

Time Series Analysis to Forecast Air Quality Indices in Thiruvananthapuram District, Kerala, India

Naveen V¹. and Anu N².

¹PG Student, Environmental Engineering & Management, Department of Civil Engineering, UKF College of Engineering & Technology, Kollam-691302, Kerala, INDIA.

²Assistant Professor, Department of Civil Engineering, UKF College of Engineering & Technology, Kollam-691302, Kerala, INDIA.

Highlights

- Air Quality Index (AQI) of Thiruvananthapuram city has been calculated
- AQI forecasting using ARIMA and SARIMA model were introduced
- Error between actual and predicted AQI has been reduced using optimization technique

ABSTRACT

Deterioration of air quality is an important issue faced by many cities in India. The increase in the number of vehicles, unrestrained burning of plastics, unacceptable construction and demolition activities and industrial activities are the main reasons for this deterioration. So it is necessary to assess the effectiveness of air quality monitoring programs for planning air pollution control actions by analyzing the trends in air quality regularly. In this study, the varying trends of ambient air quality were analyzed and forecasted in terms of Air Quality Index (AQI) based on the database monitored at different monitoring stations in Thiruvananthapuram District, Kerala, India. The air quality data from the Kerala State Pollution Control Board (KSPCB) shows that the responsible pollutant for AQI in all these stations were Respirable Suspended Particulate Matter (RSPM) due to its abundance in the atmosphere. By forecasting, we can predict the future air quality in terms of AQI or individual pollutants in order to reduce the pollutant concentration and exposure to air pollutants. For air quality forecasting, Auto Regressive Integrated Moving Average (ARIMA) and Seasonal Auto Regressive Integrated Moving Average (SARIMA) method was used. ARIMA models gave satisfactory results than the SARIMA models and these results can be combined with other models to create more accurate results.

Keywords: Air Quality Index, Air Quality Forecasting, ARIMA Model, Box-Jenkins Method, SARIMA Model, Time Series Analysis.

Abbreviations

AIC	Akaike Information Criterion
AICc	Akaike Information Criterion with a correction
AQI	Air Quality Index
AQIs	Air Quality Indices
ARIMA	Auto Regressive Integrated Moving Average
ARMA	Auto Regressive Moving Average
CPCB	Central Pollution Control Board
KSPCB	Kerala State Pollution Control Board
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
RMSE	Root Mean Square Error
RSPM	Respirable Suspended Particulate Matter
SARIMA	Seasonal Auto Regressive Integrated Moving Average

I. INTRODUCTION

Air pollution is one of the major issues affecting human health, plants, animals, agricultural crops, and ecosystem. Air quality is the state of air around us and ambient air quality is the quality of outdoor air in our surrounding environment. Certain concentrations of substances like Particulate matter (PM₁₀ and PM_{2.5}), Gases (Nitrogen oxides, Sulphur oxides and Carbon monoxide), Radioactive materials etc. can cause adverse effects on human beings, animals and environment. This atmospheric state can be defined as air pollution (Rao, 2006). Air pollution can also cause severe damages to buildings, sculptures and vegetation (Sarasamma and Narayanan, 2014). Regardless of the air pollution problems, the result of these processes increases the demand for energy and India experienced a substantial increase in fossil fuel consumption owing to the rapid growth in industrialization (Sinha and Ratogi, 2016). Due to the increased vehicular emissions, industrial emissions and combustion emissions, the

quality of air in the urban areas are deteriorating at a greater rate than in the rural areas. Since the number of vehicles are increasing day by day, the emissions from them are increasing rapidly and it is the main contributor of air pollution (Mitra and Sharma, 2002). As a result, the exposed inhabitants have a negative health effects due to the deteriorated air quality. Allergies, asthma, cancers, even death to humans and animals, and damage to plants are the main problems due to this deterioration in air quality (Lee et al., 2014; Daniel, 2014).

Due to the increased problems due to air pollution, the developed countries as well as the developing countries have increased the public awareness of air quality. Increasing and higher concentrations of air pollution can be life intimidating and thereby air pollution became one of the major issue in the world (Kurt and Oktay, 2010). Thus the authorities in these countries are regularly monitoring and forecasting the particulars about the different air pollutants in terms of air quality index. Forecasting or prediction of concentrations one or more air pollutant species in the atmosphere can be done by using air quality models.

Future air quality can be analyzed in terms of air pollutants and air quality index by using time series models. Air quality model is a significant tool in monitoring and controlling air pollutants like Carbon monoxide (CO), Particulate matter (PM₁₀ and PM_{2.5}), Sulphur dioxide (SO₂), Nitrogen dioxide (NO₂), Ozone (O₃) etc. The general way in describing air quality status is by expressing it in Air Quality Index (AQI) or Air Pollution Index (API) values (Kumar and Goyal, 2011). A simple generalized method which combines the AQI, scales and terms has been used for several years in order to assess the impact of deteriorated air quality on human health (Nur et al., 2016).

Air quality indices (AQIs) was suggested by US Environmental Protection Agency (USEPA) and it has been used in many cities to emphasize the severity of air pollution and adverse health risks due to this deterioration of air quality (Environmental Protection Agency, 2003). The variations in air quality can be straightforwardly recognized with the reference values in the air quality index system. This approach used in air quality management helps to form better air quality and for making future planning (Nagendra et al., 2007). Therefore, short-term forecasting of air quality index is mostly based on the time series analysis techniques (Argiriou, 2007). Lots of research has been conducted related to the air quality mostly concentrating on the main air pollutants like PM₁₀, PM_{2.5}, CO, NO₂, SO₂ and O₃ (Kumar and Goyal, 2011; Brunelli et al., 2007;

Ghazali et al., 2010; Kumar and Jain, 2010; Vlachogianni et al., 2011; Begum et al., 2013).

Forecasting of air quality helps to inform the public about the air quality to take necessary safety measures, to alert the authorities, industries to take necessary action and measures for reducing the emission rates. This will help further to reduce and avoid the exposure to air pollutants. Gaussian dispersion models are widely used for predicting air quality in many air pollution studies. Although the Gaussian dispersion models have some physical base, the thorough information about the various air quality parameters and source of pollutants were unknown (Chelani et al., 2002). Due to the accuracy of existing time series forecasting techniques, time series models are important in many decision making processes. Air quality forecasting is very consistent and effective in controlling the measures and can be proposed as a preventive action for regulations that are to be required (Sansuddin et al., 2011). According to many time series forecasting researches claimed that the predictive performance in combined models have been improved. Combined models are used when a single model fails to recognize the true data generating process or when a single model may not be adequate to recognize the characteristics of the time series data. In order to achieve more accurate results, hybrid models can be adopted. The purpose of these hybrid models is to reduce the risk of using an unsuitable model by combining several models together. Hybrid model of artificial neural network can be obtained using autoregressive integrated moving average (ARIMA) models and this hybrid model can produce more accurate results than the artificial neural network (Khashei and Bijari, 2010).

For public to easily be aware about how bad or good air quality for their health and to aid in data interpretation for decision making processes associated with air quality management and pollution mitigation measures, AQI is an important aspect. In this study, time series analysis has been used to analyze and forecast the varying trends of ambient air quality in terms of AQI in Thiruvananthapuram District, Kerala, India based on the database monitored at different air quality monitoring stations during the period of 2012-2015.

II. MATERIALS AND METHOD

2.1 Study Area

Thiruvananthapuram District in Kerala, India was formed on July 1, 1949 and it is the administrative capital of the Kerala state (District profile of Thiruvananthapuram, 2017). The coordinate of the region is 8.5241° N, 76.9366° E and is spread over an area of 2192 km² with 3,301,427

inhabitants as per the 2011 census report. Thiruvananthapuram is the second populated District in the Kerala State and among the 3,301,427 inhabitants, 1,529,831 inhabitants are in the rural area and 1,771,596 inhabitants are in the urban area (District Census Handbook, 2011). Thiruvananthapuram district consists of four air quality monitoring stations and they are classified into residential, rural, sensitive and industrial categories as shown in Table 1 and these stations are controlled by Kerala State Pollution Control Board (KSPCB) (Water and Air Quality Directory, 2015) shown in Figure 1.

(Figure 1)

(Table 1)

2.2 Air Quality Data

In this study, the monthly air quality data of Nitrogen dioxide (NO₂), Sulphur dioxide (SO₂), Suspended particulate matter (SPM) and Respirable particulate matter (RSPM) over a period of 2012-2015 at four stations in Thiruvananthapuram District, obtained from Kerala State Pollution Control Board has been used.

2.3 Air Quality Index

Air quality index (AQI) provides information on the levels and likely health risks of air pollution presented in an easy, understandable format (Nagendra et al., 2007). One of the most basic air quality indices to come out in literature was proposed by Marvin, H. Green, in 1966. AQI formulated by Green (known as Green Index), was based on two parameters, viz. Sulphur dioxide and Coefficient of haze (Green, 1966). Green index had limited applications because it was calculated as the arithmetic mean of two sub-indices. During the design of air quality index system, the following aspects are considered: (i) number of pollutants, (ii) mathematical functional relationship between the pollutant concentration and corresponding index, (iii) health criteria and (iv) description categories like good, moderate, poor etc. (Sharma et al., 2003a). Due to the specific design of AQI, a minimum of three pollutants is sufficient to calculate the AQI. Two steps are primarily involved in the formulation of AQI and they are: formation of sub-indices for each pollutants and aggregation of sub-indices to get an overall air quality index. Air quality index values of pollutants will be calculated and from those index values, the maximum sub-index or individual pollutant AQI value is reported as the AQI. The two reasons for the selection of maximum value are: (i) eclipsing (situation in which air pollution is underestimated by AQI) and ambiguity (situation in which air pollution is overestimated by AQI) can be eliminated, (ii) health based index cannot be

combined or weighted due to the lack of information about the health effects of combination of pollutants (synergistic effect). There are numerous AQIs used all around the world and the single thing common in air quality indices developed after 1990 is that all are based on maximum sub-index values. Descriptor categories for AQI adopted in India as Good (0-100), Moderate (101-200), Poor (201-300), Very Poor (301-400) and Severe (401-500) (Sharma et al., 2003a; Nagendra et al., 2007), are summarized in Table 2. As per Central Pollution Control Board (CPCB) of the Indian Ministry of Environment and Forests, the revised descriptor categories for AQI as Good (0-50), Satisfactory (51-100), Moderate (101-200), Poor (201-300), Very Poor (301-400) and Severe (greater than 400) (CPCB National Air Quality Index Report, 2014), are summarized in Table 3.

