iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** I **Month of publication:** January 2023

DOI: https://doi.org/10.22214/ijraset.2023.48774

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue I Jan 2023- Available at www.ijraset.com

Air Population Components Estimation in Silk Board Bangalore, India

Nikitha Masineni
M.S Ramaiah University Of Applied Sciences

Abstract: ***As we see on global arena most of the political leaders focus us on how to reduce carbon prints and make the planet a safer place to breathe. Though the focus on reducing pollution we hardly have technologies which focusses on reducing pollution, while reducing pollution is one aspect identifying source of pollution is another important aspect. Even though we intent to reduce pollution we still see there is lot of shortcomings. We intend to focus on how pollutants vary over a period of time and if there is any seasonal pattern, We also look forward to vary pollutants to see which pollutants causes variation in particle matter i.e Pm2.5 and Pm10. SARIMA modelling is used which mainly focuses on decomposing the data and giving the residual details. The RMSE is 10.97 which shows the model is efficient enough to predict the pollutants and particulate matter***
Keywords: SARIMA, ARIMA, POLLUTANT,PM2.5,PM10,Air Pollution

I. INTRODUCTION

In 2010 for example, a loss of 0.65 million healthy years and more than 0.62 million premature deaths in India were attributed to outdoor air pollution [1]. A very thorough study [2] done by the Global Burden of Diseases study published in 2017 showed that 4.2 million deaths were attributed to the influence of air pollution in 2015 out of which, 1.2 million were in India. There has been a deadly effect on the lives of people, and due to this, there is a need for accurate monitoring and reasoning about environmental phenomena and to find out effective measures to combat the damage caused by air pollution. A way to improve the understanding of how air pollution behaves throughout time is by applying prediction mechanisms. Monitoring and predicting the environment, specifically air pollution levels, is mostly done using extensive sensor networks, which are part of a greater paradigm of cyber-physical systems implemented nowadays. To tackle the Air Quality (AQ) monitoring and prediction a combination of IoT networks, contextaware concepts and machine learning techniques can be applied. In this work we combine these areas to prove that improvement can be achieved over other conventional approaches

Environmental monitoring data can be described by multivariate time series compliances generated from geo-located monitoring stations. For our scenario, urban air quality monitoring data is obtained from monitoring stations in the city which consist of many air pollutant concentration values (such as fine particles, carbon monoxide, sulphur dioxide, nitrogen oxides zone, etc),. To discuss on pollutants, Adverse health impacts from exposure to outdoor air pollutants are complicated functions of pollutant compositions and concentrations. Major outdoor air pollutants in cities include ozone (O3), particle matter (PM), sulphur dioxide (SO2), carbon monoxide (CO), nitrogen oxides (NOx), volatile organic compounds (VOCs), pesticides, and metals, among others. According to the report from the American Lung Association [10], 10 parts per billion (ppb) increase in the O3 mixing ratio might cause over 3700 premature deaths annually in the United States (U.S.). Meteorological conditions, including regional and synoptic meteorology, are critical in determining the air pollutant concentrations [14–19]. In the study by Holloway et al. [20], the O3 concentration over Chicago was found to be most sensitive to air temperature, wind speed and direction, relative humidity, incoming solar radiation, and cloud cover. Humidity is connected with air pollution, the higher humidity, the higher the concentration in air pollution. Because various particle compositions and their interactions with light were found to be the most important factors in attenuating visibility, low visibility could be an indicator of high PM concentrations. In the formation of air pollution, some clouds absorb solar radiation (e.g., O3). Therefore, these important meteorological variables were selected to predict air pollutant variation with time in our work. Our work , we focus on refined modelling for predicting hourly air pollutant concentrations on the basis of historical metrological data and air pollution data. A striking difference between this work and the previous works is that we emphasize how to regularize the model in order to improve its generalization performance and how to learn a complex regularized model from big data. Also with the current situation such as COVID we have understood that forecasting has huge dependencies on variabilities, So we have proposed a model to predict the pollutant content in case of linear variability of pollutants as well as random variation of pollutants, with the focus to see what happens if certain pollutants were limited or varied linearly to the particulate matter.

