Article

Comparative study and prediction of ambient air quality of Durgapur Industrial Belt, West Bengal using time series forecast model

S. Sarkar, **S. Sanyal**, **M. Chakrabarty**
PG Department of Conservation Biology, Durgapur Government College, Jawahar Lal Nehru Avenue, Durgapur-713214, India
E-mail: sanuban@gmail.com

Received 6 June 2022; *Accepted 15 July 2022*; *Published online 2 August 2022*; *Published 1 December 2022*

Abstract
Time series predictive forecast models can be used to monitor air levels and study, trends in ambient air quality. In this study, we have collected the data of Oxides of Sulphur (SO_x), Suspended Particulate Matter (SPM_{10}), and Oxides of Nitrogen (NO_x) in Bidhannagar, Durgapur Industrial Belt, West Bengal from the West Bengal Pollution Control Board (2007-2019). The data was analysed both season-wise and month-wise to nullify the irregularity of the data obtained.Time series model can be established only using regular data. The study mainly focused on the comparision between month-wise and season-wise models, in order to identifya prediction model that provides best predictions. The forecast was performed using the seasonal autoregressive integrated moving average (SARIMA) model. The results confirmed that among all forecast models of SARIMA (P, D, Q) (p, d, q) the season-wise model provide best predictions. Graphs indicated a higher concentration of SPM_{10} and NO_x in the winter. The season-wise model for SPM_{10} (1,0,0)(0,1,1) showed significant higher trends but similar season-wise forecast model for NO_x (1,0,0)(0,1,1) showed a decreasing trend. The levels of SPM_{10} is predicted to show an increasing trend in next four years whereas NO_x level is predicted to remain low for the next four years. The seasonal prediction models can be used in understanding the trend in ambient air quality in Durgapur Industrial Belt and facilitate in taking necessary steps to combat the prevailing environmental conditions.

Keywords air pollution; oxides of nitrogen; oxides of sulphur; suspended particulate matter; SARIMA; forecast.

Computational Ecology and Software
ISSN 2220-721X
URL: http://www.iaees.org/publications/journals/ces/online-version.asp
RSS: http://www.iaees.org/publications/journals/ces/rss.xml
E-mail: ces@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

Durgapur is a major industrial belt in West Bengal with large industries like Durgapur Steel Plant, Alloy Steel Plant, Durgapur Project Ltd., Durgapur Chemicals, Ophthalmic Glass Ltd., and L&T Cement along with several medium and small industries as documented in the State Industrial Profile of West Bengal, 2018–2019 (SIP, 2019). Besides being an industrial hub, Durgapur has advanced roadways with the National Highway passing across the city. Moreover, State Highway number nine connects four important districts of West Bengal. Industries in combination with vehicles and coal-based domestic stoves contribute to the emission of air pollutants such as Suspended Particulate Matter (SPM_{10}), Oxides of Sulphur (SO_x), and Oxides of Nitrogen (NO_x) to the ambient air. The main sources of these pollutants are internal combustion engines, fossil fuel-fired power stations, and industrial combustion (Holman, 1999). The concentration of SPM_{10} depends on every particle emitted (particles generally less than 10μm aerodynamic diameter) from industries, traffic, and burnings of domestic waste (Holman, 1999).

The increasing levels of pollution in recent years are a global concern. Air quality management and air quality limit have played a major role in ecological balancing and maintaining the temperature, humidity, and rainfall. In Romania (Pohoata and Lungu, 2017), statistical analysis using advanced Principal component analysis (PCA) and Auto-regressive integrated moving average(ARIMA)limitations of prediction models used to forecast the environmental air qualitywas assessed. ARIMA models were popularized by Box and Jenkins (1976), and therefore,these are known as Box-Jenkins models.

Moreover, Box and Tiao (1975) discussed the general transfer function model with ARIMA.ARIMA procedure analyses and forecasts equally spaced univariate time series data. ARIMA supports seasonal, subset, and factored models, ARMA errors, and rational transfer function models. Forecast models can be constructed using stochastic methods like ARIMA (Shamsnia et al., 2011). A stochastic model describes the probability structure of a sequence of observations. A statistic of N successive observations $\mathbf{z}' = (z_1, z_2, \ldots, z_N)$ is considered a sample realization. From an infinite population of such samples, a probability forecast model can be processed (Box et al., 2015). The time series model could be further used to forecast the future trends in air pollution with this modeling process. To examine the concentrations of primary air pollutants like NO, NO_2, NO_x, PM_{10}, SO_2, and ground-level O_3, factor analysis and Box-Jenkins methods were applied in a study conducted in Blagoevgrad, Bulgaria. Seasonal ARIMA (SARIMA) models present statistical approaches that allow the building of non-complex models that are effective for short-term air pollution forecasting including time series forecasting of air pollution in small urban areas (Gocheva-Ilieva et al., 2014) and useful for cautionary purposes in urban areas (Ivanov and Gocheva-Ilieva, 2013).

In India, prediction models in pollution analysis have become a major field of study in recent times. Air quality studies in various parts of India showed remarkably high levels of air pollution. An air quality study performed in Delhi (1987–2006) analyzed the effects of the measures taken to control air pollution (Firdaus and Ahmed, 2011). Subsequently, in 2017, the trends in air quality in four major cities of India, namely New Delhi, Mumbai, Chennai, and Bengaluru were evaluated using ARIMA (Yadav and Toshniwal, 2017). In 2007, the forecast models were constructed for four major air pollutants at the urban traffic site of Delhi using ARIMA (Kumar and Jain, 2010). The daily air quality index in Delhi was measured and forecasted for each season using ARIMA, Principal Component Regression (PCR), and a combination of both (Kumar and Goyal, 2011). These studies provided evidence for the increasing levels of air pollution supported by statistical models and predictions.

Previous studies in West Bengal mostly focuson the effects of air quality on the surroundings. West Bengal has seven non-attainment zones designated by the Central Government. Durgapur, Bidhannagar is a residential area in West Bengal with significant levels of biodiversity and surrounded by industries. Durgapur has been designated as a non-attainment zone due to high levels of air pollution by Central Pollution Control Board (CPCB, 2019) under the National Clean Air Programme by the Ministry of Environment and Forest (MoEF, 2019). Effects of air pollution on plants are evident from previous studies (Palit et al., 2013; Banerjee et al., 2016; Sarkar et al., 2021). Air pollution has affected humans in terms of eye and nose irritation with respiratory discomfort in the concerned study area (Nandi and Gorain, 2009). A study on the ecosystem services and the impact of industrial pollution on urban health in the concerned study site focus on a review of the effects of air pollution on human and animal populations regarding various diseases (like bronchitis, cardiopulmonary arrest, cancer, liver, and kidney dysfunction)(Banerjee et al., 2021).

In the present study predictive forecast models (SARIMA) are developed to understand the trend of ambient air quality of Durgapur Industrial Belt. To the best of our knowledge, there have been no studies reported on the time series modeling of air pollutants in this area.The data obtained from WBPCB are not in regular interval. To use this data to predict models it is important to process it in either month-wise or season-wise format. The study mainly focussed on constructing time series prediction models for major air pollutants (SPM_{10}, SO_x, NO_x) from standardised season-wise and month-wise dataset. The main aim of the study is to compare between month-wise and season-wisevariation and then predict the best forecast model. Predictive forecast models can be used to monitor air levels and study trends in ambient air quality.

2 Materials and Method

2.1 Study area

This study was performed in Bidhannagar, an area of Durgapur, West Bengal which is a residential area located at 23.5178°N, 87.3460°E. It is surrounded by various industries and experiences a heavy load of transportation. Moreover, the weather at this site is characterized by high temperatures in summer, low to heavy rainfall in monsoon, anddry winter (Choudhury et al., 2019). A satellite image of the study site is presented in Fig. 1.

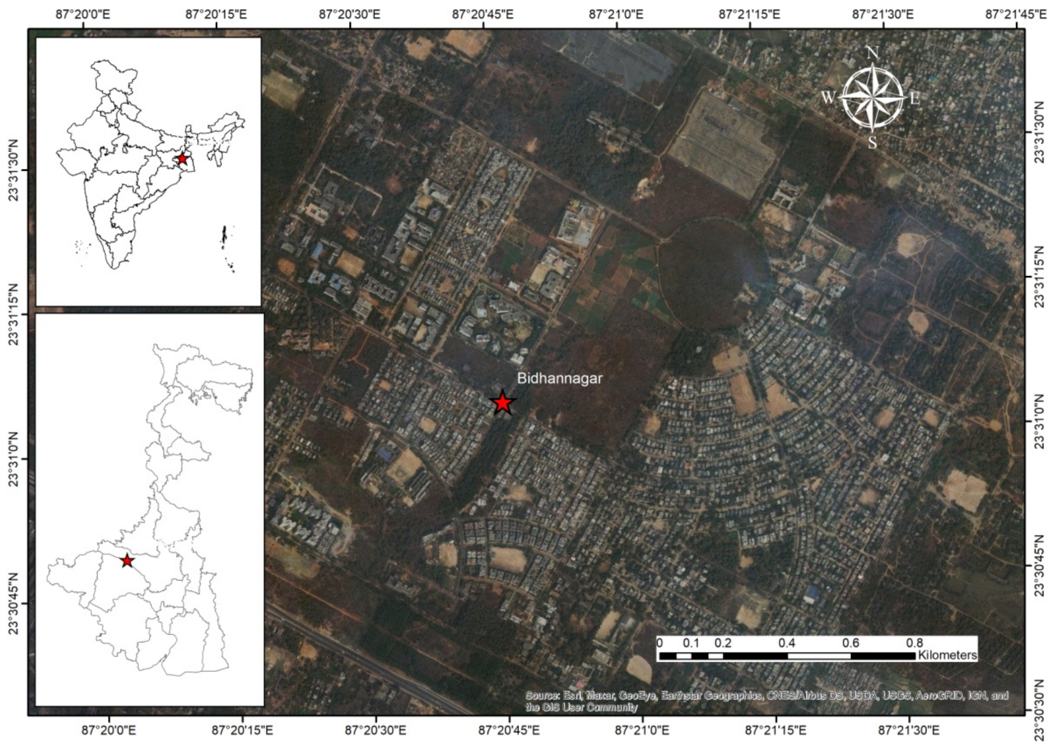


Fig. 1 Satellite map of study area.

