Journal of Applied Agriculture and Biotechnology
2017 2(2): 38-45
ISSN (print): 2415-6728
ISSN (online): 2415-6736

Research article

An integrated approach to air pollution modeling from climate change perspective using ARIMA forecasting

Faiza Naseem[1]*, Audil Rashid[2], Touqeer Izhar[2], Muhammad Irfan Khawar[2], Shabana Bano[2], Aniqa Ashraf[2] and Mian Nazish Adnan[2]

HIGHLIGHTS

- Variation in air pollutants in urban environment can be predicted by climate variables
- Temperature, humidity and precipitation are strong predictors to forecast urban air quality
- Degradation in urban air quality of Rawalpindi city is a serious threat for public health

Authors' affiliation
[1]School of Mathematical Sciences, University of Science and Technology of China, Hefei, Anhui 230026, China

[2]EcoHealth Research Lab, Department of Environmental Sciences, PMAS Arid Agriculture University, Rawalpindi, Pakistan

***Corresponding author**
Faiza Naseem
Email: naseem@mail.ustc.edu.cn

How to cite
Naseem, F., A. Rashid, T. Izhar, M.I. Khawar, S. Bano, A. Ashraf and M.N. Adnan. 2017. An integrated approach to air pollution modeling from climate change perspective using ARIMA forecasting. J. Appl. Agric. Biotechnol., 2(2): 38-45.

Key words
Air Quality Index,
Urban Air Modeling,
AIRMA,
Climate Change,
Islamabad

ABSTRACT

Exposure to air pollutants and related health implications is challenging for policy makers to overcome research gaps and effectively manage urban environment. To connect this missing link, we analyzed impact of climate variables on air pollution status to identify key determinants essential for future projections. Findings show that meteorological variables (temperature, humidity, wind, solar radiation and precipitation) have definite influence on life cycle persistence of air pollutants in urban environment. This influence was evaluated using a modeling approach 'time series expert modeler' called ARIMA which has identified strong climate predictors capable of explaining variations in air pollutants in the model output (Ljung-Box statistics=81.78; stationary R^2=0.69; p<0.000). The results showed significant impact of temperature, humidity and precipitation on spatial variability of air pollutant concentrations. In developing countries such as Pakistan where lack of political will has caused problems in the enforcement of environmental regulations, degradation of urban air quality is a serious threat for public health. Therefore, reliance of climate-signals to predict air pollutant concentration and their variability is of great significance. We conclude that in order to overcome the limitations such as absence of national air quality monitoring system due to inefficient governments, universities in Pakistan need to initiate multidisciplinary research groups where expertise of mathematical modelers, geographers, meteorologists, GIS experts and public health scientists could be integrated. In this context, current study provides useful evidence to estimate likely impacts of climate change on air pollution assessments.

1. Introduction

Air pollution modeling is a subject of strong concern. Declining air quality in urban centers has raised serious management efforts in both the developed and developing countries (Kurt and Oktay, 2010). The most challenging aspect in this context is the development of most acceptable models based on real-time scenarios. This requires an integrated effort of experts from different fields to carefully analyze, model and predict future air pollution scenarios. Since policy regulations are based on comparisons between current and future conditions, the significance of accurate predictions and modeling has become center of attention. In most of the air pollution studies, Gaussian dispersion models are emphasized to develop physical basis of dispersion models. However, emission sources and the relevant atmospheric effects are less emphasized in particular impact of weather variables are generally obscure (Chelani *et al*., 2002; Karppinen *et al*., 2000). To overcome these limitations, statisticians and the software experts have innovated robust approaches that are extensively used in models owing to which, forecasting and discovering patterns of several key pollutant in the atmosphere are now possible (Ziomas *et al*., 1995; Polydoras *et al*., 1998). For this reason, huge effort has been made to bring improvements in time series prediction modeling (Faruk, 2010) including air pollution modeling (Pisoni *et al*., 2009). Substantial progress in this direction is attributed due autoregressive integrated moving average (ARIMA) model which is an important and most widely used time series model (Jiang *et al*., 2018).