The other reason for proposing the methodology is to understand by controlling the particular type of pollutant keeping in mind the climatic variation how pollution can be controlled. The output would provide us the insight of which industry pollutants could impact in the pollution. Finally to sum up our objective most of the former machine learning works on air pollutant prediction did not consider the Pollutants how that influences pollution and also looked only into similarities between the models and only focused on improving the model performance for a single task, that is, improving prediction performance for each hour either separately or identically. Therefore, we decided to use meteorological and pollutant data to perform predictions of hourly concentrations on the basis of ARIMA models. Hence we would focus on time series prediction of pollutants with the capacity of producing results for hourly, 3 hours, 7 hours, 1 day, 7 days prediction for all the pollutants and the Particulate matter. We also try to look into linear and random variabilities in the pollutants to understand the variation in Particulate matter. To the best of our knowledge, this is the first work that has utilized ARIMA based modelling for the air pollutant prediction task. This Study used analytical approaches and optimization techniques to obtain the optimal solutions. The model's evaluation metric is the root-mean-squared error (RMSE). To present the use case of our project , lets consider a scenario where the city of Melbourne, Victoria in Australia has been keeping track of its AQ levels throughout the past 10 years, with many sensors scattered across many districts of the megalopolis. The usual information consists of meteorological factors (such as temperature, humidity, wind speed and direction, amongst others) and air pollutants (such as Particle Matter under 2.5 μm of diameter (PM2.5), Carbon Monoxide (CO), Nitrogen Dioxide (NO2), etc). These historical datasets can be used to predict future AQ levels to a certain degree of accuracy, but they cannot handle high sudden peaks of pollution or reduction in pollution occurring due to abnormal phenomena, like sudden high vehicle traffic peaks in highways, or a sudden bushfire outbreak or pandemic like Corona making zero traffic.

II. LITERATURE SURVEY

Voukantsis et al. (2007) propose a methodology to compare the meteorological data and air quality for predicting the air pollutants of interest in the urban areas based on computational intelligence methods, principal component analysis and artificial neural networks. They formulated a hybrid scheme of linear regression and ANN models for developing air quality forecasting models. Gulliver et al. (2011) proposed an air pollution model to forecast annual and Kalapanidas et al. [21] elaborated effects on air pollution only from meteorological features such as temperature, wind, precipitation, solar radiation, and humidity and classified air pollution into different levels in the system. Ni, X.Y.; Huang [22] compared multiple statistical models on the basis of PM2.5 data around Beijing, and their results implied that linear regression models can in some cases be better than the other models. MTL focuses on learning multiple tasks that have commonalities Shweta Taneja et al[23], proposed paper of Predicting Trends in Air Pollution in Delhi using Data Mining. In this Paper, They have used time series analysis method for analysing the pollution trends in Delhi and predicting about the future. The time series method includes Multilayer Perceptron and Linear Regression In Springer (2018) Paper, Air Pollution Prediction Using Extreme Learning Machine, it was a case study on Delhi ELM-based prediction was found to have greater accuracy than the existing [24] .Azid et al[25]. used principal component analysis (PCA) to analyse the major components affecting air quality and to predict the air pollutant concentration by the predictive ability of neural network.

III. METHODOLOGY

This study uses Time series data which is a sequential set of data points arranged in a chronological order. It is usually measured over successive times. It has a set of vectors which is x(t), t = 0,1,2,3,4, and so on. Here, T is the time that has been elapsed. The time series which has a single variable is called as univariate time series. The time series which has a more than one variable is called as multivariate time series. A time series can be continuous or discrete. As we go through our data it falls under discrete observation. A time series in general is supposed to be affected by four main components which are: Trend, Cyclical, Seasonal and Irregular components. The cyclical variation repeats in cycles. The duration of a cycle extends over longer period of time, usually two or more years. Most of the time series show some kind of cyclical variation. Schematically a typical business cycle can be shown in figure 1

