A Time Series Study of Ahmedabad City for Air Quality

Sweta Patel[1], Chetna Bhavsar[2]

[1]Biostatistician, Statiza Statistical Services, Ahmedabad, India

[2]Professor & Guide, Department of Statistics, School of Sciences, Gujarat University, Ahmedabad, India

Abstract: ***Urban air pollution is rapidly becoming an environmental problem of public concern worldwide. It can influence public health and local/regional weather and climate. In the present study, air quality data were collected for a period of 7 years (2011–2017) at 6 locations in Ahmedabad, a mega city in Gujarat State in western India. The data were collected by the Central Pollution Control Board (CPCB). The observed air quality data were within the permissible limits set by the Central Pollution Control Board (CPCB). In this study, time series models will be discussed to analyse future air quality and used in modelling and forecasting future air quality in Ahmedabad next 9 years. Based on ARIMA model we develop models for the air quality parameters SO_2, NO_2, RSPM and SPM. Based on that model we forecasted the future air quality parameters for the Ahmedabad city. The approach demonstrates the potential to be applied to other areas of the city also.***

Keywords: Urban air pollution, ARIMA, Forecasting, Time series

1. Introduction

In keeping abreast with Ahmedabad's rapid economic development and to meet the nation's aspiration for an improved quality of life, clean-air legislation limiting industrial and automobile emissions was adopted in 1998. The need to monitor, control and possibly predict the slope of pollutants in proximity of human settlements is essential nowadays. Both chemical and physical agents affect the health of people living and working in areas with high concentrations of these pollutants. Development and use of statistical and other quantitative methods in the environmental sciences have been a major communication between environmental scientists and statisticians [1]. In recent years, many statistical analyses have been used to study air pollution as a common problem in urban areas [2]. The common descriptive statistical approach used for air quality measurement and modelling is rather limited as a method to understand behaviour and variability of air quality [3]. Many investigators have used probability models to explain temporal distribution of air pollutants [4-5]. Time series analysis is a useful tool for better understanding of cause and effect relationship in environmental pollution [6-8]. The main aim of time series analysis is to describe movement history of a particular variable in time. Many authors have tried to detect changing behaviour of air pollution through time using different techniques [9]. Many others have tried to relate air pollution to human health through time series analysis [10-12]. Therefore, this study aims at extending time series analysis to give both qualitative and quantitative information about air pollution of six different region of Ahmedabad city, and to predict future concentrations of air pollutant.

2. Materials and Methods

There are many tools available and have been adapted by different researchers to predict air pollutants concentration. It has been observed that many multi-parameter meteorological models either under predict or over predict the air pollutants concentration. In the air quality forecasting study, time-series analysis is good choice. Many researchers has successfully used a regression technique called ARIMA model to air pollutants and its analysis [13-15].

ARIMA linear models have dominated many areas of time series forecasting. ARIMA is the most popular linear model for forecasting time series. It has enjoyed great success during the last three decades. As the application of these models is very common, it is described here briefly. The linear function is based upon three parametric linear components: autoregression (AR), integration (I), and moving average (MA). The ARIMA models also have the capability to include external independent or predictor variables. In this study, the meteorological variables were not included in the prediction model. The ARIMA model was obtained using the Times Series Forecasting System tool of the SPSS 22.0 software.

The final stage for the modelling process is forecasting, which gives results as three different options that are forecasted values and upper and lower limits that provide a confidence interval of 95%. Any forecasted values within the confidence limit are satisfactory. Finally, the accuracy of the model is checked with the Mean-Square error (MS) to compare fits of different ARIMA models. A lower MS value corresponds to a better fitting model.

In this study we have collected the data from **Central Pollution Control Board(CPCB),** Ministry of Environment & Forests, and Government of India. CPCB has established the **National Ambient Air Quality Monitoring (NAMP)** Network, covering 209 cities/towns of the country in compliance with the mandate under the Air (Prevention and Control of Pollution) Act, 1981 to collect compile and disseminate information on air quality. Air quality monitoring is an important part of the air quality management. The NAMP has been established with objectives to determine the air quality status and trends and to control and regulate pollution from industries and other source to meet the air quality standards. Under this programme they have collected data from 2011 to 2017 of Ahmedabad city in the six different places like AZL Behrampura, Cadila Bridge Narol, L. D. Engineering College, R. C. High School, Shardaben Hospital and GIDC Naroda. The pollutants they measures are Sulphur Dioxide (SO_2), Nitrogen Dioxide (NO_2), Suspended Particulate Matter (SPM) and Respirable Suspended Particulate Matter (RSPM).

