IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Analysis of Air Quality Index of Major Cities of India

[1]Devesh Raj, [2]Dr. Savita Shiwani

[1]Master of Technology, [2]Associat Professor
Department of Computer Science & Engineering
Jaipur National University, jaipur , India

Abstract: Air Pollution is a complex mixture of gases, particles, aerosols, water vapor which has originated due to human development and other natural/anthropogenic activities. Its close relation to human development, complex structure containing infinite proportions of particles and gaseous matrix makes it more challenging towards its management. Air pollution management is lie at the interface of science and public policy. These decisions involve a number of stakeholders with competing agendas and vested interests in the ultimate decision .It is then appropriate to adopt formal methods for consensus building to ensure transparent and repeatable decisions. In this paper, different method for estimating the Air Quality Index is evaluated as a tool for assessing the impact of air pollution with a case study. Air Quality Index (AQI) is such an indicator tool which is widely used worldwide and in India since last 2-3 decades. Essentially it is used for assessing the air pollution hot spots in the region for delineating management and concrete actions.

Index Terms* - *Aerosols, Air Quality Index, Air pollution management

I. INTRODUCTION

Air is one of the most essential natural resources for the existence and survival of the entire life on this planet. All forms of life including plants and animals depend on air for their basic survival. Thus, all living organisms need good quality of air which is free of harmful gases to continue their life. According to the world's worst polluted places by Blacksmith Institute in 2008 [1], two of the worst pollution problems in the world are urban air quality and indoor air pollution. The increasing population, its automobiles and industries are polluting all the air at an alarming rate. Air pollution can cause long-term and short-term health effects. It's found that the elderly and young children are more affected by air pollution. Short-term health effects include eye, nose, and throat irritation, headaches, allergic reactions, and upper respiratory infections. Some long-term health effects are lung cancer, brain damage, liver damage, kidney damage, heart disease, and respiratory disease. It also contributes to the depletion of the ozone layer, which protects the Earth from sun's UV rays. Another negative effect of air pollution is the formation of acid rain, which harms trees, soils, rivers, and wildlife. Some of the other environmental effects of air pollution are haze, eutrophication, and global climate change. Hence, air pollution is one of the most alarming concerns for us today. Addressing this concern, in the past decades, many researchers have spent lots of time on studying and developing different models and methods in air quality analysis and evaluation. Air quality evaluation has been conducted using conventional approaches in all these years. These approaches involve manual collection and assessment of raw data. The traditional approaches for air quality prediction use mathematical and statistical techniques. In these techniques, initially a physical model is designed and data is coded with mathematical equations. But such methods suffer from disadvantages like: 1) they provide limited accuracy as they are unable to predict the extreme points i.e. the pollution maximum and minimum 2) cut-offs cannot be determined using such approach 3) They use inefficient approach for better output prediction 4) the existence of complex mathematical calculations 5) equal treatment to the old data and new data But with the advancement in technology and research, alternatives to traditional methods have been proposed which use big-data and machine learning approaches. In recent times, many researchers have developed or used big data analytics models and machine learning based models to conduct air quality evaluation to achieve better accuracy in evaluation and perdition. This paper is written based on our recent literature survey and study on the existing publications which focused on air quality evaluation and prediction using these approaches. The major objective is to provide a snapshot of the vast research work and useful review on the current state-of-the-art on applicable big data approaches and machine learning techniques for air quality evaluation and predication. As a survey paper on air quality evaluation, it begins with a general introduction to the concerns with air quality, causes and effects of air pollution. In Section II, it presents an understanding on air quality evaluation standards and their need. Section III reviews and compares big data analytics models and research work on air quality evaluation. Section IV covers and compares the machine learning models and research work for air quality evaluation. Section V discusses the future research needs and directions in big data and machine learning based air quality evaluation and prediction, and conclusion Ease of Use

II AREA OF STUDY

Various AQI used in the country as well world over include synergistic effect estimation based on mean of the ratios of pollutant over guideline levels for a certain time period. These can further be classified as AQIs using various mean values viz., geometric, arithmetic mean, weighted average, logarithmic mean and break point concentration. Air Quality Index is the simplest Research J. Engineering and Tech. 6(2): April-June, 2015 2 and widely used measure of measure of overall air pollution of a region. More recently the breakpoint concentration method of measurement of AQI was proposed by CPCB which is for individual pollutants AQI estimation followed by max of these as synergistic level of AQI which may be used for decision making. This was also adopted by China and is USEPA concept of break point concentration level which they have adopted since last decade for their development. The pollutant with the highest AQI value determines the overall AQI for that hour. The four pollutants measured for the AQI are good indicators of daily air quality, but are not the only air pollutants which may cause health effects, such as air toxics pollutants. Additionally, the AQI does not account for temperature or pollen levels, which may increase sensitivity to air pollutants.

The environmental situation in Delhi over concerns of deteriorating conditions. It was estimated that about 3000 metric tons of air pollutants were emitted every day in Delhi, with a major contribution from vehicular pollution (67%), followed by coal-based thermal power plants (12%). Significant worries for human wellbeing from presentation to PM10incorporate consequences for breathing and respiratory frameworks, harm to lung tissue, cancer and premature death. Elderly persons, children and people with chronic lung disease, influenza or asthma are especially sensitive to the effects of particulate matter.

New Delhi is a busy metro politan city,the capital of India.InMay2014,ParticleMatter(PM) of size less than 2.5 micrometers in diameter measured concentrations was found to be greater than 350 micrograms per cubic meter of air in the city of New Delhi, making the city the most polluted city in the world as per WHO.

