PM_{10} prediction using Genetic Programming: A Case Study in Salt, Jordan

Hossam Faris[1], Mouhammd Alkasassbeh [2], Nazeeh Ghatasheh [3], Osama Harfoushi[1]

[1.] King Abdulla II School for Information Technology, The University of Jordan, Amman, Jordan
[2.] Computer Science Department, Mutah University, Jordan
[3.] Department of Business Information Technology, The University of Jordan, Aqaba, Jordan
hossam.faris@ju.edu.jo

Abstract: Monitoring and controlling air pollutants have been one of the main environmental concerns so far. Such concerns are highly emphasized and monitored in large cities all over the world by air quality management systems. The various polluting emissions transported by atmospheric air affect the living bodies, including the human's health, wild life, and plants. Nation wise, air pollution negatively imposes economic effects. Regularity boards, usually governmental, induce actions to reduce air pollution levels in the industrial regions; by limiting certain emission amounts and imposing air quality standards. This paper aims to put in hand a symbolic regression prediction model based on the genetic programming algorithm. The main objective of the prediction model is to predict the Particulate Matters (PM_{10}) near Salt City, Jordan. This study analyzes the recordings of five monitoring stations around Al-Fuhais cement plant between the years 2006 and 2007. The incorporated and measured meteorological input variables are the Relative Humidity, Atmospheric Temperature and Wind Speed. The results of the current study prove that genetic programming technique can performs better than other approaches tackling the same issue.

[Hossam Faris, Mouhammd Alkasassbeh, Nazeeh Ghatasheh, Osama Harfoushi. **PM_{10} prediction using Genetic Programming: A Case Study in Salt, Jordan.** *Life Sci J* 2014;11(2):86-92]. (ISSN:1097-8135). http://www.lifesciencesite.com. 12

Keywords: Air pollution; PM_{10}; Genetic Programming.

1. Introduction

Predicting air quality provides essential information for environmental management. Though, the complexity of air pollution issue makes it challenging to formulate a concrete prediction model. Such model should address many related input variables [1],[2]. Air pollution was one of the factors causing health problems for urban areas [3]. Authors in [4] stated that main air pollutants include sulfur dioxide, nitrogen dioxide, polycyclic aromatic hydrocarbons (PAH) and suspended particulate matter (PM) as dust, soot and smoke.

Particulate Matter (PM), alternatively called particle pollution, denotes the complex combination of liquid drops and tiny particles. PM consists of various elements, including acids, organic compounds, metals and dust particles. The tiny PM particles can reach human's lung through the respiratory system. Several serious health issues are related to the penetration of these particles, affecting the lungs and the heart [3],[16]. Moreover, the exposure to air pollutants for the long-term was related to having a higher probability of many types of cancer [15],[18].

PM is expressed by the mass of the particles in micrograms per one cubic meter of air (µg/m3), normally the diameter of the particles is less than 10 micrometers. Two air quality standards introduced by the United States are the PM_{10} and $PM_{2.5}$, while for industrial activities, PM_{10} was selected to be the air pollution standard because many industrial dust particles have a diameter larger than 2.5 µm [19-21].

Due to the importance of the environmental issues, there were several attempts to tackle the area of air pollution prediction [5-8]. Forecasting the daily PM_{10} average concentrations for Bordeaux metropolitan area was conducted by an "adaptive nonlinear state-space based modeling technique" in [9]. The designed model incorporated the empirical relationships among the calculated PM_{10}, main pollutants and metrological variables. In addition, an extended Kalman filter was used to compute one day ahead prediction for Bordeaux area.

In literature, researchers proposed different machine learning approaches such as Artificial Neural Networks (ANN), Fuzzy Logic (FL) and Genetic Algorithms (GAs) for building models that predict atmospheric pollutant's concentrations. In [10], authors proposed air pollution time-series based model using a Back-Propagation (BP) method. Furthermore, ANN was applied to predict various environmental issues [11],[12]. A unified approach was proposed in [13]. The approach integrates explicit and implicit knowledge in neuro-symbolic systems in a collective form composed of neuro-fuzzy and neural modules. The neural implicit knowledge modules were used to extract fuzzy rules.

