Global J. Environ. Sci. Manage.,4(4): 427-438, Autumn 2018
DOI: 10.22034/gjesm.2018.04.004

ORIGINAL RESEARCH PAPER

Statistical trend analysis and forecast modeling of air pollutants

A. Jaiswal[1,*], *C. Samuel*[1], *V.M. Kadabgaon*[2]

[1]*Department of Mechanical Engineering, Indian Institute of Technology, Varanasi, India*

[2]*Department of Mechanical Engineering, Brahmdevdata Mane Institute of Technology, Solapur, India*

Received 24 April 2018; revised 30 August 2018; accepted 30 September 2018; available online 1 October 2018

ABSTRACT: The study provides a statistical trend analysis of different air pollutants using Mann-Kendall and Sen's slope estimator approach on past pollutants statistics from air quality index station of Varanasi, India. Further, using autoregressive integrated moving average model, future values of air pollutant levels are predicted. Carbon monoxide, nitrogen dioxide sulphur dioxide, particulate matter particles as PM2.5 and PM10 are the pollutants on which the study focuses. Mann-Kendall and Sen's slope estimator tests are used on summer (February-May), monsoon (June-September) and winter (October-January) seasonal data from year 2013 to 2016 and trend results and power of the slopes are estimated. For predictive analysis, different autoregressive integrated moving average models are compared with goodness of fit statistics, and the observed results stated autoregressive integrated moving average (1,1,1) as the best-suited for forecast modeling of different pollutants in Varanasi. Autoregressive integrated moving average model (1,1,1) is also used on the annual concentration levels to predict forthcoming year's annual pollutants value. Study reveals that PM 10 shows a rising trend with predicted approximate annual concentration of 273 $\mu g/m^3$ and PM2.5, carbon monoxide, nitrogen dioxide and sulphur dioxide show a reducing trend with approximate annual concentration of 139 $\mu g/m^3$, 1.37 mg/m^3, 38 $\mu g/m^3$ and 17 $\mu g/m^3$, respectively, by the year 2030. The study predicted carbon monoxide, nitrogen dioxide and sulphur dioxide concentrations are lower and PM10 and PM2.5 concentrations are much higher to the standard permissible limits in future years also, and specific measures are required to control emissions of these pollutants in Varanasi.

KEYWORDS: *Air pollutants; Autoregressive integrated moving average (ARIMA); Forecast; Mann-Kendall; Sen's slope estimator.*

INTRODUCTION

Air pollution accounts for a number of ecological and health issues globally. Severe effects of air pollution on environmental degradation and health conditions has been observed in past years throughout the world (Bernard *et al.,* 2001; Cohen *et al.*, 2005; Emberson *et al.,* 2003; Kampa and Castanas, 2008; Pascal *et al.,* 2013). A foremost concern for air pollution is consistency in its rising rate over the previous years. For the demographic regions nearby urban cities, emissions of different pollutants and particulate matter (PM) particles in the environment has seen a rapid growth due to industrial development, high use of motorized vehicles and higher population density (Gulia *et al.,* 2015; Mayer, 1999; Rodríguez *et al.,* 2016). Due to vast population residing in the urban zones of countries, continuous measurement of pollutants, trend analysis of pollution in such areas and forecast of pollutants becomes necessary so that proper strategies can be formulated for pollution control in the regions for minimizing environmental and health effect. Because of it, most of the countries in the recent time are continuously monitoring air quality index (AQI) based on the pollution data and metrological parameters (Chaudhuri and Dutta, 2014). Monitoring for urban AQI mostly comprises tracking of pollutants like carbon monoxide (CO), particulate

*Corresponding Author Email: *ajaiswal.rs.mec13@itbhu.ac.in*
Tel.: +91 8172 873333 Fax: +91 5422 360657

Note: Discussion period for this manuscript open until January 1, 2019 on GJESM website at the "Show Article".

matters (PM), Ozone (O_3), lead (Pb) particles, Sulphur dioxide (SO_2), nitrogen dioxide (NO_2) and ammonia (NH_3) (Azmi *et al.,* 2010; Gurjar *et al.,* 2008). With available pollutants data under AQI, trend analysis and forecasting are possible through various statistical modeling techniques. Statistical modeling techniques do not depend upon traditional ways of environment predictions (Zuma-Netshiukhwi *et al.,* 2013) and chemical formulations of pollutants value in the ambient environment but use the past pollutants data for estimation of trends and forecasting of future pollutants value. The proposed study uses the statistical methods for prediction of trends and forecasting of pollutants in the ambient environment. Statistical methodologies are common in researches for predictions of trends in environmental data. Can (2017) used graphical and statistical approaches for time-series analysis of air pollutants, Rani *et al.* (2018) used past air pollution index (API) data for trend analysis using XLSAT, Jaruskova and Liska (2011) used median regression and Spearman correlation - coefficient for analysing trends in pollution due to nutrients and organic pollution, Pandolfi *et al.* (2016) used Mann- Kendall test and a multi-exponential fit centred methodology for trend prediction of particulate matter particles, Dai and Zhou (2017) used PMFG network method for assessing the spatial and temporal correlation patterns of different pollutants in air, Kumar *et al.* (2018) investigated Sulphur dioxide and Nitrogen dioxide levels in environment using dispersion modeling methodology. Through most of the statistical methodologies applied by the researchers, the most common aspect is to analyze and correlate the trends present in the pollutants data and other environmental parameters. Trend estimation highly depends upon the characteristics of data and thus are considered as a complex approach (Kisi and Ay, 2013). In the proposed study, non-parametric tests are applied for statistical analysis. Parametric approaches are considered more precise than non-parametric tests but come with a limitation of normally distributed independent data whereas non-parametric tests have no such constraints (Watthanacheewakul, 2011). The proposed study uses the non-parametric Mann-Kendall (M-K) test in addition with Sen's - slope estimator approach for trend estimations of different pollutants and autoregressive integrated moving average (ARIMA) approach for modeling the pollutants forecast. M-K and Sen's - slope estimator tests are well-established tests for estimating the rising or reducing trends for the non-parametric data (Da Silva *et al.,* 2015; Drápela and Drápelová, 2011; Gocic and Trajkovic, 2013). ARIMA modeling is a generalized approach in which the models are fit on the time-series data to predict the future values (Brocklebank *et al.,* 2018; Eymen and Köylü, 2018). The proposed study first used M-K test along with Sen's - slope estimator tests to assess trend existence in the pollutants time-series data and afterward ARIMA modeling is done to forecast the pollutants value with precision. The study objective is to examine trends and to forecast different pollutants concentrations using the data of an AQI urban station in Varanasi, India. Different pollutants considered in the study of statistical analysis are PM 2.5, and PM 10, CO, NO_2, and SO_2. M-K test and Sen's - slope estimator methodologies are used on a seasonal scale to estimate trends during summer (February - May), monsoon (June - September) and winter (October - January) seasons and ARIMA modeling is done for the yearly forecast of the pollutants considered. This study has been carried out for the district Varanasi, India in 2018 and considers the past data of year 2013 to 2016 for statistical trend assessment of air pollutants.

