



Prediction of Indoor Total Volatile Organic Compound in a University Hostel Using a Neural Network Model

F. C. Nkeshita*, A. A. Adekunle, A. Abegunrin

Department of Civil Engineering, Federal University of Agriculture, Abeokuta, Ogun State, NIGERIA.

Abstract

The study predicted the concentration of indoor total volatile organic compounds (TVOC) concentration in a randomly selected room at the Umar Kabir male Hostel located at the Federal University of Agriculture, Abeokuta, Nigeria. Readings were taken using an active sampler to measure particulate matter, PM (1.0, 2.5 and 10), TVOC, Relative Humidity (RH), Temperature and Formaldehyde. Two network types namely; feedforward back propagation and the cascaded forward back propagation were adopted randomly to predict TVOC as an output variable using data set generated from six different parameters mentioned earlier. The best performing neural network was the cascaded feedforward with a coefficient of determination of 0.98 which exhibited the lowest mean square error of 0.000124 with a network structure of 6-15-1-1. The results show the ability of Artificial Neural Networks to map inputs and outputs in complex non-linear situations such as the existence of volatile compounds in the atmosphere. It can be adopted for monitoring environmental systems by engineers and public health workers, stakeholders can use such models for initiating environmental related policies aimed at safeguarding human health.

Keywords: TVOC, back propagation, ANN modelling

1. INTRODUCTION

Indoor air pollution is fast becoming a cause of concern for environmental engineers, health workers and other stakeholders. This is partly due to the fact that the air people breathe in enclosures also influences their wellbeing. Under ambient conditions, there is a natural self-cleansing mechanism, such as the influence of rain, wind and plant life. These will remove impurities from the atmosphere thereby making the composition of the impurities in the outdoor air relatively balanced. This is not the case with indoor air where human activities may alter the air quality, and negative effects may be experienced when a room is not properly ventilated. This may result in discomfort and health issues for the occupants [1].

A number of indoor air quality researches have been carried out with some of the results emphasizing higher level of Indoor Air Quality (IAQ) pollution than Outdoor Air Quality pollution [2–4].

There are emissions that emanate from different sources within a building. These include building materials, furniture, hair sprays, air fresheners, certain objects and human activities such as cooking, fumigation, etc. These tend to alter the indoor air quality and may require more ventilation in order to negate its negative effects

[5]. Some of these emissions are termed ‘Volatile Organic Compounds’ (VOCs) which to a layman is perceived as smells or odors. VOCs are present both indoors and outdoors and include toxic substances such as toluene, chloroform, styrene, benzene, etc. VOCs are known to have certain negative health effects such as infant diarrhea and headaches in nursing mothers, sensory irritation, asthma, etc. [6]. Other health related ailments that can be attributed to TVOCs include; eye, nose and throat irritation, nausea, damage to liver, kidney and central nervous system and cancer. New constructed buildings greatly influence high level of VOC in an indoor environment as a result of plenty of new materials generating VOC particles within a short period of time [7]. Some of the TVOCs dissipate over time from a few days to six months when there is proper ventilation indoors, however, there are other gases that may linger for years before dissipation. There is a need to adopt measures aimed at assessing and modelling the occurrence of these volatile gases in indoor environments in order to document or initiate mitigation measures by policy makers and other stakeholders. This research focuses on the development of a predictive model aimed at predicting the occurrence of volatile organic compounds in an indoor environment. The occurrence of these gases in the environment are non-linear in terms of the chemical processes and environmental factors associated with it and conventional mathematical modelling may not be ade-

*Corresponding author (Tel: +234 (0)806 136 2244)

Email addresses: nkeshita@funaab.edu.ng (F. C. Nkeshita), adebolamay@gmail.com (A. A. Adekunle), ayodejiabegunrin@gmail.com (A. Abegunrin)