While in [14], GAs found the most suitable feature to formulate an air-quality model.

For the sake of protecting human health several researchers proposed machine learning based models [17],[22-23]. A machine learning tool based on ANN and SVM was proposed by [17] to predict the PM_{10} level from the principal causes to air pollution. The authors in [22] presented satisfying result for applying an ANN based model for the air pollution issue in Belgium, by analyzing the height within the boundary layer and PM_{10} values. While in [23], a Support Vector Regression (SVR) model to predict PM_{10} particles in Bankog was proven to be suitable.

Many air pollution studies were conducted over Jordan. Previously, the statistical outcomes for the concentrations of TSP matter over the period July 1986 – June 1987 based on four sampling sites [26]. The input variables included were nitrogen dioxide, carbon monoxide, sulfur dioxide, PM_{10} and TSP. Black carbon levels were studied in some Jordanian cities and industrial centers [24]. A photo-acoustic tool, having 870nm wavelength, was used to measure black carbon light absorption coefficients during the summer of 2007 and the winter of the consequent year.

During the spring of the year 2009, the authors in [27] calculated the submicron particle concentrations for urban and suburban areas in Amman city. They tried to find the differences between the concentrations with the presence and absence of dust episodes.

In this paper, we investigate the application of GP for developing a model to predict the atmospheric PM_{10} concentrations. The proposed GP model has three input variables, which are temperature, relative humidity and wind speed. The case study of this research is Al-Fuhais City in Salt, Jordan. Al-Fuhais is considered as one of the most polluted areas in Jordan since it has the largest cement plant in the country. The location of the plant is surrounded a residential area. Moreover, there is a high transportation traffic for the propose of cement shipping. It was reported by many studies that the particulate matter emitted from such industrial processes has potential effects that cause health problems.

3. Genetic Programming

GP is a powerful evolutionary computing technique for automatically solving problems, which is inspired by the theories of biological evolution. GP was first introduced by J. R. Koza [28-29]. GP has some remarkable advantages over other techniques in modeling complex and nonlinear systems in a wide range of different domains [25–27]. Some of these advantages can be summarized in the following two points:

- GP generates mathematical models in a tree structure which are less complex than models developed by other approaches like ANN. Therefore, GP models are easier to evaluate.
- No prior knowledge is needed in advance about the internal structure of the system. GP can adapt automatically to various constraints.
- GP has an identification and explanation power. GP models can give an insight into the complex relationships between model variables.

GP generates automatically mathematical models in form of tree structures through an evolutionary cycle. GP evolutionary cycle can be summarized in the following four steps:

- Initialization: the evolutionary cycle starts by randomly generating a predefined number of individuals which form the initial population. Number of individuals in a population is referred to as *population size*. Each individual is a mathematical expression and can be represented as a tree or as LISP expression. In Figure 1, we show a simple GP tree representation of the system (with output z) as given in Equation 1.

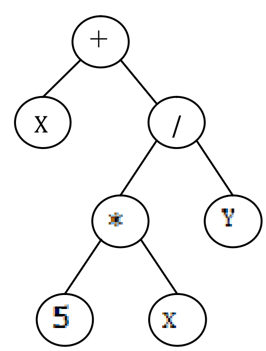

Figure 1. Simple GP mathematical model.

$$X + \left(\frac{5 * X}{Y}\right) \qquad (1)$$

Fitness Evaluation: each individual is evaluated using a specific evaluation measurement. In this work, we use the Mean Absolute Error (MAE) criteria. Fitness evaluation assesses the accuracy of the generated individuals.

- Reproduction: GP generates new population to replace the older one by selecting individuals and then applying reproduction operators. There are many selection mechanisms proposed in the literature. The most famous one is the tournament selection where a number of individuals are chosen randomly then the fittest individual among them is selected for reproduction. The most common reproduction operators are crossover and mutation, which can briefly describe as follows:

A. Crossover: this operator selects two individuals at random then a randomly chosen part of the first individual is swapped by another randomly chosen part from the second individual.
B. Mutation: this is unary operator. A random point is chosen in an individual then the subtree under this point is replaced by randomly generated new subtree. Usually, mutation is applied with probability much less than the crossover one.