MATERIALS AND METHODS

Study area and air quality data

This study is based on pollutant data of Varanasi district in Uttar Pradesh (UP) state, India. Varanasi is a major city in the UP eastern region specifically important with its tourism. In the survey of the year 2015-16, the air quality of Varanasi is considered to be the most toxic in the whole of India, and according to the 2015 data of CPCB, Varanasi does not record a single day of 'good' air quality (Safi, 2016). The past pollutants data for the study is retrieved from Central Pollution Control Board (CPCB) website for the AQI station: Ardhali Bazar, Varanasi – UPPCB with Latitude: 25.3505986, and Longitude: 82.9083074. The study area and the sampling station is shown in Fig. 1. The air pollutants which are selected in this study are the PM particles PM 10 and PM 2.5, CO, NO_2, and SO_2. The pollutants data retrieved for the study is from January 2013 to December 2016. Due to the unavailability of data for the previous year to 2013 and year 2017, the study is confined to a limited year of data set. The mean of the hourly average values

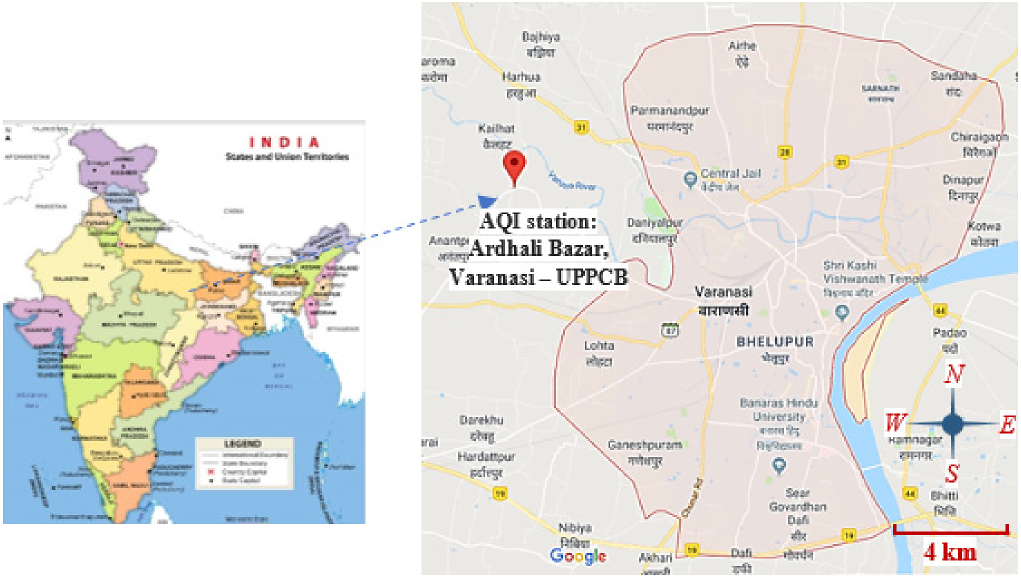

Fig. 1: Geographic location and map of air quality index station for the study area

of the pollutants for a day (24 h) is considered as the unit time in the time-series data. Time-series data of the pollutant for a year are categorized in three seasons of Varanasi with summer as February to May, monsoon as June to September and winter as October to January. Few of the data points with very absurd values are considered as outliers and are eliminated in the estimations to prevent the adverse effects of the outliers on the results (Hirsch *et al.,* 1991). Missing data and the eliminated outliers are interpolated with the nearest neighbor method using S-PLUS FinMetrics (Azmi *et al.,* 2010; Junninen *et al.,* 2004. Excel-XLSAT version 2018.5 is used as statistical software for M-K Test, Sen's slope estimator, and ARIMA modeling.

Mann – Kendall test

M-K test is a widely accepted statistical tool to predict and analyze different pollutants statistics, retrieved over time for steadily decreasing or increasing trends (Chaudhuri and Dutta, 2014; Emami *et al.,* 2018; Ma *et al.,* 2011; Vanguelov *et al.,* 2010). Introduced by Mann (1945) and reworked by Kendell (1975), the null hypothesis of this non-parametric test defines no monotonic trend in data, and alternate hypothesis states an existence of a positive, negative or non-null trend. The principal statistics value S of M-K test is calculated as Eqs. 1 and 2.

$$S = \sum_{a=1}^{z-1} \sum_{b=a+1}^{z} sgn\,(x_b - x_a) \tag{1}$$

$$sgn(x_b - x_a) = \begin{cases} 1 & if\ x_b - x_a\ is\ greater\ than\ 0 \\ 0 & if\ x_b - x_a\ is\ equal\ to\ 0 \\ 1 & if\ x_b - x_a\ is\ less\ than\ 0 \end{cases} \tag{2}$$

Where, length of data points in time-series (x_1, x_2, x_3.... x_n) is defined by z. x_a and x_b are individual pollutant data points with b is greater than a. Null hypothesis H_0 states no existence of any trend in pollutant data and all individual pollutants value of each day over the year are not in trend. Alternate hypothesis H_1 defines existence of a monotonic trend in the pollutant data points. Mean M[S] and variance Varian[S] of the principal statistics value S are calculated as Eqs. 3 and 4.

$$M[S]=0 \tag{3}$$

$$\mathrm{Varian}[S] = \frac{z(z-1)(2z+5) - \sum_{i=1}^{j} t_i(t_i - 1)(2t_i + 5)}{18} \tag{4}$$

Where j represents the tied groups number and t_i represents the data value count in i^{th} group. Standard normal test statistics W[S] is calculated using the Eq. 5

for estimating the presence of a statistically significant trend (Chattopadhyay *et al.,* 2012; Chaudhuri and Dutta, 2014; Da Silva *et al.,* 2015).

$$W[S] = \begin{cases} \frac{S-1}{\sqrt{Varian[S]}} & if\ S\ is\ greater\ than\ 0 \\ 0 & if\ S\ is\ equal\ to\ 0 \\ \frac{S+1}{\sqrt{Varian[S]}} & if\ S\ is\ less\ than\ 0 \end{cases} \tag{5}$$

The negative or positive W[S] value ensures a decreasing or increasing trend respectively. For a significance level α in a two-tailed test, H_0 is rejected, confirming the existence of a trend in the time-series data, when W[S] is greater than $W_{1-\alpha/2}$. The values of $W_{1-\alpha/2}$ for different significance level α can be obtained from standard normal distribution table (Chaudhuri and Dutta, 2014; Da Silva *et al.,* 2015). The Kendall's τ values is calculated as Eq. 6.

$$\tau = 2\frac{S^*}{z(z-1)} \tag{6}$$

in which S^* denotes the Kendall's sum, computed as $S^* = A - B$ where A represents number of chances when difference of x_b to x_a is greater than zero and B represents number of chances when difference of x_b to x_a is less than zero (Chattopadhyay *et al.,* 2012; Xu *et al.,* 2004).

