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

Reza Salehi^{1†}, 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

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2021 Korean Society of Environmental Engineers

Received August 12, 2020 Accepted November 05, 2020

[†] Corresponding author

Email: reza.salehi@polymtl.ca

Tel: +1-438-889-6591

ORCID: 0000-0003-1416-173X

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*, *Thiothrix*, and *Thiovulum* have shown great potential to oxidize H_2S under OX condition, and the genera such as *Beggiatoa*, *Thiomargarita*, *Thioploca*, *Thioalkalivibrio*, and *Thiobacillus* are capable of oxidizing H_2S under either OX or AX condition [18].

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

Initially, H_2S 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].

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 $\text{NO}_3^-/\text{HS}^-$ ratio. Note that reactions 6 and 7 are based on complete denitrification (reduction of NO_3^- to N_2 without nitrite (NO_2^-) accumulation).

Based on the stoichiometry of reactions 6 and 7, if the $\text{NO}_3^-/\text{HS}^-$ molar ratio is 0.33, HS^- can be oxidized mainly to S^0 , while HS^- oxidation to SO_4^{2-} requires an $\text{NO}_3^-/\text{HS}^-$ molar ratio of 1.33. Similar $\text{NO}_3^-/\text{HS}^-$ ratio were reported by some researchers who achieved the complete oxidation of HS^- to SO_4^{2-} with an $\text{NO}_3^-/\text{HS}^-$ molar ratio between 1.2 and 1.6 whereas HS^- oxidation ended up with S^0 at an $\text{NO}_3^-/\text{HS}^-$ 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_2S 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* AZ11 could grow at a pH as low as 0.2 for H_2S 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 underestimation 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 back-

ward 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).

$$W^{(r+1)} = W^{(r)} - \alpha \frac{\partial E}{\partial W} + \beta \cdot (W^{(r)} - W^{(r-1)}) ,$$

$$W = \bar{W}_{k \times l}, \bar{\bar{W}}_{l \times j} \quad (8)$$

where,

$$\frac{\partial E}{\partial \bar{W}_{l \times j}} = -[f(X_{i \times k}, \bar{W}_{k \times l})]^T \cdot ((Y_{i \times j} - \hat{Y}_{i \times j}) \odot g'(f(X_{i \times k}, \bar{W}_{k \times l}), \bar{\bar{W}}_{l \times j})) \quad (9)$$

$$\frac{\partial E}{\partial \bar{W}_{k \times l}} = -[X_{i \times k}]^T \cdot [((Y_{i \times j} - \hat{Y}_{i \times j}) \odot g'(f(X_{i \times k}, \bar{W}_{k \times l}), \bar{\bar{W}}_{l \times j})) \cdot [\bar{\bar{W}}_{l \times j}]^T) \odot f'(X_{i \times k}, \bar{W}_{k \times l})] \quad (10)$$

Fig. 1. Schematic representation of a typical three-layered ANN; Each circle in the layers represents a neuron; Symbols “n”, “m” and “p” refer to the number of neurons in the input layer, hidden layer and output layer, respectively; (x_1-x_n) and $(\hat{Y}_1-\hat{Y}_p)$ are the network inputs and outputs, respectively; $\bar{w}$ and $\bar{\bar{w}}$ are the synaptic weights between the input and hidden layers, and between the hidden and output layers, respectively; symbols “f” and “g” represent the activation functions for the hidden layer neurons and the output layer neurons, respectively.

