FUOYE Journal of Engineering and Technology, Volume 8, Issue 2, June 2023 ISSN: 2579-0617 (Paper), 2579-0625 (Online)

Air Pollution Forecasting using Fuzzy Time Series Models for Kaduna Metropolis, Nigeria

[1]Adenike Folaponmile, [2]Samuel. F. Kolawole and [2]Samuel N. John
[1] Department of Computer Engineering, Kaduna Polytechnic, Kaduna, Nigeria
[2]Department of Electrical and Electronic Engineering, Nigerian Defence Academy, Kaduna, Nigeria
adenikefolaponmile@kadunapolytechnic.edu.ng|sfkolawole@nda.edu.ng|samueljohn@nda.edu.ng

Received: 05-DEC-2022; Reviewed: 15-MAR-2023; Accepted: 10-JUN-2023
http://doi.org/10.46792/fuoyejet.v8i2.968

ORIGINAL RESEARCH

Abstract- Fuzzy Time Series (FTS) is able to eliminate the problem of overfitting that is fundamental to Artificial Neural Network (ANN), hence this study used air pollution data acquired from three different sampling stations in Kaduna metropolis, Nigeria, to implement FTS using the Adaptive Neuro Fuzzy Inference System (ANFIS). The fuzzy inference system (FIS) was generated by the ANFIS model using grid partitioning and subtractive clustering optimization types with backpropagation and hybrid training algorithms. The models were implemented using MATLAB 2018b software, and a total of thirteen models were developed. The resulting models were used to forecast the daily mean for the next ten days for each sampling station and for each pollutant. Carbon monoxide (CO), Nitrogen dioxide (NO_2), Sulfur dioxide (SO_2), Particulate matter, (PM2.5 and PM10) air pollutants were considered. Determination of the accuracies of the developed models in forecasting the next ten days was achieved using the error performance metrics of Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). The results of the performance metrics from the models in the same category are correlated and indicated similar trends. Comparison and analysis of the models revealed the one with the most accurate prediction for each sampling station and pollutant.

Keywords- Adaptive Neuro Fuzzy Inference System, air pollution, forecasting, fuzzy inference system, fuzzy time series.

1 INTRODUCTION

A serious global concern is subsequent increase in air pollution brought on by population growth and technological advancement. A wide range of human activities from industrial, traffic, energy production, and construction sites, as well as natural occurrences like volcanoes, forest fires, sea salt, and others, contribute to high levels of air pollution (Isazade *et al.*, 2022). The principal pollutants that affect the majority of countries are ozone, particulate matter, lead, carbon monoxide, nitrogen dioxide, sulfur dioxide, toxic compounds, etc. (Yonar & Yonar, 2023; Zahran et al., 2018). An increase in these pollutants' concentrations, particularly in metropolitan areas, has a substantial negative influence on people's health. Air pollution causes a variety of allergies, respiratory conditions, cardiovascular disorders, and diseases related to acute bronchitis (Al-jarakh et al, 2021; Masih, 2019). This environmental issue needs to be addressed because of these negative consequences on public health. Air pollution models are therefore needed to create warnings and control measures, as well as to examine potential future emission occurrences (Uthayakumar, Thangavelu, & Ramanujam, 2021; Zeinalnezhad *et al.*, 2020).

In this study, air pollution data was acquired from the three different sampling stations in Kaduna metropolis, Nigeria. These data were used to train Fuzzy Time Series (FTS) models for each sampling station. The Adaptive Neuro-Fuzzy Inference System (ANFIS) was used to implement FTS models using MatLab software.

*Corresponding Author

Section B- ELECTRICAL/COMPUTER ENGINEERING & RELATED SCIENCES
Can be cited as:
Folaponmile A., Kolawole S.F. and John S.S. (2023), "Air Pollution Forecasting using Fuzzy Time Series Models for Kaduna Metropolis, Nigeria", FUOYE Journal of Engineering and Technology (FUOYEJET), 8(2), 142-146. http://doi.org/10.46792/fuoyejet.v8i2.968

2 MATERIALS AND METHODS

2.1 IMPLEMENTING FORECASTING MODELS

The processes involved in implementing the forecasting models are:

i. Data acquisition
ii. Data preprocessing
iii. Model implementation
iv. Prediction
v. Model evaluation

