ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 9, Issue, 06, pp.53027-53031, June, 2017

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

REVIEW ARTICLE

ARTIFICIAL NEURAL NETWORK AND HYBRID ANN MODELS FOR THE PREDICTION OF PARTICULATE MATTER (PM10) CONCENTRATION, TRIVANDRUM (INDIA)

***Athira, T.**

Assistant Professor in Ernad Knowledge City Technical Campus, Manjeri, Malappuram, Kerala, India

ARTICLE INFO

Article History:
Received 03rd March, 2017
Received in revised form
12th April, 2017
Accepted 09th May, 2017
Published online 30th June, 2017

Key words:
Air pollution, Artificial neural network,
Hybrid ARIMA- ANN,
Particulate matter, Prediction,
Risk- impact asses sment,
Statistical analysis.

ABSTRACT

Air pollution rate is getting verse day by day around the world due to industrialization and urbanization. And particulate matter (PM) is considered as one of the major contributor to this increase in air pollution. Besides the deposition of trace elements in air and reduction in visibility, the direct impact of particulate matter on vegetation and human health are serious issues. In several researches a reliable relation was found between health effects and elevated concentrations of atmospheric PM10 concentration. So for the risk-impact assessments and health studies it is very important to quantify the air pollutant concentration rate in the atmosphere and forecast particulate matter concentrations. In the present study particulate matter concentration of PM10 (10 micron size) was predicted using Artificial Neural Network and a Hybrid Artificial Neural Network. The parameters which affect particulate matter concentration such as Temperature maximum, Temperature minimum, Rainfall, Relative humidity and Station level pressure were considered as the input parameters in the modeling. ANN and hybrid ANN models were developed for Trivandrum district and the model performance was compared by statistical evaluation.

Citation: Athira, 2017. "Artificial Neural Network and Hybrid ANN Models for the Prediction of Particulate Matter (PM10) Concentration, Trivandrum (India)", *International Journal of Current Research*, 9, (06), 53027-53031.

INTRODUCTION

AIR pollution is one of the serious problems that have been affecting socio-economic systems, human health, agricultural crops, forest species and ecosystems (Siew *et al*., 2008). Air pollution related issues especially respiratory illness in human has resulted in an increased public awareness of the air quality. The US Environmental Protection Agency (EPA) has set air quality standards for six criteria air pollutants: sulphur dioxide (SO_2), nitrogen dioxide (NO_2), carbon monoxide (CO), ozone (O_3), lead (Pb) and particulate matter with aerodynamic diameters less than or equal to 10 µm (PM_{10}) and 2.5 µm ($PM_{2.5}$) (Lee *et al*., 2012). Of the six air pollutants, PM is said to be the cause for the respiratory health problems associated with air pollution. Particulate matter (PM) is the term used for particles and liquid droplets that are suspended in the air (Wang and Guo, 2009). It is a combination of extremely minute particles and liquid droplets, which includes nitrates, sulphates, organic chemicals, metals, and soil or dust particles. Particulate matter from anthropogenic aerosols are about ten percent of the total mass of aerosols in the atmosphere.

***Corresponding author:* Athira, T.**
Assistant Professor in Ernad Knowledge City Technical Campus, Manjeri, Malappuram, Kerala, India

