STUDY OF CORRELATION WITH MAJOR POLLUTANTS IN AIR AND EXCESS CONCENTRATION OF SULPHUR DIOXIDE AND CARBON MONOXIDE IN AHMEDABAD

Shambhavi Kumar, Nidhay Pancholi

Pandit Deendayal Energy University, ICT department

Pandit Deendayal Energy University, ICT department

ABSTRACT

Improving the Air quality is the priority task of many governments around the world. Data analysis and machine learning will help a lot in identifying the potential places where the pollution can be controlled and the steps to be followed. Predicting the AQI can help a lot in planning out things for the future. Also, identifying which pollutants affect a particular region the most can help identify potential sources and steps to follow to bring the concentration of that pollutant under control which indirectly improves the Air Quality Index of the region. This paper tries to find out the pollutants that affect any particular city the most and predicts the AQI using the concentration of that pollutant.

Key Words: AQI, correlation, pollutants,

Linear Regression,

1. INTRODUCTTION

Air quality index has been created to check the quality of air of a country. Air quality indices measure the amount of carbon monoxide, sulphur dioxide, nitrogen dioxide, etc. present in the air. It also shows the types of gases dissolved.

The National Air Quality Index constitute 8 pollutants (PM10, PM2.5, NO2, SO2, CO, O3, NH3 and Pb) and their safe limit has been provided in the WHO Standards Document. If the amount of Sulphur Dioxide (SO2), Nitrogen Dioxide (NO2), and Carbon Monoxide (CO) exceed the national standards then the air is said to be polluted.

PM2.5 and PM10 are particulate matters with diameters of 2.5 micrometres and 10 micrometres respectively. When their limit exceeds the safety standards, people face difficulty in breathing and irritation in the eyes. The burning of fossil fuels in vehicles is the major source of CO emissions. The leaking chimneys, gas stoves and blast furnaces also release CO. SO2 and NOx is the main

constituent of emissions from electric power plants and refineries.

2. LITERATURE REVIEW

Air Quality Analysis and prediction of AQI is one of the most popularly researched topics in current times. The alarming situation of deteriorating air quality has become a major concern. The increasing concentration of obnoxious gases in air is becoming fatal for people. Air quality is estimated on the basis of its index. Air Quality Index constitutes measurement of PM2.5, PM10, NOx, SO2 and CO. These are the major pollutants present in the air. A detailed analysis of the concentration of the pollutants in the air of cities demonstrated a correlation with its Air Quality Index. A lot of research has been done on the AQI of only the major cities of the country like Delhi or Mumbai. These studies show the present conditions of the air quality and often nothing is said about the future precisely. The air quality has been categorised into 6 broad categories: good, satisfactory, moderate, poor, very poor and severe.

AQI Category, Pollutants and Health Breakpoints

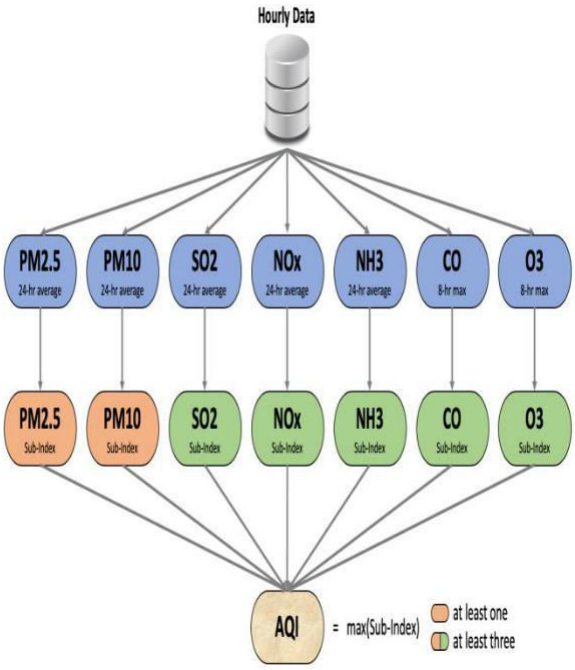
AQI Category (Range)	PM_{10} (24hr)	$PM_{2.5}$ (24hr)	NO_2 (24hr)	O_3 (8hr)	CO (8hr)	SO_2 (24hr)	NH_3 (24hr)
Good (0–50)	0–50	0–30	0–40	0–50	0–1.0	0–40	0–200
Satisfactory (51–100)	51–100	31–60	41–80	51–100	1.1–2.0	41–80	201–400
Moderately polluted (101–200)	101–250	61–90	81–180	101–168	2.1–10	81–380	401–800
Poor (201–300)	251–350	91–120	181–280	169–208	10–17	381–800	801–1200
Very poor (301–400)	351–430	121–250	281–400	209–748	17–34	801–1600	1200–1800
Severe (401–500)	430+	250+	400+	748+	34+	1600+	1800+

As the quality deteriorate so does the ranking. It goes from good to poor and up to severe.

Inhaling pollutants in such high concentrations have become hazardous for living beings.