2.2 Data analysis

The data was obtained from the official website of the West Bengal Pollution Control Board (WBPCB).The concentrations SO_x, NO_x, and SPM_{10} were recorded from 2007 to 2019 for 13 years by the WBPCB. Since the data was not recorded in regular intervals, the month-wise and season-wise average of the data was calculated. Seasons are a natural cycle whereas months and years are man-made projections of time. The seasonal analysis was performed across four seasons: Summer, Monsoon, Post-monsoon, and Winter. Scatter plots were constructed in Microsoft Excel (version 2011) for the selected pollutants to compare the ambient concentration of the pollutants with the threshold limits prescribed by the CPCB (2009). Season-wise and month-wise polar graphs were plotted to study the concentration of parameters across seasons and months, respectively. Polar graphs represent a system of coordinates in which the location of a point is determined by its distance from a fixed point at the center. Thesegraphs allow independent control over grid size and bend positions (Duncan and Kobourov, 2003) and help in better representation of the trends of data. Finally, the SARIMA model was performed to study the trends in season-wise and month-wise data. Similarly, a forecast model was constructed to understand the future trendsfor five years. Cross-validation (CV) was used for algorithm selection (Arlot and Celisse, 2010). In this CV process, the data is split, once or several times, for estimating the risk of each algorithm. Part of the data (the training sample) is used for training each algorithm, and the remaining part (the validation sample) is used for estimating the risk of the algorithm. Thus, CV selects the algorithm with the smallest estimated risk. This increases the data and model reliability. The models were constructed using 80% of the data for training and 20% of the data for validation. All statistical analysis was performed in R 4.0.5 (Team, 2021) using packages'tseries' (Trapletti and Hornik, 2019)and forecast (Hyndman et al., 2018), and plots were madewith 'ggplot2'(Wickham et al., 2016).

2.3 Modeling approach

2.3.1 ARIMA model

ARIMA is a powerful tool for predicting future trends and understanding the future circumstances and their probable impact in a given scenario. Moving Average (MA) schemes were used by Slutsky (1937).Yule (1926) was the first to introduce AutoRegressive (AR) models. Wold (1938) studied the analysis of stationary Time Series Model. ARIMA models are the most generalised models of ARMA enhanced by degree differencing from changing non-stationary data to stationary data (Ong et al., 2005). The partial autocorrelation function (PACF) is used to check the degree of association regarding the lag value.Time series are predicted and confidence intervals are generated for these forecasts based on the ARIMA model. The equation for ARIMA (p, d, q) is:

$$(1- \phi_1 B_1- \phi_2 B_2- \phi_3 B_3-\ldots\ldots \phi_i B_i)(1 - B)yt = (1 + \theta_1 B+\ldots\ldots \theta_j B_i)et$$

where yt and et are the actual value and random error at time t and ϕ_i, respectively, and θ_j are the model parameters (Box and Jenkins, 1976).

A non-seasonal ARIMA model is classified as an "ARIMA (p, d, q)" model, whereas in SARIMA (p, d, q) (P, D, Q) the seasonality of the data is also considered, where p = autoregressive terms, d = number of non-seasonal differences, and q = number of lagged forecast errors. Generally, a non-seasonal stationary time series can be modeled as a combination of past values and errors.

The generalized SARIMA equation is represented as:

$$\varphi(B)\Phi(Bs)(1 - B)d(1 - Bs)DXt = \theta(B)\Theta(Bs)Zt$$

where Φ, Θ denote polynomials in Bs of order P, Q, respectively (Chatfield, 2000).

2.3.2 Measures of goodness of fit

Goodness-of-fit statistics are used to compare modelsand evaluate the prediction performance. The measure of accuracy was performed following the below statistics.

Root Mean Square Error (RMSE)

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

Mean Absolute Percent Error (MAPE):

$$\frac{\sum_{t=1}^{n}|y_t - \hat{y}_t|}{n}$$

Mean Absolute Error (MAE):

$$\frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right|$$

The stationarity of data was checked following the augmented Dickey-Fuller test (ADF) test (Dickey and Fuller, 1979) using the Ljungs Box Q Statistics (Qm).

Ljungs Box Q Statistics (Qm):

$$Q_m = n(n+2)\sum_{k=1}^{m}\frac{{r_k}^2}{n-k} \approx {\chi^2}_{m-r}$$

The Ljung-Box test is based on the statistic where *m* is the length of the time series, r_k is the *k* autocorrelation coefficient of the residuals, and *n* is the number of lags to test.

3 Results

3.1 Statistical analysis

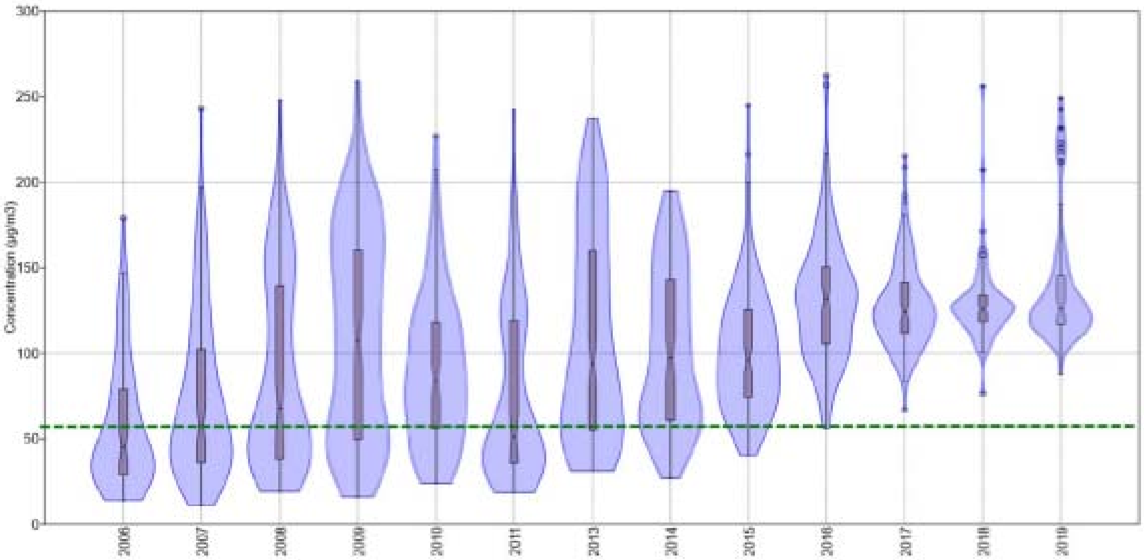
The analysis of SPM_{10}, SO_x, and NO_x was conducted both month-wise and season-wise. The data was assorted accordingly from the year 2007 to 2019, as shown in Table 1. The yearly data of the selected parameters were compared with the permissible threshold limits prescribed by the CPCB (Central Pollution Control Board, 2009), as shown in Fig. 2. The SPM_{10} levels were always higher than the permissible threshold limits and gradually increased since 2015. However, the NO_x levels were higher than the limit from 2007 till 2014. After 2014 there was a dip in the levels, and it was below the permissible limits. On the other hand, SO_x levels were always below the permissible limits.

Table 1 Data summary of Bidhannagar from 2007 to 2019 consisting of SPM_{10}, NOx, and SOx.

