International Journal of Application or Innovation in Engineering & Management (IJAIEM)
Web Site: www.ijaiem.org Email: editor@ijaiem.org
Volume 7, Issue 6, June 2018 **ISSN 2319 - 4847**

Development of SO_2 Prediction Model with Artificial Neural Network for Pune City

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

[1]Deputy Registrar, Shivaji University, Kolhapur - 416 004, India,
(Corresponding Author)

[2]Research Scholar, Walchand Institute of Technology, Solapur - 413006, India,

Abstract
In this study, an artificial neural network is proposed to predict SO_2 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 SO_2 concentrations for 6 yrs (2004-2010) as input. The subsequent SO_2 concentration for one day ahead being the output parameter of this study is estimated. The developed model is based on three-layer neural network trained by a back-propagation algorithm with number of epoch. The models accurately match the trend of SO_2 concentrations for one day ahead upto 86.9%.
Keywords : Artificial neural network, SO_2, back-propagation, transform, Generalized Feed Forward Network

1.1 Introduction
Urban air pollution is one of the major issues thrown a challenge in many parts of the world on the 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, SO_2, NO_x, CO, etc.[1],[3],[5] More attention is being given to SO_2 since they are creating not only to acute pulmonary and cardiovascular problems on humans but also degrading air quality. Health effects of SO_2 on human being can be linked with constriction of airways, bronchitis, irritation of the eyes, nose, and throat, low depth of breathing, cardiovascular complications, breathing problems and COPD.[2],[4],[6]

The prediction of air pollution levels is very useful to enables the local authorities to give a warning against 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 neural networks 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 values and can solve complex problems. Owing to their wide range of applicability and their ability to learn complex and non-linear relationships-including noisy or less precise information, NNs are very well suited to solving problems in environmental engineering and in particular, analyzing air pollution [7].

1.2 Material and Method
Pune city along with its industrial twin Pimpri-Chinchwad as well as the three cantonment towns of Pune, Khadki and Dehu Road having population of about 3.13 millions located in the western region of Maharashtra, India. Pune 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 population, vehicular traffic and industrialization has increased in environmental degradation, particularly air quality in Pune city. According to national ambient air quality standards the limit for SO_2 is 80 $\mu g/m^3$ however, average observed values of 6 year data is upto 72 $\mu g/m^3$ which is within the limit, but having increasing trend of SO_2 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 SO_2 concentration has been utilize to develop a time series model for Pune city. The daily data of SO_2 from 15th Sept 2004 to 23rd Dec 2010 of about 6 yrs has been obtained for Pune city from MPCB[8] and also meteorological variables such as temperature, relative humidity and wind speed obtained from IMD[9]. The daily SO_2 concentration consisting of 2291 data points are used for modeling of prediction of SO_2 concentration for one day ahead.

With this data, most suitable "Best Fit Network" of ANN architecture and its associated training technique is developed for prediction of SO_2 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.
The methodology adopted in developing the ANN model in the present study is shown in Fig. 1 with 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 for non-transformed and transformed data sets, out of these models, best ANN model is investigated in detail. It has fund 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 along with statistical measures for short term prediction of SO_2. 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 SO_2 concentration observed in Pune city of Maharashtra, India by using observed historical data on daily basis from the period from 15th Sept 2004 to 23rd Dec 2010 of SO_2, 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.

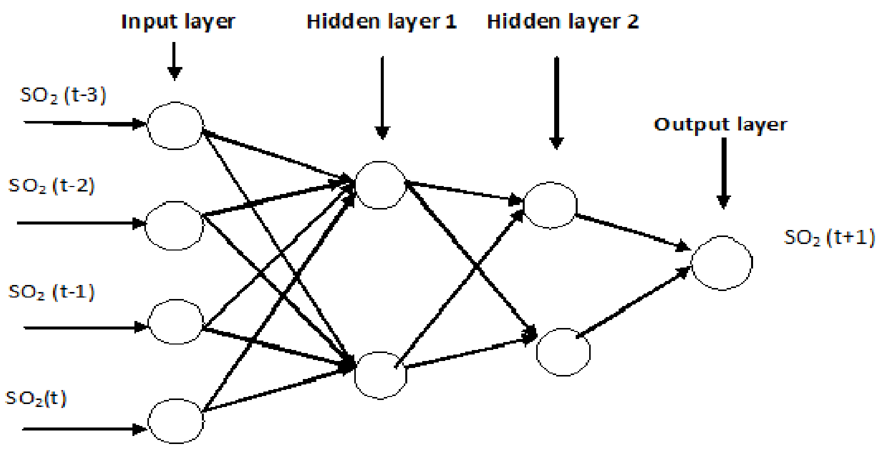


Fig. 1: Typical Neural Network Architecture.