- Termination condition: the evolutionary cycle of GP stops iterating when the maximum number of individuals is reached or an individual with a required level of fitness is reached.

4. Materials and methods

A. Area of Study

The case study of this work is Al-Fuhais city located 21.5 km west of Amman. The area of the city is about 30 square kilo meters. The height is 550m to 1200m above mean sea level, and the total population is about 20,000. The annual average temperature in the city ranges from 18-220c. The coldest month in the year is January, where the average temperature reaches 6^0C, while the warmest month in the year is August with the average temperature of 24^0C. The precipitation may reach 600mm. The average wind speed throughout the year is 5.46 knot.

Al-Fuhais has the biggest cement plant in Jordan with the annual product of 2 million tons per year. Dust is emitted from different cement manufacturing processes and activities, including mining, milling, transport, handling of raw materials and kiln operation. In addition to dust emitted from vehicles and trucks. Since Al-Fuhais cement plant is surrounded by residential areas, imitations can cause potential health problems [31], [32].

B. Monitoring Instruments

Measuring process is carried out by using High-Volume Sampler (HVS) GS3210 and Beta attenuation monitors in order to measure the PM_{10} concentrations. HVS is shown in Figure 2. In our case, the HVS is installed on the terrace of the buildings, and the instrument was placed in a parallel position to the ground. The HVS was used as a vacuum cleaner; it drew a volume of air through a filter that captured suspended particles. The sampler is operated at flow rates of 1m3/min collected on pre-weighed glass fiber filters. The filters weighing as accurately as possible since the ambient humidity is the common cause of producing wrong filter weights. The run was held for a period of 24 hours and at the end of the sampling period. The filter paper was removed, and a daily reading for the instrument was taken.

Figure 2. High-volume sampler GS3210.

C. Monitoring Stations

A monitoring program for PM_{10} was deployed in order to collect data and assess the environment situation in the area. The Faculty of Agricultural Technology, Al-Balqa Applied University (BAU), Salt, Jordan used five monitoring stations to measure PM_{10} atmospheric concentrations. The five stations were selected up to a distance of 7 km around the plant and cover different types of environments and surround the main industrial activities in the area.

D. Dataset description

A complete dataset of the meteorological factors and air pollution parameters PM_{10}, with particulate matter of a diameter less than 10µm was carried out in the five stations around Al-Fuhais cement plant over the one-year period started on 26 November 2006 to 25 November 2007. Ambient air quality was monitored with a sampling frequency of 24 hours. The meteorological data include temperature, relative humidity and wind speed, which were obtained from Al-Salt meteorological station (the nearest station to the study area). For the purpose of evaluation of the GP model, 2/3 of the data were used for the training phase and develop the GP model while the other 1/3 were used for testing the model. Figures 3 and 4 show the training and testing data respectively.