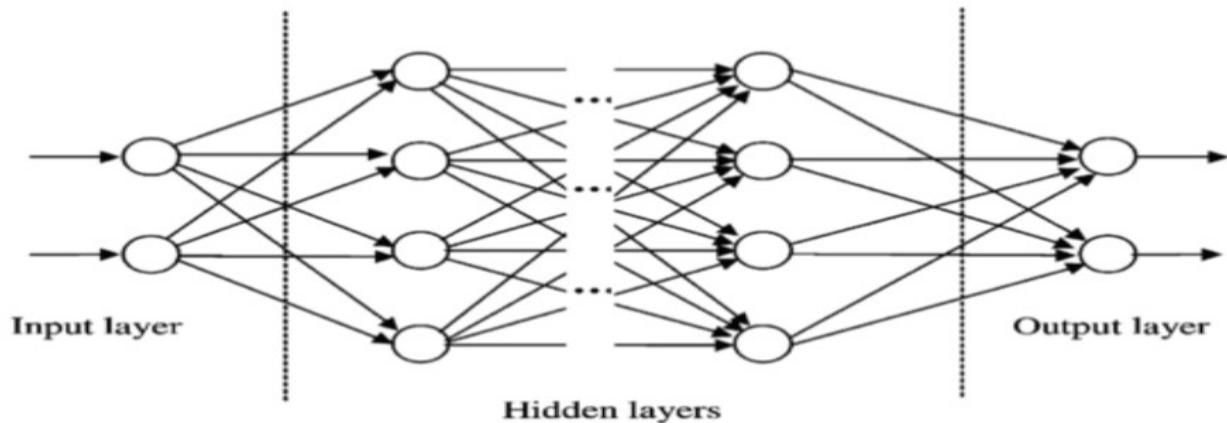


Figure 1: A schematic diagram displaying multilayer feed forward neural network [8].

quate enough to handle non-linear processes. Artificial neural network is being adopted in many fields of engineering and the sciences and has proved to be a reliable tool because of its relative ease of simulating, modeling and predicting the performance of a complex process [9]. This study is to develop an ANN-based model by using Feedforward Back propagation and cascaded forward algorithms that can effectively predict the occurrence of indoor air quality parameters such as total volatile organic compounds in an enclosure. An example is shown in Fig. 1.

2. METHODOLOGY

2.1. Study Area and Dataset Collection

The data was collected from a single room at the Umar Kabir male Hostel located at the Federal University of Agriculture, Abeokuta, South-western Nigeria on 7.232160°N , 3.432887°E . The set of data that was used in this study are particulate matters (1.0, 2.5 and 10), Total Volatile Organic Compound (TVOC), Relative Humidity, Daily Ambient Temperature and Formaldehyde. Sample data for Indoor Air Quality were gathered from an active sampler.

Data collection was carried out within an eighteen day duration in the evening when it is estimated that maximum activities would be occurring in the hostel rooms. A single room was adopted for taking indoor air quality readings on the upper floor of the hostel which was used to simulate the typical activity that takes place in a room occupied by students based on the restrictions made by the school authorities. The single room was also adopted to enable a simpler analysis during the modeling process. A similar research on an office room had been previously reported [2]. The measurements were taken at a standard breathable level of (1.0 - 1.5m) in real time. The sampler was allowed to adjust with the indoor environmental conditions for a minimum duration of fifteen minutes. The dataset is listed in Table 1.

2.2. The ANN Model

The Artificial Neural network (ANN) adopts a human pattern of solving problems by imitating the way the human biological neural network does. Human beings solve problems by relating to previous experiences and initiating corrections. The same approach is done by the ANN which looks for patterns in a given dataset, learns these patterns and create the ability to create a futuristic pattern related to the initial patterns originally learnt [10]. A neural network comprises a system of neurons which are connected to a network by a set of weights. The neuron takes a number of inputs, weighs, sums them up, adds a bias and uses the results as a basis for the argument for a single-valued function (transfer function) which results in the neuron's output [11]. For this study, the Feed forward back propagation algorithm was adopted. In which the input data was fed to the network and a corresponding output of the neural network was generated. The generated output was compared with the desired output and the error was computed. These input variables include two comfort variables (Relative humidity and Temperature) and four indoor pollutants (PM1.0, PM2.5, PM10 and Formaldehyde). One hidden layer was used and a set of neurons in the hidden layer were chosen in the model to optimize the ANN performance. The output layer of the prediction model adopted TVOC concentrations as the output variable.

2.3. Training, Testing and Validation

The neural network training tool used for the predictive modelling was implemented using MATLAB 2018 software. The given set of data was divided into three groups, 70% of the data was used in training the networks, 15% was used as the validation set while the remaining 15% were used in testing the networks. The mean square error (MSE) was adopted as the statistical criteria for measuring the performance of the network while the coefficient of determination was used to measure the relationship between the variables. The dataset was normalized before it was fed into

Table 1: Input data.

