Estimation of Air Pollutants using Time Series Model at Coalfield Site of India

A Choudhary[1*], P Kumar[2]

[1]Environment, Climate Change and Public Health, Utkal University Bhubaneswar, Odisha, India
[2]School of Environmental Sciences, Jawaharlal Nehru University, New Delhi, India

* Corresponding author's e-mail: choudharyarti12@gmail.com

Abstract. Assessment of air pollutants and quality is an intricate task because of dynamic nature, unpredictability and high inconsistency in space and time. In this study, a time series moving average (MA) model is employed to estimate air pollutants ($PM_{2.5}$, PM_{10}, NO_2, NO_X, O_3, SO_2 and CO) over the coalfield site of India. The estimated O_3 with Adj. R^2 = 0.958 was identified as the most accurate estimation followed by other estimated pollutants. Though, results for the estimated $PM_{2.5}$ (Adj. R^2 = 0.950) and NO_2 (Adj. R^2 = 0.949) were found almost similar to the results of O_3 (Adj. R^2 = 0.958). The estimated CO with Adj. R^2 = 0.887 was identified lower among all the estimated pollutants was also found very well. The existing results of the study demonstrate that MA model permits us to precisely estimate daily basis pollutant concentrations, for the different sites of India.

1. Introduction

Coal mining in India plays a huge part in the monetary development of the country; however, it debases the environment and air quality. The different sources like drilling, blasting, coal handling plants, losses from exposed overburden dumps, and workshops[1], loading and unloading exposed pit faces[2], and road transport[3,4] etc. are the reasons of air pollution over coal mines. These atmospheric pollutants decline air quality and eventually influence nearby environment [5-7]. Most of the studies are mainly focused on SO_2, NOx, HC, CO, CO_2 and particulate matter (SPM, RSPM, etc.) and almost for short duration[8,9]. Due to the significant importance of $PM_{2.5}$, PM_{10}, NO_2, NO_X, O_3, SO_2 and CO in air pollution, we present daily MA time series model estimation over the Talcher coal mine site of India. Current increase of India's economy, transportation and industry with the improvement of urbanization, environmental pollution issues have slowly become noticeable, but this is contrary to people's vision of seeking a high-quality life[10]. Therefore, aerosol studies over mining regions are critical predominantly according to the climatic perspective. In the present study, the moving average (MA) modeling approach has been adapted to estimate the different pollutants over coal mining site of India. Air quality information takes into account the presentation of different examinations, most normal is measurable ones, to discover general examples and conditions for various time spans and connections between noticed air attributes. The traditional methods like PCA (Principle Component Analysis) and factor investigation are significant measurable apparatuses regularly utilized in ecological sciences revealed in a few exploration articles. Distinctive parametric strategies generally utilized for times series examination and guaging, moving normal (MA) models is one of them, referred to likewise as Box–Jenkins stochastic models [11].

The principle benefits of the Box-Jenkins approach are [11,12]: (I) Its materialness for demonstrating and anticipating for all intents and purposes any time series, which is fixed or can be diminished to fixed through a differencing method; (ii) The ability to extract all the trends and serial correlations in the data with a minimized sequence of white noise (shock) through inclusion in one general model equation that gets to the basis of historical data development; (iii) The strategy has been fused into numerous standard programming bundles like SPSS, Statistica, R and numerous others (Comparison of Statistical Packages 2013), which accelerates and works with the demonstrating system essentially. Detriments of this strategy incorporates that Box–Jenkins models are exact, a recognizable proof assessment finding system should consistently be done. Likewise, for time series investigation, essentially extra 50–100 perceptions are required [11-12]. On a basic level, the cycles of air contamination in the air are unequivocally administered by meteorology (e.g., see Jacobson 2005). Nonetheless, in supposed univariate models, it is expected that the last convergence of air toxins in the climate is the endproduct of all the mind boggling exchange of meteorology, science, transport, dissemination and so on. Accordingly, the consolidated data of their impact on air toxin fixation is contained in the relating time series in a stochastic way. With this methodology estimations are rearranged and performed just utilizing the time series of the contamination without unequivocal consideration of meteorological or different estimations.