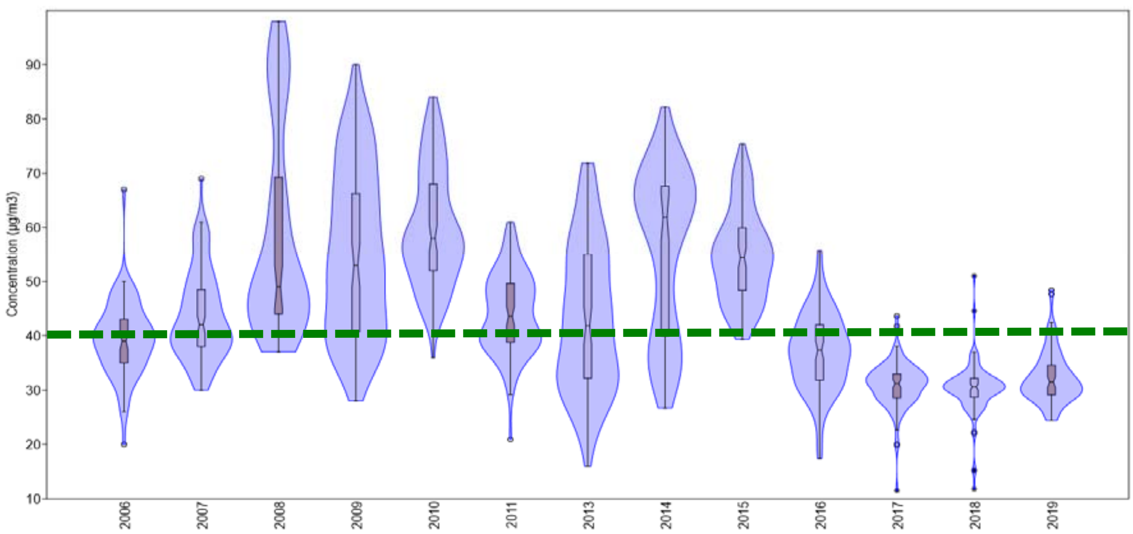
	SPM_{10} (µg/m³)				**NOx (µg/m³)**				**SOx (µg/m³)**			
Year	Max	Min	Mean	St. Dev	Max	Min	Mean	St. Dev	Max	Min	Mean	St. Dev
2007	142.56	29.88	73.15	39.73	58.67	33.56	43.60	7.24	6.13	4.33	5.05	0.51
2008	176.33	30.50	90.89	51.76	89.63	42.00	58.85	19.09	6.56	4.70	5.48	0.62
2009	197.50	39.43	109.83	53.78	76.11	36.00	54.89	14.17	12.44	5.57	8.14	2.17
2010	149.22	36.33	89.57	37.03	78.78	46.67	59.71	10.14	8.63	5.56	7.09	0.91

2011	152.04	29.81	80.65	47.78	57.08	35.77	44.78	6.66	8.47	4.14	5.85	1.60
2012	173.50	38.21	93.28	50.01	62.07	30.76	44.19	9.57	9.59	4.77	6.95	1.78
2013	194.95	46.61	105.91	52.24	67.06	25.74	43.60	12.49	10.71	5.39	8.05	1.97
2014	148.27	50.55	101.98	38.31	72.10	34.08	56.25	14.64	11.38	5.95	8.52	1.83
2015	147.23	63.55	101.57	29.15	68.52	44.66	54.61	7.43	10.75	6.23	7.81	1.39
2016	156.05	90.76	131.39	22.42	43.89	28.23	37.06	5.34	16.55	8.84	13.00	2.65
2017	170.61	107.50	128.11	20.12	33.92	29.15	30.98	1.73	14.08	8.99	11.70	1.57
2018	151.41	114.53	127.90	11.17	34.23	25.48	30.25	2.49	15.13	9.91	12.85	1.57
2019	205.24	115.60	135.38	24.08	39.46	29.01	31.91	2.78	15.32	9.45	10.73	1.71

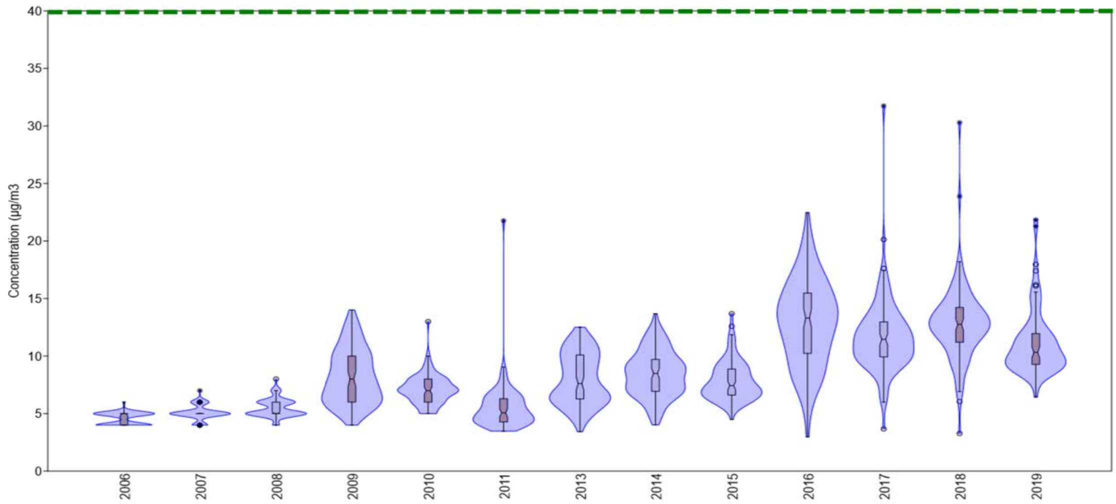
Max=Maximun; Min=Minimum; St. Dev=Standard Deviation.



a) SPM_{10}



b) NO_x



c) SO_x

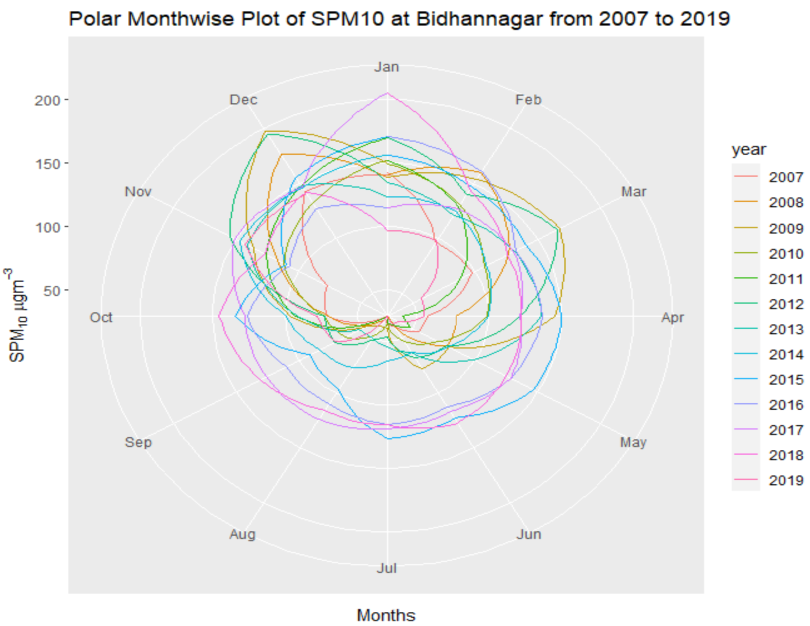
Fig. 2 Annual concentration of a) SPM_{10}, b) NO_x and c) SO_x.

3.2 Polar Graphs

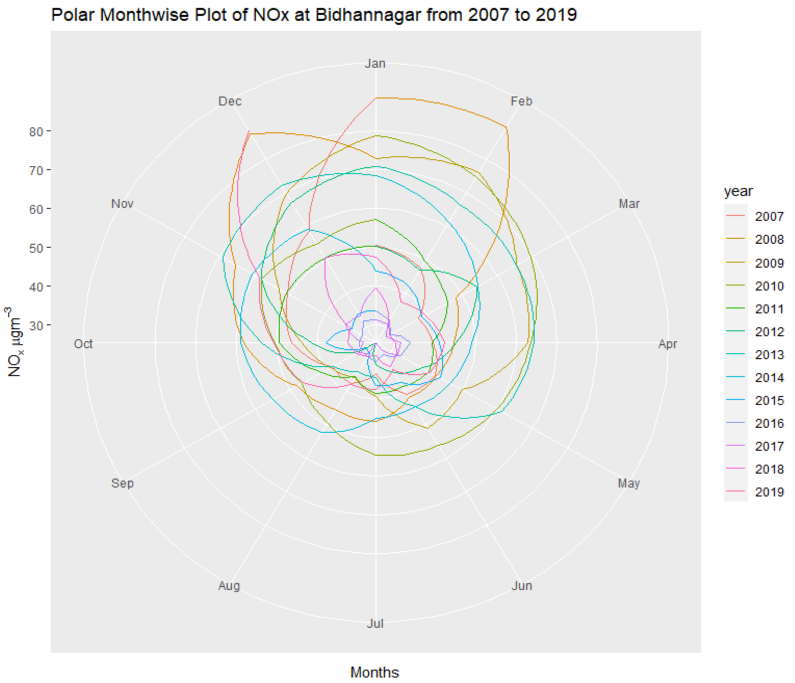
Polar graphs were plotted to represent the distribution of pollutants across months and seasons. Fig. 3(1) and Fig. 3(2) show the month-wise and season-wise polar graphs, respectively, from the year 2007 to 2019. The changes are represented circularly to indicate the cyclic pattern of the months and seasons. The x-axis represents the months and seasons in each plot, respectively. The y-axis represents the concentration of the parameter under study.