PM 1.0 ($\mu\text{g}/\text{m}^3$)	PM 2.5 ($\mu\text{g}/\text{m}^3$)	PM 10 ($\mu\text{g}/\text{m}^3$)	TVOC	Rel. Hum. (%)	Temperature ($^{\circ}\text{C}$)	HCHO (mg/m^3)
6	9	9	1.431	80	25.7	0.185
6	8	9	0.806	76	26.3	0.133
3	5	5	3.54	73	36.5	0.731
3	4	4	1.411	71	26.8	0.183
6	9	9	0.751	70	26.8	0.106
6	8	6	0.825	69	27.1	0.179
3	4	4	3.463	72	26.6	0.335
3	4	4	3.448	71	27.1	0.315
1	2	2	0.82	69	27.5	0.115
1	4	2	0.756	67	26.6	0.108
3	4	5	0.702	68	26.7	0.102
14	19	21	0.728	70	26.7	0.109
19	23	26	0.611	70	27.3	0.082
52	69	75	0.649	69	27.9	0.098
14	13	13	0.833	67	27.9	0.181
3	5	6	0.649	68	26.4	0.1
3	4	4	0.985	69	25.8	0.191
3	5	6	0.853	70	26.3	0.122

PM -Particulate matter, TVOC - Total Volatile Organic Compound, REL HUM – Relative Humidity.

Table 2: Input Data.

S/No	Network Structure	Transfer Function	MSE	NN type	R	R ²
Case 1	6-13-1-1	Tansig	0.00345	FFBP	0.97	0.94
Case 2	6-15-1-1	Tansig	0.0000855	FFBP	0.96	0.92
Case 3	6-10-1-1	Logsig	0.0022	FFBP	0.99	0.98
Case 4	6-13-1-1	Logsig	0.00137	FFBP	0.99	0.98
Case 5	6-15-1-1	Logsig	0.00282	FFBP	0.99	0.98
Case 6	6-20-1-1	Logsig	0.0128	FFBP	0.97	0.94
Case 7	6-10-1-1	Tansig	0.000206	Cascaded FFBP	0.99	0.98
Case 8	6-13-1-1	Tansig	0.00232	Cascaded FFBP	0.99	0.98
Case 9	6-15-1-1	Tansig	0.000124	Cascaded FFBP	0.99	0.98
Case 10	6-10-1-1	Logsig	0.000272	Cascaded FFBP	0.83	0.69
Case 11	6-13-1-1	Logsig	0.000156	Cascaded FFBP	0.99	0.98
Case 12	6-15-1-1	Logsig	0.00294	Cascaded FFBP	0.99	0.98
Case 13	6-20-1-1	Logsig	0.01129	Cascaded FFBP	0.99	0.98

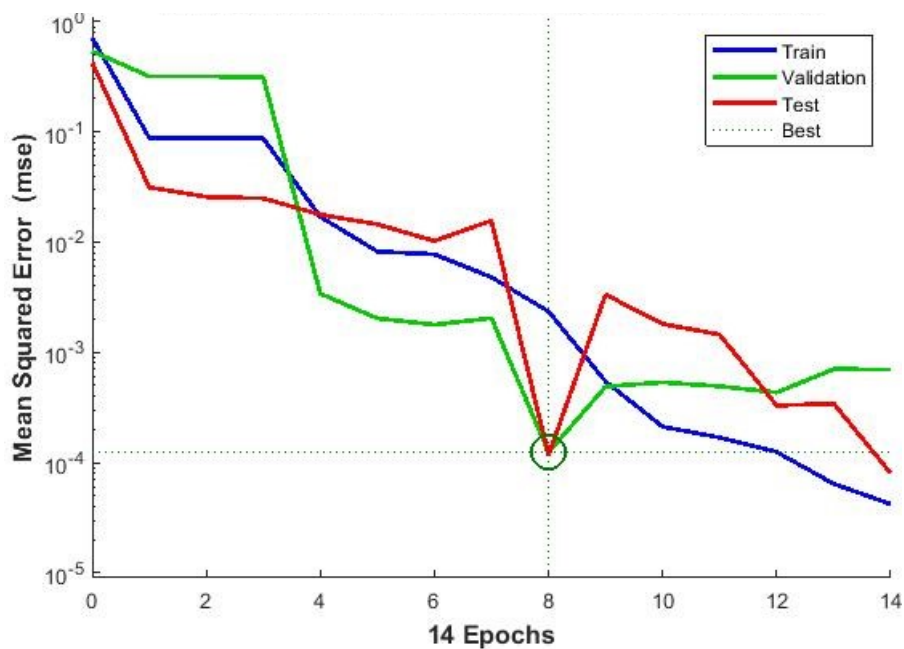


Figure 2: Best validation for network structure 6-15-1-1.

