Analysis of Air Pollutant Levels in India

Bharani Ujjaini Kempaiah
Computer Science Dept.
PES University
Bangalore, India
ukbharani@gmail.com

Bhavya Charan
Computer Science Dept.
PES University
Bangalore, India
bhavya.charan.edu@gmail.com

Ruben John Mampilli
Computer Science Dept.
PES University
Bangalore, India
rubenjohn1999@gmail.com

***Abstract*— Air pollution, one of the most serious problems in the world, is a major environmental risk to health. WHO estimated that every year, air pollution causes the premature death of around 7 million people worldwide. Through this project, we try to focus on analysing the air quality trends in India, over the span of 2016 - 2018. Concentrations of some of the major air pollutants such as SO_2, NO_2, CO, PM_{10}, $PM_{2.5}$ and O_3 are used to arrive at conclusions about the air quality. This project aims at identifying various factors that could contribute to the changing air quality over time.**

Keywords— Air Pollution, Pollutants, Air Quality index

I. Introduction

Air pollution in India is emerging as a major factor that seems to contribute to many of the health hazards that people are facing. Air pollution occurs when harmful or excessive quantities of substances including gases, particles, and biological molecules are introduced into the Earth's atmosphere. With the increasing number of vehicles used in metropolitan cities, the quality of air we breathe every second is deteriorating and is playing a harmful role in the overall balance of the ecosystem. While pollution levels across some of the most polluted cities in India have improved in the past couple of years, several new pollution hotspots have emerged. Thus, through this project we attempt to look into the pollution levels at various regions across India and try to spot areas where the trends in pollution levels are more alarming and require attention.

Air pollution is caused by the release of various pollutants. Some of the major pollutants and their sources are as follows:

1) Sulphur dioxide (SO_2) - A gas produced from burning coal and also some industrial processes, such as smelting of metals. It is a major contributor to smog and acid rain. Sulfur dioxide can cause lung diseases.

2) Ozone (O_3) - At the ground level, it is a pollutant with highly toxic effects. Vehicles and industries are the major sources. Ozone makes our eyes itch, burn, and water. It lowers our resistance to colds and pneumonia.

3) Nitrogen oxide (No_x) - Causes smog and acid rain. It is produced from burning fuels. Nitrogen oxides can make children susceptible to respiratory diseases in winters.

4) Suspended particulate matter (SPM) - Consists of solids in the air in the form of smoke, dust, and vapour that can remain suspended for extended periods. The finer of these particles, when breathed in can lodge into our lungs and cause lung damage and respiratory problems.

5) Carbon monoxide (CO) - Gas produced by the incomplete burning of carbon-based fuels or combustion of natural and synthetic products such as cigarettes. It lowers the amount of oxygen that enters our blood.

Air Quality Index : The large monitoring data results in information that don't give a clear picture about how good or bad the air is. Thus the concept of an Air Quality Index (AQI) has been developed. An AQI is defined as an overall scheme that transforms weighted values of individual air pollution related parameters (SO_2, CO, visibility, etc.) into a single number or set of numbers. The Sub-indices for individual pollutants are calculated using its 24-hourly average concentration value (8-hourly in case of CO and O_3)

$$Index = \frac{\text{Pollutant Concentration}}{\text{Pollutant Standard Level}} \times 100$$

The worst sub-index is the AQI for that location. The National Ambient Air Quality Standards of India are shown.

Pollutant	Time Weighted Average	Concentration in Ambient Air	
		Industrial, Residential, Rural and Other Areas	Ecologically Sensitive Area
Sulphur Dioxide (SO_2), $\mu g/m^3$	Annual* 24 hours**	50 80	20 80
Nitrogen Dioxide (NO_2), $\mu g/m^3$	Annual* 24 hours**	40 80	30 80
Particulate Matter (size less than 10 µm) or PM_{10} $\mu g/m^3$	Annual* 24 hours**	60 100	60 100
Particulate Matter (size less than 2.5 µm) or $PM_{2.5}$ $\mu g/m^3$	Annual* 24 hours**	40 60	40 60
Ozone (O_3) $\mu g/m^3$	8 hours* 1 hour**	100 180	100 180
Lead (Pb) $\mu g/m^3$	Annual* 24 hours**	0.50 1.0	0.50 1.0
Carbon Monoxide (CO) mg/m^3	8 hours* 1 hour**	02 04	02 04