Sen's - slope estimator test

This test, also termed as Theil–Sen slope test, is a widely used statistical tool for non-parametric data to estimate the power of trend, detected through the M-K test (Caloiero *et al.,* 2017; Eymen and Köylü, 2018). Developed by Theil, 1950 and Sen, 1968, this is a median-based tool which evaluates the slope of the trend through a linear model. If there are m number of pollutant data points in a time-series (X_1, X_2, X_3…… X_m) and X_a and X_b are the pollutant values at time instance a and b such that b > a, then variance of the residual is computed as Eqs. 7 and 9.

$$T_i = \frac{X_b - X_a}{b-a}\ for\ i = 1,2,3 \dots m \tag{7}$$

Median of all T_i values, denoted as T_{med}, is the Sen's slope estimator and is calculated as equation 8. The sign of T_{med} reveals the upward or downward trend of the data and its numeral denote the trend steepness.

$$T_{med} = \begin{cases} T_{\frac{m+1}{2}} & if\ m = odd \\ \frac{T_{\frac{m/2}{2}} + T_{\frac{(m+2)/2}{2}}}{2} & if\ m = even \end{cases} \tag{8}$$

The trend prediction of the pollutants through M-K test depends upon the significance level α, and there is a possibility of the existence of trends with other significant levels. So through the Sen's slope estimator, the changing rates can be assessed for the pollutants which shows no trend in M-K Test.

Autoregressive integrated moving average (ARIMA)

The ARIMA model, developed for prediction and estimation of future values in univariate time-series data, was introduced by Box and Jenkins (1976). ARIMA includes a combination of several time-series techniques to give a better representation and analysis of time-series data. Auto regression (AR), differencing order integration (I) and moving average (MA) collectively makes ARIMA(p,d,q) model in which p is the order of auto regression model, d is for differencing order integration, and q is the moving average model order. In the first step of modeling methodology, time-series data are checked whether it is stationary or not. Dickey-Fuller (D-F) test is used in the paper to check the data (Dickey and Fuller, 1979). If the data is stationary, the model moves in the second step else the data is made stationary by difference. In the next step, p, d, q possible values are estimated using correlogram of autocorrelation and partial autocorrelation functions (ACF and PACF). In next stage, for determining the adequacy of the model, the values of Akaike information criteria (AIC), and other error estimation measures are assessed over the best-suited goodness of fit statistics to select appropriate ARIMA model order. For the idea of order determination of the ARIMA model in the provided study, various goodness of fit statistics criteria observed which, other than AIC, includes sum of squared errors (SST), root mean squared deviation (RMSD), W-N Variance, mean absolute percentage deviation (MAPD) and final prediction error (FPE). With the chosen model, the last step involves estimation of forecasted values for the provided time-series data. A generalized expression of ARIMA (p,d,q) can be given as Eq. 9.

$$\phi(\beta)\nabla^d f_t = \theta(\beta)e_t \tag{9}$$

Where, $\phi(\beta)$and $\Theta(\beta)$ represent the polynomial of degree p and q respectively, β is a backward-shift operator, ∇ is difference operator, f_t is pollutants parameter at time t and e_t is the error term at time t.

RESULTS AND DISCUSSIONS

In this section of the study, results estimation and analysis of its inferences are carried out for the Mann-Kendall test, Sen's slope estimator test and, ARIMA modeling of time-series pollutants data of AQI sampling station of district Varanasi. Table 1 displays the results of the Mann Kendall test for different pollutants data retrieved from AQI station of Varanasi and Fig. 2 shows the seasonal trend graph of different pollutants from the year 2013 to 2016. The horizontal axis of the seasonal trend graphs in Fig. 2 denotes the 24 h average of the months of the particular season for the year. The winter period includes the months of October, November, December, and January, so the trend graph of PM 10 of winter period only shows the three months values up to December 2016.

For the retrieved data of carbon monoxide in the summer season, the p-value is below the significance level, with a negative value of Kendall's tau, so the null hypothesis is rejected confirming the alternate hypothesis of acceptance of trend in the time-series data. Similarly, for monsoon also, tau value of CO is -0.700 (negative) and the p-value is 0.00018 which is below 0.05 ensuring the existence of a negative trend. For the winter period, as p-value is 0.260, that is more to 0.05, H_0 is established, and no trend exists for CO over provided years. For the pollutant NO_2, it can be observed from the results of Table 1 that for all the three periods, summer, monsoon, and winter, the corresponding p-values are 0.392, 0.753 and 0.964 respectively, which is more to significance value 0.05, H_0 is accepted, confirming no trend in data. For pollutant SO_2, only in monsoon, a trend can be observed with a negative orientation. In the other two seasons, no trend is estimated due to higher p values of 0.392 and 0.499. In case of particulate matter particles of size up to 10 µm (PM 10), H_0 is rejected for all three seasons, confirming the existence of a trend in the pollutant data. Kendall's tau τ positive value for summer and winter season indicates a rising trend in data and the negative τ value for monsoon indicates a decreasing trend. For PM 2.5, only monsoon season shows a trend with a p-value of 0.010 with a decreasing ratio confirmed by the negative τ value. For summer and winter, pollutant data of PM 2.5 shows no trend. The values of Sen's slope estimator T_{med} is also provided in Table 1, and outcomes of Sen's slope test validates the M-K test results. T_{med} value is also calculated for those pollutants seasonal data in which no trend exists. This is for the reason that the hypothesis in M-K test is established over a significance level α and there is a possibility of the presence of trend, and thus the trend slope possibility, beyond α. In the proposed study, α is kept

Table 1: Results of M-K and Sen's slope estimator test on different pollutants data of Varanasi over a seasonal scale from the year 2013 to 2016

Pollutants	Seasonality	Kendall's tau value τ	p-value	s-value	Trend Estimation	Sen's slope T_{med}
CO ($\frac{mg}{m^3}$)	Summer	-0.383	0.043	-46.00	Trend	-0.031
	Monsoon	-0.700	<0.0001	-84.00	Trend	-0.021
	Winter	-0.217	0.260	-26.00	No trend	-0.025
NO_2 ($\frac{\mu g}{m^3}$)	Summer	0.167	0.392	20.00	No trend	0.556
	Monsoon	-0.067	0.753	-8.00	No trend	-0.107
	Winter	0.017	0.964	2.00	No trend	0.077
PM 2.5 ($\frac{\mu g}{m^3}$)	Summer	-0.233	0.224	-28.00	No trend	-2.737
	Monsoon	-0.483	0.010	-58.00	Trend	-7.802
	Winter	0.167	0.392	20.00	No trend	4.699
PM 10 ($\frac{\mu g}{m^3}$)	Summer	0.467	0.013	56.00	Trend	9.07
	Monsoon	-0.500	0.008	-60.00	Trend	-6.656
	Winter	0.417	0.027	50.00	Trend	10.05
SO_2 ($\frac{\mu g}{m^3}$)	Summer	-0.167	0.392	-20.00	No trend	-0.259
	Monsoon	-0.733	<0.0001	-88.00	Trend	-0.765
	Winter	-0.133	0.499	-16.00	No Trend	-0.107