(Table 2 and Table 3)

2.4 ARIMA and Seasonal ARIMA Model

ARIMA is the abbreviation for Autoregressive Integrated Moving Average model, which consists of Autoregressive (AR) models, Integrated (I) models and Moving average (MA) models (Cryer and Chan, 2008). Yule introduced Auto Regressive (AR) models first and later Walker generalized the AR models. Moving average (MA) models were first developed by Slutsky and Wold provided theoretical base for combined Auto Regressive Moving Average (ARMA) process. Later, Box and Jenkins extensively studied the time series approach and effectively combined the relevant information to use time series ARIMA models (Box and Jenkins, 1970). ARIMA models are a class of complex linear models that is capable of representing stationary as well as non-stationary time series. These models do not involve independent variable in their structure and make use of the data in the time series to develop forecasts. For stationary time series data, AR and MA models are appropriate, while for non-stationary time series data, I models are required (Cryer, 1986). For determining the best model for time series data by Box-Jenkins methodology, tentative identification, parameter estimation and diagnostic checking procedures must be considered (Hanke and Wichern, 2008). An ARIMA model forecasts a value in a response time series as a linear combination of its own previous values, previous errors and current values. It uses an interactive approach of detecting a suitable model from a wide ranging class of models and the selected model is then checked against the historical data to see if it precisely describing the time series. ARIMA models can be extended to include seasonal components of a time series data,

thereby forming seasonal ARIMA model (SARIMA). ARIMA and SARIMA models are additions of ARMA model in order to include more accurate dynamics, i.e., non-stationarity in mean and seasonal patterns.

2.5 Methodology

The purpose of this study is to analyze the air quality index using the time series approach and the approach used is based on Box-Jenkins method which is referred as Autoregressive Integrated Moving Average (ARIMA) model. Because of the effective results in forecasting field, a lot of researchers use this model in many areas of research (Khashei and Bijari, 2010). Air quality index is calculated based on the AQI formula (USEPA, 1999) given below:

$$I_p = \left[\frac{(I_{Hi} - I_{Lo})}{(BP_{Hi} - BP_{Lo})} \right] [C_p - BP_{Lo}] + [I_{Lo}] \dots(2.1)$$

where,

- I_p Individual AQI for pollutant 'P'
- C_p Daily mean concentration of pollutant 'P'
- I_{Hi} Sub index value or individual AQI value corresponding to BP_{Hi}
- I_{Lo} Sub index value or individual AQI value corresponding to BP_{Lo}
- BP_{Hi} Breakpoint or nearby higher value greater than or equal to C_p
- BP_{Lo} Breakpoint or nearby lower value less than or equal to C_p

The sub index values and breakpoint values can be obtained directly from Table 3 based on the observed pollutant concentration at the monitoring station.Box-Jenkins method or ARIMA model consists of Autoregressive (AR) models, Integrated (I) models and Moving average (MA) models (Cryer and Chan, 2008). If 'Y' is a variable and 't' is the date, then the value of variable 'Y' at time 't' is 'Y_t'. Difference equation is an expression relating a variable 'Y_t' to its previous values. The linear first order difference equation, which expresses 'Y_t' as a linear function of 'Y_{t-1}' and 'w_t' can be written as in Equation 2.2 given below.

$$Y_t = \phi Y_{t-1} + w_t \dots(2.2)$$

when, $t = 0$, then $Y_0 = \phi Y_{-1} + w_0$

$$t = 1, \text{ then } Y_1 = \phi Y_0 + w_1$$

$$t = 2, \text{ then } Y_2 = \phi Y_1 + w_2$$

similarly,

when, $t = t$, then $Y_t = \phi Y_{t-1} + w_t$

For pth order, the difference equation can be written as in Equation 2.3.

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \dots + \phi_p Y_{t-p} + w_t \dots(2.3)$$

In ARIMA model, for autoregressive (AR) models, forecast depends on observed values in proceeding time period. But in the case of moving average (MA) models, the forecast depends on preceding values of the errors instead of on the variables itself. AR model is appropriate for stationary time series data pattern and the pth order autoregressive model i.e. AR(p) can be written as in Equation 2.4 given below.

$$Y_t = \psi + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \dots + \phi_p Y_{t-p} \dots(2.4)$$

where,

'p' is the order of autoregressive part

'ψ' is a constant

'φ₁, φ₂, φ₃, ..., φ_p' are the autoregressive model parameters

Similarly, the moving average model of qth order i.e. MA(q) can be written as in Equation 2.5.

$$Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \theta_3 \varepsilon_{t-3} - \dots - \theta_q \varepsilon_{t-q} \dots(2.5)$$

where,

'q' is the order of moving average part

'μ' is a constant

'θ₁, θ₂, θ₃, ..., θ_q' are the moving average model parameters

By combining autoregressive (AR) and moving average (MA) terms, Autoregressive moving average (ARMA) model can be developed. It can be represented as ARMA(p,q), where 'p' and 'q' are the order of autoregressive and moving average part respectively. ARMA(p,q) can be written as in Equation 2.6.

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \dots(2.6)$$

The input data series for ARIMA needs to be stationary and by stationary means constant mean, variance, autocorrelation through time. To make the data stationary, the non-stationary data should be differenced. In the case of practical purpose, the difference would be one or two usually (d ≤ 2). Differencing non-stationary series one or more times to attain stationary and doing so produces an integrated (I) model (Peter and Richard, 2002; Robert and David, 2010). Therefore the data series 'Y_t' after differencing is said to follow an integrated autoregressive moving average model and is represented as ARIMA (p,d,q), where 'p' is the order of autoregressive part 'q' is the order of moving average part 'd' refers to how many times the data needs to be differenced to produce a stationary series

By considering d = 1, we can obtain ARIMA (p,d,q) = ARIMA (p,1,q), with difference of 'Y_t' that can be written as shown in Equation (2.7).

$$\Delta_d Y_t = W_t \text{ or } W_t = Y_t - Y_{t-1} \dots(2.7)$$

when, $W_t = Y_t - Y_{t-1}$
 $W_{t-1} = Y_{t-1} - Y_{t-2}$
 $W_{t-2} = Y_{t-2} - Y_{t-3}$

similarly,

when, $t = t - p$

then, $W_{t-p} = Y_{t-p} - Y_{t-p-1}$

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \dots(2.8)$$

substituting,

$$W_t = \phi_1 (Y_{t-1} - Y_{t-2}) + \phi_2 (Y_{t-2} - Y_{t-3}) + \dots + \phi_p (Y_{t-p} - Y_{t-p-1}) + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \dots(2.9)$$

$$Y_t - Y_{t-1} = \phi_1 Y_{t-1} - \phi_1 Y_{t-2} + \dots + \phi_p Y_{t-p} - \phi_p Y_{t-p-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \dots(2.10)$$

therefore,

$$Y_t = (1 + \phi_1) Y_{t-1} + (\phi_2 - \phi_1) Y_{t-2} + (\phi_3 - \phi_2) Y_{t-3} + \dots + (\phi_p - \phi_{p-1}) Y_{t-p} + \phi_p Y_{t-p-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \dots(2.11)$$

Seasonal time series data can be analyzed by using Box-Jenkins method. When seasonal components are involved the model will be called as Seasonal Autoregressive Integrated Moving Average (SARIMA) Model. This seasonal ARIMA model can be represented as SARIMA (p,d,q) (P,D,Q)^S, where the lowercase shows the non-seasonal part and the uppercase shows the seasonal part. SARIMA (p,d,q) (P,D,Q)^S can be written as in Equation 2.12 (James, 2012):

$$\phi_p(B)\phi_p(B^S)(1-B)^d(1-B^S)^D Y_t = \theta_q(B)\theta_q(B^S)\varepsilon_t \dots(2.12)$$

where,

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

$$\phi_p(B^S) = 1 - \phi_1 B^S - \phi_2 B^{2S} - \dots - \phi_p B^{pS}$$

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

$$\theta_q(B^S) = 1 - \theta_1 B^S - \theta_2 B^{2S} - \dots - \theta_q B^{qS}$$

where,

- 'p' is the order of autoregressive part
- 'q' is the order of moving average part
- 'd' is the order of difference
- 'P' is the order of seasonal autoregressive part
- 'Q' is the order of seasonal moving average part
- 'D' is the order of seasonal difference
- 'B' is the back shift operator
- 'S' is the length of seasonal period

In this study, air quality data of Sulphur dioxide (SO₂), Nitrogen dioxide (NO₂), Respirable suspended particulate matter (RSPM) and Suspended particulate matter (SPM) over a period of 2012-2015 at Thiruvananthapuram District, obtained from Kerala State Pollution Control Board (KSPCB) has been used. The computations involved in this study have been computed by

using Minitab 17 (Version: 17.3.1), R Software (Version: 3.3.2) and R-Studio (Version: 1.0.136).

