Prediction of Respirable Particulate Matter (PM_{10}) Concentration using Artificial Neural Network in Kota city

Shikha Saxena
Civil engineering department
Rajasthan Technical University
Kota, India
Johrisaxena@gmail.com

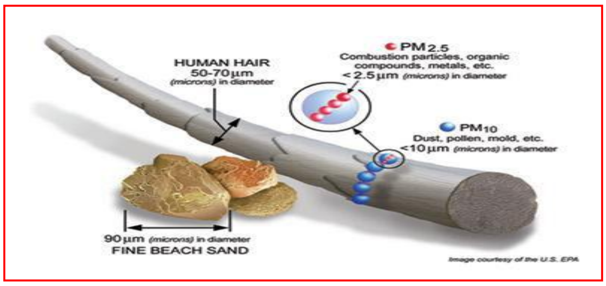
Anil K Mathur
Civil engineering department
Rajasthan Technical University,
Kota, India
shaileekittu@gmail.com

***Abstract*— Recent years concerns related with ambient air quality is prominent due to increments in the entropy and ozone layer depletion. Green house gas emission from the industries is the key contributing factor in the increase in carbon foot prints. The accurate prediction of hazardous gases in the environment can be beneficial information to initiate the corrective strategies for reduction in carbon foot prints. This paper presents a supervised learning based prediction engine for prediction of Respirable Suspended Particulate Matter (RSPM). "Supervised learning" takes a known set of input data and known responses to the data, and seeks to build a predictor model that generates reasonable predictions for the response of new data. Data of 2012,2013 and 2014 of an industrial area of Kota city is employed to train, test and validate four different topologies of the neural networks namely Feed Forward Neural Network (FFNN), Layer Recurrent Neural Network (LRNN),Nonlinear autoregressive Exogenous (NARX) and Radial Basis Function Neural Network (RBFN). A meaningful comparison between these topologies revealed that RBFN is a suitable topology for prediction engine. PM10 constitutes solid and liquid suspended particles having an aerodynamic diameter up to 10 µm (micro meter). It is a common pollutant among all sectors namely transports domestic, industries and manufacturing so it is the major common pollutant need to be taken under consideration in terms of air pollution. Most of the cities of India exceed PM10 levels according to the National ambient air quality standard (NAAQS). PM10 is responsible for heart and lungs diseases. To decrease the mortality rate due to these pollutant Effective countermeasures need to be taken. The precise prediction of pollutant needed to alert the population.**

Keywords: Prediction, PM10, Artificial neural network.

Introduction

India, a developing country has rapid and haphazard development leads to air pollution. Air quality have progressively deteriorated due to urbanization, industrialization, uncontrolled increase in vehicles, construction debris, waste incineration and lack of knowledge are the major causes of pollution. Previous studies suggested that in Jaipur, various pollutants are employed (PM10, PM2.5, SO2, NO2, CO) in determining air quality index and it was found that value of Air Quality Index (AQI) for PM10 fall in the range of 154-362, which was maximum. It shows poor air quality with reference to particulate matter. Particulate matter is a major cause of all respiratory problems. Particulate or particulate matter is microscopic solid or liquid matter suspended in earth's atmosphere. PM10 is the particulate matter of diameter less than 10 micrometer. They have impact on climate and precipitation that adversely affect human health. The daily average RSPM concentration prediction has become priorities for environmental health researchers to access the impact of air on the health concern to decrease the mortality rate. In TEMUCO, CHILE epidemiological studies have been conducted to find out the relationship between the air quality and health recent study in Chile found the strong relationship between PM10 and mortality cases (1997-2002) over 65 years old. The accurate model for forecasting the PM levels to alert the population to initiate preventive measures to control pollutant levels. European Union set two limit values for PM10. The mean daily concentration of PM10 may not exceed 50µg/m3 and mean annual concentration of PM10 may not exceed



40µg/m3. Fig.1 shows the size of particulate matter.

Fig.1 Particulate matter

PM10 has primary and secondary origins. The primary source of PM10 is combustion process, vehicular emissions (carbonaceous compound) from exhaust emission, abrasion of tire, re-suspension of dust particle etc. The secondary sources are formed due to condensation of vapors and oxidation of SO2 to H2SO4 and NO2 to HNO3. The future scenario has been calculated for the RSPM only in the absence of comprehensive emission inventory. As RSPM is the common pollutant across all sectors namely transport, domestic, manufacturing industries and power. Fig @ 2 shows the projections of the contribution of various sectors of which transport sector is worse.