at 5 % for the calculation of results. The Sen's slope results presented in Table 1 confirms the results of M-K Test and shows the similar slope orientations. The T_{med} value of CO for all the three seasons shows a negative slope for the trends (-0.031, -0.21 and -0.025). M-K Test for the NO_2 data shows no trend, and the Sen's slope estimator values predicted a positive slope with value 0.556 and 0.077 for summer and winter seasons and negative slope with value 0.107 for the monsoon season. The SO_2 slope estimator values for all the three months shows a negative slope in the provided years. Winter season of the presented years for PM 10 and PM2.5 shows a positive slope confirming an increasing trend in the data of both the pollutants. The summer season for PM10 observed a positive slope in the trend and a negative slope trend for the monsoon season. It can be observed from the results of Table 1, that almost all the pollutants show a decreasing slope of a trend in monsoon season from the year 2013 to 2016. From the report of rainfall statistics of India for the year 2013 to 2016, India Meteorological Department, Ministry of Earth Science, it can be observed an increasing trend in the observed rainfall in district Varanasi in monsoon season with detected rainfall as 772.6 mm, 683.4 mm,

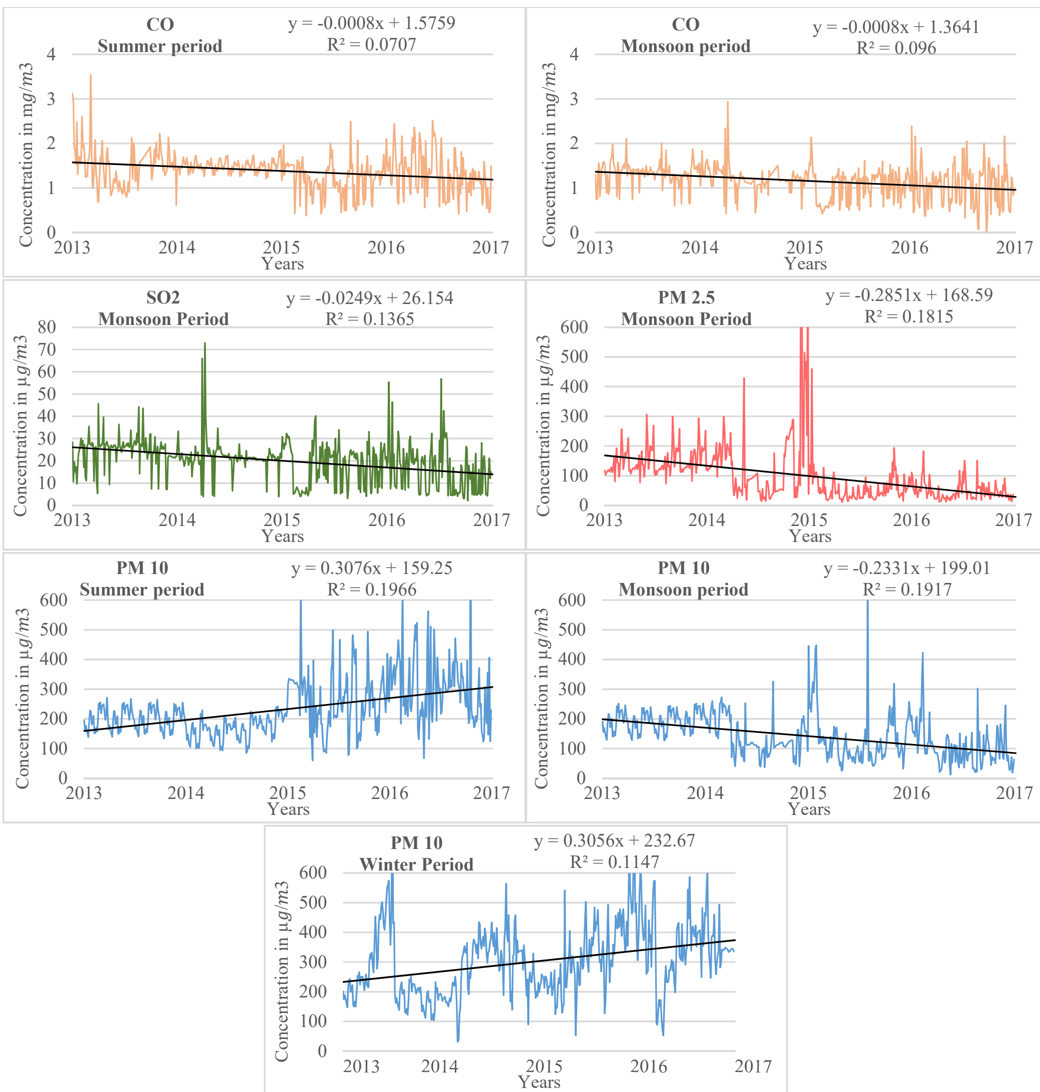

Fig. 2: Seasonal trend graph of different pollutants for the period from year 2013 to 2016

722.6 mm and 1145.4 mm (Rainfall statistics of India, 2016). The result in Table 1 confirms the negative correlation between the rainfall and humidity with pollutants level (Aleksandropoulou and Lazaridis, 2004; Jayamurugan *et al.,* 2013; Shukla *et al.,* 2008), as with increasing trend of monsoon for year 2013 to 2016 a negative trend in all the pollutants has been observed. Results of Fig. 2 show that the seasonal trend lines of different pollutants for the year 2013 to 2016 have a different slope than that of the Sen's slope estimator test provided in Table 1. This is because the trend shown in Fig. 2 refers to the linear change in the mean concentrations of the pollutants value while Sen's slope line models the linear change of median concentrations of the pollutants value. The coefficient of determinations or R squared value in most of the trend plots of Fig. 2 is low because of the scattered data points as the individual data point's values represent the 24 h average of the pollutants values. With the estimated outcomes of M-K test and Sen's slope estimator test, ARIMA time-series model is fitted on the available pollutant data of 24 h mean. Using the D-F test, the data is checked with null hypothesis H_0 which shows the existence of unit root in time-series data and alternative hypothesis H_1 which shows no unit root confirming a stationary time-series data. Computed p-value in D-F test for the time-series data of each of the pollutants in different season comes lower to a significance level of 0.05, due to which H_0 is rejected, and data confirms to be stationary and suitable for applying ARIMA (p,d,q) model. For a significance level of 95 %, three models ARIMA (1,0,0) ARIMA (1,0,1) and ARIMA (1,1,1) are selected and checked over the goodness of fit statistics for choosing a best-suited model. Table 2 gives the comparisons of the goodness of fit statistics for different pollutant time-series data of different seasons.

