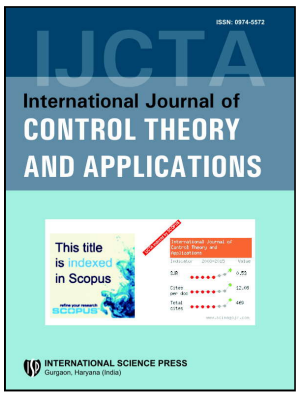

International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 10 • Number 19 • 2017

Prediction of Pollutant Oxide of Nitrogen Component in Delhi City using Artificial Neural Network

Monika Singh[1], S.P. Mahapatra[2], Sudhir Nigam[3], Kuhu Gupta[4] and Mridu Sahu[5]

[1&2] Department of Chemistry, National Institute of Technology, Raipur, Chhattisgarh, India,
Emails: moniica.singh21@gmail.com, spmahapatra.chy@nitrr.ac.in
[3] Department of Civil, Laxmi Narain College of Technology and Science, Bhopal, Madhya Pradesh, India,
Email: nigam.sudhir@hotmail.com
[4&5] Department of Information Technology, National Institute of Technology, Raipur, Chhattisgarh, India,
Emails: psingh.cs@nitrr.ac.in, kuhu.gupta08@gmail.com

Abstract: India is a developing economy which is facing a dangerous problem of air pollution. The main causes of urban air pollution are emissions from industrial and domestic fossil fuels combustion and vehicular emissions. Among all the pollutants released, .Oxide of Nitrogen is one of the dangerous components of polluted air. Apart from various health problems such as lung dysfunction, asthma, bronchitis and irritation of the lungs, Oxide of Nitrogen causes environmental hazards such as damage and destruction of watercourses, vegetation, construction material and eutrophication. Therefore, it is imperative to estimate the concentration of the harmful pollutants in the air. Different analysis and computational methods are developed to predict the concentration of pollutants in order to improve the quality of air. In the proposed work, an artificial neural network have been developed successfully and applied to atmospheric pollution modeling and in air quality determination.

Keywords: air pollution, Oxide of Nitrogen, concentration, prediction, Artificial Neural Network

Corresponding author: Dr .S.P Mahaptra

Email spmahapatra.chy@nitrr.ac.in

1. INTRODUCTION

In India, growth of industries as well as enormous use of vehicles, the concentration of unwanted substances in our environment is increasing day by day and exceeding the thresholds of pollution. The major causes of air pollution are migration of peoples, large number of industries, and high birthrate over death rate, vehicle exhaust and lack of awareness about the after effects of air pollution [1]. Air pollution is due to heavy transportation, untreated industrial wastes released free in environment, forest fires, burning of fuels, and toxic gases from diesel generators. Air pollutants such as Carbon Monoxide (CO) and Hydrocarbons (HC) and Oxide of Nitrogen

(NO), total suspended solids (TSP) non-methane volatile organic compounds Sulphur di oxide are released into the atmosphere [2]. The motion of an air pollutant that take less than the average time for transport is termed as dispersion while the mixing of air pollutant and is termed as dilution. The meteorological conditions such as wind direction, wind speed turbulence, and atmospheric stability influence dispersion and dilution of an air pollutant [3]. Chemical reaction between the pollutants also depends on ambient weather conditions. The dispersion and dilution of air pollutants causes its transmission thus, reducing the air quality. These pollutants are oxidized in the atmosphere to form radicals, like, SO42- and NO3- which reacts with ammonium (NH4+) to form aerosols. The aerosols along with primary pollutants are deposited in the form of wet and dry deposition. Deposition is the downward movement of pollutants in the atmosphere by processes such as, precipitation, scavenging, and sedimentation. In this way, pollutants are released into the ground surface causing more damage to flora and fauna. This entire process as shown in Figure 1. has backed many air pollutants in creating a paucity of fresh air to breathe, healthy food to eat and pure water to drink. The proposed work, emphasizes on atmospheric pollutant know as Oxides of Nitrogen (NOx). Oxide of Nitrogen are a family of poisonous, irritant and highly reactive gases. Having compounds and derivatives in the family of Oxide of Nitrogen, includes nitrogen di oxide, nitric acid, nitrous oxide, nitrates, and nitric oxide. During combustion when fuel is burnt at high temperatures, the reaction of nitrogen and oxygen in the air produce Oxide of Nitrogen (NOx) [4]. About 50% of NOx is emitted by vehicular emissions, 40% from electric power plant boilers and additional emissions are added by anthropogenic sources such as incinerators, gas turbines and diesel engines as shown in Figure 2. Lightning, forest fires, grass fires, trees, bushes, grasses, and yeasts are some biogenic sources of Oxide of Nitrogen. NOx is not only an independent harmful pollutant, but also reacts in atmosphere to form ozone and acid rain [5]. Nitrogen di oxide with volatile organic compounds involves in photochemical induced catalytic production of tropospheric ozone that leads to the formation of summer smog and increase in level of ozone. The steps of formation of smog is discussed in the following reactions-

