TRENDANALYSIS OF STREAMING AIR QUALITY DATA OF FOUR TIER I CITIES IN INDIA

A DISSERTATION

Submitted in partial fulfillment of the requirement for the award of the degree

of

MASTER OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

by

Alka Yadav (Roll No. 13535002)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY, ROORKEE

ROORKEE- 247667(INDIA)

MAY, 2016

Declaration of Authorship

I, Alka Yadav, declare that this thesis titled, ’Trend Analysis of Streaming Air Quality of Four Tier I Cities in India’ and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: ______________________

Date: ________________________

ABSTRACT

Time series data mining refers to the process of extracting relevant information from time series data. Trend analysis of time series data refers to a set of techniques which help to reveal hidden patterns in time series data and it is important as it helps to make prediction based on past data. Time series mining finds its applications in a number of fields like medical, finance, weather forecasting and many more. In this report trend analysis of air quality data has been discussed.

Air pollution has been a major concern of study as it can have adverse impact on human health as well as ecosystem. Not much research has been carried out on air quality data of India. The cities in India have been classified into Tier I, Tier II and Tier III on the basis of their population by Reserve Bank of India. This report aims at doing the trend analysis of air quality of Tier I cities in India. The Tier I cities included are New Delhi, Mumbai, Chennai and Bengaluru. The air quality data included parameters are SO_2, NO_2, NO_x, NO, CO, O_3, temperature, relative humidity, PM2.5 and PM10.

The air quality data collected has a large number of dimensions and hence to reduce dimensionality, after a number of experiments, principal component analysis has been found to be the best technique for dimensionality reduction to reduce the number of variables under consideration. The results of interpretation of principal component analysis have been used to provide a useful description that can be interpreted in terms of sources of air pollution. After removing seasonality from air quality data ARIMA model has been applied to better understand the data and also predict future values in the time series data. GARCH model has also been applied on the data and the results of ARIMA and GARCH model are quite comparable. In previous studies, ARIMA has been applied only on static data but in this study SDA (Streaming Data ARIMA) has been proposed that applies ARIMA model on streaming data and estimates how various parameters of ARIMA model change with every iteration based on window size of streaming data. The results of SDA and ARIMA on static time sequence are compared and have been found to be very promising.

Keyword Index: Air Quality Data India, Tier I cities in India, pollutants, trend analysis, streaming data ARIMA

CERTIFICATE

This is to certify that the statement made by the candidate is correct to the best of my knowledge and belief.

DATE: …………………………………　　　SIGNED: …………………………..

(DR. DURGA TOSHNIWAL)

ASSOCIATE PROFESSOR

DEPT. OF CSE, IIT ROORKEE

Acknowledgements

I would like to thank my project guide Dr. Durga Toshniwal for her guidance and helping me carry out the thesis work. I would also like to thank my family for providing their support throughout the work. I would also like to thank my friend Neeraj Kumar who helped me in obtaining the data without which this work wouldn’t have been possible.

CONTENTS

List of Tables

List of Figures

CHAPTER 1 INTRODUCTION AND MOTIVATION

1.1 Overview

Air quality data collected from Central Pollution Control Board, India website has been analyzed for the concentrations of various pollutants involving CO, O_3, etc. in order to identify the trends and sources of air pollution. The government of India has divided the Indian cities into three tiers on the basis of population by Reserve Bank of India in order to allocate house rent allowance to the employees employed in these cities. The Tier-I cities in India are Ahmedabad, Bengaluru, Chennai, Delhi, Hyderabad, Kolkata, Mumbai and Pune. In this report, air quality data of four Tier I cities have been collected from Central Pollution Control Board, India. The cities included are Delhi, Mumbai, Bengaluru and Chennai. The hourly data of various pollutants collected for these cities include SO_2, NO_2, NO_x, NO, CO, O_3, temperature, relative humidity, PM2.5 and PM10 over a range from 2006 to 2016.

In this report trend analysis of hourly air quality data of Delhi, Mumbai, Bengaluru and Chennai from 2006 to 2016 has been carried out. It involves the graphical analysis of hourly data of pollutants like SO_2, NO_2, NO_x, NO, CO, O_3, temperature, relative humidity, PM2.5 and PM10 and identifying seasonal patterns of pollutants and then identifying various sources of air pollution responsible for that pattern. In this report photochemical activity resulting in the observed behavior of pollutants has also been illustrated.