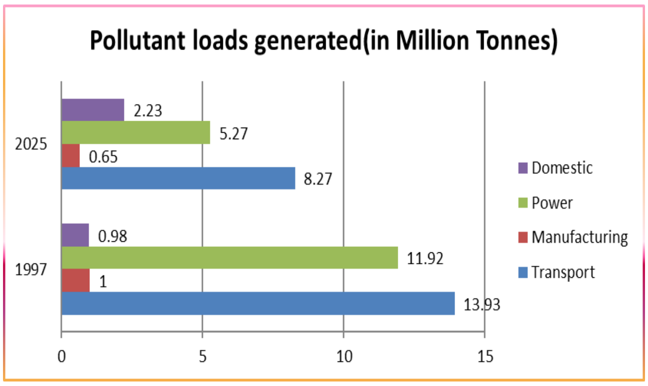


Fig.-2 Future projections of the pollutant load from the various sectors in India

2. Area of study

Kota city is well known as education hub with over 16 lakh habitants. It is situated on the bank of Chambal River in the southern part of Rajasthan. It is the third largest city after Jaipur and Jodhpur. The area of city is 527 sq km and the cartographic coordinates are 25.18º N 75.83º E. It has semi arid climate with high temperature throughout the year. Maximum temperature reaches up to 48o C, which subsides in monsoon season. The average rainfall in Kota is 660.6 mm. Majority of region is rural in character.

As far as the concentration of PM10 (coarse particles) is concerned, Jodhpur has the highest concentration. It has 196 g/m³, while Jaipur 155 g/m³, Kota 146 g/m³, Udaipur 143 g/m³ and Alwar has the lowest air pollution in comparison to other Rajasthan's cities examined at 86 g/m³.

The existing industries (large and medium scale in Kota district), the traffic volume within the towns, along the highways and on the earthen roads, burning of fossil fuel and fire wood in residential areas are probable sources of pollution. In total there are fifteen industrial estates in Kota District and that have been developed by RIICO. There are number of Cement plants, Heavy Chemical Plants, Synthetic Fiber Plant, Thermal Power Plants, Solvent extraction Plant etc. in Kota. The ambient air quality monitoring was carried out in at three stations for RSPM parameter. Air quality map Fig.3 shows RSPM level in Kota district as given below.

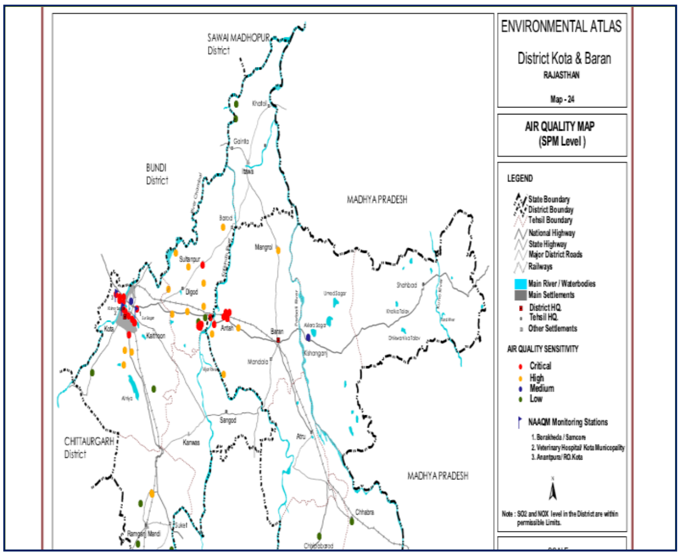


Fig. 3 – SPM level in Kota district

The charts show the variation of concentration of particulate matter for year 2012, 2013 and 2014 respectively. One metrological station (Samcore) has taken into account for frequency distribution.