2.6 Performance Measures

Primary criterion in selecting the order of ARIMA model is by using Akaike Information Criteria (AIC) (Allan and Chih, 1998; Peter and Richard, 2002; Burnham and Anderson, 2004). AIC is a measure of the relative quality of various statistical models and AIC_c is Akaike Information Criteria with a correction for finite sample sizes. Among the numerous values of AIC_c, the smallest value of AIC_c, are selected as the finest model to be used in forecasting purpose (James, 2012). Model accuracy and performance was identified using the Akaike Information Criteria (AIC_c) with a correction, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). Equations for calculating MAE, MAPE and RMSE are as follows.

$$AIC = -2L + 2k \dots(2.13)$$

where,

- 'L' is the log-likelihood
- 'k' is the number of estimated parameters in the model (i.e., number of variables + intercept)
- 'n' is the sample size

$$AIC_c = -2L + 2k + \frac{2k(k+1)}{(n-k-1)} \dots(2.14)$$

$$MAE = \frac{\sum_{t=1}^n |y_t - \hat{y}_t|}{n} \dots(2.15)$$

$$MAPE = \frac{\sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|}{n} \times 100 \dots(2.16)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}} \dots(2.17)$$

where,

‘ y_t ’ is the actual value

‘ \hat{y}_t ’ is the predicted value

‘ n ’ is the number of observations

III. RESULTS AND DISCUSSION

AQI is determined by selecting the maximum operator scheme, i.e. out of all the considered pollutants, the maximum sub-index value becomes the overall index or air quality index. The air quality data from the Kerala State Pollution Control Board shows that the responsible pollutant for AQI in all these stations is respirable suspended particulate matter (RSPM) due to its abundance in the atmosphere. From the AQI calculation, the air quality reported at these four stations from 2012-2015 comes under the satisfactory (51 - 100) category. An average AQI of 60 was observed in the year 2012 and this was the highest average during the four years. The AQI for four stations shows that from 2012 to 2014, the AQI values were decreasing. But in the year 2015 the average AQI values were showing an increasing trend. Air quality data from the four stations in Thiruvananthapuram District shows that the air quality is starting to deteriorating at a consistent rate. Pollutant concentrations of SO₂, NO₂, and Suspended particulate matter were well below the limits specified by the Central Pollution Control Board (CPCB). As a result, respirable suspended particulate matter is mainly responsible for the AQI values in all cases i.e. out of the pollutant’s individual AQI values, the index values of RSPM was high.

The time series plots of air quality index of the four stations are shown in Figure 2 to Figure 5. Time series plot showed that data is non-stationary in all cases. For the seasonal pattern, stationary data series was obtained by taking D=1 with S=12 and for non-seasonal pattern, it was obtained by taking d=1 (Hanke and Wichern, 2008).

(Figure 2 to Figure 5)

After the tentative identification, parameter estimation and diagnostic checking procedures, best suited ARIMA model was developed. AICc values of the all the possible ARIMA models were obtained automatically are summarized in Table 4 to Table 7 and from those values the best fit ARIMA model was selected and the accuracy was checked using the measures like MAE, MAPE and RMSE. AICc, MAE, MAPE and RMSE values were summarized in Table 8 and

Table 9. Certain cases R Studio fails to detect the right ARIMA model automatically and in such cases the ARIMA model was determined manually using R Studio.

(Table 4 to Table 9)

Also by using Ljung-Box test and by inspecting the p-value of co-efficient, the best model was developed. Since the correlogram showed that none of the sample autocorrelation for lags 1-35 exceeds the significance bounds and the p-value for the Ljung-Box test is approximately 0.9 for all the cases, we can conclude that there is exceptionally small evidence for non-zero autocorrelation in the forecast errors at lag 1-35. After the best fit ARIMA model determination, the model was used to predict the air quality index during the year 2012-2015. The actual and predict values for the year 2012-2015 for the stations are shown in Figure 6 to Figure 9 and also summarized in Table 10 to Table 13. Error i.e. difference between the actual and predicted values in most cases varies from -3 to +3 and higher error was observed during August 2012 (Error = -7) and August 2013 (Error = 9).

(Figure 6 to Figure 9)

(Table 10 to Table 13)

Based on the best fit ARIMA model, the air quality index for the year 2016 was forecasted for these four stations. The time series plots of actual values and predicted values for the different stations using Minitab are shown in Figure 10 to Figure 13. Based on the forecast, this ARIMA method is capable in monitoring and forecasting the ambient air quality. From the comparison between actual and predicted values, the air quality index value of these four stations comes under satisfactory AQI category.

(Figure 10 to Figure 13)

IV. CONCLUSION

In this study, the time series analysis and forecasting was used to analyze the air quality index in Thiruvananthapuram District, Kerala, India. Generally, the air quality index recorded at all stations between the years 2012-2015 were within satisfactory (51-100) AQI range. Mean of the air quality index recorded in all stations were satisfactory and the major pollutant in the area is respirable suspended particulate matter. Effect and concentrations of Sulphur dioxide and Nitrogen dioxide recorded at all the stations were well below the limits.

Based on the forecasting results, we can say that the ARIMA method is capable of monitoring and forecasting the air quality condition. The actual and predicted AQI values plot for the four stations demonstrates the capability of ARIMA model in the time series

analysis of air quality data. In summary, this method provided satisfactory results and has been proven more effective in forecasting and analyzing air quality index. Even though the time series models are very useful in prediction of future air quality, these models cannot replicate the spatial scenario, information about the sources of pollutants and dispersion parameters, it would be better to incorporate the statistical models to the air dispersion models for the forecasting of air quality. On the other hand, the deterministic air quality models that include spatial scenario may not work perfectly when it comes to future air quality prediction with respect to the statistical models. It is expected that the combination of both models i.e. combined models may improve the forecasting or prediction ability and the spatio-temporal scenario of these air quality models, thereby helping the policy makers for regulating air pollution problems. In order to achieve more accurate results, hybrid models can be adopted and these models can also reduce the risk of using an unsuitable model by combining several models together. Artificial Neural Network (ANN) are computer algorithms designed to simulate the human brain in terms of pattern recognition and they are well suited for forecasting because it can handle the complex pollutant formation processes with ease. The hybrid model of ANN can be obtained using ARIMA models and this hybrid model can produce more accurate results than the artificial neural network.

COMPETING INTERESTS

All authors declare that they have no actual or potential competing financial interest.

ACKNOWLEDGEMENTS

We thank Kerala State Pollution Control Board (Regional and District Office, Thiruvananthapuram, Kerala, India) for necessary data provided. We acknowledge the critical comments from anonymous reviewers and editor.

ETHICAL CONSIDERATIONS

Ethical issues (including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, redundancy, etc.) have been completely observed by authors.

REFERENCES

- [1]. Allan, D.R.M., Chih, L.T., 1998. Regression and time series model selection, World Scientific Publishing Company, Singapore.
- [2]. Argiriou, A.A., 2007. Use of Neural Networks for Tropospheric Ozone time series approximation and forecasting- A review. *Atmospheric Chemistry and Physics* 7, 5739-5767.
- [3]. Begum, B.A., Hopke, P.K., Markwitz, A., 2013. Air pollution by fine particulate matter in Bangladesh. *Atmospheric Pollution Research* 4, 75-86.
- [4]. Box, G.E.P., Jenkins, G.M., 1970. Time series analysis, forecasting and control, Holden-Day, San Francisco.
- [5]. Brunelli, U., Piazza, V., Pignato, L., Sorbello, F., Vitabile, S., 2007. Two-days ahead prediction of daily maximum concentrations of SO₂, O₃, PM₁₀, NO₂, CO in the urban area of Palermo, Italy. *Atmospheric Environment* 41(14), 2967-2995.
- [6]. Burnham, K.P., Anderson, D.R., 2004. Multimodel interface: understanding AIC and BIC in model selection. *Sociological Methods and Research* 33, 261-304.
- [7]. Central Pollution Control Board (CPCB) National Air Quality Index Report, 2014. PR Division on behalf of Dr. A.B. Akolkar., Member Secretary.CPCB, Delhi.
- [8]. Chelani, A.B., Rao, C.V.C., Phadke, K.M., Hasan, M.Z., 2002. Prediction of Sulphur Dioxide concentration using artificial neural networks. *Environmental Modelling and Software* 17(2), 159-166.
- [9]. Cryer, J.D., 1986. Time series analysis, Duxbury Press, United States of America.
- [10]. Cryer, J.D., Chan, K.S., 2008. Time series analysis with applications in R, 2nd Ed., Springer-Verlag, New York.
- [11]. Daniel, V., 2014. Fundamentals of air pollution, 5th Ed., Academic Press, San Francisco.
- [12]. District Census Handbook (DCHB), Series-33, Part XII-B, Census Organization, Directorate of Census Operations, Kerala, IN, 2011.
- [13]. District profile of Thiruvananthapuram city, Thiruvananthapuram Corporation. Available:http://www.trivandrum.gov.in/district_profile.php accessed on April 2017.
- [14]. Environmental Protection Agency, 2003, Air Quality Index: A Guide to Air Quality and Your Health, *Environmental Protection Agency Office of Air Quality Planning and Standards*, Research Triangle Park.
- [15]. Ghazali, N.A., Nor, A.R., Ahmad, S.Y., Noor, F.F.M.Y., Sansuddin, N. and Wesam, A.A.M., 2010. Transformation of Nitrogen Dioxide into Ozone and prediction of Ozone concentrations using multiple linear regression techniques. *Environmental Monitoring and Assessment* 165(1), 475-489.

