International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 4, Issue 6, June-2018

Artificial Neural Network model for Respirable Particulate Matter Concentration Prediction

Mane S. J.[1] Sonaje N. P.[2]

[1]Research Scholar, Walchand Institute of Technology, Solapur – 413 006, India, msachin77@yahoo.co.in
[2]Deputy Registrar, Shivaji University, Kolhapur - 416 004, India, nitinsonaje@yahoo.co.in

***Abstract-** In this paper, an artificial neural network is proposed to predict RSPM concentrations for Pune city, the major industrial metropolis of Maharashtra, India. The developed artificial neural network models involve meteorological parameters viz., temperature, relative humidity and wind speed and historical data on observed RSPM concentrations for 6 yrs (2004-2010) as input. The subsequent RSPM concentration for one day ahead being the output parameter is estimated. The developed model is based on three-layer neural network trained by a back-propagation algorithm with number of epoch. The developed model accurately match the trend of RSPM concentrations for one day ahead upto 90.5%.*

***Keywords-** Artificial neural network, RSPM, back-propagation, transform, Multilayer Perceptron Network*

I. INTRODUCTION

Due to industrialization and globalization population density of Indian cities has been increasing day by day which lead to environmental pollution and deterioration of quality of life. Many epidemiological studies have demonstrated the association of air pollution with a deterioration of human health due to elevated concentration of different air pollutant such as RSPM, RSPM, NO_x, CO, etc.[1],[3],[5],[6] More attention is being given to RSPM which has strong relation between lung function, asthma and other cardio related problems.[8] RSPM are creating not only to acute pulmonary and cardiovascular problems on humans but also degrading air quality. Health effects of RSPM on human can be linked congestive heart failure, heart attacks, Chronic Obstructive Pulmonary Disease (COPD). [2],[4],[7]

The prediction of air pollution levels is much useful to enables the authorities to issue a warning of high pollutant concentrations, which may be detrimental for health. Different methods and techniques are used to predict pollutant concentrations such as linear regression methods, fixed box methods, Holt's linear method, computational fluid dynamics (CFD) simulation, artificial intelligence (AI), etc. Artificial intelligence method includes genetic programming, fuzzy logic, artificial neural network (ANN), etc. Computing with ANN is one of the faster growing fields in the artificial intelligence, largely because ANNs can be trained to identify nonlinear patterns between input and output so as to predict the futuristic values. Owing to their wide range of applicability and their ability to learn complex and non-linear relationships-including noisy or less precise information, ANNs are very well suited to solving problems in environmental engineering and in particular, analyzing air pollution [6],[8].

II. MATERIAL AND METHOD

In a broad sense, Pune city comprises with its industrial twin Pimpri-Chinchwad and three cantonment towns of Pune, Khadki and Dehu Road, having population of about 3.13 millions located in the western region of Maharashtra, India. It has a hot semi-arid climate with average temperatures ranging between 19°C and 33°C. Pune is a major industrial metropolis known for its manufacturing and automobile industries with large industrial area covering nearly 10000 industries. An adverse consequence of the rapid rise in urban population, vehicular traffic and industrialization has degraded environment, particularly air quality [6]. According to national ambient air quality standards the limit for RSPM is 100μg/m^3, however, average observed values of 6 year data is upto 370 μg/m^3 which is dangerous to human health. Also, having increasing trend of RSPM has intimated alarming situation. Thus controlling this air pollutant parameter is highly necessary and hence its prediction is of high importance. The potential of three-layered MLP, GFN, R and TLR networks trained by back-propagation algorithm in forecasting RSPM concentration has been utilize to develop a time series model for Pune city. The daily data of RSPM from Sept 15, 2004 to Dec 23, 2010 of about 6 yrs has been obtained for Pune city from Maharashtra Pollution Control Board[9] and also meteorological variables such as temperature, relative humidity and wind speed obtained from Indian Meteorological Department[10]. The daily RSPM concentration consisting of 2291 data points are used for modeling of prediction of RSPM concentration for one day ahead.