This increasing trend in particulate matter concentration in the atmosphere and related human health effects have led many countries to establish several laws and regulations to air quality and the required standard of emission levels. Monitoring and controlling of criteria air pollutants are an expensive work which requires skilled personnel (Kandya and Mohan, 2009). The development of air quality models for forecasting particulate matter concentration in the atmosphere can provide early warning to the popula tion and help to take a decision regarding abatement measures and air quality management (Grivas and Chalouakou, 2006). So particular importance and forecasting. The studies related to air quality monitoring and model development in Kerala were very low compared to other states. It presumed that air quality was never been an issue in the state. But based on the PM_{10} concentration data available from the NAAQMS in statewide, it can be seen that PM_{10} concentration exceeds ambient air quality standards and the frequency of that has increased recently. The effective techniques used for PM forecasting are multiple linear regression, artificial neural networks, fuzzy logic and time series Autoregressive integrated moving average (ARIMA). In which ARIMA is good in forecasting linear patterns of time series while ANN gives better prediction results for nonlinear patters (Wang and Guo, 2009). The PM_{10} concentrations values in the air and the meteorological data usually possess both linear and nonlinear patters. So while using any of the above mentioned methods alone for prediction, it will give less accurate prediction. A hybrid model is a combination of both ARIMA and ANN models, which makes use of the good side of both the models and improve the forecast accuracy. Experimental results and past studies have indicated that the hybrid model can be used as an effective tool to improve the PM forecasting accuracy obtained by either of the models used separately (Feng *et al*., 2015).

Methodology

In the present study an artificial neural network model and hybrid ARIMA – ANN model were developed for predicting PM_{10} concentration. From the water and air quality directory published by Kerala state pollution control board, high variation of PM_{10} concentration in Thiruvananthapuram district was observed. The data includes PM concentration from the four monitoring stations in Trivandrum. The meteorological parameters such as maximum temperature, minimum temperature, relative humidity, station level pressure, and rainfall were used for the model development in MATLAB Version R 2013a (Grivas and Chalouakou, 2006).

Study Area

Thiruvananthapuram, the capital city of Kerala was selected as the study area which is situated between north latitudes 8˚17´ to 8˚54´ and east longitudes 76˚41´ to 77˚17´. According to state urbanization report of 2011, its urban content has increased by 20 % from 2001 to 2011. The Kerala State Pollution Control Board is monitoring ambient air quality at four stations in Thiruvananthapuram. The station points are cosmopolitan hospital, SMV School, Veli and Plamood. Cosmopolitan hospital and SMV school stations belong to sensitive category and having a dense vehicular traffic all the time. Veli is an industrial zone of notable activity and Plamoodu station is a residential area majorly affected by traffic-related sources, frequent congestion of the neighboring roads and intense human activity. The map of the study area is shown in Figure 1.

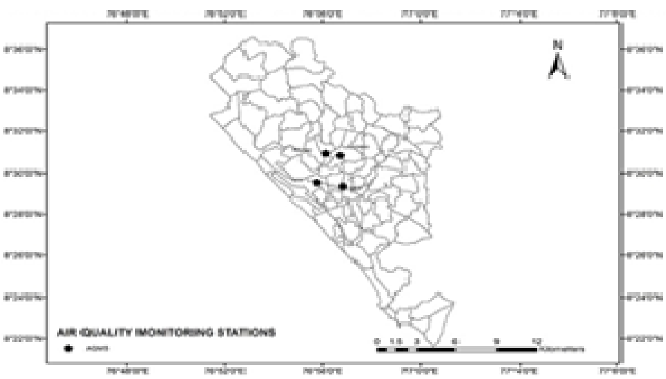

Fig. 1. Study Area

Data collection

The study was based on the data measured for the five year period, from 1 January 2011 to 31 December 2015. It was observed from many studies that the particulate matter concentration in the atmosphere is mainly influenced by the parameters like, background PM_{10} concentration, maximum temperature, minimum temperature, rainfall, station level pressure and relative humidity. Daily average PM_{10} concentration and background PM_{10} concentration data were collected from Kerala state pollution control board (KPCB). The meteorological variables selected for the study were maximum temperature (degree celsius), minimum temperature (degree celsius), relative humidity (percentage), barometric pressure (bar), and amount of rainfall (mm). These data were obtained from Indian meteorological department (IMD), Thiruvananthapuram for the above said period.

Trend Analysis

The trend of PM_{10} concentration for the four stations during the last five years were analyzed. The trend was plotted using Gretl software.