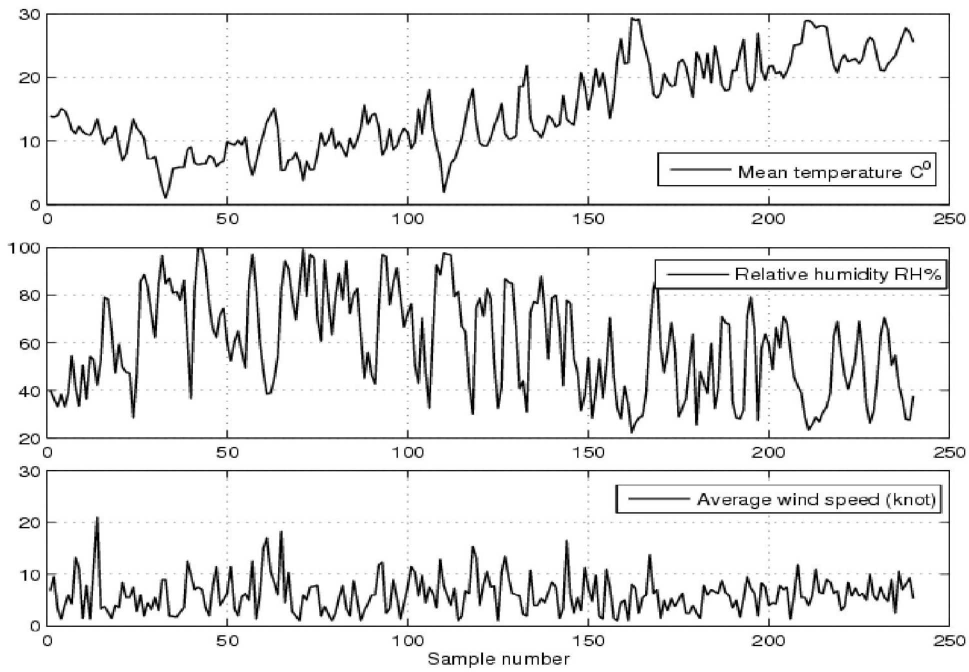

Figure 3. Measurements of input variables in training dataset.

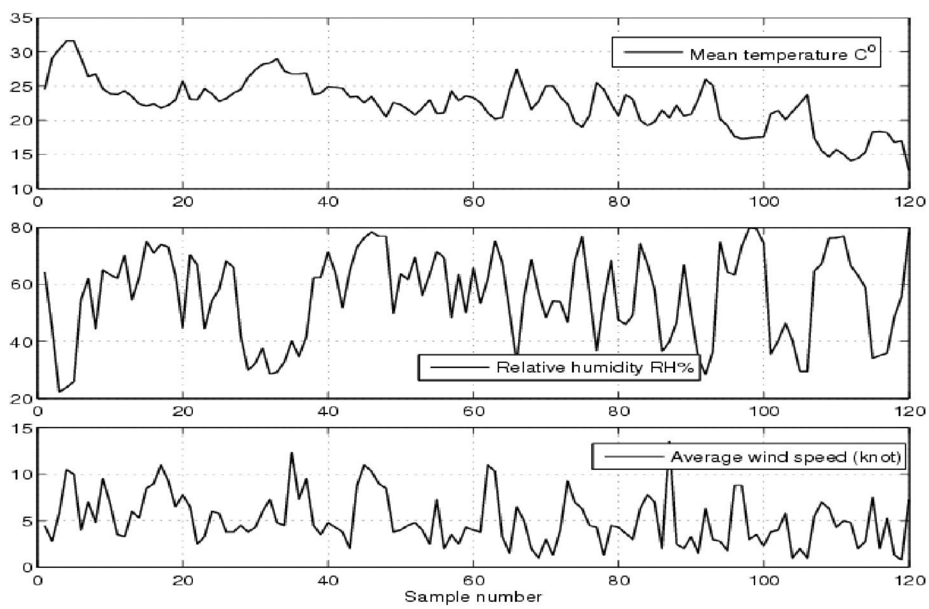

Figure 4. Measurements of input variables in testing dataset.

3. Experiments and results

A. Evaluation methods

To assess the performance of the developed GP tree model for PM_{10}, Mean Squared Error (MSE) and Mean Absolute Error (MAE) are used to measure how close the measured values to the predicted ones. RMSE and MAE are computed as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y-\hat{y})^2 \qquad (2)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y-\hat{y}| \qquad (3)$$

where y and $\hat{y}$ are the actual and the estimated PM_{10} values based on the GP model and n is the number of samples used in the experiments, respectively.

B. Modeling framework

HueristicLab framework was used to apply the GP approach and the experiments designed in this research. HeuristicLab is a framework for heuristic and evolutionary algorithms developed by members of the Heuristic and Evolutionary Algorithms Laboratory (HEAL).

The data set described earlier was loaded into Heuristiclab then GP was applied with parameters tuned as shown in Table 1. After a run of 200 generations GP was able to converge to a model with MSE value of 212.76 and MAE value of 10.58 for the testing case. Actual and predicted values for both cases training and testing are shown in Figure 4. In Table 2, results are compared with those obtained by ANN model developed earlier in [30]. GP shows better evaluation results compared to ANN model.

Table 1. GP Tuning Parameters.