With this data, "Best Suitable Network" of ANN architecture and its associated training technique is developed for prediction of RSPM concentration, one day ahead. The developed network is made up of number of interconnected nodes (processing elements), arranged into three basic layers viz., input, hidden and output layers.

Four types of networks used to develop model by using Tanhaxon as transfer function and learning rule as LevenbergMarquardt. Authors has developed large number of ANN models, out of these models, best suitable ANN model is investigated in detail. It has been found that three types of architecture viz., 4-2-0-1, 4-2-1-1, 4-2-2-1 with training and testing data in the ratio of 50:50, 60:40, 70:30, 80:20 and 90:10 with an epoch values of 500, 1000, 2000 and 3000 are utilized for detailed investigation.

Hence, in the present study ANN has been used for short term prediction of RSPM. For the purpose of prediction study, Multi Layer Perception (MLP) Network, Generalized Feed Forward (GFF) Network, Recurrent (R) Network and Time Lagged Recurrent (TLR) Network models of ANN are used to forecast one day ahead RSPM concentration observed in Pune city of Maharashtra, India by using observed historical data on daily basis from the period from Sept 15, 2004 to Dec 23, 2010 of RSPM, along with meteorological variables viz., temperature, relative humidity and wind speed.

The sample network of 4-2-2-1 architecture is as shown in Fig 1 below[6].

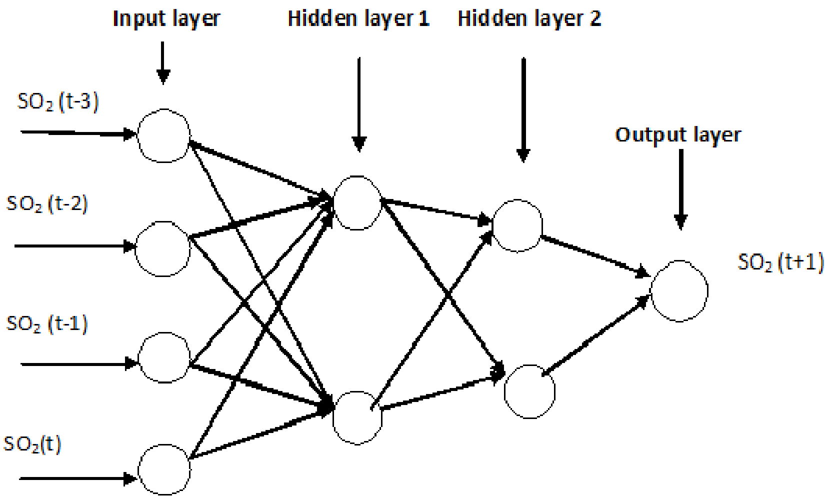


Fig. 1: Typical Neural Network Architecture.

As mentioned earlier, number of trials has been performed on the observed data and finally it is concluded that following set of combination of parameters has been considered for detailed experimentation,

Network	Architecture	Training : Testing	Epoch
MLP GFF R TLR	4 - 2- 0- 1 4 - 2- 1- 1 4 - 2- 2- 1	50 : 50 60 : 40 70 : 30 80 : 20 90 : 10	500 1000 2000 3000

With the above combinations, for each network, 60 experiments and thus in all total 240 experiments has been carried out. Out of these, "Suitable Models" are chosen for further experimentation with the help of statistical tools such Coefficient of Correlation (R), Mean Absolute Error (MAE), Root Mean Square Error (RMSE). Maximum coefficient of correlation as well as minimum value of MAE and RMSE gives the "Best Suitable Model" for further experimentation purpose.

The "Best Suitable Model" obtained as above has been analyzed by transforming the input into logarithmic, square root and moving average values along with and without transformed meteorological parameter viz., humidity, temperature and wind speed. After getting the "Best Suitable Model" from above, further experimentation has been carried out to get the "Best Architect" with the same statistical method.