3.2.1 Month-wise Polar graphs

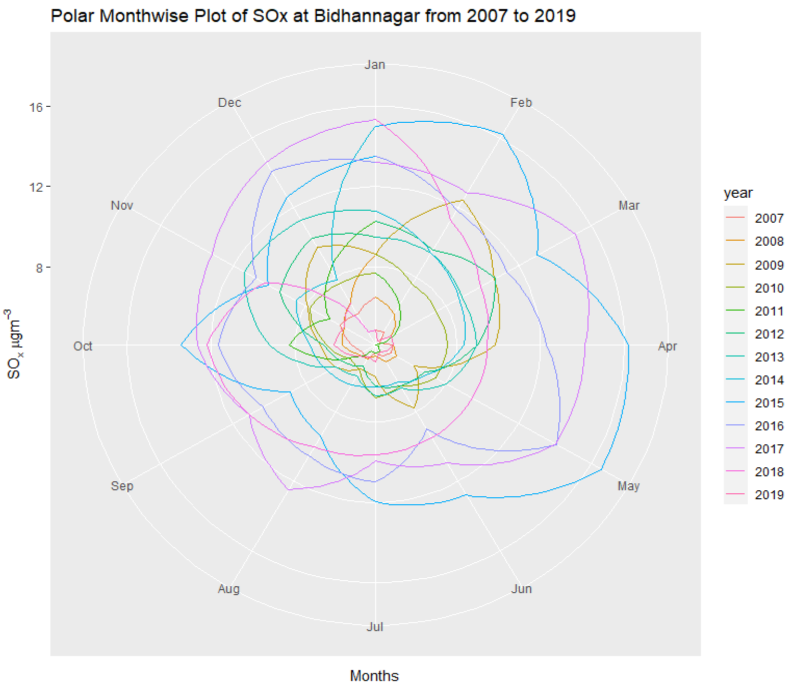
The month-wise polar graphs for all the three parameters in Fig. 3(1) show random distribution with a very minor sharp peak in January and December.



a) SPM_{10}



b) NO_x

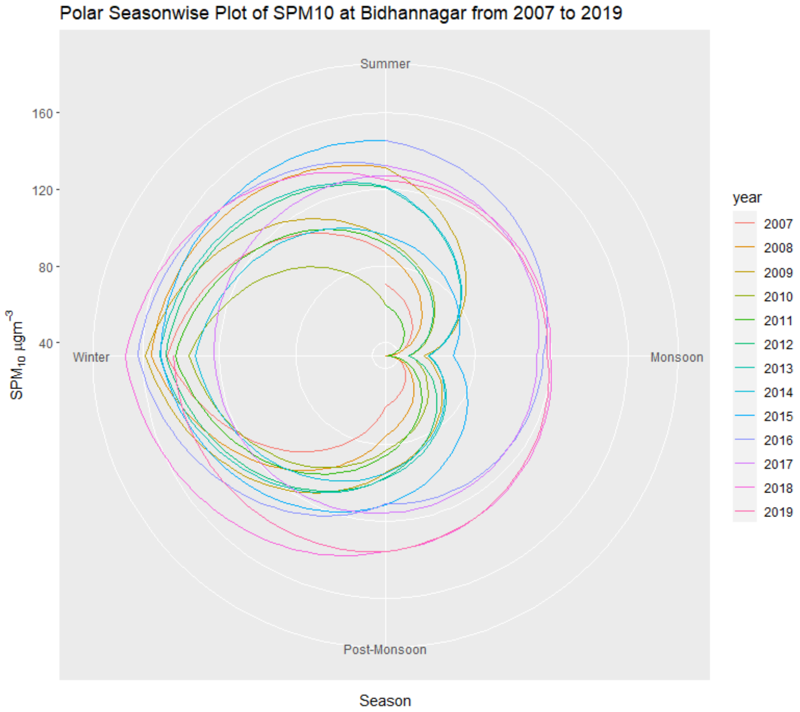


c) SO_x

Fig. 3(1) The month-wise polar plots at Bidhannagar from 2007 to 2019. a) SPM_{10}, b) NO_x, c) SO_x.

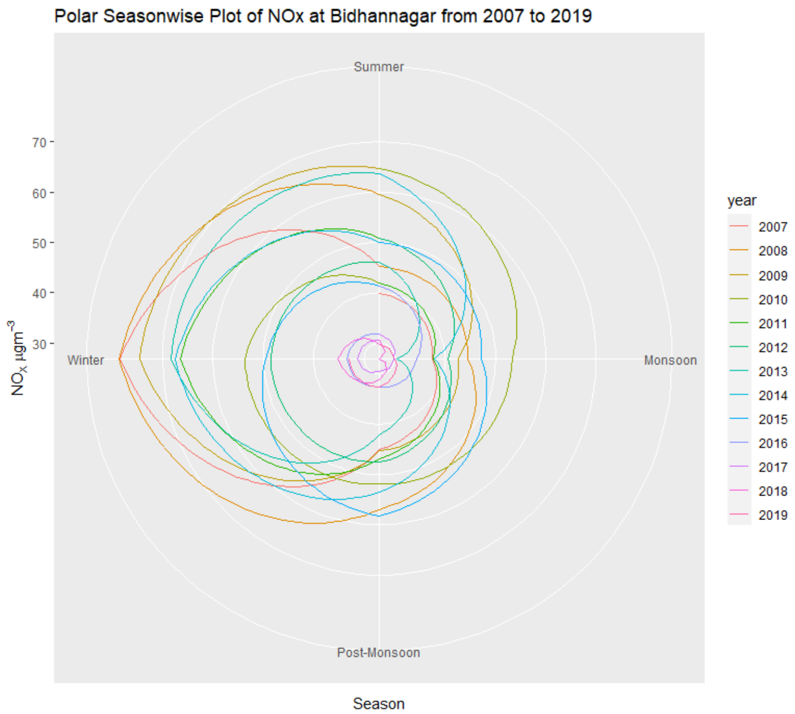
3.2.2 Season-wise Polar Graphs

In Fig. 3(2), both SPM_{10} and NO_x graphs show a consistent clustering and tapering towards the winter season every year whereas the monsoon experiences the least concentration with a blunt end. The SO_x graph does not indicate any preference for seasons.



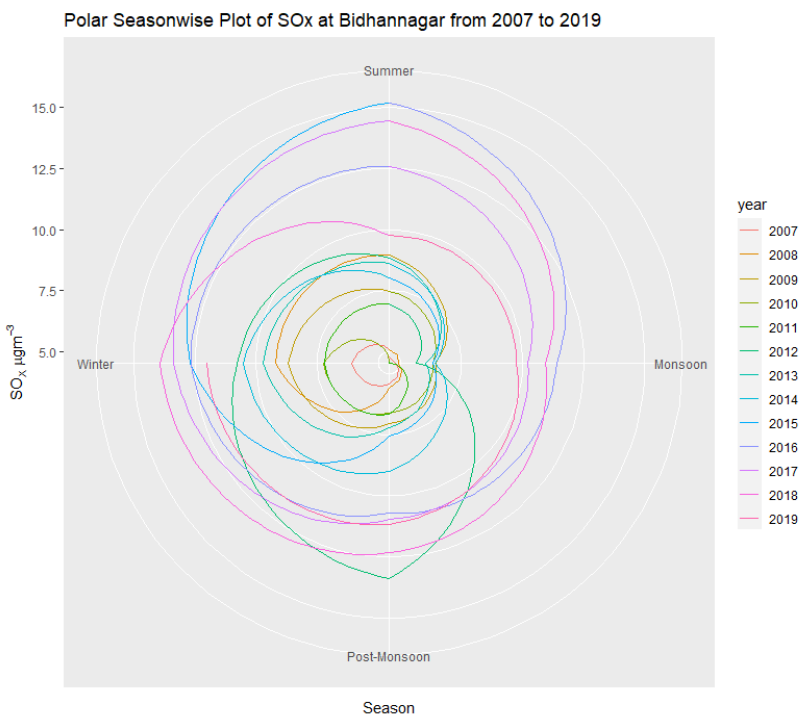
Polar Seasonwise Plot of SPM10 at Bidhannagar from 2007 to 2019
Summer
Winter
Monsoon
Post-Monsoon
160
120
80
40
SPM10 µgm-3
Season
year
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019

a) SPM_{10}



Polar Seasonwise Plot of NOx at Bidhannagar from 2007 to 2019
Summer
Winter
Monsoon
Post-Monsoon
70
60
50
40
30
NOx µgm-3
Season
year
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019

b) NO_x



c) SO_x

Fig. 3(2) The season-wise polar plots at Bidhannagar from 2007 to 2019 a) SPM_{10}, b) NO_x, c) SO_x.

3.3 Time series using ARIMA

Since the annual scatter plot showed that the levels of SO_xwere within the permissible threshold limits, running a forecast trend for SO_x levels would not help study pollution trends. On the other hand, SPM_{10} levels remained high in the study period and NO_x levels decreased in recent years. Therefore, SPM_{10} and NO_x data are used to perform season-wise and month-wise predictions using ARIMA models. The average monthly data was sorted and computed. Table 2 summarizes the season-wise and month-wise ARIMA models.