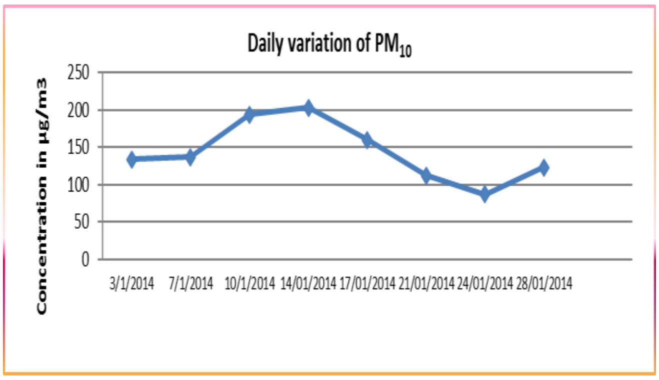


Fig -4 Daily variation of PM10(2012)

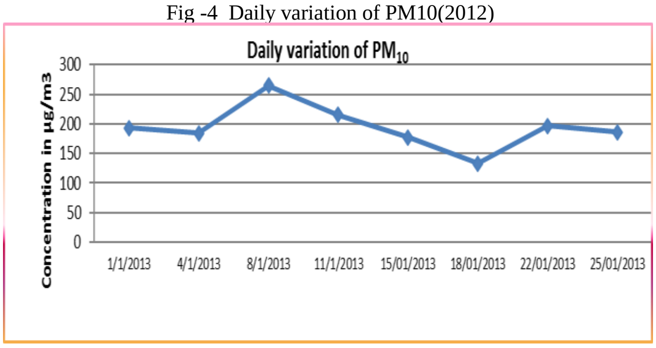


Fig .5 – Daily variation of PM10 (2013)

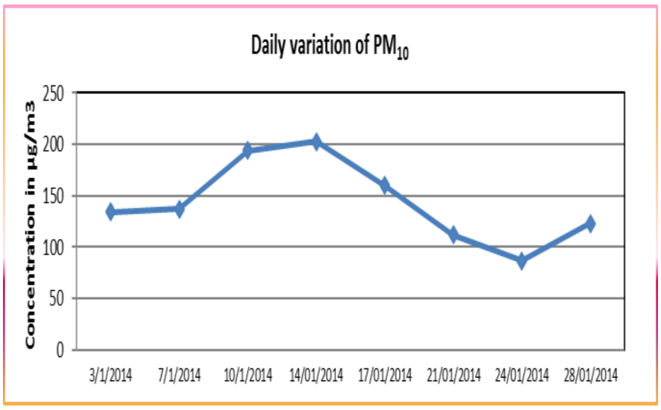


Fig.6 – Daily variation of PM10 (2014)

The prediction is indispensable to prevent dangerous situation. In this regard prediction is done by four different topologies FFNN, RBFN, NARX, and LRNN of ANN.

3. Artificial neural network

Artificial neural networks (ANN) are a branch of artificial intelligence developed in 1950's aiming at imitating the biological brain architecture and have been frequently used as a nonlinear tool in recent atmospheric and air quality forecasting studies. To deal with non-linear systems, especially when theoretical models are difficult to be constructed ANN is very beneficial. ANNs are flexible and nonparametric modeling tools that can perform any complex function mapping with arbitrarily desired accuracy. Modeling with ANN covers a learning (training, validation) and a testing process using historical data by determining nonlinear relationships between the variables in input and output data sets. The basic structure of the ANNs are composed of input and output neurons with weights of interconnection placed in different layers, and internal transfer functions of them. Several artificial neural network were used for ambient air quality forecasting analysis. Feed forward network architectures were applied to forecast PM10 concentration in Targoviste town from Tambovita country. Some feed forward were also applied in forecasting of air pollutants in urban regions.

An Artificial neural network architecture as shown in Fig. 7 comprises of three layers; Input layer, Hidden layer, Output layer. There is two independent datasets of RSPM concentrations of year 2012 and 2013 respectively in Input layer. Three neurons in one hidden layer for prediction. Output layer has one datasets of RSPM 2014. The hybrid model can effectively improve the forecasting accuracy obtained by either of the models used separately.