The different AQI categories are - Good (0–50), Satisfactory (51–100), Moderately polluted (101–200), Poor (201–300), Very Poor (301–400), Severe (401-500)

II. PRIOR WORK

A. Approach

The Greenpeace India organisation has performed analysis regarding pollution levels throughout the country[1]. The published report states and supports the fact that air pollution is not an issue just in the National Capital Region of New Delhi but it is a national problem that is costing the economy an estimated 3% of GDP. They worked on data collected from the SPCB and examined the annual average PM_{10}. Their work consisted of focus on specific cities in the states of Andhra Pradesh, Bihar, Chandigarh, Chhattisgarh, Gujarat, Haryana, Jharkhand, Karnataka, Madhya Pradesh, Maharashtra, Odisha, Punjab, Rajasthan, TamilNadu, Telangana, Uttar Pradesh and Uttarakhand.

B. Results

- There are no cities in northern India complying with WHO and NAAQS standards for PM_{10} concentrations, and most of the cities are critically polluted.
- Except for a few places in Southern India which compiled with the NAAQ standard, the entire country is facing a public health crisis due to high air pollution levels.
- The report also used the study conducted by Chaudhary et al[2] which carried out source apportionment study to identify the biggest contributors to the overall particulate matter concentrations.

A. Approach

This study focuses on the air quality in the Delhi region, conducting a detailed analysis from 2009-2017 [7]. Descriptive analysis and predictive analysis has been used to study the trends of various pollutants like SO_2, NO_2, Particulate matter, O_3, CO and benzene and predict the future trends. The data has been collected from CPCB. Collected data has been pre processed using steps like parsing of dates, noise removal, cleaning, training and scaling. Further, descriptive analysis has been carried out on two different platforms- Rstudio and Tableau. For observing the forecasted results, predictive analysis has been done.

B. Results

- Descriptive analysis shows drastic increase in PM_{10} level. NO_2 and $PM_{2.5}$ have evidently increased.
- Carbon monoxide is predicted to reduce by 0.169 mg/m^3.
- The trend of nitrogen dioxide (NO_2) in air is predicted to drastically increase for the coming years by 16.77μg/m^3.
- A gradual rise in NO_2, often breaching the safety standard is expected. Ozone (O_3) is estimated to increase by 6.11 μg/m^3 in the upcoming years. Benzene trend and the predicted trend seems to reduce by 1.33 mg/m^3 .
- The observed result indicates that pollutants like NO_x, PM_{10}, $PM_{2.5}$ are likely to drastically increase in the future, while SO_2 levels may increase marginally in future. O_3 levels will increase in initial years. Though the amount of CO and Benzene are showing reducing trend.

A. Approach

The National Morbidity, Mortality and Air Pollution Study (NMMAPS) characterizes the effects of airborne particles (PM_{10}) alone and in combination with gaseous air pollutants, in a large number of cities [8]. Multiple locations were selected based on the specific criteria of population size and availability of PM_{10} data from the US Environmental Protection Agency's Aerometric Information Retrieval System (AIRS) database. And a time-series study of mortality effects was conducted. The morbidity analysis also used a unified analytic method to examine the association of PM_{10} with hospitalization of those 65 years of age or older.

B. Results

- The results of both the 20 cities and 90 cities analyses are generally consistent with an average approximate 0.5% increase in overall mortality for every 10 mg/m^3 increase in PM_{10} measured the day before death.
- The results of the morbidity analysis were consistent with an approximate 1% increase in admissions for cardiovascular disease and about a 2% increase in admissions for pneumonia and chronic obstructive pulmonary disease for each 10 mg/m^3 increase in PM_{10}.

A. Approach

Five years data on CO, NO, NO_2, O_3, smoke and SO_2 concentrations recorded at one air - pollution monitoring station in the city of Athens were analysed using principal component analysis. Separate analyses were undertaken for summer and winter periods. PCA was also applied to meteorological data concerning relative humidity, temperature, sunshine duration, wind velocity and wind direction.

B. Results

It was found that the main principal components extracted from the air pollution data were related to gasoline combustion, oil combustion and ozone interactions. The most prominent principal components from the meteorological data were related to dry conditions (summer period) and high-speed southwestern winds (for both periods). Finally, canonical correlation analysis determined relationships between the two different datasets. The main relationship was between total pollution and high humidity in combination with the low-velocity wind.

