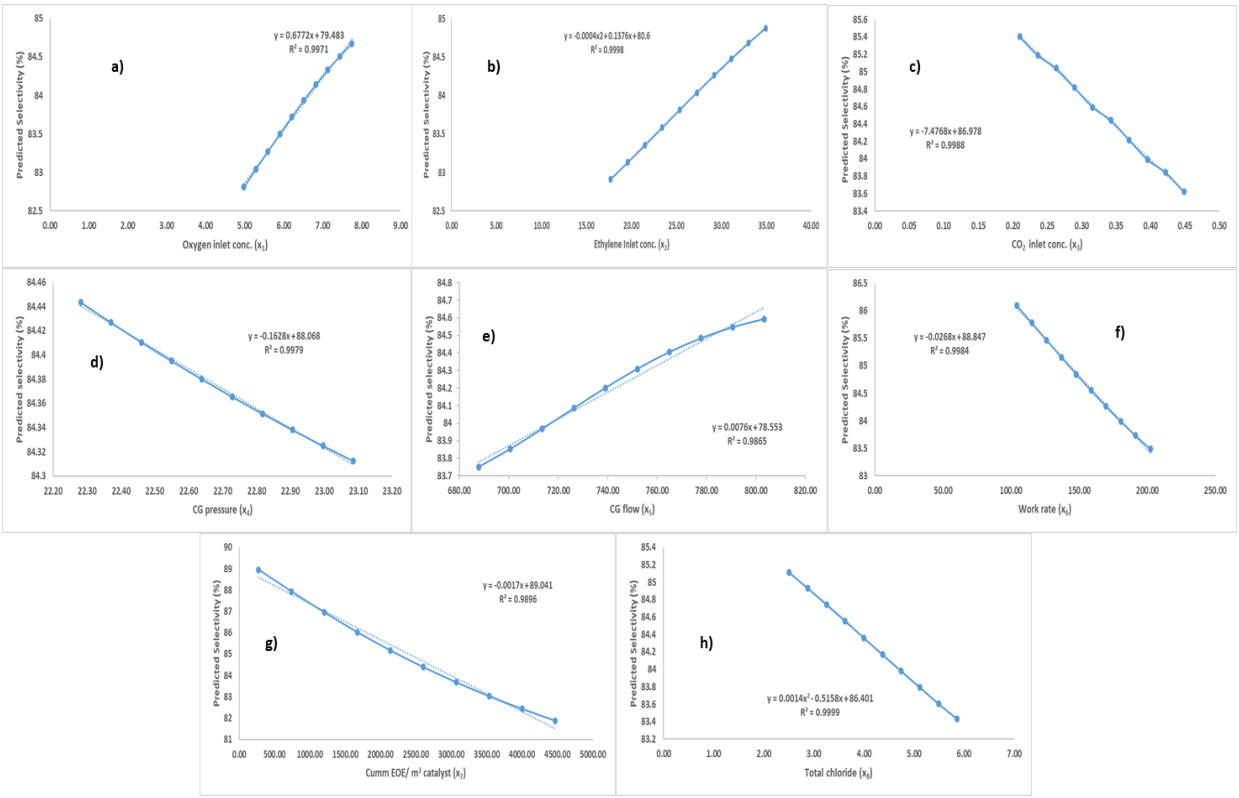

FIGURE 7: Influence of each variable on selectivity

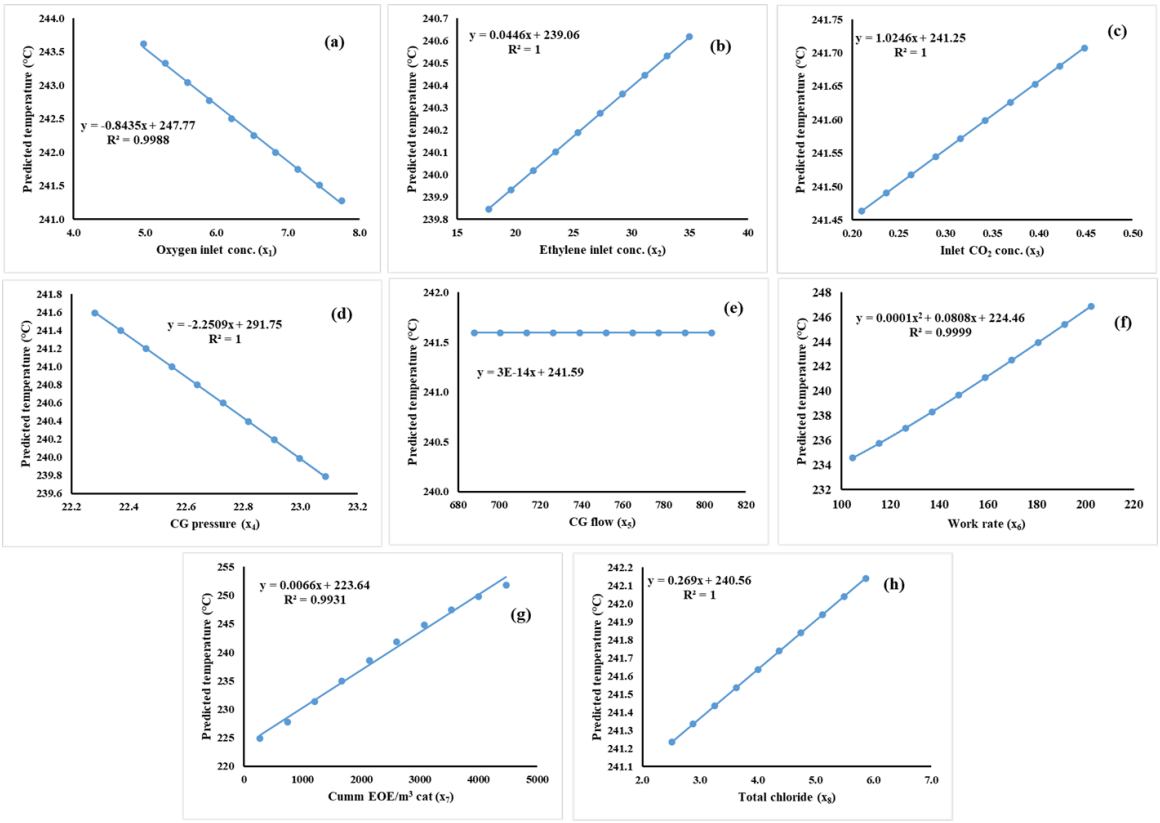

FIGURE 8: Influence of each variable on reactor temperature

5.2. Optimization

5.2.1. Optimization through MOGWO

Once the reliable models were successfully developed, the models were subjected to optimization. Purpose of optimization is to find the optimum value of rector operating parameters to achieve maximum selectivity and minimum temperature simultaneously. One of the critical tasks for optimization of any process is fixing of search space at which the optimal process conditions are to be found out. Therefore, before running the optimization a lower bound and upper bound of the process variables were fixed in consultation with plant engineers. As plant conditions are dynamic in nature, the total one operational year was divided into three periods (SOR, MOR and EOR) and at each running period, optimizations were performed. The lower bounds (LB) and upper bounds (UB) considered in these three cases have been depicted in Table 4.

Table 4. Lower bounds and upper bounds for optimization

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
Case 1 : SOR								
LB	3377	6.37	27.41	1.32	0.15	18.94	5.13	3455.62
UB	3387	9.00	35.12	2.50	0.24	19.95	9.22	4585.31
Case 2: MOR								
LB	5377	6.37	27.41	1.52	0.15	18.94	6.13	3455.62
UB	5387	9.00	35.12	2.50	0.24	19.95	9.22	4585.31
Case 3: EOR								
LB	8377	6.37	27.41	1.82	0.15	18.94	7.63	3455.62
UB	8387	9.00	35.12	2.50	0.24	19.95	9.22	4585.31