This study aimed to identify relationships between climate variations and air pollution status in an urban area using integrated approach based on computer application and statistical tools of analysis. Using this approach, we tried to develop a time series model to predict air quality time series data and its application in air regulation decision making.

2. MATERIALS AND METHOD

In the present study, detailed air quality data was acquired from Pakistan Environmental Protection Agency (Pak EPA). For the information about meteorological parameters, Pakistan Meteorological Department (PMD) was contacted. These sources were consulted on regular basis for the acquisition of secondary data for urban areas of Islamabad city. However, we included in our analysis the information that extended over 24-months i.e. January 2009 to December 2010 to build air quality index and its subsequent utilization in time series analysis. The indicator parameters used to monitor air quality included NO_2 ($\mu g/m^3$); SO_2 ($\mu g/m^3$); O_3 (ppb); total suspended particulate matter TSP ($\mu g/m^3$) and pollen counts (in thousands). Although a number of other variables related to air pollution have been used to construct air quality indices in different studies (Cogliani, 2001; Van den Elshout *et al*., 2008) but we give priority to those variables that indicate the risk of adverse health effects in epidemiologic studies. On the other hand, the meteorological variables used in this study were as follows: Daily minimum temperature (T_{min}), daily maximum temperature (T_{max}), relative humidity (rh), wind speed (wsp), light intensity (lgi), and precipitation (ppt).

$$A_p = \left[\frac{(A_i - A_o)}{(TL_i - TL_o)}\right](RC_p - TL_o) + A_o \qquad \text{Eq. 1}$$

Where

A_p refers to as air quality index (AQI) for the pollution 'p'

RC_p = real-time ambient concentration of the air pollutant 'p'

TL_i = threshold limit given in Table 1 that is equal or greater than RC_p

TL_o = threshold limit given in Table 1 that is equal or less than RC_p

A_i = the air quality sub-index value corresponding to TL_i

A_o = the air quality sub-index value corresponding to TL_o

Prior to use ARIMA model, the assumptions regarding multi-collinearity and autocorrelation of variables were solved to minimize the error. The purpose of using ARIMA was to predict air quality of Islamabad urban area using meteorological variables as predictors. Since the seasonal effect was not in focus and the prediction was made on whole year basis therefore no constant terms was included in the model. The other assumption was that future value of the five variables used must be a linear function with numerous past observations and random errors. Once these assumptions were met, the further modeling approach involved three important steps i.e. identification of the model, estimation of parameters and diagnostic checking. We selected Akaike Information Criterion (AIC) that gave the best fit model. The mathematical algorithm for the AIC is as follows

$$\text{AIC} = n(\ln((2\pi \text{RSS})/n+1)+2m \qquad \text{Eq. 2}$$

where *m* is the order of non-seasonal autoregression number estimated in the model and RSS stands for sum of squared residuals.

$$(1-\Phi_1 B^1)y_{1t} = \delta_1 + a_{1t} \qquad \text{Eq. 3}$$

for Brisbane (Australia)

$$(1-\Phi_1 B^1 - B^5)y_{2t} = \delta_2 + a_{2t} \qquad \text{Eq. 4}$$

for Islamabad we employed the five variables, all were based on climate parameters.

3. RESULTS AND DISCUSSION

The air pollution status is an indicator of environmental health and deterioration in ambient air quality has direct consequences on public health (Ali *et al*., 2017). In Pakistan, over the period of last two decades, there is substantial increase in number of vehicles on roads and this trend of rising personal cars appears a prime cause of urban pollution and declining air quality in country (Rashid *et al*., 2018). The mean values obtained after developing the air quality index for Islamabad city show a slight variation in two years for different parameters analyzed. Index values for SO_2 and total suspended particulate matter from 2009 to 2010 indicate decline in air quality (Table 1). However, with respect to ozone concentration, there was not much difference observed for both the years (Table 2). The other noticeable parameters include oxides of nitrogen and pollen counts whose mean values were not available in literature and hence limited information was available to describe the category 'good' in terms of air quality. Despite of that, we have observed the significant difference between the concentration of other key parameters for the two years in Islamabad which shows a trend of rising air pollution in 2010 (Table 2).