This enhanced concentration of PM of various sizes in atmosphere is due to increasing automobile exhaust and increase of coal fueled factories in the cities. According to WHO report, Delhi is the worst polluted city in the world and the major source of the particulate matter i.e. solid and liquid particles of diameter less than 2.5 micrometer are the smoke coming out from industries in the city.

2.1 ANALYSIS AND STUDY OF DATA

Ozone: Ozone is a gas found in the air we breathe .Several people who are active outdoor are sensitive to ozone as ozone levels are high outdoors. When you do physical activity, it causes faster and deep breathing which draws more ozone into the body. Good ozone is present naturally in the Earth's upper atmosphere. Bad ozone forms near the ground when pollutants react chemically in sunlight

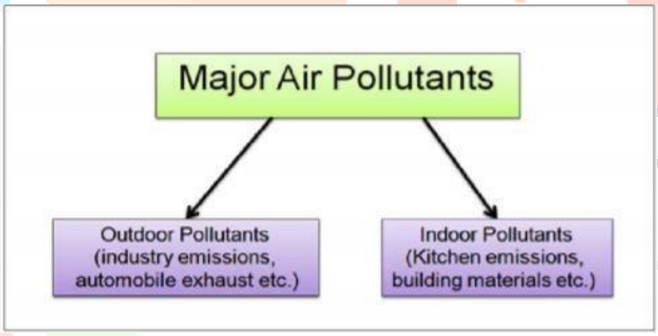

FIGURE.2.1 : Types of air pollutants

Particulate Pollution: Particle pollution (also known as "particulate matter") consists of a mixture of solids and liquid droplets. Some particles are emitted directly; others form when pollutants emitted by various sources react in the atmosphere. People with heart or lung disease are more sensitive to particulate pollution. Particulate pollution comes in 2 sizes: Fine particles (those 2.5 micrometers or less in diameter) and Coarse particles (particles between 2.5 and 10 micrometers).

Sulfur Dioxide: Sulfur dioxide, a colorless, reactive gas, is produced when sulfur- containing fuels such as coal and oil are burned .Major source so sulfur dioxide include power plants ,refineries ,and industrial boilers .It is more commonly found near industrial complexes.

Carbon Monoxide: Carbon monoxide is an odorless, colorless gas. It forms when the carbon in fuels does not completely burn. The major sources of carbon monoxide are vehicles exhaust, industrial processes, natural sources such as wildfires. Carbon monoxide can enter into the body through lungs and binds to hemoglobin. It is a substance in blood that carries oxygen to cells. In the body ,it (carbon monoxide) reduces the amount of oxygen reaching the body's organs and tissues.

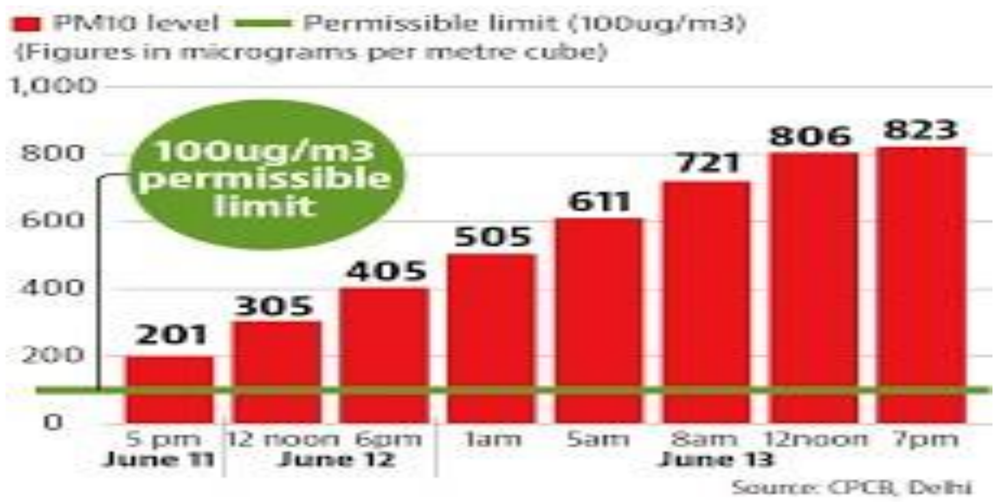

FIG. 2.2: PM (LIMITS OF DELHI)

AQI Category (Range)	PM_{10} (24hr)	$PM_{2.5}$ (24hr)	NO_2 (24hr)	O_3 (8hr)	CO (8hr)	SO_2 (24hr)	NH_3 (24hr)	Pb (24hr)
Good (0-50)	0-50	0-30	0-40	0-50	0-1.0	0-40	0-200	0-0.5
Satisfactory (51-100)	51-100	31-60	41-80	51-100	1.1-2.0	41-80	201-400	0.5-1.0
Moderately polluted (101-200)	101-250	61-90	81-180	101-168	2.1-10	81-380	401-800	1.1-2.0
Poor (201-300)	251-350	91-120	181-280	169-208	10-17	381-800	801-1200	2.1-3.0
Very poor (301-400)	351-430	121-250	281-400	209-748	17-34	801-1600	1200-1800	3.1-3.5
Severe (401-500)	430+	250+	400+	748+	34+	1600+	1800+	3.5+

