Artificial Neural Network Analysis of Sulfide Production in A Moroccan Sewerage Network

Asmae El Brahmi, Souad Abderafi, Rachid Ellaia

Mohammadia School of Engineers, Mohammed V University in Rabat, Av. Ibn Sina, BP. 765 Rabat, Morocco
Correspondence: E-mail: asmae.elbrahmi@gmail.com

Abstracts

Sulfide in urban wastewater leads to the formation of hydrogen sulfide and its release in the air. This molecule is an odorous compound, representing an annoyance and health threat for workers and the nearby population. In order to prevent hydrogen sulfide emission, it is necessary to evaluate sulfide concentration in sewage water and identify environmental key parameters that enhance sulfide production. In this study, Artificial Neural Network (ANN) method was used to analyze the presence of this substance in a Moroccan sewerage network. Experimental data of wastewater composition of Tangier sites (north of Morocco) were used for the training, testing, and validating the ANN model. The results showed satisfactory capability of ANN to predict sulfide concentration in aqueous phase, reaching value of 89%. Dissolved oxygen and temperature have the most significant impact on sulfide production. The obtained model can be the first step towards monitoring sulfide for building up in sewers and consequently applying it into an appropriate treatment.

© 2021 Tim Pengembang Jurnal UPI
1. INTRODUCTION

Studies on sulfide formation in sewer systems have been carried out since a long time. Although widely studied, the concerns associated with sulfide are still up to date (Zhang et al., 2008). The passage of the effluent under anaerobic conditions (absence of oxygen) degrades its quality and increases the middle of sepsis. In such a medium, sulphates present in urban wastewater can be reduced to sulfide by sulphate-reducing bacteria (Delgado et al., 1999). Thereby Several parameters contribute to sulfide production such as pH, temperature, nutrients, hydraulic retention time, presence of biofilm on the pipe surface and oxidation-reduction potential (Firer et al., 2008). This sulfide under certain conditions gives rise to hydrogen sulfide (H₂S) which is causing a critical problem in the sewage system. Hydrogen sulfide is a chemical substance that is present and/or formed in sewers having different characteristics: colorless, flammable, with a “rotten egg” odor. So, it has adverse effects on health, structures and the environment (Aguilar et al., 2004). At high concentration in the air, sulfide has significant toxic effects on olfactory nerves, lungs, eyes, brain and respiration (Lambert et al., 2006; Hughes et al., 2009). It is also responsible for the degradation of equipment and pipelines of sanitation networks (Parande et al., 2006) (Boumehraz et al., 2018). So, it is necessary to follow the sulfide concentration and identify the parameters that are the origin of its production. An experimental measurement of this concentration is time consuming, costly, and quite difficult (Hughes et al., 2009). Therefore, it would be interesting to use a reliable model for its evaluation. Among those available in the literature, we cite particularly, the model developed by Nielsen et al., (1998) that describes the sulfide production rate from biofilm surfaces in Danish pressure mains. Pomeroy model was developed for sulfide generation in filled pipe sewers in Pasadena (Pomeroy, 1959). Boon and Lister have proposed equations for predicting sulfide production in sewer pipe that transport raw wastewater with high organic matter. Recently, a model is proposed to predict sulfide generation under sewer-like conditions of Malaysian sewage (Penang). The study indicates that sulfide generation is strongly related to site-specific sewage characteristics. All of these equations relate the amount of sulfide build-up with chemical hydraulic and environmental parameters. These parameters include waste water composition, sulfate concentration, temperature and pipe dimensions. These models cannot be applied to different sites, because of their particularities (Vourch et al., 2008; Koukia et al., 2009).

Recently, we have successfully evaluated the production of hydrogen sulfide based on the physicochemical properties of wastewater from a site located in the central-western part of Casablanca the largest city in Morocco, using an artificial neural network (El Brahmi & Abderafi, 2020). In this work, the same methodology was followed to evaluate and test the effect of some parameters on the production of sulfide in another site of Moroccan sewerage network.