- [16]. Green, M.H., 1966. An air pollution index based on sulphur dioxide and smoke shade. *Journal of Air Pollution Control Association* 16(12), 703-706.
- [17]. Hanke, J.E., Wichern, D.W., 2008. Business forecasting, 9th Ed., Pearson United States of America.
- [18]. James, D.H., 2012. Time series analysis, Levant Books, Princeton University Press, United States of America.
- [19]. Khashei, M., Bijari, M., 2010. An artificial neural network (p,d,q) model for time series forecasting. *Expert Systems with Applications* 37(1), 479-489.
- [20]. Kumar, A., Goyal, P., 2011. Forecasting of daily air quality index in Delhi. *Science of the Total Environment* 409, 5517-5523.
- [21]. Kumar, U., Jain, V., 2010. ARIMA forecasting of ambient air pollutants (O₃, NO, NO₂ and CO). *Stochastic Environmental Research and Risk Assessment* 24(5), 751-760.
- [22]. Kurt, A., Oktay, A.B., 2010. Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks. *Expert Systems with Applications* 37, 7986-7992.
- [23]. Lee, J.Y., Lee, S.B., Bae, G.N., 2014. A review of the association between air pollutant exposure and allergic diseases in children. *Atmospheric Pollution Research* 5, 616-629.
- [24]. Mitra, A.P., Sharma, C.S., 2002. Indian aerosols: present status. *Chemosphere* 49, 1175-1190.
- [25]. Nagendra, S.M.S., Venugopal, K., Steven, L.J., 2007. Assessment of air quality near traffic intersections in Bangalore city using air quality indices. *Transportation Research Part D* 12, 167-176.
- [26]. Nur, H.A.R., Muhammad, H.L., Suhartono., Mohd, T.L., 2016. Evaluation performance of time series approach for forecasting air pollution index in Johor, Malaysia. *Sains Malaysiana* 45(11), 1625-1633.
- [27]. Peter, J.B., Richard, A.D., 2002. Introduction to time series and forecasting, 2nd Ed., Springer-Verlag, New York.
- [28]. Rao, C.S., 2006. Environmental pollution control engineering, 2nd Ed., New Age International Publishers, India.
- [29]. Robert, H.S., David, S.S., 2010. Time series analysis and its applications-with R examples, 3rd Ed., Springer-Verlag, New York.
- [30]. Sansuddin, N., Ramli, N.A., Yahaya, A.S., Yusof, N.F., Ghazali, N. A. and Madhoun, W.A., 2011. Statistical analysis of PM₁₀ concentrations at different locations in Malaysia. *Environmental Monitoring and Assessment* 180(1), 573-588.
- [31]. Sarasamma, J.D., Narayanan, B.K., 2014. Air quality assessment in the surroundings of KMML industrial area, Chavara. *Aerosol and Air Quality Research* 14, 1769-1778.
- [32]. Sharma, M., Pandey, R., Maheshwari, M., Sengupta, B., Shukla, B.P., Mishra, A., 2003b. Air quality index and its interpretation for the city of Delhi. *Clean Air: International Journal on Energy for a Clean Environment* 4(3), 83-98.
- [33]. Sharma, M., Pandey, R., Maheshwari, M., Shukla, B.P., Gupta, N.K., Johri, S., 2003a. Interpretation of air quality data using an air quality index for the city of Kanpur, India. *Journal of Environmental Engineering and Science* 2(6), 453-462.
- [34]. Sinha, A., Rastogi, S.K., 2016. Collaboration between central and state government and environmental quality: Evidences from Indian cities. *Atmospheric Pollution Research* 8(2), 285-296.
- [35]. USEPA, Air Quality Index Reporting; Final Rule, 1999. Federal Register, Part III, CFR Part 58.
- [36]. Vlachogianni, A., Kassomenos, P., Karppinen, A., Karakitsios, S., Kukkonen, J., 2011. Evaluation of a multiple regression model for the forecasting of the concentrations of NO_x and PM₁₀ in Athens and Helsinki. *Science of the Total Environment* 409(8), 1559-1571.
- [37]. Water and Air Quality Directory, 2015, Kerala State Pollution Control Board, Kerala, India.

List of Tables

1. **Table 1.** Air quality monitoring stations in Thiruvananthapuram
2. **Table 2.** Proposed sub-index and breakpoint pollutant concentration for Indian air quality index (24 hr. avg.)
3. **Table 3.** Proposed sub-index and breakpoint pollutant concentration for Indian air quality index (24 hr. avg.)
4. **Table 4.** Possible models and AICc values for station: Veli Filatex
5. **Table 5.** Possible models and AICc values for station: SMV School
6. **Table 6.** Possible models and AICc values for station: Cosmopolitan Hospital
7. **Table 7.** Possible models and AICc values for station: Pettah
8. **Table 8.** ARIMA model for different stations
9. **Table 9.** MAE, MAPE and RMSE for the selected ARIMA model

10. **Table 10.** Actual and predicted AQI values for the year 2012
11. **Table 11.** Actual and predicted AQI values for the year 2013
12. **Table 12.** Actual and predicted AQI values for the year 2014
13. **Table 13.** Actual and predicted AQI values for the year 2015

List Figures

1. **Fig.1** Study area and sampling locations in Thiruvananthapuram District, Kerala, India
2. **Fig.2** Air quality index from 2012-2015 for Veli Filatex station
3. **Fig.3** Air quality index from 2012-2015 for SMV School station
4. **Fig. 4** Air quality index from 2012-2015 for Cosmopolitian Hospital station

5. **Fig. 5** Air quality index from 2012-2015 for Pettah station
6. **Fig. 6** AQI actual v/s predicted from 2012-2015 for Veli Filatex station
7. **Fig. 7** AQI actual v/s predicted from 2012-2015 for SMV School station
8. **Fig. 8** AQI actual v/s predicted from 2012-2015 for Cosmopolitian Hospital station
9. **Fig. 9** AQI actual v/s predicted from 2012-2015 for Pettah station
10. **Fig. 10** AQI forecast for the year 2016, Veli Filatex station
11. **Fig. 11** AQI forecast for the year 2016, SMV School station
12. **Fig. 12** AQI forecast for the year 2016, Cosmopolitian Hospital station
13. **Fig. 13** AQI forecast for the year 2016, Pettah station

Table 1. Air quality monitoring stations in Thiruvananthapuram (Water and Air Quality Directory, 2015)

Station	Location Type	Latitude	Longitude
Veli Filatex	Industrial	8°29'529"N	76°55'764"E
SMV School	Sensitive	8°36'299"N	76°56'759"E
Cosmopolitian Hospital	Sensitive	8°31'000"N	76°56'159"E
Pettah	Residential	8°29'651"N	76°55'761"E

Table 2. Proposed sub-index and breakpoint pollutant concentration for Indian air quality index (24 hr. avg.) (Nagendra et al., 2007)

Sl. No.	SO ₂	NO ₂	SPM	RSPM	Index values	Descriptor
1	0 - 80	0 - 80	0 - 200	0 - 100	0 - 100	Good ¹
2	81 - 367	81 - 180	201 - 260	101 - 150	101 - 200	Moderate ²
3	368 - 786	181 - 564	261 - 400	151 - 350	201 - 300	Poor ³
4	787 - 1572	565 - 1272	401 - 800	351 - 420	301 - 400	Very poor ⁴
5	> 1572	> 1272	> 800	> 420	401 - 500	Severe ⁵

¹ Good: Acceptable air quality, however, there may be a moderate health distress for a very small number of people due to some pollutants.

² Moderate: Sensitive group members may experience health effects.

³ Poor: Sensitive group members may experience more severe health effects.

⁴ Very poor: Triggers health alert, everyone may experience more severe health effects.

⁵ Severe: Triggers health warnings of emergency situations (Sharma et al., 2003a, 2003b).

Table 3. Proposed sub-index and breakpoint pollutant concentration for Indian air quality index (24 hr. avg.) (CPCB National Air Quality Index Report, 2014)

Sl. No.	SO ₂	NO ₂	PM _{2.5}	PM ₁₀	Index values	Descriptor
1	0 - 40	0 - 40	0 - 30	0 - 50	0 - 50	Good
2	41 - 80	41 - 80	31 - 60	51 - 100	51 - 100	Satisfactory
3	81 - 380	81 - 180	61 - 90	101 - 250	101 - 200	Moderate
4	381 - 800	181 - 280	91 - 120	251 - 350	201 - 300	Poor
5	801 - 1600	281 - 400	121 - 250	351 - 430	301 - 400	Very poor
6	> 1600	> 400	> 250	> 430	> 400	Severe

Table 4. Possible models and AICc values for station: Veli Filatex

Station	ARIMA	AICc	Selected ARIMA
Veli Filatex	(0,1,0) with drift	245.176	ARIMA (1,1,1)
	(1,1,0)(1,0,0) ^[12]	248.098	
	(0,1,1)(0,0,1) ^[12]	247.377	
	(1,1,1)	243.233	
	(0,1,0)(1,0,0) ^[12] with drift	247.429	
	(0,1,0)(0,0,1) ^[12] with drift	247.435	
	(1,1,0) with drift	245.707	
	(0,1,1) with drift	245.019	

Table 5. Possible models and AICc values for station: SMV School

Station	ARIMA	AICc	Selected ARIMA
SMV School	(2,0,2)(1,0,1) ^[12] with non-zero mean	241.749	ARIMA (2,0,1)
	(1,0,0)(1,0,0) ^[12] with non-zero mean	236.757	
	(0,0,1)(0,0,1) ^[12] with non-zero mean	240.164	
	(1,0,0)	235.409	
	(1,0,0)(0,0,1) ^[12] with non-zero mean	236.603	
	(1,0,0)(1,0,1) ^[12] with non-zero mean	239.098	
	(2,0,0) with non-zero mean	235.288	
	(2,0,1) with non-zero mean	235.163	
	(2,0,1)(1,0,0) ^[12] with non-zero mean	236.644	
	(2,0,1)(0,0,1) ^[12] with non-zero mean	236.557	
	(2,0,1)(1,0,1) ^[12] with non-zero mean	239.307	
	(1,0,1) with non-zero mean	236.144	
	(3,0,1) with non-zero mean	237.732	
(2,0,2) with non-zero mean	237.638		

Table 6. Possible models and AICc values for station: Cosmopolitan Hospital

Station	ARIMA	AICc	Selected ARIMA
Cosmopolitan Hospital	(1,0,0)(1,0,0) ^[12] with non-zero mean	223.743	ARIMA (1,0,0)
	(0,0,1)(0,0,1) ^[12] with non-zero mean	224.386	
	(1,0,0) with non-zero mean	222.493	
	(1,0,0)(0,0,1) ^[12] with non-zero mean	223.840	
	(2,0,0) with non-zero mean	224.331	
	(1,0,1) with non-zero mean	223.881	