From the results of Table 2 and statistical analysis of the residual plots, the ARIMA (1,1,1) model has the least error estimation values in the goodness of fit statistics and thus appears to be best suited for the forecasts of

Table 2: Goodness of fit statistics of different ARIMA models for pollutants with 95 % confidence interval

Pollutant	Goodness of fit statistics	Summer			Monsoon			Winter		
		ARIMA (1,0,0)	ARIMA (1,0,1)	ARIMA (1,1,1)	ARIMA (1,0,0)	ARIMA (1,0,1)	ARIMA (1,1,1)	ARIMA (1,0,0)	ARIMA (1,0,1)	ARIMA (1,1,1)
CO	SST	84.69	68.69	62.64	91.20	62.75	59.78	154.44	125.03	116.25
	MAPD	24.93	24.90	23.95	43.90	56.38	52.89	26.30	27.94	26.64
	W-N Variance	0.17	0.14	0.13	0.18	0.12	0.12	0.31	0.25	0.236
	FPE	0.17	0.14	0.13	0.18	0.12	0.12	0.31	0.25	0.23
	AIC	536	438	392	572	393	368	832	731	693
	RMSD	0.41	0.37	0.36	0.43	0.35	0.35	0.56	0.50	0.48
NO_2	SST	48958	48888	44640	11439	11211	10390	42607	41540	38688
	MAPD	14.25	14.37	14.98	11.89	11.99	12.17	15.50	16.09	17.67
	W-N Variance	101.78	101.63	93.00	23.44	22.97	21.33	86.60	84.43	78.79
	FPE	102.20	102.06	93.39	23.53	23.06	21.42	86.95	84.77	79.11
	AIC	3596	3597	3544	2931	2924	2879	3598.	3588	3543
	RMSD	10.08	10.08	9.64	4.84	4.79	4.61	9.30	9.18	8.87
PM 2.5	SST	2859757	2089733	2004323	3352988	2676713	2640077	3208941	2894414	2781344
	MAPD	37.35	36.21	37.17	45.65	47.66	52.20	27.280	26.28	26.77
	W-N Variance	5945.44	4344.56	4175.67	6870.87	5485.06	5421.10	6522.23	5882.95	5664.65
	FPE	5970.21	4362.66	4193.10	6899.09	5507.59	5443.41	6548.80	5906.91	5687.77
	AIC	5550.49	5402.88	5371.68	5701	5594	5576	5723	5674	5643
	RMSD	77.10	65.91	64.61	82.89	74.06	73.62	80.76	76.70	75.26
PM 10	SST	3824108	2987062	2796227	1762330	1526310	1454165	2840257	2754909	2517953
	MAPD	22.91	24.49	625.02	30.90	32.97	34.49	18.99	19.46	19.76
	W-N Variance	7950	6210	5825	3611	3127	2985	5772	5599	5128
	FPE	7983	6235	5849	3626	3140	2998	5796	5622	5149
	AIC	5690	5574	5531	5388	5320	5285	5664	5651	5594
	RMSD	89.16	78.80	76.32	60.09	55.92	54.64	75.97	74.82	71.61
SO_2	SST	31579	22470	22022	54347	37103	36414	54860	43352	40844
	MAPD	34.09	38.07	37.69	52.23	53.96	57.96	43.55	48.70	49.47
	W-N Variance	65.65	46.71	45.88	111.36	76.03	74.77	111.50	88.11	83.18
	FPE	65.92	46.91	46.07	111.82	76.34	75.08	111.95	88.47	83.52
	AIC	3383	3223	3205	3690	3507	3491	3720	3608	3571
	RMSD	8.10	6.83	6.77	10.55	8.71	8.64	10.55	9.38	9.12

the pollutants value of Varanasi. Fig. 3 shows the actual pollutant values and the forecasted pollutant values with ARIMA (1,1,1) model at a 95 % confidence interval for the year 2013 to 2016 for different seasons. In Fig. 3, ARIMA (1,1,1) plot is compared with the original data points of different pollutants where the blue line represents the 24 h mean concentrations of observed pollutants value and the red line represents the ARIMA (1,1,1) model values. The time step unit on the horizontal axis refers to the 24 h mean point in the

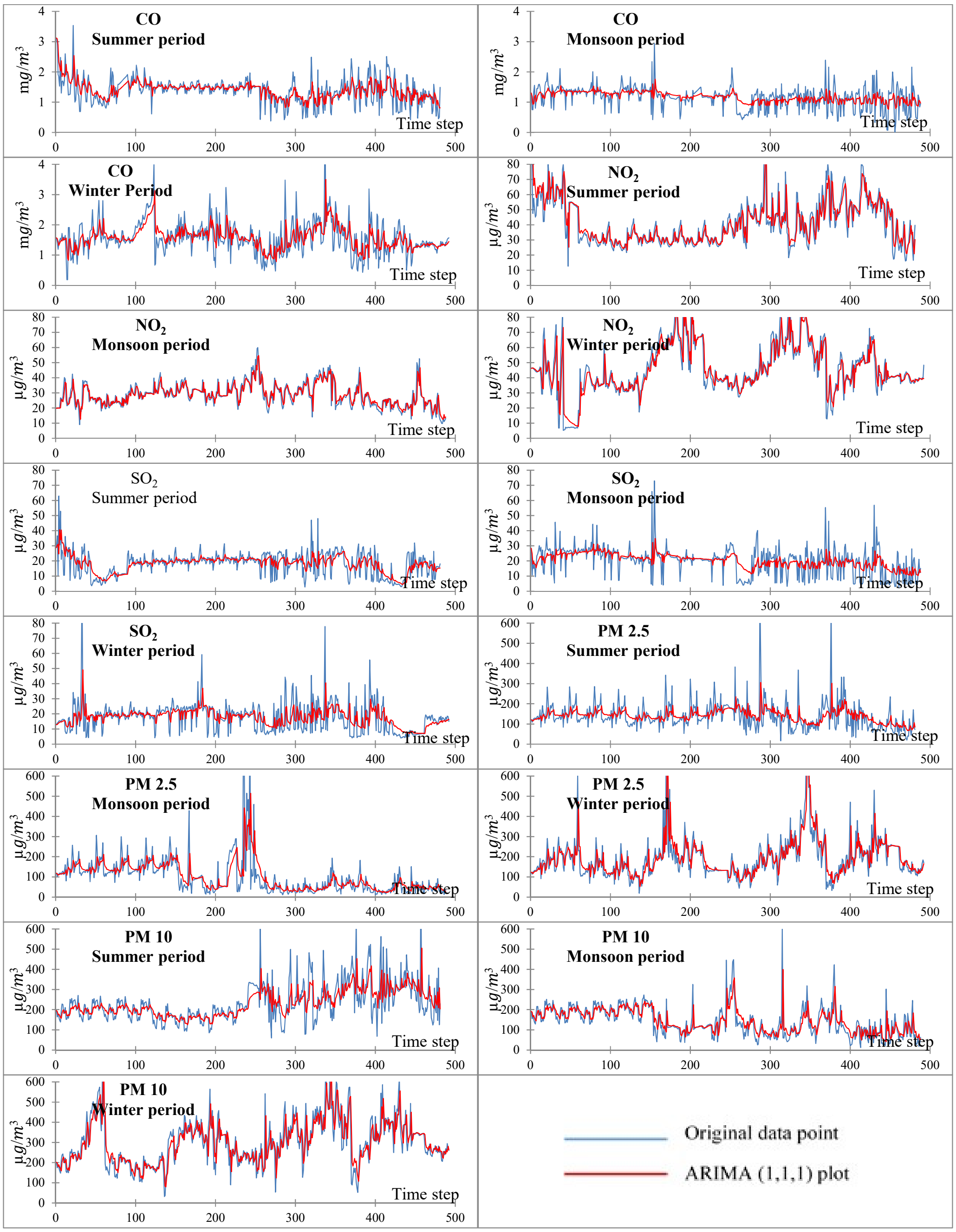

Fig. 3: Pollutants original data point and ARIMA (1,1,1) plot for different seasons from the year 2013 to 2016 of Varanasi

