Predicting the performance of a desulfurizing bio-filter using an artificial neural network (ANN) model

Reza Salehi†, Retno Ambarwati Sigit Lestari²

1Independent Researcher, 3495 Saint-Dominique, Montreal, Quebec H2X 2X5, Canada
2Chemical Engineering Department, 17 Agustus 1945 University, Semarang, Jawa Tengah, Indonesia

ABSTRACT
The aim of this study was to develop a model for predicting the performance of a desulfurizing bio-filter (BF), without requiring prior information about H2S biodegradation kinetics and mechanism. A single hidden layer artificial neural network (ANN) model was developed and validated using the gradient descent backpropagation (GDBP) learning algorithm coupled with a learning rate and a momentum factor. The ANN model inputs were gas flow rate, residence time, and axial position in the BF bed. The removal efficiency of H2S was the model output. Various structures for ANN model, differing in the number of hidden layer neurons, were trained and an early stopping validation technique, the K-fold cross-validation, was used to determine the optimal structure with the best generalization ability. The modeling results showed that there was a good agreement between the experimental data and the predicted values, with a determination coefficient (R²) of 94%. This implies that the ANN model might be an attractive and useful alternative tool for forecasting the performance of desulfurizing BFs.

Keywords: Artificial neural network, Backpropagation algorithm, Bio-filter, Cross-validation, Hydrogen sulfide removal

1. Introduction
Biogas, an alternative energy source, is mostly produced from both anaerobic wastewater treatment processes [1, 2] and anaerobic digestion of organic material such as food waste, agricultural and industrial residues, animal manure, and sewage sludge [3]. The biogas can contain hydrogen sulfide (H2S) from a trace amount [4] to extremely high values ranging between 20,000 and 40,000 ppm [5, 6]. Therefore, good care must be taken when biogas is used in internal combustion engines (ICEs) and boilers for electricity and thermal energy generation because the corrosive nature of H2S, especially in the presence of water vapor, causes the metal equipment to wear down. As the recommended H2S concentration acceptable in ICEs and boilers are below 100 ppm and 1,000 ppm, respectively [7], H2S-rich biogas must be treated before it can be used for heat and electricity generation.

H2S removal from biogas can be performed using physical/chemical techniques based on (1) adsorption on a solid with a high-surface area such as activated carbon and iron oxides or (2) absorption in water or chemical solvents such as sodium hydroxide and aqueous solutions of bivalent metal sulfates [8-10]. The main disadvantage associated with the adsorption techniques is that the adsorbents, as they are saturated over time, must be expensively regenerated or treated as a hazardous waste. Moreover, the adsorbents are often expensive [11, 12]. Regarding the absorption techniques (water/chemical scrubbing), they generate a contaminated liquid stream that needs to be treated. Therefore, a post-treatment step is required, which results in high operational costs. Other disadvantages are high consumption of water and high cost of chemical solvents [13, 14]. In contrast, biogas desulfurization using biofiltration is an attractive and potential alternative to the physicochemical techniques because of its low operational cost and low environmental impact [15]. In desulfurizing bio-filters (BFs), a stream of biogas containing H2S passes through a biofilm of sulfide oxidizing bacteria (SOB) immobilized on a nutrient-rich porous packed media. H2S diffuses from the gas phase into the biofilm where it is metabolically consumed by SOB and degraded to elemental sulfur (S0) and/or sulfate (SO4²⁻)[16]. Chemotrophic SOB are the most widely studied group of microorganisms for biological desulfurization processes. They drive energy from chemical reactions such as oxidation of H2S or S0 under different environmental conditions, aerobic (OX) or anoxic (AX), and utilize carbon dioxide (CO2) or organic...
carbon compounds as the carbon source [15]. An OX condition is one in which dissolved oxygen is present as an electron acceptor, while an AX condition is one in which nitrate (NO$_3^-$) is served as electron acceptor [17]. Chemotrophic SOB from the genera Acidithiobacillus, Thermothrix, Thioalkalispira, Thiolithrix, and Thiouvulum have shown great potential to oxidize H$_2$S under OX condition, and the genera such as Beggiatoa, Thioanaragata, Thioploca, Thioalkalivibrio, and Thiobacillus are capable of oxidizing H$_2$S under either OX or AX condition [18].