2. METHODS

In this section, we gave a physicochemical characterization of the studied sewerage network and describe the methodology used to develop the model that predicts the concentration of sulfide according to the main wastewater parameters.

2.1 Wastewater data characterization

The experimental data were provided by the national society of management and...
treatment of the polluted waters. These data characterized the gravity sewer lines located in different sites Tangier city Figure 1. The area is characterized by a diverse topography with the presence of hills to the north, and a plain in the western side, the attitude varies between 2 and 13 m. The area is also characterized by a diversified land mosaic the plot varies from 100 m to 50 ha plots. Treatment plant located behind the port, at the main discharge of the city, this emplacement presents two major constraints: the proximity to an unsteady cliff, and contiguity to the sea. The station is built on a maritime platform of 10850 m² on the sea and located west of the entrenchment of Tangier port’s jetty. Figure 1 presents the global sewage network in Tangier city where the green lines present old infrastructures, while the red one are new and the different rejection sites are presented by red stars. These stars precise the 11 rejection sites, from where different samples of wastewater were collected. After different analyzes performed in laboratory scale, by network managers, wastewater characterization data was obtained Table 1. This characterization gives all the necessary information about the key parameters, which were used for sewage characterization, included concentrations of sulfide (S²⁻), temperature (T), pH, Conductivity (Cd), Dissolved Oxygen (DO), Chemical Oxygen Demand (COD) and Total Suspended Solids (TSS).

The effluent temperature is between 19.4 and 32.7 °C with an average of 24.9 °C. The recorded temperature values are slightly above 30°C, considered as the limit value for the discharge of wastewater into the natural environment. Similarly, these values are much lower than 35 °C, considered as an indicative limit value for water intended for irrigation. The pH of all wastewater samples varies between 7 and 9 with an average of 7.5. The recorded pH values can be considered to meet the direct discharge standards in the environment. The maximum value of the conductivity recorded is 13290 μS/cm and the minimum value is 1131 μS/cm with an average value of 3182.78 μS/cm. This electrical conductivity value is greater than 2700 μS / cm considered as the limit value of direct discharge of wastewater, but it's is only attributed to the great value recorded in one of the 11 rejection sites. The DO of the sewerage system shows that the minimum and maximum extremes of the dissolved oxygen content are 0.1 mg/L and 7.02 mg/L, respectively, with an average concentration of 1.21 mg /L. This value is greater than 0.5 which shows that these waters are not low in oxygen. The TSS represents all the mineral and organic particles contained in the wastewater. Its concentrations in the wastewater analyzed ranged from 69 to 2644 mg /L with an average value of 544 mg / L that is less than the reject limit value. COD allows the concentration of organic or inorganic matter, dissolved or suspended in water, to be assessed through the quantity of oxygen necessary for their total chemical oxidation. Effluents are characterized by COD concentrations ranging from 186 to 1920 mg O₂/L with an average of 863 mg O₂ /L. This value is lower than that of the direct rejection standard (1000 mg O₂/L). The amounts of sulfide recorded range from 0.128 to 8.2 mg / l, with an average value of 2.37 mg / l, higher than the value of 1 mg / L supposed to be the limit value for direct discharge.

In general, physicochemical parameters are in the average range, and relatively, do not exceed the general limit values of the discharges in the receiving environment, which does not represent a risk of environmental pollution for the latter. Apparent maximum or minimum values of some parameters may be at the root of sulfide production. To evaluate the importance of these parameters, the Artificial Neural Network methodology was followed.
Artificial neural networks method

Artificial Neural Networks (ANN) approach is proposed to analyze the production of sulfide in sanitation networks. This method was chosen because the major processes found in the chemical engineering are unfortunately non-linear processes. In that case ANNs can learn non-linearities in the systems. ANN could perform better than statistical models based on the regression analysis and is tolerant to the noise in data. It was successfully used to predict the performance of a wastewater treatment plant by different researchers (Hamed et al., 2004; Tümer & Edebali, 2015). It was also the best method in comparison with Adaptive Neuro Fuzzy Inference System method (Vasseghian et al., 2016). Recently, ANN has undergone numerous applications in chemical engineering (Schmitz et al., 2006; Karaci et al., 2016; Nourani et al., 2017; Nourani et al., 2018). In this section this
method and its different steps of applications were described.