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]

# obs	Input variable			Output variable		# obs	Input variable			Output variable	
	x ₁	x ₂	x ₃	y (ppm)	y (%)		x ₁	x ₂	x ₃	y (ppm)	y (%)
1	8550	0	0	179.62	0.00	33	18810	0	40	30.53	26.33
2	8550	0	20	163.90	8.75	34	18810	0	60	21.78	47.44
3	8550	0	40	119.67	33.38	35	18810	0	80	2.50	93.97
4	8550	0	60	98.24	45.31	36	18810	2	0	25.03	0.00
5	8550	0	80	59.00	67.15	37	18810	2	20	20.04	19.94
6	8550	2	0	163.90	0.00	38	18810	2	40	13.71	45.23
7	8550	2	20	138.22	15.67	39	18810	2	60	10.05	59.85
8	8550	2	40	96.26	41.27	40	18810	2	80	1.94	92.25
9	8550	2	60	39.11	76.14	41	18810	4	0	10.97	0.00
10	8550	2	80	23.44	85.70	42	18810	4	20	8.90	18.87
11	8550	4	0	142.48	0.00	43	18810	4	40	6.78	38.20
12	8550	4	20	104.40	26.73	44	18810	4	60	4.59	58.16
13	8550	4	40	49.00	65.61	45	18810	4	80	1.71	84.41
14	8550	4	60	24.15	83.05	46	23940	0	0	144.55	0.00
15	8550	4	80	4.06	97.15	47	23940	0	20	130.68	9.60
16	13680	0	0	87.85	0.00	48	23940	0	40	127.84	11.56
17	13680	0	20	82.69	5.87	49	23940	0	60	111.74	22.70
18	13680	0	40	75.60	13.94	50	23940	0	80	100.00	30.82
19	13680	0	60	53.01	39.66	51	23940	2	0	149.13	0.00
20	13680	0	80	41.49	52.77	52	23940	2	20	122.30	17.99
21	13680	2	0	89.99	0.00	53	23940	2	40	110.00	26.24
22	13680	2	20	73.31	18.54	54	23940	2	60	86.00	42.33
23	13680	2	40	56.96	36.70	55	23940	2	80	68.70	53.93
24	13680	2	60	37.89	57.90	56	23940	4	0	118.17	0.00
25	13680	2	80	15.49	82.79	57	23940	4	20	105.82	10.45
26	13680	4	0	89.51	0.00	58	23940	4	40	85.76	27.43
27	13680	4	20	60.98	31.87	59	23940	4	60	54.27	54.07
28	13680	4	40	45.92	48.70	60	23940	4	80	29.11	75.37
29	13680	4	60	22.17	75.23						
30	13680	4	80	5.06	94.35						
31	18810	0	0	41.44	0.00						
32	18810	0	20	39.56	4.54						

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”.

$$\hat{Y}_{i \times j} = g(f(X_{i \times k}, \bar{W}_{k \times l}) \cdot \bar{W}_{l \times j}) \quad (11)$$

$$E = \frac{1}{2} (Y_{i \times j} - \hat{Y}_{i \times j})^2 = \frac{1}{2} (Y_{i \times j} - \hat{Y}_{i \times j}) \odot (Y_{i \times j} - \hat{Y}_{i \times j}) \quad (12)$$

The notations in equations (8-12) are as follows.

- $\bar{W}_{k \times l}$ and $\bar{W}_{l \times j}$ are the matrices assigned to the synaptic weights entering and leaving the hidden layer neurons, respectively
- E: Network error function
- α : Convergence speed of the algorithm, more commonly known as learning rate
- β : Influence of (r-1)th iteration on the synaptic weights update at rth iteration
- The symbols “g” and “f” denote the activation function of the neurons within the output layer and the hidden layer,

respectively

- $X_{i \times k}$: Network input matrix
- Symbol “ $\odot$ ” represents an element wise multiplication of two matrices
- Superscript “T” in $[.]^T$ refers to the transpose of matrix $[.]$
- $Y_{i \times j}$ and $\hat{Y}_{i \times j}$ 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 *Thiobacillus* isolated from the sludge of the municipal wastewater treatment plant in Srandakan (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 (x_1), residence time (x_2) and axial position in the BF bed (distance from the BF inlet) (x_3). 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).