With the help of MOGWO code in MATLAB, a pareto diagram was developed for case 2 (MOR) (Fig. 8). All the points in the pareto curve in Figure. 9 represents the pareto non-dominated optimal solution in catalyst selectivity and reactor temperature space and all Pareto optimal solutions on the curve are considered equally good. As seen from the curve, if someone try to increase selectivity, he must sacrifice reactor temperature. Reverse is also true, i.e., lower temperature can be achieved at the cost of lower selectivity. Other than these solutions, there is no process condition which can increase selectivity without sacrificing temperature and vice versa. Some of the pareto optimal solutions have been also tabulated in Table 5 for reference. From the output of MOGWO the optimum values of input parameters can be obtained which will maximize the selectivity and minimize the temperature. Four solutions in table 5 are plotted as four small black circles on Figure 9 and represents trade-off between two conflicting objectives. All the four solutions are equally good, and it is up to the plant operation engineer which criteria among selectivity or temperature they will give priority. Based on that, they can choose the optimum input parameters from the Pareto optimal solution. For example, if selectivity has more priority than reactor temperature for a particular plant, then operators should choose solution number 4 in Table 5 which will help to achieve 86.1% selectivity by sacrificing the temperature (241.53 °C). On the other hand, if the temperature has more priority than selectivity then solution number 1 in Table 4 should be chosen which will give a lower temperature by sacrificing selectivity. If temperature and selectivity have equal priority, then solution number 2 of Table 5 should be chosen.

