TIME SERIES ANALYSIS OF AMBIENT AIR QUALITY AT ITO INTERSECTION IN DELHI (INDIA)

Inderjeet Kaushik*1 and Rinki Melwani2

1. Department of Civil Engineering, Institute of Technology, Banaras Hindu University, Varanasi (INDIA)
2. Department of Civil Engineering, Indian Institute of Technology, Delhi (INDIA)

Received August 4, 2007    Accepted November 28, 2007

ABSTRACT

Air pollution is an important and popular topic of concerns at ITO intersection in Delhi as health impact concerns have been raised caused by vehicle exhausts and industrial exhausts in recent years. Sulphur dioxide (SO2), nitrogen dioxide (NO2), carbon dioxide (CO2) and respirable suspended particulate matters (RSPM) are major air pollutants caused by the dominant usage of diesel, petrol and CNG fuels by vehicles and industries. These major air pollutants impose harmful impact on human health. In this study, the varying trends of ambient air quality and the levels of related air pollutants are analyzed based on the database monitored at Bahadurshah Jafar marg near ITO intersection, air quality monitoring station. Seasonal Autoregressive Integrated Moving Average (ARIMA) approach, implemented by Box–Jenkins is used to forecast the levels of ambient air quality parameters. The performance evaluations of the adopted models are carried out on the basis of correlation coefficient (R²) and root mean square error (RMSE). The results indicate that the seasonal ARIMA model provide reliable and satisfactory predictions for the air quality parameters and expected to be an alternative tool for practical assessment and justification.

Key Words : ARIMA, Stationary series, Seasonal variation, ITO intersection, RMSE

INTRODUCTION

The alarming growth of atmospheric pollution has led many countries in the world to establish severe laws and regulations defining air quality and the required standard of emission levels. Particular attention is devoted to urban areas where the problems of atmospheric air pollution are especially worrying, because the output of pollutants are high and the number of people exposed to health hazards are growing constantly. In this context, many public administrations have installed several monitoring networks or stations able to carry out, in real time, qualitative and quantitative information about the characteristics of the environment of urban centers. Particular importance is given to both atmospheric pollution and air quality evaluation and forecasting1-3.

It is well known that air quality depends on different factors, including the population density, the volume of traffic, the energy demand and the physical characteristics of the

*Author for correspondence
territory (i.e. geographical conformation). The development of a coherent picture of national environmental trends and conditions requires the collection of sufficient data, the statistical analysis and integration of the information and the provision of complete, accurate and understandable presentations\(^2\).

The availability of measured environmental data useful in determining the air pollution level, is often not enough to describe the temporal and spatial behavior of all pollutant emissions (in most cases much of the data do not exist or are not available). Moreover, there are constant variations in pollution and meteorological conditions, so the overall picture of the environment is frequently not clear. For this reason, it is very important to integrate the available measurements with the evaluation of pollutant concentrations, estimated by suitable mathematical models allowing to represent the atmospheric process and to investigate on eventual missing and/or incorrect data\(^3\).

**Model Formulation**

Daily levels of air pollutants were analyzed using time series approach. Time series models have been extensively studied by Box and Jenkins\(^1\) and as there names have frequently been used synonymously with general ARIMA process applied to time series analysis and forecasting. Auto Regressive (AR) models were first introduced by Yule and later generalized by Walker, while Moving Average (MA) models were first introduced by Slutsky. Wold provided theoretical foundation for combined Auto Regressive Moving Average (ARMA) process\(^4\).

Box and Jenkins have effectively put together in a comprehensive manner, the relevant information required to understand and use time series ARIMA models. A detailed strategy for the construction of linear stochastic equation describing the behavior of time series was examined\(^5\).

Consider the function \(Z\) represents forecasted air pollutant level at time \(t\). \(Y\) is series of observed data of air pollutants at intersection at time \(t\). If series is stationary then it can be represented as:

\[
\nabla^q Z_t = \nabla^q Y_t \quad \ldots \ldots \(1\)
\]

Where \(\nabla\) is a back shift operator.

If series \(Y\) is not stationary then it can be reduced to a stationary series by differencing a finite number of times.

\[
\nabla^q Z_t = \nabla^q (1 - B)^d Y_t \quad \ldots \ldots \(2\)
\]

Where \(d\) is a positive integer, and \(B\) is back shift operator on the index of time series such that \(B^t Y_t = Y_{t-d}\) and so on. Thus further equation (2) can be simplified into following equation:

\[
(1-\Phi_1 B-\Phi_2 B^2- \ldots - \Phi_p B^p) Z_t = \theta_0 + (1-\Theta_1 B-\Theta_2 B^2- \ldots - \Theta_q B^q) a_t \quad \ldots \ldots \(3\)
\]

Where \(a\)'s a sequence of identically distributed uncorrelated deviates, referred to as "white noise".