$$\phi(S_{in}) = \frac{2}{1 + \exp(-2S_{in})} - 1 \quad 0 \leq \phi(S_{in}) \leq 1 \quad (13)$$

where, " S_{in} " denotes the sum of the weighted inputs entering a neuron, and $\phi(S_{in})$ 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].

$$x_{i,N} = \frac{x_i - x_{i,min}}{x_{i,max} - x_{i,min}} \quad (14)$$

where, " $x_{i,N}$ " represents the normalized value of " x_i ", which is the actual value of input variable " i ". " $x_{i,min}$ " and " $x_{i,max}$ " 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 (R^2)-that is a goodness-of-fit measure for the model-defined by equations (15) and (16), respectively. The closer R^2 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 R^2 value is equal to 1.0 and RMSE value is equal to 0.

$$RMSE = \sqrt{\frac{1}{s \times p} \sum_{i=1}^s \sum_{j=1}^p (a(i,j) - \hat{a}(i,j))^2} \quad (15)$$

$$R^2 = 1 - \frac{\sum_{i=1}^s \sum_{j=1}^p (a(i,j) - \hat{a}(i,j))^2}{\sum_{i=1}^s \sum_{j=1}^p (a(i,j) - \bar{a})^2} \quad (16)$$

where, $a(i,j)$ denotes the elements of matrix $Y_{s \times p}$ assigned to the desired output values; $\hat{a}(i,j)$ represents the elements of matrix $\hat{Y}_{s \times p}$ 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.

Fig. 2. Schematic illustration of K-fold cross-validation (RMSE: root mean squared error; each colored rectangular box denotes N_{Tr}/K input-output pairs).

- 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 H_2S removal efficiency. The authors developed a mathematical model as a predictive tool for the assess-

ment of changes in H_2S 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 H_2S from the gas stream into the biofilm and biological oxidation of H_2S 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 H_2S 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 α and β were adjustable. The best values for α and β 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. 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. Architecture of the ANN model used in this study; " $x_{i,N}$ " and " y_N " are the normalized values of " x_i " and " y "; See Table 1 for " x_i " and " y " notations and values, and Table S1 for " $x_{i,N}$ " and " y_N " values; $\bar{w}$ and $\bar{w}$ 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. 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].

Fig. 6. Correlation between measured values of H_2S removal efficiency in the desulfurizing BF and the corresponding predicted values using the proposed ANN model during (a) training phase and (b) testing phase; solid line indicates the best-fit line.

Fig. 7. Correlation between measured values of H_2S removal efficiency in the desulfurizing BF and the corresponding predicted values over the whole of training and testing datasets (60 observations); filled triangles denote the data points obtained using the proposed ANN model in this study; hollow circles indicate data points taken from the mathematical model of Lestari et al. [12]; solid and dashed lines are the best fit lines obtained based on the results of the proposed ANN model in this study and the mathematical model developed by Lestari et al. [12], respectively.

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_2S biodegradation kinetics and mechanism.

Nomenclature

$a(i, j)$	elements of matrix Y assigned to the desired output values
$\hat{a}(i, j)$	elements of matrix $\hat{Y}$ 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
HLNs	hidden layer neurons
ICE	internal combustion engine
MEA	mono-ethanolamine
N_{Tr}	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
$\bar{w}$	synaptic weights between the input layer and hidden layer of the network
$\bar{W}_{k \times l}$	" $\bar{w}$ " in matrix form, where subscripts "k" and "l" stand for the number of input variables and the number neurons within the hidden layer, respectively
$\bar{\bar{w}}$	synaptic weights between the hidden layer and output layer of the network
$\bar{\bar{W}}_{l \times j}$	" $\bar{\bar{w}}$ " 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 ($\text{g m}^{-3} \text{ h}^{-1}$)
x_2	residence time (h)
x_3	axial position in the BF bed (distance from the BF inlet) (cm)
$x_{i,\max}$	maximum value of " x_i "
$x_{i,\min}$	minimum value of " x_i "
$x_{i,N}$	normalized value of " x_i "
$X_{i \times k}$	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_2S removal efficiency (%)
y_N	normalized value of " y "
$Y_{i \times j}$	"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_2S removal efficiency (%)
$\hat{Y}_{i \times j}$	"ŷ" in matrix form, where subscripts "i" and "j" are the data set size and the number of output variables, respectively