2.2 DATA ACQUISITION

Data acquisition was achieved through the development of an IoT-based air pollution data acquisition system. The system consists of a sensor node comprising of DHT11, MICS 6814, PMS5003 and 2SH12 sensors interfaced directly with an Arduino Uno microcontroller. The Microcontroller is interfaced with an internet gateway which transmits sensor data to ThingSpeak cloud platform remotely. DHT11 measures temperature in degree Celsius (°C) and humidity in percentage (%) which are used as metrological inputs. MICS 6814 measures CO in parts per million (ppm), and NO_2 in ppm, 2SH12 measures SO_2 in ppm while PMS5003 measures Particulate Matter 2.5 (PM2.5) and Particulate Matter 10 (PM10) both in ug/m^3. Three separate hardware devices were deployed to three different locations within the city. The areas where the hardware devices were deployed are Kakuri industrial layout, Ahmadu Bello Way commercial center, and Gonin Gora residential station. The acquired data retrieved from the ThingSpeak cloud server in CSV format and saved on a local storage device in Excel format, was primarily numerical. The acquired datasets used in this study are daily ambient air temperature, daily relative humidity, and daily concentrations of CO, NO_2, PM2.5, PM10, and SO_2 pollutants. The data was acquired for a duration of 12 months, from September 1, 2021, to August 31, 2022.

2.3 Data Preprocessing

Data preprocessing is necessary because it enhances consistency, makes datasets more complete, and reconfigures the data. Additionally, it contributes to producing more accurate forecasts. All data preprocessing activities were carried out using MS Excel. Outliners were removed from the resulting data entries using MS Excel formatting toolboxes. Missing values were replaced using linear interpolation method. Normalization is scaling original numerical values to fit into a specified range. In this study, the values were normalized between 0 and 1. Normalization was done using equation 1:

$$\text{Normalization } (x_n) = \frac{x - data_{min}}{datamin_{max}} \quad (1)$$

Where, x_n is the normalized value of the data x, x is the value to be normalized, $data_{min}$ is the minimum value in the dataset and $data_{max}$ is the maximum value in the dataset.

2.4 Model Implementation

A machine learning algorithm that has been applied to data produces a model. The models were implemented using MATLAB 2018b software. FTS models were implemented using the Fuzzy Logic Toolbox. Fuzzy Logic Toolbox software provides a command-line function (anfisedit) and an interactive app (Neuro-Fuzzy Designer) for training an adaptive neuro-fuzzy inference system. Based on the training data, ANFIS creates a fuzzy inference system (FIS) and establishes the parameters of the membership function (MF). In addition, the training data is used to adaptively generate the fussy rule base. In modelling ANFIS, two training algorithms were used separately; these are the back propagation algorithm and the hybrid algorithm. The back-propagation uses the gradient descent approach for all parameters, while the hybrid method uses gradient descent for the parameters related to the input membership functions and least squares estimation for the parameters related to the output membership. The fuzzy inference system is generated either by grid partitioning using Gaussian membership function, or subtractive clustering.

The parameters used in modelling FTS are displayed in table 1. As shown in the table, four ANFIS models were implemented and designated as ANFIS1, ANFIS2, ANFIS3 and ANFI4. The individual parameters for these models are listed in table 1. In summary, ANFIS 1 refers to the model using grid partitioning, Gaussian membership function and hybrid algorithm, ANFIS 2 refers to the model using grid partitioning, Gaussian membership function and backpropagation algorithm, ANFIS 3 refers to the model using subtractive clustering and hybrid algorithm ANFIS 4 refers to the model using subtractive clustering and backpropagation algorithm. A Graphical User Interface (GUI) for a network loaded with data is shown in Figure 1.

70% of the dataset was used for training, 15% for testing and 15% for validation. The inputs are the individual independent meteorological variables of humidity and temperature while the individual measured pollutant is the target. The various output are the dependent variables which are the individual predicted pollutant (CO, NO_2, PM2.5, PM10 and SO_2). Figure 2, shows ANFIS's model structure for a single model. This structure comprises of two inputs, one output and nine rules.

