Proceedings of the Fifth International Conference on Computing Methodologies and Communication (ICCMC 2021)
IEEE Xplore Part Number: CFP21K25-ART

Data Analytics based Statistical Analysis of Air Pollution in the Major Cities of Karnataka

Sachin Bhat*, Gopika C B, Shetty Namratha Anil, Shreya H P, Prajwala Shetty
Shri Madhwa Vadiraja Institute of Technology and Management, Bantakal – 574115, Karnataka, India
*sachinbhat88@gmail.com

2021 5th International Conference on Computing Methodologies and Communication (ICCMC) | 978-1-6654-0360-3/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICCMC51019.2021.9418342

Abstract - **Air pollution is the discharge of contaminations into the air that is damaging to human wellbeing and the planet. Air contamination is one of the big issues to deal with for any nation. In South Asia, they rank it as the sixth most risky killer. People won't realize the damaging outcomes of hassle if he/she has not experienced it. Pollutants comprising particulate matter (PM) or nitrogen dioxide (NO2) can cause respiratory ailments or cardiovascular sicknesses. World Health Organization tested that 11.6% of worldwide sudden demise may follow back to air pollution. The Air Quality Data is capable to perceive the patterns and identify correlating elements on key ranges of Air Pollution. To our knowledge, this analysis offers to deal with distinguishing a pollution source as starting from an individual area considering collected records. This gives a helpful strategy for air quality administration, and the outcome may be noteworthy to ecological and monetary issues. In this article, the air quality information of the Indian state, Karnataka were analyzed to find standards or styles which might deliver our perception into how extreme the trouble can be measured.**

Keywords— Air pollution, pollution control board, statistics analysis.

I. INTRODUCTION

Air pollution is an ecological and social hassle because it prompts many antagonistic outcomes. The presence of particulates or other destructive materials in the environment reasons air contamination, which results in human mortality, and harm to other living organisms which includes creatures and food crops. Air pollution may start from anthropogenic or natural sources [1]. The traditional air pollution tracking process installs a few very precise, expensive stations at symbolic areas. Their inadequate analysis, the low geographical verdict is not enough to measure urban air contamination and its effects on human wellbeing and surroundings. Advances in minimal effort transportable air pollution sensors enabled air pollution tracking implementation to gauge air contamination at peak spatiotemporal resolution. Pollution affects the standard ways of life and general wellbeing. Pollution particles that incorporate particulate matter (PM), ozone (O3), carbon monoxide (CO) or nitrogen dioxide (NO2) can be the origin of respiratory sufferings, or cardiovascular ailments. Inspection of World Health Organization tested that 11.6% of overall demise is because of atmosphere pollutants [2]. An endeavor has been made to see air contamination hotspots for over 50 years. In this article, an endeavor is made to reveal insight into the source of air pollutions. Today many developing nations, for example, China and India, and once in a while even propelled nations endure exceptional air pollution, which has related to fast urbanization and development. Poor air grade is also equally intense and produces consequences to human wellness. The past research has shown that unveiling fine-particulate air contamination has negative results on cardio-pulmonary well-being and is a significant threat for non-communicable diseases, for example, stroke, interminable obstructive pneumonic ailment, and lung disease. Contamination can likewise affect neighborhood and territorial climate and the atmosphere, with upwardly reflecting and captivating approaching solar radiation and changing cloud effects. In each year, all around, 3.7 million unexpected losses were because of atmosphere air contamination. In India, this amount is about 1.2–2 million passing every year. During most recent years, although the authority has taken some strict measures, the strength, rate of occurrence, and impact scale of air contamination have now not radically improved.

II. LITERATURE REVIEW

Few current techniques for air pollution have been depicted as beneath. In the plug and sense gadget approach, it uses more than one sensor with local co-ordinate, AQI LED marker has activated as steady with toxins degree and the ongoing pollution level pictured with the utilization of a line diagram [4,12]. The author in [5] uses distributed sensor data registering, it uses dispensed knowledge for sensor hubs and uses a spatial database for areas. Since 2011, repeated incidences of smog in China have become a reason for routinely show up as a significant theme in the media and on atmosphere sites. Along these lines, significant research in different fields has concentrated on air contamination, endeavoring to tackle the issue with various approaches. On many occasions, the strategies are not instinctive and are hard for administrative authorities and the overall population to comprehend. The visual investigation of air contamination with patio-fleeting information is an answer that makes complex information reasonable because graphical portrayal is moderately instinctive [6,13]. The present contamination estimation philosophy uses costly hardware at fixed areas or devoted versatile gear. It uses the crude information gained to