The main advantage of such a study is that it gives the operation engineer/ DCS panel operators a strategy to run the reactor in optimum condition in real-time. In a running plant, since operators have no idea about Pareto optimal solution, operators try to optimize the plant heuristically based on their experience and knowledge. One such real-life operating point is plotted and named as 'A' in Figure. 9. As the operating point, A is not on the Pareto curve, it is not the optimum point and improvement scope exists both in selectivity and temperature. From this plot in Figure. 9, it can be concluded that the operator can improve the plant operating conditions in two ways: selectivity can be improved keeping the temperature same as before, or reactor temperature can be decreased keeping the catalyst selectivity as it is as shown by two arrows on Figure 9. The only action operator must do is to run the MOGWO with proper bound in real-time, and MOGWO will provide a set of optimum operating conditions that the operator needs to set in the plant. This is the potential improvement area suggested by the Pareto optimal solution.

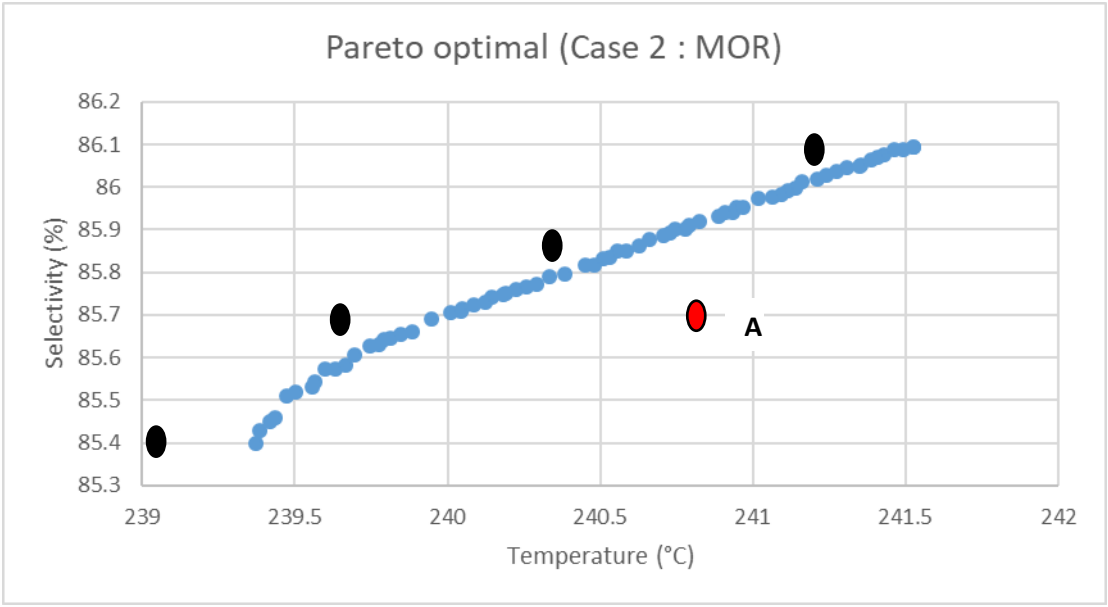

FIGURE 9. Pareto diagram of catalyst selectivity vs. reactor temperature

Table 5. Pareto optimal solution

x1	x2	x3	x4	x5	x6	x7	x8	Selectivity	Temperature
7.75	18.25	0.31	22.64	782.15	161.09	2572.84	2.59	85.40	239.37
7.75	28.09	0.31	22.54	785.62	161.08	2572.84	2.52	85.71	240.01
7.74	28.18	0.31	22.17	785.93	161.09	2572.84	2.52	85.92	240.82
7.74	28.71	0.31	21.86	785.96	161.07	2572.84	2.53	86.10	241.53