Fig. 2.3: Air Quality and Break Points

2.2 AIR QUALITY EVALUATION

Air quality evaluation is an important way to monitor and control air pollution. The characteristics of air supply affect its suitability for a specific use. A few air pollutants, called criteria air pollutants, are common throughout the United States. These pollutants can injure health, harm the environment and cause property damage. The current criteria pollutants are: 1) Carbon Monoxide (CO) 2) Lead (Pb) 3) Nitrogen Dioxide (NO2) 4) Ozone (O3) 5) Particulate matter (PM) 6) Sulfur Dioxide (SO2). The Air Quality System (AQS) contains ambient air pollution data collected by EPA, state, local, and tribal air pollution control agencies from over thousands of monitors. AQS also contains meteorological data, descriptive information about each monitoring station (including its geographic location and its operator), and data quality assurance/quality control information. AQS data is used to assess air quality, assist in Attainment/Non-Attainment designations, evaluate State Implementation Plans for Non-Attainment Areas, perform modeling for permit review analysis, and other air quality management functions. AQS information is also used to prepare reports for Congress as mandated by the Clean Air Act A. Air Quality Shtandards Office of air quality planning and standards (OAQPS) manages EPA programs to improve air quality in areas where the current quality is unacceptable and to prevent deterioration in areas where the air is relatively free of contamination. To accomplish this task, OAQPS establishes the National Ambient Air Quality Standard (NAAQS) for each of the criteria pollutants. There are two types of standards - primary and secondary. 1) Primary standards: They protect against adverse health effects; 2) Secondary standards: They protect against welfare effects, such as damage to farm crops and vegetation and damage to buildings. Because different pollutants have different effects, the NAAQS standards are also different. Some pollutants have standards for both long-term and short-term averaging times. The short-term standards are designed to protect against acute or short-term health effects, while the long-term standards were established to protect against chronic health effects. Because different pollutants have different effects, the NAAQS [3] standards are also different and some of them are shown in Table I. Some pollutants have standards for both long-term and short-term averaging times. The short-term standards are designed to protect against acute, or short-term, health effects, while the long-term standards were established to protect against chronic health effects. Modeling of atmospheric pollution phenomena till now has been based mainly on dispersion models that provide approximation of the complex physicochemical processes involved. While the sophistication and complexity of these models have increased over the years, use of these techniques in the frame of real-time atmospheric pollution monitoring seems not totally suitable in terms of performance, input data requirements and compliance with the time constraints of the problem. Instead, human experts' knowledge has been primarily applied in Air Quality Operational Centers for the real-time decisions required, while mathematical models have been used mostly for off-line studies of the phenomena involved. As per them, air pollution phenomena have been measured by using physical reality as the start point. And then, for example, these data traditionally have been coded into differential equations. However, these kinds of techniques have limited accuracy due to their inability to predict extreme events.

We have one important parameter called air quality index (AQI) which quantifies air quality in a region as shown in Table II. It is a number used by government agencies to communicate to the public how polluted the air is currently or how polluted it is forecasted to become. As the AQI increases, an increasingly large percentage of the population is likely to be exposed, and people might experience increasingly severe health effects. Different countries have their own air quality indices, corresponding to different national air quality standards.

Computation of the AQI requires an air pollutant concentration over a specified averaging period, obtained from an air monitor. Taken together, concentration and time represent the dose of the air pollutant. Health effects corresponding to a given dose are established by epidemiological research. Air pollutants vary in potency, and the function used to convert from air pollutant concentration to AQI varies by pollutant. Its air quality index values are typically grouped into ranges. Each range is assigned a descriptor, a color code, and a standardized public health advisory.

The AQI can increase due to an increase of air emissions (for example, during rush hour traffic or when there is an upwind forest fire) or from a lack of dilution of air pollutants. On a day when the AQI is predicted to be elevated due to fine particle pollution, an agency or public health organization might:

- advise sensitive groups, such as the elderly, children, and those with respiratory or cardiovascular problems, to avoid outdoor exertion.
- declare an "action day" to encourage voluntary measures to reduce air emissions, such as using public transportation.
- recommend the use of masks to keep fine particles from entering the lungs

During a period of very poor air quality, when the AQI indicates that acute exposure may cause significant harm to the public health, agencies may invoke emergency plans that allow them to order major emitters (such as coal burning industries) to curtail emissions until the hazardous conditions abate.

Most air contaminants do not have an associated AQI. Many countries monitor ground-level ozone, sulphur dioxide, carbon monoxide, and nitrogen dioxide and calculate air quality indices for these pollutants.

The definition of the AQI in a particular nation reflects the discourse surrounding the development of national air quality standards in that nation. Website allowing government agencies anywhere in the world to submit their real-time air monitoring data for display using a common definition of the air quality index has recently become available

POLLUTANTS	SOURCES
SULPHUR DIOXIDE(SO2)/SULPHUR OXIDES(SOX)	Power plants, sulphuricacid manufacture, boilers,ore refining, petroleum refining.
SUSPENDED PARTICULATE MATTER,SPM(from sulphates and nitrates)	Fine particles which are added either man-made or naturally. Automobile, power plants, boilers, Industries requiring crushing and grinding such as quarry, cement.
Lead	Naturally occurring, produced by lead smelters, contained in old paints and plumbing. Also in ore refining, battery manufacturing and automobiles.
Chlorine	Chlor-alkali plants, manufacturer of polyvinyl chloride(PVC)resins, bleaching powder and many other chemicals.
Fluorides	Fertilizer,aluminum refining, nuclear industry,steelindustry, oil refineries
Oxides of Nitrogen NO,NO2,NOx	Automobiles, power plants, nitric acid manufacture,
Peroxyacetyl nitrate, PAN	Secondary pollutant
Persistent organic pollutants(POP's)	Produced through industrial processes and waste incineration.
Formaldehyde	Secondary pollutant
Ozone	Secondary pollutant, formed from chemical reaction during sunlight.
Carbon Monoxide	Automobiles,fromcombustion processes low in oxygen, burning wood, coal, fuel (cars).
Hydrogen Sulphide	Pulp and paper, petroleum refining
Hydrocarbons	Automobiles,petroleum refining
Ammonia	Used to fertilize crops and emitted from this agricultural process and farm animals.
Carbon dioxide	From volcanic activity and hot springs, combustion processes, cars and plants.