2.2.1 Method description

ANN approach is a biological inspiration based on various characteristics of the brain functionality. A neural network consists of a very large number of small identical processing units called artificial neurons. The units are interconnected by unidirectional links that act like axons and dendrites of their biological counterparts. The multilayer structure of networks is relatively simple, it consists of an input layer an output layer and one or more hidden layers. Inputs to the network (predictors) are passed from the input layer of neurons, through the hidden layers of neurons, to the output layer where they become predictions. Neurons in the input layer do no more than disperse all predictors to each neuron in the hidden layer. The network operates by applying weights to values as they pass from one layer to the next and calculating outputs for each of the neurons in all other layers. Links exist, only between cells in a layer and the cells of the next layer (Zhang & Friedrich, 2003). In each unit of the hidden layer, variables are combined linearly. Neural network applies a non-linear transformation to each of these combinations that determine the transfer function of the network. The activation function used, for connections between neurons determine the transfer function of the network. Finally, the resulting values of the hidden units are combined in linear ways to get the predicted value.

Briefly, ANN technique is a nonlinear modeling approach used in different domains (Farobie & Hasanah, 2016). An ANN method provides a dynamic relationship between inputs and outputs and bypass underlying complexity inside the system where the inputs of the system are the independent variables and the outputs are the dependent variables. Therefore, it is important for the user to have a good understanding of the science behind the underlying system to provide the appropriate input and, consequently, to support the identified relationship.

2.2.2 Artificial neural network model

The output of a neuron is computed by using the network design that worked with a two-layer feed-forward network (one hidden layer). The mathematical model of the basic ANN structure is given in equation (1):

\[ y_i = f\left(\sum w_{ij}x_j + b_i\right) \]  

(1)

where \( x_j \) is the input signal transmitted to the neuron \( i \); \( w_{ij} \) is the weight coefficients between the external input and the neuron \( i \); \( b_i \) is bias; \( f \) is a transfer function of the neuron and \( y_i \) is output value.

The input signals to each neuron are weakened or strengthened through their multiplication to weight coefficients. The biases are activation thresholds that are added to the production of inputs and their particular weight coefficients. The net output of each neuron passes through a transfer function of the neuron. This latter is called activation function comes in different forms, such as linear, logarithmic sigmoid, hyperbolic tangent sigmoid and radial basis transfer functions (Khayet & Cojocaru, 2013). In the present study, we choose, a "tangent sigmoid" function for the hidden layer, because it is continuous and relatively easy to compute (as is its derivative). It maps the outputs away from extremes and it introduces nonlinear behavior to the network. This function used in the first layer is written in equation (2):

\[ \tan \text{sig}(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \]  

(2)

For the second layer the Purelin transfer function is used and defined in equation (3)

\[ f(x) = x \]  

(3)
2.2.3 Steps in designing network

To design the network, the calculation must be performed via numerical computational tools by following different steps. After selection and preparation of samples, the first step is to develop the neural network structure. The next step is to use an optimization technique to estimate the weights and biases in such a way that the output of the network is as close as possible to the target values, through much iteration. This optimization strategy is known as the training process and the neural network learns the relation between the inputs and the outputs of a dynamic system. There are various learning algorithms to train neural networks. In this work, the Levenberg–Marquardt method is used because is suitable algorithm for a small and moderate total number of the net weights and the most efficient calculation procedure adapted for learning (Godini et al., 2011). In the end, the network is used after validation and test. Some statistical methods can be used for the comparison between the net output and the training data to evaluate the performance of a neural network in the learning. These methods indicate how the network predictions are close to the target values and what adjustment should be applied to the learning weighting algorithm, at each iteration. To validate our results, we use the most commonly employed criterion for comparison, which are the correlation coefficient ($R^2$), the Mean Absolute Error (MAE) and the Mean Root Square Error (MRSE) defined respectively by equations (4), (5) and (6) respectively.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (V_{exp,i} - V_{cal,i})^2}{\sum_{i=1}^{n} V_{exp,i}^2}$$  \hspace{1cm} (4)

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |V_{exp,i} - V_{cal,i}|$$  \hspace{1cm} (5)

$$MRSE = \frac{1}{n} \sqrt{\sum_{i=1}^{n} (V_{exp,i} - V_{cal,i})^2}$$  \hspace{1cm} (6)

where $n$ is the number of data points in the entire data set; $V_{cal}$ and $V_{exp}$ are the calculated and experimental value, respectively.