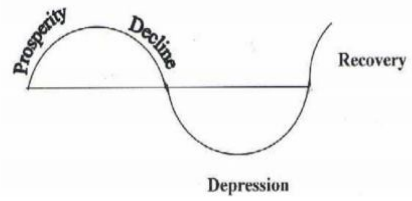


Figure 1:Data Flow in Timeseries

Irregular or random variations in a time series are caused by unpredictable influences, which are not regular and also do not repeat in a particular pattern; which can be caused by floods, war, etc. There is no defined statistical technique for measuring random fluctuations in a time series. The models time series is Multiplicative (Eq1) and Additive models (Eq2)

$$Y(t) = T(t) \times S(t) \times C(t) \times I(t) \quad - (1)$$

$$Y(t) = T(t) + S(t) + C(t) + I(t) \quad - (2)$$

Y(t is the observation and) T(t ,) S(t ,) C(t) I(t)are respectively the trend, seasonal, cyclical and irregular variation at time .t. For the Time series, In the multiplicative model, it is assumed that four components are not necessarily independent and they can affect each other but in additive model it is assumed that they are independent. To visualize the basic pattern of the data, usually a time series is represented by a graph, where the observations are plotted against corresponding time. Below we show time series plots in Fig 2

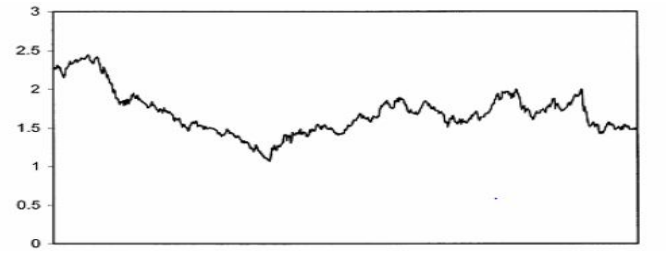


Figure 2: Timeseries Plot

In, time series model there are two models that are widely used, which is called Autoregressive (AR) and Moving Average (MA) models.

This study uses models such as Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA). The model that generalizes ARMA and ARIMA, is called as Autoregressive Fractionally Integrated Moving Average (ARFIMA). Seasonal Autoregressive Integrated Moving Average (SARIMA) model is used for seasonal forecasting of the time series. ARIMA expects data that is either not seasonal or has the seasonal component removed and it has three new hyperparameters to specify the autoregression (AR), differencing (I) and moving average (MA) for the seasonal component of the series, as well as an additional parameter for the period of the seasonality The autoregressive component of the ARIMA model is denoted by AR(p); where p is called as the parameter which enables the number of lagged series, it also includes variables like seasonality and exogenous- which is very poweful.

$$y_t = k + \sum_{n=1}^{p} a_n \, y_{t-n} + epsilon(t) \quad - (3)$$

Increasing the p parameter means going back and adding more timestamps adjusted by their own multipliers and we have to use Moving average (MA9(q)).Since the ARIMA model assumes that the time series is stationary, we cannot use it for Air pollution as it changes seasonally

$$y_t = k + \sum_{n=1}^{p} a_n \, y_{t-n} + \sum_{n=1}^{q} b_n \, K_{t-n} + \sum_{n=1}^{P} \mu_n \, Y_{t-sn} + epsilon(t)$$

IV. RESULTS

The below section has the results discussed, The below plot in figure 3 , shows the variation of data with respect to particulate matter and time . If we observe the data there is a seasonal trend which can be seen, there is a peak and dip in pollution

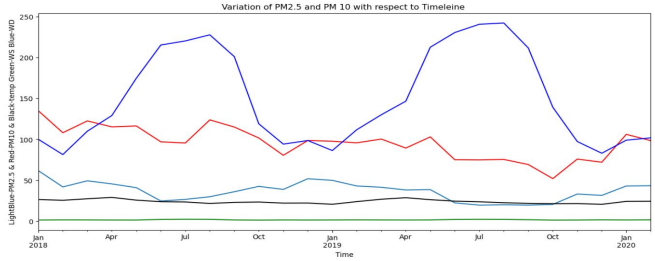


Figure 3:Seasonal Trend

The model SARIMA has been trained and the regression result for both the PM particle is shown in Figure 4 and Figure 5, On training the model has seen a RMSE value of 10.97 which iterates there could be a variation of (+/-) 10.97 in the predicted value .As we are looking at the AQI index of pollution this is under an acceptable limit and hence the model can be used to predict variation .