III. EXPERIMENTATION

For prediction of RSPM concentration, one day ahead (t+1), in each architecture, four observations of RSPM for t, t-1, t-2 and t-3 has been considered as an input. Similarly, when metrological parameters are added, then only 't' value of RSPM with humidity (H), temperature (T) and wind speed (W) are considered as input.

To obtain the "Best Suitable Model, with the help of ANN, for prediction of RSPM concentration, one day ahead (t+1), following steps are adopted,

1. For each network, one architecture, one value of epoch and five training and testing combination sets has been considered. With the help of statistical tools, "Best Model" has been obtained for each such set by using input values of observed RSPM for t, t-1, t-2 and t-3 without metrological parameters. One such "Suitable Model" for each set is obtained and tabulated in Table 1 below,

Table 1:- Models for RSPM values without metrological parameters with each set of epoch

Network	Architecture	Epoch	Training Testing	MAE	MSE	RMSE	R
MLP	4-2-1	500	60:40	17.565	627.458	25.049	0.897
	4-2-1-1	1000	60:40	17.191	613.709	24.773	0.899
	4-2-1	2000	60:40	17.565	627.483	25.050	0.897
	4-2-1-1	**3000**	**60:40**	**17.316**	**602.242**	**24.541**	**0.900**
GFF	4-2-1	500	60:40	17.192	624.557	24.991	0.895
	4-2-1-1	**1000**	**60:40**	**16.945**	**601.862**	**24.533**	**0.900**
	4-2-1-1	2000	60:40	17.591	649.103	25.478	0.891
	4-2-1	3000	60:40	19.109	703.698	26.527	0.881
R	4-2-1	500	60:40	22.190	839.543	28.975	0.878
	4-2-1	1000	90:10	17.204	516.339	22.723	0.893
	4-2-1	2000	90:10	19.563	645.593	25.409	0.859
	4-2-1	3000	60:40	19.109	703.698	26.527	0.881
TLR	4-2-2-1	500	70:30	23.537	1062.468	32.596	0.819
	4-2-2-1	1000	70:30	19.808	785.774	28.032	0.869
	4-2-1-1	2000	60:40	21.179	834.743	28.892	0.874
	4-2-1-1	3000	60:40	17.917	672.770	25.938	0.889

2. From above table 1, for all networks, all architectures and one value of epoch has been considered and with the help of statistical tools, "Best Suitable Model" has been obtained and the results are tabulated in table 2 below,

Table 2:- Best Suitable Models for RSPM values without metrological parameters for each set of epoch

Network	Architecture	Epoch	Training Testing	MAE	MSE	RMSE	R
MLP	4-2-1	500	60:40	17.565	627.458	25.049	0.897
GFF	**4-2-1-1**	**1000**	**60:40**	**16.945**	**601.862**	**24.533**	**0.900**
MLP	4-2-1	2000	60:40	17.565	627.483	25.050	0.897
MLP	4-2-1-1	3000	60:40	17.316	602.242	24.541	0.900

3. For each network, one architecture, one value of epoch and five training and testing combination sets has been considered. With the help of statistical tools, "Best Model" has been obtained for each such set by using input values of observed RSPM for t, and all three metrological parameters. One such "Suitable Model" for each set is obtained and tabulated in Table 3 below,

Table 3:- Models for RSPM values with metrological parameters with each set of epoch