A number of statistical techniques including principal component analysis have been carried out. Principal Component Analysis (PCA) has been applied to the four Tier I cities and the PCA results have been used to identify associations between various parameters of the dataset [1]. The factors being considered for checking and validating pollutant emissions include seasonal and time-of-day factors along with photochemical activity leading to transformations producing Ozone and nitrogen dioxide. After application of principal component analysis on standardized air quality data, varimax rotation has been performed on the components identified and then association of factors or identified components to sources of air pollution has been made.

In [8] and [9] the application of Autoregressive Integrated Moving Average (ARIMA), implemented by Box-Jenkins approach to predict air quality data has been mentioned. The ARIMA model involves three steps namely model identification, parameter estimation and performance evaluation. The various measures of model performance identification are done in terms of mean square error, mean

absolute error, mean absolute percentage error etc. After that the model can be used to predict future values of air quality data based on previously observed values. In [10] the usage of GARCH model to analyze air quality data is mentioned. The results of GARCH and ARIMA model have been compared and found to be quite comparable and it ensures the correctness of the model parameters predicted.
ARIMA [16] and GARCH have been the traditional and most popular models for static time sequence but these models have not been applied on streaming data. In this report ARIMA [16] model has also been applied on streaming air quality data and the proposed approach is named as SDA (streaming data ARIMA). The ARIMA model is applied on a stream of fixed window size and then for every consecutive and cumulative stream fixed window size the model parameters are analyzed and compared with the model parameters of whole dataset. The results of SDA and ARIMA on static time sequence have been compared and found to be quite comparable.

1.2 Motivation

There are a number of factors affecting our environment adversely. Air pollution has emerged as one of the bigger threats to the environment of India. In this field much research has been carried out across the world but in India not much research is done. The seriousness of the issue and the positive impact which can be caused due to the outcome of this study is the major source of motivation to analyze the past data of ambient air quality of major cities of India. The factors of motivation can be listed as below:

- There is a need to identify the trends in the concentration of air pollutants over a number of years and use them to identify the major sources of air pollution.
- Trend analysis of air quality data can help to identify hidden patterns and associations among various pollutants which are not visually obvious.

1.3 Problem Statement

This report focuses on analyzing the air quality data of four Tier I cities in India. The cities included are Delhi, Mumbai, Bengaluru and Chennai. The main objectives of the study involve:

- To analyze the levels of major pollutants like SO_2, NO_2, O_3, NO_x, NO, CO, etc. in four above mentioned Tier I cities in India and their variations over the period of time.

- To identify various meteorological and photochemical processes responsible for the trends obtained.
- Use statistical methods like principal component analysis to reduce dimensionality and identify hidden patterns.
- To provide meaningful interpretation of principal component analysis results to find associations of identified components to sources of air pollution.
- To do trend analysis on static air quality data and analyze the results.
- To do trend analysis on streaming air quality data and assess the results obtained.

1.4 Organization of Report

The rest of the report is organized as follows. The chapter 2 illustrates the literature reviews done; chapter 3 contains the proposed framework for the work done. Chapter 4 illustrates the techniques used to analyze the air quality data and illustration of the experimental results obtained. Chapter 5 contains the discussion on our research so far and the conclusion drawn from the discussion. In the last reference section all the research papers which are referred for this study is listed down.

CHAPTER 2 LITERATURE REVIEW

In [1] one year air quality data from an urban roadside location in Central London has been analyzed using statistical techniques like principal component analysis and also uses Gaussian plume models. This study illustrates a high pollution time of winter where there was a steep increase in the pollutants was detected and the reason for this was elevated nitrogen dioxide concentrations. It also shows the principal component analysis application to find the factors and then associate these factors with the sources of air pollution.

In [2], a study has been carried out to investigate the relationship among nitric oxide, nitrogen dioxide, oxides of nitrogen, Ozone, and carbon monoxide for two year data of an urban site in Delhi, India. This study uses simple statistical techniques like mean and trend line fitting to identify trends and relationships among the pollutants.

In [3] Sarath Guttikunda et al. have done the urban air pollution analysis by considering six cities of India namely Pune, Chennai, Indore, Ahmedabad, Surat and Rajkot. In their report they have given detailed analysis of the pollution level in the considered city and their impact on the health of people. This is a detailed report about air quality data and it mentions causes of air pollution in the cities, the adverse health impacts that can be caused by air pollution as well as proposes what-if scenarios for 2020.