Table 1. ANFIS Network Parameters

Network/Parameter	ANFIS 1	ANFIS 2	ANFIS 3	ANFIS 4
Type of Inference system	Sugeno	Sugeno	Sugeno	Sugeno
Fuzzy Inference System	Grid partitioning	Grid partitioning	Subtractive	Subtractive
Training algorithm	Hybrid	Back	Hybrid	Back
Number of inputs	2	2	2	2
Number of outputs	1	1	1	1
Membership function	Gaussian	Gaussian	Not available	Not available
Training Epoch	100	400	100	400

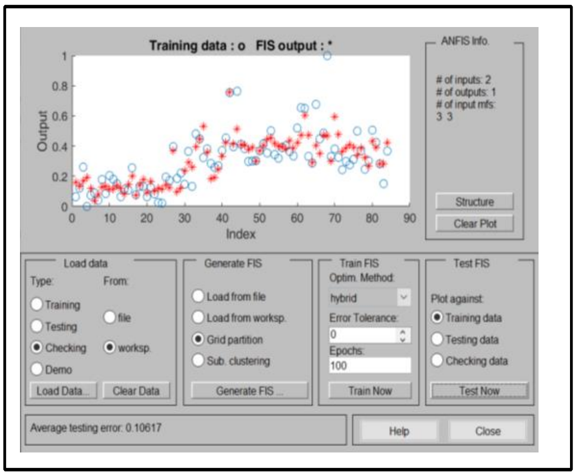

Fig. 1: Neuro-Fuzzy Designer for ANFIS Network Loaded with Data

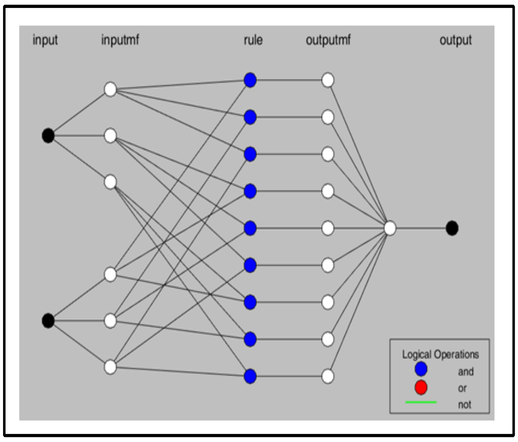

Fig. 2: ANFIS's Model Structure for a single model

http://doi.org/10.46792/fuoyejet.v8i2.968 http://journal.engineering.fuoye.edu.ng/

2.5 Prediction

The developed models were used to forecast air pollutant's concentration for the next ten days. This implies predicting future events using the created model. The forecast for the daily mean of the subsequent ten days was from September 1 to 10, 2022, using the created models. For ANFIS model prediction, the command "output = evalfis(fis,input)" was used at MatLab's command window. This command evaluates the fuzzy inference system, "fis" for the input values in "input" and returns the resulting output values in "output".

2.6 Model Evaluation

MAE, and RMSE are the parameters used to determine the accuracy of the models. The MAE gives an idea about the magnitude of the error. MAE is given by equation 2. The Mean Absolute Error of N forecasting results is defined by:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|F_i - A_i| \qquad (2)$$

RMSE tends to put heavier weight on larger errors. It is the square root of the average of the error squares and is given by:

$$RMSE = \sqrt{\frac{1}{N} x \sum_{i=1}^{N}(F_i - A_i)^2} \qquad (3)$$

Where, F_i is the forecast value and A_i is the actual value.

3 Results and Discussion

To evaluate the developed ANFIS models, the accuracy of each model was determined using error performance evaluation metrics of MAE and RMSE. These error performances results are presented in Fig.3 to 5. On the charts, the terms designated "RES" indicate results for the Gonin Gora residential sampling station, "IND" refers to results for the Kakuri industrial sampling station, and "COM" refers to results from the Ahmadu Bello Way commercial sampling station.