Development of ANN model

The development of ANN model was done in three stages, namely, training, testing and validation. Seventy percentage of the total dataset were used for training, fifteen percentage for testing and the remaining fifteen percentage for validation. PM_{10} concentrations and all the meteorological variables such as maximum temperature, minimum temperature, rainfall, station level pressure and relative humidity were normalized into representative range of values (0, 1) for the computation using ANN. The dataset was normalized and then fed to modeling. The number of input parameters were selected as six and the number of output neurons as one. The number of neurons in the hidden layer was fixed by trial and error method. After creating the network, transfer function was assigned to each connection. Transfer function used here was "transig" and the training function was "trainlm". The Levenberg–Marquardt algorithm or the "trainlm" is a highly reliable region-based method with a hyperspherical trust region. The performance function was chosen as mean square error (MSE). Training parameters like "goal", "epoch", "min gradient" was set. The model was trained by changing the number of hidden neurons. Architecture with least MSE value was selected as the best architecture.

Development of Hybrid ARIMA- ANN model

A hybrid ARIMA- ANN model was developed to use the good sides of each model to capture different patterns in the air quality data. This was developed as a combination of both time series ARIMA model and ANN model for the prediction of PM_{10} concentration. The time series data consists of both linear and nonlinear patterns. For the hybrid model development time series data was first decomposed into linear and nonlinear components using moving average filter method which gives linear part of the time series and nonlinear residual. This operation was performed using Gretl software. For the prediction using hybrid model, first the linear dataset was fed in ARIMA model and the residuals are predicted using ANN model.

Development of ARIMA model

For the development of ARIMA model the dataset was first stationarized which gives the order of differencing. The determination of the order of differencing is needed to stationarize the series and remove the gross features of seasonality. Normally, the correct amount of differencing is the lowest order of differencing that yields a time series which fluctuates around a well-defined mean value and whose autocorrelation function (ACF) plot decays fairly rapidly to zero, either from above or below. If the series still exhibits a long-term trend, or otherwise lacks a tendency to return to its mean value, or if its autocorrelations are positive out to a high number of lags then it needs a higher order of differencing (3). The differencing was done by augmented dickey fuller (ADF) unit root test and a significance level of 10% was assumed. The obtained p-value was compared with the significance level which decides the order of differencing. After differencing Partial Autocorrelation function (PACF) and Autocorrelation function (ACF) were plotted to identify p and q values. ACF plot gives moving average (MA) part and PACF gives autoregressive part (AR) of time series. The time series data can be predicted with the p, d and q values (Kaushik and Melwani, 2007).

Development of ANN with the residual

Obtained the residuals after smoothing of time series data. The nonlinear residual dataset was fed into ANN model without normalization. Rest of the procedure was same as that of the ANN model mentioned (Feng, 2015).

Model Performance Evaluation

Statistical analysis was carried out to identify the performance of the models developed. For identification of the models performance, the criteria chosen were, Mean Absolute percentage Error (MAPE), Mean Absolute Error (MAE), Mean Square Error (MSE) and Root Mean Square Error (RMSE). Model with the smallest values of MAPE, MAE, MSE and RMSE was selected as the best model in forecasting of PM (Siew *et al*., 2008).

RESULT AND DISCUSSION

In the present study two different models namely, ANN and hybrid ARIMA- ANN were developed for the PM_{10} prediction. The result obtained from trend analysis and ANN and hybrid ARIMA- ANN modelling are explained in the following section.

Trend Analysis

The variation of PM_{10} concentration in the four stations are shown in figure below. The graph shows a high variation of PM10 concentration over time. S3 ambient monitoring station in the industrial zone showed high PM_{10} concentration where the PM concentration reached the marginal values of NAAQS. S4 station in the residential zone showed low concentrations of PM over the time.

Development of ANN model

ANN models were developed for the four ambient air monitoring stations in Thiruvananthapuram district. The best architecture of ANN for the prediction of PM_{10} concentration was selected by following trial and error approach. The architecture with least MSE value and high R value was selected as the best architecture. The architecture 6- 20- 1 resulted with least MSE value and high R value.