2. Methodology

The various pollutants ($PM_{2.5}$, PM_{10}, NO_2, NO_X, O_3, SO_2 and CO) data were collected from CPCB website during (25 December 2016–25 May 2021) for the Talcher site of Odisha. Sample pretreatment applied for example data set should be tested for stationary before modeling. In this respect, MA time series model was applied to estimate daily basis air pollutants over Talcher coalmine region of India. MA model is a widespread methodology for the modeling of univariate time series data. The mathematical form of MA (1) model of order 1 is portrayed as following:

$$y_t=c+\varepsilon_t+\theta_1\varepsilon_{t-1}$$

Where, c is the mean of the series, θ_1 is the parameters of the model, the ε_t, ε_{t-1} are white noise error terms. For MA (1) model: $-1<\theta_1<1$. The auto correlation function with a significant autocorrelation at lag 1 is an indicator of MA (1) model.

Sample pretreatment applied for example data set should be tested for stationary before modeling. The selection of the models was performed based upon combined criteria: (1) Minimum BIC; (2) Minimum RMSE; (3) Maximum R^2; (4) Minimum significance. After the order determination and parameter estimation, the applicability of the model established should be tested. In the event that the model mistake is background noise, acquired model is qualified if not; the request re-assurance and boundary re-assessment are required. Model outputs are compared with testing data using a statistical measure of errors such as root mean square error (RMSE), Adj. R^2, t-value, and standard error (SE) are defined by equation 1 to 4.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left[f(t)-\hat{f}(t)\right]^2} \quad (1)$$

$$R^2_{adj} = 1-\left[\frac{(1-R^2)\,(n-1)}{n-k-1}\right] \quad (2)$$

$$t-value = \frac{mean1-mean2}{\frac{(n1-1)x\,var1^2+(n2-1)xvar2^2}{n1+n2-2}\;x\sqrt{\frac{1}{n1}+\frac{1}{n2}}} \quad (3)$$

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} \quad (4)$$

where n is the number of data points, $f(t)$ is the tth measurement, and $\hat{f}(t)$ is its corresponding prediction; k is no of independent regression excluding constant; mean 1 and mean 2 are the average values of each of the sample sets; var1 and var2 are the variance of the sample sets; n1 and n2 are the records in each sample set; $\overline{x}$ is mean value of the sample and σ is standard deviation.

3. Results and discussion

$PM_{2:5}$, also known as fine particles, and PM_{10}; O_3, SO_2, NO_2, NO_X and CO are the other main pollutants that affect the quality of the atmospheric environment. In addition, $PM_{2:5}$ in the air can cause great health risks to the human body and even affect the climate [13]. In this investigation MA model was used for the estimation of these important air pollutants. Overall the estimated O_3 was found as Adj. R^2 = 0.958 with RMSE = 2.799, the most accurate estimation. Though, results for the estimated $PM_{2.5}$ (Adj. R^2 = 0.950; RMSE = 6.903) and NO_2 (Adj. R^2 = 0.949; RMSE = 2.246) were found almost similar to the results of O_3 (Adj. R^2 = 0.958; RMSE = 2.799). The scattered plots for the estimation of $PM_{2.5}$, PM_{10}, NO_2, and NO_X using MA model are shown in the Figure 1. The major concerns of pollution during summer are coal-fired heating. In winter, especially in the residential regions, heating, coal consumption increases significantly, and the local pollutant emission concentration increases. Some of the biomass burning in winter, which is also one of the ‘‘culprits’’ that can cause regional pollution in winter [14]. Increased exhaust emissions from traffic vehicles and unfavourable meteorological conditions were also found an important cause of heavy pollution. The detailed statistical analysis about the various estimated pollutants are described in Table 1. The scattered plots for the estimation of SO_2, O_3 and CO using MA model are shown in the Figure 2.