additionally extrapolate the degree and grouping of contamination through scattering models. This is a coarse-grained framework where the contamination estimations are few and far in the middle. The far-reaching organization of this estimation worldview has been compelled by its restrictive expense. In expansion, it is alluring to approach continuous estimations to have the option to rapidly dissect and distinguish disturbing degrees of toxins. As of now, access to such information has constrained if not missing. It is accessible to and observable by just rare sorts of people who are very much educated regarding the matter of contamination [7].

III. MATERIALS AND METHODS

Pollution control is a hard assignment in the current situation. The absolute initial step to control pollution is to distinguish the source of contamination. The urban territories are more contaminated than provincial because of the high population density. Data mining is used to break down the information and to focus on the air quality data. The primary point of the data mining process is to remove the helpful data from the dossier of information and shape it into a reasonable structure for the future [8].

3.1 Dataset preparing process:

The air quality records are assembled by having the subsequent attribute between 2016 and 2019 from the pollution control board, Karnataka. This information is a more accurate rendition of the authentic day-to-day encompassing air quality information gathered by the pollution Control board. In this investigation, structure esteem patterns, amount and distribution of pollution, month-to-month midpoints, poison proportions, and relationship coefficients between different toxins have been used to gauge the status of encompassing air quality. Also, hourly midpoints are used for -plots of NO2, SO2, PM2.5, and PM10 [14].

3.2 Site Description:

The Central Pollution Control Board (CPCB) of India is a statutory organization under the Ministry of Environment, Forest, and Climate Change. They developed it in 1974 underneath the Water Act, 1974. The CPCB is besides relied upon with the powers and limits underneath the Air Act, 1981. It fills in as an area of improvement and also provides specific commitments to the Administration of Condition and Boondocks. It coordinates the activities of the State Pollution Control boards by providing specific help and guidance and settles inquiries among them.

Central Pollution Control Board is executing a nation-wide programme of ambient air quality monitoring known as the National Air Quality Monitoring Programme (NAMP). The framework joins 621 working stations covering 262 urban networks/towns in 29 states and 5 Affiliation Spaces of the country. Under N.A.M.P., four air poisons Sulfur Dioxide (SO2), Nitrogen Oxide (NO2), Suspended Particulate Matter (SPM) and Respirable Suspended Particulate matter (RSPM/PM10) has perceived for standard checking at all the territories. The checking of meteorological parameters, for instance, wind speed and wind heading, relative humidity (RH), and the temperature have similarly joined with the observations of air quality. It invigorates this information on Air Quality at ITO every week.

The analysis of the following facts was accomplished for our work as State, Location, Location checking station, No2-Nitrogen Dioxide, So2-Sulfur Dioxide, PM-2.5 and -particulate issue, RSPM (Respirable Suspended Particulate Matter), Date, Agency, and Type.Overview of the task chosen for records assessment:

a) State: State whose air quality information (AQI) to be analyzed.
b) Location: City whose AQI to be estimated.
c) Location_monitoring_station: It shows the region of the observing territory.
d) NO2: The amount of Nitrogen Dioxide estimated. It is an intermediate in the industrial synthesis of nitric acid. It is one of the most hazardous air pollutants.
e) SO2: The amount of Sulfur Dioxide assessed. SO2 is a gaseous air pollutant composed of sulfur and oxygen. Coal and oil normally fuse sulfur blends, and their ignition makes sulfur dioxide.
f) pm2.5: It speaks to the expense of particulate matter estimated.
g) RSPM: Respirable Suspended Particulate Issue. Health effects are primarily caused by a respirable portion of particulate matter which is in the size range of 0.5 to 10 microns.
h) Date: examining the date
i) Agency: Name of the association that deliberate the records.
j) Type: The area where the estimation has been made [9].