$$N_2 + O_2 \rightarrow 2NO$$

$$2NO + O_2 \rightarrow 2NO_2$$

$$NO_2 \rightarrow NO + (O)$$

$$(O) + O_2 \rightarrow O_3$$

$$NO + O_3 \rightarrow NO_2 + O_2$$

$$NO_2 + O_2 + \text{unburnt hydrocarbons} \rightarrow CH_3COOONA \text{ (Peroxy Acetyl Nitrate)}$$

Acid rain is one of the other important effects of NOx. Acid rain can cause damage of materials building and historic monuments and change in water nature to acidic, damage of lungs tissue, and even premature death [6]. One of major member of NOx, Oxide of Nitrogen which is unpleasant in odor and cause eye irritation and breathing trouble, asthma and bronchitis.NO2 reacts with hemoglobin and leads to a condition called Methemoglobinemia. This is also known as Blue baby syndrome. It reduces the oxygen carrying capacity of blood. The increased concentration causes respiratory problems such as inflammation the lining lungs and reduces immunity to lungs infections, this cause problem like wheezing, coughing, colds and flu and bronchitis [7]. All of these facts indicate an obvious need to reduce NOx emissions. For successful remedies from harmful pollutants needs analysis at the base and identification of pollution sources, conveying ability of pollutants in the atmosphere, anticipation and its controlling methodologies, and economic impact on the particular area. After the consequences obtained from such an analysis, the government and pollution control board and authorities should take necessary steps to curb the hazardous air pollution of that particular area. Different analysis and computational methods are developed to know and predict the concentration of pollutants respectively. Analysis of regional air pollution problems and their solutions is effectively performed through the application of computer modeling techniques. The aim of the paper is to forecast the values of nitrogen oxide in Delhi using Artificial Neural Network.