In an OX desulfurizing BF, the following reactions (1-4) occur [19].

$$\begin{align*}
H_2S(g) & \rightarrow H_2S(aq) \\
H_2S(aq) & \leftrightarrow H^+(aq) + HS^-(aq) \\
HS^-(aq) + 0.5 O_2(aq) & \rightarrow S^0(s) + OH^-(aq) \\
HS^-(aq) + 2 O_2(aq) & \rightarrow SO_4^{2-}(aq) + H^+(aq)
\end{align*}$$

Initially, H$_2$S is absorbed in the biofilm (reaction 1), followed by its dissociation to HS$^-$ and hydrogen ion (H$^+$) (reaction 2), which is dependent on pH. Then after, HS$^-$ is biologically oxidized to S$^0$ and/or SO$_4^{2-}$ depending on the availability of oxygen (reactions 3 and 4). Under limited-oxygen conditions, reaction (3) proceeds in which the oxidation of HS$^-$ ends up with production of S$^0$. In contrast, in the presence of sufficient amount of oxygen, the final product resulting from HS$^-$ oxidation is SO$_4^{2-}$ (reaction 4). An O$_2$/HS$^-$ molar ratio of 0.7 would favor the production of S$^0$ as the dominant final product, while an O$_2$/HS$^-$ molar ratio of greater than 1.0 is required to obtain a significant conversion of HS$^-$ to SO$_4^{2-}$ [20-22]. It should be mentioned that when HS$^-$ is limited instead of oxygen, and S$^0$ is present, SO$_4^{2-}$ is formed according to reaction (5) [10].

$$S^0(s) + 1.5 O_2(aq) + H_2O(l) \rightarrow SO_4^{2-}(aq) + 2 H^+(aq)$$

In an AX desulfurizing BF, HS$^-$ oxidation can be represented by reactions 6 and 7 [10], in which the production of S$^0$ and/or SO$_4^{2-}$ depends on the NO$_3^/$HS$^-$ ratio. Note that reactions 6 and 7 are based on complete denitrification (reduction of NO$_3^-$ to N$_2$ without nitrite (NO$_2^-$) accumulation).

$$\begin{align*}
HS^-(aq) + \frac{1}{3} NO_3^-(aq) & \rightarrow S^0(s) + \frac{1}{6} N_2(g) + OH^-(aq) \\
HS^-(aq) + \frac{4}{3} NO_3^-(aq) & \rightarrow SO_4^{2-}(aq) + \frac{2}{3} N_2(g) + H^+(aq)
\end{align*}$$

Based on the stoichiometry of reactions 6 and 7, if the NO$_3^/$HS$^-$ molar ratio is 0.33, HS$^-$ can be oxidized mainly to S$^0$, while HS$^-$ oxidation to SO$_4^{2-}$ requires an NO$_3^/$HS$^-$ molar ratio of 1.33. Similar NO$_3^/$HS$^-$ ratio were reported by some researchers who achieved the complete oxidation of HS$^-$ to SO$_4^{2-}$ with an NO$_3^/$HS$^-$ molar ratio between 1.2 and 1.6 whereas HS$^-$ oxidation ended up with S$^0$ at an NO$_3^/$HS$^-$ molar ratio of less than 0.4 [23-26].

As mentioned above, in both OX and AX desulfurizing BFs, the principal products of HS$^-$ oxidation are SO$_4^{2-}$ and S$^0$. If S$^0$ is the dominant end-product, the major drawback is the accumulation of S$^0$ within the BF media, which causes clogging and increases the pressure drop across the BF column [27]. In case where the complete oxidation of HS$^-$ occurs, the BF medium pH would be expected to lower into acidic zone due to the formation of SO$_4^{2-}$ and H$^+$. However, several studies have shown that some groups of SOB could resist and maintain their activity under acidic environment. For instance, Acidithiobacillus thiooxidans showed high activity for degradation of H$_2$S at a pH of 1.5-3.0 [28-30]. Thiobacillus thiooxidans could tolerate a pH swing between 2.0 and 0.5, and Acidithiobacillus thiooxidans AZ71 could grow at a pH as low as 0.2 for H$_2$S oxidation [30].

In recent years, several modeling studies have been conducted to predict the performance of the desulfurizing BFs [12, 31-33]. The developed models are complex, with a large number of parameters because biofiltration is a multi-phase system in which physical, chemical, and biological reactions occur. In addition to the complexity, the mass transfer and kinetic parameters are often difficult to measure accurately [34]. Furthermore, most of the models have been simplified by incorporating some assumptions [12, 33], which may lead to an understimation of the model output [12].