----------FOR MODEL 2.5 PM----------

OLS Regression Results

Dep. Variable:	PM2.5		
µg/m³ R-squared:	0.479		
Model:	OLS	Adj. R-squared:	0.478
Method:	Least Squares	F-statistic:	1007.
Date:	Sun, 20 Nov 2022	Prob (F-statistic):	0.00
Time:	21:39:15	Log-Likelihood:	-78217.
No. Observations:	18672	AIC:	1.565e+05
Df Residuals:	18654	BIC:	1.566e+05
Df Model:	17		
Covariance Type:	nonrobust		

Figure 4:Regression for PM2.5

----------FOR MODEL 10 PM----------

OLS Regression Results

Dep. Variable:	PM10		
µg/m³ R-squared:	0.501		
Model:	OLS	Adj. R-squared:	0.501
Method:	Least Squares	F-statistic:	1103.
Date:	Sun, 20 Nov 2022	Prob (F-statistic):	0.00
Time:	21:39:15	Log-Likelihood:	-92802.
No. Observations:	18672	AIC:	1.856e+05
Df Residuals:	18654	BIC:	1.858e+05
Df Model:	17		
Covariance Type:	nonrobust		

Figure 5: Regression for PM10

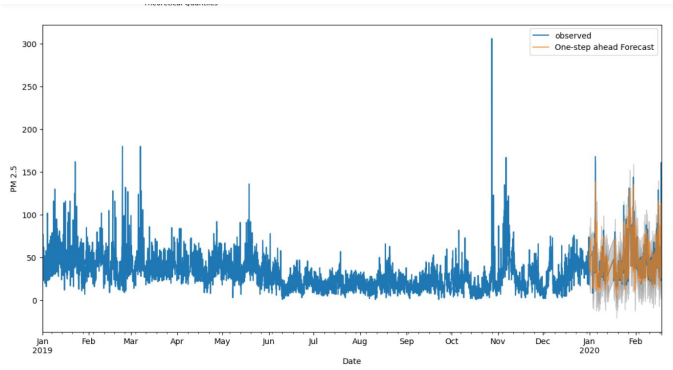


Figure 6: Actual vs Predicted Value

Figure 6 shows how there is a data trend in Actual value and predicted value, As we see the output graph both values are almost overlapping. This shows us the model is able to give a accurate result

V. CONCLUSION

With the available data we have included seasonal parameter and combined it with regression model to get the expected pollutants, further this model also helps us to identify which pollutant can control the particulate matter. Future as a enhanced scope we can use real time information to predict the variability in parameter and hence give more accurate AQI index which has a offset of less than 15 minutes .

REFERENCES

[1] Yin, P., et al. (2017). Particulate air pollution and mortality in 38 of China's largest cities: time series analysis. Bmj, 667(March), p. j667. ISSN 0959-8138, doi:10.1136/bmj.j667, url: http://www.bmj.com/lookup/doi/10.1136/bmj.j667.