Fig. 3. ARIMA (1,1,1) model can be used for predicting and forecasting the future pollutants values which can aid the decision makers for planning steps to mitigate the pollutants which shows a rise in trend and are above the permissible standard limits. ARIMA (1,1,1) model is also used on the annual average concentration value of the different pollutants in Varanasi, and a forecast is provided in Table 3. As per the outcomes of M-K and Sen's slope estimator test for seasonal data, forecasted annual concentration values of pollutants also show the similar trend by applying the ARIMA (1,1,1) model on the annual data.

It can be observed from Table 3 that there is a reducing trend in annual concentrations of CO, NO_2, SO_2 and PM 2.5, whereas PM 10 shows a rising trend in the annual concentrations of Varanasi. Also, Table 3 illustrates much higher concentrations of PM2.5 and PM 10, to that of the permissible standard annual concentration limits, that is, 40 μg/m^3 of PM2.5 and 60 μg/m^3 of PM10. CO and NO_2 are though under the permissible standard annual concentration limits 2 mg/m^3 and 40 μg/m^3 respectively, and are showing a decreasing trend, but still are very close to the permissible limits and the results from Fig. 3 shows that both the pollutants have frequently crossed the acceptable limits during the year 2013 to 2016. The ARIMA (1,1,1) model prediction of SO_2 shows a better condition and the pollutants in the future years are predicted satisfactorily below to the annual permissible limits of 50 μg/m^3 in Varanasi (National Air Quality Index, 2014; Permissible Level for Pollutants, 2017). The result of the study helps to assess the conditions of different air pollutants in Varanasi in recent past years. It can be inferred from the results of M-K and Sen's slope estimator tests presented in Fig. 2 and Table 1, that more control measures are required for pollutants especially for particulate matter 10 and nitrogen dioxide. Results shows that PM10 and NO_2 are increasing in past years of Varanasi for summer and winter period and thus better policies are required such as improved road traffic conditions, limiting vehicular pollutions by better vehicle types as these are the main sources of PM10 and NO_2 (CAI-Asia Factsheet, 2010; Lenschow *et al.,* 2001). After the introduction of BSES IV environment standard vehicles, the Indian government has somewhat limited the growth of traffic-related NO_2 and PM 10 emissions (Bansal and Bandivadekar, 2013; Hilboll *et al.,* 2017) but still, the positive trend in results indicates the need of better strategies for countering such pollutants. CO, SO_2 and PM2.5, though shows a decreasing tendency in previous years but the low magnitude of their slopes indicates that these pollutants also required specific measures for systematic controlling. Inferences from the results of the ARIMA model gives an estimate that PM10 and PM2.5 are a bigger concern in the coming years and will require specific measures to control its emissions. The study summarizes that PM10 with increasing trend and higher concentrations, and PM2.5

Table 3: Annual concentration of pollutants forecasted by ARIMA (1,1,1) of Varanasi up to the year 2030

Year	CO (in $\frac{mg}{m^3}$) Forecast Value	NO_2 (in $\frac{\mu g}{m^3}$) Forecast Value	PM 2.5 (in $\frac{\mu g}{m^3}$) Forecast Value	PM 10 (in $\frac{\mu g}{m^3}$) Forecast Value	SO_2 (in $\frac{\mu g}{m^3}$) Forecast Value
2013	1.503	38.646	158.151	259.47	19.6
2014	1.503	38.55	158.08	261.984	19.681
2015	1.497	38.694	153.693	272.126	20.778
2016	1.287	39.992	134.939	292.079	18.74
2017	1.298	39.483	142.207	311.983	15.477
2018	1.328	38.899	135.207	272.284	16.519
2019	1.31	38.383	134.605	262.985	16.477
2020	1.322	38.634	137.233	270.224	16.755
2021	1.332	38.715	138.395	272.349	16.971
2022	1.34	38.742	138.909	272.973	17.141
2023	1.347	38.75	139.136	273.156	17.273
2024	1.353	38.753	139.236	273.21	17.376
2025	1.357	38.754	139.28	273.225	17.457
2026	1.361	38.754	139.3	273.23	17.52
2027	1.364	38.754	139.309	273.231	17.569
2028	1.366	38.754	139.313	273.232	17.607
2029	1.368	38.754	139.314	273.232	17.637
2030	1.37	38.754	139.315	273.232	17.66

with decreasing trend but still higher concentration is the primary concern in Varanasi. NO_2 has an increasing trend in seasonal perspective but on annual concentrations shows a reducing trend and is under the accepted concentration levels along with CO and SO_2.

CONCLUSION

The study presented in the paper provides a statistical analysis of trends in the atmospheric pollutants of the district Varanasi, and further, a forecasting model is formulated to predict different pollutants concentrations in the forthcoming years. M-K and Sen's slope estimator tests are applied to past pollutants data retrieved from AQI service station of Varanasi and ARIMA (p,d,q) model is applied for predictive analysis. Results of M-K test shows the existence of a trend in some of pollutants data in different seasons and the outcomes of Sen's slope estimator test defined power of the trends. ARIMA (1,1,1) model resulted in being best suited for predicting the future pollutant levels by comparing the goodness of fit statistics. Results of ARIMA (1,1,1) model on the annual concentration of pollutants shows an increasing trend in PM 10 pollutant and decreasing trend in PM 2.5, CO, SO_2 and NO_2. PM 10 shows a rising trend with predicted approximate annual concentration of 273 µg/m^3 and; PM 2.5, CO, NO_2 and SO_2 show a reducing trend with approximate annual concentration of 139 µg/m^3, 1.37 mg/ m^3, 38 µg/m^3 and 17 µg/m^3, respectively, by the year 2030. The results predicted high concentration with levels above standard permissible limits of PM 10 and PM 2.5 in upcoming years in Varanasi, so explicit measures are required to control these pollutants.

ACKNOWLEDGMENT

Authors are grateful to Indian Institute of Technology (BHU) for facilitating the necessary support to carry out the research. Authors are also thankful to UP pollution control board (UPPCB) and central pollution control board (CBCB), Government of India, for making the pollutants data available for the study.

