28th WEDC Conference Kolkata (Calcutta), India, 2002

SUSTAINABLE ENVIRONMENTAL SANITATION AND WATER SERVICES

Neuro-Ensemble for Air Quality Prediction

Sameer Sharma, A. K. Dikshit and S. V. Barai, India

Introduction

Air pollutants exert a wide range of impacts on biological, physical, and economic systems. Their effects on human health are of particular concern. The decrease in respiratory efficiency and impaired capability to transport oxygen through the blood caused by a high concentration of air pollutants may be hazardous to those having pre-existing respiratory and coronary artery disease (Rao and Rao, 2000). Consequently, it has become a vital task to accurately keep track of the variation of ambient air pollution levels in urban areas.

Natural phenomena are mostly a time series with some degree of randomness. Pollutants in the atmosphere may disperse or concentrate during varied time periods. Previous studies (Giorgio and Piero, 1996) have indicated that the date of ambient air quality are stochastic time series, thereby making it possible to make a short-term forecast on the basis of historical data. Though models may be imperfect, they are the best tool for use in all aspect of air quality planning where prediction is a major component such as for emission control (Melas et al., 2000), accidental release of pollutant, land-use planning, traffic planning (Hadjiiski and Hopke, 2000), planning of measurement programs (Rao and Rao, 2000) analyses of measurements/trends and episode forecasting (melas et al., 2000).

The neural networks (Principe and Kuo, 1995) have emerged out to be more flexible, less assumption dependent and adaptive methodology in environment related areas such as rainfall runoff modeling, stream flow forecasting (Thirumalaiah and Deo, 1998), ground water modeling water management policy, precipitation forecasting, hydrologic (AXCE, 2000a, 200b), remote sensing and GIS related activities, real time control of water treatment plants, water quality and air quality management (Boznar et al., 1993), adsorbent beds design (Basheer and Najjar, 1996), and hazardous waste management.

The present study investigates the advantage of ensemble of neural networks (Haykin, 2000) for forecasting the air pollution. The aim is to find accurate air quality predictors, which can work with low number of data sets and should be robust enough to handle data with noise and errors. The objectives of the study are as follows :

- To implement various available variations of neural network models for predicting air quality.
- To collect suitable data sets for multiple air quality parameters - containing daily average pollutant concentrations at a specific location.
- To conduct exhaustive simulations using developed models with hourly data.
- To perform comparative study to identify suitable air quality prediction model(s) for hourly (short-term) data.
- To develop ensemble model for the developed air quality prediction models.

Why Ensemble of Neural Networks?

Whether a model is neural networks model (Haykin, 2000) selected systematically or not, it is clear that the remaining neural networks models that were developed and discarded contain potentially valuable information about the problem. It has become clear that prediction models could also be improved by the combination of multiple models into one, an approach called ensemble modeling. Many engineering applications have been demonstrated using ensembles (Barai and Reich, 1999). There are two kinds of approaches for ensemble modeling: *Opportunistic and Principled.* The opportunistic approach emerges from the usual iterative data modeling process. During this process, various models are studied, the problem is gradually better understood. Finally we select the best model and discard others. Instead of discarding these models ensemble modeling combines them into one as shown in Figure 1. If the models are quite different (i.e. diverse) and reasonably good, the ensemble will improve upon the best of them. We demonstrate this approach in the present context. The principal approach seeks to generate systematically a set of as accurate as possible and diverse models from which a single model is composed. The model of this kind is described elsewhere (Barai and Reich, 1999).

Implementation of Various Neural Networks Models

For the present study following various neural network models have been implemented in MATLAB software (Math works, 2000) using Neural Networks Tool box (Demuth and Beale, 1992).