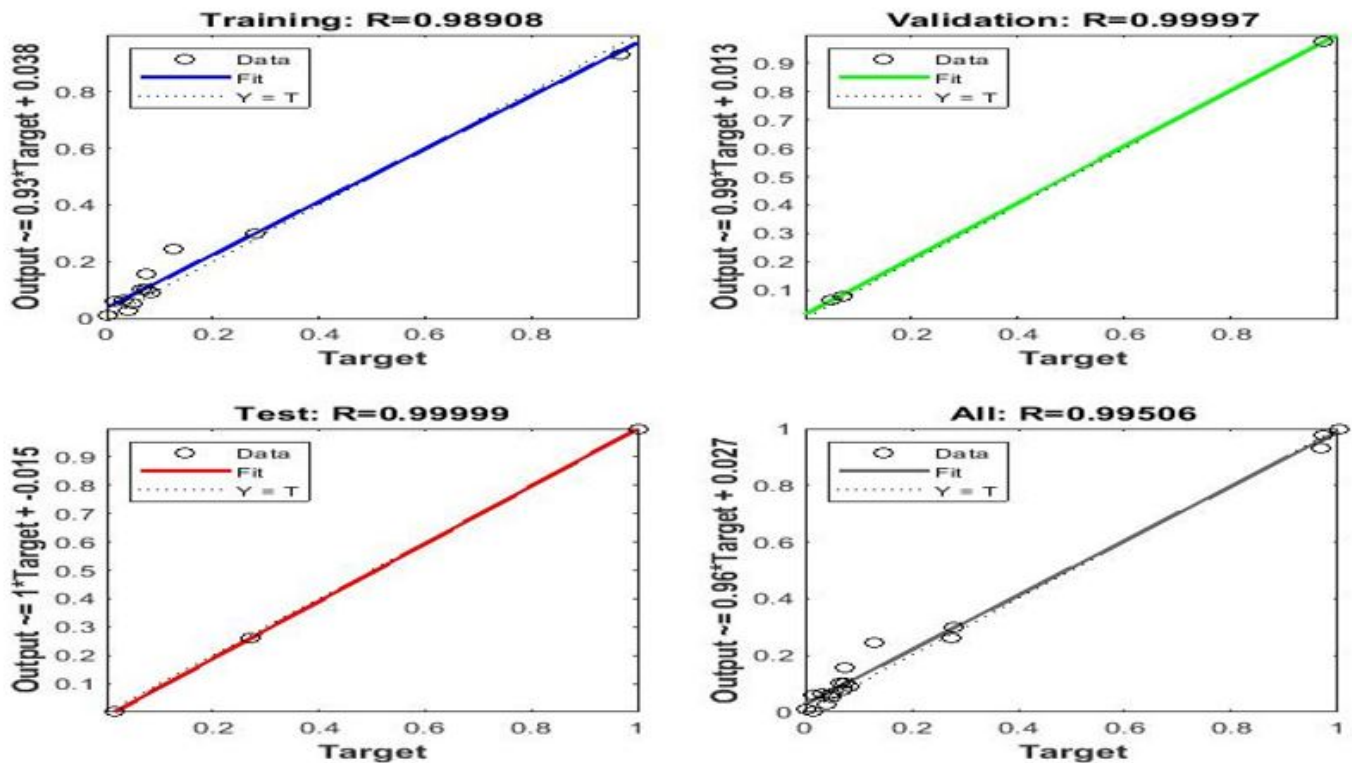


Figure 3: Regression analysis for network structure 6-15-1-1.

the neural network. The Levenberg-Marquardt method was used to train the input data and also to minimize sum-of-square error functions. Two neural network types were adopted; the feed forward back propagation and cascaded forward back propagation algorithms. Two transfer functions were also adopted; LOGSIG and TANSIG. The training function used was TRAINLM and this remain fixed for all the network structure. Different neural network structures were created between 10 and 20 hidden neurons and both the training functions and network types were randomly chosen to simulate the target variable.

Over-fitting is a problem for large networks with small data sets, such as the indoor air variable dataset used in this study. Early stopping method was employed to avoid over-fitting. A network training function that updates weight and bias values according to gradient descent momentum and an adaptive learning rate was employed with early stopping.