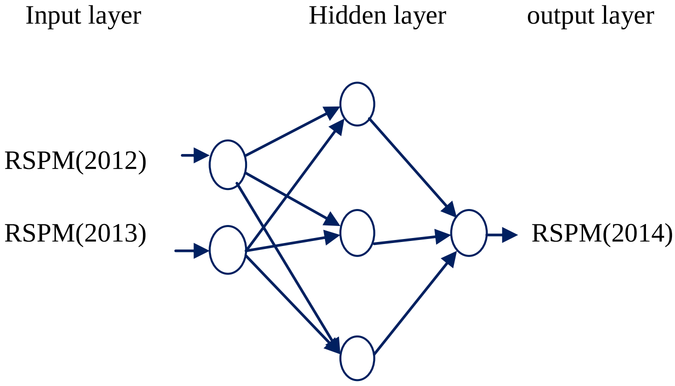


Fig.7 Artificial neural network architecture

The previous studies on statistical modeling of PM10 shows that non linear techniques outperform linear techniques for the temporal prediction of PM10.The frequency distribution of training datasets of PM10 values may strongly influence the results of modeling. However the utilization of uniformly distributive training data result more precise model for prediction.

3.1 Data Analysis:

Usually a large training data set will produce more precise result. However, the trend and pollutant may change overtime therefore more current data might produce better forecasting model.

The database was generated through RPCB official site. Daily data of RSPM from January 2012 to December 2014 for one monitoring station i.e. Samcore glass have been used in this study. Time period is the only predictor variable and daily RSPM concentration as outcome variable.

As the RSPM values (variables) showed a wide range of numerical values, the data set was normalized within a range of (0.1,0.9) using (equation 1) to avoid the overflow of network due to large or small weights produced for the data set considered and to eliminate the influence of dimensions of data on the network . Variables were coded into numerical values before normalization of the data sets.

$$X_{norm} = \frac{(X_{max} - X_{sample}) \times 0.8 + 0.2 X_{min}}{X_{max} - X_{min}} \quad \text{......} \ (1)$$

Where X norm is the normalized value, X sample is the observed value, and X min and X max are the minimum and maximum values of X sample. The output data were transformed back to the real values after prediction.

In this comparative study, Prediction has been done by these four different topologies FFNN, RBFN, NARX, and LRN of ANN. Four different topologies were trained with same input and target data. Four different criteria were used to evaluate the effectiveness of each model i.e. mean square error

, mean absolute error ,sum absolute error and sum square error. The process is shown through Fig. 8 as given below

Fig.8 – Process chart

3.2 Criteria of comparing different architectures

Mean square error(MSE), mean absolute error (MAE), sum square error(SSE) ,sum absolute error(SAE) and are widely acknowledged as effective error indices in precise evaluation in air pollution modeling. Therefore, these are used to judge the reliability in prediction of concentration of SPM in this study. These parameters evaluation are calculated by using the model prediction of SPM concentration .The equation for calculating each of their statistical parameters are as follows: is the predicted concentration by the ANN model and is the corresponding measured concentration.

$$MAE = \frac{\sum | Output_i - Target_i |}{n} \qquad .....(2)$$

$$SAE = \sum | Output_i - Target_i | \qquad ... (3)$$

$$MSE = \frac{\sum (Output_i - Target_i)^2}{n} \qquad ....(4)$$

$$SSE = \sum (Output_i - Target_i)^2 \qquad ... ...(5)$$

4. Results

In this work a supervised learning architecture is proposed to evaluate the ambient air quality of Kota city. The data taken over here is from industrial area (Samcore). In first architecture the inputs of the neural topologies are the raw measurement of the Responded RSPM values of year 2012 and 2013 and the network targets are the RSPM measurements if 2014. For better fitting of the data the data are normalized between 0.1 to 0.9 values. The data of pollution board are employed to frame the prediction model.Four different topologies of neural network namely FFNN, LRNN, NARX and RBFN are utilized and their performances are compared as prediction engines. The simulation work is carried out in MATLAB environment.

Fig. 9, 10, 11 shows the results in terms of prediction error indices of the above said neural topologies. From the results it can be concluded that RBFN topology is more precise as compared with other three topologies.