Population size	1000
Number of generation	500
Selection mechanism	Tournament
Max. tree depth	12
Crossover probability	90%
Mutation probability	15%
Maximum tree depth	12
Maximum tree length	150
Operators set	{+,-,/,*,log, exponential}

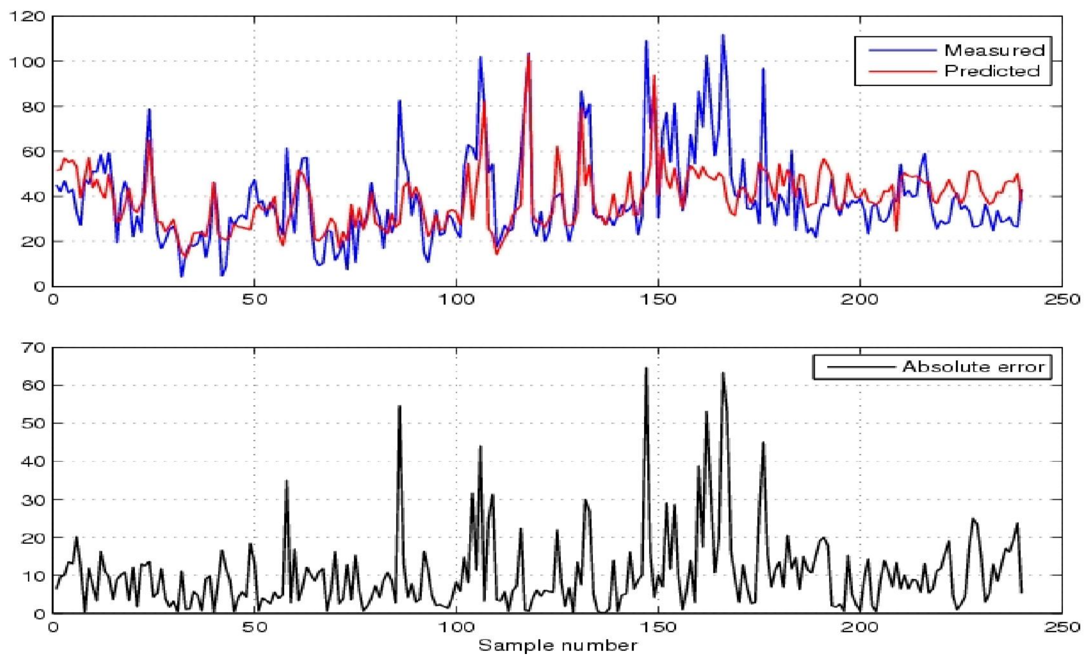

Figure 5. Actual and predicted values for PM_{10} based on the developed GP model.

Table 2. Evluation results for GP and ANN models

	GP		ANN [30]	
	Training	Testing	Training	Testing
MSE	222.85	212.76	155.81	219.79
MAE	11.88	10.58	9.3084	11.208

6. Conclusion

Several studies and reports pointed out that airborne particulate matter (PM) can cause health problems. For that reason, predicting PM concentrations is important. In this paper, we investigated the application of Genetic Programming for developing a prediction model for PM_{10} air pollution parameter in Salt, Jordan. The developed GP model has only three input variables, which are the following meteorological parameters: Temperature (Temp), Relative Humidity (RH), Wind Speed (WS) as inputs. GP showed reasonable prediction power and out performed a neural network based model from the literature used for the same case study.

Acknowledgements:

Authors would like to thank the Faculty of Agricultural Technology, Al-Balqa Applied University (BAU) for providing the necessary data collected at five monitoring stations to measure PM_{10} concentrations at Al-Fuhais, Salt, Jordan.

Corresponding Author:
Dr. Hossam Faris
Department of Business Information Technology
The University of Jordan
Amman, Jordan
E-mail: hossam.faris@ju.edu.jo