3.1 Discussion of Results

Figures 3 to 5 show the results of MAE and RMSE for the residential, industrial and commercial sampling stations for the CO, NO_2, PM2.5, PM10 and SO_2. As seen on the charts, the lower the heights of the bars, the lower the error, and the better the accuracy of the model. In Figure 3, for the residential sampling station, ANFIS 2 is the best performing model for forecasting air pollution in this particular area. In Figure 4, ANFIS 3 is the best performing model for forecasting air pollution in this particular area. In Figure 5, ANFIS 2 is the best-performing model for forecasting air pollution in this particular area.

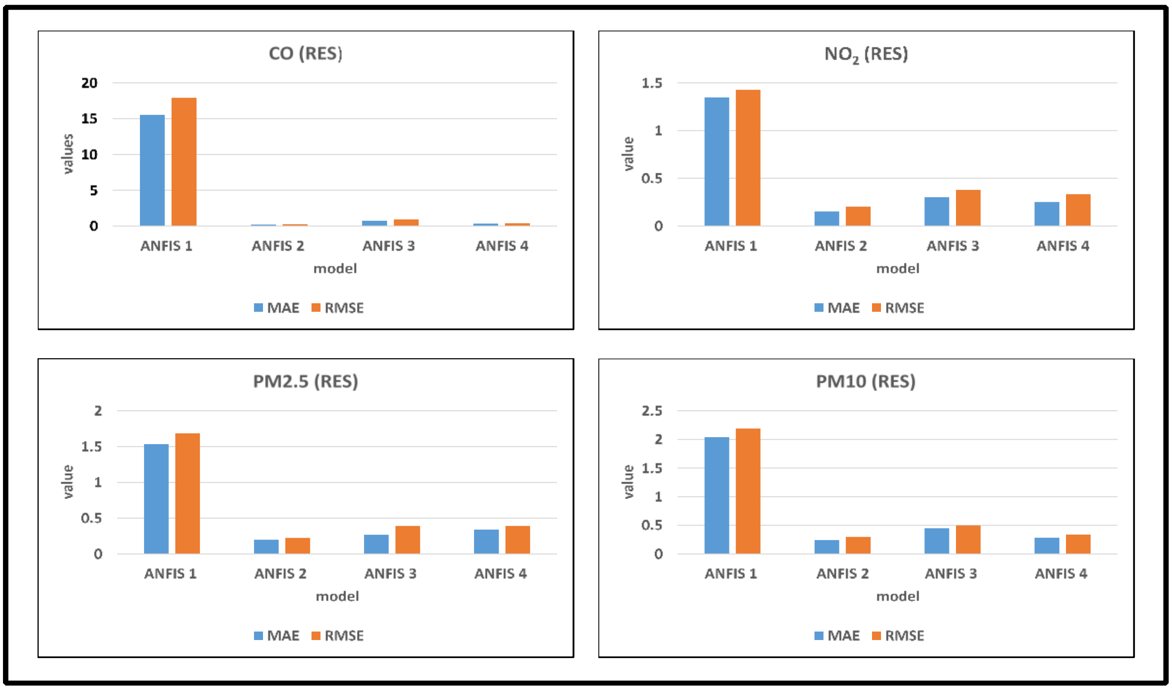

Fig. 3: Error Performance Result for the Residential Sampling Station

http://doi.org/10.46792/fuoyejet.v8i2.968 http://journal.engineering.fuoye.edu.ng/