TABLE 2.1: AIR POLLUTANTS AND THEIR SOURCES

2.3 EFFECTS OF AIR POLLUTION

As we know, the occurring life existence on Earth is in danger due to rapid increase of various effects such as global warming, many health disorders that are found I human body as well in animals. Even the food, fruits and vegetables we eat is not safe, it also get affected by various air pollutants. Thus, the study shows that the human beings, animals, plants and the environment surrounding them are all affected by the air pollution and its harmful pollutants. Air pollutants effects on human health and on environment, summarized in the table 2.2.

POLLUTANTS	EFFECT ON HUMAN HEALTH	EFFECT ON ENVIRONMENT
Sulphur Dioxide(so_2)/ Sulphur Oxides(so_x)	➢ Respiratory problems. ➢ Causes irritation to the eyes and skin. ➢ Lung cancer ➢ Cardiovascular diseases and mortality rate.	➢ Damage vegetation. ➢ Causes Corrosion. ➢ Acid Rain. ➢ Plant and water damage. ➢ Aesthetic damage
Suspended Particulate Matter	➢ Increase in cancer and deaths from respiratory illness. ➢ Decrease in lung fluctuation. ➢ Causes problems like bronchitis, asthma, pneumonia etc.	➢ Acid deposition in air and water. ➢ Haze in natural areas. ➢ Change in nutrient cycles. ➢ Loss of ecosystem services.
Lead	➢ Destructive behavioural changes, learning disabilities and permanent brain damage. ➢ Loss of appetite. ➢ Stomach discomfort and/or constipation. ➢ Insomnia.	➢ Causes Lead Poisoning to soil organisms. ➢ Lead intoxication is a common disease found in animals.
Chlorine	➢ Results in coughing, chest pain and also cause water retention in the lungs.	➢ Chlorine affects the marine life. ➢ It causes damage to ozone layer.
Oxides of Nitrogen NO, NO_2, NO_x	➢ Nitrogen dioxide can irritate the lungs and lower resistance to respiratory infections such as influenza. ➢ Reducing the oxygen carrying capacity of the blood. ➢ Decreased functioning of the thyroid gland. ➢ Causes Asthma,	➢ Adverse effects on both terrestrial and aquatic ecosystem. ➢ Animals exposed to nitrogen dioxide may experience destruction of cilia, alveolar tissue disruption. ➢ Nitrogen Oxides are precursors for acid precipitation and ozone and each of which cause injury to plants.
Peroxyacetyl nitrate, PAN	➢ Causes irritation to eyes, respiratory system. ➢ People with heart and lung disease, children and the elderly are the most susceptible to PAN.	➢ Crops are affected. ➢ Winds can carry it to rural regions and pristine areas cause an ecological impact. ➢ Damage to clothes, plants and rubber articles. ➢ Agricultural and forests are affected.
Formaldehyde	➢ It can cause watery eyes, burning sensations in the eyes, and difficulty in breathing. ➢ Health effects include eye, nose, and throat irritation; wheezing and coughing.	➢ It breaks into carbon monoxide and formic acids which affects the environment. ➢ Causes acid rain.
Carbon Monoxide	➢ At low concentrations, fatigue in healthy people and chest pain in people with heart disease. ➢ At higher concentrations, impaired vision and coordination. ➢ Can cause flu-like symptoms.	➢ Its presence affects concentrations of other greenhouse gases including methane, ozone and carbon dioxide. ➢ It causes biomass burning.
Hydrogen Sulphide.	➢ It causes neurological, cardiovascular, metabolic, reproductive problems.	➢ Effects on the life of fishes and other marine organisms.
Ammonia	➢ When combined with water can injure and damage the respiratory tract. ➢ Problems with Bronchitis, asthma and cough	➢ Soil and freshwater acidification. ➢ aerosol production affecting global radioactive forcing(the greenhouse effect)
Carbon dioxide	➢ Causes headaches, dizziness, breathing problems, sweating, tiredness, increased heart rate, blood pressure, coma, asphyxia, and convulsions.	➢ A Greenhouse gas, responsible for global warming of Earth.
Ozone	➢ Induction of respiratory symptoms. ➢ Decrements in lung function. ➢ Also, causes throat irritation.	➢ It reduces crop yield and forest growth, and damages the appearance of leaf.

TABLE 2.1: AIR POLLUTION EFFECTS

III AIR QUALITY EVALUATION

Now a day, it is important to the society to look for Awareness of daily levels of air pollution. AQI is a tool which is used to report the overall air quality status and trends based on a specific standard. In India we are using CPCB Standard for calculating air quality index or environment pollution index. This index gives an idea about the environmental status as air quality. And also tells the general public to understand how clean or pollute air is breathe daily.

Overall this index can be used to give meaningful evaluation of air pollution to the common man. It also helps to identify the air pollution control policies or control equipment can reduce level of dominating pollutant. AQI is representing the cumulative effect of all the pollutant to show overall air quality status in better way. The AQI of specific pollutant is derived mainly from the physical measurement of pollutant like PM_{10}, $PM_{2.5}$, NO_2 and SO_2 etc. In the present study, six different methods were used to calculate ambient air quality index.