References

1. N. S. Leksmono, J. W. S. Longhurst, K. A. Ling, T. J. Chatterton, B. E. A. Fisher, and J. G. Irwin, "Assessment of the relationship between industrial and traffic sources contributing to air quality objective exceedences: a theoretical modelling exercise," Environ. Model.Softw., vol. 21, pp. 494–500, April 2006.
2. V. Mallet and B. Sportisse, "Air quality modeling: From deterministic to stochastic approaches," Comput. Math. Appl., vol. 55, pp. 2329– 2337, May 2008.
3. S. Bert Brunekreef and S. T. Holgate, "Air pollution and health," THE LANCET, vol. 360, pp. 1233 – 1242, October 2002.
4. H. Bai, Y. Ma, X. Ai, H. Li, P. Liu, and D. Cang, "Chemical and morphological properties of particulate matter generated from coal fired circulating fluidized bed boiler," in Proceedings of the 2011 Third International Conference on Measuring Technology and Mechatronics Automation (ICMTMA '11), Volume 01 , (Washington, DC, USA), pp. 708–711, IEEE Computer Society, 2011.
5. G. R. Carmichael, A. Sandu, T. Chai, D. N. Daescu, E. M. Constan- tinescu, and Y. Tang, "Predicting air quality: Improvements through advanced methods to integrate models and measurements," J. Comput. Phys., vol. 227, pp. 3540–3571, March 2008.
6. R. S. Sokhi, H. Mao, S. T. G. Srimath, S. Fan, N. Kitwiroon, L. Luhana, J. Kukkonen, M. Haakana, A. Karppinen, K. Dick van den Hout, P. Boulter, I. S. McCrae, S. Larssen, K. I. Gjerstad, R. San Jose´, J. Bartzis, P. Neofytou, P. van den Breemer, S. Neville, A. Kousa, B. M. Cortes, and I. Myrtveit, "An integrated multi-model approach for air quality assessment: Development and evaluation of the oscar air quality assessment system," Environ. Model.Softw., vol. 23, pp. 268– 281, March 2008.
7. S. Finardi, R. De Maria, A. D'Allura, C. Cascone, G. Calori, and F. Lollobrigida, "A deterministic air quality forecasting system for torino urban area, italy," Environ. Model.Softw., vol. 23, pp. 344–355, March 2008.
8. K. W. Appel, R. C. Gilliam, N. Davis, A. Zubrow, and S. C. Howard, "Overview of the atmospheric model evaluation tool (AMET) v1.1 for evaluating meteorological and air quality models," Environ. Model.Softw., vol. 26, pp. 434–443, April 2011.
9. A. Zolghadri and F. Cazaurang, "Adaptive nonlinear state-space mod- elling for the prediction of daily mean PM_{10} concentrations," Environ. Model.Softw., vol. 21, pp. 885–894, June 2006.
10. G. Nunnari and F. Cannavo´, "A new cost function for air quality modeling," J. VLSI Signal Process. Syst., vol. 49, pp. 281–290, November 2007.
11. M. Oprea and A. Matei, "The neural network-based forecasting in environmental systems," WSEAS Trans. Sys.Ctrl., vol. 5, pp. 893– 901, Dec. 2010.
12. M. Oprea and A. Matei, "Applying artificial neural networks in environmental prediction systems," in Proceedings of the 11th WSEAS International Conference on Automation & Information (ICAI'10), (Stevens Point, Wisconsin, USA), pp. 110–115, World Scientific and Engineering Academy and Society, 2010.
13. C.-D. Neagu, N. Avouris, E. Kalapanidas, and V. Palade, "Neural and neuro-fuzzy integration in a knowledge-based system for air quality prediction," Applied Intelligence, vol. 17, pp. 141–169, July 2002.
14. E. Kalapanidas and N. Avouris, "Feature selection for air quality forecasting: a genetic algorithm approach," AI Commun., vol. 16, pp. 235–251, December 2003.
15. C. Mensink, F. Deutsch, L. Janssen, R. Torfs, and J. Vankerkom, "Comprehensive modelling of PM_{10} and $PM_{2.5}$ scenarios for Belgium and Europe in 2010," in Proceedings of the 6th international conference on Numerical methods and applications (NMA'06), (Berlin, Heidelberg), pp. 466–474, Springer-Verlag, 2007.
16. N. Kunzli, R. Kaiser, S. Medina, M. Studnicka, O. Chanel, P. Filliger, M. Herry, F. J. Horak, V. Puybonnieux-Texier, and P. Que´nel, "Public-health impact of outdoor and traffic-related air pollution: a european assessment," THE LANCET, vol. 356, pp. 795–801, September 2000.
17. G. Raimondo, A. Montuori, W. Moniaci, E. Pasero and E. Almkvist. "A machine learning tool to forecast PM10 level." In Proc. of the AMS 87th Annual Meeting, San Antonio, TX, USA. 2007.
18. M. Amann, I. Bertok, J. Borken-Kleefeld, J. Cofala, C. Heyes, L. Hoglund-Isaksson, Z. Klimont, B. Nguyen, M. Posch, P. Rafaj, R. Sandler, W. Scho¨pp, F. Wagner, and W. Winiwarter, "Cost-effective control of air quality and greenhouse gases in europe: Modeling and policy applications," Environ. Model.Softw., vol. 26, pp. 1489–1501, Dec. 2011.