Table 1. The mean air quality index values showing the variations among air pollutants with their threshold ranges for the year 2009

Description of index value	**NO_2 (µg/m^3)**	**SO_2 (µg/m^3)**	**O_3 (ppb)**	**TSP‡ (µg/m^3)**	**PC* (000s)**
Good [< 5]	4.73	2.09	1.76	3.79	—
Moderate [6 – 10]	7.86	6.08	6.11	8.73	9.79
Poor [11 – 15]	10.24	11.76	10.45	14.85	—
Very poor [16 – 20]	18.67	15.98	—	18.28	16.02
Severe [> 20]	23.12	—	20.32	—	21.54

‡ Total suspended particles (including PM_{10})
* Pollen count (in thousands)

Table 2. The mean air quality index values showing the variations among air pollutants with their threshold ranges for the year 2010

Description of index value	NO_2 (μg/m^3)	SO_2 (μg/m^3)	O_3 (ppb)	TSP‡ (μg/m^3)	PC* (000s)
Good [< 5]	—	3.90	1.63	—	—
Moderate [6 – 10]	7.35	—	—	9.67	7.02
Poor [11 – 15]	14.87	11.86	10.44	14.03	14.11
Very poor [16 – 20]	17.41	15.49	16.41	—	18.68
Severe [> 20]	29.67	23.52	22.65	26.83	30.52

‡ Total suspended particles (including PM_{10})
* Pollen count (in thousands)

Our results further indicate that there was no significant correlation between residuals obtained for individual air quality parameter (Table 3). Our results show that model is well fit with observations and it fulfills the assumptions of ARIMA such as independence and homoscedasticity. The normality of autocorrelated residuals a_t was also satisfied. The best-fit model based on diagnostic checking has identified temperature, humidity and light intensity as component of air quality that have significant influence on model prediction for the year 2009. For the following year, a climate variable i.e. light intensity was replaced precipitation.

Table 3. Correlation matrix for residuals of air quality parameters

Parameters	NO_2	SO_2	O_3	TSP	PC
NO_2	1				
SO_2	0.21	1			
O_3	0.15	0.19	1		
TSP	–0.07	–0.21	0.05	1	
PC	0.11	–0.03	0.11	–0.16	1

The trend identified in 2009 for the air pollution pattern indicate that the worst quality of air persisted for first 100 days as it shows peaks of poor to very poor air quality index (Figure 1). However, moderate to good air quality appeared in cycles later on particularly in the mid of the year when temperature approaches maximum. This probably has relevance from seasonal variations as during July and August, monsoon approaches due to which heavy rain fall causes lot of air pollutants to settle down. Moreover, the suspended particulate matter also has been observed in these days to a minimum (Ghauri *et al*., 2007). The region of South East Asia, in which Pakistan is located, has strong monsoon mediated weather systems that can bring substantial change in the overall patterns. Studies conducted in this region have highlighted the effects of summer, monsoon and post-monsoon winter seasons on air quality prediction (Kumar and Goyal, 2011). This trend of improved air quality during monsoon was also evident in 2010 (Figure 2).