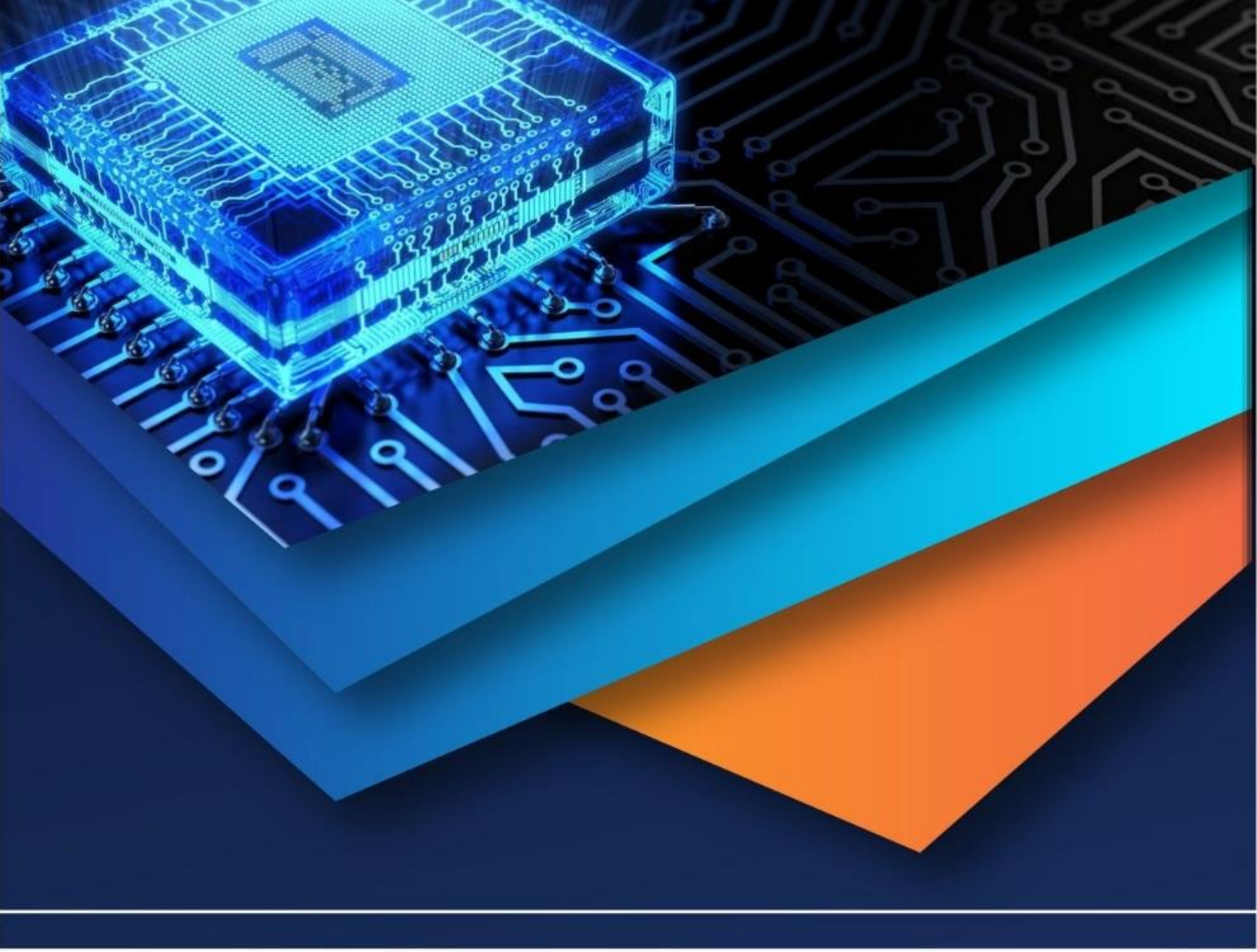
[2]] Cohen, A.J., et al. (2017). Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. The Lancet, 389(10082), pp. 1907–1918. ISSN 1474547X, doi:10.1016/S0140-6736(17)30505-6, url: http://dx.doi.org/10.1016/ S0140-6736(17)30505-6.

[3] Kraak, M.J.; Ormeling, F. Cartography: Visualization of Spatial Data; Guilford Press: New York, NY, USA ,2011

[4] Guo, D.; Chen, J.; MacEachren, A.M.; Liao, K. A visualization system for space-time and multivariate patterns (vis-stamp). IEEE Trans. Vis. Comput. Graph. 2006, 12, 1461–1474.

[5] . Long, Y.; Wang, J.; Wu, K.; Zhang, J. Population Exposure to Ambient PM 2.5 at the Subdistrict Level in China. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2486602 (accessed on 27 August 2014).