Table 7. Possible models and AICc values for station: Pettah

Station	ARIMA	AICc	Selected ARIMA
Pettah	(0,1,0) with drift	245.176	ARIMA (0,1,0)
	(1,1,0)(1,0,0) ^[12] with drift	248.098	
	(0,1,1)(0,0,1) ^[12] with drift	247.377	
	(0,1,0)	243.233	
	(0,1,0)(1,0,0) ^[12] with drift	247.429	
	(0,1,0)(0,0,1) ^[12] with drift	247.435	
	(1,1,0) with drift	245.707	
	(0,1,1) with drift	245.019	

Table 8. ARIMA model for different stations

Station	ARIMA Model	AIC	AICc	BIC
Veli Filatex	ARIMA (1,1,1)	243.14	243.23	244.99
SMV School	ARIMA (2,0,1)	233.73	235.16	243.09
Cosmopolitan Hospital	ARIMA (1,0,0)	221.95	222.49	227.56
Pettah	ARIMA (0,0,1)	181.42	181.97	187.04

Table 9. MAE, MAPE and RMSE for the selected ARIMA model

Station	ARIMA Model	MAE	MAPE	RMSE
Veli Filatex	ARIMA (1,1,1)	2.121	3.932	2.945
SMV School	ARIMA (2,0,1)	1.877	3.465	2.469
Cosmopolitan Hospital	ARIMA (1,0,0)	1.818	3.464	2.286
Pettah	ARIMA (0,0,1)	1.184	2.303	1.496

Table 10. Actual and predicted AQI values for the year 2012

Month-Year	Filatex-Veli		SMV School		Cosmopolitan Hospital		Pettah	
	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted
Jan-12	63	66	57	58	50	52	50	51
Feb-12	67	65	52	57	53	53	50	51
Mar-12	64	65	63	57	56	55	52	52
Apr-12	64	64	54	57	52	54	53	52
May-12	64	62	54	55	53	53	51	51
Jun-12	61	57	51	51	49	49	50	48
Jul-12	59	58	53	51	52	50	51	50
Aug-12	49	56	51	52	57	54	51	51
Sep-12	57	55	54	52	52	54	50	50
Oct-12	57	57	51	51	51	53	49	51
Nov-12	60	57	54	52	53	54	53	52
Dec-12	59	57	54	53	53	52	51	52

Table 11. Actual and predicted AQI values for the year 2013

Month-Year	Filatex-Veli		SMV School		Cosmopolitan Hospital		Pettah	
	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted
Jan-13	60	56	55	53	52	53	52	52
Feb-13	59	60	53	53	52	54	52	52
Mar-13	61	57	53	56	55	55	53	53
Apr-13	61	58	56	52	54	53	53	54
May-13	60	58	55	52	54	54	54	52
Jun-13	51	55	51	51	48	50	49	50
Jul-13	51	52	50	51	48	50	47	50
Aug-13	55	46	53	50	50	51	50	50
Sep-13	57	55	52	52	50	48	50	49
Oct-13	58	55	54	50	52	50	52	50

Nov-13	59	57	54	52	57	53	52	52
Dec-13	61	57	56	53	53	55	53	52

Table 12. Actual and predicted AQI values for the year 2014

Month-Year	Filatex-Veli		SMV School		Cosmopolitian Hospital		Pettah	
	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted
Jan-14	53	57	54	55	53	53	53	52
Feb-14	53	56	59	54	53	55	53	52
Mar-14	53	55	54	58	54	56	53	54
Apr-14	52	54	52	54	53	53	52	54
May-14	52	52	53	52	53	53	52	52
Jun-14	50	46	51	49	50	49	50	49
Jul-14	49	47	48	50	47	51	49	50
Aug-14	49	47	49	49	47	51	49	51
Sep-14	49	49	50	49	48	46	49	50
Oct-14	50	49	50	48	49	48	50	51
Nov-14	49	50	50	50	48	50	49	52
Dec-14	51	50	50	50	49	47	51	51

Table 13. Actual and predicted AQI values for the year 2015

Month-Year	Filatex-Veli		SMV School		Cosmopolitian Hospital		Pettah	
	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted
Jan-14	53	46	53	50	51	48	53	51
Feb-14	52	49	54	51	55	52	52	51
Mar-14	54	50	58	54	57	56	54	52
Apr-14	55	51	58	53	58	55	55	53
May-14	51	52	55	54	55	57	51	53
Jun-14	49	47	53	51	52	52	49	50
Jul-14	53	47	57	52	56	52	53	49
Aug-14	53	48	57	54	57	58	53	52
Sep-14	51	50	56	56	53	56	51	52
Oct-14	55	51	54	55	53	54	55	52
Nov-14	55	52	55	55	54	55	55	54
Dec-14	52	54	55	55	53	53	52	54

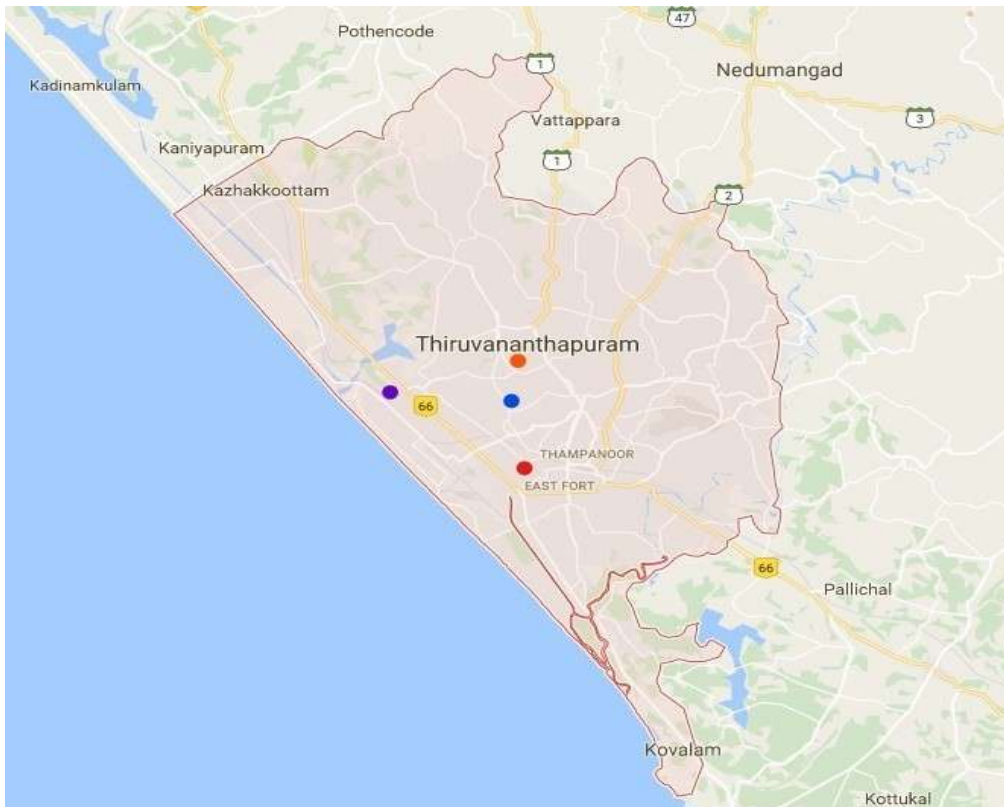


Fig.1 Study area and sampling locations in Thiruvananthapuram District, Kerala, India (©Google Maps, 2017)