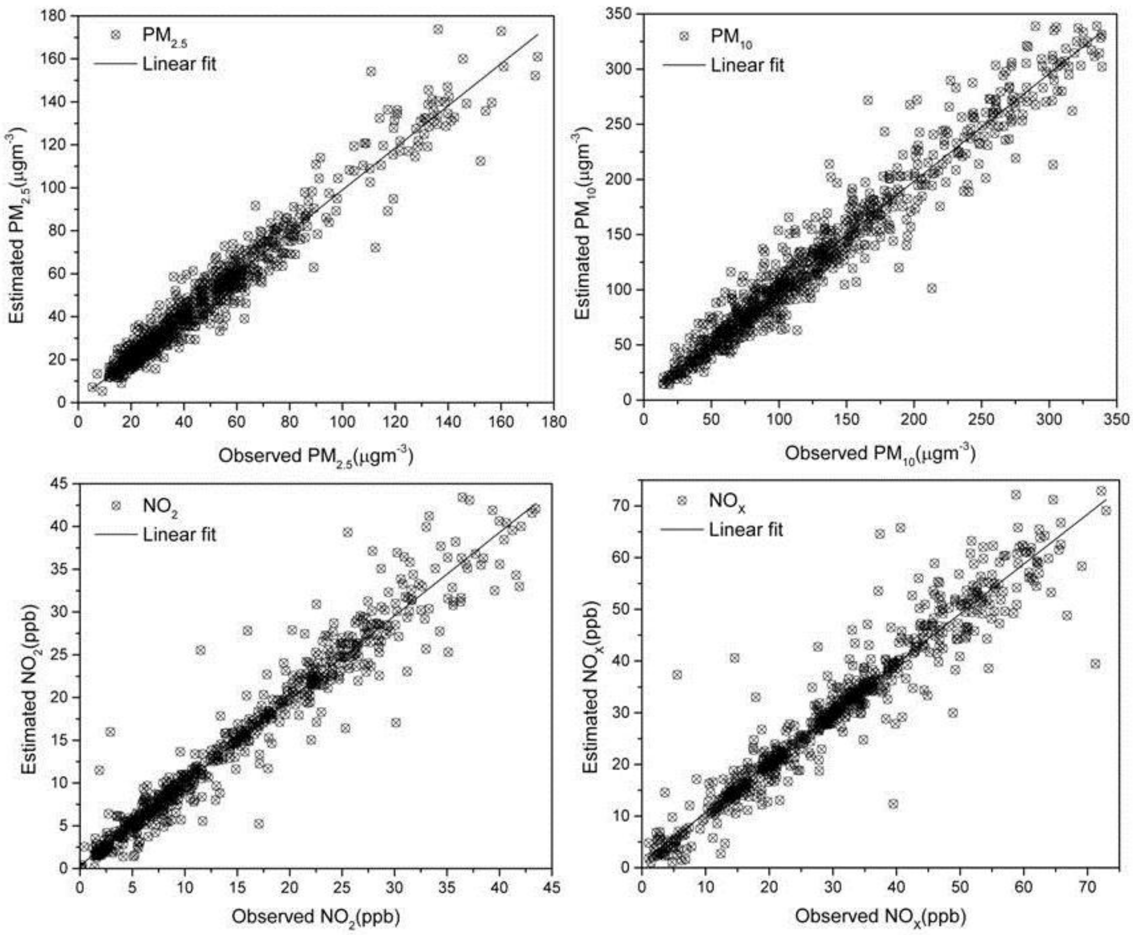

Figure 1: Scattered plots for the estimation of $PM_{2.5}$, PM_{10}, NO_2, and NO_X using MA model.

Table 1: The detailed statistical analysis about the various estimated pollutants.

SN	Pollutants	Parameters								Statistics		
		Intercept				Slope				No of points	Adj.R^2	RMSE
		Value	SE	t-value	p-value	Value	SE	t-value	p-value			
1	$PM_{2.5}$	1.097	0.448	2.451	0.014	0.979	0.008	123.441	0.000	797	0.950	6.903
2	PM_{10}	3.611	1.271	2.842	0.005	0.974	0.008	114.681	0.000	797	0.943	18.492
3	NO_2	0.334	0.140	2.385	0.017	0.974	0.008	119.654	0.000	756	0.949	2.246
4	NO_X	1.032	0.324	3.188	0.001	0.963	0.009	97.514	0.000	756	0.926	4.299
5	CO	0.090	0.020	4.431	1.062E-5	0.946	0.012	80.993	0.000	835	0.887	0.193
6	O_3	0.383	0.169	2.266	0.024	0.979	0.007	137.194	0.000	820	0.958	2.799
7	SO_2	0.980	0.248	3.948	8.571E-5	0.959	0.009	98.183	0.000	793	0.924	2.674