3. Results and Discussion

We have four (SO2, NO2, SPM and RSPM) pollutants in this analysis. We analysed it one by one. First we consider the SO2 and then we plot the graph of t (time) vs. Y(t) (Observation). From the graph we need to check the Stationarity. If it is non stationary then we need to make it stationary. Then we have to find the value of p and q. From the graph of ACF and PACF we can assume the value of p and q for the ARIMA model. And then from trial and error we need to find one model that fits the time series and we can forecast the values further from the ARIMA model. This procedure we have done for all four air pollutants.

We can see from the Table 1 that shows the average of different six locations of the air quality data of the year 2011 to 2017 of SO_2, NO_2, RSPM and SPM. From that values we can say that SO_2 level is within the limit specified by CPCB. In fact it was low than the limit. Similarly NO_2 from all the six locations were within the limit and low than the desire limit. Similarly RSPM and SPM were within the limit and low than the desire limit.

Table 1: Average of air quality data year wise with respect to location

Location	Year	So2	No2	RSPM	SPM
Cadila Bridge, Narol	2011	14.62	28.54	147.55	349.32
	2012	11.83	24.36	117.09	269.96
	2013	14.46	25.83	101.52	233.94
	2014	12.52	20.78	82.23	204.72
	2015	18.63	23.06	89.81	200.09
	2016	17.86	21.89	79.56	183.11
	2017	12.95	21.84	75.82	174.57
Naroda	2011	14.80	52.81	152.12	360.07
	2012	13.29	27.02	143.27	328.32
	2013	16.73	30.61	148.78	349.47
	2014	12.98	21.37	127.51	336.68
	2015	17.48	22.88	129.10	295.39
	2016	19.85	25.98	179.50	508.38
	2017	20.69	35.74	105.28	242.17
Shardaben Hospital	2011	11.82	25.04	82.58	197.32
	2012	10.18	21.02	87.54	195.66
	2013	11.97	20.27	85.19	196.37
	2014	12.33	19.00	79.57	201.99
	2015	14.33	20.10	87.27	193.80
	2016	15.64	19.84	83.33	191.00
	2017	11.62	17.48	64.25	147.57
Behrampura	2011	11.54	24.24	82.60	197.66
	2012	9.77	21.15	89.09	202.05
	2013	11.10	19.48	84.26	193.11
	2014	12.31	19.61	82.14	194.67
	2015	16.39	20.83	85.96	192.76
	2016	16.38	19.98	86.44	181.44
	2017	12.07	18.90	62.38	142.82
L. D. Eng.	2011	11.64	22.38	99.79	233.07
	2012	8.54	17.94	69.78	154.99
	2013	8.57	14.52	60.06	137.04
	2014	12.37	18.25	73.02	176.58
	2015	13.44	18.08	82.29	185.44
	2016	11.84	17.33	59.25	134.38
	2017	10.59	14.54	60.06	132.72
R. C. High School	2011	11.80	25.13	85.35	202.75
	2012	9.79	21.33	95.49	217.10
	2013	11.38	19.35	86.69	197.59
	2014	11.85	19.79	81.09	197.84
	2015	14.60	19.75	88.41	196.67
	2016	15.48	20.28	89.70	182.80
	2017	12.71	17.66	67.97	155.18

We run different ARIMA models for air pollutant SO_2 like ARIMA (0,1,2), ARIMA (1,0,3), ARIMA (0,1,3) and ARIMA (1,0,3) from SPSS and then select the ARIMA model (0,1,3) best fitted model. The ARIMA model parameters Table 2 displays values for all of the parameters in the model, with an entry for each estimated model labelled by the model identifier. We already know from the model statistics table that there are zero significant predictors. From the Table 2 we can obtain model to predict the value for forecasting the future values for SO_2.

$$Yt = 0.522Y_{t-1} + 0.097\ Y_{t-2} + 0.096\ Y_{t-3}$$

Table 2: Shows the ARIMA Model (0, 1, 3) for the air pollutant SO_2

					Estimate	SE	t	Sig.
SO_2-Model_1	SO_2	Natural Log		Difference	1	-	-	-
			MA	Lag 1	0.522	0.019	27.984	0.000
				Lag 2	0.097	0.021	4.605	0.000
				Lag 3	0.096	0.019	5.136	0.000

We run different ARIMA models for air pollutant NO_2 like ARIMA (0,1,3), ARIMA (1,0,3) and ARIMA (2,0,3) from SPSS and then select the ARIMA model (0,1,3) best fitted model. The ARIMA model parameters Table 3 displays values for all of the parameters in the model, with an entry for each estimated model labelled by the model identifier. From the table 3 we can obtain model to predict the value for forecasting the future values for NO_2.