3.1 Methods to find Air Quality Index

Method I:

Air quality Index (AQI) is calculated based on the arithmetic mean of the ratio of concentration of pollutants to the standard value of that pollutant such as PM_{10}, $PM_{2.5}$, NO_2 and SO_2. The average is then multiplied by 100 to get the AQI index. AQI was then compared with rating scale (Kaushik et al., 2006). For individual pollutant AQI was calculated by the following formula

$AQI = (C/Cs)*100$

Where

AQI = Air Quality Index

C= the observed value of the air quality parameters pollutant (PM10, PM2.5, NO2 and SO2)

Cs= CPCB standard for residential Area (CPCB, 2009)

Method II:

In this procedure AQI is calculated by taking the geometric mean of the ratio of concentration of pollutants to the standard value of that pollutant such as PM_{10}, $PM_{2.5}$, NO_2 and SO_2. AQI was then compared with rating scale. (Ravikumar et. al., 2014)

Method III:

Oak Ridge National Air Quality Index (ORNAQI) is used for the relative ranking of an overall air quality status. Over all AQI was estimated by the following mathematical equation developed by the Oak Ridge National Laboratory (ORNL), USA is given below.

$$AQI = [39.02 \sum \frac{C}{C_s}]^{0.967}$$

Air quality Index then measured and compared with relative ORAQI values (Bhuyan et al. 2010).

Method IV:

Air Quality Index was done for combining qualitative measures with qualitative concept of the environment. The individual air quality index here is calculated as follow:

$AQI = (W * C)/Cs$

Where

AQI = Air Quality Index

W= Weighted of Pollutant

C= the observed value of the air quality parameters pollutant (PM_{10}, $PM_{2.5}$, NO_2 and SO_2)

Cs= CPCB standard for residential Area (CPCB, 2009)

Method V:

Air Quality Index was done based on dose response relationships of pollutants to obtain break point concentration.(USEPA, 2006,CPCB 2014) The individual air quality index for a given pollutant concentration (Cs) as based on linear segmented principle is calculated as

$$I_P = [\{\frac{(I_{HI} - I_{LO})}{(B_{HI} - B_{LO})}\} * (C_P - B_{LO})] + I_{LO}$$

Where

B_{HI} = Breakpoint concentration greater or equal to given Concentration

B_{LO} = Breakpoint concentration smaller or equal to given Concentration

I_{HI} = AQI value corresponding to B

I_{LO} = AQI value correspond to B

Finally;

AQI=Max (I_P) (where p=1, 2, 3…n; denotes n pollutants)

Data obtained from monitoring of ambient air at Residential site is used to calculate the air quality index (air pollution index) for critical parameter. Different AQI were estimated for various months and varying results were observed ranging from good to unacceptable for the same set of data. This may be due to eclipsing effect of the values used in the formulas. The statistical theory behind these AQI makes it more prone to variations viz. the use of means from simple arithmetic to logarithmic and weighted averages to use of breakpoint concentration as basis of estimation. As reported in USEPA, CPCB, 2014, the breakpoint concentration based AQI is more robust and can be used for decision making. Accordingly, the AQI values are calculated based on Break Point concentration for 24 hourly averages for PM10, PM2.5, SO2 and NO2 concentrations and are categorized as satisfactory to moderate during the study period at the residential site.

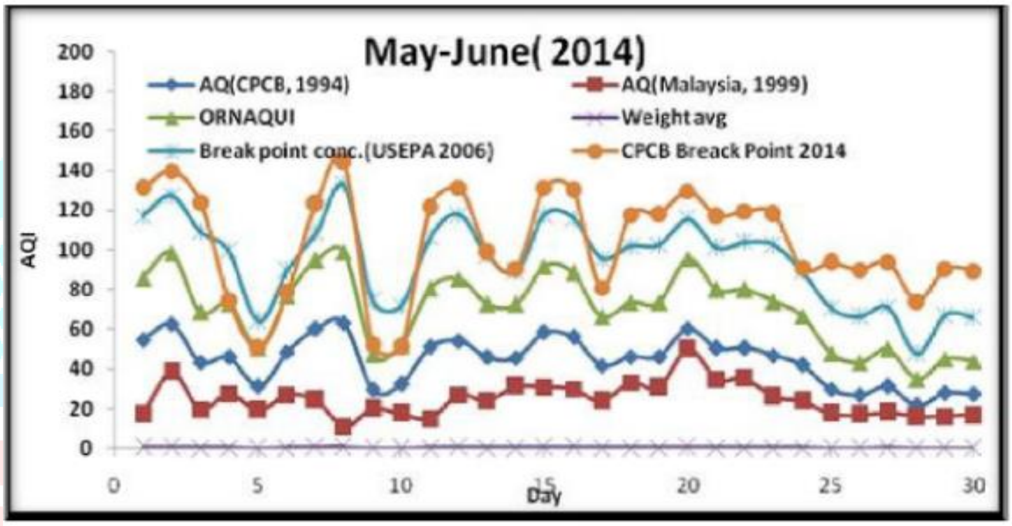

Fig. 3.2: **Variation in air quality index in the study area during May-June (2014)**