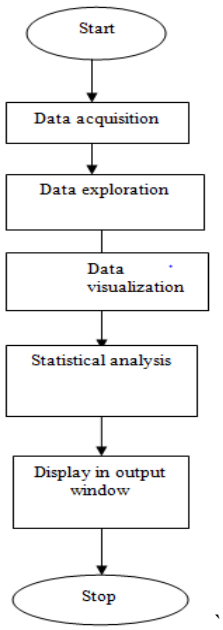


Fig. 1. Flowchart

A. Data acquisition

Information Sources: For this examination, the contaminated information from more than one air pollution monitoring

station has been taken into consideration. The stations in the most extreme dirtied zones were situated. Another explanation behind settling on those stations is to show the intricacy and heterogeneity in estimating the poison patterns. The measurements from the Pollution Control Board site were accumulated [10].

B. Data exploration

Some insights on these records are the wide style of entries in every column is taken. Data is preprocessed to remove the unwanted attributes like the location of the monitoring station, organization, and date.

C. Data visualization

Here, the grouping of so2, no2, rspm, pm-2.5 in One-of-a sort zones, using the bar plots were plotted. The intention is to represent and avoid the speculations to obtain helpful discoveries. The representation system is instinctive and reliable through information quality checking and data imparting to multi-point view air pollution charts. These strategies permit the information without issues by the overall population and move or help further investigations in different fields. By taking care of graphical portrayals of complex records, ordinary plots together with disperse plots which is used to investigate time-arrangement records and to depict the relationship among different factors in air pollution investigation.

D. Statistical analysis

By performing statistical analysis for the dataset to check whether those capabilities have a few relations. Initially, the aid of plotting the scatter plot for each characteristic is an in-depth analysis. It is possible to plot correlation matrix changes of the concentration of air [11].

E. Displaying in output window

Displaying the plots of performed evaluation in the result section.

IV. RESULTS AND DISCUSSION

The gathered information is obtained from pollution control board of the respective state and various other sources and its breakdown is as shown in table 1. Data is collected between the periods of 2016-2019. Here, a fixation for every parameter is manifested. The heat map is obtained by showing the parameters considered during various years. The measurable examination of information is performed to check relations between these highlights.

Table 1. Description of dataset with respect to location and samples

Name of the state	City	Number of samples
Karnataka	Bangalore	495
	Mysore	24
	Mangalore	12
	Mandya	12
	Hasan	12
	Hubli	105
	Dharwad	105
	Gulbarga	79
	Belgaum	68
	Domlur	26
	Davanagere	206
	Ranebennuru	104
	Raichur	60
	Bijapur	68
	Chitradurga	104
	Shimoga	102
	Kolaar	17
	Bidar	25
	Tumkur	85
Tamilnadu	Chennai	36
Maharashtra	Nagpur	66
	Latur	11
Kerala	Kozhikode	11
Gujarat	Baroda	6
	Surat	1
	Ahmadabad	7

4.1 Analyzing data for Karnataka

The data is collected from 715 residential spots and 800 industrial areas across the state. Fig 2 shows the number of entries for residential and industrial and other areas.

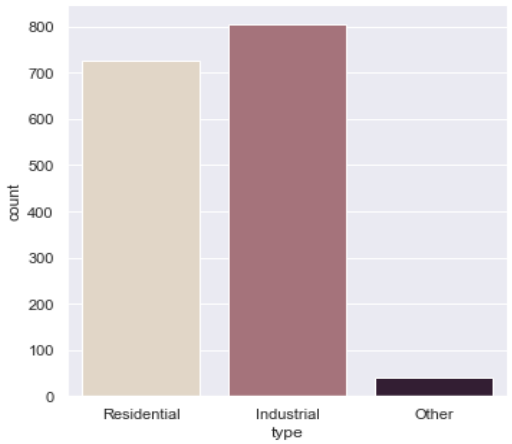


Fig. 2. Data collection from different parts of the state

From the dataset, the given territory is arranged into two classes such as industrial and residential. Here, the information is investigated for Karnataka for various conditions of India.

Based on the figures above, the records were documented at the primary centre near the industrial blog because it has the most comprehensive range of sections.

4.2 Concentration of SO_2 and NO_2

Sulpher dioxide risks lung health and a major cause of chronic asthma in most patients with breathing difficulties. It also affects the habitat suitability of plant species in the environment. This fig. 3 explains approximately the so2 concentration of different Indian states. Nitrogen Dioxide in

brownish color and acidic odor causes bronchitis and immunity-related disorders. Karnataka has an average of 30 ppm NO2 concentration as given in fig. 4 depicting the concentration level of various states. This level causes mild throat irritation to those who inhale.

