IOP Conf. Series: Materials Science and Engineering **396** (2018) 012016 doi:10.1088/1757-899X/396/1/012016

# A Review on Various Techniques used in Predicting Pollutants

**Aiswarya Baby**

P.G Scholar, Department of Computer Science and Engineering, St Joseph's College of Engineering and Technology,Palai, India

E-mail: aiswaryapbaby@gmail.com

**Aneena Ann Alexander**

Assistant Professor, Department of Computer Science and Engineering, St Joseph's College of Engineering and Technology,Palai,India

E-mail: aneenaalex@gmail.com

**Abstract.** Pollution is now becoming a biggest problem in today's age. In numerous nations, because of the high number of vehicles that are accessible on the streets, air contamination has progressed toward becoming a difficult issue. This eventually leads to emission of different gases. So there is a need to predict the amount of emission beforehand. There have been various research commitments in the region of anticipating toxins. Artificial intelligence (AI) are often employed where a problem requires estimation or prediction. Most of the methods used are based on neural network, fuzzy logic, hybrid methods and others. This paper discusses findings and limitations of existing contributions in the field of pollution prediction.
**keywords**:Artificial Intelligence, pollution prediction, Neural Network, Multilayer Perceptron.

## 1. Introduction

Artificial Intelligence(AI) is playing an increasing role in todays life. Intelligence is considered as the ability to collect reason and knowledge about knowledge to solve complex problems. In the near future intelligent machines will replace human capabilities in many areas. Artificial intelligence is the study and developments of intelligent machines and software that can learn, reason, communicate, gather knowledge, perceive and manipulate the objects. It makes machines more helpful and more brilliant. It works with the help of scientific theorems (if then statements and logics) and artificial neurons (artificial neural network). AI technologies have matured to the point in offering real practical benefits in many of their applications. Major Areas in Artificial Intelligence include Natural Language Processing ,Expert Systems, Neural Computing ,Robotics and Sensory Systems, Computer Vision and Scene Recognition, Speech Understanding, Intelligent Computer Aided Instruction.

A large number of tools are used to solve the most difficult problems in computer science. Some of the methods are Search and optimization, Logic , Probabilistic methods for uncertain reasoning , Classifiers and statistical learning methods, Neural networks, Deep feed forward neural networks, Deep recurrent neural networks, Control theory, Languages, Evaluating progress.

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

## 2. Literature Review

### *2.1. Greenhouse Gas Estimation*

Environmental pollution is the major challenge that the world faces today. The transportation system is one of the greatest contributors to the deterioration of the environmental situation because vehicles are the major contributors to greenhouse gas emissions. In this manner, transportation is one of the real sources of the greenhouse gas emissions. Two major approaches that are mainly used to estimate the GHGs emissions, are bottom-up models and top-down models proposed by the IPCC [1]. In top-down. approach the entire carbon emission is calculated from the total fuel consumption in the city. Later this total carbon emission is divided among each economic sector, such as transportation sector. Whereas in bottom-up approach from each economic sector the fuel consumption is calculated. Then according to the carbon emission parameters of the fuel, the carbon emission amount is calculated.

However the two models have many uncertainties because of uncertainties in the empirical assumptions and the uncertainty in the primary data such as fuel consumption, carbon content, and oxidation factors[2]. The top-down approach mainly focuses on market interactions. It doesnt give much importance to the technological detail in the energy sector. While the major focuses of bottom-up approach is on substitutability of individual energy technologies and it lack macro-economic feedbacks between other economic sectors and the energy.

### *2.2. Greenhouse Gas Prediction*

*2.2.1. Neuro-Fuzzy InferenceTechniques :* An neuro-fuzzy inference system (ANFIS) consolidates both neural networks and fuzzy systems. A neuro-fuzzy system depends on a fuzzy system that is trained by a learning algorithm derived from neural network theory. The learning technique works on nearby data, and results in narrow modifications in the underlying fuzzy system. Using neuro-fuzzy inference system an environmental monitoring system is designed to predict the pollutant levels in Italy [3]. Input data are those collected from sensors environmental parameters. The model efficiently predicted the pollutant levels. The criteria used for evaluation was Root Mean Square Error (RMSE). This evaluation criterion can stress the models performance.