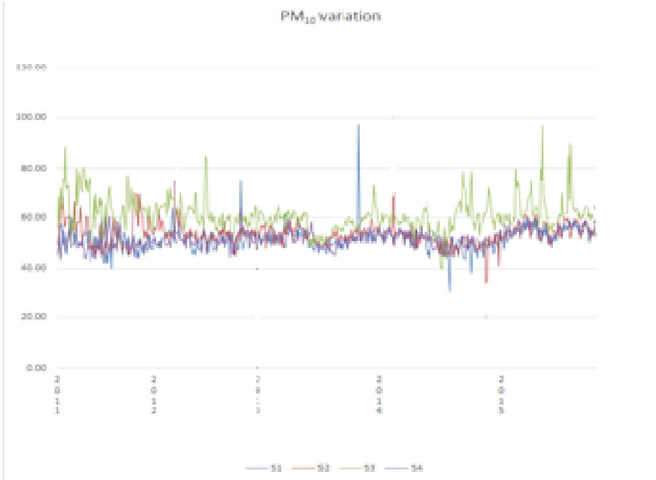

Fig. 2. Trend analysis plot for Thiruvananthapuram stations

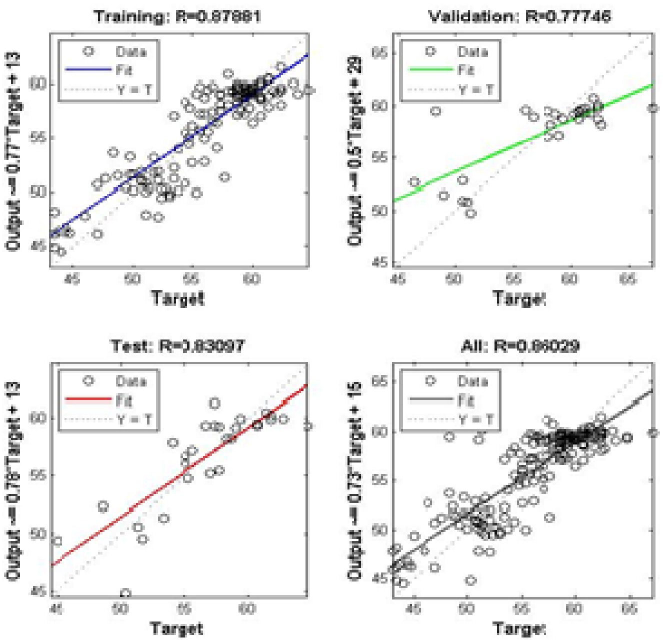

Fig. 3. Regression plot of ANN Trivandrum

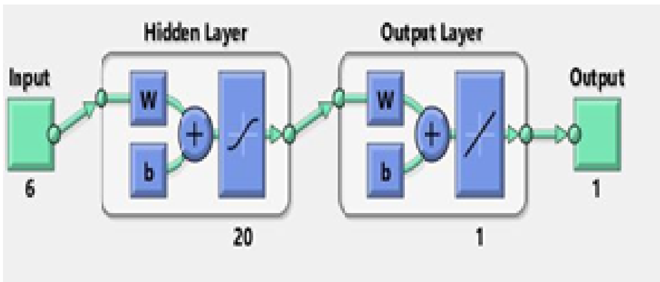

Fig. 4. Typical architecture

Development of ARIMA model

PM concentration data from the four stations shows a varying trend, hence ADF unit root was performed to identify the order of differencing. The p-values obtained after first ADF unit root test were all greater than the significance level. First differencing was carried out for the entire dataset and resulted p- values less than significance level in ADF unit root test, so no further differencing needed.