The ARIMA goodness of fit prediction models for each pollutant shows a four-year forecast in blue with the best fit model in red and finally the time series for 13 years in black for a better understanding of the plot. The month-wise forecast model shows a high rate of fluctuation throughout the year for both SPM_{10} and NO_x. Although the model though shows a decrease in the levels of NO_x, future prediction of trends is difficult. Therefore, the month-wise prediction and future trend analysis for SPM_{10} is not conclusive.

On the contrary, the season-wise model shows a distinct trend for SPM_{10} over the years. The graph clearly shows that there is an increase in the SPM_{10} levels over the years. The prediction shows a steep increase in the future. The NO_x level over the time scale shows a decrease since 2016. The trend follows minor fluctuations but the levels are below permissible limits.

3.3.1 Month-wise ARIMA

3.3.1.1 Suspended Particulate Matter (SPM_{10})

The month-wise SPM_{10} time series plot, the autocorrelation function (ACF), and PACF from the year 2007 to 2019 are represented in Fig. 4(1).

Corresponding to the time series model the month-wise prediction model for the period of the next five years with the best fit line for the goodness of fit model (1,1,1)(0,1,1) is represented in Fig. 4(2).

Table 2 Summary of ARIMA models.

Parameter	Model Criteria	Month-wise Model	Season-wise Model
SPM_{10}	ARIMA Model	(1, 1, 1) (0, 1, 1)	(1, 0, 0) (0, 1, 1)
	p-value	0.2551	0.2377
NOx	ARIMA Model	(2, 1, 2) (0, 0, 1)	(1, 0, 0) (0, 1, 1)
	p-value	0.7518	0.3191

3.3.1.2 Oxides of Nitrogen (NO_x)

The time series plot of 13 years with the ACF and PACF plot for the model (2,1,2)(0,0,1) of month-wise NO_x is shown in Fig. 4(3). ARIMA (2,1,2)(0,0,1) showed the goodness of fit model with the best fit model, and a prediction of the next five years is represented in Fig. 4(4).

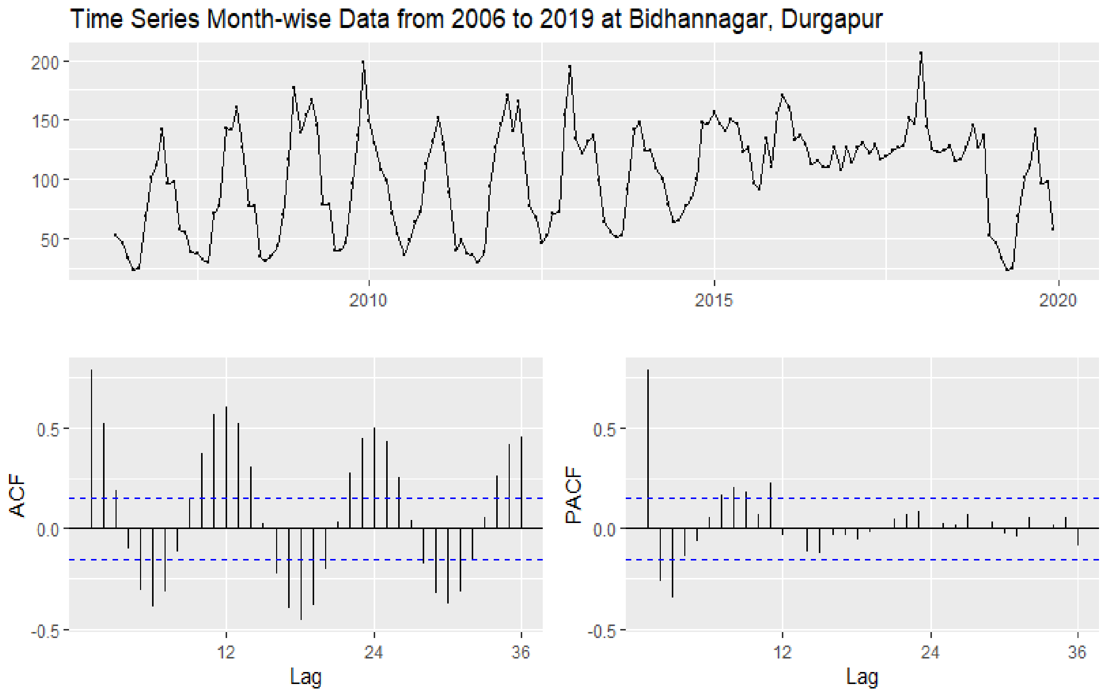


Fig. 4(1) Month-wise trend of SPM_{10}.

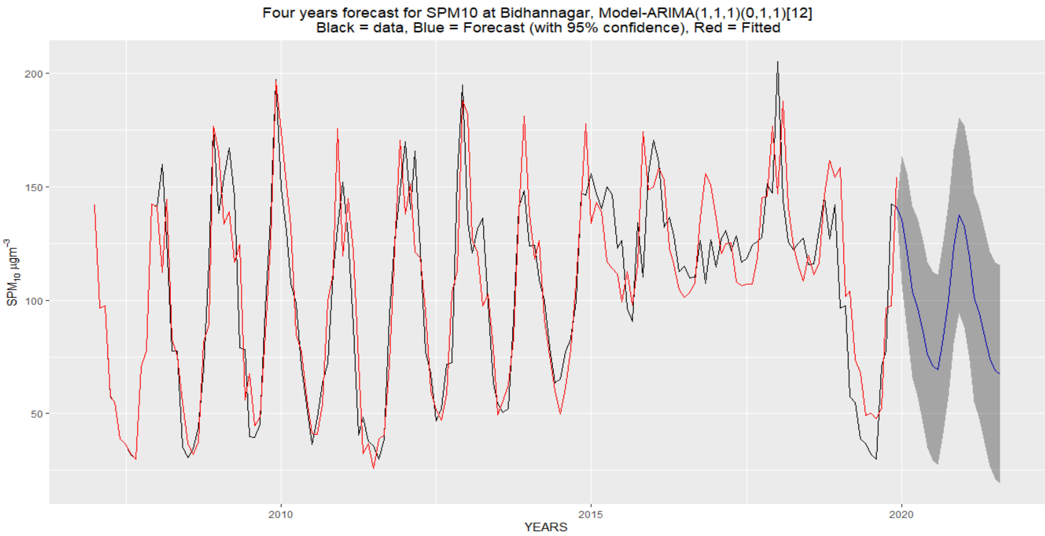


Fig. 4(2) Forecast ARIMA model (1,1,1)(0,1,1) of SPM_{10}.

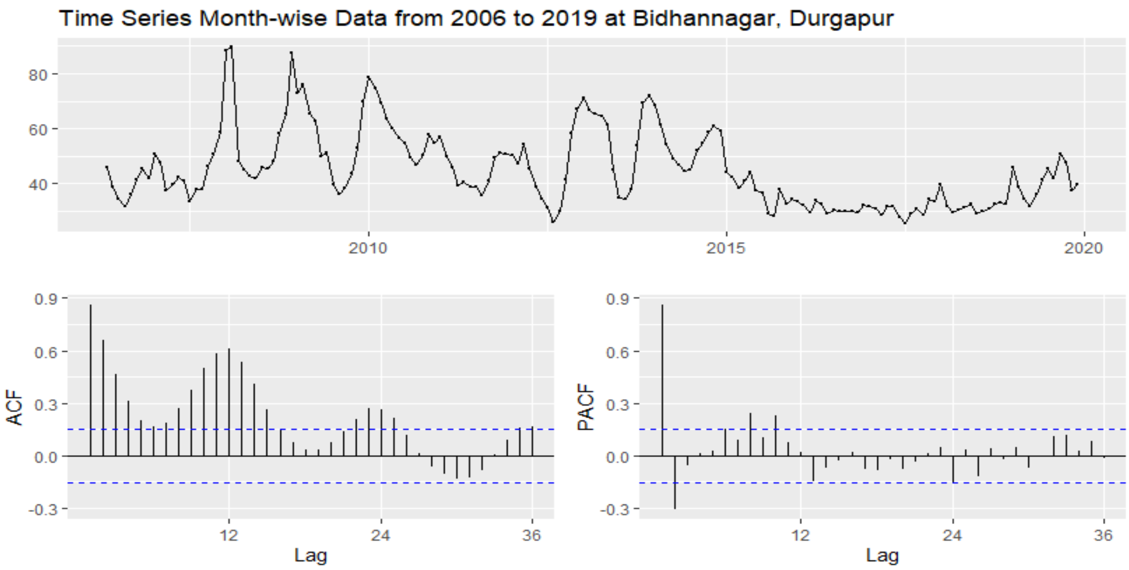


Fig. 4(3) Month-wise trend of NO_x.