*2.2.2. Ordered Weighted Averaging based Time Series Model :* OWA is a multicriteria evaluation procedure. The OWA methodology relies upon a few parameters, which can be specified by fuzzy quantifiers. OWA based time series model are used to predict future occasions in light of past occasions. The amount of Ozone emitted in Taiwan is predicted using this method [4]. Here the data from air monitoring stations are used. Ozone is selected as the attribute since it is the air pollutant species of concern. This strategy combines the merits of both time series method (TSM) and ordered weighted averaging (OWA). The criteria used for evaluation was RMSE and Mean Square Error (MSE) . This can stress the models performance and takes a lot of computational time.

*2.2.3. Grey Dynamic Model :* The grey dynamic model performs well in modelling and forecasting. The Grey Dynamic Model was utilized to forecast the oil pollutants in China [5]. Since the oil utilization in China is expanding step by step, the development of oil stations is in the ascendant. Concentrations of CH,CO, NOx are used as input data. These information are then pre-processed and are utilized to train the Grey model. The problem here is that the dataset used are not diverse.

*2.2.4. Response Surface Methodology :* Response surface methods (RSM) are information based techniques. They perform well when there is a rich tapestry of accurate information provided.

Due to their ease of use as well as dexterity of being applied across the board of industrial applications They have gained huge fame. This technique was utilized to foresee the NOx emission [6] and input data like coal quality, total air supply, oxygen concentration and air velocity are considered. Four different Response surface models were used and the model with quadratic terms had the best prediction. The challenge for RSM is to demonstrate the continuum and to capture precisely the rich collection of physical phenomena.

*2.2.5. Matern Function Based Extended Fractional Kalman Filtering :* This method is used to addresses the problem of predicting air pollution emissions [7]. Uncertainty in data involves measurement error or error related to instrument precision .Here the input data are the observed station data. But Kalman Filtering techniques suffer from a few impediments, for example, that identified with the need of exact stochastic models of sensor random errors and its immunity to noise, observability.

*2.2.6. Digital Imaging, Contourlet Transform and Radial Basis Function Network Techniques :* Digital images are electronic snapshots taken of a scene or scanned from documents, for example, photographs, printed texts, and artwork. Through the combination of flame radical imaging, contourlets transform, and radial basis function network techniques, NOx emissions in a biomass ignition process is predicted [8]. Using a spectroscopic imaging system the images of four flame radicals (OH*, CN*, CH* and C2*) are captured. Based on contourlet coefficients the features of the images are then identified .These features are used to train the radial basis function network techniques(RBF). The The fundamental disservice is that this is time consuming.

*2.2.7. Regression Model :* The economy of a country can be improved by rapid industry development. But this leads to the consequences of heavy CO2 emissions. So there is a need for the CO2 emission prediction. Regression model was used for the CO2 emission prediction purpose [9]. The relationships among variables can be estimated by this method.

Another Regression model application is in the estimation and forecast of pollutants in road tunnels [10]. The sensors collect data on traffic parameters, such as traffic flow , speed , truck ratio , and others as well as environmental variables, like atmospheric pressure , air flow velocity , air flow direction. The data collected are pre-processed and is then applied to the Regression model. The relationship that is established between variables based on some limited data may not hold great if more and more data are taken into consideration.

*2.2.8. FCM-HMM Clustering and TS Fuzzy Inference Algorithm :* A multi model frame is constructed based on FCM-HMM clustering and TS fuzzy inference multi-model to predict the air pollution index [11]. Data used are meteorological and climate factor. But HMM frequently have various unstructured parameters and dependencies between hidden state cannot be expressed.

*2.2.9. Support Vector Machine :* Support vector machine (SVM) is a discriminative classifier and is used for the prediction of the co2 level [12]. Input data is the co2 amount which is emitted by burning coal and when generating electricity. The data is normalized, cross validated and then applied to the support vector machine. The major drawback of the support vector approach lies in choice of the kernel.

*2.2.10. Neural Network Technique :* Various gases have been predicted using Neural Network technique (NN). Neural networks are parallel computing devices. They are essentially an attempt to make a computer model of the brain. Its fundamental goal is to build up a system that perform

various computational tasks faster than the traditional systems. A multilayer perceptron neural network based model for short prediction of SO concentrations was developed for the stations around the SoStanj Thermal Power Plant [13]. The input features are fed into the model and backpropagation algorithm is used to train the model. One major drawback observed is that forecast of the so2 gas is for a brief timeframe because of the limited amount of data.