$$Yt = 0.495Y_{t-1} + 0.077\ Y_{t-2} + 0.086\ Y_{t-3}$$

Table 3: Shows the ARIMA Model (0,1,3) for the air pollutant NO_2

					Estimate	SE	t	Sig.
NO_2-Model_1	NO_2	Natural Log	Difference		1	-	-	-
			MA	Lag 1	0.495	0.019	26.507	0
				Lag 2	0.077	0.021	3.727	0
				Lag 3	0.086	0.019	4.601	0
	Location	Natural Log	Numerator	Lag 0	0.37	0.103	3.607	0
			Difference		1	-	-	-

Same like SO_2 and NO_2, We run different ARIMA models for air pollutant RSPM, like ARIMA (0,1,3), ARIMA (1,1,3) and ARIMA (2,0,3) from SPSS and then select the ARIMA model (0,1,3) best fitted model. The ARIMA model parameters Table 4 displays values for all of the parameters in the model, with an entry for each estimated model labelled by the model identifier. From the Table 4 we can create an equation to predict the value for forecasting the future values for RSPM.

$$\mathbf{Yt = 0.335Y_{t-1} + 0.200\ Y_{t-2} + 0.145\ Y_{t-3}}$$

Table 4: ARIMA Model (0,1,3) for the air pollutant RSPM

					Estimate	SE	t	Sig.
RSPM-Model_1	RSPM	No Transformation	Difference		1	-	-	-
			MA	Lag 1	0.335	0.019	18.083	0
				Lag 2	0.2	0.019	10.395	0
				Lag 3	0.145	0.019	7.808	0

Same like SO_2, RSPM and NO_2, We run different ARIMA models for air pollutant RSPM, like ARIMA (0,1,3), ARIMA (1,1,3) and ARIMA (2,0,3) from SPSS and then select the ARIMA model (0,1,3) best fitted model. The ARIMA model parameters Table 5 displays values for all of the parameters in the model, with an entry for each estimated model labelled by the model identifier. From the Table 5 we can create an equation to predict the value for forecasting the future values for SPM.

$$\mathbf{Yt = 0.311Y_{t-1} + 0.197\ Y_{t-2} + 0.151\ Y_{t-3}}$$

Table 5: ARIMA Model (0,1,3) for the air pollutant SPM

					Estimate	SE	t	Sig.
SPM-Model_1	SPM	No Transformation	Difference		1	-	-	-
			MA	Lag 1	0.311	0.019	16.792	0
				Lag 2	0.197	0.019	10.335	0
				Lag 3	0.151	0.019	8.137	0

From the above table we can predict the values of four air quality parameters SO_2, NO_2, RSPM and SPM till year 2020. We analyse the data of these four air quality parameters of the year 2011 to 2017. So based on ARMIA models we can predict the values till 2012 and we can say from the table that the given vales for the air quality parameters are well within the standard limits given by CPCB. We can say that for the analysis and prediction for the future of air quality data ARIMA model of time series and forecasting is the best.

4. Conclusion

In this Air quality study we have data of Ahmedabad city during 2011 – 2017 of six different areas. We compared air pollutant parameters with the National Ambient Air Quality Standards (NAAQS Notification dated 18th November 2009). Most of the parameters are within the limits. Air pollutant SO_2 is within the specified limit during the entire period. Time series ARMIA model (0,1,3) is the perfect model for the forecasting. Which shows the value of p = 0, d = 1 and q = 3. Similarly air pollutant NO_2 has almost all the values were within the prescribed limit during the entire period of study which was 2005-2011. From the time series and forecasting we have ARIMA model (0,1,3) which shows the value of p = 0, d = 1 and q = 3 and the perfect model for the time series of NO_2. RSPM is the air pollutant and the value of that parameter is well within the described limit for National Air Quality Programme. We had applied time series and forecasting for the air pollutant RSPM. We got the ARIMA model (0,1,3) for the value of p = 0, d = 1 and q = 3. The all values of air pollutant SPM were almost within the limit. We had applied time series and forecasting for the air pollutant RSPM. We got the ARIMA model (0,1,3) for the value of p = 0, d = 1 and q = 3. Based on the above all equations and ARIMA models we can conclude that ARIMA model (0,1,3) is the perfect model for all four air pollutants for Ahmedabad city during 2011 to 2017.