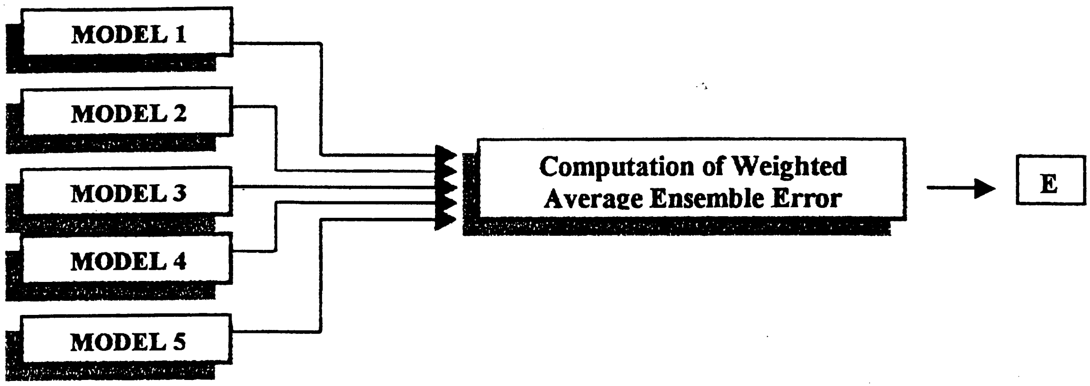

Figure -1 : Opportunistic Ensemble Model

- Recurrent Network Model (RNM)
- Change Point Detection Model (CPDM)
- Sequential Network Construction Model (SNCM)
- Self Organizing Feature Maps Model (SOFM)
- Combined NARMA and Self Organizing Feature
- Maps Model (NARMA_SOFM)
- Moving Window Model (MWM)
  - MWM2 – With two values of the past input data
  - MWM3 – With three values of the past input data
  - MWM4 – With four values of the past input data
  - MWM5 – With five values of the past input data

The details about the models and its performance capability can be found elsewhere (Sharma, 2002). As discussed in previous section, we ensemble all the models to forecast air-quality of problem domain.

Data Collection and Properties

The data for three parameters namely RPMA (Respiratory Particulate Matter Average), SO_2 (sulphur dioxide) and NO_2 (nitrogen dioxide) is collected for Delhi State at nine locations. These data are daily average concentrations for last two years from 3/7/2000 to 20/8/2001. This data set has been downloaded from Tata Energy Research Institute web site <www.teri.in>. However, only the data for Ashram Chowk has been used for carrying out simulation studies outlined elsewhere (Sharma, 2002). Statistical properties of the data is given in Table 1.

Table 1. Statistical properties of dataset.

Parameter Property	RPMA	SO_2	NO_2
Mean	172.83	9.587	77.18
Std. Dev	119.88	3.7	31.59
Median	141	9	72

Results and Discussion

Out of 110 data points of the time series, initial 80 points were used for training the models discussed in section 3.0. Remaining 30 points were used to predict air quality parameters. Network parameters and data are not reported here due to space restriction, however can be referred elsewhere (Sharma, 2002). The general observations made during the study are as follows:

- The models studied for this case (section 3), in general could predict with modest accuracy (Table 2). However, among all the models implemented, NARAMA_SOM based model has performed extremely well in comparison to other models. This performance could be attributed to an integrated approach of statistical model with neural networks model. Here SOFM performance decreases due to the formation of many clusters. The performances of Moving Window Model increase as the numbers of past input data are increased. This may be attributed to the fact that network learns the history much more.
- The case study demonstrated an example of a daily average emission data prediction using various neural networks model for a reasonable size dataset. Models in general have performed reasonably well even though data was chaotic by nature.

Table 2 : Neural networks models performance

Air-quality Parameter	RPMA Emissions	SO_2 Emissions	NO_2 Emissions
RNM	56.76	48.63	43.5
CPDM	45.36	41.83	38.9
SNCM	33.45	37.79	35.87
NARMA_BPX	25.29	10.92	18.56
SOFM	25.6	30.73	28.94
NARMA_SOFM	**21.66**	**8.82**	**17.98**
MWM2	25.71	25.51	22.93
MWM3	32.38	28.87	28.61
MWM4	36.51	21.97	30.94
MWM5	43.52	31.59	32.57

To improve the model performance, we studied further and developed weighted average ensemble model combining the discussed models (section 3). The models were combined with equal weights for predicting air quality parameters (Figure 1). Typical results for future predictions of RPMA and SO_2 are shown in Figure 2 and 3 respectively. From the figures, it is clear that ensemble models predict reasonably close to the real air quality data for future trend.

Summary

In this paper, the study was carried out on air quality forecasting using various neural network models. The study was focused at preliminary investigation of single variable based time series prediction. The investigation was carried out for short-term air quality data set, which were daily average data. The models studied in this study were easily implemented and could deliver fast prediction at run time, unlike other modeling techniques. The models can very well deal with input noise and uncertainty. To improve the model performance, ensemble modeling approach was explored and found to be suitable for the present application. The encouraging results clearly indicate that the neural networks can be used as a tool for real time air quality management and prediction.