A. Approach

One project focuses on conducting a time series analysis of air pollution levels which involves identification of long-term variation in mean and cyclical components [6]. The model is based on a step-wise approach to time series analysis applied to the daily average concentrations of strong acid and black smoke. The results show that the trends are mostly random and the application of a sequential decomposition model to the air quality data is an adequate method for identifying periodicities.

A. Approach

Another project aims to investigate various big-data and machine learning based techniques for air quality forecasting in diverse conditions.
The big data analytics approach for studying, evaluating, and predicting air quality is made possible due to the availability of environmental sensing networks and sensor data.
This paper reviews the published research results relating to air quality evaluation and prediction using methods of artificial intelligence, decision trees, support vector machines, deep learning etc.. Moreover, the paper classifies and compares the applied big data analytics approaches and big data based prediction models for air quality assessment. Furthermore, the paper also throws light on some of the challenging areas and future research needs.

III. PROBLEM STATEMENT

Analysing the trends in pollutant levels across India for a duration of 3 years from 2015-18 using suitable visualisations and features extracted from the data from Open Air Quality data measured sensors installed across numerous cities in India.

A. Assumptions

In this project we assume the data is representative of the situation in the nation from 2015-2018. There can be missing data during a particular day but we will be focused on aggregating monthly data and thus we assume there will be no striking impact if some data is missing. We are working with a limited amount of data of the recent past which cannot be used to generalise the overall trend in pollution levels. The insights gained from our work will definitely shine a light on the current pollutant levels.

B. Data

- We are currently working with a subset[3] of the data collected by OpenAQ[4]
- OpenAQ is an open source air quality project that is answering the question “What would happen if all the world’s air quality data is made available for the public to explore”.
- Our dataset consists of 6363284 rows and 11 columns.

location	city	country	utc	local	parameter	value	latitude	longitude	attribution
<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<dbl>	<dbl>	<dbl>	<fct>
iplomatic Post: Hyderabad	Hyderabad	IN	2016-01-03T18:30:00.000Z	2016-01-04T00:00:00+05:30	pm25	61	17.44346	78.47489	[{"name":"EPA AirNow DOS","url":"http://airnow.gov/index.cfm?action=airnow.global_summary"}]
iplomatic Post: Chennai	Chennai	IN	2016-01-03T18:30:00.000Z	2016-01-04T00:00:00+05:30	pm25	28	13.05237	80.25193	[{"name":"EPA AirNow DOS","url":"http://airnow.gov/index.cfm?action=airnow.global_summary"}]
iplomatic Post: Mumbai	Mumbai	IN	2016-01-03T18:30:00.000Z	2016-01-04T00:00:00+05:30	pm25	127	19.06602	72.86870	[{"name":"EPA AirNow DOS","url":"http://airnow.gov/index.cfm?action=airnow.global_summary"}]
iplomatic Post: Kolkata	Kolkata	IN	2016-01-03T18:30:00.000Z	2016-01-04T00:00:00+05:30	pm25	337	22.54714	88.35105	[{"name":"EPA AirNow DOS","url":"http://airnow.gov/index.cfm?action=airnow.global_summary"}]
iplomatic Post: New Delhi	Delhi	IN	2016-01-03T18:30:00.000Z	2016-01-04T00:00:00+05:30	pm25	374	28.59810	77.18907	[{"name":"EPA AirNow DOS","url":"http://airnow.gov/index.cfm?action=airnow.global_summary"}]
iplomatic Post: Chennai	Chennai	IN	2016-01-03T19:30:00.000Z	2016-01-04T01:00:00+05:30	pm25	32	13.05237	80.25193	[{"name":"EPA AirNow DOS","url":"http://airnow.gov/index.cfm?