To tackle these challenges, an artificial neural network (ANN) model was developed to predict the performance of a desulfurizing BF based on the experimental data reported in Lestari et al. [12]. The neural networks have advantage over the conventional mathematical methods for modeling of complex biological systems, because (1) they are based only on an actual measured set of input and output variables, without requiring prior information about the interrelationship between the variables, and (2) they possess strong generalization and prediction ability [13, 35]. To the best of the authors' knowledge, the application of ANN in desulfurizing BF modeling has been few reported so far [13, 36-38], despite the wide use of the desulfurizing BFs. Therefore, this study was conducted to enrich this research gap. The simulation results obtained in this study was compared with that of the mathematical model introduced by Lestari et al. [12].

The rest of this paper is organized as follows. Section 2 presents ANN modeling approach along with a brief description of the gradient descent backpropagation algorithm. Section 3 covers the methodology including the source of the data used for this study, identification of the ANN structure and the ANN modeling process (training and testing phases). Section 4 discusses the results of the modeling exercises. Finally, the paper ends with conclusions presented in section 5.

## 2. Artificial Neural Network (ANN)

ANN, a sort of machine learning, is a powerful computational technique in simulating and predicting the behavior of complex processes in a similar way to that of human brain [35, 39]. An ANN has a multiple-layer structure with one layer for each input and output, and one or more layers between the input and output layers, called hidden layers, with the aim of performing complex computations on the input data before directing them to the output layer. Each layer is composed of a number of simple processors...
called cells, more commonly known as nodes or neurons. The number of neurons in the input layer and the output layer equals the number of input variables and the number of output variables, respectively. However, the number of neurons in the hidden layers is an adjustable parameter that is optimized via trial and error method. Any neuron either in the input layer or in the hidden layer sends out a weighted input to each neuron in the downstream layer. A weighted input refers to the input multiplied by its associated synaptic weight. A synaptic weight is the strength of the connection between two neurons. The input layer neurons have no processing function. In other words, they only act as buffers to distribute the input data to the neurons within the hidden layer. However, in the hidden layer, each neuron applies two processing functions on its input. Initially, the weighted inputs are aggregated and an activation function is then applied on the aggregated value to generate its output. In a similar manner, the output layer neurons produce the network output, which is then compared to the target output to evaluate the network error [13, 40, 41].

Let us consider a three-layer ANN, illustrated in Fig. 1, consisting of one input layer, one hidden layer, and one output layer.

A crucial part of the neural network modeling is to choose an appropriate algorithm to train the network, in other words, to adjust the network weights in such a way that the error function of the network is minimized. The gradient descent backpropagation (GDBP) algorithm, introduced by Rumelhart et al. [42], has been widely and successfully used for training numerous multi-layer neural networks. The GDBP is an iterative process in which each iteration (epoch) is composed of two passes, namely the forward pass and the backward pass. In the forward pass, the neurons within each layer of the network deliver their outputs forward to the next layer neurons until the final output of the network is obtained. Then, the error between the actual output and the network output is computed. If the error is greater than the pre-specified threshold, the backward pass starts. In the backward pass, the partial derivative of the error function with respect to all the weights in the network are propagated from the output layer backward to the input layer to adjust the network weights. The forward and backward passes are repeated until either the maximum number of the training epochs is reached or the network error is within the pre-specified threshold [42]. It should be mentioned that since the GDBP algorithm deals with the gradient of the network error function at each epoch, the error function must be a differentiable and continuous function. To guarantee this, the neurons activation function can be, for instance, the sigmoid or the hyperbolic tangent function. According to the literature, the GDBP algorithm is computationally efficient, and conceptually straightforward to implement. Furthermore, this algorithm is accurate and offers a rapid convergence when the momentum factor, which helps avoid the oscillation of the weights that might happen as the training algorithm proceeds, and the learning rate are properly chosen [43-46]. Hence, in this study, the GDBP algorithm with a learning rate and a momentum factor was used to train the proposed neural network model.

Using the GDBP, the network weights are modified according to equation (8).

$$W^{(t+1)} = W^{(t)} - a \frac{\partial E}{\partial W} + \beta (W^{(r)} - W^{(r-1)})$$

where,

$$\frac{\partial E}{\partial W_{ij}} = -[f(X_{ij}, W_{ij})] \cdot Y_{i} - Y_{i}$$

$$\circ g'(f(X_{ij}, W_{ij}), \bar{W}_{ij})$$

$$\frac{\partial E}{\partial \tilde{W}_{ij}} = -[X_{ij}]^T \cdot [((Y_{i} - Y_{i})]$$

$$\circ g'(f(X_{ij}, \tilde{W}_{ij}), \bar{W}_{ij}) \cdot [\bar{W}_{ij}]^T$$

$$\circ f'(X_{ij}, \tilde{W}_{ij})$$

(9)
Table 1. Axial H₂S Concentration and Corresponding H₂S Removal Efficiency along the BF Bed as a Function of Gas Flow Rate and Residence Time; Data from Lestari et al. [12]