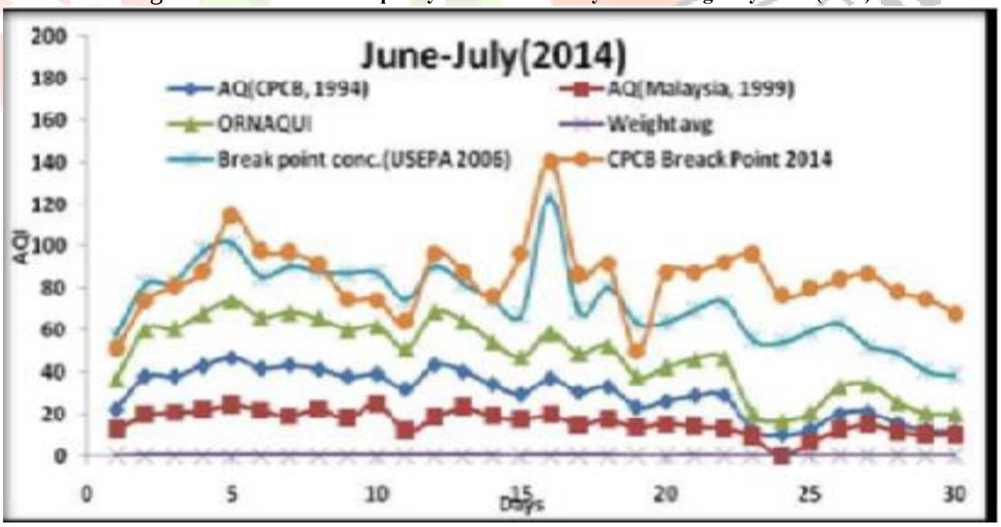

Fig. 3.3: **Variation in air quality index in the study area during June-July (2014)**

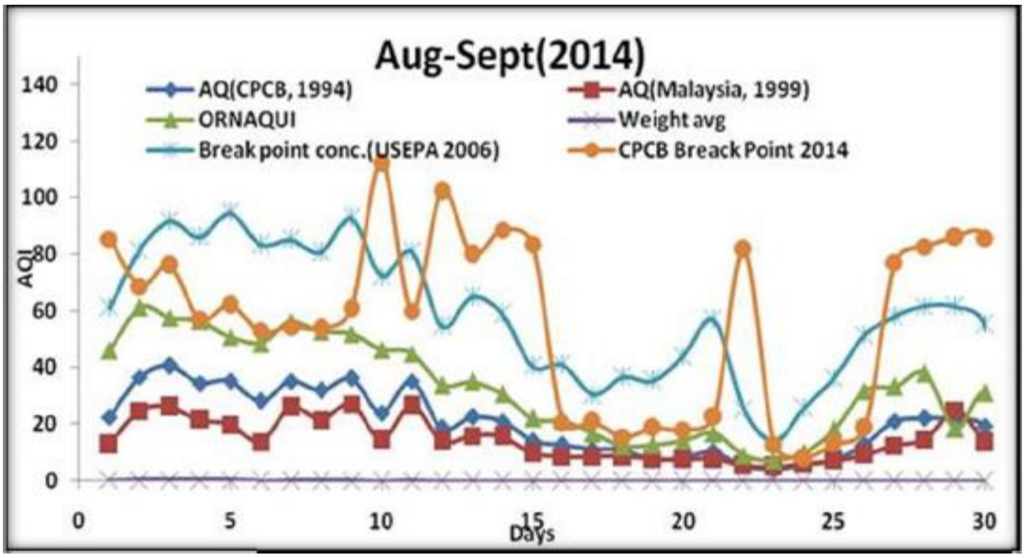

Fig. 3.4: **Variation in air quality index in the study area during Aug- Sept (2014)**

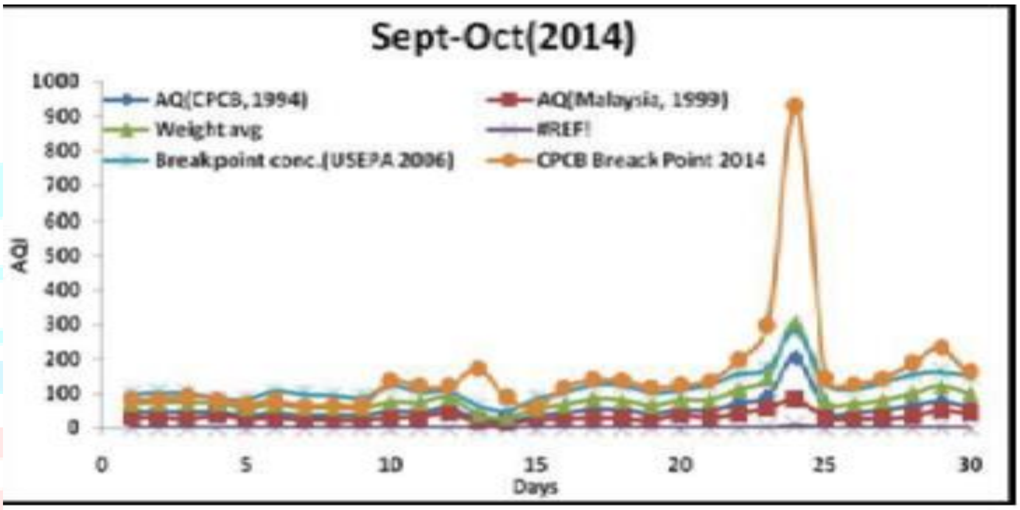

Fig. 3.5: **Variation in air quality index in the study area during Sept-Oct (2014)**