Network	Architecture	Epoch	Training Testing	MAE	MSE	RMSE	R
MLP	**4-2-1-1**	**500**	**60:40**	**17.019**	**602.768**	**24.551**	**0.903**
	4-2-1-1	1000	60:40	17.119	617.385	24.847	0.901
	4-2-1	**2000**	**60:40**	**17.134**	**610.209**	**24.702**	**0.902**
	4-2-1-1	**3000**	**60:40**	**17.115**	**617.283**	**24.845**	**0.901**
GFF	4-2-1	500	60:40	17.541	631.350	25.127	0.900
	4-2-1	**1000**	**60:40**	**17.370**	**621.481**	**24.930**	**0.901**
	4-2-1	2000	60:40	18.392	657.183	25.636	0.895
	4-2-1-1	3000	60:40	17.898	674.388	25.969	0.891
R	4-2-1	500	70:30	20.445	813.306	28.519	0.861
	4-2-1	1000	60:40	20.859	799.608	28.277	0.862
	4-2-1-1	2000	50:50	19.732	771.206	27.771	0.874
	4-2-2-1	3000	90:10	16.327	494.994	22.248	0.900
TLR	4-2-2-1	500	50:50	21.077	863.948	29.393	0.859
	4-2-1	1000	50:50	22.989	945.857	30.755	0.841
	4-2-1	2000	90:10	17.639	576.303	24.006	0.876
	4-2-2-1	3000	60:40	19.551	776.441	27.865	0.870

4. From above table 3, for all networks, all architectures and one value of epoch has been considered and with the help of statistical tools, "Best Suitable Model" has been obtained and the results are tabulated in table 4 below,

Table 4:- "Best Suitable Models" for RSPM values with metrological parameters with each set of epoch

Network	Architecture	Epoch	Training Testing	MAE	MSE	RMSE	R
MLP	**4-2-1-1**	**500**	**60:40**	**17.019**	**602.768**	**24.551**	**0.903**
GFF	4-2-1	1000	60:40	17.370	621.481	24.930	0.901
MLP	4-2-1	2000	60:40	17.134	610.209	24.702	0.902
MLP	4-2-1-1	3000	60:40	17.115	617.283	24.845	0.901

5. After analyzing table 2 and table 4, the "Most Suitable Model" has been obtained for RSPM with and without metrological parameters and are tabulated in table 5 below,

Table 5:- "Most Suitable Model" for RSPM with and without metrological parameters

Network	Architecture	Epoch	Training Testing	MAE	MSE	RMSE	R	
MLP	4-2-1-1	3000	60:40	17.316	602.242	24.541	0.900	RSPM without metrological parameters
MLP	**4-2-1-1**	**500**	**60:40**	**17.019**	**602.768**	**24.551**	**0.903**	**RSPM with metrological parameters**

It is observed from Table 5 above that for prediction of RSPM concentration, one day ahead (t+1), highest efficiency with minimum mean absolute error (MAE) and minimum root mean square error (RMSE) is found for MLP Network with 4-2-1-1 architecture with epoch of 500 having training to testing data in the ratio of 60:40 gives the efficiency of 90.3% is the "Most Suitable Model" having normal values of RSPM pollutant with metrological parameters as an input.

6. The "Most Suitable Model" obtained from table 5 has been analyzed by transforming the input values of RSPM to logarithmic, square root and moving average to normalize input data by using input for t, t-1, t-2 and t-3 without metrological parameters and are tabulated in Table 6 below,

Table 6:- Models for RSPM values without metrological parameters with transformed values of RSPM

Transform Type	Architecture	Epoch	Training Testing	MAE	MSE	RMSE	R
Moving Average	4-2-1-1	500	60:40	18.873	749.647	27.380	0.872
Logarithmic	**4-2-1-1**	**500**	**60:40**	**17.442**	**626.142**	**25.023**	**0.895**
Square root	4-2-1-1	500	60:40	18.758	825.055	28.724	0.860

7. The "Most Suitable Model" obtained from table 5 has been analyzed by transforming the input values of RSPM to logarithmic, square root and moving average to normalize input data with normal values of meteorological parameter and are tabulated in Table 7 below,

Table 7:- Models for RSPM values with transformed values of RSPM and normal values of metrological parameters

Transform Type	Architecture	Epoch	Training Testing	MAE	MSE	RMSE	R
Logarithmic	4-2-1-1	500	60:40	17.477	642.570	25.349	0.898
Square Root	**4-2-1-1**	**500**	**60:40**	**17.841**	**662.155**	**25.732**	**0.904**
Moving Average	4-2-1-1	500	60:40	20.048	804.845	28.370	0.865