The results of ARIMA model for a period of 24-months is given in Figure 3. The peaks obtained over time duration of 365 days for the years 2009 show that air quality was in the range of poor to very poor. Although trend of air pollution is distinguishable in low to high cycles throughout the year but the prediction of peaks based on ARIMA model were able to identify and explain these changes. Ljung-Box statistics (Table 4) and stationary R^2 has indicated considerable strength of model to predict the variations attributed by climate predictors. As most of the residual autocorrelation functions were observed within the confidence limit, it seems reasonable to believe that 24-months lag phase is good to predict patterns of variations and in air pollution on the basis of climate predictor variables temperature, humidity and light intensity (Table 4). The Ljung-Box statistics values (23.98, P=0.002) for these variables shows overwhelming evidence that model is in agreement with the observed data.

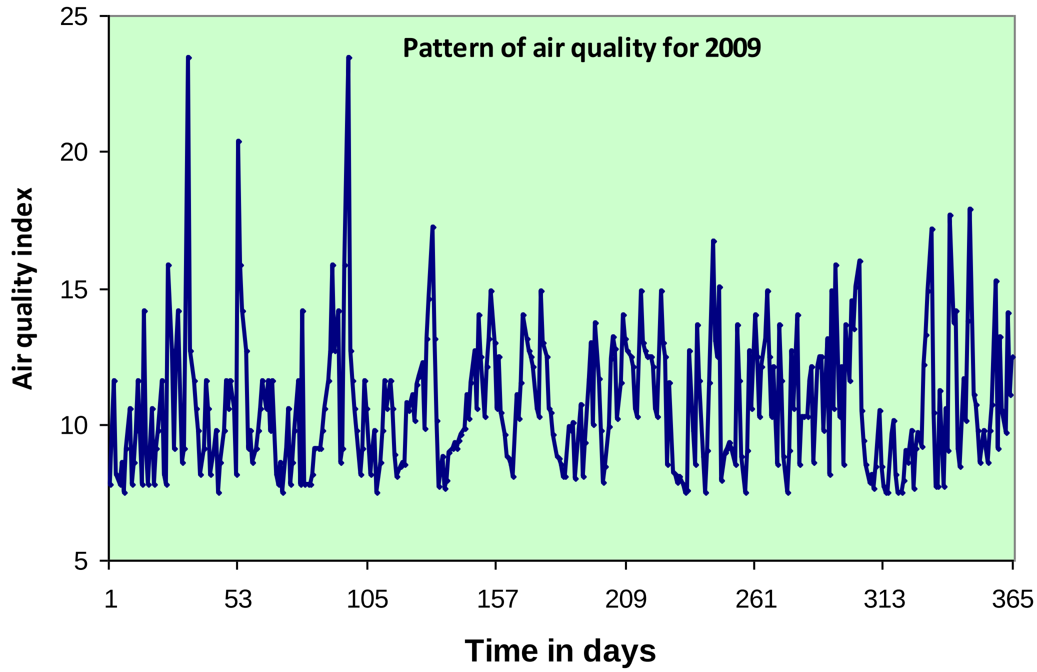

Figure 1. Index of air quality for Islamabad based on pollution estimates of five parameters for the year 2009