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

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

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

The "Best Fit 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 Fit Model" from above, 240 experiments, further experimentation has been carried out to get the "Best Architect" with the same statistical method.

1.3 Experimentation

For prediction of SO_2 concentration, one day ahead (t+1), in the architecture four observations of SO_2 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 SO_2 with humidity (H), temperature (T) and wind speed (W) are considered as input.

To obtain the "Best Fit Model, with the help of ANN, for prediction of SO_2 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 SO_2 for t, t-1, t-2 and t-3 without metrological parameters. One such "Best Model" for each set is obtained and tabulated in Table 1 below,

Table 1: Models for SO_2 values without metrological parameters with each set of epoch

Network	Architecture	Epoch	Training Testing	MSE	RMSE	MAE	R
MLP	4-2-1	500	50:50	66.217	8.137	4.908	0.832
	4-2-1	1000	50:50	52.402	7.239	4.525	0.867
	4-2-1	**2000**	**50:50**	**48.637**	**6.974**	**4.488**	**0.865**
	4-2-2-1	3000	50:50	50.878	7.133	4.427	0.864
GFF	**4-2-1**	**500**	**50:50**	**49.169**	**7.012**	**4.603**	**0.867**
	4-2-2-1	**1000**	**50:50**	**46.901**	**6.848**	**4.531**	**0.869**
	4-2-1	2000	50:50	55.069	7.421	4.639	0.863
	4-2-1	**3000**	**50:50**	**45.380**	**6.736**	**4.384**	**0.879**
R	4-2-2-1	500	50:50	93.791	9.685	7.238	0.800
	4-2-1	1000	50:50	56.535	7.519	4.931	0.844
	4-2-2-1	2000	50:50	63.776	7.986	5.165	0.848
	4-2-2-1	3000	50:50	61.280	7.828	5.419	0.833
TLR	4-2-1-1	500	60:40	31.583	5.620	3.893	0.771
	4-2-1	1000	50:50	78.499	8.860	5.961	0.792
	4-2-2-1	2000	50:50	60.837	7.800	4.992	0.828
	4-2-1-1	3000	60:40	27.939	5.286	3.472	0.794

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 Model" has been obtained and the results are tabulated in table 2 below,

Table 2: Best Models for SO_2 values without metrological parameters for each set of epoch

Network	Architecture	Epoch	Training Testing	MSE	RMSE	MAE	R
GFF	4-2-1	500	50:50	49.169	7.012	4.603	0.867
GFF	4-2-2-1	1000	50:50	46.901	6.848	4.531	0.869
MLP	4-2-1	2000	50:50	48.637	6.974	4.488	0.865
GFF	**4-2-1**	**3000**	**50:50**	**45.380**	**6.736**	**4.384**	**0.879**

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 SO_2 for t, and all three metrological parameters. One such "Best Model" for each set is obtained and tabulated in Table 3 below,

Table 3: Models for SO_2 values with metrological parameters with each set of epoch

Network	Architecture	Epoch	Training Testing	MSE	RMSE	MAE	R
MLP	**4-2-1-1**	**500**	**50:50**	**71.313**	**8.445**	**5.377**	**0.821**
	4-2-2-1	1000	50:50	73.273	8.560	5.587	0.830
	4-2-2-1	**2000**	**50:50**	**56.415**	**7.511**	**5.186**	**0.857**
	4-2-1	3000	50:50	63.897	7.994	5.368	0.835
GFF	4-2-1	500	60:40	28.908	5.377	3.653	0.807
	4-2-1	**1000**	**50:50**	**59.411**	**7.708**	**5.174**	**0.846**
	4-2-1-1	2000	50:50	53.354	7.304	4.756	0.843
	4-2-1-1	**3000**	**50:50**	**54.747**	**7.399**	**4.669**	**0.837**
R	4-2-1-1	500	50:50	102.285	10.114	6.873	0.762
	4-2-1	1000	50:50	81.374	9.021	6.964	0.779

	4-2-2-1	2000	50:50	65.537	8.095	6.014	0.815
	4-2-1-1	3000	50:50	123.280	11.103	8.187	0.743
TLR	4-2-1	500	50:50	88.870	9.427	6.127	0.753
	4-2-1	3000	50:50	80.365	8.965	5.872	0.785
	4-2-2-1	1000	50:50	86.016	9.274	5.828	0.773
	4-2-1	2000	50:50	64.045	8.003	5.483	0.828