19. T. Slini, A. Kaprara, K. Karatzas, and N. Moussiopoulos, “PM10 forecasting for thessaloniki, greece,” Environ. Model.Softw., vol. 21, pp. 559–565, April 2006.
20. C. Silibello, G. Calori, G. Brusasca, A. Giudici, E. Angelino, G. Fossati, E. Peroni, and E. Buganza, “Modelling of PM10 concentrations over Milano urban area using two aerosol modules,” Environ. Model.Softw., vol. 23, pp. 333–343, March 2008.
21. X. Liu, Y. Liang, W. Zhou, X. Peng, X. Tang, and Y. Wang, “Pollution characteristics of heavy metal in the PM10 of Nanchang City,” in Proceedings of the 2011 Second International Conference on Digital Manufacturing & Automation, ICDMA ’11, (Washington, DC, USA), pp. 859–862, IEEE Computer Society, 2011.
22. J. Hooyberghs, C. Mensink, G. Dumont, F. Fierens, and O. Brasseur, "A neural network forecast for daily average of PM10g concentrations in belgium," Atmospheric Environment, vol. 39, no. 18, pp. 3279-3289, 2005.
23. S. Arampongsanuwat and P. Meesad. "Prediction of PM10 using Support Vector Regression." In 2011 International Conference on Information and Electronics Engineering IPCSIT, Vol. 6, pp. 120-124. 2011.
24. K. M. Hamasha, M. S. Almomani, M. Abu-Allaban, and W. P. Arnott, “Study of black carbon levels in city centers and industrial centers in Jordan,” Jordan Journal of Physics, vol. 3, no. 1, pp. 1–8, 2010.
25. A. A. Al-Hasaan, Air Pollution Monitoring in Amman: A Collaborative Project. Royal Scientific Society of Jordan Report, 1989.
26. A. A. Al-Hasaan, T. F. Dann, and P. F. Brunet, “Air pollution monitoring in amman, jordan,” J. Air Waste Manage. Assoc., vol. 11, pp. 814–816, 1992.
27. T. Hussein, R. A. Al-Ruz, T. Petaja, H. Junninen, D. E. Arafah, K. Hameri, and M. Kulmala, “Local air pollution versus shortrange transported dust episodes: A comparative study for submicron particle number concentration,” Aerosol and Air Quality Research, vol. 11, p. 109119, 2011.
28. J. Koza, “Evolving a computer program to generate random numbers using the genetic programming paradigm,” in Proceedings of the Fourth International Conference on Genetic Algorithms, Morgan Kaufmann, La Jolla,CA, 1991.
29. J. R. Koza, Genetic Programming: On the Programming of Computers by Means of Natural Selection. The MIT Press, 1992.
30. Mouhammd Al-kasassbeh, Alaa F. Sheta, Hossam Faris, and Hamza Turabieh. Prediction of PM10 and TSP air pollution parameters using artificial neural network autoregressive, external input models: A case study in salt, jordan. Middle-East Journal of Scientific Research, 14(7):999–1009, 2013.
31. B. Radakovic and T. Krope, “The effect of ambient air pollution on human health,” in Proceedings of the 2nd IASME/WSEAS international conference on Energy and environment, EE’07, (Stevens Point, Wisconsin, USA), pp. 253–256, World Scientific and Engineering Academy and Society (WSEAS), 2007.
32. G. Koop, R. McKitrick, and L. Tole, “Air pollution, economic activity and respiratory illness: Evidence from canadian cities, 1974-1994,” Environ. Model.Softw., vol. 25, pp. 873–885, July 2010.

1/12/2014