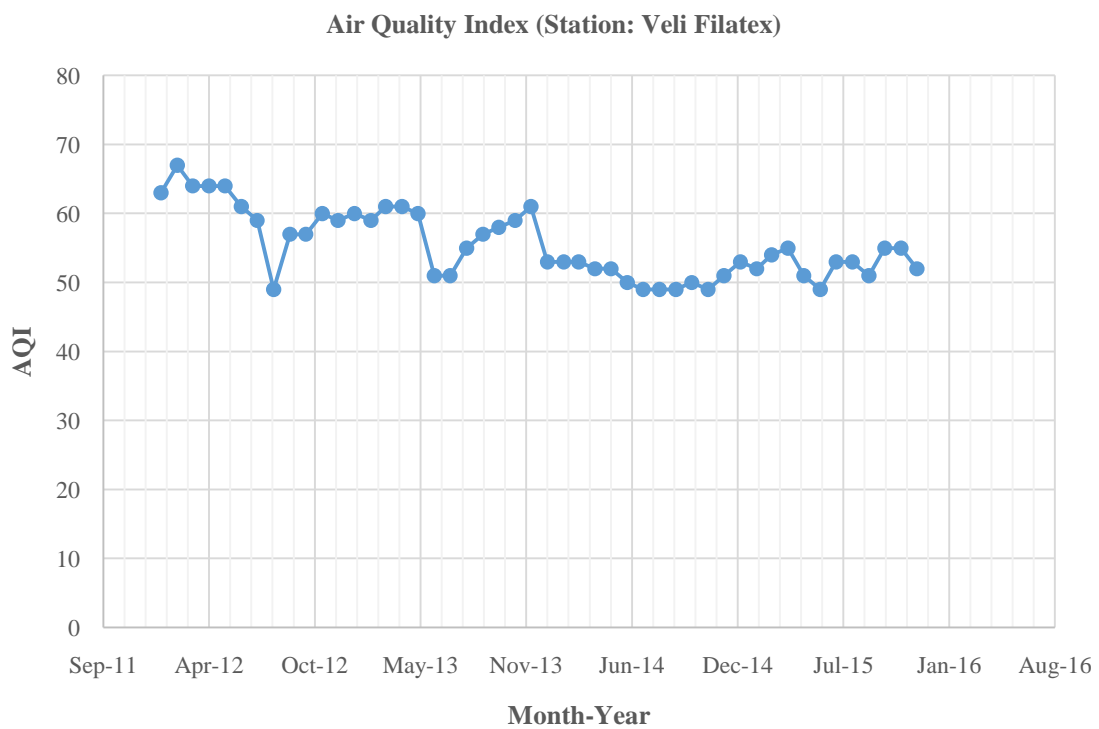


Fig.2 Air quality index from 2012-2015 for Veli Filatex station

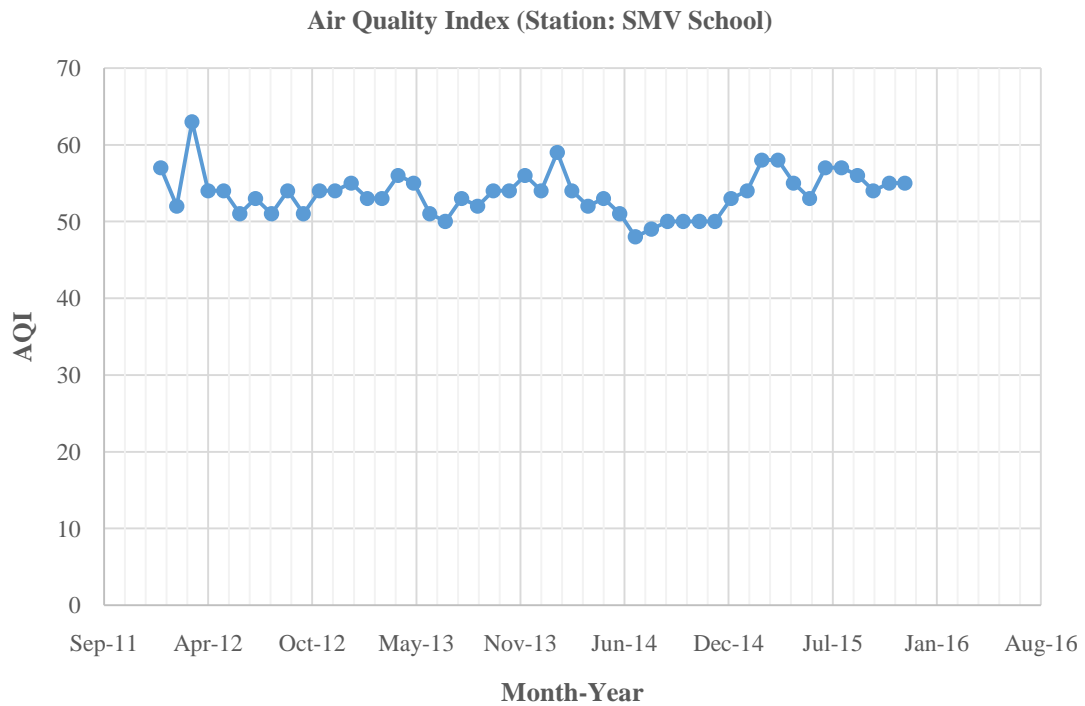


Fig.3Air quality index from 2012-2015 for SMV School station

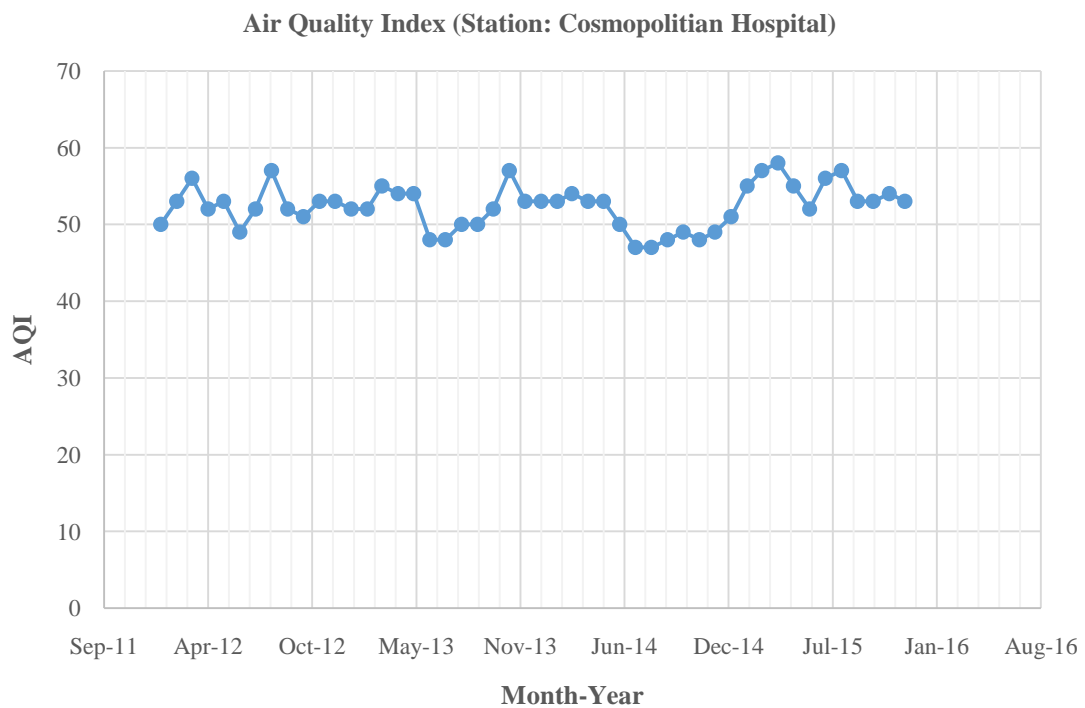


Fig.4Air quality index from 2012-2015 for Cosmopolitian Hospital station

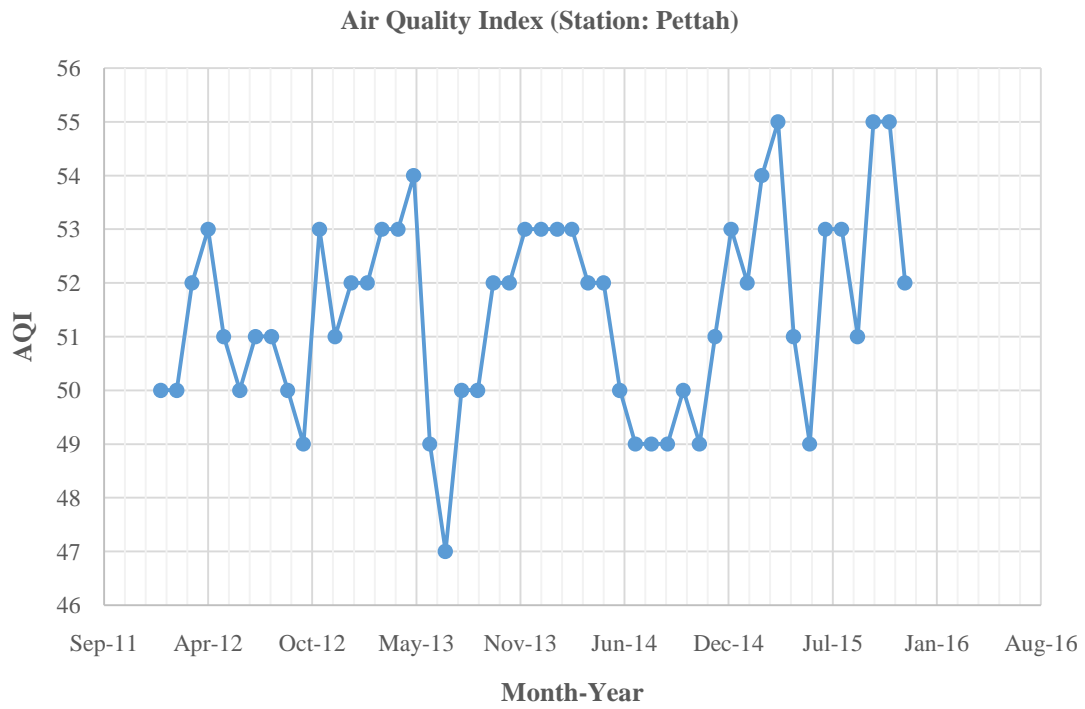


Fig. 5 Air quality index from 2012-2015 for Pettah station