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 Model" has been obtained and the results are tabulated in table 4 below,

Table 4: "Best Models" for SO_2 values with metrological parameters with each set of epoch

Network	Architecture	Epoch	Training Testing	MSE	RMSE	MAE	R
MLP	4-2-1-1	500	50:50	71.313	8.445	5.377	0.821
GFF	4-2-1	1000	50:50	59.411	7.708	5.174	0.846
MLP	**4-2-2-1**	**2000**	**50:50**	**56.415**	**7.511**	**5.186**	**0.857**
GFF	4-2-1-1	3000	50:50	54.747	7.399	4.669	0.837

5. After analyzing table 2 and table 4, the "Best Model" has been obtained for SO_2 with and without metrological parameters and are tabulated in table 5 below,

Table 5: "Best Model" for SO_2 with and without metrological parameters

Network	Architecture	Epoch	Training Testing	MSE	RMSE	MAE	R	Remark
GFF	**4-2-1**	**3000**	**50:50**	**45.380**	**6.736**	**4.384**	**0.879**	Without metrological parameters
MLP	4-2-2-1	2000	50:50	56.415	7.511	5.186	0.857	With metrological parameters

It is observed from Table 5 above that for prediction of SO_2 concentration, one day ahead (t+1), highest efficiency with minimum difference between root mean square error (RMSE) and mean absolute error (MAE) is found for GFF Network with 4-2-1 architecture with epoch of 3000 having training to testing data in the ratio of 50:50 gives the efficiency of 87.9% is the "Best Fit Model" having normal values of SO_2 pollutant without metrological parameters as an input.

6. The "Best Model" obtained from table 5 has been analyzed by transforming the input values of SO_2 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. With the help of statistical tools, "Best Model" has been obtained and are tabulated in Table 6 below,

Table 6: Models for SO_2 values without metrological parameters with transformed values of SO_2

Transform Type	Architecture	Epoch	Training Testing	MSE	RMSE	MAE	R
Logarithmic	**4-2-1**	**3000**	**50:50**	**53.116**	**7.288**	**4.543**	**0.863**
Square Root	4-2-1	3000	50:50	69.688	8.348	4.938	0.822
Moving Average	4-2-1	3000	50:50	51.822	7.199	4.690	0.848

7. The "Best Model" obtained from table 5 has been analyzed by transforming the input values of SO_2 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 SO_2 values with transformed values of SO_2 and normal values of metrological parameters

Transform	Architecture	Epoch	Training	MSE	RMSE	MAE	R

Type			Testing				
Logarithmic	**4-2-1**	**3000**	**50:50**	**64.835**	**8.052**	**5.117**	**0.825**
Square Root	4-2-1	3000	50:50	119.407	10.927	6.207	0.604
Moving Average	4-2-1	3000	50:50	66.550	8.158	5.619	0.816

8. The "Best Model" obtained from table 5 has been analyzed by transforming the input values of SO_2 and meteorological parameters to logarithmic, square root and moving average to normalize input data and are tabulated in Table 8 below,

Table 8: Models for SO_2 values with transformed values of SO_2 and metrological parameters

Transform Type	Architecture	Epoch	Training Testing	MSE	RMSE	MAE	R
Logarithmic	4-2-1	3000	50:50	65.687	8.105	5.122	0.823
Square Root	4-2-1	3000	50:50	106.877	10.338	5.927	0.662
Moving Average	**4-2-1**	**3000**	**50:50**	**61.339**	**7.832**	**5.409**	**0.834**

9. After analyzing table 6, 7 and table 8, the "Best Model" has been obtained for SO_2 with and without metrological parameters by transforming the input values and tabulated in Table 9 below,

Table 9: "Best Fit Model" for SO_2 with and without metrological parameters by transforming the input values

Network	Architecture	Epoch	Training Testing	MSE	RMSE	MAE	R	Remark
GFF	**4-2-1**	**3000**	**50:50**	**53.116**	**7.288**	**4.543**	**0.863**	**Logarithmic of SO_2 without metrological**
GFF	4-2-1	3000	50:50	64.835	8.052	5.117	0.825	Logarithmic of SO_2 with normal values of meteorological parameters
GFF	4-2-1	3000	50:50	61.339	7.832	5.409	0.834	Moving average of SO_2 and meteorological parameters

10. After analyzing table 5and table 9, the "Best Model" has been obtained for SO_2 with and without metrological parameters by transforming the input values and tabulated in Table 10 below,