<table>
<thead>
<tr>
<th># obs</th>
<th>Input variable</th>
<th>Output variable</th>
<th># obs</th>
<th>Input variable</th>
<th>Output variable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x₁</td>
<td>x₂</td>
<td>x₃</td>
<td>y (ppm)</td>
<td>y (%)</td>
</tr>
<tr>
<td>1</td>
<td>8550</td>
<td>0</td>
<td>0</td>
<td>179.62</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>8550</td>
<td>0</td>
<td>20</td>
<td>163.90</td>
<td>8.75</td>
</tr>
<tr>
<td>3</td>
<td>8550</td>
<td>0</td>
<td>40</td>
<td>119.67</td>
<td>33.38</td>
</tr>
<tr>
<td>4</td>
<td>8550</td>
<td>0</td>
<td>60</td>
<td>98.24</td>
<td>45.31</td>
</tr>
<tr>
<td>5</td>
<td>8550</td>
<td>0</td>
<td>80</td>
<td>59.00</td>
<td>67.15</td>
</tr>
<tr>
<td>6</td>
<td>8550</td>
<td>2</td>
<td>0</td>
<td>163.90</td>
<td>0.00</td>
</tr>
<tr>
<td>7</td>
<td>8550</td>
<td>2</td>
<td>20</td>
<td>138.22</td>
<td>15.67</td>
</tr>
<tr>
<td>8</td>
<td>8550</td>
<td>2</td>
<td>40</td>
<td>96.26</td>
<td>41.27</td>
</tr>
<tr>
<td>9</td>
<td>8550</td>
<td>2</td>
<td>60</td>
<td>39.11</td>
<td>76.14</td>
</tr>
<tr>
<td>10</td>
<td>8550</td>
<td>2</td>
<td>80</td>
<td>23.44</td>
<td>85.70</td>
</tr>
<tr>
<td>11</td>
<td>8550</td>
<td>4</td>
<td>0</td>
<td>142.48</td>
<td>0.00</td>
</tr>
<tr>
<td>12</td>
<td>8550</td>
<td>4</td>
<td>20</td>
<td>104.40</td>
<td>26.73</td>
</tr>
<tr>
<td>13</td>
<td>8550</td>
<td>4</td>
<td>40</td>
<td>49.00</td>
<td>65.61</td>
</tr>
<tr>
<td>14</td>
<td>8550</td>
<td>4</td>
<td>60</td>
<td>24.15</td>
<td>83.05</td>
</tr>
<tr>
<td>15</td>
<td>8550</td>
<td>4</td>
<td>80</td>
<td>4.06</td>
<td>97.15</td>
</tr>
<tr>
<td>16</td>
<td>13680</td>
<td>0</td>
<td>0</td>
<td>87.85</td>
<td>0.00</td>
</tr>
<tr>
<td>17</td>
<td>13680</td>
<td>0</td>
<td>20</td>
<td>82.69</td>
<td>5.87</td>
</tr>
<tr>
<td>18</td>
<td>13680</td>
<td>0</td>
<td>40</td>
<td>75.60</td>
<td>13.94</td>
</tr>
<tr>
<td>19</td>
<td>13680</td>
<td>0</td>
<td>60</td>
<td>53.03</td>
<td>39.66</td>
</tr>
<tr>
<td>20</td>
<td>13680</td>
<td>0</td>
<td>80</td>
<td>41.49</td>
<td>52.77</td>
</tr>
<tr>
<td>21</td>
<td>13680</td>
<td>2</td>
<td>0</td>
<td>89.99</td>
<td>0.00</td>
</tr>
<tr>
<td>22</td>
<td>13680</td>
<td>2</td>
<td>20</td>
<td>73.31</td>
<td>18.54</td>
</tr>
<tr>
<td>23</td>
<td>13680</td>
<td>2</td>
<td>40</td>
<td>56.96</td>
<td>36.70</td>
</tr>
<tr>
<td>24</td>
<td>13680</td>
<td>2</td>
<td>60</td>
<td>37.89</td>
<td>57.90</td>
</tr>
<tr>
<td>25</td>
<td>13680</td>
<td>2</td>
<td>80</td>
<td>15.49</td>
<td>82.79</td>
</tr>
<tr>
<td>26</td>
<td>13680</td>
<td>4</td>
<td>0</td>
<td>89.51</td>
<td>0.00</td>
</tr>
<tr>
<td>27</td>
<td>13680</td>
<td>4</td>
<td>20</td>
<td>60.98</td>
<td>31.87</td>
</tr>
<tr>
<td>28</td>
<td>13680</td>
<td>4</td>
<td>40</td>
<td>45.92</td>
<td>48.70</td>
</tr>
<tr>
<td>29</td>
<td>13680</td>
<td>4</td>
<td>60</td>
<td>22.17</td>
<td>75.23</td>
</tr>
<tr>
<td>30</td>
<td>13680</td>
<td>4</td>
<td>80</td>
<td>5.06</td>
<td>94.35</td>
</tr>
<tr>
<td>31</td>
<td>18810</td>
<td>0</td>
<td>0</td>
<td>41.44</td>
<td>0.00</td>
</tr>
<tr>
<td>32</td>
<td>18810</td>
<td>0</td>
<td>20</td>
<td>39.56</td>
<td>4.54</td>
</tr>
</tbody>
</table>