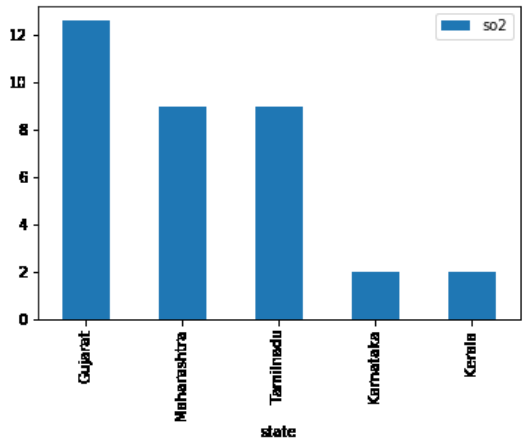


Fig. 3. SO_2 concentration

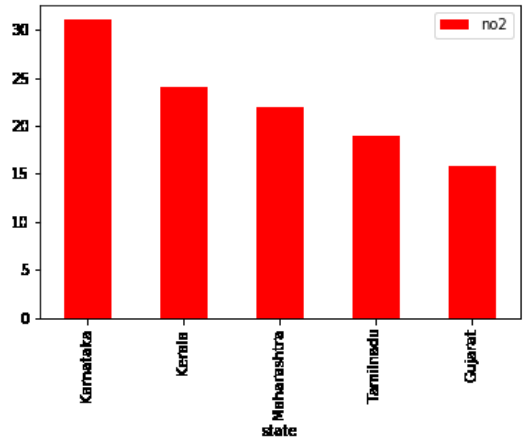


Fig. 4. NO_2 concentration

Cities of Karnataka with different so2 and NO2 concentrations are given in fig. 5. It is estimated that Mangalore has the highest and Belgaum has the lowest so2 concentration. It is also observed that no2 concentration is highest in Bangalore and lowest in Davanagere among the major cities.