6. Conclusion

From the existing plant operating data, this study uses Artificial Neural Networking to construct an accurate model of a commercial ethylene oxide reactor. An ANN creates a closed model equation that is portable and may be used in a plant control system. The true value of this research is that it has produced an explainable model equation that is very accurate and provides insights into the process. The produced model equations are based on the underlying physics of the process and are in line with the observations and experiences of the plant engineer. After that, the developed model equations are used to construct multi-objective pareto optimum solutions that optimize selectivity while lowering reactor temperature and thus ensure profit maximization.

REFERENCES

[1] Lahiri, S.K., 2020. Profit Maximization Techniques for Operating Chemical Plants. John Wiley & Sons.

[2] Lakshminarayanan, S., Fujii, H., Grosman, B., Dassau, E., Lewin, D.R., 2000. New product design via analysis of historical databases. Comput. Chem. Eng. 24, 671–676. https://doi.org/10.1016/S0098-1354(00)00406-3

[3] Grosman, B., Lewin, D.R., 2002. Automated nonlinear model predictive control using genetic programming. Comput. Chem. Eng. 26, 631–640. https://doi.org/10.1016/S0098-1354(01)00780-3

[4] Searson, D., Willis, M., Montague, G., 2007. Co-evolution of non-linear PLS model components. J. Chemom. 21, 592–603. https://doi.org/10.1002/cem.1084

[5] Barati, R., Neyshabouri, S.A.A.S., Ahmadi, G., 2014. Development of empirical models with high accuracy for estimation of drag coefficient of flow around a smooth sphere: An evolutionary approach. Powder Technol. 257, 11–19. https://doi.org/10.1016/j.powtec.2014.02.045

[6] Floares, A.G. and Luludachi, I., 2014. Inferring transcription networks from data. In Springer Handbook of Bio-/Neuroinformatics (pp. 311-326). Springer, Berlin, Heidelberg.

[7] Horne, B.G., 1993. Progress in supervised neural networks. Signal Process. Mag. IEEE 10, 8–39.

[8] Hagan, M.T., Menhaj, M.B., 1994. Training Feedforward Networks with the Marquardt Algorithm. IEEE Trans. Neural Networks 5, 989–993. https://doi.org/10.1109/72.329697

[9] Adeli, H., Hung, S.L., 1994. An Adaptive Conjugate Gradient Learning Algorithm. Appl. Math. Comput. 62, 81–102.

[10] Zhang, N., 2009. An online gradient method with momentum for two-layer feedforward neural networks. Appl. Math. Comput. 212, 488–498. https://doi.org/10.1016/j.amc.2009.02.038

[11] Zhang, J.R., Zhang, J., Lok, T.M., Lyu, M.R., 2007. A hybrid particle swarm optimization-back-propagation algorithm for feedforward neural network training. Appl. Math. Comput. 185, 1026–1037. https://doi.org/10.1016/j.amc.2006.07.025

[12] Gori, M., Tesi, A., 1992. On the problem of local minima in backpropagation. IEEE Trans. Pattern Anal. Mach. Intell. https://doi.org/10.1109/34.107014

[13] Shaw, D., Kinsner, W., 1996. Chaotic simulated annealing in multilayer feedforward networks. Can. Conf. Electr. Comput. Eng. 1, 265–269. https://doi.org/10.1109/ccece.1996.548088

[14] Koh, C.S., Hahn, S.Y., 1994. Detection of Magnetic Body using Artificial Neural Network with Modified Simulated Annealing. IEEE Trans. Magn. 30, 3644–3647. https://doi.org/10.1109/20.312730

[15] Montana, D.J., Davis, L., 1989. Training Feedforward Neural Networks Using Genetic Algorithms. Proc. 11th Int. Jt. Conf. Artif. Intell. - Vol. 1 89, 762–767.

[16] Kiranyaz, S., Ince, T., Yildirim, A., Gabbouj, M., 2009. Evolutionary artificial neural networks by multi-dimensional particle swarm optimization. Neural Networks 22, 1448–1462. https://doi.org/10.1016/j.neunet.2009.05.013

[17] Settles, M., Rylander, B., 2002. Neural network learning using particle swarm optimizers. Adv. Inf. Sci. Soft Comput. 224–226.

