FORECASTING OF AMBIENT AIR QUALITY FOR VICTTORIYA HOSPITAL AREA BENGALURU

Vijayakumar¹, Dr. Maya Naik² and Col Rajshekar S Hiremath³

¹PG Student Civil Engg Dept. BMSCE, Bengaluru, India
²Professor Civil Dept. BMSCE, Bengaluru, India
³Phd Student Civil Engg Dept. BMSC, Bengaluru, India

Abstract- The statistical models are used in the prediction of air pollution parameter by using different models. The Autoregressive Integrated Moving Average (ARIMA) model is used in the analysis of SO₂ NO₂ and PM₁₀. The collected data is to be maintained in chronological order. The diagnostic check of the model is carried out by using ACF and PCF values. Based on the ACF and PCF values the order of the model is identified, which give the best fitting model for that particular air pollution parameter. The SO₂ showing the decreasing pattern with time and is much more within the permissible limit. The NO₂ not showing any much more variation but is within the permissible limit. The PM₁₀ is initially high and has value more than the permissible value with showing increasing in its trend.

KeyWords- Autoregressive Integrated Moving Average, Autocorrelation Function, Partial Autocorrelation Function.

I. INTRODUCTION

The air pollution is higher in its level particularly in developed and developing country, the reasons for this is mainly due increasing in number of Industries, Vehicles and in moreover the increase in population. Development of such statistical tools and are used in forecasting of air pollution parameter will became a communication between statisticians and to the Environmental scientist. By using the forecast value environmental scientists can analyse and come with a more ideas about the situation and behaviour of the air parameters. Bengaluru is the metropolitan city and now day it facing worst situation in the air quality status.

In forecasting of future value mainly done in the two way such as one is numerical way and second one is statistical way. The most used methods out of these two are mainly of statistical models, these models plays vital rule in forecasting the future values. The major terms used in the models are explained as below. As for Bengaluru, the research and development of time series modelling and forecasting for monitoring air pollution are still limited.

The Box & Jenkins methodology has been used and ARIMA (Auto-Regressive Integrated Moving Average) technique has been adopted for analysis. The modeling has been performed on “R” software. The following brief definitions will enable a better understanding of the Time Series Analysis and reasons behind selection of particular models for the same:

Box and Jenkins Methodology- This is the basic principle which is used behind the forecasting of ARIMA model. The procedures which is used in the model is mainly consists of model selection, parameter estimation and model checking. The steps which are included in this process are explained as below:

Identification of the parameters of the model: The ARIMA model uses a univariate time series analysis. The ACF and PCF values are used in finding order of the model. The AR process consists of lagged dependent variables and MA is a function of the number of lagged errors terms. The series is to be differentiated to d times to convert from non-stationary parameter into the stationary parameter and I equals d for that series.
Estimation of the parameters: This can be done in two ways, the first one is Maximum Likely hood estimation and the second one is the conditional maximum likely hood estimation (MLE). The numerical optimization techniques are used in MLE for estimating the parameter and which is preceded by a simple OLS regression. The ARIMA model analysis uses MLE for the parameter estimation. The best fitting model can be identified by using AIC value.

Performing diagnostic checks: This test is carried out to determine uncorrelated things, which is present in the data. It is found that in reality, every data which consist some noise that is identified. This noise is usually called as white noise. Therefore, if the statistics values are close one which shows good results.

Forecasting: This is the final step which is carried out to predict the future values. This can be done in two ways such as the static forecasting and dynamic forecasting. The dynamic forecast model which are more complex and also need chemical relation with each other in the data and also done for longer horizon data. In my project I basically used statistics forecasting for the ARIMA model. Because it is only uses single time period ahead of the series.

II. OBJECTIVE AND SCOPE

The main objective of this paper is to predict the future values SO2 NO2 PM10 by using ARIMA MODEL.ARIMA MODEL is a statistical tool which is used to identify the time series pattern of the data and finally identified the best suitable order for that particular data based on the AIC and BIC values.

III. DATA COLLECTION AND DESCRIPTION

Karnataka State Pollution Control Board regulates the air pollution stations at different locations that provide air pollution data. For data analysis, the below location is considered for the forecasting the future value of SO2 NO2 and PM10 (RSPM). The locations are mentioned below

i. Victoria hospital located at 12˚ 57’ 47.2¨ N latitude, 77˚ 34’ 30.3¨ E longitude, is a sensitive area.

IV. RESULTS

Experiments were carried out to model the maximum daily SO2 NO2 and PM10, for the collected data from the KSPCB. The one year daily data is carried out to predict the future value for the Air pollution parameter.