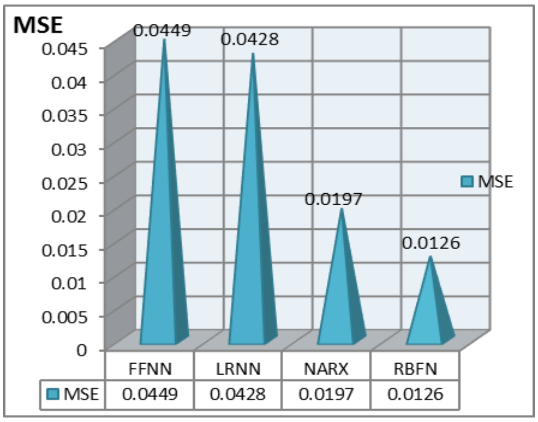


Fig. 9 Minimum square error

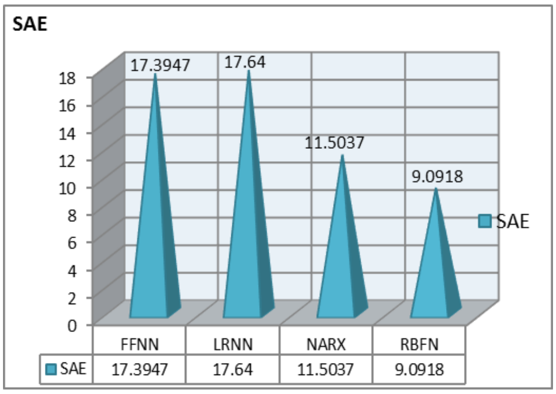


Fig. 10 Sum Absolute Error

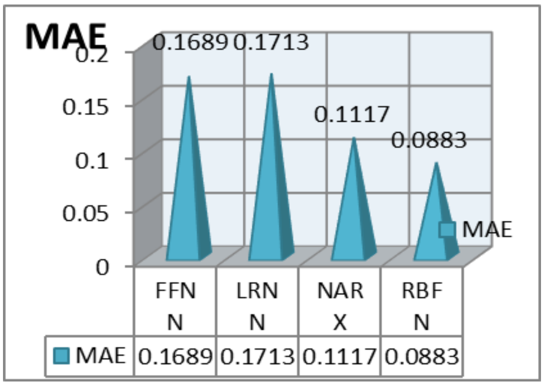


Fig .11 Minimum Average Errors

Table 1 shows the values of normalized prediction results of RSPM from four different topologies for thirteen unseen data patterns.

Table 1 UNSEEN DATA PREDICTION RESULT

ACTUAL	FFNN	LRNN	NARX	RBFN
0.466667	.832162	0.458746	0.601561	0.568054
0.476667	0.793361	0.457517	0.516663	0.404968
0.666667	0.893821	0.470808	0.671218	0.684265
0.696667	0.872437	0.460413	0.635587	0.550476
0.553333	0.769307	0.458332	0.531483	0.493835
0.393333	0.434272	0.464697	0.559091	0.467468
0.310000	0.833743	0.458297	0.576201	0.449890
0.430000	0.799239	0.457287	0.511665	0.386413
0.423333	0.818282	0.457625	0.541155	0.416687
0.373333	0.652413	0.459696	0.463798	0.444030
0.296667	0.269376	0.476543	0.496497	0.444030
0.263333	0.601911	0.460347	0.400401	0.409851
0.273333	0.348913	0.469084	0.517628	0.498718

Fig. 12,13,14,15 shows the graph between actual values and model outputs. As we can see in Fig. 15 the values of output are more near to the actual values. This is clearly amply from the figures that radial basis function network results are more precise than any other network.