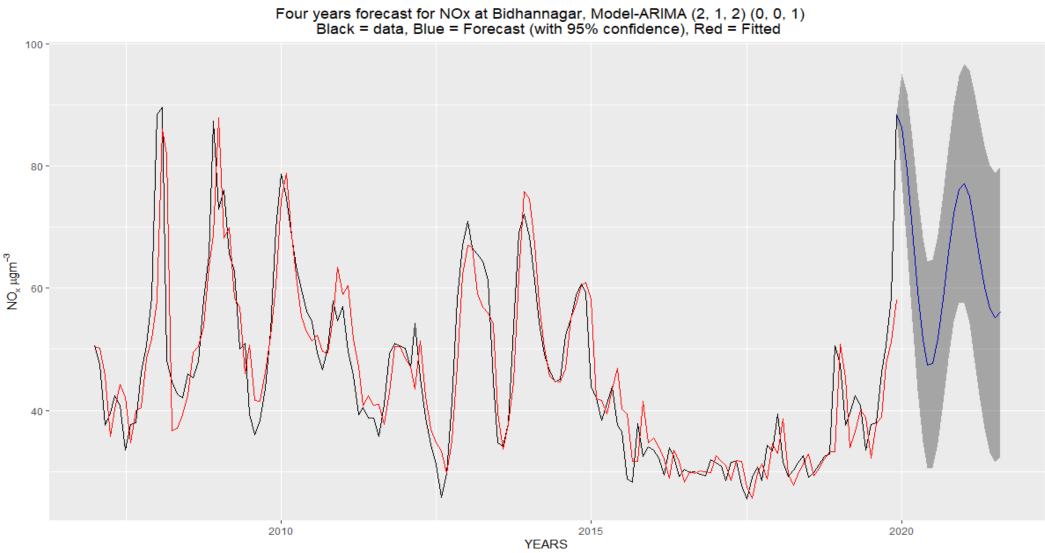


Fig. 4(4) Forecast ARIMA model (2,1,2)(0,0,1) of NO_x.

3.3.2 Season-wise ARIMA

3.3.2.1 Suspended Particulate Matter (SPM_{10})

The season-wise representation of the SPM_{10} time series from 2007 to 2019 along with its ACF and PACF plots for the goodness of fit model (1,0,0)(0,1,1) is shown in Fig. 5(1).

ARIMA (1,0,0)(0,1,1) showed the best fit model for the SPM_{10} season-wise prediction model with four years of the forecast is represented in Fig. 5(2).

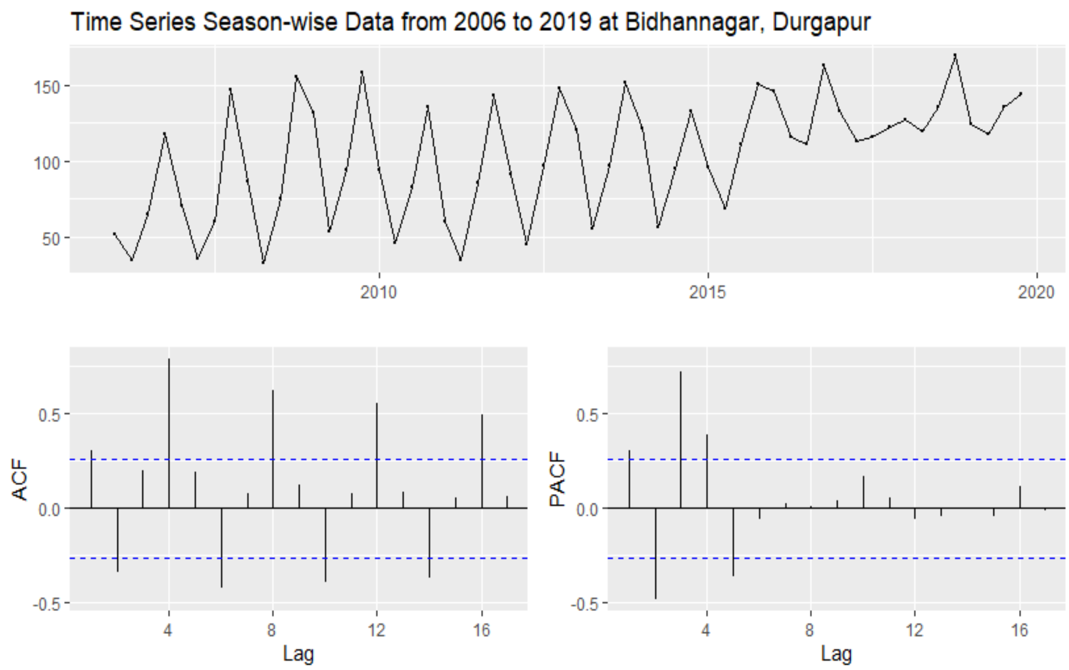


Fig. 5(1) Season-wise trend of SPM_{10}.

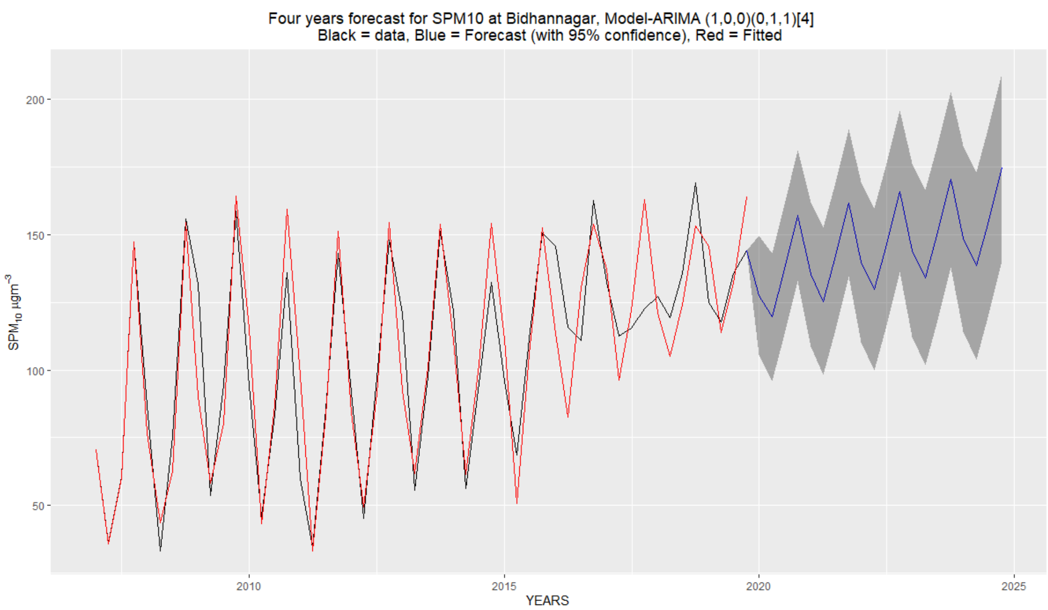


Fig. 5(2) Forecast ARIMA model (1,0,0)(0,1,1) of SPM_{10}.

3.3.2.2 Oxides of Nitrogen (NO_x)

The ACF and PACF plots of NO_x and the time series plot from 2007 to 2013 for 13 years are shown in Fig. 5(3).

ARIMA (1,0,0)(0,1,1) showed the goodness of fit model for NO_x season-wise plot with the four years prediction in Fig. 5(4).

CV was performed both in the case of season-wise and month-wise analysis.The complete data set was divided into two subsets: the training set and test set comprising 80% and 20% of the data, respectively. The training set was used to prepare the model.The test set was used to check the efficiency of the model and validate the forecast model. Table 3 represents the summary of the models'validation. The validation table indicates that the RMSE, MAPE, and MAE values are consistently low in the season-wise model compared to the month-wise model (Table 3). In the case of NO_x, the training set values for the season-wise model are higher, but the test set values are lower than those in the month-wise models.

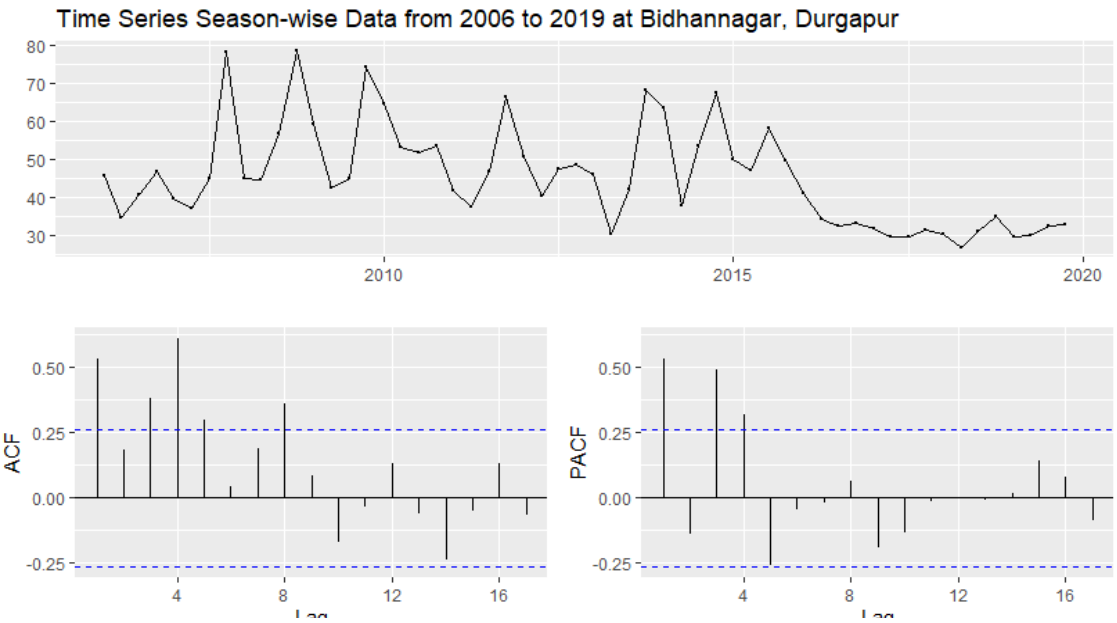


Fig. 5(3) Season-wisetrend of NO_x.