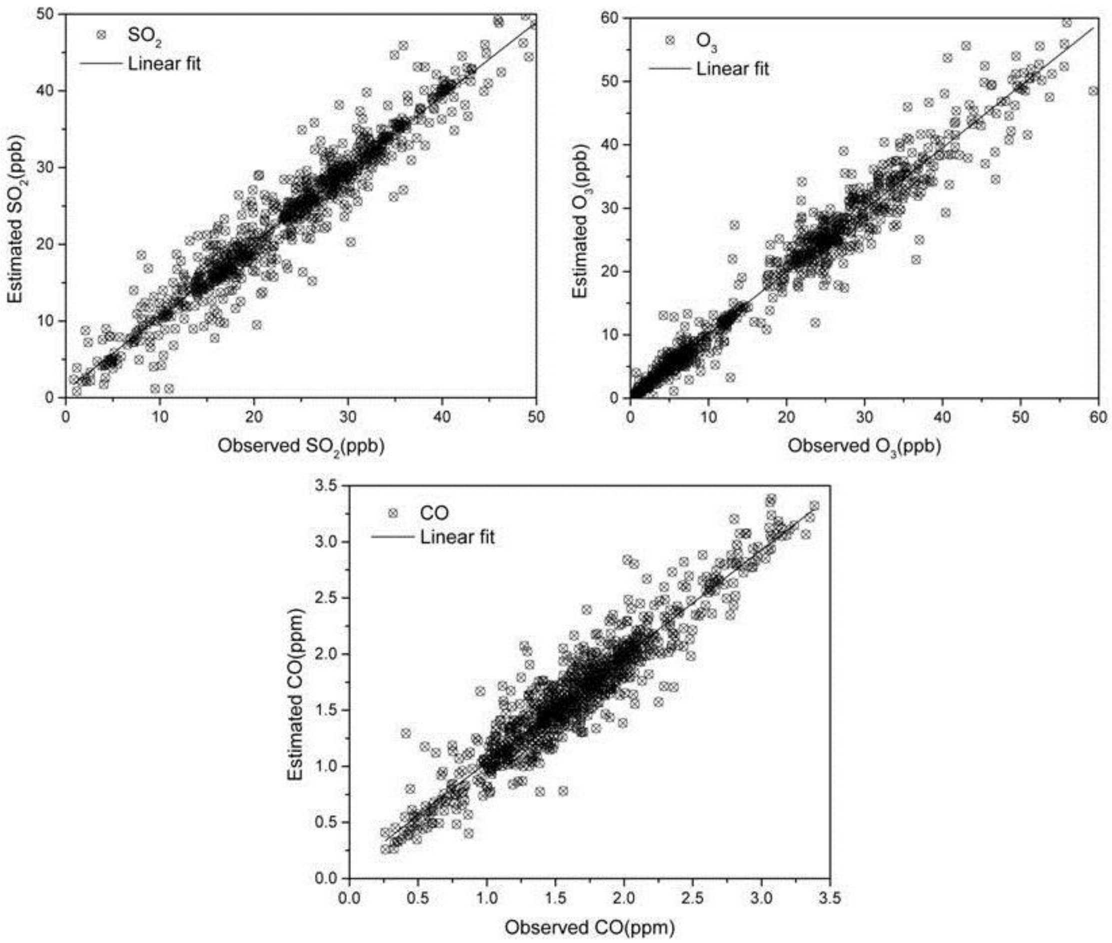

Figure 2: Scattered plots for the estimation of SO2, O3, and CO using MA model.