Development of hybrid ARIMA- ANN model

After predicting PM concentrations in ARIMA the nonlinear residual was obtained and were fed into the ANN. The best architecture selected with least MSE value and high value was obtained. 6- 11- 1 was the best architecture with least MSE value.

Validation of the models

The predicted results obtained from both ANN model and hybrid ARIMA- ANN models compared. The results are given below.

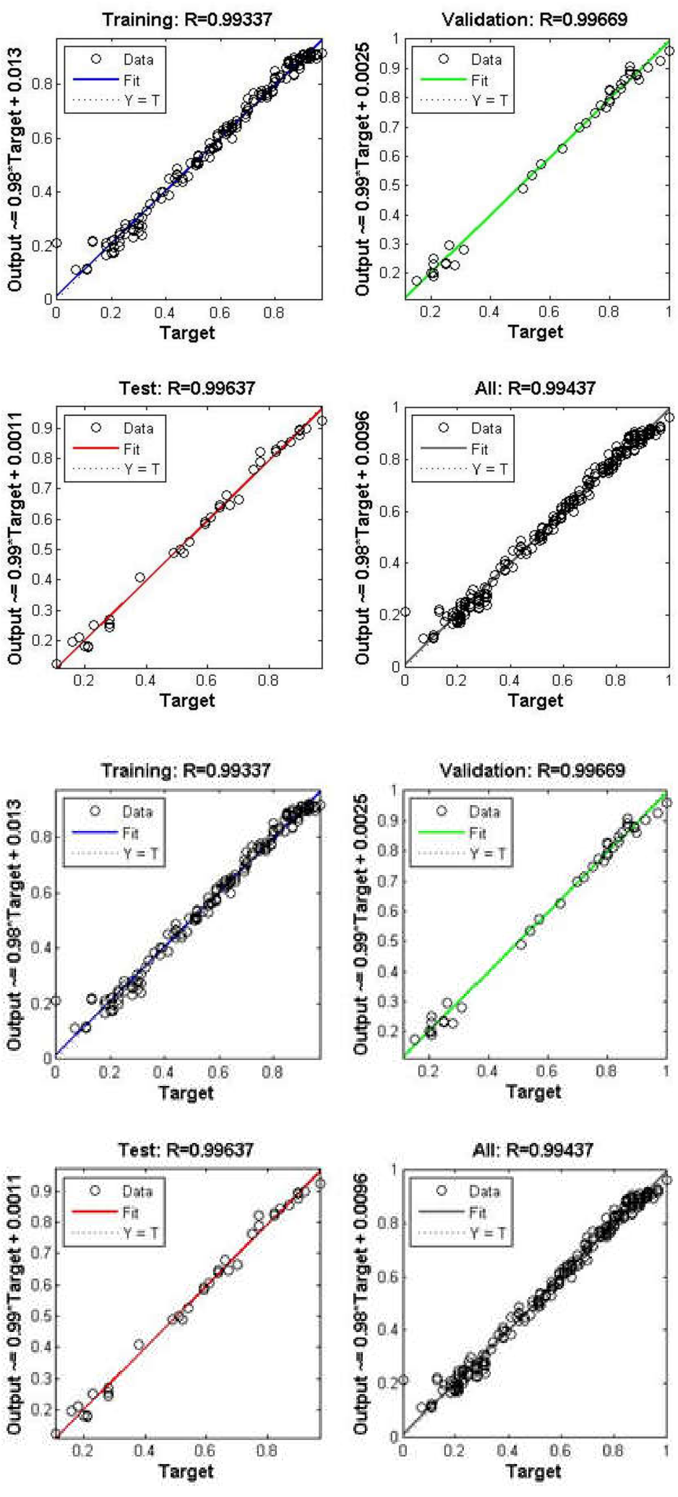

Fig. 5. Regression plot of hybrid model

Performance evaluation

A statistical analysis for the performance evaluation of model developed was yielded following result in Table. The sensitivity analysis of hybrid model yielded low values for all the errors.