Combining equations (2) and (3) yields the basic Box-Jenkins models for non stationary time series.

\[
(1-\Phi_1 B-\Phi_2 B^2- \ldots - \Phi_p B^p) (1-B)^d Y_t = \theta_0 + (1-\Theta_1 B-\Theta_2 B^2- \ldots - \Theta_q B^q) a_t \quad \ldots \ldots \(4\)
\]

Equation (4) represents an ARIMA process of order \((p,d,q)\).

Seasonal ARIMA model represented as follows for a stationary series i.e. differencing parameters \((d\) and \(d_s = 0)\) equal to Zero, used for forecasting air pollutants level.

\[
\nabla^p \nabla^s Z_t = \nabla^p \nabla^s Y_t \quad \ldots \ldots \(5\)
\]

Where \(p_s\) and \(q_s\) are the seasonal parameters corresponding to AR and MA process.

Model of type of equation (5) was fitted to given set of data using an approach consists of mainly three steps (a) identification (b) estimation (c) application (forecasting) or diagnostic checking. At the identification stage tentative values of \(p,d,q\) and \(p_s,d_s,q_s\) were chosen. Coefficients of variables used in model...
were estimated. Finally diagnostic checks were made to determine, whether the model fitted adequately describes the given time series. Any inadequacies discovered might suggest an alternative form of the model, and whole iterative cycle of identification, estimation and application was repeated until a satisfactory model was obtained.

**Study site and Data collection**

ITO intersection at Delhi was chosen for studying air pollutants level. Air pollutants data monitored on Bahadurshah Jafar marg near intersection approximately fifty meters away. At this intersection daily thousands of vehicles pass, railway line and power plant near by it increases the pollution level to a significant level. Therefore to analyze the pollution level, air pollutants levels were studied. A model was developed to analyze daily pollutants level using time series approach. To develop the model two years (January 2001 to December 2002) pollutants data was used.

**RESULTS AND DISCUSSION**

Daily air pollutants data was analyzed using time series analysis. Traffic flow varies daily; therefore on the basis of variation of traffic flow seven days i.e. week was selected as a period for the function. Seasonal ARIMA model was selected for forecasting of levels air pollutants. Statistics of models are shown in Table1. Since daily traffic flow varies which

<table>
<thead>
<tr>
<th>Air quality parameter</th>
<th>ARIMA model (pdq,p,d,q)</th>
<th>$R^2$</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SO$_2$</td>
<td>(201,202)</td>
<td>0.65</td>
<td>2.52</td>
</tr>
<tr>
<td>NO$_2$</td>
<td>(202,202)</td>
<td>0.60</td>
<td>10.78</td>
</tr>
<tr>
<td>RSPM</td>
<td>(302,302)</td>
<td>0.69</td>
<td>61.70</td>
</tr>
</tbody>
</table>

Comparisons of observed and predicted levels of air quality parameters SO$_2$, NO$_2$ and RSPM are shown below in figure 1-3 respectively.

![Figure 1: Comparison of observed and predicted levels of SO$_2$](image-url)
results into variation of air pollutants level. Therefore accuracies of these predictions further can be improved by taking into account daily traffic volume, emission of power plant etc.

**CONCLUSION**

The approach suggested in this paper enables environmental engineers to analyze and to characterize concentration measurements of air pollutants with time, in very high road-traffic areas. In particular, such approach allows to carry out a good estimate of analyzed substances and to reconstruct the eventual missing data in time domain. By using such techniques, it is possible to work with a continuous and valid set of data, with the further aim of reducing measurement uncertainty.

**RECOMMENDATIONS**

In this model we have correlated air pollutants level with time only, considering a week as period. There are many variables which should be taken into consideration for
more accurate forecasting like traffic volume with their classifications and corresponding emissions, and other factors like emission by trains passing nearby railway line and emissions of power plants nearby which release a major portion of pollutants into air. Collection of these factors allows us to study the analysis of impacts that how composition and volume of traffic flow affects air pollution level and to assess the impact of CNG vehicles replacing petrol and diesel vehicles.

REFERENCES

SAVE THE ENVIRONMENT

● Good Environment is good health
● Air Pollution causes health hazards
● Recycle every drop of Water
● Environment is gods gift, preserve it
● Grow tree, Feel the Environment free