The higher concentrations of aerosol optical depth (AOD) were experiential during the monsoon season while lowest seasonal AOD values during post-monsoon seasons over the coal mining spot [14]. The estimated CO (Adj. R^2 = 0.887; RMSE = 0.193) was found lower among all the estimated pollutants was also found very well. The estimated SO_2 (Adj. R^2 = 0.924; RMSE = 2.674) also indicates the better estimation followed by other estimated pollutants. The confidence intervals at 95% level were calculated to test the statistical significance using t-test and p-value (<0.05). At long last, an estimated pollutants was determined which showed that the model assessment was quite well over the Talcher region of India. Almost all the estimated results were found statistically significant at $p<0.05$.

4. Conclusion

In the present study, air pollutants estimation becomes a reliable tool to reduce the negative impact of environmental pollution on health and to formulate more complete prevention policies. The MA time series model was found appropriate for the air pollution assessment with high upsides of Adj. R2 and lower RMSE esteems. The time series air pollutants estimation results using MA model are identified significant to extrapolate the current time series for future prediction of pollutants loading over the coal mining areas in India.

Acknowledgement

Arti Choudhary acknowledge to Rashtriya Uchchatar Shiksha Abhiyan (RUSA) PDF scheme, order no. RUSA-1041-2016(PDF-XVIII) 25986/2020, for the financial support under RUSA 2.0 project. PK acknowledge to UGC, for Dr. DSK-PDF awarded vide UGC sanction No.F.4-2/2006 (BSR)/ES/18-19/0041.

References

[1] CMRI 1998 Determination of emission factor for various opencast mining activities report GAP/9/EMG/MOEF/97 Central Mining Research Institute Environmental Management Group, Dhanbad.

[2] Chaulya SK 2005 Air quality status of an open pit mining area in India *Environ Monit Assess* **105** 369-389.

[3] Choudhary A and Gokhale S 2016 Urban real-world driving traffic emissions during interruption and congestion *Transp. Res. D. Transp. Environ.* **43** 59-70.

[4] Choudhary A and Gokhale S 2019 On-road measurements and modelling of vehicular emissions during traffic interruption and congestion events in an urban traffic corridor *Atmos. Pollut. Res.* **10** 480-492.

[5] Crabbe H, Beaumont R, and Norton D 2000 Assessment of air quality, emissions and management in a local urban environment *Environ. Monit. Assess.* **65** 435-442.

[6] Kumar P, Pratap V, Kumar A, Choudhary A, Prasad R, Shukla A, ... and Singh AK 2020 Assessment of atmospheric aerosols over Varanasi: physical, optical and chemical properties and meteorological implications *J. Atmos. Sol.-Terr. Phys.* **209** 105424.

[7] Kumar P, Kapur S, Choudhary A, and Singh AK 2021 Spatiotemporal variability of optical properties of aerosols over the Indo-Gangetic Plain during 2011–2015 *Indian J. Phys.* 1-13.

[8] Chaulya SK 2004 Spatial and temporal variations of SPM, RPM, SO_2 and NO_x concentrations in an opencast coal mining area *J. Environ. Monit.* **6** 134-142.

[9] Choudhary A, Kumar P, Gaur M, Prabhu V, Shukla A, and Gokhale S 2020 Real world driving dynamics characterization and identification of emission rate magnifying factors for auto-rickshaw *Nat. Environ. Pollut. Technol.* **19** 93-101.

[10] Choudhary A and Gokhale S 2019 Evaluation of emission reduction benefits of traffic flow management and technology upgrade in a congested urban traffic corridor *Clean Technol. Environ. Policy* **21** 257-273.

[11] Box GEP, and Jenkins GM 1976 Time Series Analysis; Forecasting and Control, Holden-Day Inc. U.S.A.

[12] Pankratz A 1983 Forecasting with Univariate Box-Jenkins Models: Concepts and Cases. John Wiley, New York.

[13] Schwartz J 1996 Air pollution and hospital admissions for respiratory disease *Epidemiol.* **7** 20–28.

[14] Soni K, Parmar KS and Kapoor S 2015 Time series model prediction and trend variability of aerosol optical depth over coal mines in India *Environ. Sci. Pollut. Res.* **22** 3652-3671.