[18] Zhang, C., Li, Y., Shao, H., 2000. A new evolved artificial neural network and its application. Proc. World Congr. Intell. Control Autom. 2, 1065–1068. https://doi.org/10.1109/wcica.2000.863401

[19] van den Bergh, F., Engelbrecht, A.P., 2000. Cooperative Learning in Neural Networks using Particle Swarm Optimizers. South African Comput. J. 26, 84–90.

[20] Zhang, C., Shao, H., Li, Y., 2000. Particle swarm optimization for evolving artificial neural network. Proc. IEEE Int. Conf. Syst. Man Cybern. 4, 2487–2490. https://doi.org/10.1109/icsmc.2000.884366

[21] Mirjalili, S., Sadiq, A.S., 2011. Magnetic Optimization Algorithm for training Multi Layer Perceptron. 2011 IEEE 3rd Int. Conf. Commun. Softw. Networks, ICCSN 2011 42–46. https://doi.org/10.1109/ICCSN.2011.6014845

[22] Si, T., Hazra, S., Jana, N.D., 2012. Artificial neural network training using differential evolutionary Algorithm for classification. Adv. Intell. Soft Comput. 132 AISC, 769–778. https://doi.org/10.1007/978-3-642-27443-5_88

[23] Leung, Y., Gao, Y., Xu, Z. Ben, 1997. Degree of population diversity - A perspective on premature convergence in genetic algorithms and its Markov chain analysis. IEEE Trans. Neural Networks 8, 1165–1176. https://doi.org/10.1109/72.623217

[24] Hrstka, O., Kučerová, A., 2004. Improvements of real coded genetic algorithms based on differential operators preventing premature convergence. Adv. Eng. Softw. 35, 237–246. https://doi.org/10.1016/S0965-9978(03)00113-3

[25] Liang, J.J., Qin, A.K., Suganthan, P.N., Baskar, S., 2006. Comprehensive learning particle swarm optimizer for global optimization of multimodal functions. IEEE Trans. Evol. Comput. 10, 281–295. https://doi.org/10.1109/TEVC.2005.857610

[26] Gao, W. feng, Liu, S. yang, Huang, L. ling, 2012. Particle swarm optimization with chaotic opposition-based population initialization and stochastic search technique. Commun. Nonlinear Sci. Numer. Simul. 17, 4316–4327. https://doi.org/10.1016/j.cnsns.2012.03.015

[27] Moslemipour, G., Lee, T.S., Rilling, D., 2012. A review of intelligent approaches for designing dynamic and robust layouts in flexible manufacturing systems. Int. J. Adv. Manuf. Technol. 60, 11–27. https://doi.org/10.1007/s00170-011-3614-x

[28] Mirjalili, S., Mirjalili, S.M., Lewis, A., 2014. Grey Wolf Optimizer. Adv. Eng. Softw. 69, 46–61. https://doi.org/10.1016/j.advengsoft.2013.12.007

[29] Faris, H., Aljarah, I., Al-Betar, M.A., Mirjalili, S., 2018. Grey wolf optimizer: a review of recent variants and applications. Neural Comput. Appl. 30, 413–435. https://doi.org/10.1007/s00521-017-3272-5

[30] Saremi, S., Mirjalili, S.Z., Mirjalili, S.M., 2015. Evolutionary population dynamics and grey wolf optimizer. Neural Comput. Appl. 26, 1257–1263. https://doi.org/10.1007/s00521-014-1806-7

[31] Nadimi-Shahraki, M.H., Taghian, S., Mirjalili, S., 2021. An improved grey wolf optimizer for solving engineering problems. Expert Syst. Appl. 166, 113917. https://doi.org/10.1016/j.eswa.2020.113917

[32] Mirjalili, S., Saremi, S., Mirjalili, S.M., Coelho, L.D.S., 2016. Multi-objective grey wolf optimizer: A novel algorithm for multi-criterion optimization. Expert Syst. Appl. 47, 106–119. https://doi.org/10.1016/j.eswa.2015.10.039

[33] Pal, S., Chowdhury, S., Hens, A., & Lahiri, S. K. (2022). Artificial intelligence based modelling and multi-objective optimization of vinyl chloride monomer (VCM) plant to strike a balance between profit, energy utilization and environmental degradation. Journal of the Indian Chemical Society, 99(1), 100287.