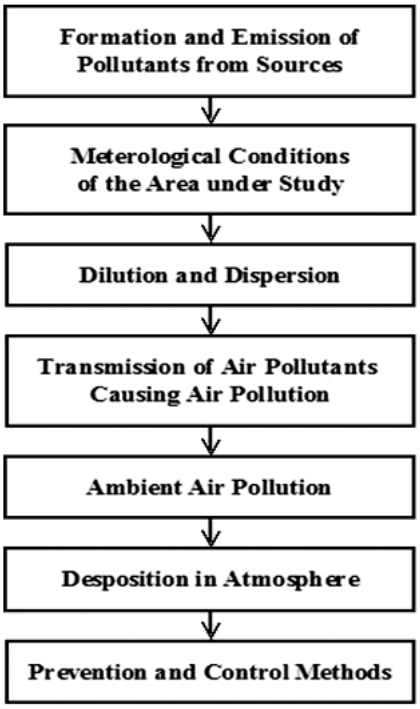

Figure 1: Block Diagram of Proposed Methodology

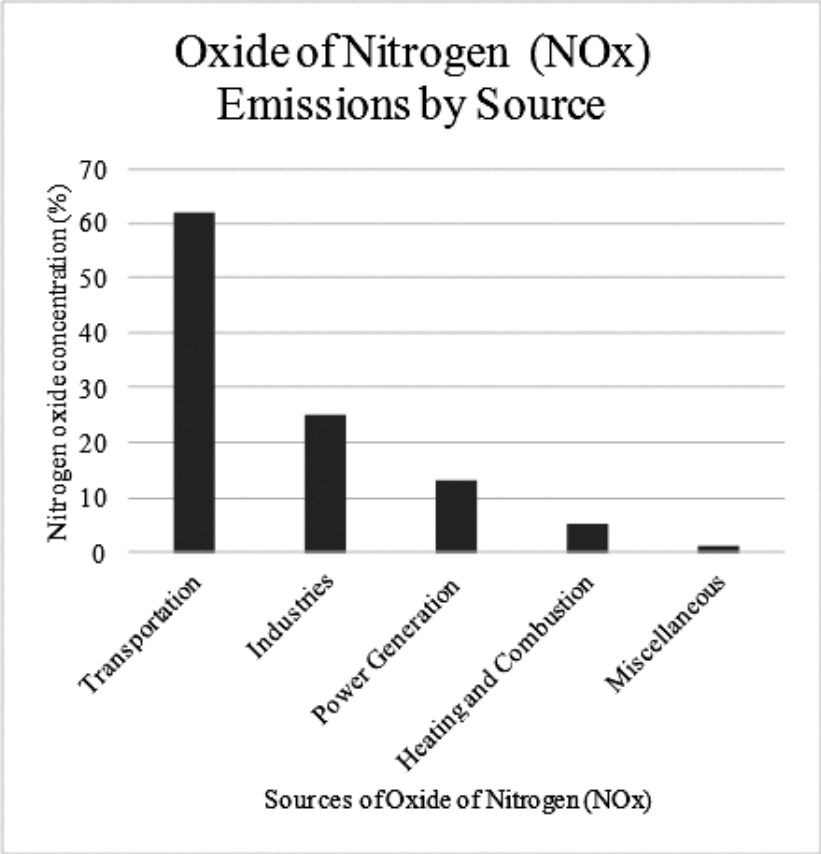

Figure 2: Sources of Oxide of Nitrogen (NOx)

2. POLLUTION IN DELHI- A MAJOR CONCERN

Every year air pollution kills 1.5 million people. On conducting a survey of 1600 cities all across the globe, Delhi, the capital city of India is announced to be at top of the charts of pollution [8]. Similar studies on mortality from Delhi stated that increase in natural cause mortality is due to increased concentration of air pollutants. Rapid rate of industrialization and vehicular emissions has aggravated the chronic respiratory diseases among people of Delhi [9]. These urban challenges has contributed in exceeding the permissible limits of harmful pollutants, such as, Oxide of Nitrogen, Sulphur dioxide, ozone and particulate matter in Delhi [10]. Consistent efforts are being made by the environmentalists by emphasizing the damage caused to human health by air pollution. Several steps has been taken to improve the air quality in Delhi through improvement programs, control instruments and awareness programs over the past ten years but all in vain [11]. Scientific and evolutionary methodologies must be adopted to control air pollution in Delhi.

3.1. Air Quality Index (AQI)

Air Quality Index (AQI) is estimated through a mathematical expression to give a numerical quantity [12]. The value of the index ranges from 0-500 and it is subdivided into six range to which six categories of air quality is represented as shown in Table 1 and Table 2 [13]. The National Air Monitoring Programme, a flagship programme under which the Central Pollution Control Board (CPCB; http:// www.cpcb.delhi.nic.in) has been reporting ambient air quality in Delhi. This monitoring programme has generated large amount of data on concentration of air pollutants. The volume of this data has been an obstacle in conveying the status of air quality, therefore Air

Table 1
Level of Health Concern: (averaged over 24 hours)
(Source: Environement Protection Agency)

Index Values	*Levels of Health Concern*	*Health Concern Statements*
0-50	Good	None
51 - 100*	Moderate	None
101 – 150	Unhealthy for Sensitive Groups	People with asthma should consider limiting Outdoor exertion.
151 – 200	Unhealthy	Children, asthmatics ,and people with heart or lung disease should limit outdoor exertion
201 – 300	Very Unhealthy	Children, asthmatics, and people with heart or lung disease should avoid outdoor exertion; everyone else should limit outdoor exertion
301 - 500	Hazardous	Children, asthmatics, and people with heart or lung disease should remain indoors; everyone else should avoid outdoor exertion

Table 2
Break point concentration for NOx

Mapping with recently proposed IND: AQI (CPCB) for NOX	
Category	Break point conc. for NO2(ug/m3)-24 Hrs
AQI	Conc.
Good	40
Satisfactory	80
Moderately polluted	180
Poor	280
Very poor	400

Quality Index (AQI) is introduced [14]. Based on compact evaluation of the concentration five major air pollutants, that are, ground-level ozone, particulate matter, carbon monoxide, sulfur dioxide, and nitrogen dioxide There is greater danger to public health for higher values of AQI. For India, Environment Protection Authority (EPA) has set AQI value of 100 as national air quality standard for the pollutant. AQI values 100 are regarded as detrimental to people suffering from respiratory diseases while values below 100 are satisfactory.