Symbols

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

Greek letters

α	learning rate
β	influence of $(r-1)^{\text{th}}$ iteration on the synaptic weights update at r^{th} iteration
$\phi(\cdot)$	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

- Jain M. Anaerobic Membrane Bioreactor as Highly Efficient and Reliable Technology for Wastewater Treatment-A Review. *Adv. Chem. Eng. Sci.* 2018;8:82-100.
- Yee TL, Rathnayake T, Visvanathan C. Performance Evaluation of a Thermophilic Anaerobic Membrane Bioreactor for Palm Oil Wastewater Treatment. *Membr.* 2019;9:55.
- Zhang P. Biogas Recovery from Anaerobic Digestion of Selected Industrial Wastes. In: Nageswara-Rao M, Soneji JR, eds. *Advances in Biofuels and Bioenergy*. London: IntechOpen; 2018. p. 251-271.
- Yentekakis IV, Goula G. Biogas Management: Advanced Utilization for Production of Renewable Energy and Added-value Chemicals. *Front. Environ. Sci.* 2017;5(7):1-18.
- Salehi R, Chaiprapat S. Single/triple stage biotrickling filter treating a H_2S -rich biogas stream: statistical analysis of the effect of empty bed retention time and liquid recirculation velocity. *J. Air Waste Manag. Assoc.* 2019;69:1429-1437.
- Santos-Clotas E, Cabrera-Codony A, Castillo A, Martín MJ, Poch M, Monclús H. Environmental Decision Support System for Biogas Upgrading to Feasible Fuel. *Energies* 2019;12:1546.
- Do K-U, Nghiem T-D, Kim SD, et al. Development of an iron-based adsorption system to purify biogas for small electricity generation station in Vietnam: A case study. In: Chan HY, Sopian K. eds. *Renewable Energy in Developing Countries: Local Development and Techno-Economic Aspects*. Cham: Springer; 2018. p. 155-184.
- Kulkarni MB, Ghanegaonkar PM. Hydrogen sulfide removal from biogas using chemical absorption technique in packed column reactors. *Glob. J. Environ. Sci. Manag.* 2019;5:155-166.
- Xiao C, Ma Y, Ji D, Zang L. Review of desulfurization process for biogas purification. In: IOP Conference Series: Earth Environ. Sci.; 22-25 December 2017; Singapore. 100 012177.
- Fischer ME. Biogas Purification: H_2S Removal using Biofiltration [dissertation]. Waterloo: University of Waterloo; 2010.
- Jaber MB, Couvert A, Amrane A, Le Cloirec P, Dumont E. Hydrogen sulfide removal from a biogas mimic by biofiltration under anoxic conditions. *J. Environ. Chem. Eng.* 2017;5: 5617-5623.
- Lestari RAS, Sediawan WB, Syamsiah S, Sarto S, Teixeira JA. Hydrogen sulfide removal from biogas using a salak fruit seeds packed bed reactor with sulfur oxidizing bacteria as biofilm. *J. Environ. Chem. Eng.* 2016;4:2370-2377.
- López Gómez ME. Biofiltration of volatile compound mixtures from pulp and paper industries [dissertation]. A coruña : Universidade da Coruña; 2015.
- Montebello AM. Aerobic biotrickling filtration for biogas desulfurization [dissertation]. Cerdanyola del Vallès: Universitat Autònoma de Barcelona; 2013.
- Pokorna D, Zabranska J. Sulfur-oxidizing bacteria in environmental technology. *Biotechnol. Adv.* 2015;33:1246-1259.
- Omri I, Bouallagui H, Aouidi F, Godon J-J, Hamdi M. H_2S gas biological removal efficiency and bacterial community diversity in biofilter treating wastewater odor. *Bioresour. Technol.* 2011;102:10202-10209.
- Tchobanoglous G, Burton FL, Stensel HD. 2003. Wastewater engineering treatment and reuses. 4th ed. New York: McGraw-Hill; 2003.p.607-623.
- Lin S, Mackey HR, Hao T, Guo G, Van Loosdrecht MCM, Chen G. Biological sulfur oxidation in wastewater treatment: A review of emerging opportunities. *Water Res.* 2018;143:399-415.
- Fortuny M, Guisasola A, Casas C, Gamisans X, Lafuente J, Gabriel D. Oxidation of biologically produced elemental sulfur under neutrophilic conditions. *J. Chem. Technol. Biotechnol.* 2010;85: 378-386.
- Klok JB, Van den Bosch PLF, Buisman CJ, Stams AJ, Keesman KJ, Janssen AJ. Pathways of sulfide oxidation by haloalkaliphilic bacteria in limited-oxygen gas lift bioreactors. *Environ. Sci. Technol.* 2012;46:7581-7586.
- Van den Bosch PLF, Van Beusekom OC, Buisman CJ, Janssen AJ. Sulfide oxidation at halo-alkaline conditions in a fed-batch bioreactor. *Biotechnol. Bioeng.* 2007;97:1053-1063.
- Janssen A, Sleyster R, van der Kaa C, Jochemsen A, Bontsema J, Lettinga G. Biological sulphide oxidation in a fed-batch reactor.