[6] Rohde, R.A.; Muller, R.A. Air pollution in China: Mapping of concentrations and sources. PLoS ONE 2015, 10, e0135749. 16. Sicard, P.; Serra, R.; Rossello, P. Spatiotemporal trends in ground-level ozone concentrations and metrics in France over the time period 1999–2012. Environ. Res. 2016, 149, 122–144.
[7] Huan, L.; Hong, F.; Feiyue, M. A Visualization Approach to Air Pollution Data Exploration—A Case Study of Air Quality Index (PM2.5) in Beijing, China. Atmosphere 2016, 7, 35.
[8] Chung, K.L.; Qu, H.; Chan, W.Y.; Guo, P.; Xu, A.; Lau, K.H. Visual Analysis of the Air Pollution Problem in Hong Kong. IEEE Trans. Vis. Comput. Graph. 2007, 13, 1408–1415.
[9] Zhang, Y.L.; Cao, F. Fine particulate matter (PM2.5) in China at a city level. Sci. Rep. 2015, 5. doi:10.1038/srep14884
[10] American Lung Association. State of the Air Report; ALA: New York, NY, USA, 2007; pp. 19–27.
[11] Environmental Protection Agency (EPA). Region 5: State Designations, as of September 18, 2009. Available online: https://archive.epa.gov/ozonedesignations/web/html/region5desig.html (accessed on 17 December 2017).
[12] Hinds, W.C. Aerosol Technology: Properties, Behavior, and Measurement of Airborne Particles; John Wiley & Sons: Hoboken, NJ, USA, 2012
[13] Soukup, J.M.; Becker, S. Human alveolar macrophage responses to air pollution particulates are associated with insoluble components of coarse material, including particulate endotoxin. Toxicol. Appl. Pharmacol. 2001, 171, 20–26.M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.
[13] Kalkstein, L.S.; Corrigan, P. A synoptic climatological approach for geographical analysis: Assessment of sulfur dioxide concentrations. Ann. Assoc. Am. Geogr. 1986, 76, 381–395.
[14] Comrie, A.C. A synoptic climatology of rural ozone pollution at three forest sites in Pennsylvania. Atmos. Environ. 1994, 28, 1601–1614.
[15] Eder, B.K.; Davis, J.M.; Bloomfield, P. An automated classification scheme designed to better elucidate the dependence of ozone on meteorology. J. Appl. Meteorol. 1994, 33, 1182–1199
[16] Zelenka, M.P. An analysis of the meteorological parameters affecting ambient concentrations of acid aerosols in Uniontown, Pennsylvania. Atmos. Environ. 1997, 31, 869–878.
[17] Laakso, L.; Hussein, T.; Aarnio, P.; Komppula, M.; Hiltunen, V.; Viisanen, Y.; Kulmala, M. Diurnal and annual characteristics of particle mass and number concentrations in urban, rural and Arctic environments in Finland. Atmos. Environ. 2003, 37, 2629–2641.
[18] Jacob, D.J.; Winner, D.A. Effect of climate change on air quality. Atmos. Environ. 2009, 43, 51–63
[19] Holloway, T.; Spak, S.N.; Barker, D.; Bretl, M.; Moberg, C.; Hayhoe, K.; Van Dorn, J.; Wuebbles, D. Change in ozone air pollution over Chicago associated with global climate change. J. Geophys. Res. Atmos. 2008, 113, doi:10.1029/2007JD009775.
[20] Kalapanidas, E.; Avouris, N. Short-term air quality prediction using a case-based classifier. Environ. Model. Softw. 2001, 16, 263–272.
[21] Manisha Bisht and K.R. Seeja," Air Pollution Prediction Using Extreme Learning Machine: A Case Study on Delhi.", Springer(2018)
[22] P. Jiang, Q. Dong, and P. Li, "A novel hybrid strategy for PM2.5 concentration analysis and prediction," Journal of Environmental Management, vol. 196, pp.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)