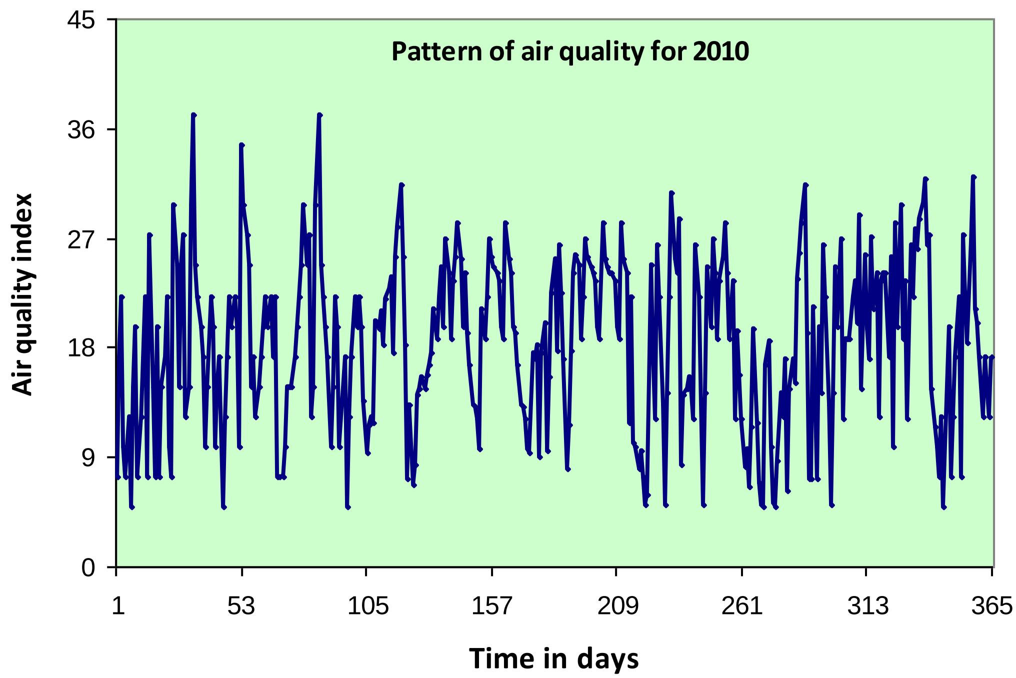

Figure 2. Index of air quality for Islamabad based on pollution estimates of five parameters for the year 2010

Similarly, for the year 2010, except precipitation, the other two variables were common and show good significant strength (R^2 0.5, *P*=0.005) in prediction the air quality (Table 4). Therefore, use of such weather related parameters could be effectively used in determining the future patters of air quality on time scale. The relationship between observed and model fit data for the year 2010 further revealed that most of the peaks obtained in ARIMA model has been explained (Figure 4). As reported earlier that ARIMA forecast air pollution (Siew *et al.*, 2008), our study indicates that the model is sufficiently strong to predict the air quality.

Table 4. Summary and fit statistics for the ARIMA models developed to predict air quality

ARIMA Model	Number of Predictors	Predictor variables	Model Fit statistics	Ljung-Box Q			Number of Outliers
			Stationary R^2	Statistics	DF	Sig.	
Fit 2009 Model_1	3	Temperature, humidity & light intensity	0.482	23.987	4	0.002	1
Fit 2010-Model_1	3	Temperature, humidity & precipitation	0.500	35.834	4	0.005	0
Overall time series strength			0.690	81.779	–	0.000	–