- Biotechnol. Bioeng.* 1995;47:327-333.
23. Dolejs P, Paclík L, Maca J, Pokorna D, Zabranska J, Bartacek J. Effect of S/N ratio on sulfide removal by autotrophic denitrification. *Appl. Microbiol. Biotechnol.* 2015;99:2383-2392.
24. An S, Tang K, Nemati M. Simultaneous biodesulphurization and denitrification using an oil reservoir microbial culture: effects of sulphide loading rate and sulphide to nitrate loading ratio. *Water Res.* 2010;44:1531-1541.
25. Mahmood Q, Zheng P, Cai J, et al. Comparison of anoxic sulfide biooxidation using nitrate/nitrite as electron acceptor. *Environ. Prog. Sustain. Energy* 2007;26:169-177.
26. Cardoso RB, Sierra-Alvarez R, Rowlette P, Flores ER, Gomez J, Field JA. Sulfide oxidation under chemolithoautotrophic denitrifying conditions. *Biotechnol. Bioeng.* 2006;95:1148-1157.
27. Jaber MB, Couvert A, Amrane A, Rouxel F, Le Cloirec P, Dumont E. Biofiltration of high concentration of H₂S in waste air under extreme acidic conditions. *New Biotechnol.* 2016;33:136-143.
28. Sun S, Jia T, Chen K, Peng Y, Zhang L. Simultaneous removal of hydrogen sulfide and volatile organic sulfur compounds in off-gas mixture from a wastewater treatment plant using a two-stage bio-trickling filter system. *Front. Environ. Sci. Eng.* 2019;13:Article 60.
29. Namgung H-K, Song J. The Effect of Oxygen Supply on the Dual Growth Kinetics of *Acidithiobacillus thiooxidans* under Acidic Conditions for Biogas Desulfurization. *Int. J. Environ. Res. Public Health* 2015;12:1368-1386.
30. Lee EY, Lee NY, Cho K-S, Ryu HW. Removal of hydrogen sulfide by sulfate-resistant *Acidithiobacillus thiooxidans* AZ11. *J. Biosci. Bioeng.* 2006;101:309-314.
31. Jaber MB, Couvert A, Amrane A, Rouxel F, Le Cloirec P, Dumont E. Biofiltration of H₂S in air-Experimental comparisons of original packing materials and modeling. *Biochem. Eng. J.* 2016;112:153-160.
32. Liu D, Feilberg A, Jørgen Hansen M, Pedersen CL, Nielsen AM. Modeling removal of volatile sulfur compounds in a full-scale biological air filter. *J. Chem. Technol. Biotechnol.* 2016;91:1119-1127.
33. Jiang X, Tay JH. Removal mechanisms of H₂S using exhausted carbon in biofiltration. *J. Hazard. Mater.* 2011;185:1543-1549.
34. Kraakman NJR, Rocha-Rios J, van Loosdrecht MCM. Review of mass transfer aspects for biological gas treatment. *Appl. Microbiol. Biotechnol.* 2011;91:873-886.
35. Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, Arshad H. State-of-the-art in artificial neural network applications: A survey. *Heliyon* 2018;4:e00938.