Table 6: Prediction values for SO_2, NO_2, RSPM and SPM till 2020

Year	SO_2	NO_2	RSPM	SPM
2012	13.55	20.99	74.08	169.17
2013	13.65	21.10	74.20	169.30
2014	13.75	21.21	74.32	169.43
2015	13.85	21.32	74.44	169.56
2016	13.95	21.43	74.56	169.69
2017	14.05	21.44	74.68	169.82
2018	14.15	21.55	74.80	169.95
2019	14.25	21.66	75.00	170.08
2020	14.35	21.77	75.12	170.21

Based on the study we can say that the Ahmedabad mega-city area is expanding rapidly; and human and animal populations, vehicular traffic, industrialization, and per-capita energy consumption are increasing. These developments are increasing atmospheric aerosol concentrations which, in turn, are increasing ambient air pollution substantially. It is certain that these increases will make adverse effects on climate as well as health, and both personal and social wealth of the people living in Ahmedabad and in downwind regions. So, it can be concluded that a strict implementation of adequate abatement measures and environmental regulations is urgently necessary. Finally we can say that from the study time series model used in forecasting is an important tool in monitoring and controlling the air quality condition. It is useful to take quick action before the situations worsen in the long run. In that case, better model performance is crucial to achieve good air quality forecasting. Moreover, the pollutants must in consideration in analysis air pollution data.

5. Acknowledgements

Authors like to thank Gujarat Pollution Control Board (GPCB), Mr. A. Sudhakar and Ms. Smiriti of Centre for pollution Control Board (CPCB), Delhi and Dr. G. G. Shah from NDDB for their support and encouragement.

References

[1] Hertzberg A. M. and Frew L. (2003) Can public policy be influenced? Environmentrics**.** 14: 1-10.
[2] Lee, C. K. (2002) Multiracial characteristics in air Pollutant concentration time series. Water Air Soil Pollution. 135: 389-409.
[3] Voigt K., Welzl G. and Bruggemann R. (2004) Data analysis of environmental air pollutant monitoring. Water Air Soil Pollution. 156: 234-241.
[4] Bencala, K. E. and Seinfeld J. H. (1979) On frequency distribution of air pollutant concentrations. Atmosphere Environment. 10: 941-950.
[5] Yee E., and Chen, R. (1997) A simple model for the probability density functions of concentration fluctuations in atmospheric plumes.Atmosphere Environment. 31:991-1002.
[6] Schwartz J. and Marcus A. (1990) Mortality and air pollution in London: a time series analysis. American Journal of Epidemiology**.** 131: 85-194.
[7] Salcedo R. L. R., AlvimFerraz M., Alves C. and Martins F. (1999) Time series analysis of air pollution data. Atmosphere Environment. 33: 2361-2372.
[8] Kyriakidis P.C. and Journal A. G. (2001) Stochastic modelling of atmospheric pollution: a special time series framework. Part II: application to monitoring monthly sulphate deposition over Europe. Atmosphere Environment. 35: 2339-2348.
[9] Hies T., Treffeisen R., Sebald L. and Reimer E. (2003) Spectral analysis of air pollutants. Part I: elemental carbon time series. Atmospheric Environment. 34: 3495-3502.
[10] Gouveia, N. and Fletcher T. (2000) Time series analysis of air pollution and mortality: Effects by cause, age and socioeconomic status. Journal of Epidemiology Communication Health. 54:750-755.
[11] Roberts S. (2003) Combining data from multiple monitors in air pollution mortality time series studies. Atmosphere Environment. 37: 3317-3322.
[12] Touloumi G., Atkinson R. and Terte A.L. (2004) Analysis of health outcome time series data in epidemiological studies. Environmetrics. 15:101-117.
[13] Saffarini G. and Odat S. (2008) Time series analysis of air pollution in Al-Hashimeya Town Zarqa. Journal of Earth and Environmental Sciences. 1: 63-72.
[14] Ismail M. and Mohd Z. I. (2011) Time series analysis of surface ozone monitoring records in Kemaman, Malaysia. American Journal of Environmental Science. 40: 411-417.
[15] Kumar A. and P. Goyal. (2011) Forecasting of daily air quality index in Delhi. Science of the Total Environment.409: 5517-5523.

Author Profile

Dr. Sweta Patel receivedM.Phil and PhD degrees in Statistics from School of Science, Gujarat University in 2005 & 2015 respectively. From 2007-2017 she was working in Pharmaceutical Education Research & Development(PERD) Centre, Ahmedabad as a Biostatistician. Currently, she is working as a Biostatistician & Business Development at Statiza Statistical Services, Ahmedabad, India

Dr. C. D. Bhavsar received her PhD degree from School of Science, Gujarat University, Ahmedabad. She is currently working as a Reader at Department of Statistics, Gujarat University. Her expertise is Multivariate Techniques and many more.