3. RESULTS AND DISCUSSION

This study was aimed at developing an ANN-based model by using Feedforward Back propagation and cascaded forward algorithms. Two network types namely; the feedforward back propagation and the cascaded forward back propagation were used randomly to predict Total Volatile Organic Compound as an output variable using data set generated from six different parameters namely; particulate matters (1.0, 2.5 and 10), Total Volatile Organic Compound (TVOC), Relative Humidity, Daily Ambient Temperature and

Formaldehyde as input variables. The model was able to detect relationships between input and output variables. This type of modeling is necessary for predicting the occurrence of gaseous compounds that may be detrimental to human health especially when living in an enclosure where these compounds from both natural and anthropogenic sources occur. The results obtained from the developed neural network models after training are displayed in Table 2. It can be seen from the tables that the log-sigmoid transfer functions performed better with the cascaded forward algorithms when compared to the tan-sigmoid transfer functions. The number of neurons influenced the performance of the networks with network structures of 15 neurons showing a better reduction in the root mean square errors and increasing the coefficient of determination. It can be seen from the table that cases 9 and 11 outperformed other neural network models but however, the better of the two was case 9, a cascaded feedforward type of network with a coefficient of determination of 0.98 which exhibited the lowest mean square error of 0.000124 with a network structure of 6-15-1-1 as shown in Fig. 2 and 3. Other network structures performed well but had lower R values for either training or testing or validation. It showed a better reduction in root mean square error when compared to the prediction carried out in an office room as reported in [2]. The result thus indicated that a very good relationship between the input and output variables was established and could be used to predict the performance of the neural network with the aim of predicting TVOC concen-

trations in the hostel room.

4. CONCLUSION

The predictive model of Total Volatile Organic Compound was successfully carried out using different neural network structures with the best being a NN with structure 6-15-1-1 when the cascaded feedforward network type was used with a training function TRAINLM and transfer function LOGSIG. This type of modeling has shown the ability of the ANN to map inputs and outputs in complex non-linear situations such as the existence of volatile compounds in the atmosphere.

It can be adopted for monitoring environmental systems by engineers and public health workers, stakeholders can use such models for initiating environmental related policies aimed at safeguarding human health. There is still a need to create room for further improvement of this type of model by incorporating other variables and increasing the number of times readings are taken to reflect different weather and times of the day.

References

- [1] World Health Organization WHO, "Air quality guidelines for Europe," WHO Regional publications, European Series No. 23, Copenhagen, 1987.
- [2] C. Jouvan, P. Pratama, Safrilah, and I. Mohammad, "The prediction of indoor air quality in office room using artificial neural network," in *AIP Conference Proceedings*, 1977, 020040.
- [3] S. C. Lee, "Comparison of Indoor and Outdoor Air Quality at Two Staff Quarters in Hongkong," *Journal of Environment International*, vol. 23, no. 97, pp. 791–797, 1997.
- [4] P. Blondeau, V. Iordache, O. Poupard, D. Genin, and F. Allard, "Relationship between outdoor and indoor air quality in eight French schools," *Indoor Air*, vol. 15, no. 1, pp. 2–12, 2005.
- [5] W. Liang, C. Wang, C. Yang, and X. Yang, "Volatile organic compounds in different interior construction stages of an apartment," *Building and Environment*, vol. 81, p. 380387, 2014.
- [6] A. Farrow, H. Taylor, K. Northstone, and J. Golding, "Symptoms of mothers and infants related to total volatile organic compounds in household products," *Arch Environ Health*, vol. 58, no. 10, pp. 633–641, 2003.
- [7] S. Wang, H. Ang, and M. Tade, "Volatile organic compounds in indoor environment and photocatalytic oxidation: State of the art," *Environment International*, vol. 33, no. 5, pp. 694–705, 2007.
- [8] S. A. Kalogirou, "Artificial intelligence for the modeling and control of combustion processes: a review," *Journal of Progress in Energy and Combustion Science*, vol. 29, no. 6, pp. 515–566, 2003.
- [9] S. Prakash, A. Manikandan, L. Govindarajan, and V. Vijayagopal, "Prediction of biosorption efficiency for the removal of copper (ii) using artificial neural networks," *Journal of Hazardous Materials*, vol. 152, pp. 1268–1275, 2008.
- [10] W. Daosud, P. Thitayasook, A. Arpornwichanop, P. Kit-tisupakorn, and M. Hussain, "Neural network inverse model-based controller for the control of a steel pickling process," *Computers & Chemical Engineering*, vol. 29, pp. 2110–2119, 2005.
- [11] D. Strik, A. Domnanovich, L. Zani, R. Braun, and P. Holubar, "Prediction of trace compounds in biogas from anaerobic digestion using the MATLAB neural network toolbox," *Environmental Modelling & Software*, vol. 20, no. 6, pp. 803–810, 2005.