CONFLICT OF INTERESTS

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

ABBREVIATIONS

%	Percent
A	Number of cases when $x_b - x_a > 0$
ACF	Autocorrelation function
AIC	Akaike Information Criteria
API	Air pollution index
AQI	Air quality index
AR	Auto-regression
ARIMA	Auto regressive integrated moving average
B	number of cases when $x_b - x_a < 0$
CO	Carbon monoxide
CPCB	Central pollution control board
d	Degree of differencing
D-F	Dickey-Fuller
e_t	Error term at time t
FPE	Final prediction error
f_t	Pollutants parameter at time t
h	Hour
H_0	Null hypothesis
H_1	Alternate hypothesis
I	Differencing order integration
j	Tied groups numbers
m	Length of data point in time-series in Sen's slope estimator test
mg/ m^3	Milligram per cubic meter
MA	Moving average
MAPD	Mean absolute percentage deviation
M-K	Mann-Kendall
M[S]	Mean value of S
NH_3	Ammonia
NO_2	Nitrogen dioxide
O_3	Ozone
p	Order of auto-regressive model
PACF	Partial autocorrelation function
Pb	Lead
PM	Particulate matter
PMFG	Planar maximally filtered graph
q	Order of moving-average model
RMSD	Root mean squared deviation
S	Principal statistics value of M-K test
S^*	Kendall's sum
SO_2	Sulphur dioxide
SST	Sum of squared errors

T_i	Variance of residual for Sen's slope estimator
t_i	Number of data value in the i^{th} group
T_{med}	Sen's slope estimator
UP	Uttar Pradesh
UPPCB	UP pollution control board
Varian[S]	Variance value of S
W[S]	Standard normal test statistics
X_a *and* X_b	Individual pollutant data values in the time-series where b>a in Sen's slope estimator test
x_a *and* x_b	Individual pollutant data values in the time-series where b>a in M-K test
z	Length of data point in time-series in M-K test
α	Significance level
β	Backward-shift operator
$\mu g/m^3$	Microgram per cubic meter
μm	Micrometer
τ	Mann-Kendall's tau value
$\Theta(\beta)$	Polynomial of degree q
$\phi(\beta)$	Polynomial of degree p
∇	Difference operator

REFERENCE

Aleksandropoulou, V.; Lazaridis, M., (2004). Spatial distribution of gaseous and particulate matter emissions in Greece. Water Air Soil Pollut., 153(1-4): 15-34 **(20 pages)**.

Azmi, S.Z.; Latif, M.T.; Ismail, A.S.; Juneng, L.; Jemain, A.A., (2010). Trend and status of air quality at three different monitoring stations in the Klang Valley, Malaysia. Air Qual. Atmos. Health., 3(1): 53-64 **(12 pages)**.

Bansal, G.; Bandivadekar, A., (2013). Overview of India's vehicle emissions control program. ICCT, Beijing, Berlin, Brussels, San Francisco, Washington.

Bernard, S.M.; Samet, J.M.; Grambsch, A.; Ebi, K.L.; Romieu, I., (2001). The potential impacts of climate variability and change on air pollution-related health effects in the United States. Environ. Health Perspect., 109(Suppl 2): 199-209 **(11 pages)**.

Box, G.E.; Jenkins, G.M., (1976). Time series analysis: forecasting and control. Revised Ed. Holden-Day.

Brocklebank, J.C.; Dickey, D.A.; Choi, B., (2018). SAS for forecasting time series. SAS institute.

CAI-Asia Factsheet, (2010). Nitrogen dioxide (NO2): status and trends in Asia. No. 14.

Caloiero, T.; Coscarelli, R.; Ferrari, E.; Sirangelo, B., (2017). Trend analysis of monthly mean values and extreme indices of daily temperature in a region of southern Italy. Int. J. Climatol., 37(S1): 284-297 **(14 pages)**.

Can, A., (2017). Time Series Analysis of air pollutants for Karabük province. In ITM web of conferences Vol. 9, p. 02002. EDP Sciences. AMCSE, Rome, Italy 5-7 October.

Chattopadhyay, G.; Chakraborthy, P.; Chattopadhyay, S., (2012). Mann–Kendall trend analysis of tropospheric ozone and its modeling using ARIMA. Theor. Appl. Climatol., 110(3): 321-328 **(8 pages)**.

Chaudhuri, S.; Dutta, D., (2014). Mann–Kendall trend of pollutants, temperature and humidity over an urban station of India with forecast verification using different ARIMA models. Environ. Monit. Assess., 186(8): 4719-4742 **(24 pages)**.

Cohen, A.J.; Ross Anderson, H.; Ostro, B.; Pandey, K.D.; Krzyzanowski, M.; Künzli, N.; Gutschmidt, K., Pope, A.; Romieu, I.; Samet, J.M.; Smith, K., (2005). The global burden of disease due to outdoor air pollution. J. Toxicol. Environ. Health, Part A, 68(13-14): 1301-1307 **(7 pages)**.

Da Silva, R.M.; Santos, C.A.; Moreira, M.; Corte-Real, J.; Silva, V.C.; Medeiros, I.C., (2015). Rainfall and river flow trends using Mann–Kendall and Sen's slope estimator statistical tests in the Cobres river basin. Nat. Hazard., 77(2): 1205-1221 **(17 pages)**.

Dai, Y.H.; Zhou, W.X., (2017). Temporal and spatial correlation patterns of air pollutants in Chinese cities. PloS One, 12(8): 1-24 **(24 pages).**

Dickey, D.A.; Fuller, W.A., 1(979). Distribution of the estimators for autoregressive time series with a unit root. J. Am. Stat. Assoc., 74(366a): 427-431 **(5 pages)**.

Drápela, K.; Drápelová, I., (2011). Application of Mann-Kendall test and the Sen's slope estimates for trend detection in deposition data from Bílý Kříž (Beskydy Mts., the Czech Republic) 1997-2010. Beskydy, 4(2): 133-146 **(14 pages)**.

Emami, F.; Masiol, M.; Hopke, P.K., (2018). Air pollution at Rochester, NY: long-term trends and multivariate analysis of upwind SO2 source impacts. Sci. Total Environ., 612: 1506-1515 **(10 pages)**.

Emberson, L.; Ashmore, M.; Murray, F., (2003). Air pollution impacts on crops and forests: a global assessment. Imperial College Press.

Eymen, A.; Köylü, Ü., (2018). Seasonal trend analysis and ARIMA modeling of relative humidity and wind speed time series around Yamula Dam. Meteorol. Atmos. Phys.: 1-12 **(12 pages)**.

Gocic, M.; Trajkovic, S., (2013). Analysis of changes in meteorological variables using Mann-Kendall and Sen's slope estimator statistical tests in Serbia. Global Planet. Change, 100: 172-182 **(11 pages)**.

Gulia, S.; Nagendra, S.S.; Khare, M.; Khanna, I., (2015). Urban air quality management-a review. Atmos. Pollut. Res., 6(2): 286-304 **(19 pages)**.

Gurjar, B.R.; Butler, T.M.; Lawrence, M.G.; Lelieveld, J., (2008). Evaluation of emissions and air quality in megacities. Atmos. Environ., 42(7): 1593-1606 **(14 pages)**.