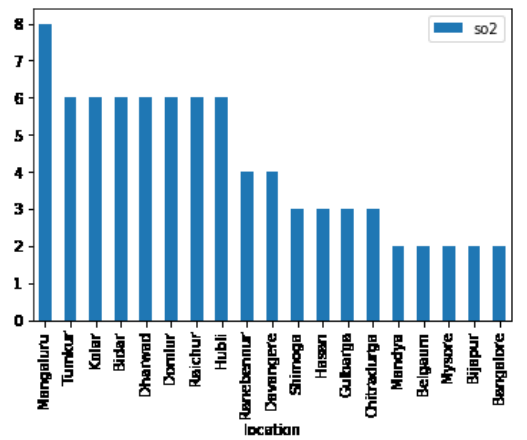


Fig. 5. SO_2 vs. location

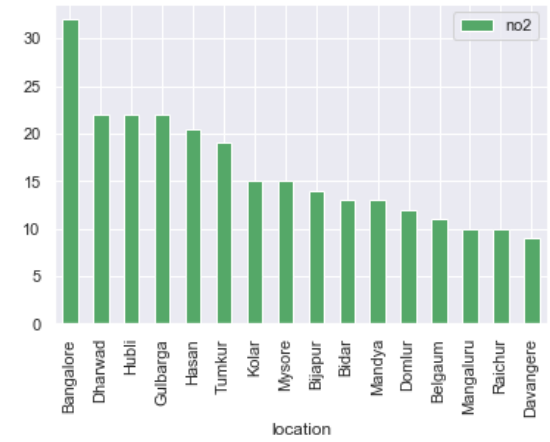
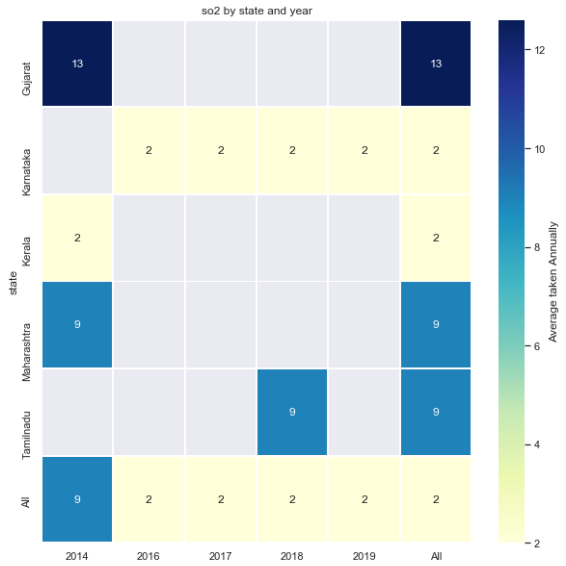
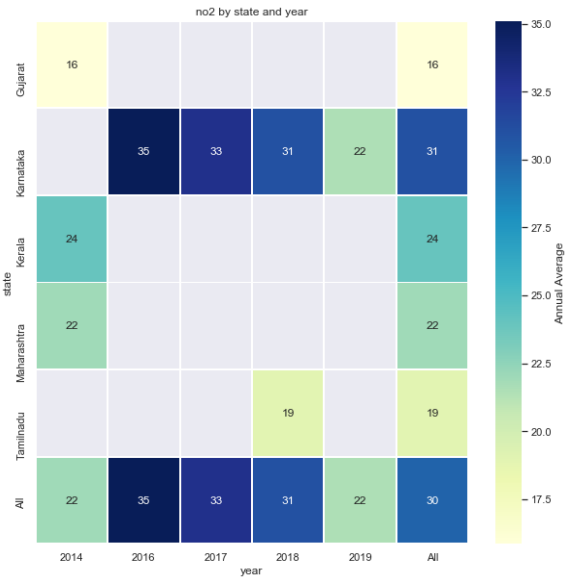


Fig. 6. NO_2 vs. location



(a)



(b)

Fig. 7 (a) and (b). Comparison of SO2 and NO2 levels of various states

The above examination in fig 6 shows that sulphur dioxide has been high in 2014 in some areas of Karnataka yet has diminished in the ongoing years. Compared to the national average, Gujarat, Maharashtra, and Tamilnadu produced the highest sulphur dioxide when compared to the other developed states. Whereas, Karnataka is one among the highest producer of Nitrogen Dioxide consistently throughout estimation. It has observed a decline in NO2 over the past 4 years.

4.3 Rspm and pm2.5 concentrations:

RSPM is a particulate matter with its diameter smaller than 2.5 µms. Its rise is seen in all places in the observed period from 40-120µg/m^3. Being an industrial hub, Tumkur comprises the highest rspm concentration and Chitradurga with the lowest among the major places of Karnataka (fig. 7). Also, Mangaluru and Bengaluru have always been at great risk in cases of PM2.5 levels (fig. 8). Fig 9 shows the comparison of these 2 pollution particulate levels of different states throughout observation between 2015 and 2019. It can be concluded that Karnataka's RSPM level is slightly higher than the national average.