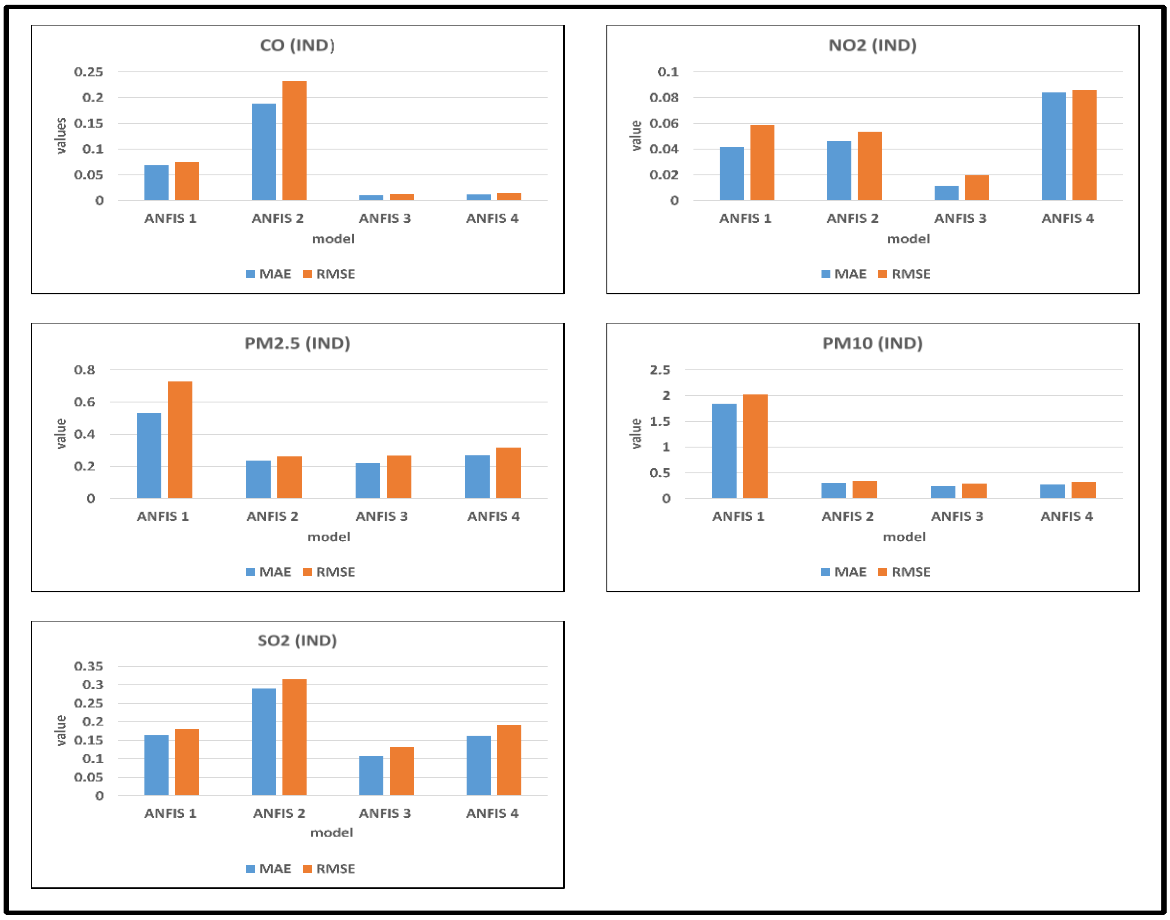

Fig. 4: Error Performance Result for the Industrial Sampling Station.