4. MATERIALS AND TECHNIQUES

4.1. Data

Data is collected from capital of India, Delhi. The metropolitan city, being the busiest place of India has the combination of both commercial and industrial hub. Over the past 10 years, the air quality in Delhi has undergone several changes. CPCB (Central Pollution Control Board) has an automatic monitoring station at ITO SQUARE in New Delhi. At this station respirable suspended particles, carbon monoxide, ozone, Sulphur di- oxide, nitrogen dioxide and suspended particulate matter are being monitored and information is weekly updated. The data used for this experiment is collected from CPCB's website.

4.2. Artificial Neural Network (ANN)

An Artificial Neural Network (ANN) is a set of interconnected processing elements known as neurons that have weights associated with them [15]. The biological neural network is an inspiration for the creation of artificial neural network. An ANN is categorized into three layers. These three are the input layer for taking the inputs, the hidden layer for passing the weighted outputs and the output layer [16]. On increasing the number of neurons in the hidden layer, the neural network achieves more flexibility and accuracy in processing at the cost of increased complexity. Artificial Neural Network has aided the scientific community from many years by performing a series of tasks, like pattern recognition, classification and time series forecasting because of several types of artificial neural network. There are classified into two broad categories –

4.2.1. Feed-Forward Artificial Neural Network

A feed forward network is a non-recurrent neural network based on one rule: information must flow forward from input layer through the hidden layer to output layer with no cycles or loops [17]. There can be any number of hidden layers or connections between neurons. The input data is given to the input layer for performing computation based on weighted sum of inputs. These new values are input to the next layer. The process continues until it reaches the output layer. The output is determined by a threshold transfer function and there are no constraints on choosing the type of transfer function in feed forward network. This type of network has its applications in assessing and identifying input patterns and data mining.

4.2.2. Feed-Back Artificial Neural Network

A feed-back neural network is a recurrent network in which the information is fed back into itself. The information flows in both directions with no restriction of loops which enables it to possess a dynamic temporal behavior [18]. In order to achieve the state of equilibrium, it changes continuously to give best evolved results. These networks are applied for associative memories and optimization problems.

4.3. Neural Network as a Predictor

Artificial neural network is type of computational process that is inspired by working of human brain and computer. In the past years, neural networks have been used for carrying out cognitive tasks such as, identifying a familiar face, natural language processing and handwritten text recognition. Thus, ANN has been successful in aiding empirical modelling for finding solution to a broad spectrum of problems [19]. The reason for its popularity

is the capability of ANN to find hidden relationships even if the problem is poorly understood and completely unknown [20]. However, recent advancements in artificial neural network has been used as a solution in fields such as finance, ecological science, business, industry and science. Meanwhile, Artificial Neural Network emerged as a powerful tool for forecasting of time series data. The intricate topography and heat phenomena of urban air pollution make the use of deterministic model inconvenient but few distinctive features of ANN's make them appealing alternatives for prediction purposes [21]. Contrary to traditional models, neural network models are data driven self-adaptive procedures which recognize the inherent functional relationships between input and output based on previous examples. This approach is suited for problems with enough data but whose solutions require knowledge. Artificial neural networks can even generalize since they correctly deduce about the unspecified data from the sample which contains noisy information. ANN are even considered as nonlinear models because it does not require prior knowledge about relationships in the dataset. While dealing with non-linearity of data they can even behave as functional approximators by approximating continuous function to any desired accuracy [22]. In addition to all this advantages of using neural networks for prediction one of the most compelling advantages of neural network prediction is its accuracy which is higher than any other models.