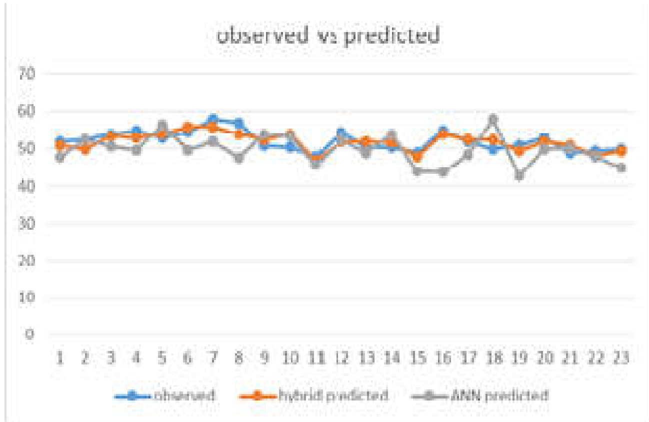

Fig. 6. Validation plot for Trivandrum

Table 1. Statistical analysis results

	ANN	Hybrid ANN
MAPE	19.13	4.2589
MAE	8.6871	2.589
MSE	0.00665	0.0000923
RMSE	10.9683	5.358

Conclusion

Increase in particulate matter (PM_{10}) concentration in capital city Trivandrum is a serious issue that needs to be given serious attention, as it is one of the most important factors that contributes to the quality of life and living. The two models developed in the present study aimed in the prediction of PM_{10} concentration of Thiruvananthapuram district. The trend analysis showed high variation in the PM concentration across the four stations in Thiruvananthapuram which implies the relevance of this study. The hybrid ARIMA- ANN model showed much better prediction accuracy over ANN model. Both the linear as well as the nonlinear patterns of PM concentrations were forecasted by the hybrid model. The statistical analysis of hybrid model yielded lesser values for all the errors like, MAPE, MAP, MSE and RMSE than ANN model.

Acknowledgment

The author thank the Head of the Department and other staff members of Department of Civil Engineering, college of Engineering, Thiruvananthapuram for providing necessary infrastructure for the study. My sincere thanks to Kerala State Pollution control board Thiruvananthapuram for providing relevant data.

REFERENCES

Feng, X., Li, Q., Hou, J., Jin, L. and Wang, J. 2015. "Artificial Neural Networks Forecasting of PM2.5 Pollution Using Air Mass Trajectory Based Geographic Model and wavelet transformation," *Atmospheric Environment*, pp. 118-128.

Grivas, G. and Chalouakou, A. 2006. "Artificial Neural Network Models for prediction of PM10 Hourly Concentrations, in the Greater Area of Athens, Greece," *Atmospheric Environment*, pp. 1216-1229.

Kandya, A. and Mohan, M. 2009. "Forecasting the Urban Air Quality Using Various Statistical Techniques," *The Seventh International Conference on Urban Climate*, Japan.

Kaushik, I. and Melwani, R. 2007. "Time series Analysis of Ambient Air Quality at ITO Intersection in Delhi (India)," *Journal of Environmental Research and Development,* Vol. 2, pp. 268-272.

Lee, M. H., Rahman, N. H. A., Suhartono, M. T., Latif, M. E., Nor and Kamisan, N. A. B. 2012. "Seasonal ARIMA for Forecasting Air Pollution Index: A case Study," *American Journal of Applied Sciences*, pp. 570-578.

Siew, L. Y., Chin, L. Y. and Jin Lee, P. M. 2008. "ARIMA and Integrated ARFIMA Models for Forecasting Air Pollution Index in Shah Alam, Selangor," *The Malaysian Journal of Analytical Sciences*, Vol. 12, pp. 257-263.

Wang, W. and Guo, Y. 2009. "Air Pollution PM2.5 Data Analysis in Los Angeles Long Beach with Seasonal ARIMA Model," *International Conference on Energy and Environment Technology*.

****