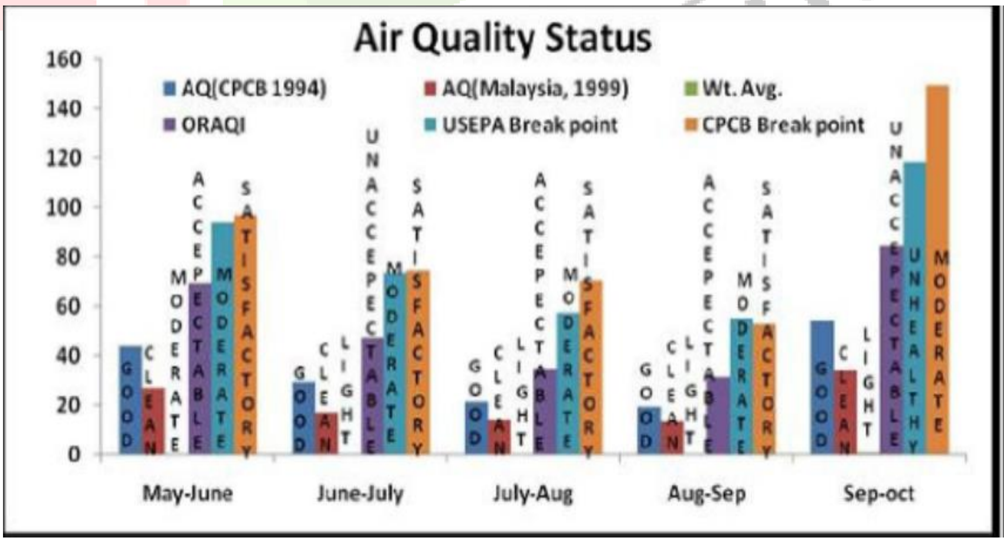

Fig. 3.6: **Classification of air quality index in the study area during study time**

The diurnal variation of different AQI has been shown in figure 3.2, figure 3.3, figure 3.4, and figure 3.5 for the month of May-June, June-July, July-August, August- September 2014 respectively .AQI values calculated for PM_{10} is found in the Satisfactory, $PM_{2.5}$ in Poor and NO_2 and SO_2 in Good category during May-June 2014. While they were found as PM_{10}: Satisfactory, $PM_{2.5}$: Moderate and NO_2 and SO_2 in Good category during June-July 2014. The AQI values calculated for PM_{10} is found in Good category, $PM_{2.5}$ in the Moderate, NO_2 and SO_2 is coming in the range of Good in July-August 2014. AQI values calculated for PM_{10} is coming in the Good, $PM_{2.5}$ is coming in the Satisfactory, NO_2 and SO_2 is coming in the range of Good in August-September. AQI values calculated for PM_{10} is coming in the

Satisfactory, $PM_{2.5}$ is coming in the Poor, NO_2 and SO_2 is coming in the range of Good in September-October 2014. The overall Air Quality Index was found to fall under the category of satisfactory to moderately polluted area (figure.3.6).

3.2 OCCURRENCE OF POLLUTANTS

In order to study the frequency and occurrence of individual pollutant in diurnal variation study factor analysis method through SPSS 13.0 software is used (Anikender et al. 2011). The frequency of occurrence of different pollutant has been shown in figure 8. It has been observed that occurrence of particulate matter is more as compared to other pollutant in all the seasons from May to October 2014. PM_{10} is found more polluting parameter as compared to $PM_{2.5}$, SO_2 and NO_2 with variance of average as 62%.

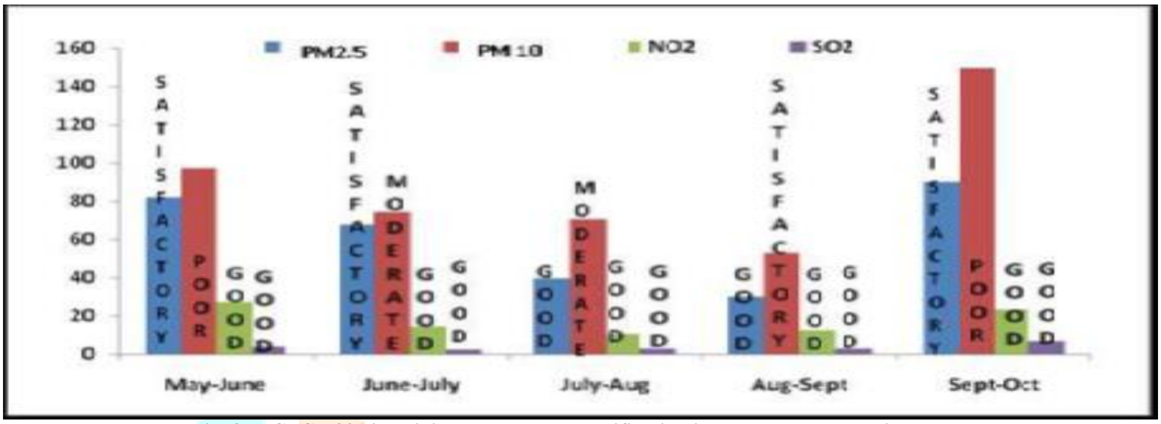

Fig. 3.7: CPCB 2014 Individual pollutant classification in the study area during study

Based on break point concentration, AQI of individual pollutant has been shown in figure. 3.7. It has been observed that Particulate matter ($PM_{2.5}$) is satisfactory, SO_2 and NO_2 is Good from May- October 2014. While PM_{10} is to be appear in moderate to poor from May- October 2014. This PM_{10} is factor which is the major pollutant for causing the overall air quality reduction. Source of PM_{10} may be thermal power plant, small medium scale industry and vehicle etc. PM_{10} may cause lot of respiratory problem to human health (Ekpenyong et. al. 2012).