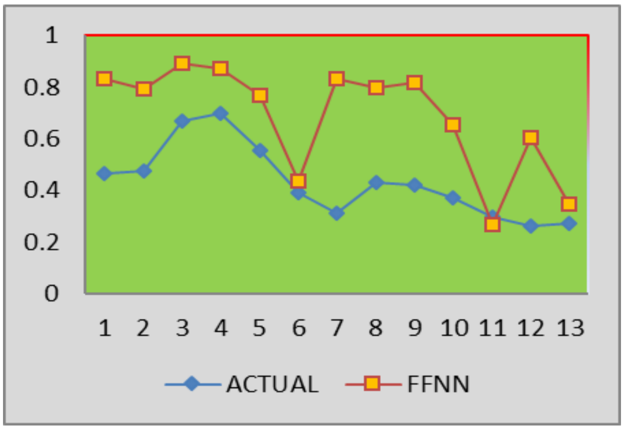


Fig.12: Comparison of FFNN model Predicted values from Actual values

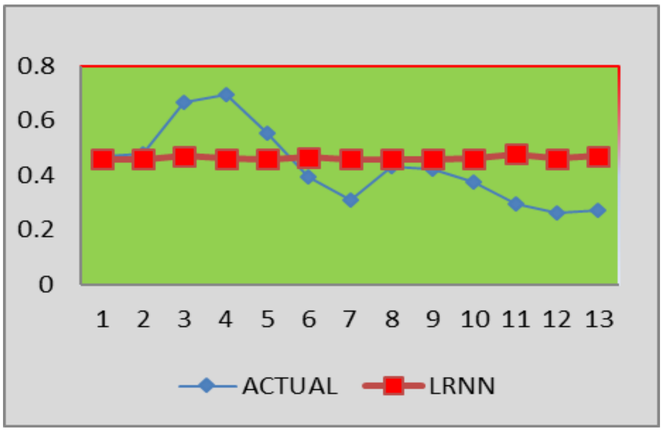


Fig.13: Comparison of LRNN model Predicted values from Actual values

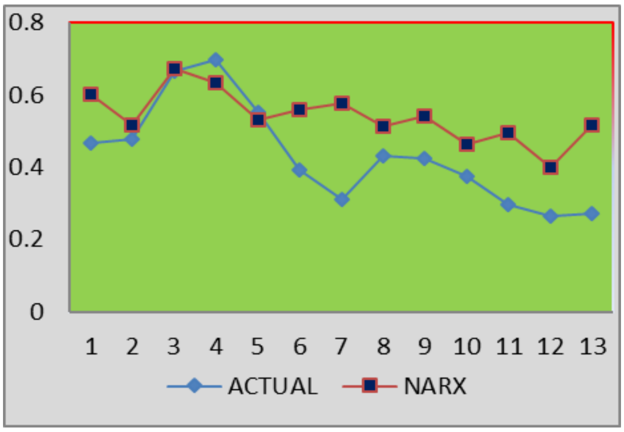


Fig.14: Comparison of NARX model Predicted values from Actual values

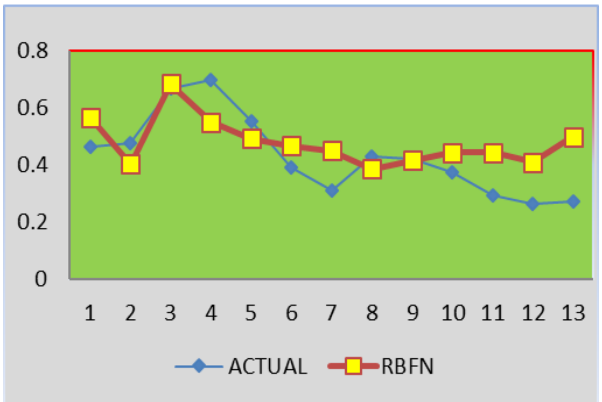


Fig.15: Comparison of RBFN model Predicted values from Actual values

Following points emerged from this analysis.

(i)The ambient air quality is a burning issue as per environment concern and for the betterment of life. A precise prediction engine is well appreciated in this regard. While observing the results from the prediction engines it can be concluded that the values of RSPM predicted by RBFN falls very closer to the actual values. Hence it can be concluded that RBFN topology is the suitable topology for the prediction engine.

(ii)The worst prediction is performed by LRNN topologies as the predicted values by this topology are placed quite far. Moreover the noise sensitivity of these recurrent approaches put a big question mark on the performance as prediction engine.

(iii)After calculating the values of standard deviations and average of these unseen data pattern is observed that the actual values of average RSPM and standard deviations in those are 0.3712 and 0.089 however for FFNN it is 0.53 and 0.1, for LRNN it is 0.22 and 0.2, for NARX 0.64 and 0.11 and for RBFN these values are 0.43 and 0.06. The values of these statistical indices of RBFNs are quite close to the real values. Hence it can be concluded that RBFN is a suitable topology for the prediction of RSPM.

5. Conclusion

Environmental concerns are cropping up with every passing day due to increase in the entropy of the earth. With this, the employment of accurate prediction engines is inevitable to predict the quality of ambient air. This paper proposes architecture of supervised learning to predict the RSPM in industrial zone of Kota city. Following are the major findings of this research work.