Using ARIMA Model the Box- Jenkins Transfer function approach is used based on the maximum likelihood occurrence. Figure 1 Figure 2 and Figure 3 show variation of raw air pollution parameter data with time for SO2 NO2 and PM10 respectively. The Chronological pattern of the data is maintained by filling missing data with the nearest neighbouring value. In figure 1 the SO2 is almost constant which not varying as such for the last month. The NO2 shows increase in the starting month and decreasing in the middle of the year and slight increasing at the end of the year. The PM10 have showing varying a lot with the time and it exceeding the value more than the standard value.
V. TIME SERIES ANALYSIS FOR VICTORIYA HOSPITAL AREA

5.1 Time series Analysis for SO$_2$

Figure 4: SACF and SPCF plots for SO$_2$
The above figure shows the variation of SACF and SPACF with a time interval of in logarithmic way for the past one year data. By using these values the best fitting model is selected for the given data. The ARIMA (0,1,1) model which represents the best model for the SO$_2$ parameter. The values which are generated by the ARIMA model are given in the Table-1.

The predicted future value of SO$_2$ using the formula is given below:

$$X_t = \hat{\phi}_1 X_{t-1} + \hat{\phi}_2 X_{t-2} + a_t + \hat{\theta}_1 a_{t-1} + \hat{\theta}_2 a_{t-2} + \hat{\theta}_3 a_{t-3}$$

| $\hat{\phi}_1$ | $\hat{\phi}_2$ | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\theta}_3$ | $\hat{\sigma}^2$ | AIC |
|----------------|----------------|----------------|----------------|----------------|----------------|-----|
| 1.6336         | 0.9419         | 2.2371         | 1.9683         | 0.6174         | 0.2089         | 471 |

**Figure 5: Diagnostic test of model for the SO$_2$**

The In figure-11, the diagnostic check is carried out to identify the sequence of standard residuals, K values varying from 2 to 15 for the p-values of the Lung-Box test. Therefore from the above figure-5, we can conclude that which satisfies the assumptions of the model.

**Table 1: Forecast value of SO$_2$ for the lead time of 6 days**

| Time  | 1/1/16 | 2/1/16 | 3/1/16 | 4/1/16 | 5/1/16 | 6/1/16 |
|-------|--------|--------|--------|--------|--------|--------|
| SO$_2$| 1.9966 | 1.9939 | 1.9923 | 1.9926 | 1.9937 | 1.9961 |

**Figure 6: Forecast value of SO$_2$ for the lead time of 6 days for the fitted model**

The above graph the SO$_2$ concentration µg/m$^3$ is in y direction and months in x direction represents, the forecasted value of SO$_2$ with the help of prediction model “R” software. The graph shows the SO$_2$ trend over the past 12 months as well as the forecasted value of 6 days from 1$^{st}$
January 2016 to 6th January 2016. The blue line indicates the forecasted value and the values are almost follows the same pattern of past data. The forecasted values are well within the Limits prescribed by the Pollution Control Board.

5.2 Time series analysis for the NO₂

![Figure 7: SACF and SPCF plots for NO₂](image)

The above figure represents the sample auto correlation function and sample partial auto correlation function plots for given past data (1year data), with the help of above graph the type of model will be selected. For the NO₂ parameter, ARIMA (0,1,2) Model is suitable. The values obtained from the ARIMA models which have shown in the below table are used to predict the future value of conductivity parameter using the formula: 

\[ X_t = \hat{\phi}_1 X_{t-1} + \hat{\theta}_1 a_{t-1} \]

The Model is  

|   | \( \hat{\phi}_1 \) | \( \hat{\theta}_1 \) | \( \hat{\sigma}^2 \) | AIC |
|---|------------------|------------------|------------------|-----|
|   | 0.3471           | -0.8569          | 31.91            | 2298.24 |

![Figure 8: shows the SACF and SPCF plots for NO₂](image)
The above graph shows the diagnostic tools in one display, a sequence of standardized residuals, the sample autocorrelation function of the residuals, and p-values for the Lung-Box test statistic for a whole range of values of K from 2 to 15. From all the three plots, we can conclude that there are no violations in the assumptions.

Table 2: Forecast value of NO$_2$ for the lead time of 6 day

| Time       | 1/1/16 | 2/1/16 | 3/1/16 | 4/1/16 | 5/1/16 | 6/1/16 |
|------------|--------|--------|--------|--------|--------|--------|
| NO$_2$     | 25.929 | 25.525 | 26.2423| 26.245 | 26.4983| 26.76209|

The above graph represents the forecasted value of NO$_2$ using Computation prediction model of “R” software. The graph shows the Conductivity trend over the past 12 months as well as the forecasted value of 6 days from 1$^{st}$ January 2016 to 6$^{th}$ January 2016. The blue line indicates the forecasted value and the values are almost follows the same pattern of past data. The forecasted values are well within the Limits prescribed by the Pollution Control Board.