Notes: x₁: gas flow rate (g m⁻³ h⁻¹); x₂: residence time (h); x₃: distance from the BF inlet (cm); y: axial H₂S concentration (or H₂S removal efficiency) along the BF bed; “obs” stands for “observation”.

\[
\begin{align*}
\hat{y}_{i|x} &= g( f(X_{i|k}; \hat{W}_{i|x}), \hat{W}_{i|x} ) \\
E &= \frac{1}{2} (Y_{i|x} - \hat{Y}_{i|x})^2 = \frac{1}{2} (Y_{i|x} - \hat{Y}_{i|x}) \odot (Y_{i|x} - \hat{Y}_{i|x})
\end{align*}
\]

The notations in equations (8-12) are as follows.
- \(X_{i|k}\): Network input matrix
- Symbol “\(\odot\)” represents an element wise multiplication of two matrices
- Superscript “T” in \([.]^T\) refers to the transpose of matrix \([.]\)
- \(Y_{i|x}\) and \(\hat{Y}_{i|x}\) are the matrices assigned to the desired output values and the output values computed using the network, respectively

Subscripts:
- k: number of input variables
- l: number of neurons within the hidden layer
- j: number of output variables
- i: dataset size
3. Methodology

3.1. Data Acquisition

Data were obtained from the study of Lestari et al. [12], who operated a lab-scale BF for desulfurization of a biogas stream containing 10-180 ppm of H2S. Salak fruit seeds (SFS) was used as the BF packing material, which was immobilized with SOB from the genus *Thiodacillus* isolated from the sludge of the municipal wastewater treatment plant in Surakarta (Yogyakarta, Indonesia). The BF, with an inner diameter of 8 cm, had a total height and a packing height of 100 cm and 80 cm, respectively. A series of experiments were carried out to evaluate the performance of the BF, in term of H2S removal efficiency, as function of the axial distance from the BF inlet (0-80 cm), gas flow rate (8,550 to 23,940 g m⁻³h⁻¹) and residence time (up to 4 h) (Table 1).

3.2. ANN Structure

In this study, an ANN model was developed to predict the performance of a BF for H2S removal from a gas stream. The ANN model was composed of three layers: input layer, hidden layer and output layer. The input layer contained three neurons, one each for gas flow rate (x₁), residence time (x₂) and axial position in the BF bed (distance from the BF inlet) (x₃). The output layer generating the model output had one neuron that was H2S removal efficiency (y). In the hidden layer, the optimal number of neurons was determined via trial and error method. The hyperbolic tangent function, one of the most widely used activation function in ANN models as it allows the model to learn non-linear relationships, was used for the neurons within the hidden and the output layers. The mathematical definition of the hyperbolic tangent function is given by equation (13).

\[
\phi(S_m) = \frac{2}{1 + \exp(-2S_m)} - 1 \quad 0 \leq \phi(S_m) \leq 1
\]  

where, “S_m” denotes the sum of the weighted inputs entering a neuron, and \( \phi(S_m) \) represents the output of that neuron.

3.3. ANN Modelling Process

The ANN modelling process was performed in two phases including training and testing. In order to speed up the model convergence and increase prediction accuracy, the data needs to be within a small range. As seen in Table 1, the input data used in this study were either in the order of thousands or in single digits. Hence, the min-max technique (Eq. (14)) was applied using which all the input data points fall within the range [0,1].

\[
x_{i,N} = \frac{x_i - x_{i,\text{min}}}{x_{i,\text{max}} - x_{i,\text{min}}}
\]  

where, “x_{i,N}” represents the normalized value of “x_i”, which is the actual value of input variable “i”, “x_{i,\text{min}}” and “x_{i,\text{max}}” are the minimum and maximum values of input variable “i”, respectively.