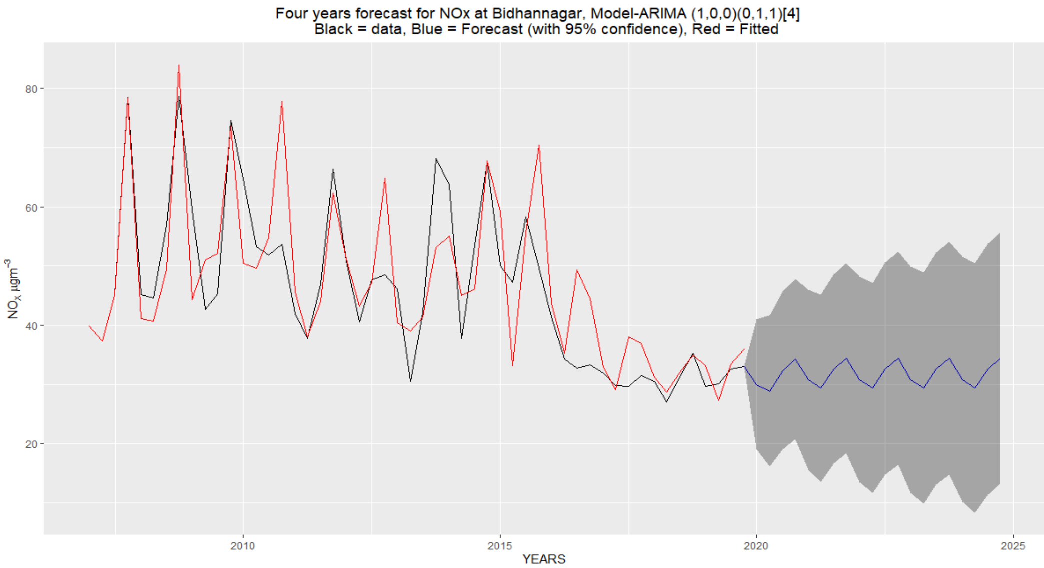


Fig. 5(4) Forecast ARIMA model (1,0,0)(0,1,1) of NO_x.

Table 3 The validation values of the models.

		Month-wise			Season-wise		
		MAPE	RMSE	MAE	MAPE	RMSE	MAE
SPM_{10}	Training set	14.89	18.19	13.63	9.38	13.35	9.04
	Test set	35.65	49.77	44.08	16.47	24.74	21.18
NOx	Training set	10.09	7.13	5.02	13.42	8.89	6.37
	Test set	49.02	15.69	14.94	28.90	9.82	8.99

4 Discussion

4.1 Comparative study of month-wise and season-wise polar graphs

It can be concluded that month-wise polar graphs (Fig. 3(1)) do not clearly indicate a particular month with high levels of air pollution. But in comparison, the season-wise polar graphs (Fig. 3(2)) provided a better representation of the seasonality of the pollutants. SPM_{10} graph showsconsistent clustering and tapering towards the winter season with a blunt end in the monsoon season over the years. This indicates a high level of SPM_{10} concentration during winter. Similar projections were seen for NO_x graph as well. But the levels of NO_x concentration seemed to be lower than SPM_{10} in Durgapur.

Different studies carried out globally including India has established that levels of NO_x, SO_x and SPM_{10} remain high in Industrial Belt (Thangam et al., 2016; Gopu et al., 2021; Abhilash et al., 2018; Naveen and Anu, 2017; Nadeem et al., 2020). The present study also reflects similar kind of observations in case of SPM_{10} and NO_x. But surprisingly, levels of SO_x have been found to be within the prescribed limits of CPCB. The probable reasons could be the absence of source of SO_x or the implementation of dry/wet scrubbers for chimneys as studied in various industries in India (Mittal et al., 2012). But further study is required to confirm such findings.

A study on mapping air quality in various metropolitan cities suggests that winter inversion leads to higher values of pollutants that decrease in monsoon due to precipitation (Dubey et al., 2021). A similar meteorological study in Delhi showed that irrespective of the levels of emission of the pollutants, winter has the highest pollution, and the months of November, December, and January experienced the highest emission (Guttikunda and Gurjar, 2012). Our study also has conformity with these findings.

4.2 Comparative study of month-wise and season-wise time series models

The prediction models indicate better results for season-wise models (Fig. 5(2) and Fig. 5(4)) compared to month-wise models (Fig. 4(2) and Fig. 4(4)). There is an increasing trend in the levels of concentration of SPM_{10} (Fig. 5(2)) for the next 4 years.Durgapur is continuously developing in terms of industries viz. steel, alloy, cement, etc (SIP, 2019). Such kind of industrial growth can be the possible reason for the higher levels of SPM_{10} as predicted in our model. The levels of NO_xconcentrations (Fig. 5(4)) has a decreasing trend and predicts to remain same in next 4 yearsi.e.within permissible limits as prescribed by CPCB. It can be hypothesized from the current study that the decrease in NO_x levels may be due to the modified mechanisms implemented in various industries to check the emissions. The decrease in the level of NO_x could be due to the reduction in fossil fuel combustion and biomass burning which is considered to be essential source of NO_x (Delmas et al., 1997). Along with this, regular monitoring and maintenance of the automobiles could be a probable reason for the decrease in the levels of NO_x. The implementation of biodiesel and introduction of

Exhaust Gas Recirculation system (EGR) in vehicles reduces the NO_x emission by controlling the temperature (Rajendran et al., 2020).

Table 2 provides the best fit model in both month-wise and season-wise predictions. Table 3 summarizes the validation of the best fit model to justify the models to be accurate and precise.

The study of the seasonality trend of the pollutants from the predicted models is believed to provide a scientific and logical representation of the conditions of the ambient air quality. In a previous study, time series model prediction and trend variability of aerosolswas conducted at various sites in the Coal Belts of India (Soni et al., 2015). The Raniganj Coalfields had the highest aerosol optical density out of the 11 study sites which concludes ARIMA to be a satisfactory prediction model. A study used ARIMA models to forecast the probable trends of SPM_{10} in cities of Turkey,one of the most polluted countries of Europe, and identify the most and least polluted areas and take precautionary measures accordingly (Cekim, 2020). A study conducted in the global level urban areas showedthe effects of climate change and mitigation measures. The study concluded that particle mass ($PM_{2.5}$) and particle number concentrations (PNC) were impacted differently based on the changing climate and the mitigation measures.Thus, both parametersmust be considered in urban air quality management plans (Jesus et al., 2020). The ARIMA based seosonal modelling used in the current study also predicts that levels of SPM_{10} will remain high in the next four yearsand levels of NO_x will gradually decline in the coming four years.

Forecast modeling using regression analysis has helped to understand future threats that can occur due to air pollution.As shownin the pollution study in the Nanded city in Maharashtra, ARIMA modeling was used as a potent method of prediction in developing a smart city (Kulkarni et al., 2018). Similar prediction models could be used to convert a planned city like Durgapur into a smart city which is the basic premise of the current study.

5 Conclusion

High concentrations of particulate matter and other pollutants in an Industrial belt like Durgapur, which is surrounded by Raniganj Coalfields, have increased over the years. An industrial area is always consideredto be polluted due to huge emissions and effluents.The present study has inferred that though SPM_{10} levels will show an increasing trend in future, NO_x levels will gradually decrease, whereas SO_x will remain under permissible limits.As SPM_{10} levels are alarmingly high and if left unchecked, it will lead to human respiratory disorders and biodiversity loss. Our study shows that season-wise prediction models using ARIMA to be more suitable and reliable for understanding trends and forecasting. The study also projects season-wise dataset to be more suitable for air quality modelling. It also solves the problem of managing irregular dataset in time series model.Season-wise modelsprovide an accurate understanding of the seasonal variation of air pollutants. ARIMA-based predictions will enable the identification of reasons for air pollution in Durgapur and help in finding feasible solutions for addressing air pollution especially in case of SPM_{10}. Our work is the first time series prediction study for air pollution in the Durgapur Industrial Belt, West Bengal. This study will form the baseline for future studies using suitable prediction models and willhelp inimplementing policy measures that canaid in reduction of air pollutionand related health problems.