Table 10: "Best Fit Model" for SO_2 with and without metrological parameters as well as transform values of all input values

Network	Architecture	Training Testing	Epoch	MSE	RMSE	MAE	R	Remark
GFF	**4-2-1**	**50-50**	**3000**	**45.380**	**6.736**	**4.384**	**0.879**	**SO_2 without metrological parameters**
GFF	4-2-1	50-50	3000	53.116	7.288	4.543	0.863	Logarithmic of SO_2 without metrological

It is observed from Table 10 above that for prediction of SO_2 concentration, one day ahead (t+1), highest efficiency with minimum root mean square error (RMSE) and mean absolute error (MAE) is found for GFF network with 4-2-1 architecture with epoch of 3000 having training to testing data in the ratio of 50:50 gives efficiency of 87.9% is the "Best Model" having normal values of SO_2 pollutant without metrological parameters as an input.

11. On getting the "Best Fit Model" for SO_2 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 Fit Model" of GFFN for transformed values of SO_2 without metrological parameters as input values for different architecture

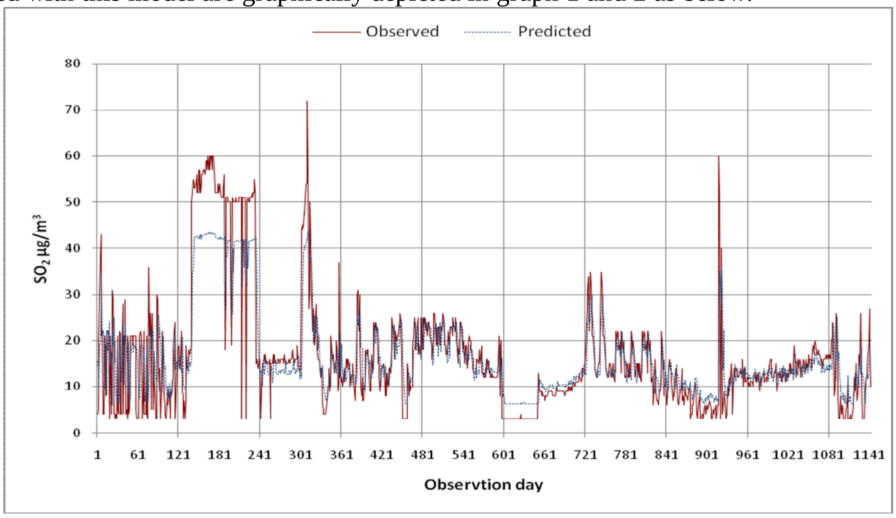
Network	Architecture	Epoch	Training Testing	MSE	RMSE	MAE	R
GFF	**4-2-1**	**3000**	**50:50**	**45.380**	**6.736**	**4.384**	**0.879**
GFF	5-2-1	3000	50:50	63.591	7.974	4.901	0.834
GFF	6-21	3000	50:50	46.001	6.782	4.417	0.874
GFF	7-2-1	3000	50:50	52.401	7.239	4.578	0.844
GFF	8-2-1	3000	50:50	54.216	7.363	4.564	0.843
GFF	9-2-1	3000	50:50	50.495	7.106	4.641	0.856
GFF	10-2-1	3000	50:50	58.339	7.638	5.120	0.822
GFF	11-2-1	3000	50:500	54.202	7.362	4.936	0.842
GFF	12-2-1	3000	50:50	55.696	7.463	4.775	0.837
GFF	13-2-1	3000	50:50	55.603	7.457	4.746	0.836

12. After analyzing table 11, the "Best Fit Model" has been obtained for SO_2 with and without metrological parameters by transforming the input values and tabulated in Table 12 below,