4.4. Package description used for Neural Network Modelling-

In time series analysis wherever data is measured or recorded, problems concerning missing values occur. Due to unavoidable reasons, data is often lost, not measured or considered unusable. In large datasets, there are frequent occurrences of missing values that could create problems in data processing and analysis. Therefore, missing values must be handled reasonably [23]. R is a widely used statistical programming language which uses the basic concepts of statistics and mathematics to solve problems of large datasets. In statistics, the process of solving the well-known problem of missing values is known as imputation and R offers a lot of packages that has imputation functions. In R, missing values are represented as NA which is a reserved word. For working with missing values in univariate time series, R has introduced "zoo" package. This package has several functions, such as, na.aggregate () which replaces missing values with aggregated values, na.approx() replaces with interpolated values, na.locf() replaces the missing value with most recent non-NA prior to it and na.StructTS() for filling NA values using seasonal Kalman filter [24]. R has outperformed several statistical languages in not only manipulating and transforming present data or past data but also predicts future values. In the proposed work, artificial neural network is employed for predicting the concentration of Oxide of Nitrogen. "nnet" package in R aids in building feed-forward neural networks with single hidden layer [25].

4.5. Classification Algorithms

A Classifier is an algorithm that executes classification. It maps input values to a category using a mathematical function. Classification is the paradigm of finding the common features in a large dataset and categories into different classes according to a classification model [26]. The large dataset is divided into training and testing set. The training set is the representative of the larger dataset with similar set of attributes and a class variable [27]. In order to perform classification, the training data is analyzed to form correct description for each class. These descriptions are the key to categorize future test data to build classification rules for each class [28]. These algorithms are basically used to predict a category for a binomial or multinomial classification. There are numerous classification algorithms based on decision trees, logistic regression, Naïve Bayes and support vector machines.

4.6. Algorithm

Algorithm: Oxide of Nitrogen (NOx) concentration prediction using neural network

Input: D

D = Extracted Oxide of Nitrogen (NOx) Component from Air Pollution database

Output: Prediction of Oxide of Nitrogen (NOx)

Begin

Step 1: Data Pre-processing [Handling of missing values using "zoo" package of R and discretization on the scale of 1 to 7.]

Step 2: Division of Dataset into Training set (60%) and Testing set (40%).

Step 3: Applying neural network model for prediction[Building a feed forward network using "nnet" package of R]

Step 4: Prediction of Oxide of Nitrogen (NOx) concentration for the next 20 days.

Step 5: Identifying the existence of pollution based on binary classification.

Step6: Applying Classification Algorithms for pollution identification.

Step 7: Evaluate performance of classifier

End

The dataset taken from CPCB (Central Pollution Control Board) consists of the concentrations of pollutants for 365 days. The pollutants are Sulphur oxide, PM 2.5, PM 10, carbon monoxide, Oxide of Nitrogen and Ozone. The pollutant to be predicted, i.e., NOx is assumed to be X. All the values are discretized on the scale of 1 to 7 while the missing values are replaced with mean of that particular pollutant values using aggregate function of zoo package. The dataset is divide into Training dataset which is given from Day 1 to Day 300 and testing dataset is from Day 301 to Day 320. In the proposed work, artificial neural network is adapted to predict the concentration of NOx. The nnet package of R language is used to train the model. The neural network used has three layers. In the input layer, all other pollutants are provided as inputs while in the hidden layer (single hidden layer), there will be neurons which we can vary according to the requirement. In this work, two neurons are taken in the hidden layer. No. of maximum iteration can also be varied. The threshold is set to 3, that is, if the output of the hidden layer is greater than 3, the model predicts the presence of pollution while for values less than 3 there is no pollution. The neural network forms a binary classifier in the input layer by setting the threshold for prediction. To measure the accuracy of the model, numerous classifiers are employed to evaluate the accuracy of the presence of pollution.

Figure 3: Block Diagram of Proposed Methodology

5. RESULTS AND DISCUSSION

The result generated from proposed methodology is shown in Table 3, which shows predicted and actual value of Oxide of Nitrogen using neural network and the obtained prediction accuracy is 83%. The predicted values of NOx are binary in nature which aids in identifying the presence of pollution using classification algorithms. The classification task is completed using numerous classification algorithms. The rule based classifiers such as

Table 3
Observed and Actual Value of Nox

Observed Data of NOx	*Actual Data of NOx*
2	3
2	5
2	6
2	6
2	5
1	5
2	6
2	6
2	6
2	5
2	5
1	3
2	4
2	5
2	4
2	5
2	5
2	5
2	5
2	4