Fig : Snapshot of our dataset

C. Approach

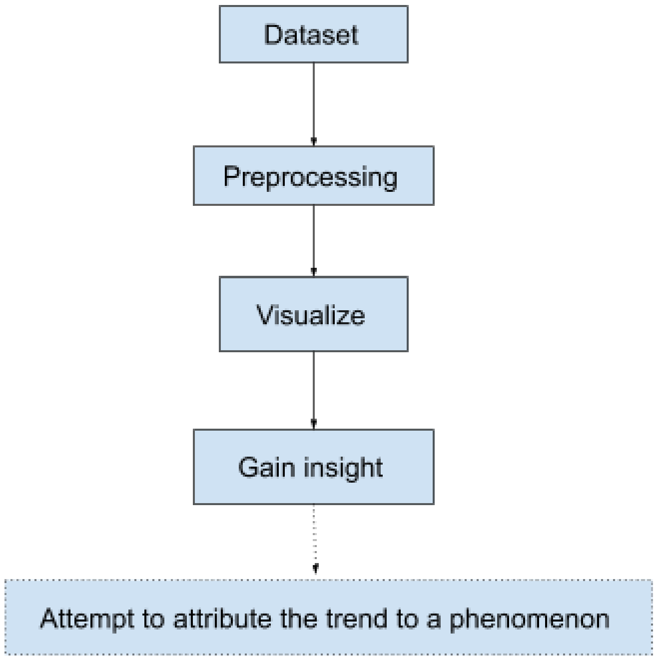

Fig : Basic Workflow representation

The above figure represents the approach we will be taking to meet our goals. It involves extensive cycles of the above workflow.

We will try to support any changes observed with suitable data obtained from other resources as we proceed. We will be attempting to observe whether changes in pollutant levels are accompanied by changes in life expectancy, mortality rates, number of vehicles sold, weather and temperature attributes.

Though extensive amount of work has been done in this area, it has been localised to particular regions with a lot of work being done on the India capital city New Delhi because of its toxic air quality which is stated by a survey carried out by the WHO as one of the worst in any major city across the world.

ABBREVIATIONS USED

SPCB State Pollution Control Board
CPCB State Pollution Control Board
OpenAQ Open Air Quality
NO_2 Nitrogen dioxide
SO_2 Sulphur dioxide
CO Carbon monoxide
O_3 Ozone
PM_{10} Particulate matter having diameter <= 10 micrometers
$PM_{2.5}$ Particulate matter having diameter <= 2.5 micrometers
WHO World Health organisation
PCA Principal Component Analysis
NAAQS National Ambient Air Quality Standards

References

[1] Sunil Dahiya, Lauri Myllyvirta and Nandikesh Sivalingam, GreenPeace, India.
https://secured-static.greenpeace.org/india/Global/india/Airpoclypse--Not-just-Delhi--Air-in-most-Indian-cities-hazardous--Greenpeace-report.pdf

[2] Chowdhury, Zohir; Zheng, Mei and Russell, Armistead, 2004, "Source Apportionment and Characterization of Ambient Fine Particles in Delhi , Mumbai , Kolkata , and Chandigarh" Georgia Institute of Technology, Atlanta Georgia, https://smartech. gatech.edu/bitstream/ handle/1853/10872/E20-H76_736587.pdf

[3] Kaggle link to our dataset *https://www.kaggle.com/ruben99/air-pollution-dataset-india20162018*

[4] OpenAQ dataset *https://openaq-data.s3.amazonaws.com/index.html*

[5] Guttikunda, S.K. and P. Jawahar, 2014. "Characterizing Patna's Ambient Air Quality and Assessing Opportunities for Policy Intervention", UrbanEmissions.Info (Ed.), New Delhi, India, http:// shaktifoundation.in/wpcontent/uploads/2014/02/ AQM-in-Patna-2014-07-15- Final-Report.pdf

[6] Romualdo L. Salcedo, Monaliza Antunes Ferraz, Célia Alves, Fernando Gomes Martins. https://www.researchgate.net/publication/248342740_Time-series_analysis_of_air_pollution_data

[7] Nidhi Sharma, Shweta Taneja, Vaishali Sagar, Arshita Bhatt https://www.sciencedirect.com/science/article/pii/S1877050918307555

[8] Jonathan M Samet, Scott L Zeger, Francesca Dominici, Frank Curriero, Ivan Coursac, Douglas W Dockery, Joel Schwartz, and Antonella Zanobetti https://pdfs.semanticscholar.org/cbc8/77035ce570ba12affe0916b1a06a387d1224.pdf

[9] Impact of Regional climate change and future emission scenarios on surface O_3 and $PM_{2.5}$ over India https://www.atmos-chem-phys.net/18/103/2018/acp-18-103-2018.pdf