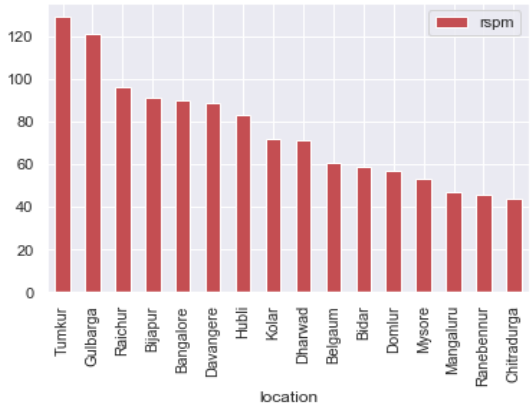


Fig. 8. rspm vs. cities

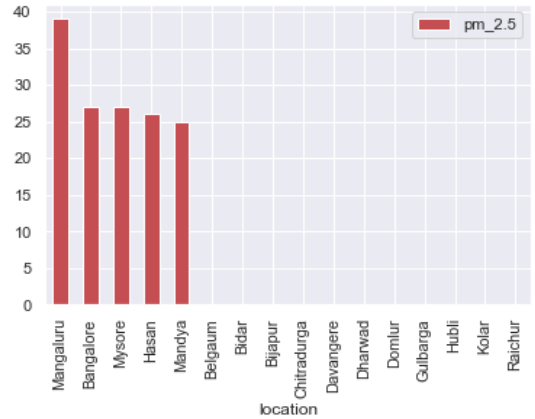


Fig. 9. Pm2.5 vs cities

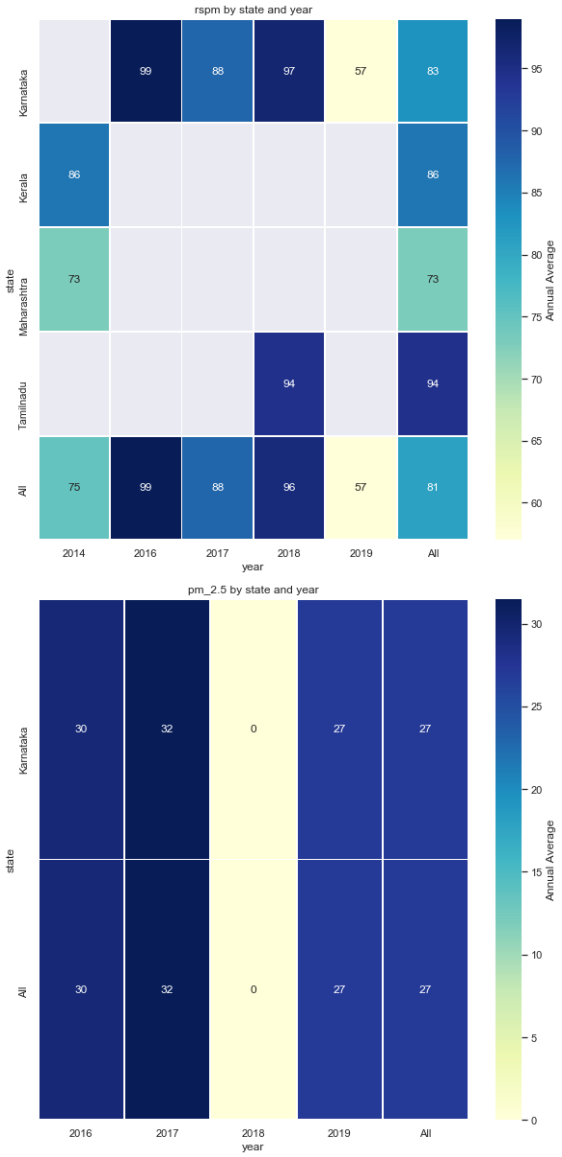


Fig. 10. Comparison of RSPM and PM2.5 levels among various states

Fig 11 helps in plotting the correlation individual each feature. Because of the enormous amount of null value in the PM2.5 attribute of the database, its factual noteworthiness is low. So2 and no2 values are exceptionally close to the origin, which implies that both are low for the vast majority of the observations. It can be observed that no2 and so2 have some degree comparative examples with different highlights.