8. The "Most Suitable Model" obtained from table 5 has been analyzed by transforming the input values of RSPM and meteorological parameters both to logarithmic, square root and moving average to normalize input data and are tabulated in Table 8 below,

Table 8:- Models for RSPM values with transformed values of RSPM and metrological parameters

Transform Type	Architecture	Epoch	Training Testing	MAE	MSE	RMSE	R
Logarithmic	4-2-1-1	500	60:40	17.581	645.711	25.411	0.900
Square Root	**4-2-1-1**	**500**	**60:40**	**18.038**	**673.355**	**25.949**	**0.905**
moving average	4-2-1-1	500	60:40	19.891	796.243	28.218	0.869

9. After analyzing table 6, 7 and table 8, the "Most Suitable Model" has been obtained for RSPM with and without metrological parameters by transforming the input values and tabulated in Table 9 below,

Table 9:- "Most Suitable Model" for RSPM with and without metrological parameters by transforming the input values

Network	Architecture	Epoch	Training Testing	MAE	MSE	RMSE	R	Input Type
MLP	4-2-1-1	500	60:40	17.442	626.142	25.023	0.895	Logarithmic of RSPM without metrological
MLP	4-2-1-1	500	60:40	17.841	662.155	25.732	0.904	Square root of RSPM with normal values of meteorological parameter
MLP	**4-2-1-1**	**500**	**60:40**	**18.038**	**673.355**	**25.949**	**0.905**	**Square root of RSPM and meteorological parameter**

10. After analyzing table 5 and table 9, the "Most Suitable Model" has been obtained for RSPM with and without metrological parameters by transforming the input values and tabulated in Table 10 below,

Table 10:- "Most Suitable Model" for RSPM with and without metrological parameters as well as transform values of all input values

Network	Architecture	Epoch	Training Testing	MAE	MSE	RMSE	R	Input Type
MLP	4-2-1-1	**500**	60:40	17.019	602.768	24.551	0.903	RSPM with meteorological parameters
MLP	**4-2-1-1**	**500**	**60:40**	**18.038**	**673.355**	**25.949**	**0.905**	**Square root of RSPM and square root of meteorological parameter**

It is observed from Table 10 above that for prediction of RSPM concentration, one day ahead (t+1), highest efficiency with slightly more root mean square error (RMSE) and mean absolute error (MAE) is found for MLP network with 4-2-1-1 architecture with epoch of 500 having training to testing data in the ratio of 60:40 gives efficiency of 90.5% is the "Most Suitable Model" having Square root transformation of RSPM pollutant and metrological parameters as an input.

11. On getting the "Most Suitable Model" for RSPM prediction for one day ahead (t+1), as per table 10 above, the same has been analyzed for various architecture and tabulated in Table 11 below,

Table 11: "Best Suitable Model" of MLP for Square root transformed values of RSPM and metrological parameters as input values for different architecture

Network	Architecture	Epoch	Training Testing	MAE	MSE	RMSE	R
MLP	**4-2-1-1**	**500**	**60:40**	**18.038**	**673.355**	**25.949**	**0.905**
MLP	5-2-1-1	500	60:40	17.175	615.553	24.810	0.899
MLP	6-2-1-1	500	60:40	17.686	649.413	25.484	0.895
MLP	7-2-1-1	500	60:40	17.956	666.376	25.814	0.895
MLP	8-2-1-1	500	60:40	17.692	640.772	25.313	0.894
MLP	9-2-1-1	500	60:40	19.332	754.453	27.467	0.878
MLP	10-2-1-1	500	60:40	18.083	685.765	26.187	0.876