3. RESULTS AND DISCUSSION

3.1 ANN Model Reliability

The network was trained by 109 samples of the 11 sites. The purpose is to estimate the concentration of sulfide $S^{2−}$ based on parameters that contribute to its production in sewers. For the network design, we worked with a two-layer feed-forward network (one hidden layer), having as inputs the six parameters suspected to be the most influencing parameters in the sulfide formation according to literature (Taleb et al., 2005), which are $T$ (K), pH, Cd ($\mu$s/cm), DO (mg/L), TSS (mg/L) and COD (mgO$_2$/L). We tried to minimize the number of nodes in the hidden layer so that we have the minimum complexity, finally, ten nodes were enough to get a satisfactory regression. The output of our network is the sulfide concentration $S^{2−}$ (mg/L). Figure 2 shows the optimized structure of the implemented neural network. This figure shows that the identified neural network has ten neurons in the hidden layer and one neuron in the output layer, with a tangent-sigmoid and linear transfer functions in the hidden and the output layers, respectively.

The regression results obtained are given in Table 2. The correlation coefficient $R^2$ for training (70% of data) is 0.94. As for testing set (15% of data) it equals 0.73. Finally for the validation set (15% Data) it was 0.91. The correlation coefficient for the model is equal to 0.89. These results indicate that there is a satisfactory correlation between the target
and the result, if we consider the experimental error. The concentration of sulfide ($S^2$) as function of $T$, pH, Cd, DO, TSS and COD was predicted using ANN with MAE and MRSE equal to 0.53 and 0.08, respectively, which can be regarded as satisfactory results in Figure 2.

The performance of the model was also obtained by comparing the predicted values of sulfide concentration to the experimental one. This comparison is realized by plotting the residual error between experimental and predicted concentration of sulfide in function of predicted values. Figure 3 represents this comparison and shows that the points are distributed randomly around the zero axes and the errors not exceed the interval between -3.4 and 3.4 mg/L. As shown in Figure 3, the values occupy clearly both sides of the normal, so the residual distribution confirms the performance of ANN model. Statistical analysis of the proposed ANN model shows that ANN has produced a very satisfactory model in term of precision and reliability (see Figure 3).

### 3.1 Variable importance in ANNs

There are several methods for quantifying variables importance in artificial neural networks. The Connection Weight Approach was chosen because it provides the best overall methodology for accurately quantifying variable importance according to the comparison made by Olden et al. (Olden & Jackson, 2002), it is calculated in equation (7):

$$RI_i = \frac{r_i}{\sum_i r_i} \times 100$$  \hspace{1cm} (7)

where $RI_i$ is the relative importance of the variable $i$ and $r_i$ is the relative contribution of each input neuron to the outgoing signal of each hidden neuron given by equation (8):

$$r_i = \frac{|c_i|}{\sum_i c_i}$$  \hspace{1cm} (8)

where $C_i$ is the contribution of each input neuron to the output via each hidden neuron, calculated as the product of the input hidden connection ($W_i$) and the hidden output connection ($W_0$) shown in equation (9) is written as follow:

$$c_i = W_i \times W_0$$  \hspace{1cm} (9)

To determine factors influencing sulfide formation, in this particular data, the obtained matrix containing inputs-hidden-output neuron connection weights is given in Table 3. This table enables us to scale the importance of $T$, pH, Cd, DO, COD and TSS, using equations (7-9).