IV EXPERIMENTAL WORK

This research work is focused to find out how our country's global warming is getting affected by the pollution from industries or from the harmful gases coming from electronic items. We basically have data of all the pollution causing gases of major cities in India. So as to find the air quality index by doing processing over it using hadoop and R language and analyzing it for getting the acquired results. Moreover we are trying to analyze the temperature of those cities so as to find how much air pollution exists in that city as well as how to control them so as to control the global warming and destructions caused by global warming. We have data for this which contains the temperature of the cities from 1960 to 2018 so with the help of them we can find how pollution in affecting day by day. And we have data of total greenhouse gas emissions of all the countries so by which we can conclude how greenhouse gases affecting the world from 1960 till 2018 and fir that we have to find out the solutions how can we control greenhouse gas emissions.

Data Description:-

There is a data of harmful gases in all cities of India which are affecting the global warming of the country and with that too temperature of the cities from 1960 to 2018 in csv format._And also have data of total greenhouse gas emissions of all the countries so by which we can conclude how greenhouse gases affecting the world from 1960 till 2018 in csv format only.
These files are extracted and loaded into Hive. The data is further processed, transformed, and analyzed to get which area is mostly affected by which gas and what could be stopped so that the air pollution get decreased and we get a good air quality index.

V ISSUES, CHALLENGES, AND NEEDS

San Francisco Bay selected to conduct air quality evaluation for using selected big data analysis approaches and machine learning models. Based on literature and experience, we highlight some research issues, challenges, and future needs. **Issues #1: Data quality and validation issue** - There are lots of sensor data quality issues which affect the accuracy of air quality evaluation and assessments due to device faults, battery issues, and sensor network communication problems. This brings the first need below.

Need #1: Research demand in big data quality assurance – As pointed out in [22] by Jerry Gao et al., there is a strong need in big data quality assurance research in data quality modeling, automatic real-time validation methods, and tools to increase the accuracy of air quality evaluation.

Issue #2: Real-time air quality monitor and supervision for air resources - As the advance of smart sensing and IoT, more and more environmental sensors (including air sensors and networks) have been installed and deployed for many air resources. However, there is a lack of integrated real-time big data based air quality evaluation and monitor environments for smart cities to support dynamic air quality evaluation, monitor, and supervision management. Air in a city could be considered as a multi-level air system which is impacted by multiple factors like pollutant emission levels, location, time, wind speed etc. The air quality on all levels usually affect each other.

Need #2: Research and development of real-time air quality monitor and evaluation systems supporting air quality evaluation and analysis on multiple levels.

This demand is caused by the lack of the existing research work addressing the air quality impacts on different levels due to air pollution from a special air source. This suggests the demand on an integrated real-time air quality monitor and evaluation system based on sensor networks and IoT infrastructures at the different levels.

Issue #3: Big data modeling issues for dynamic air quality monitor and analysis at the different levels for smart cities - Most published research work applied big data analytics approaches and used one specific machine learning technique for air resource at specific level in a limited location (or a region) during an interested time. The air system for a future smart city must support real-time air quality monitor, evaluation, and prediction.

Need #3: This implies that we need to develop integrated and dynamic air quality models using hybrid machine learning models to address these factors: a) the nature of dynamic wind flow, b) both single-input time series and multiple input time series, c) dynamic quality impacts on different atmospheric levels.

VI .CONCLUSION & FUTURE SCOPE

With the advancement of IoT infrastructures, big data technologies, and machine learning techniques, real-time air quality monitor and evaluation is desirable for future smart cities. This paper reports our recent literature study, reviews and compares current research work on air quality evaluation based on big data analytics, machine learning models and techniques. Finally, it highlights some observations on future research issues, challenges, and needs The future aim is to examine the interaction of pollutant mixtures and weather on health and health inequalities, now and in the context of future air quality and climate policies.

REFRENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] K. Elissa, “Title of paper if known,” unpublished.

[4] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[5] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[6] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

[7] V. M. Niharika and P. S. Rao, “A survey on air quality forecasting techniques,” International Journal of Computer Science and Information Technologies, vol. 5, no. 1, pp.103-107, 2014.

NAAQS Table. (2015). [Online]. Available:

https://www.epa.gov/criteria-air-pollutants/naaqs-table

[8] E. Kalapanidas and N. Avouris, “Applying machine learning techniques in air quality prediction,” in Proc. ACAI, vol. 99, September 1999.

[9] Ajit Noonia, Vikas Verma, Ruchika Khandelwal and Kushagra Gautam, “Sentimental Analysis on Twitter using Pig and Hive” International Journal of Innovative Technology and Exploring Engineering, vol. 9, issue 3, Jan 2020

[10] D. J. Nowak, D. E. Crane, and J. C. Stevens, “Air pollution removal by urban trees and shrubs in the United States,” Urban Forestry & Urban Greening, vol. 4, no. 3, pp. 115-123, 2006.

[11] T. Chiwewe and J. Ditsela, “Machine learning based estimation of Ozone using spatio-temporal data from air quality monitoring stations,” presented at 2016 IEEE 14th International Conference on Industrial Informatics (INDIN), IEEE, 2016.

[12] Y. Zheng, X. Yi, M. Li, R. Li, Z. Shan, E. Chang, and T. Li, “Forecasting fine-grained air quality based on big data,” in Proc. The 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2267-2276, August 10, 2015.

[13] J. A. Engel-Coxa, C. H. Hollomanb, B. W. Coutantb, and R. M. Hoffc, “Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality,” Atmospheric Environment, vol. 38, issue 16, pp. 2495–2509, May 2004.

[14] Questioning smart urbanism: Is data-driven governance a panacea? (November 2, 2015). [Online]. Available: http://chicagopolicyreview.org/2015/11/02/questioning-smart-urbanis m-is-data-driven-governance-a-panacea/

[15] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.