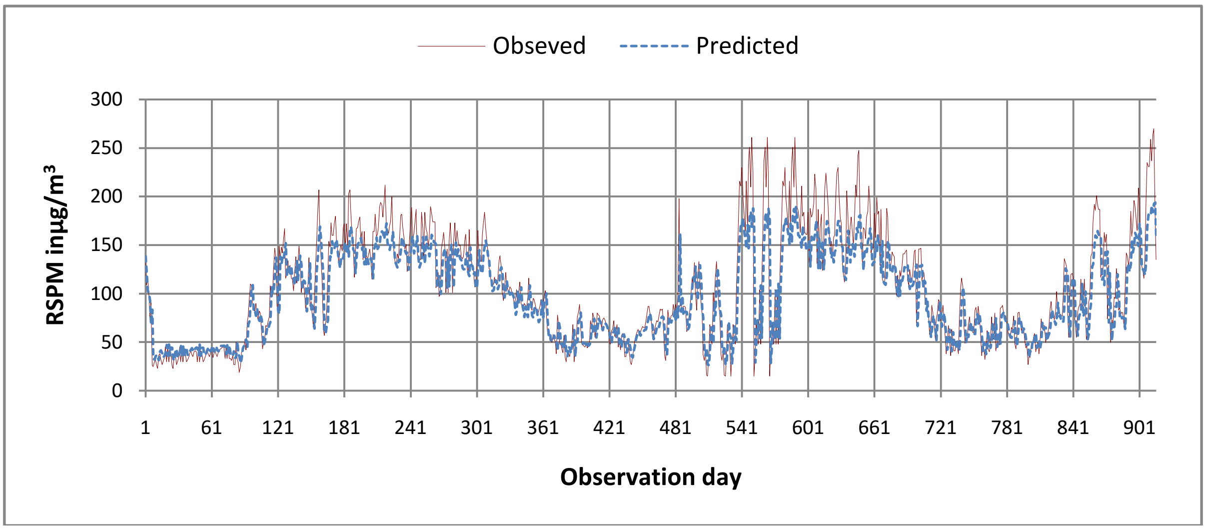
12. After analyzing table 11, the "Most Suitable Model" has been obtained for RSPM with and without metrological parameters by transforming the input values and tabulated in Table 12 below,

Table 12:- "Most Suitable Model" for RSPM with and without metrological parameters

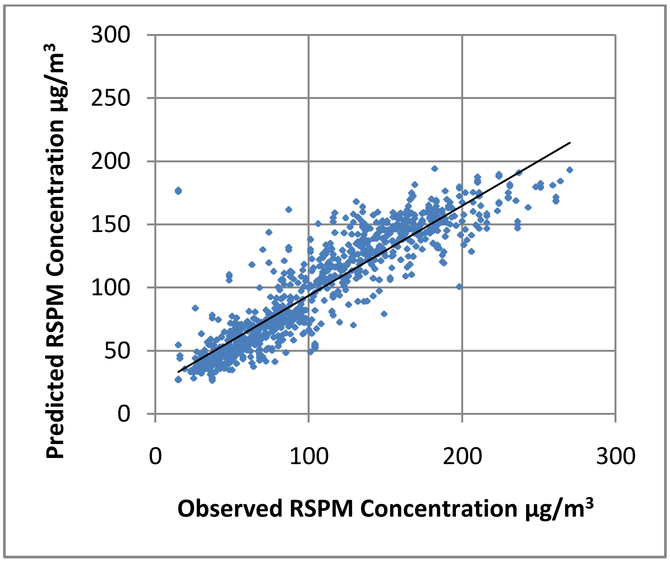
Network	Architecture	Epoch	Training Testing	MAE	MSE	RMSE	R
MLP	4-2-1-1	500	60:40	18.038	673.355	25.949	0.905

It is observed from Table 12 above that the "Most Suitable Model" for prediction of RSPM concentration, one day ahead (t+1), highest efficiency is found for MLP network with 4-2-1-1 architecture with epoch of 500 having training to testing data in the ratio of 60:40 gives efficiency of 90.5% having square root transformed value of observed RSPM for t, t-1, t-2 and t-3 with square root transformed value of metrological parameters as an input.

The results obtained with above model are graphically depicted in graph 1 and 2 as below.