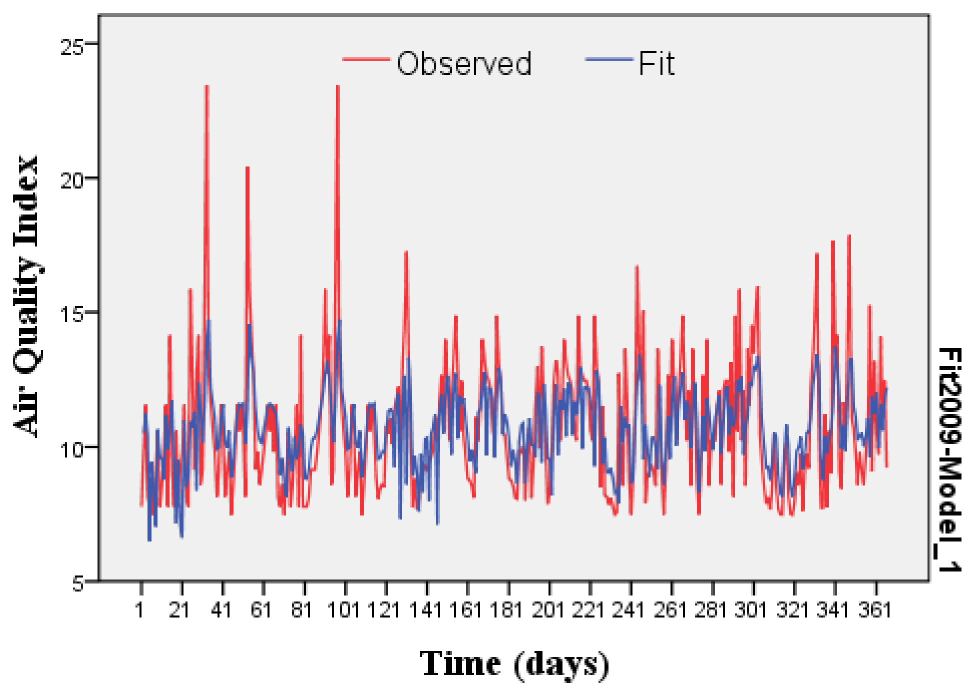

Figure 3. ARIMA-modeling of air quality index pattern after fitting temperature, relative humidity and light intensity as significant predictors

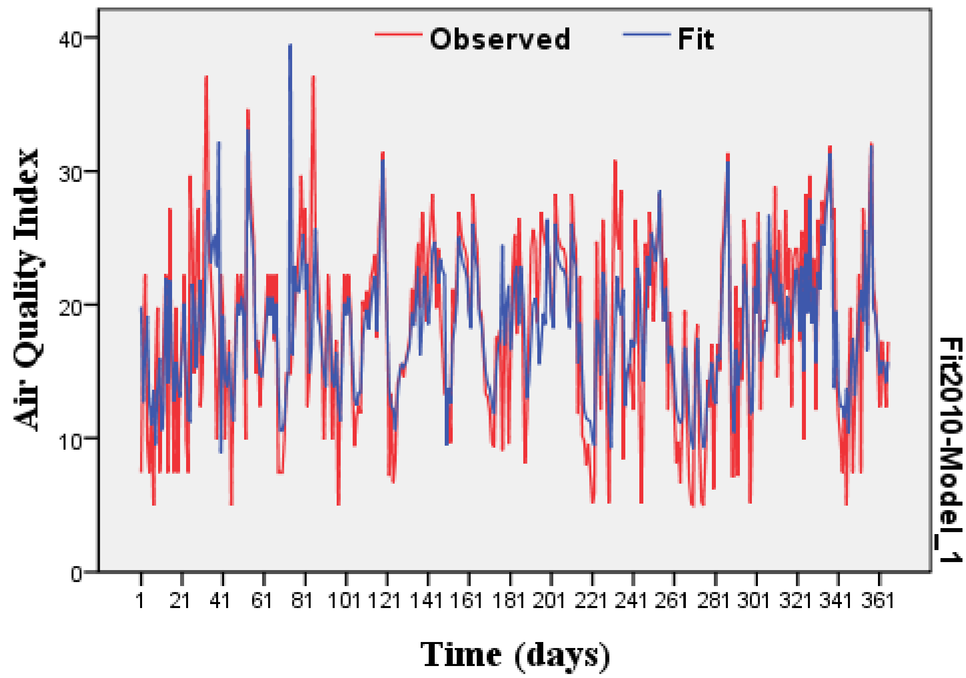

Figure 4. ARIMA-modeling of air quality index pattern after fitting temperature, relative humidity and precipitation as significant predictors

4. CONCLUSIONS

Accurate prediction of time series data is fascinating not only to develop innovative models but to provide sound scientific basis in order to formulate policies for environmental management. We are confronted with both the linear and non-linear patterns of air pollutant data so upon using ARIMA model, the strength of time series approach was fully demonstrated. Throughout the year the concentration of air pollutants in Islamabad urban areas have definite patterns that were noticeable in 2009 and 2010. The models for the years exhibited that uncertainties in weather related parameters could cause decline in air quality that could seriously affect the inhabitants of Islamabad city. Thus we stress upon continuous monitoring of meteorological variables on a wide range of locations especially where air pollution monitoring is carried out. This type of arrangement will not only improve the quality of raw input data but also enable the modelers of more accurately predict the air quality status of urban cities. The findings of this study have special significance for government because the predictors identified by ARIMA models can be effectively used by authorities for decision making. In the coming years, when temperature or rainfall patterns deviate from normal course, its influence on air pollutants can be predicted and this information is extremely important not only for environmental decision makers but also for general public to be well aware of the health risks imminently lying ahead of them and to adapt safety measure against air pollution related health problems.