In order to achieve good generalizing capabilities of the model, one of the most important tasks that should be solved is the selection of the patterns used for neural network-based model training [14]. Two types of pattern selection procedures were developed: a meteorological knowledge - based strategy and a Kohonen neural network-based strategy. Patterns selection reduce the computational cost. It also improve the models generalizing capabilities and hence enhance the quality of prediction. The result of these two strategies showed a high increase of the models generalizing capabilities. The strategy based on Kohonen neural network is certainly more suitable for use in other research fields such as speech recognition, or otherwise where huge data bases are available.

Semi-supervised learning is a class of machine learning techniques that make use of both unlabeled and labeled data for training - normally a small amount of labeled data with a large amount of unlabeled data. Semi-supervised learning (SSL) approach was used to predict air quality information using input data such as traffic flow, POIs , meteorology, structure of road networks, and human mobility [15]. This method provides accurate air quality information. But verifying SSL assumptions or mathematically formalizing them is very difficult.

Feed-forward ANNs enable signals to travel one way only that is from input to output. There are no loops. Feed-forward ANNs tend to be straightforward networks that associate inputs with outputs. Feedback networks can have signals travelling in both directions by introducing loops in the network. Feedback networks are powerful and can get extremely complicated. Using a neural network model based on feed-forward neural networks and recurrent neural network accurate forecasting of fine particulate matter concentration is done [16]. Here the dataset used are not diverse. So there occurs a problem of over fitting.

Multi Layer perceptron (MLP) is a feedforward neural network with one or more layers between input and output layer. Backpropagation learning algorithm is used to train this type of network. MLPs are widely used for recognition, pattern classification, prediction and approximation. Such a network is used for the transportation carbon emission prediction in the cities [17 ]. The model showed improved accuracy. Multi Layer Perceptron can solve problems which are not linearly separable.

**Table 1.** Comparison Table

| Techniques | Pros | Cons | Performance metrics |
|---|---|---|---|
| Neuro-Fuzzy InferenceTechniques | • ability to change the qualitative aspects of human knowledge into the process of precise quantitative analysis | • time consuming | • efficiency |
| Grey Dynamic Model | • useful for dealing with poor, incomplete, and uncertain information | • low prediction precision | • accuracy |

| Techniques | Pros | Cons | Performance metrics |
|---|---|---|---|
| Ordered Weighted Averaging based Time Series Model | • allows decision-makers to change the form of attribute combinations from a minimum-type combination through all intermediate types to a maximum-type combination. | • takes a lot of computational time | • efficiency and accuracy |
| Response Surface Methodology | • simplicity and easiness | • difficult to model the continuum and to capture accurately the rich repertoire of physical phenomena. | • predictive accuracy |
| Matern Function Based Extended Fractional Kalman Filtering | • results to a faster convergence in terms of iterations compared to traditional methods | • Space and computation complexity | • accuracy |
| Digital Imaging, Contourlet Transform and RBF Network Techniques | • RBF is quick to train<br>• the hidden layer of RBF is easier to interpret than the hidden layer in an MLP | • Although the RBF is quick to train,after training is finished , when it is used it is slower than a MLP | • stability |
| Regression Model | • Easy to implement and interpret | • involves very lengthy and complicated procedure of calculations and analysis. | • accuracy |
| FCM-HMM Clustering and TS Fuzzy Inference Algorithm | • fairly readable<br>• HMM generally provides better compression than a simple Markov model | • training can be quite slow<br>• suffers heavy computational burdens thereby becoming very time-consuming | • improved prediction performance |
| Support Vector Machine | • prediction accuracy is generally high<br>• robust | • long training time<br>• difficult to understand the learned function | • accuracy |

| Techniques | Pros | Cons | Performance metrics |
|---|---|---|---|
| NN Technique | • no prior knowledge of the data generating process is needed for implementing NN<br>• problem of model misspecification does not occur | • addition of too many hidden units incites the problem of over fitting the data | • accuracy |

## 3. Conclusion

In this paper, various methods that are used to predict the greenhouse gas emission were discussed. Artificial Intelligence plays a very important role in forecasting various greenhouse gas emission. In forecasting models, the major concern is with accuracy. More and more complex models have been introduced to gain accuracy. It is found that Multi Layer perceptron is more efficient than other techniques in terms of accuracy. In future, various efficient hybrid algorithms/model should be substituted in place of complex models to decrease the execution time of the model.