Graph 1: Time series plot of observed and predicted RSPM values for one day ahead (t+1) with "Best Suitable Model"



Graph 2: Comparison of observed and predicted RSPM values plot for one day ahead (t+1) with "Best Suitable Model"

IV. OBSERVATIONS

It is observe from the above table 10, 11 and 12 that the metrological parameters and transform values of input do affect on the prediction value of RSPM. For prediction of RSPM concentration, one day ahead (t+1), by using input values of square root transformation of observed RSPM concentration for t and square root transformation of meteorological parameter gives the optimal results as shown in table 12.

V. CONCLUSION

For prediction of RSPM concentration, one day ahead (t+1), highest efficiency is found for MLP network with 4-2-1-1 architecture with epoch of 500 having training to testing data in the ratio of 60:40 gives efficiency of 90.5% is the "Most Suitable Model" having square root transformed values of RSPM pollutant with square root transformed metrological parameters as an input.

FUTURE SCOPE

Since the prediction of RSPM concentration, one day ahead (t+1) has been analyzed in the present set of experiment, similar prediction can be done for fifteen days ahead (t+15) to thirty days ahead (t+30) and the Most Suitable Model can be found out.

ACKNOWLEDGEMENT

Authors are very much thankful to Principal, Walchand Institute of Technology, Solapur and Principal, Dr. D. Y. Patil, College of Engineering, Akurdi, Pune for providing computing facilities. Authors are also thankful to the local authorities of MPCB and IMD for providing the data.

REFERENCES

[1] Bacharova, L., Fandakova, K., Bratinka, J., Budinska, M., Bachar, J., and Gudaba, M., "The Association between Air Pollution and the Daily Number of Deaths: Findings from the Slovak Republic Contribution to the APHEA Project," *Journal of Epidemiology Community Health*, vol. - 50(S1), p.g. S19-S21, 1996.

[2]. Bartzokas A., Kassomenos P., Petrakis M., Celessides C., "The Effect of Meteorological and Pollution Parameters on the Frequency of Hospital Admissions for Cardiovascular and Respiratory Problems in Athens", *Indoor and Built Environment*, vol. - 13, p.g. 271-275, 2004.

[3]. Lvovsky, K., "Economic Costs of Air Pollution with Special Reference to India" *National Conference on Health and Environment Delhi, India,* July 7-9, 1998.

[4]. Schwartz, J., Slater, D., Larson, T.V., Pierson, W.E., Koenig, J.Q., "Particulate air pollution and hospital emergency room visits for asthma in Seattle", *American Review of Respiratory Disease*, vol. - 147, p.g. 826-831, 1993.

[5]. Singh, G., Pal, A. K., Tiwari, A., "Air Pollution and its Impact on Social Spectrum with Special References to KORBA Coal Field on Chhattisgarh India", *First International conference on MSECCMI, New Delhi*, p.g. 549-558, 2007.

[6] Sonaje N. P., Mane S. J., "Development of SO2 Prediction Model with Artificial Neural Network for Pune City," *International Journal of Application or Innovation in Engineering and Management (IJAIEM)*, Vol.- 7, Issue - 6, p.g.69-76, 2018.

[7]. Tecer, L.H., "Prediction of RSPM and PM Concentrations in a Coastal Mining Area Zonguldak, Turkey Using an Artificial Neural Network", *Environmental Studies*, vol.- 16(4), p.g. 633-638, 2007.

[8]. Timonen, K.L., Pekkanen, J., Tiittanen, P., Salonen, R.O., "Effects of Air Pollution on Changes in Lung Function Induced by Exercise in Children with Chronic Respiratory Symptoms", *Occupational. Environmental Medicine*, vol. - 59(2), p.g. 129-134, 2002.

[9]. http://mpcb.mah.nic.in/envtdata/envtair.php retrieved on 14 Oct 2017.

[10]. http://www.imd.gov.in retrieved on 28 Apr 2018.

.