Similarly, the target values were normalized using equation (14) to be included within the operational range of the activation function [0-1].

After data normalization, the entire dataset was randomly split into two different subsets as follows.

1) Training dataset: It was used to identify the optimal number of hidden layer neurons (HLNs) and to adjust the model synaptic weights to minimize the error function. The training dataset was further divided into “K” subsets in order to use the cross validation technique to determine the optimal number of iterations (epochs) at which the model training should be stopped.

2) Testing dataset: It was served to assess the accuracy and predictive capability of the model after training phase was complete.

There are no general rules to measure the ratio between the size of training and testing datasets. However, many researchers have reported that the training dataset size is in the range 60-85% of the entire dataset [47, 48]. Hence, out of 60 experimental observations (input-output data pairs) applied to the ANN model in this study, 45 and 15 observations were assigned to the training and testing, respectively. The training and testing datasets, in form of normalized, for the ANN model are given in Table S1.

3.3.1. ANN model training

ANN model training is an iterative process through which the model learns the input-output behavior. The training phase was performed using the GDBP learning algorithm (Eq. (8)-(12)). The model performance was assessed using two statistical indices, namely root mean squared error (RMSE)-that represents the square root of the average squared differences between the target value and the model output value-and coefficient of determination (R²)-that is a goodness-of-fit measure for the model-defined by equations (15) and (16), respectively. The closer R² value to unity, and the smaller RMSE value (closer to zero), the better the model fits the data. In other words, the model perfectly fits the data when R² value is equal to 1.0 and RMSE value is equal to 0.

\[
\text{RMSE} = \sqrt{\frac{1}{s \times p} \sum_{i=1}^{s} \sum_{j=1}^{p} (a(i,j) - \hat{a}(i,j))^2}
\]  

\[
R^2 = 1 - \frac{\sum_{i=1}^{s} \sum_{j=1}^{p} (a(i,j) - \hat{a}(i,j))^2}{\sum_{i=1}^{s} \sum_{j=1}^{p} (a(i,j) - \bar{a})^2}
\]  

where, a(i,j) denotes the elements of matrix \( Y_{\text{exp}} \) assigned to the desired output values; \( \hat{a}(i,j) \) represents the elements of matrix \( \hat{Y}_{\text{exp}} \) assigned to the output values computed using the network; \( \bar{a} \) is the mean value of a(i,j); parameters “s” and “p” stand for the dataset size and the number of output variables, respectively.

The ANN model training always faces a big dilemma that is after how many epochs the training phase should be stopped because an over-trained model, due to learning the noise instead of signals, may show poor performance on an/the unseen dataset. The stop-training criterion based on the K-fold cross-validation is one the most common procedure to avoid the model from an overtraining. An example of a K-fold cross-validation, graphically demonstrated in Fig. 2, is described as follows [49].

1) The training dataset is partitioned into “K” disjoint folds where each fold contains the same number of samples.
2) “K” runs are performed such that within each run, the model is trained on (K-1) folds.
3) The trained model is then evaluated on the remaining fold (termed as validation fold) to estimate its RMSE.
4) The average of RMSE on the validation folds is plotted against the number of epochs.

The averaged RMSE usually decreases through the initial course of training and begins to increase as soon as the network starts over-fit the data. In other words, when RMSE stops decreasing while the number of epochs increases, the training phase should be stopped.

3.3.2. ANN model testing

After the model training phase was complete, the trained model was tested against the testing dataset (unseen during the training phase) to assess the predictive capability of the model. It should be mentioned that when the model training and testing phases were complete, the output values were anti-normalized to their actual values.

3.4. Software

The ANN modelling was performed using a code written in MATLAB environment (version 8.3, R2014a, MathWorks Inc., USA). The graphs were generated using Microcal Origin software (version 6.0, Microcal Software Inc., Massachusetts, USA).

4. Results and Discussion

Lestari et al. [12] conducted a series of experiments, as briefly described in the subsection 3.1, to evaluate the performance of a desulfurizing BF in term of H2S removal efficiency. The authors developed a mathematical model as a predictive tool for the assessment of changes in H2S removal efficiency with respect to the gas flow rate, axial distance from the BF inlet, and residence time (experimental observations represented in Table 1). The mathematical model, took into account mass transfer of H2S from the gas stream into the biofilm and biological oxidation of H2S in the biofilm, consisted of a set of ordinary differential equations (ODEs). The solution was obtained by Runge-Kutta method. The authors found a good agreement between the experimental data and the model-predicted values, with $R^2$ value of 86.56% [12].