Values calculated for quantifying these variables are represented in Figure 4. This figure shows that, it’s clear that the Dissolved Oxygen is the most important parameter in sulfide formation, followed by temperature. The third scaled parameter is the pH followed by the conductivity and the total suspended solid those that have the same influence on the production of sulfide. The chemical oxygen demand is in the last rank. In fact, the presence of dissolved oxygen leads the aerobic degradation reactions of the organic matter and more generally the biological balance of the water. In wastewater treatment systems, its complete disappearance is generally accompanied by the appearance of $H_2S$ in the air, resulting from the reduction of sulfur compounds present in the effluents. On the other hand, its presence inhibits the denitrifying activities of the specialized flora. So, monitoring these parameters is very important to prevent sulfide formation.
Asmae El Brahmi, et al. Artificial neural network analysis of sulfide production...

Figure 2. The optimized ANN structure

Table 2. Regression results of testing, training and validation sets

<table>
<thead>
<tr>
<th>Data</th>
<th>R²</th>
<th>MAE (%)</th>
<th>RMSE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>0.9435</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Testing</td>
<td>0.7536</td>
<td>0.53</td>
<td>0.08</td>
</tr>
<tr>
<td>Validation</td>
<td>0.9111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>0.8964</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Residual distribution, for predicted sulfide concentration
**Figure 4.** Bar plots of the relative importance (%) of each parameter, for predicting sulfide concentration with the ANN developed model

**Table 3.** Matrix containing inputs-hidden-output neuron connection weights

<table>
<thead>
<tr>
<th>Neuron</th>
<th>Inputs</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T</td>
<td>Ph</td>
</tr>
<tr>
<td>1</td>
<td>1.6669</td>
<td>-2.8640</td>
</tr>
<tr>
<td>2</td>
<td>2.5090</td>
<td>1.3623</td>
</tr>
<tr>
<td>3</td>
<td>1.8214</td>
<td>0.23754</td>
</tr>
<tr>
<td>4</td>
<td>-1.1246</td>
<td>-0.7384</td>
</tr>
<tr>
<td>5</td>
<td>0.5714</td>
<td>0.3018</td>
</tr>
<tr>
<td>6</td>
<td>-1.2584</td>
<td>0.0219</td>
</tr>
<tr>
<td>7</td>
<td>-1.0126</td>
<td>-2.6517</td>
</tr>
<tr>
<td>8</td>
<td>1.2274</td>
<td>0.6109</td>
</tr>
<tr>
<td>9</td>
<td>-1.0729</td>
<td>-0.0589</td>
</tr>
<tr>
<td>10</td>
<td>0.8909</td>
<td>1.2167</td>
</tr>
</tbody>
</table>

### 4. CONCLUSION

The aim of this work was to develop a model for predicting and analyzing the production of sulfide in sanitation networks of Moroccan Tangier city. The artificial neural network technique for modeling the presence of sulfide in these sites was used. The network was trained by collected data which characterize eleventh sites. For the network design, we worked with a two-layer feed-forward network, having as inputs the six parameters suspected to be the most significant parameters in the sulfide formation, which are: Temperature, pH, Conductivity, dissolved oxygen, total suspended solid and chemical oxygen demand. The optimal neural network configuration for the estimation of sulfide
concentration has one hidden layer with ten neurons and has been trained by Levenberg Marquardt algorithm. The performance of ANN architecture was evaluated using $R^2$, MAE and MSE those are equal to 99%, 0.53 and 0.08, respectively. Variable importance indicates that sulfide production is strongly related to dissolved oxygen concentration, temperature followed by pH. The COD effect on sulfide formation was in the last rank, after Cd and TSS. There is thus a satisfactory procedure which makes it possible to analyze the formation of sulfide and leads to better management of the sewerage network for their treatment.

5. ACKNOWLEDGEMENTS

The authors gratefully acknowledge national society of management and treatment of the polluted waters for their valuable helpful.

6. AUTHORS’ NOTE

The author(s) declare(s) that there is no conflict of interest regarding the publication of this article. Authors confirmed that the data and the paper are free of plagiarism.

7. REFERENCES