Hilboll, A.; Richter, A.; Burrows, J.P., (2017). NO2 pollution over India observed from space–the impact of rapid economic growth, and a recent decline. Atmos. Chem. Phys. Discuss.

Hirsch, R.M.; Alexander, R.B.; Smith, R.A., (1991). Selection of methods for the detection and estimation of trends in water quality. Water Resour. Res., 27(5): 803-813 **(11 pages)**.

Jaruskova, D.; Liska, I., (2011). Statistical analysis of trends in organic pollution and pollution by nutrients at selected Danube river stations. J. Environ. Monit., 13(5): 1435-1445 **(11 pages)**.

Jayamurugan, R.; Kumaravel, B.; Palanivelraja, S.; Chockalingam, M.P., (2013). Influence of temperature, relative humidity and seasonal variability on ambient air quality in a coastal urban area. Int. J. Atmos. Sci.: **(7 pages).**

Junninen H.; Niska H.; Tuppurainen K.; Ruuskanen J.; Kolehmainen M., (2004). Methods for imputation of missing values in air quality data sets. Atmos. Environ., 38: 2895–2907 **(13 pages)**.

Kampa, M.; Castanas, E., (2008). Human health effects of air pollution. Environ. Pollut., 151(2): 362-367 **(5 pages)**.

Kendall M.G., (1975). Rank correlation methods. Griffin, London.

Kisi, O.; Ay, M., (2014). Comparison of Mann–Kendall and innovative trend method for water quality parameters of the Kizilirmak river, Turkey. J. Hydrol., 513: 362-375 **(14 pages)**.

Kumar, D.S.; Bhushan, S.H; Kishore, D.A., (2018). Atmospheric dispersion model to predict the impact of gaseous pollutant in an industrial and mining cluster. Global J. Environ. Sci. Manage., 4(3): 351-358 **(8 pages)**.

Lenschow, P.; Abraham, H.J.; Kutzner, K.; Lutz, M.; Preuß, J.D.; Reichenbächer, W., (2001). Some ideas about the sources of PM10. Atmos. Environ., 35: S23-S33 **(11 pages)**.

Ma, J.; Hung, H.; Tian, C.; Kallenborn, R., (2011). Revolatilization of persistent organic pollutants in the Arctic induced by climate change. Nat. Clim. Change, 1(5): .255-260 (**6 pages)**.

Mann, H.B., (1945). Nonparametric tests against trend. Econometrica, 13: 245–259 **(15 pages)**.

Mayer, H., (1999). Air pollution in cities. Atmos. Environ., 33(24-25): 4029-4037 **(9 pages)**.

National Air Quality Index, (2014). CPCB, ministry of environment, forests & climate Change, GoI.

Pandolfi, M.; Alastuey, A.; Pérez, N.; Reche, C.; Castro, I.; Shatalov, V.; Querol, X., (2016). Trends analysis of PM source contributions and chemical tracers in NE Spain during 2004–2014: a multi-exponential approach. Atmos. Chem. Phys., 16(18): 11787-11805 **(19 pages)**.

Pascal, M.; Corso, M.; Chanel, O.; Declercq, C.; Badaloni, C.; Cesaroni, G.; Henschel, S.; Meister, K.; Haluza, D.; Martin-Olmedo, P.; Medina, S., (2013). Assessing the public health impacts of urban air pollution in 25 European cities: results of the Aphekom project. Sci. Total Environ., 449: 390-400 **(11 pages)**.

Permissible level for pollutants, (2017). Ministry of environment, forest and climate change, GoI.

Rainfall statistics of India 2013- 2016, (2016). India Meteorological Department, Ministry of Earth Science.

Rani, N.L.A.; Azid, A.; Khalit, S.I.; Juahir, H.; Samsudin, M.S., (2018). Air pollution index trend analysis in Malaysia, 2010-15. Pol. J. Environ. Stud., 27(2): 801–807 **(7 pages).**

Rodríguez, M.C.; Dupont-Courtade, L.; Oueslati, W., (2016). Air pollution and urban structure linkages: evidence from European cities. Renewable Sustainable Energy Rev., 53: 1-9 **(9 pages)**.

Safi, M., (2016). Air quality in holy city of Varanasi 'most toxic in India'. The Guardian, Decmber 12 2016.

Sen, P.K., (1968). Estimates of the regression coefficient based on Kendall's tau. J. Am. Stat. Assoc., 63(324): 1379-1389 **(11 pages)**.

Shukla, J.B.; Misra, A.K.; Sundar, S.; Naresh, R., (2008). Effect of rain on removal of a gaseous pollutant and two different particulate matters from the atmosphere of a city. Math. Comput. Modell., 48(5-6): 832-844 **(13 pages)**.

Theil, H., (1950). A rank-invariant method of linear and polynomial regression analysis I II III, Nederlandse Akademie Wetenschappen Proc. Series A 53:386–392 (Part I), 521–525 (Part II), 1397–1412 (Part III) **(16 pages)**.

Vanguelova, E.I.; Benham, S.; Pitman, R.; Moffat, A.J.; Broadmeadow, M.; Nisbet, T.; Durrant, D.; Barsoum, N.; Wilkinson, M.; Bochereau, F.; Hutchings, T., (2010). Chemical fluxes in time through forest ecosystems in the UK–soil response to pollution recovery. Environ. Pollut., 158(5): 1857-1869 (23 pages).

Watthanacheewakul, L., (2011). Comparisons of power of parametric and nonparametric test for testing means of several Weibull populations. In Proceedings of the International Multi-Conference of Engineers and Computer Scientist. Hong Kong, 16-18 March.

Xu, Z.X.; Chen, Y.N.; Li, J.Y., (2004). Impact of climate change on water resources in the Tarim river basin. Water Resour. Manage., 18(5): 439-458 (20 pages).

Zuma-Netshiukhwi, G.; Stigter, K.; Walker, S., (2013). Use of traditional weather/climate knowledge by farmers in the South-western free state of South Africa: Agrometeorological learning by scientists. Atmos., 4(4): 383-410 (28 pages).

AUTHOR (S) BIOSKETCHES

Jaiswal, A., Ph.D. Candidate, Research Scholar, Department of Mechanical Engineering, Indian Institute of Technology, Varanasi, India. Email: *ajaiswal.rs.mec13@itbhu.ac.in*

Samuel, C., Ph.D., Associate Professor, Department of Mechanical Engineering, Indian Institute of Technology, Varanasi, India. Email: *csamuel.mec@itbhu.ac.in*

Kadabgaon, V.M., B.E., Department of Mechanical Engineering, Brahmdevdata Mane Institute of Technology, Solapur, India. Email: *vinkadabgaon@gmail.com*

HOW TO CITE THIS ARTICLE

Jaiswal, A.; Samuel, C.; Kadabgaon, V.M., (2018). Statistical trend analysis and forecast modeling of air pollutants. Global. J. Environ. Sci. Manage., 4(4): 427-438.

DOI: 10.22034/gjesm.2018.04.004

url: http://www.gjesm.net/article_32677.html