36. Rene ER, Estefanía López M, Kim JH, Park HS. Back Propagation Neural Network Model for Predicting the Performance of Immobilized Cell Biofilters Handling Gas-Phase Hydrogen Sulphide and Ammonia. *BioMed Res. Int.* 2013;463401:9.
37. Rene ER, Kim JH, Park HS. An intelligent neural network model for evaluating performance of immobilized cell biofilter treating hydrogen sulphide vapors. *Int. J. Environ. Sci. Tech.* 2008;5: 287-296.
38. Elias A, Ibarra-Berastegi G, Arias R, Barona A. Neural networks as a tool for control and management of a biological reactor for treating hydrogen sulphide. *Bioprocess. Biosyst. Eng.* 2006;29:129-136.
39. Allam Z. Achieving Neuroplasticity in Artificial Neural Networks through Smart Cities. *Smart Cities* 2019;2:118-134.
40. Alemu HZ, Wu W, Zhao J. Feedforward Neural Networks with a Hidden Layer Regularization Method. *Symmetry* 2018;10:525.
41. Zaghoul MS, Hamza RA, Iorhemen OT, Tay JH. Performance prediction of an aerobic granular SBR using modular multilayer artificial neural networks. *Sci. Total Environ.* 2018;645:449-459.
42. Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL, eds. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume I: Foundations*. Massachusetts: The MIT Press; 1986.p.318-362.
43. Nawi NM, Hamzah F, Hamid NA, Rehman MZ, Aamir M, Ramli AA. An Optimized Back Propagation Learning Algorithm with Adaptive Learning Rate. *Int. J. Adv. Sci. Eng. Inf. Technol.* 2017;7:1693-1700.
44. Salameh WA, Otair MA. Efficient Training of Neural Networks Using Optical Backpropagation with Momentum Factor. *Int. J. Comput. Appl.* 2008;30:167-172.
45. Yu L, Wang S, Lai KK. Forecasting Foreign Exchange Rates Using an Adaptive Back-Propagation Algorithm with Optimal Learning Rates and Momentum Factors. In: Hillier FS, ed. *Foreign exchange rate forecasting with artificial neural networks*. New York: Springer & Business Media; 2007.p.65-84.
46. Vogl TP, Mangis JK, Rigler AK, Zink WT, Alkon DL. Accelerating the Convergence of the Back-Propagation Method. *Biol. Cybern.* 1988;59:257-263.
47. Pisa I, Vilanova R, Santín I, Vicario J, Morell A. Artificial Neural Networks Application to Support Plant Operation in the Wastewater Industry. In: 10th Doctoral Conference on Computing, Electrical and Industrial Systems (DoCEIS); 8-10 May 2019; Costa de Caparica. p. 257-265.
48. Elkiran G, Nourani V, Abba SI, Abdullahi J. Artificial intelligence-based approaches for multi-station modelling of dissolve oxygen in river. *Glob. J. Environ. Sci. Manag.* 2018;4: 439-450.
49. Fedotenkova M. Extraction of multivariate components in brain signals obtained during general anesthesia [dissertation]. Nancy Grand-Est: Univ. of Lorraine; 2016.