5.3. Time series analysis of Victoriya hospital for PM$_{10}$

The above figure shows the variation of SACF and SPACF with a time interval of in logarithmic way for the past one year data. By using these values the best fitting model is selected for the given data. The ARIMA (1,1,1) model which represents the best model for the PM$_{10}$
parameter. The values which are generated by the ARIMA model are given in the Table-3. By using these values we can get a clear idea about the coming time pattern and value of the air pollution parameter. These values are used to predict the future value of PM\textsubscript{10} parameter by using the below formula \( X_t = \hat{\phi}_1 X_{t-1} + a_t + \hat{\theta}_1 a_{t-1} \), ARIMA (1,1,1)

| \( \hat{\phi}_1 \) | \( \hat{\theta}_1 \) | \( \hat{\sigma}^2 \) | AIC |
|------------------|------------------|------------------|-----|
| 0.3310           | -0.7627          | 1550             | 3711.35 |

Figure 11: SACF and SPCF plots for PM\textsubscript{10}

In figure-11, the diagnostic check is carried out to identify the sequence of standard residuals, K values varying from 2 to 15 for the p-values of the Lung-Box test. Therefore from the above figure-11, we can conclude that which satisfies the assumptions of the model.

Table 3: Forecast value of PM\textsubscript{10} for the lead time of 6 day

| Time   | 1/1/16 | 2/1/16 | 3/1/16 | 4/1/16 | 4/1/16 | 5/1/16 |
|--------|--------|--------|--------|--------|--------|--------|
| PM\textsubscript{10} | 230.722 | 231.391 | 231.612 | 231.686 | 231.713 | 231.718 |

Figure 12: Forecast value of PM\textsubscript{10} from ARIMA(1,1,1)
The above graph represents variation of PM$_{10}$ with respect to time for the last one year and also predicted the future value of the preceding 6 days, which is from the 1$^{st}$ January 2016 to 6$^{th}$ January 2016 by using “R” software. The blue line which indicated in the figure-12 shows the forecasted values and these values almost follows the same pattern of past data.

VI. CONCLUSION

From the above graphs it clearly shows that the SO$_2$ and NO$_2$ showing a decreasing pattern and these values are so less than the permissible value prescribed by the KSPCB. The standard value for the SO$_2$, NO$_2$, and PM$_{10}$ is given as 80µg/m$^3$, 80µg/m$^3$ and 100 µg/m$^3$ on daily average bases, as per the NAAQS. The PM$_{10}$ values it was initially high at month of starting 2015 and showing a consistent decreasing pattern with showing a peak value at certain intervals. By using R software the best fitting ARIMA model is identified by using AIC values. Higher the AIC values better in the obtained result. The AIC values for the SO$_2$, NO$_2$ and PM$_{10}$ out of these three, the NO$_2$ and PM$_{10}$ giving a best result as compared to SO$_2$.Such kind of these models will helps to take decision and make important steps in minimizing the predominant air pollution parameter. The hospitals and the surrounding area is to be treated as very sensitive with respect to health issues so it’s very necessary to predict the air quality in time to time and take major steps to control in their magnitude. Therefore the time series model in predicting future values plays a important rule in monitoring and controlling the air quality conditions. ARIMA(0,1,1), ARIMA(0,1,2) and ARIMA(1,1,1) are the best model which suits, to forecast the oxide of sulphur and nitrogen and PM$_{10}$ respectively for the Victoria hospital area.

REFERENCES

[1] Renhao Jin, Tao Liu, and Sha Wang “ The application of ARIMA model on 2014 Air Quality Index in Miyun County, Beijing, China”, Global Journal of Mathematics, Vol. 5, No. 2, ISSN: 2395-4760, 2015.

[2] Pooja Chauhan, “An overview on air pollution in India at present scenario”, ISSN- 2350-0530(O) ISSN- 2394-3629, 2015.

[3] Kartikeya Jha, Nishita Sinha, Shriniwas S. Arkatkar, Ashoke K. Sarkar, “Time Series Analysis of Traffic Data Using AADT Data from California, USA”, MTTER Volume 2, Issue 1, PP. 60-67, 2013.

[4] MdHisyam Lee, NurHaizum Abd. Rahman, “Seasonal ARIMA for Forecasting Air Pollution Index: A Case Study”, American Journal of Applied Sciences, ISSN 1546-9239, 2012.

[5] Ian Mitchell Marshall, Gil Nonato Santos, “Air Pollution in Bangalore, India: A Six-Year Trend and Health Implication Analysis”, 2014.

[6] G. Saffarini and S Odat, “Time Series Analysis of Air Pollution in Al-Hashimeya Town Zarqa, Jordan”, Jordan Journal of Earth and Environmental Sciences, Volume 1, Number 2, ISSN 1995-6681, 2008.