Acknowledgment

Authors acknowledge the Department of Higher Education, Science & Technology, and Biotechnology, Government of West Bengal (Science & technology branch and biotechnology branch) Sanction no. 208(Sanc.)/ST/P/S&T/5G-17/2018 dated 20/02/2019 for the necessary financial aid to carry out the work. We are also grateful to Dr. Debnath Palit, Principal, Durgapur Government College for the infrastructural support to carry out the work. The author is thankful to Dr. Rajeshree Sanyal for editing the paper. The author also thanks Dr.Aloke Kumar Mukherjee, SACT-I, Department of Conservation Biology, Durgapur Government College for helping in the editing of the paper.This work was supported by Department of Higher Education, Science & Technology and Biotechnology, Govt. of West Bengal (WBDST), India (Sanction Order No. 208(Sanc.)/ST/P/S&T/5G-17/2018 dated 20/02/2019). Author Moitreyee Chakrabarty has received the grant from WBDST.

References

Abhilash MS, Thakur A, Gupta D, Sreevidya B. 2018. Time series analysis of air pollution in Bengaluru using ARIMA model. Ambient Communications and Computer Systems, 413-426

Arlot S, Celisse A. 2010. A survey of cross-validation procedures for model selection. Statistics Surveys, 4: 40-79

Banerjee S, Banerjee A, Palit D. 2021. Ecosystem services and impact of industrial pollution on urban health: evidence from Durgapur, West Bengal, India. Environmental Monitoring and Assessment, 193(11): 1-8

Banerjee S, Palit D, Banerjee A. 2016. Changing strategies of biochemical & physiological features of selected plant species on effect of air pollution in Eastern Steel City, Durgapur, India. International Journal of Current Microbiology and Applied Sciences, 5(9): 733-741

Box GE, Tiao GC. 1975. Intervention analysis with applications to economic and environmental problems. Journal of the American Statistical Association, 70(349): 70-79

Box GEP, Jenkins GM. 1976. Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco, California. USA

Box GEP, Jenkins GM, Reinsel GC, Ljung GM. 2015. Time Series Analysis: 362 Forecasting and Control. John Wiley & Sons, USA

Cekim HO. 2020. Forecasting PM10 concentrations using time series models: a case of the most polluted cities in Turkey. Environmental Science and Pollution Research, 27(20): 25612-25624

Central Pollution Control Board (CPCB). 2009. National Air Quality Standards. New Delhi, India

Central Pollution Control Board (CPCB). 2019. National Air Quality Monitoring Programme (NAMP). India

Chatfield C. 2000. Time-Series Forecasting. Chapman and Hall/CRC, Boca Raton, FL, USA

Choudhury D, Das K, Das A. 2019. Assessment of land use land cover changes and its impact on variations of land surface temperature in Asansol-Durgapur Development Region. The Egyptian Journal of Remote Sensing and Space Science, 22(2): 203-218

de Jesus AL, Thompson H, Knibbs LD, Kowalski M, Cyrys J, Niemi JV, Kousa A, Timonen H, Luoma K, Petäjä T, Beddows D. 2020. Long-term trends in PM2. 5 mass and particle number concentrations in urban air: The impacts of mitigation measures and extreme events due to changing climates. Environmental Pollution, 263: 114500

Delmas R, Serça D, Jambert C. 1997. Global inventory of NOx sources. Nutrient Cycling In Agroecosystems, 48(1): 51-60

Dickey DA, Fuller WA. 1979. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American statistical association, 74(366a): 427-431

Dubey R, Bharadwaj S, Zafar MI, Biswas S. 2021. Collaborative air quality mapping of different metropolitan cities of India. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 43: 87-94

Duncan CA, Kobourov SG. 2003. Polar Coordinate Drawing of Planar Graphs with Good Angular Resolution. Journal of Graph Algorithms and Applications, 7(4): 311-333

Firdaus G, Ahmad A. 2011. Changing air quality in Delhi, India: determinants, trends, and policy implications. Regional Environmental Change, 11(4): 743-752

Gocheva-Ilieva SG, Ivanov AV, Voynikova DS, Boyadzhiev DT. 2014. Time series analysis and forecasting for air pollution in small urban area: an SARIMA and factor analysis approach. Stochastic environmental research and risk assessment, 28(4): 1045-1060

Gopu P, Panda RR, Nagwani NK. 2021. Time series analysis using ARIMA model for air pollution prediction in Hyderabad city of India. Soft Computing and Signal Processing, 47-56

Guttikunda SK, Gurjar BR. 2012. Role of meteorology in seasonality of air pollution in megacity Delhi, India. Environmental monitoring and assessment, 184(5): 3199-3211

Holman C. 1999. Sources of air pollution. In: Air pollution and Health. 115-148, Academic Press, USA

Hyndman RJ, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, Petropoulos F, Razbash S, Wang E, Yasmeen F. 2018. Forecast: Forecasting functions for time series and linear models (Computer software manual). R package version 8.4.

Ivanov A, Gocheva-Ilieva S. 2013. Short-time particulate matter PM10 forecasts using predictive modeling techniques. AIP Conference Proceedings, 1561(1): 209-218

Kulkarni GE, Muley AA, Deshmukh NK, Bhalchandra PU. 2018. Autoregressive integrated moving average time series model for forecasting air pollution in Nanded city, Maharashtra, India. Modeling Earth Systems and Environment, 4(4): 1435-1444

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

Kumar U, Jain VK. 2010. ARIMA forecasting of ambient air pollutants (O_3, NO, NO_2 and CO). Stochastic Environmental Research and Risk Assessment, 24(5): 751-760

Ministry of Environmnt and Forest and Climate Change (MoEFCC). 2019. National Clean Air Programme (NCAP). India

Mittal ML, Sharma C, Singh R. 2012. Estimates of emissions from coal fired thermal power plants in India. International Emission Inventory Conference, 13-16

Nadeem I, Ilyas AM, Uduman PS. 2020. Analyzing and forecasting ambient air quality of Chennai city in India. Geography, Environment, Sustainability, 13(3): 13-21

Nandi PK, Gorain GC. 2009. People at Durgapur City. Indian Journal of Environmental Protection, 29(7): 597-604

Naveen V, Anu N. 2017. Time series analysis to forecast air quality indices in Thiruvananthapuram District, Kerala, India. International Journal of Engineering Research and Application, 7(6): 66-84

Ong CS, Huang JJ, Tzeng GH. 2005. Model identification of ARIMA family using genetic algorithms. Applied Mathematics and Computation, 164(3): 885-912

Palit D, Kar D, Misra P, Banerjee A. 2013. Assessment of air quality using several bio monitors of selected sites of Durgapur, Burdwan district by air pollution tolerance index approach. Indian Journal of Scientific Research, 4(1): 149-152

Pohoata A, Lungu E. 2017. A complex analysis employing ARIMA model and statistical methods on air pollutants recorded in Ploiesti, Romania. Revista de Chimie, 68(4): 818-823

Rajendran R, Gomez JP, Javed MM, Subbiah G. 2020. Reduction of NOx emissions with low viscous biofuel using exhaust gas recirculation technique. AIP Conference Proceedings, 2311(1): 020026

Sarkar S, Mondal K, Sanyal S, Chakrabarty M. 2021. Study of biochemical factors in assessing air pollution tolerance index of selected plant species in and around Durgapur industrial belt, India. Environmental Monitoring and Assessment, 193(8): 1-11

Shamsnia SA, Shahidi N, Liaghat A, Sarraf A, Vahdat SF. 2011. Modeling of weather parameters using stochastic methods (ARIMA model)(case study: Abadeh Region, Iran). InInternational conference on environment and industrial innovation, 12(1): 282-285

Soni K, Parmar KS, Kapoor S. 2015. Time series model prediction and trend variability of aerosol optical depth over coal mines in India. Environmental Science and Pollution Research, 22(5): 3652-3671

State Industrial Profile of West Bengal 2018-2019. 2019. Micro, Small & Medium Enterprises – Development Institute Ministry of MSME Government of India. http://www.msmedikolkata.gov.in/uploads/2021/03/SIP-2018-19.pdf

Team RC. 2021. R: A language and environment for statistical computing (4.0. 5) (Computer software). R Foundation for Statistical Computing.

Thangam ED, Narayanan R, Aedla R. 2016. Pollution dispersion modeling for concentrations of PM, SOX and NOX around Manali region, India. International Journal of Earth Sciences and Engineering, 9(3): 584-595

Trapletti, A., & Hornik, K. 2019. tseries: Time Series Analysis and Computational Finance. 452 https://CRAN.R-project.org/package=tseries, R package version 0.10-47

WBPCB. 2019. Report on daily ambient air quality. West Bengal Pollution Control Board. http://www.wbpcb.gov.in/html/aaq_info.shtml

Wickham H, Chang W, Wickham MH. 2016. Package ‘ggplot2’. Create elegant data visualisations using the grammar of graphics, 2(1): 1-89

Wold H. 1938. A study in the analysis of stationary time series. Doctoral dissertation. Almqvist & Wiksell, Stockholm, Sweden

Yadav A, Toshniwal D. 2017. Extracting Patterns and Variations in Air Quality of Four Tier I Cities in India. IEEE World Congress on Services (SERVICES). 17-20, India

Yule GU. 1926. Why do we sometimes get nonsense-correlations between Time-Series? a study in sampling and the nature of time-series. Journal of the Royal Statistical Society, 89(1): 1-64