REFERENCES

Ali, M. U., A., Rashid, B. Yousaf and A., Kamal 2017. Health outcomes of road-traffic pollution among exposed roadside workers in Rawalpindi City, Pakistan. Human and Ecological Risk Assessment: An International Journal, 23(6): 1330-1339.

Chelani, A. B., C.C. Rao, K.M. Phadke and M.Z. Hasan. 2002. Prediction of sulphur dioxide concentration using artificial neural networks. Environmental Modelling & Software, 17(2): 159-166.

Cogliani, E., 2001. Air pollution forecast in cities by an air pollution index highly correlated with meteorological variables. Atmospheric Environment, 35(16): 2871-2877.

Faruk, D. Ö., 2010. A hybrid neural network and ARIMA model for water quality time series prediction. Engineering Applications of Artificial Intelligence, 23(4): 586-594.

Ghauri, B., A. Lodhi and M. Mansha. 2007. Development of baseline (air quality) data in Pakistan. Environmental Monitoring and Assessment, 127(1-3): 237-252.

Jiang, S., C. Yang, J. Guo and Z. Ding. 2018. ARIMA forecasting of China’s coal consumption, price and investment by 2030. Energy Sources, Part B: Economics, Planning, and Policy, 1-6.

Karppinen, A., J., Kukkonen, T., Elolähde, M., Konttinen, T., Koskentalo and E. Rantakrans. 2000. A modelling system for predicting urban air pollution: model description and applications in the Helsinki metropolitan area. Atmospheric Environment, 34(22): 3723-3733.

Kumar, A. and P. Goyal. 2011. Forecasting of daily air quality index in Delhi. Science of the Total Environment, 409(24): 5517-5523.

Kurt, A. and A. B. Oktay. 2010. Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks. Expert Systems with Applications, 37(12): 7986-7992.

Pisoni, E., M., Farina, C. Carnevale and L. Piroddi. 2009. Forecasting peak air pollution levels using NARX models. Engineering Applications of Artificial Intelligence, 22(4-5): 593-602.

Polydoras, G. N., J. S. Anagnostopoulos and G. C. Bergeles. 1998. Air quality predictions: dispersion model vs Box-Jenkins stochastic models. An implementation and comparison for Athens, Greece. Applied Thermal Engineering, 18(11): 1037-1048.

Rashid, A., A., Irum, I. A., Malik, A., Ashraf, L., Rongqiong, G., Liu, H., Ullah, M. U. Ali and B. Yousaf. 2018. Ecological footprint of Rawalpindi; Pakistan's first footprint analysis from urbanization perspective. Journal of Cleaner Production, 170: 362-368.

Siew, L. Y., L. Y. Chin and P. M. J. Wee. 2008. ARIMA and integrated ARFIMA models for forecasting air pollution index in Shah Alam, Selangor. *Malaysian Journal of Analytical Sciences*, *12*(1): 257-263.

van den Elshout, S., K. Léger and F. Nussio. 2008. Comparing urban air quality in Europe in real time: A review of existing air quality indices and the proposal of a common alternative. Environment International, 34(5): 720-726.

Ziomas, I. C., D., Melas, C. S., Zerefos, A. F. Bais and A. G. Paliatsos. 1995. Forecasting peak pollutant levels from meteorological variables. Atmospheric Environment, 29(24): 3703-3711.