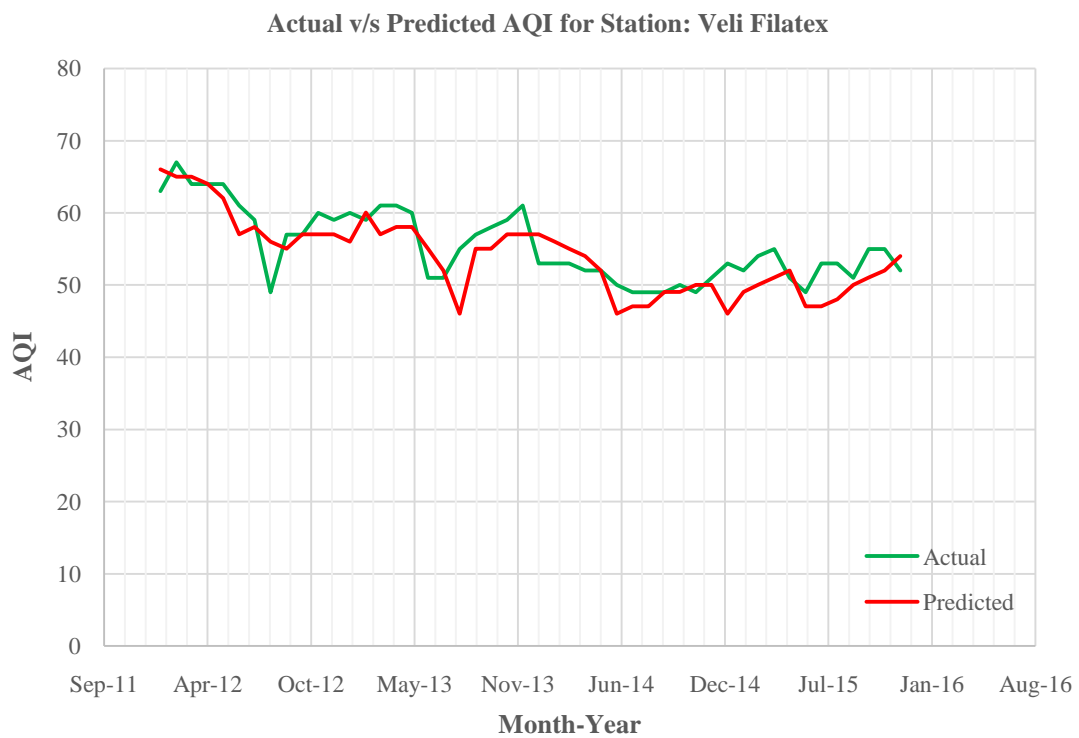


Fig.6 AQI actual v/s predicted from 2012-2015 for Veli Filatex station

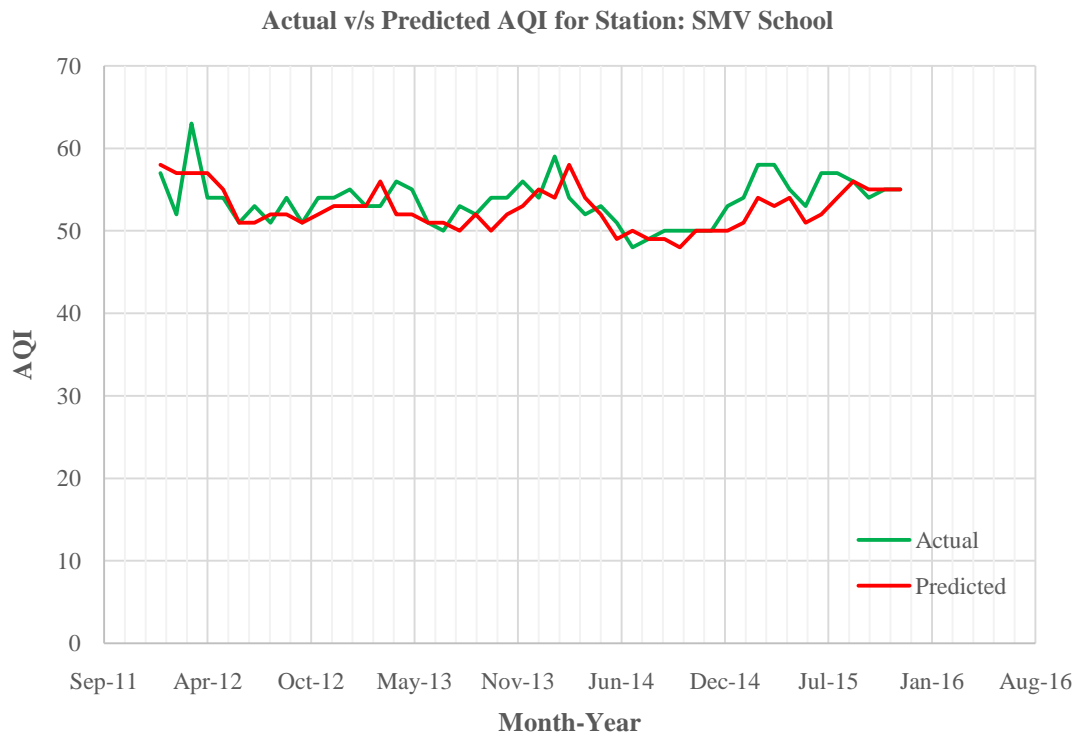


Fig.7AQI actual v/s predicted from 2012-2015 for SMV School station

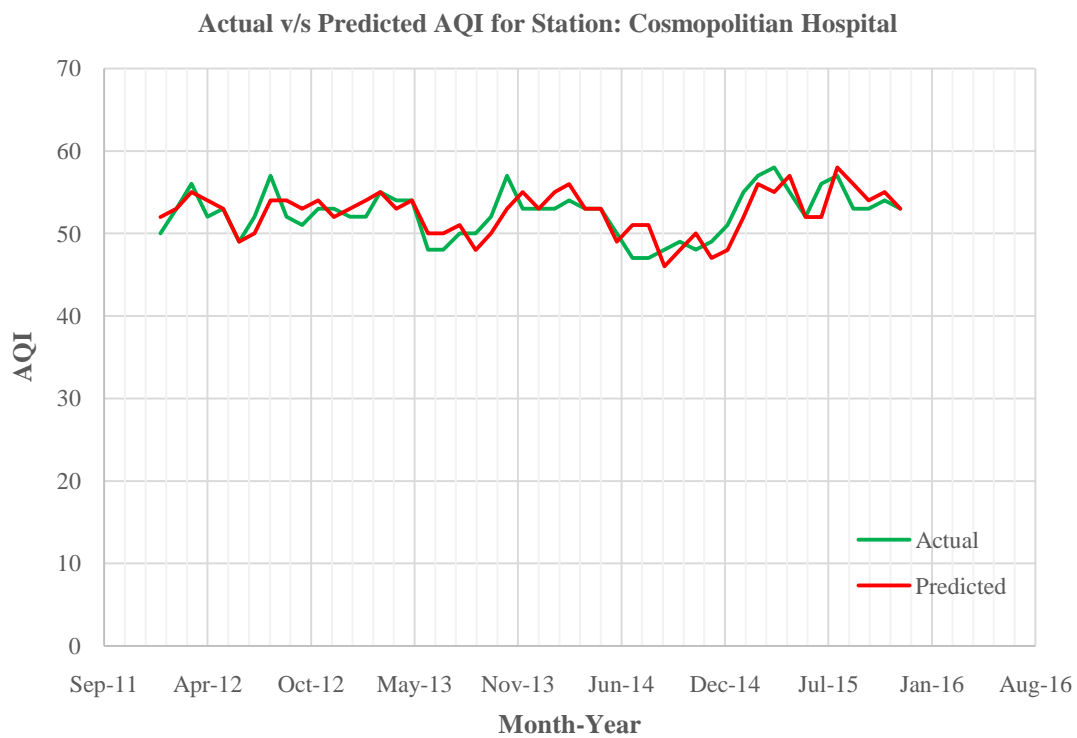


Fig.8AQI actual v/s predicted from 2012-2015 for Cosmopolitian Hospital station

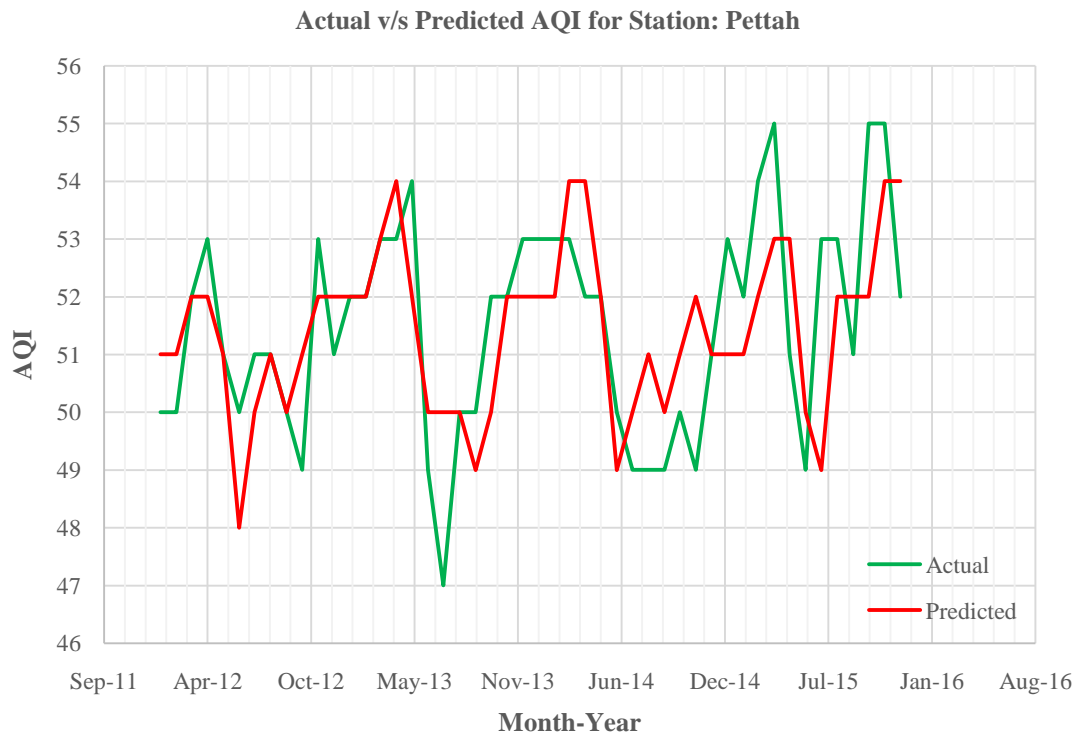


Fig. 9 AQI actual v/s predicted from 2012-2015 for Pettah station