Table 4
Comparing Accuracy using various Classifiers

Classification Algorithms	*Accuracy (%)*
IBk (K-Nearest Neighbor)	98.6301
K Star	95.8904
Locally Weighted Learning (LWL)	94.5205
Naive Bayes	93.1507
Logistics	97.9452
J48	96.5753
Decision Table	84.9315
OneR	85.6164
ZeroR	82.1918
JRip	96.5753
PART	94.5205
Decision Stump	85.6164
Bayesian Network	96.5753
Hidden Naïve Bayes	95.8904
Multilayer Perceptron	99.3151
Stochastic Gradient Descent	100.00
Simple Logistics	99.3151
Logistic Model Tree	99.3151
Random Forest	99.3151
Random Tree	99.3151

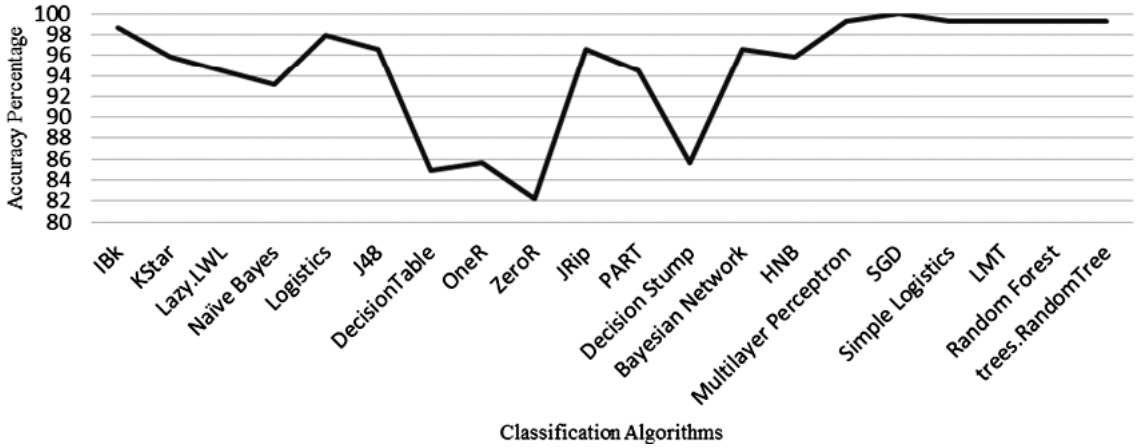

Figure 4: Graph for Comparing Accuracy using various Classifiers

OneR, ZeroR and JRip shows less accuracy while lazy and Bayes algorithms give moderate accuracy. But, the Stochastic Gradient Descent function based classifier and decision trees based algorithms outperforms other classifiers present in machine learning environment. The function based classifier Stochastic Gradient Descent gives 100% accuracy while decision tress such as Random Tree and Random Forest give 99.3151% Classification accuracy using various classifiers is shown in Table 4 and graph comparing accuracy of classifiers is shown in Figure 4.

6. CONCLUSION

In the presented work, the study of prediction of Oxide of Nitrogen (NOx) in environment from the Delhi metropolitan dataset is conducted. The present dataset is obtained from the CPCB website databank. Oxide of Nitrogen concentration in the air of Delhi is alarming and has created great havoc therefore the work is focused on the prediction of this pollutant. Artificial neural networks has been applied to predict the concentration of the pollutant. The accuracy was found to be 83% and Kappa value 0.8333. The binary predicted values give better accuracy on the application of classification algorithms. The study shows that discrete values in a dataset of any domain can give prediction results. As observed in the results, the decision trees classifiers and function based classifiers give better accuracy as compared to rule based, Bayes and Lazy classifiers. The proposed methodology in the work is not just limited to one pollutant, that is, Oxide of Nitrogen but could be applied to any pollutant hampering the quality of air in any city. The adopted methodology can have wide application not just in the field of atmospheric sciences but also in the ecological and chemical sciences.

REFERENCES

[1] Sharma, Sumit, et al. "Measures to Control Air Pollution in Urban Centres of India." TERI Policy Brief (2016): 1-20.

[2] Ravindra, Khaiwal, et al. "Air Pollution in India: Bridging the Gap between Science and Policy." Journal of Hazardous, Toxic, and Radioactive Waste (2015): A4015003.

[3] Janhäll, Sara. "Review on urban vegetation and particle air pollution–Deposition and dispersion." Atmospheric Environment 105 (2015): 130-137.