In this study, using a code written in MATLAB environment, an ANN model was developed and validated for predicting the performance of a desulfurizing BF based on the experimental data reported in Lestari et al. [12]. The inputs to the ANN model were gas flow rate, residence time and axial position in the BF bed, and the model output was H2S removal efficiency. Seventy five percent of the original (parent) dataset (45 input-output data pairs) devoted to the training process in order to build the model, and the remaining 25% of the parent dataset (15 input-output data pairs) was applied to evaluate generalization capability of the trained model. Two statistical error measures such as $R^2$ and RMSE were used to assess the model performance. In addition, the predictive capability of the proposed ANN model was compared with that of the mathematical model introduced by Lestari et al. [12].

Various ANN structures differing in the number of HLNs, from 1 to 12, were trained. The training phase for each ANN structure was conducted according to the following steps:
1) Each ANN structure was initially fed with the normalized training dataset, and the synaptic weights were randomly assigned small values between zero and one.
2) The GDBP learning algorithm was utilized during which the parameters $\alpha$ and $\beta$ were adjustable. The best values for $\alpha$ and $\beta$ were determined by changing their values from 0.01 to 1.0 (0.01, 0.1, 0.2, ..., 0.9, and 1.0).
3) The training phase continued until RMSE (or its corresponding $R^2$) value remained constant, otherwise the training was stopped at 10000 epochs.

4) The ANN performance was evaluated in terms of RMSE and $R^2$ given by Eqs. (15) and (16), respectively. The higher $R^2$ (or lower RMSE), the better the ANN fits the data. Fig. 3 illustrates the ANN training RMSE and $R^2$ curves versus the number of HLNs. From Fig. 3, as the number of HLNs increased, there was an increase in $R^2$ value and a decrease in RMSE value attaining $R^2$ and RMSE values of 99.24% and 0.028, respectively, with 9 HLNs. Further increase in the number of HLNs, from 9 to 12, did not result in a significant enhancement in the ANN performance. Hence, the optimal structure of the ANN was found to be 3-9-1 (3 neurons in the input layer, 9 HLNs and 1 neuron in the output layer) as illustrated in Fig. 4.

![Fig. 3](image1)

**Fig. 3.** Training curves in terms of $R^2$ and RMSE to determine the optimal number of HLNs for the ANN model (HLNs: hidden layer neurons).

![Fig. 4](image2)

**Fig. 4.** Architecture of the ANN model used in this study; “$x_{in}$” and “$y_{n}$” are the normalized values of “$x$” and “$y$”; See Table 1 for “$x$” and “$y$” notations and values, and Table S1 for “$x_{in}$” and “$y_{n}$” values; $w$ and $w'$ are the synaptic weights between the input and hidden layers, and between the hidden and output layers, respectively; $\hat{y}$ represent the predicted $H_2S$ removal efficiency by ANN model; GDBP: Gradient descent backpropagation.

After obtaining the optimal structure for the ANN model, 3-9-1 topology, it was assessed whether that model was correctly trained or an under/overtraining occurred. An undertraining leads to a poor training performance, and an overtraining sometimes negatively affects the model generalizability. In other words, the model performance and generalization capability depend on the number of epochs at which the training phase is stopped. To figure out when is the best to stop the ANN model training, a 9-fold cross-validation method was applied to the training dataset. Fig. 5 displays the validation RMSE curve computed by the 9-fold cross-validation along with the training RMSE curve as a function of the number of epochs. In Fig. 5, any point on the validation curve represents the averaged RMSE on the validation folds (termed generalization RMSE). When the generalization RMSE begins to rise, the training should be stopped. As depicted in Fig. 5, the ANN training was stopped at epoch 500.

![Fig. 5](image3)

**Fig. 5.** Training and validation RMSE curves as a function of the number of epochs for the proposed 3-9-1 ANN model.

The performance of the proposed ANN model was evaluated by constructing the scattered diagram of the predicted versus measured values of $H_2S$ removal efficiency with respect to the training and testing datasets (Fig. 6(a) and Fig. 6(b)). In Fig. 6 (a) and Fig. 6(b), the solid line is the best-fit line indicating 93.45% and 95.11% correlation between the measured and predicted values for the training and testing phases, respectively. These results imply that the proposed ANN model was successfully trained with a good predictive performance.