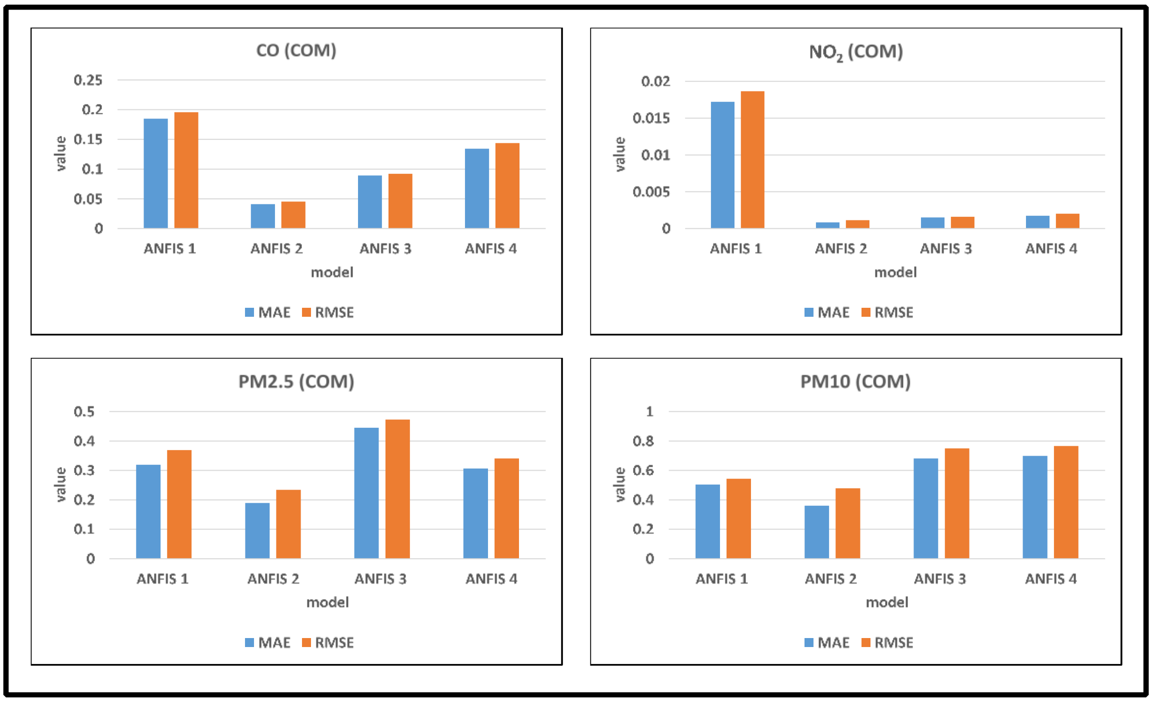

Fig. 5: Error Performance Result for the Commercial Sampling Station.

4 Conclusion

In this study, a total of thirteen different ANFIS models were implemented for CO, NO_2, PM2.5, PM10, and SO_2 pollutants representing three different sampling stations in Kaduna metropolis, Nigeria. The accuracy of each model was determined using error performance evaluation metrics of MAE and RMSE. Based on the comparative analysis of the error performance metrics, the results indicated that ANFIS 2 performed best in the residential and commercial sampling stations, while ANFIS 3 performed best in the industrial sampling station. Conversely, the relatively low values of the error performance metrics indicated that ANFIS is a good technique for forecasting air pollution in Kaduna, Nigeria. For further studies, a comparative study between ANFIS and other computational intelligence techniques such as ANN using the same data should be carried out.

References

Al-jarakh, T. E., Hussein, O. A., Al-azzawi, A. K.,& Mosleh, M. F. (2021). Design and implementation of IoT based environment pollution monitoring system: A Case Study of Iraq. In *IOP Conf. Series: Materials Science and Engineering*, 1105(1), 1 – 21. DOI:10.1088/1757-899X/1105/1/012037

Isazade V., Qasimi A., Seraj K., Isazade E., (2022). Spatial Modeling of Air Pollutant Concentrations Using GWR and ANFIS Models in Tehran City. *Environmental Contaminants Reviews*, 5(2):78-84. DOI: 10.26480/ecr.02.2022.78.84

Masih, A. (2019). Machine learning algorithms in air quality modelling. *Global Journal of Environmental Science and Management*, 5(4), 515 – 534. DOI:10.22034/gjesm.2019.04.0

Uthayakumar, H., Thangavelu, P., & Ramanujam, S. (2021). Forecasting of outdoor air quality index using adaptive neuro fuzzy inference system. *Journal of Air Pollution and Health*. 6 (3):161-170. https://doi.org/10.18502/japh.v6i3.8228

Yonar A. & Yonar H., (2023) Modeling air pollution by integrating ANFIS and Metaheuristic algorithms. *Modeling Earth Systems and Environment,* 9(2), 1621-1631 https://doi.org/10.1007/s40808-022-01573-6

Zahran, A. A., Ibrahim, M. I., Ramadan, A. E. D., & Ibrahim, M. M. (2018). Air Quality Indices, Sources and Impact on Human Health of PM 10 and PM 2.5 in Alexandria Governorate, Egypt. *Journal of Environmental Protection*, 9(12), 1237-1261. DOI: 10.4236/jep.2018.912078

Zeinalnezhad, M., Chofreh, A. G., Goni, F. A., & Kleme, J. J. (2020). Air pollution prediction using semi-experimental regression model and Adaptive Neuro-Fuzzy Inference System. *IEEE 2019 4th International Conference on Smart and Sustainable Technologies (SpliTech) - Split, Croatia*, (pp. 1 – 3). doi:10.23919/SpliTech.2019.8783075

http://doi.org/10.46792/fuoyejet.v8i2.968 http://journal.engineering.fuoye.edu.ng/