The air pollution is a major concern as it is affecting the lives of the people. (1) AN article stated that air pollution results in health hazards like respiratory disorder, neuropsychiatric, hematologic, immunologic, cardiovascular and reproduction system as they have the maximum exposure to the pollutant inhaled. These adverse effects should me minimised as soon as possible. But, first there is a need to adopt a standard method to calculate the level of pollution. Thus, Air Quality Index is officially formulated to determine levels of pollutants in the atmosphere.

Analysis and data collection has been done on the pollutants and their concentration in air. This complex data is not enough to convey the quality of air to common people. As a solution to this problem, Air quality index concept was introduced. Air quality index can be defined as a number stating the presence of pollutant levels in atmosphere and its level determining the effect on health (Thom and Ott, 1976; Bortnick et al., 2002; Murena, 2004).The AQI is calculated using the process shown below,



In a research article (2) a prediction of AQI has been done based on the levels of SO2, NOx, PM2.5, PM10, CO, NH3 etc. in the atmosphere. The predictions in this method are done using data of all the pollutants. However the below method works on the foundation that any single pollutant for a respective city tends to have a high amount of correlation with the AQI of the city and hence after finding out that pollutant the AQI of the city can be predicted.

3. PROPOSED METHODOLOGY

Here, the data has been collected from the website of the Central Pollution Control Board (CPCB) of India. The data contains daily information about the concentration of different pollutants and the calculated Air Quality Index (AQI) from 26 Major Indian cities from 2015 until June of 2020. The AQI is calculated using concentration of different pollutants such as Sulphur Dioxide (SO2), Nitrogen Dioxide (NO2), Carbon monoxide (CO), PM2.5, PM10, Ammonia (NH3) and Ozone (O3).

The process followed for the Regression analysis,

- The data contains hourly information about the concentration of pollutants and hence an average of the 24 hour data was found to calculate the concentration of pollutant during the whole day.
- For all 26 Cities in the data the pollutant whose concentration had the maximum linear correlation with the Air quality Index (AQI) was found.
- Using the pollutant which had the maximum correlation with the AQI, a Linear Regression model was trained using cross validation.
- The average of the Negative Mean Absolute Error and the standard Deviation of the 3 Fold cross validation were calculated to identify how the linear models perform.

	City	Pollutant	Correlation	Score	Standard Deviation
0	Ahmedabad	CO	0.841498	-107.752513	20.844226
1	Aizawl	PM2.5	0.949636	-10.453597	5.643016
2	Amaravati	PM2.5	0.911025	-20.134686	1.871572
3	Amritsar	PM10	0.876050	-22.830112	2.893393
4	Bengaluru	PM10	0.599427	-19.020642	6.479643
5	Bhopal	PM2.5	0.898222	-21.434590	3.268817
6	Brajrajnagar	PM2.5	0.891627	-26.113216	1.996033
7	Chandigarh	PM2.5	0.924146	-19.122199	5.514069
8	Chennai	PM2.5	0.685568	-27.905842	6.683187
9	Coimbatore	O3	0.452185	-16.232050	2.824481
10	Delhi	PM10	0.884864	-47.696420	3.911338
11	Ernakulam	CO	0.476778	-18.519461	7.217113
12	Gurugram	PM10	0.861250	-42.905658	7.663343
13	Guwahati	PM10	0.865925	-34.971339	4.183714
14	Hyderabad	PM2.5	0.797735	-23.261708	7.255073
15	Jaipur	PM2.5	0.852730	-19.709042	4.407013
16	Jorapokhar	PM10	0.878490	-35.909749	8.123310
17	Kochi	NOx	0.637799	-23.807238	3.999768
18	Kolkata	PM10	0.973276	-18.922634	3.902815
19	Lucknow	PM2.5	0.918568	-35.898568	3.073689
20	Mumbai	PM2.5	0.936100	-13.649716	0.858295
21	Patna	PM2.5	0.923235	-38.070640	4.612879
22	Shillong	PM2.5	0.847550	-13.172820	0.436641
23	Talcher	PM10	0.917865	-34.519755	3.066102
24	Thiruvananthapuram	PM2.5	0.814297	-11.688229	0.329443
25	Visakhapatnam	PM2.5	0.891224	-18.763763	1.605359

Fig 1- The pollutant with maximum correlation, the correlation, the average of negative mean absolute error and the standard deviation for 3-fold cross validation.

All cities have a pollutant that has a high amount of linear correlation with the AQI and using that the AQI can be predicted with very less amount of error. But in the case of Ahmedabad even though Carbon monoxide has a correlation of 0.84 but still it has an average negative mean absolute error of -107.7 which is pretty high as compared to the other cities.

ANALYSIS

Observations from Fig 1,

- 14 out of the 26 cities have the maximum linear correlation with PM2.5.
- 7 out of the 26 have maximum linear correlation with PM10.
- All cities except Ahmedabad have an average negative mean absolute error in the range of -10 to -50.