The prediction capability of the ANN model developed in this study, which was trained with the aid of the GDBP algorithm coupled with a learning rate and a momentum factor, was compared with that of Lestari’s mathematical model [12] in Fig. 7. As shown in Fig. 7, the accuracy of the ANN model ($R^2 = 93.8\%$) was superior compared to Lestari’s mathematical model [12], which showed $R^2$ value of 86.6%. The higher accuracy of the ANN model, could be attributed to the fact that it was based only on the measured set of the input and output variables, without making any assumption about the interrelationship between the variables, whereas the mathematical model of Lestari et al. [12] was developed based on two fundamental simplifying assumptions; (1) pseudo-steady state flow condition for the gas phase, and (2) the uniformity of $H_2S$ concentration in the biofilm at a given axial position from the BF inlet. As mentioned earlier, simplifying mathematical models may result in an underestimation of the model output [12].
5. Conclusions

This study dealt with a modeling exercise performed on the available published work of Lestari et al. [12], in order to investigate the capability of the artificial neural network (ANN) technique in forecasting the performance of a desulfurizing bio-filter (BF). From the modeling results obtained in this study, a single hidden layer ANN model with 9 hidden neurons and the hyperbolic function showed a predictive accuracy higher than that of the mathematical model developed by Lestari et al. [12], with the coefficient of determination ($R^2$) value of 93.83% and 86.56%, respectively. This implies that the ANN model could be an attractive and useful tool that is worth considering for predicting the desulfurizing BF's performance as it can be effectively implemented, without requiring prior information about H$_2$S biodegradation kinetics and mechanism.

Nomenclature

- $a_{i,j}$ elements of matrix $Y$ assigned to the desired output values
- $\tilde{a}_{i,j}$ elements of matrix $\tilde{Y}$ assigned to the output values computed using the network;
- $\bar{a}$ mean value of $a_{i,j}$
- ANN artificial neural network
- AX anoxic
- BF bio-filter
- E error function
- $f(.)$ activation function of neurons within the hidden layer
- $f'(.)$ derivative of $f(.)$ with respect to the synaptic weights
- $g(.)$ activation function of the output layer neuron
- $g'(x)$ derivative of $g(.)$ with respect to the synaptic weights
- GDBP gradient descent backpropagation
- HLN hidden layer neurons
- ICE internal combustion engine
- MEA mono-ethanolamine
- $N_T$ training dataset size
- ODE ordinary differential equation
- OX aerobic
- ppm part per million
- $R^2$ coefficient of determination
- RMSE root mean squared error
- $S_{in}$ sum of the weighted inputs entering a neuron either in the hidden layer or in the output layer
- SFS salak fruit seeds
- SOB sulfide oxidizing bacteria
- $\tilde{w}$ synaptic weights between the input layer and hidden layer of the network
- $w_{k,i}$ “$\tilde{w}$” in matrix form, where subscripts “$k$” and “$i$” stand for the number of input variables and the number neurons within the hidden layer, respectively
- $\tilde{W}$ synaptic weights between the hidden layer and output layer of the network
- $\tilde{w}_{l,j}$ “$\tilde{W}$” in matrix form, where subscripts “$l$” and “$j$” denote the number neurons within the hidden layer and the number of output variables, respectively
$x_1$ gas flow rate (g m$^{-3}$ h$^{-1}$)
$x_2$ residence time (h)
$x_3$ axial position in the BF bed (distance from the BF inlet) (cm)
$x_{\text{max}}$ maximum value of “$x_i$”
$x_{\text{min}}$ minimum value of “$x_i$”
$X_{i,k}$ network input in for of matrix, where “i” and “k” subscripts represent the data set size and the number of input variables
$y$ measured H$_2$S removal efficiency (%)
$y_N$ normalized value of “$y_i$”
$y_{(i)}$ “y” in matrix form, where subscripts “$i$” and “$j$” are the data set size and the number of output variables, respectively
$\hat{y}$ predicted H$_2$S removal efficiency (%)
$\tilde{y}_{(i)}$ “$\tilde{y}$” in matrix form, where subscripts “$i$” and “$j$” are the data set size and the number of output variables, respectively

**Symbols**

$[.]^T$ transpose of matrix $[,]$
$\odot$ element wise multiplication of two matrices

**Greek letters**

$\alpha$ learning rate
$\beta$ influence of $(r-1)$th iteration on the synaptic weights update at $r$th iteration
$\phi(.)$ hyperbolic tangent activation function

**Author Contributions**

R.S. (Ph.D., Independent Researcher) developed the model, conducted and analyzed simulations, and wrote the manuscript. R.A.S.L (Associate Professor) revised the manuscript.

**References**