## 4. References

[1] I. P. on Climate Change, 2006 IPCC Guidelines for National Greenhouse Gas Inventories. Intergovernmental Panel on Climate Change, 2016.

[2] Detlef P.vanVuuren , MoniqueHoogwijk , TerryBarker , KeywanRiahi , StefanBoeters , *Comparison of top-down and bottom-up estimates of sectoral and regional greenhouse gas emission reduction potentials*,Science Direct on 2009.

[3] F.CM. orabito, M.Versaci, *Prediction and Estimation of Atmospheric Pollutant Levels by Soft Computing Approach* ,Proceedings. IJCNN 01. International Joint on 2001.

[4] Sue-Fen Huang,Ching-Hsue Cheng *Forecasting The Air Quality Using OWA Based Time Series Model*,International Conference on Machine Learning and Cybemetics on 2008.

[5] Zhao Xiao-gang, Zhou Yi, Zhao Jian-yu *Study on the Prediction of Air Pollutants in Urban Oil Depot Based on Grey Dynamic Model* ,International Symposium On Water Resource and Environmental Protection on 2011.

[6] Ligang Zheng, Shuijun Yu, Minggao Yu , *Prediction of Nitrogen Oxides from Coal Combustion by Using Response Surface Methodology* ,International Conference on Computer Distributed Control and Intelligent Environmental Monitoring on 2012.

[7] S. Metia, S. D. Oduro and Q. P. Ha *Air Pollution Prediction Using Matern Function Based Extended Fractional Kalman Filtering* ,2014 13th International Conference on Control Automation Robotics Vision (ICARCV) on 2014.

[8] Nan Li, Gang Lu, Xinli Li, Yong Yan,*Prediction of Pollutant Emissions of Biomass Flames Using Digital Imaging, Contourlet Transform and Radial Basis Function Network Techniques* ,IEEE International Instrumentation and Measurement Technology Conference Proceedings on 2014.

[9] Tan Chun Ho, Sim Chong Keat Mohd Zubir Mat Jafri and Lim Hwee San ,*A Prediction Model for CO2 Emission From Manufacturing Industry and Construction in Malaysia*,International Conference on Space Science and Communication on 2015.

[10] Ondrej Pribyl, Pavel Pribyl , *On the estimation of pollutants in road tunnels based on traffic data* ,ELEKTRO on 2016.

[11] Ling Wang , Xi-yuan Xiao , Jian-yao Meng, *Prediction of Air Pollution Based on FCM-HMM Multimodel* , 35th Chinese Conference(CCC) on 2016.

[12] Chairul Saleh, Nur Rachman Dzakiyullah, Jonathan Bayu Nugroho ,*Carbon dioxide emission prediction using support vector machine*, Materials Science and Engineering 114 on 2016 .

[13] Primoi Mlakar and Marija Boinar ,*Perceptron Neural Network - Based Model Predicts Air Pollution*,Intelligent Information System, 1997. IIS 97. Proceedings on 1997.

[14] Marija Boinar, Jamova , *Pattern Selection Strategies for a Neural Network -Based Short Term Air Pollution Prediction Model* ,Intelligent Information System, 1997. IIS 97. Proceedings on 1997.

[15] Y. Zheng, F. Liu, and H.-P. Hsieh ,*U-air: when urban air quality inference meets big data*, in Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining.ACM, 2013.
[16] Mihaela Oprea1, Marian Popescu, Sanda Florentina Mihalache, *A Neural Network Based Model for PM2.5 AirPollutant Forecasting* ,20th International Conference on System Theory, Control and Computing (ICSTCC) on 2016.
[17] Xiangyong,Karoru,Dong,Chen and Hai Ji; *Predicting Transportation Carbon Emission with Urban Big Data"* IEEE International Journal on 2017.