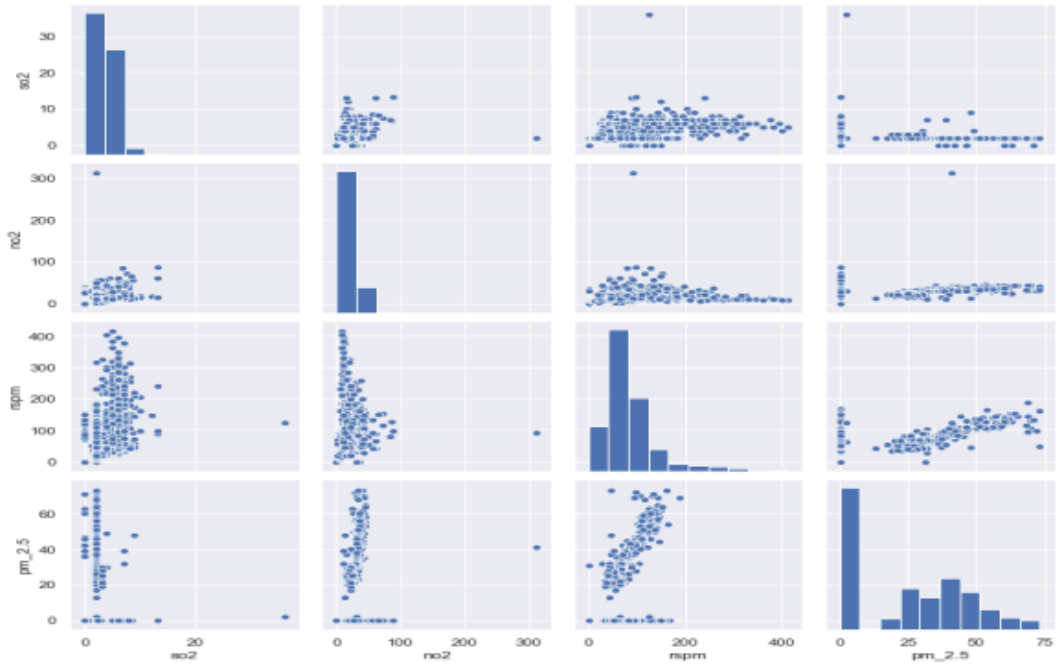


Fig. 11. Disperse plot for every section

V. CONCLUSION

From the above assessment, a nation had a top-level of pollutions, still, there have been a few districts in the states that have not contaminated. From the heat map, it can be concluded that a couple of locales has been highly polluted. The reason for the decline may be mindfulness in occupants and governance arrangements. From the above data investigation approach, the inferred information assessment is a basic angle for a higher future. The strategy changed into absolutely records-driven received upheld by genuine examples. It is intriguing to perceive how records assessment and the everyday occasions are reasonable and how insights assessment is perhaps used to manage large issues.

References

[1] Marinov, Marin B., et al. "Air quality monitoring in urban environments." 2016 39th International Spring Seminar on Electronics Technology (ISSE). IEEE, 2016.

[2] Maag, Balz, Zimu Zhou, and Lothar Thiele. "A survey on sensor calibration in air pollution monitoring deployments." IEEE Internet of Things Journal 5.6 (2018): 4857-4870.

[3] Y. Chen, W. Han, Y. Xiong, W. Wang, and L. Tong, "A New Air Pollution Source Identification Method Based on Remotely Sensed Aerosol and Improved Glowworm Swarm Optimization", IEEE Journal of selected topics in applied earth observations and remote sensing,2017, ISBN: 1939-1404

[4] Gokul, V., and Sitaram Tadepalli. "Implementation of a WiFi based plug and sense device for dedicated air pollution monitoring using IoT." 2016 Online International Conference on Green Engineering and Technologies (IC-GET). IEEE, 2016

[5] Wang, Wei, et al. "Distributed sensor data computing in smart city applications." 2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM). IEEE, 2017.

[6] Li, Huan, Hong Fan, and Feiyue Mao. "A visualization approach to air pollution data exploration—a case study of air quality index (PM2.5) in Beijing, China." Atmosphere 7.3 (2016): 35.

[7] Devarakonda, Srinivas, et al. "Real-time air quality monitoring through mobile sensing in metropolitan areas." Proceedings of the 2nd ACM SIGKDD international workshop on urban computing. 2013

[8] Agarwal, Shivam. "Data mining: data mining concepts and techniques." 2013 International Conference on Machine Intelligence and Research Advancement. IEEE, 2013

[9] Rai, Rajeshwari K. "Analysis of Air Pollution." International Journal of Advance Research, Ideas and Innovations in Technology 3.4 (2017).

[10] Srivastava, Chavi, Shyamli Singh, and Amit Prakash Singh. "Estimation of Air Pollution in Delhi Using Machine Learning Techniques." 2018 International Conference on Computing, Power and Communication Technologies (GUCON). IEEE, 2018

[11] Biswas, Jhumoor, et al. "An analysis of ambient air quality conditions over Delhi, India from 2004 to 2009." Atmospheric and Climate Sciences 1.04 (2011): 214.

[12] Parmar, G., Lakhani, S., & Chattopadhyay, M. K. (2017, October). An IoT based low cost air pollution monitoring system. In 2017 International Conference on Recent Innovations in Signal processing and Embedded Systems (RISE) (pp. 524-528). IEEE.

[13] Zheng, Yu, Furui Liu, and Hsun-Ping Hsieh. "U-air: When urban air quality inference meets big data." Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. 2013.

[14] Matějíček, Luboš, Pavel Engst, and Zbyněk Jaňour. "A GIS-based approach to spatio-temporal analysis of environmental pollution in urban areas: A case study of Prague's environment extended by LIDAR data." Ecological Modelling 199.3 (2006): 261-277.