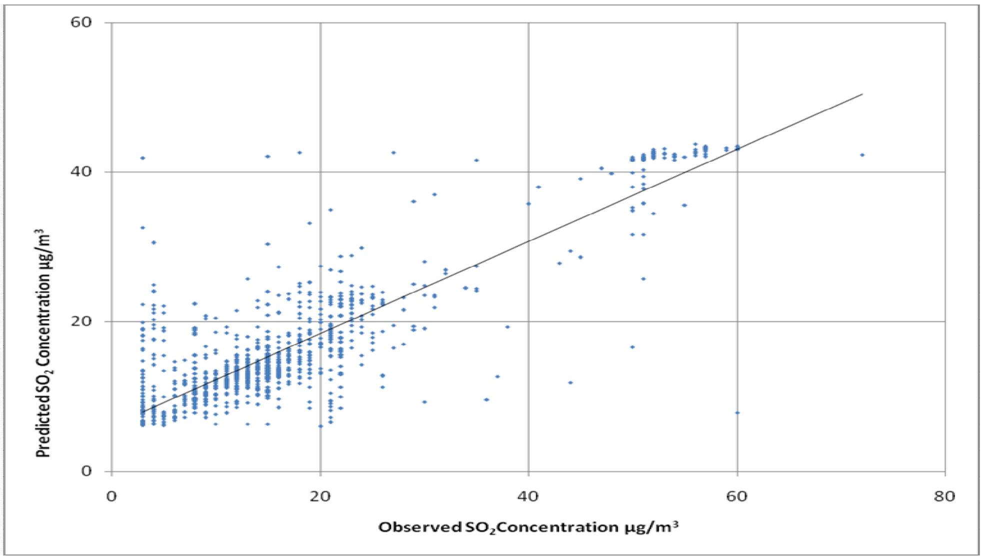
Table 12: "Best Fit Model" for SO_2 with and without metrological parameters

Network	Architecture	Training Testing	Epoch	MSE	RMSE	MAE	R
GFF	**4-2-1**	**3000**	**50:50**	**45.380**	**6.736**	**4.384**	**0.879**

It is observed from Table 10 above that the "Most Efficient Model" for prediction of SO_2 concentration, one day ahead (t+1), highest efficiency is found for GFF network with 4-2-1 architecture with epoch of 3000 having training to testing data in the ratio of 50:50 gives efficiency of 87.9% is the "Best Fit Model" having normal input values of observed SO_2 for t, t-1, t-2 and t-3 without metrological parameters as an input.
The results obtained with this model are graphically depicted in graph 1 and 2 as below.



Graph 1: Time series plot of observed and predicted SO2 values for one day ahead (t+1) with "Best Fit Model"



Graph 2: Comparison of observed and predicted SO2 values plot for one day ahead (t+1) with "Best Fit Model"

1.4 Observations

It is observe from the above table 10, 11 and 12 that the metrological parameters and transform values of input do not affect on the prediction value of SO_2. For prediction of SO_2 concentration, one day ahead (t+1), by using input values of observed SO_2 for t, t-1, t-2 and t-3 gives the optimal results as shown in table 12.

1.5 Conclusion

For prediction of SO_2 concentration, one day ahead (t+1), highest efficiency with minimum root mean square error (RMSE) and mean absolute error (MAE) is found for GFF network with 4-2-1 architecture with epoch of 3000 having training to testing data in the ratio of 50:50 gives efficiency of 87.9% is the "Best Fit Model" having normal values of SO_2 pollutant without metrological parameters as an input.

1.6 Future Scope

Since the prediction of SO_2 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 best fit model can be found out.

Acknowledgement

Authors are very much thankful to Principal, WIT, 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., 1996, "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.

[2.] Bartzokas A., Kassomenos P., Petrakis M., Celessides C., 2004. The effect of meteorological and pollution parameters on the frequency of hospital admissions for cardiovascular and respiratory problems in Athens, *Indoor and Built Environment*, vol13, p.g.271-275.

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

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

[5.] Singh, G., Pal, A. K., Tiwari, A., 2007, "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.

[6.] Tecer, L. H., 2007, "Prediction of SO_2 and PM Concentrations in a Coastal Mining Area Zonguldak, Turkey Using an Artificial Neural Network", *Environmental Studies*, vol 16(4), p.g. 633-638.

[7.] Timonen, K.L., Pekkanen, J., Tiittanen, P., Salonen, R.O., 2002, "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.
[8.] http://mpcb.mah.nic.in/envtdata/envtair.php retrieved on 12 Oct 2017
[9.] http://www.imd.gov.in retrieved on 24 Apr 2018.

AUTHORS

Capt. (Dr.) N.P. Sonaje received the M.E. degree from Gujarat University, Ahmadabad and Ph.D. degree in Civil Engineering from Shivaji University, Kolhapur. He was Captain in Indian Army up to 2001. He worked as a Registrar, Solapur University, Solapur and working as Deputy Registrar in Shivaji, University, Kolhapur. He has vast experience of working on various administrative posts. He is member of various professional organizations as well as NAAC Peer Team.

Prof. S. J. Mane received the B.E. and M.E. degrees in Civil Engineering from Shivaji University, Kolhapur. He is working as Assistant Professor in Civil Engineering at D. Y. Patil College of Engineering, Akurdi, Pune. He is member of various technical bodies.