(i)The historical data of RSPM values of years 2012 and 2013 are taken as potential input features of different supervised learning models. The targets of these models are taken as RSPM of 2014.

(ii)The meaningful analysis of the prediction accuracy of four neural topologies (FFNN, LRNN, NARX and RBFN) is carried out with the calculation of three different error indices namely MSE, SSE and SAE.

(iii)It is observed that RBFN is most accurate and outperformed over rest of three topologies as the prediction errors minimum for this topology.

The application of hybrid approaches which includes time series modeling and supervised learning models lays in future scope.

REFERENCES

[1] Sharma Sanjeev Kumar , Sharma Kriti, “Ambient Air Quality Status of Jaipur City, Rajasthan, India”, International Research Journal of Environment Sciences ,ISSN 2319–1414Vol. 5(1), 43-48, January (2016)

[2] Sanhueza, P., Vargas, C., Mellado, P., 2005. Impact of air pollution by fine particulate matter (PM10) on daily mortality in Temuco, Chile. Revista Medica De Chile 134, 754–761

[3] Luis A, Dı´az-Robles, “A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile” Atmospheric environment 42(2008)8331- 8340

[4] Statistical Modeling Approaches for PM10 Predictionin Urban Areas; A Review of 21st-Century Studies Hamid Taheri Shahraiyni and Sahar Sodoudi .

[5] R.K. Pacharuri, “ Environment- The concerns and strategies”

[6] District Environmental atlas(Environmental consideration),A report (kota district)

[7] M.W Gardner and S.R Dorling (1998), “Artificial neural network (The multilayer perceptron)”Article in atmospheric environment vol 32, No.14/15 pp 2627-2636, 1998.

[8] De Gennaro G, Trizio L, Gilio AD, Pey J, Pérez N, Cusack M, et al. Neural network model for the prediction of PM10 daily concentrations in two sites in the Western Mediterranean. Science of The Total Environment. 2013;463-464(0):875–883.

[9] Fatih Taşpınar , "Improving artificial neural network model predictions of daily average PM10 concentrations by applying principle component analysis and implementing seasonal models"

[10] D. Dunea, V. Moise, "Artificial Neural Networks as Support for Leaf Area Modelling in Crop Canopies", New Aspects of computers, Proceedings of the 12th WSEAS International Conference on COMPUTERS,2008, pp. 440-446.

[11] Mihaela Oprea, Alexandra Mateid, "Applying Artificial Neural Networks in Environmental Prediction Systems"Department of informatics ,Romania.

[12] D. Dunea, E. Lungu, M. Oprea, Investigating dataprocessing and modelling for urban air qualityforecasting: A comparison between statistical andneural network approaches, EUROPEAN MEETINGPOINT: Energy for Development, Beja, Portugal,2007.

[13] C. Nicolescu, G. Gorghiu, D. Dunea, L. Buruleanu,V. Moise, Mapping Air Quality: An Assessment of the Pollutants Dispersion in Inhabited Areas to Predict and Manage Environmental Risks, WSEAS Transactions on Environment and Development, Issue 12, Volume 4, 2008.

[14] M. Oprea, A case study of knowledge modelling inan air pollution control decision support system, AiCommunications, IOS Press, Vol. 18, No.4, 2005, pp. 293-303.

[15] M. Oprea, Some Ecological PhenomenaForecasting by Using an Artificial Neural Network, Proc. of the 16th IASTED Int. Conf. Applied Informatics, Garmisch-Partenkirchen, 1998, pp. 30-33.

[16] M. Oprea, C. Nichita, D. Dunea, Applications ofArtificial Intelligence in Environmental Protection, (in Romanian), University of Ploiesti Publishing'House, 2008

[17] Slini T, Kaprara A, Karatzas K, Moussiopoulos N. PM10 Forecasting for Thessaloniki, Greece. Environmental Modelling & Software. 2006;21(4):559–565.

[18] Kolehmainen, M., H. Martikainen, and J. Ruuskanen. Neural networks and periodic components used in air quality forecasting. Atmospheric Environment, 35:815–825, 2001.

[19] Dı´az-Robles, "A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile" Atmospheric environment 42(2008)8331- 8340