[4] Stern, Arthur C., ed. Fundamentals of air pollution. Elsevier, 2014.

[5] Aneja, Viney P., et al. "Measurements and analysis of criteria pollutants in New Delhi, India." Environment International 27.1 (2001): 35-42.

[6] Goyal, S. K., et al. "Understanding urban vehicular pollution problem vis-a-vis ambient air quality–case study of a megacity (Delhi, India)." Environmental monitoring and assessment 119.1-3 (2006): 557-569.

[7] Hameed, Sultan, and Jane Dignon. "Changes in the geographical distributions of global emissions of NOx and SOx from fossil-fuel combustion between 1966 and 1980." Atmospheric Environment (1967) 22.3 (1988): 441-449.

[8] Cropper, Maureen, et al. "The health effects of air pollution in Delhi, India." World Bank Policy Research Working Paper 1860 (1997).

[9] Véron, René. "Remaking urban environments: the political ecology of air pollution in Delhi." Environment and Planning A 38.11 (2006): 2093.

[10] Chhabra, Sunil K., et al. "Ambient air pollution and chronic respiratory morbidity in Delhi." Archives of Environmental Health: An International Journal 56.1 (2001): 58-64.

[11] Guttikunda, Sarath K., and Giuseppe Calori. "A GIS based emissions inventory at 1 km× 1 km spatial resolution for air pollution analysis in Delhi, India." Atmospheric Environment 67 (2013): 101-111.

[12] Mohan, Manju, and Anurag Kandya. "An analysis of the annual and seasonal trends of air quality index of Delhi." Environmental monitoring and assessment 131.1-3 (2007): 267-277.

[13] Gurjar, B. R., et al. "Evaluation of emissions and air quality in megacities." Atmospheric Environment 42.7 (2008): 1593-1606.

[14] Kumar, Anikender, and P. Goyal. "Forecasting of daily air quality index in Delhi." Science of the Total Environment 409.24 (2011): 5517-5523.

[15] Nagendra, SM Shiva, and Mukesh Khare. "Modelling urban air quality using artificial neural network." Clean Technologies and Environmental Policy 7.2 (2005): 116-126.

[16] Zito, Pietro, Haibo Chen, and Margaret C. Bell. "Predicting real-time roadside CO and concentrations using neural networks." IEEE Transactions on Intelligent Transportation Systems 9.3 (2008): 514-522.

[17] Hill, Tim, et al. "Artificial neural network models for forecasting and decision making." International Journal of Forecasting 10.1 (1994): 5-15.

[18] Jha, Girish Kumar. "Artificial neural networks." Indian Agricultural Research Institute (2007): 1-10.

[19] Kaur, Tarlochan, Sanjay Kumar, and Ravi Segal. "Application of artificial neural network for short term wind speed forecasting." Power and Energy Systems: Towards Sustainable Energy (PESTSE), 2016 Biennial International Conference on. IEEE, 2016.

[20] Maier, Holger R., and Graeme C. Dandy. "The use of artiûcial neural networks for the prediction of water quality parameters." Water Resour Res 32.4 (1996): 1013-1022.

[21] Kaastra, Iebeling, and Milton Boyd. "Designing a neural network for forecasting financial and economic time series." Neurocomputing 10.3 (1996): 215-236.

[22] Zhang, Guoqiang, B. Eddy Patuwo, and Michael Y. Hu. "Forecasting with artificial neural networks:: The state of the art." International journal of forecasting 14.1 (1998): 35-62.

[23] Moritz, Steffen, et al. "Comparison of different Methods for Univariate Time Series Imputation in R." arXiv preprint arXiv:1510.03924 (2015).

[24] Zeileis, Achim, and Gabor Grothendieck. "zoo: S3 infrastructure for regular and irregular time series." arXiv preprint math/0505527 (2005).

[25] Ripley, B. 2016. Package 'nnet'. R package version 7.3-12.

[26] Lotte, Fabien, et al. "A review of classification algorithms for EEG-based brain–computer interfaces." Journal of neural engineering 4.2 (2007): R1.

[27] Chen, Ming-Syan, Jiawei Han, and Philip S. Yu. "Data mining: an overview from a database perspective." IEEE Transactions on Knowledge and data Engineering 8.6 (1996): 866-883.

[28] Aggarwal, Charu C., ed. Data classification: algorithms and applications. CRC Press, 2014.