Reference

1. ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000a), "Artificial neural networks in hydrology. I: preliminary concepts", *J. Hydrologic Engineering,* ASCE, 5(2), 115-123.
2. ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000b), "Artificial neural networks in hydrology. II: preliminary concepts", *J. Hydrologic Engineering,* ASCE, 5(2), 124-137.
3. Barai, S. V. Reich, Y. (1999), "Ensemble modeling or selecting the best model: Many could be better than one", *Artificial Intelligence for Engineering Design, Analysis and Manufacturing,* 13,377-386.
4. Basheer, A., and Najjar, M. (1996), "Predicting dynamic response of adsorption columns with Neural Nets", *J. Computing in Civil Engineering,* ASCF, 10(1), 211-220.
5. Boznar, M., Lesjak, M., and Mlakar, P (1993), "A neural network-based method for short term prediction of ambient SO_2 concentrations in highly polluted industrial areas of complex terrain", *Atmos. Environ.,* 27B, 221-230.
6. Demuth, H., and Beale, M. (1992), *"Neural Networks Toolbox Manual",* Math Works Inc.
7. Giorgio, F., and Piero, M. (1996), "Mathematical

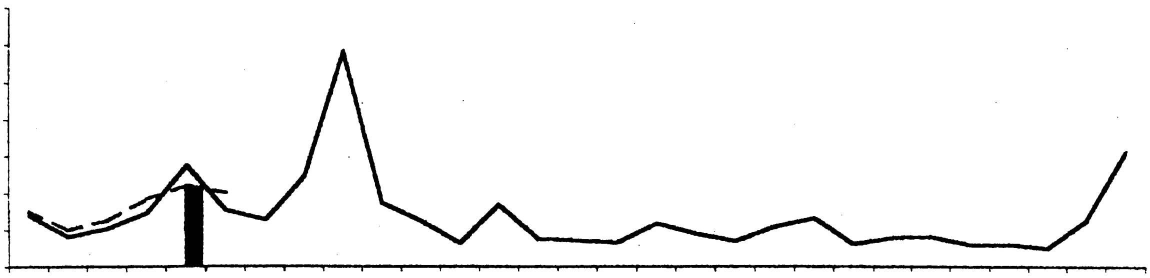

Figure -2 : Emsemble model performance of RPMA prediction

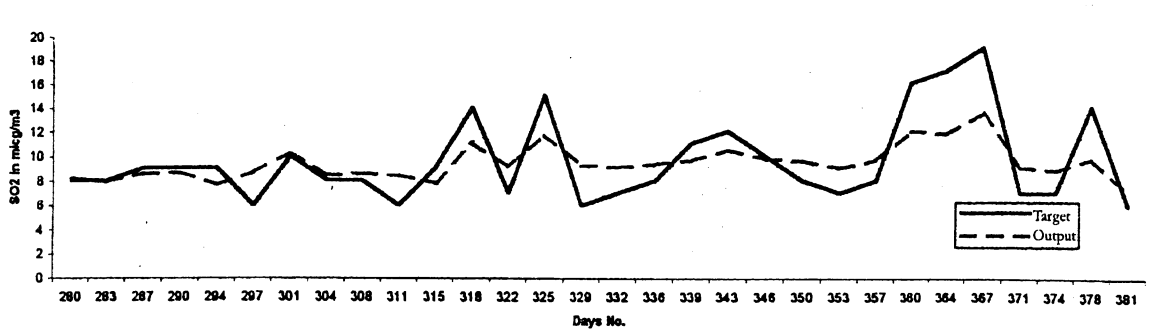

Figure -3 : Ensemble model performance of SO_2 prediction

models for planning and controlling air quality", *Proceedings of IIASA Workshop*, 17.

8. Hadjiiski, L., and Hopke, P. (2000), "Application of artificial neural networks to modeling and prediction of ambient ozone concentrations", *J., Air and Waste Manage. Assoc.*, 50, 894-901.
9. Haykin, S., (2000), "Neural Networks", Second Edition, Addison Wesely Longman.
10. Math Works (2000), *"MATLAB User's Guide"*, Math Works Inc.
11. Melas, D., Kioutsioukis, I., and Ziomas, I.C. (2000), "Neural network models for predicting peak photochemical pollutant levels", *J. Air and Waste Manage. Assoc.*, 50, 495-501.
12. Principe, J., and Kuo, J. M. (1995), "Dynamic modeling of chaotic time series with neural networks", *Proc. of Neural Infor. Proc. Sys.*, 7.
13. Rao, M. N., and Rao, H. V. (2000), *"Air Pollution"*, Tata-McGraw-Hill, New Delhi.
14. Sharma, S. (2000), *"Neuro Air Quality Predictors"*, M. Tech. Thesis, IIT Khargpur.
15. Thirumalaiah, K., and Deo, M. C. (1998), "River stage forecasting using artificial neural networks", Journal of Hydrologic Engineering, ASCE, 3(1), 26-32.
16. Thirumalaiah, K., and Deo, M. C. (2000), "Hydrological forecasting using neural networks", *Journal of Hydrologic Engineering, ASCE*, 5(2). 180-189.

Sameer Sharma, Former Post-graduate Student

A. K. Dikshit, Associate Professor

S. V. Barai, Assistant Professor, Corresponding Author, Department of Civil Engineering, Indian Institute of Technology, Kharagpur 721 302, India