On investigating further, the reason behind the large value of error in the Ahmedabad city was found. The reason being,

1. Many outliers in the AQI of Ahmedabad.

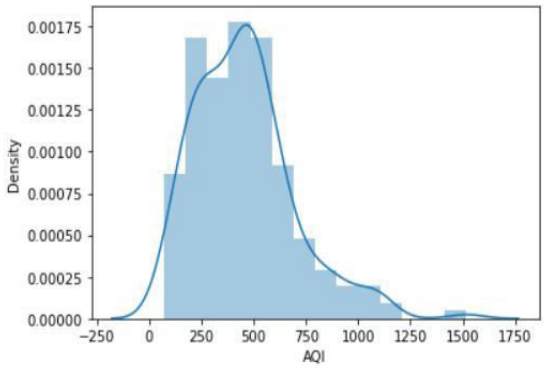


Fig 2 – The histogram for the distribution of Air quality index in Ahmedabad

2. The concentration of Carbon monoxide and Sulphur dioxide is very high as compared the other cities. To identify this relation a weekly average was calculated for all the cities for the year 2019.

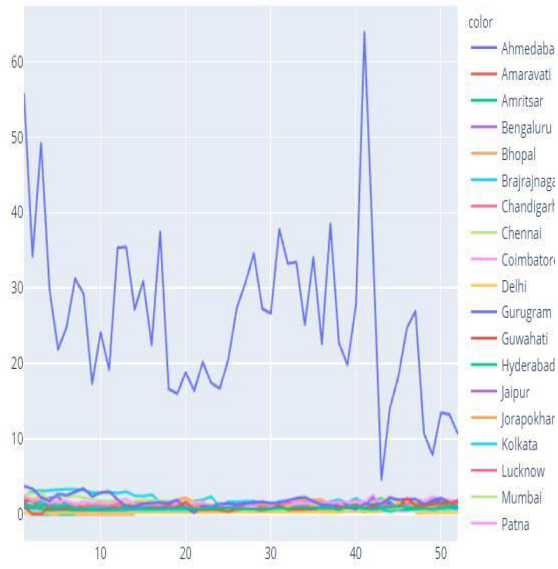


Fig 3- The weekly average of the concentration of Carbon monoxide for the year 2019.

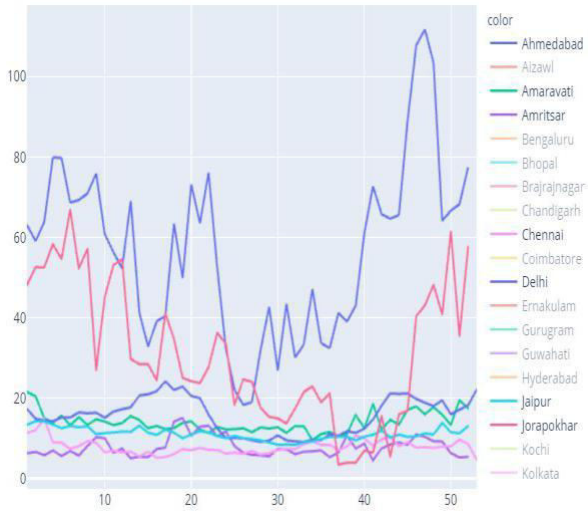


Fig 4- The weekly average of the concentration of Sulphur Dioxide for the year 2019.

FUTURE APPLICATIONS

The above study demonstrates the clear correlation between the AQI and one pollutant that is majorly present in the atmosphere. This could help in precise prediction of the AQI as it shows a linear dependency. Also, the sources of pollution that emit the dominant pollutant can be identified and also the result of the decrease in that pollutant is available in advance which further creates an incentive. This creates a great opportunity for the local authorities as now they have a direct indication of how their work will affect the AQI. The AQI here is ranging from good to very poor, so immediate actions should be taken to curb the growing pollution.

4. CONCLUSION

The linear correlation of pollutants and AQI has been identified. Hence using this AQI of any particular city on any particular day can just be predicted using just one pollutant. Also the city of Ahmedabad has very high concentration of Carbon monoxide and sulphur dioxide which is extremely high compared to any other city in India and if these continue to exceed it could really start to affect people's health seriously.

REFERENCES

1. Effects of air pollution on human health and practical measures for prevention in Iran by Adel Ghorani-Azam, Bamdad Riahi-Zanjani, and Mahdi Balali-Mood.
2. ANALYSIS AND PREDICTION OF AIR QUALITY INDEX IN INDIA DURING PRE AND POST COVID PANDEMIC USING MACHINE

LEARNING ALGORITHMS, 2020
by Dr K B Priya Iyer and Dharshini V.

3. Calculating AQI (Air Quality Index) Tutorial | Kaggle
https://www.kaggle.com/rohanrao/calculating-aqi-air-quality-index-tutorial
4. Air Quality Index – A Comparative Study for Assessing the Status of Air Quality by Shivangi Nigam, B P S Rao, Navneet Kumar, V A Mhaisalkar in May,2016.
At https://www.researchgate.net/
5. Air Quality Index (AQI) in Ahmedabad City, Gujarat, India by Mr. P.V. Upadhyay, March, 2019
6. www.jagranjosh.com// what is air quality index and how is it calculated.
Created on November 6, 2019