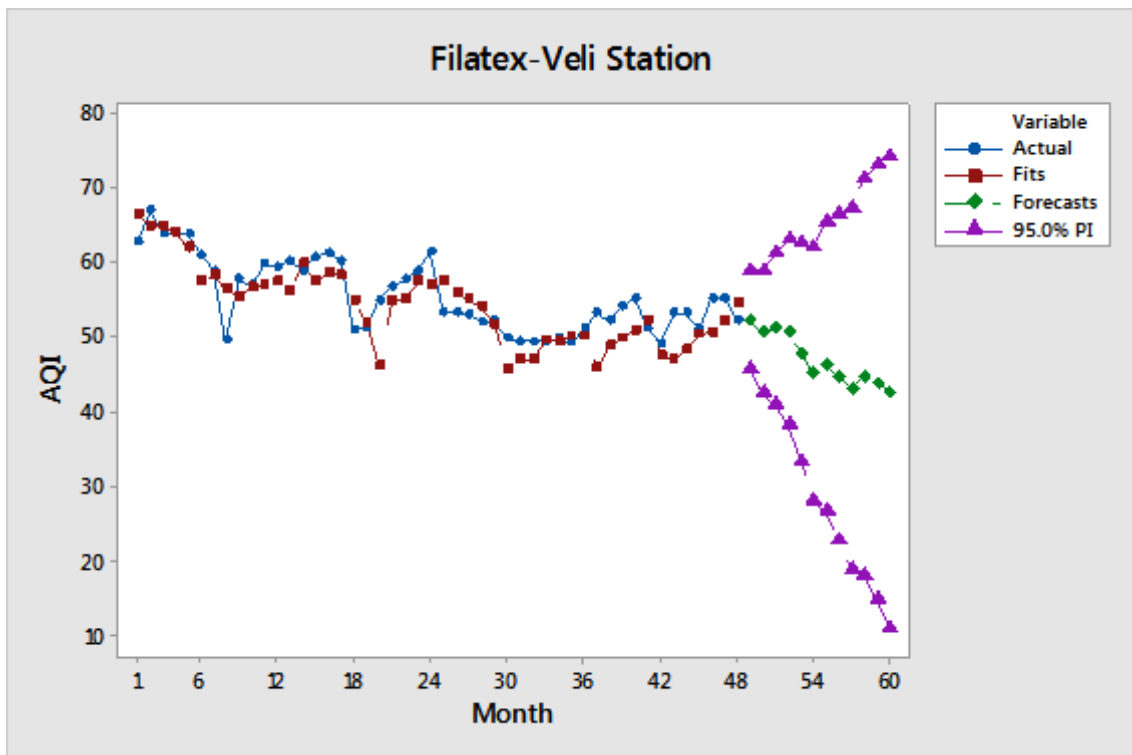


Fig. 10 AQI forecast for the year 2016, Veli Filatex station

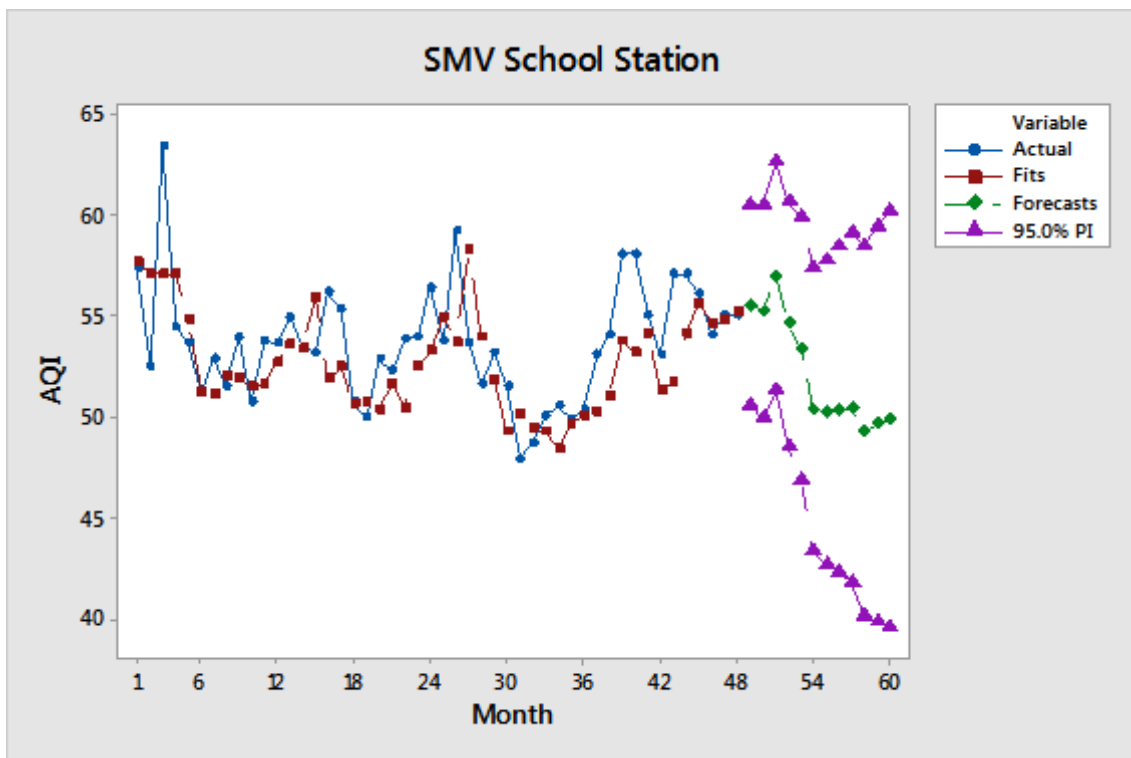


Fig. 11 AQI forecast for the year 2016, SMV School station

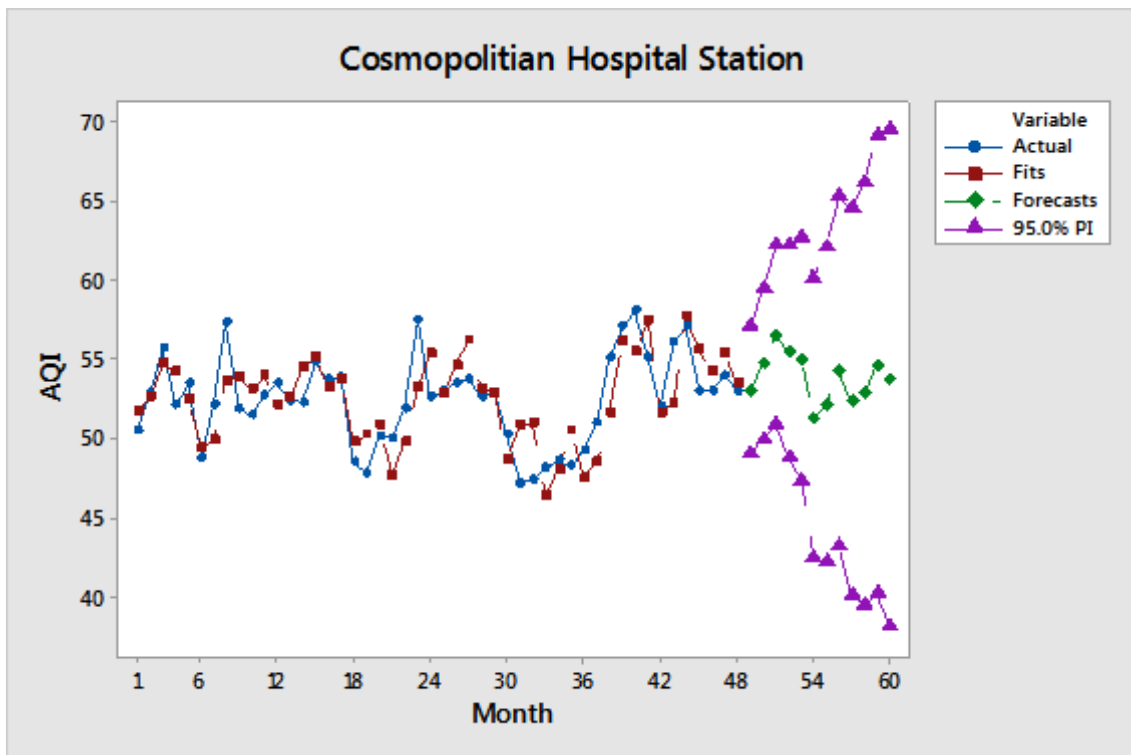


Fig. 12 AQI forecast for the year 2016, Cosmopolitan Hospital station

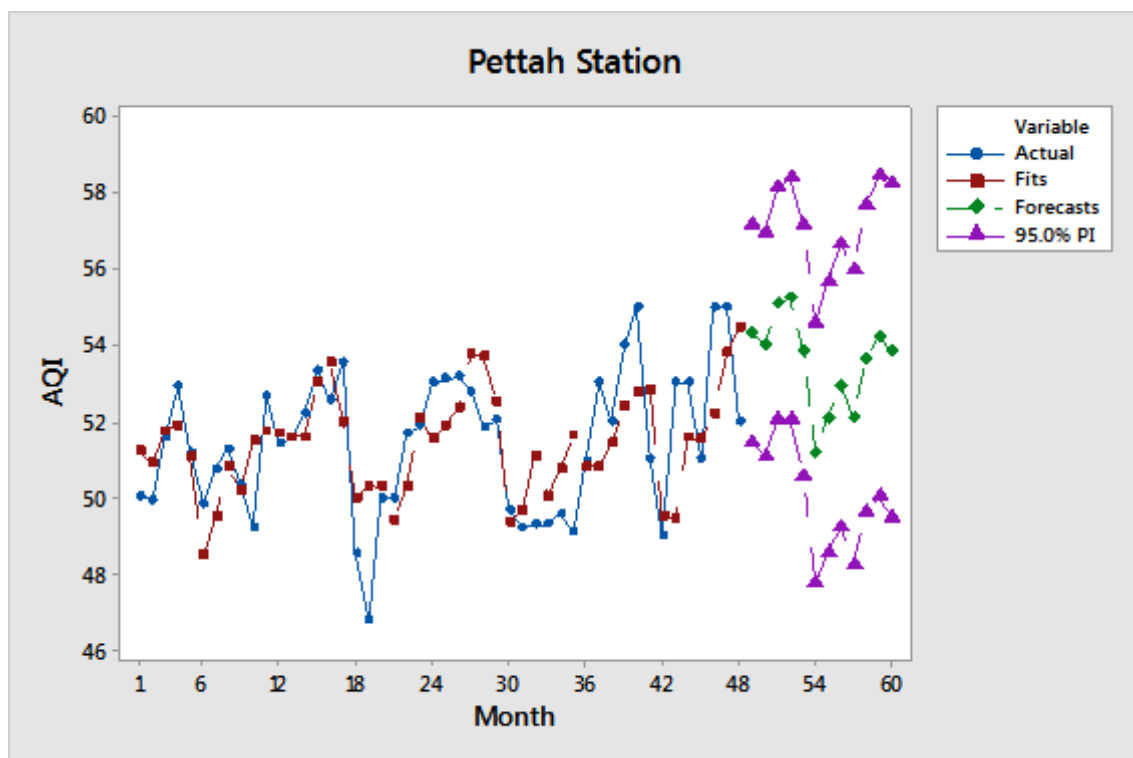


Fig. 13AQI forecast for the year 2016, Pettah station