A number of studies related to trend analysis of air quality data have been carried out like the ones mentioned in [4], [5] and [1]. These studies have been done on the air quality data of Iran, China and Central London and used graphical and statistical techniques to analyze the trends. In India [2],[3], [6] and [8] are few of the studies carried out in order to identify the relation among various pollutants and also mapping the identified components to sources of air pollution has been done.

The work in [7] and [11] mentions the detailed application of principal component analysis on time series data. The study carried out in [11] uses daily air pollution data to do principal component analysis and then provides associations of the identified factors to the sources of air pollution and then these factors are associated to mortality rates in the country. In [12] mentions the application of Autoregressive Integrated Moving Average (ARIMA), implemented by Box-Jenkins approach to predict air quality data. In this Air pollution index, which is used as a measure of air quality in Malaysia, has been used to predict air quality in a region. The model used in this study is formed in three steps namely model identification, parameter estimation and performance evaluation. There are various measures to identify model performance in terms of mean square error, mean absolute error, mean absolute percentage error etc. After that the model can be used to predict future values of air quality data based on previously observed values.

In [9] mentions the application of Autoregressive Integrated Moving Average (ARIMA), implemented by Box-Jenkins approach to predict air quality data. In [9] Air pollution index, which is used as a measure of air quality in Malaysia, has been used to predict air quality in a region. The model used in this study is formed in three steps namely model identification, parameter estimation and performance evaluation. There are various measures to identify model performance in terms of mean square error, mean absolute error, mean absolute percentage error etc. After that the model can be used to predict future values of air quality data based on previously observed values. In [8] Inderjeet Kaushik et al. have carried out trend analysis of air quality using ARIMA model. In this study the levels of various air pollutants like carbon dioxide, nitrogen oxides etc. are analyzed and a model based on Box-Jenkins approach has been developed to predict air quality.

In [10] GARCH model has been used to predict the air quality trends in future for 8 monitoring sites in Taiwan. This study uses a combination of ARCH, GARCH and multivariate analysis to predict associations between various pollutants and also for variation of air quality trend prediction.

CHAPTER 3 PROPOSED WORK

In this report the trend analysis of air quality data of four Tier I cities in India has been carried out. The hourly data for various pollutants like NO, CO, PM2.5, etc. has been collected from Central Pollution Control Board, India for four Tier I cities in India namely Delhi, Mumbai, Bengaluru and Chennai.

The hourly data collected contained some missing values i.e. for every pollutant there were some days when the concentration recordings were unavailable. In order to make the data to be used for doing trend analysis, imputation of missing values has been done using the linear interpolation method. Three methods were identified to fill in the missing values. First method being deleting the records with missing values but this method leads to distortion of data so this method was not used. Second method being using mean substitution i.e. replace missing values with nearby mean values or the whole mean value of the particular parameter, this method preserves the mean but at the same time leads to a low standard deviation hence this was not used. Third method being imputation ,several tests were made so that it doesn't lower the standard deviation and also the mean value is not disturbed , hence interpolation has been used to fill in the missing values.

After that the data has been normalized using z-score method. The next step involves the application of principal component analysis on normalized data to reduce dimensionality of data. After that the results of principal component analysis are used to identify the correlated components which are then undergone to rotations to map to sources of air pollution.

ARIMA and GARCH model has been applied on the air quality data and the parameters of both the models are compared to establish the correctness of the parameters. Streaming Data ARIMA (SDA) has been proposed in which ARIMA model has been applied on streaming data of fixed window size and the changes in the model parameters are analyzed. The results of SDA and ARIMA on static time sequence has been compared and the results have been found to be quite comparable.

CHAPTER 4 EXPERIMENTAL RESULTS

4.1 Dataset Description

The air quality data is obtained on hourly basis for various pollutants over a range of year 2006 to 2016. The data collected involved some missing values which have been interpolated to convert it to a usable dataset. The cities being considered in the report include New Delhi (stations included are Anand Vihar, ITO and IGI Airport), Mumbai (stations are Bandra and Airoli), Chennai (stations are IIT and Alandur) and Bengaluru (stations are BTM and Peenya). The parameters being considered in the analysis of Tier I cities are PM2.5, relative Humidity, SO_2 (Sulphur Dioxide), PM10 (Particulate Matter), O_3 (Ozone), NO_x (Oxides of nitrogen), NO_2 (Nitrogen Dioxide), NO (Nitric Oxide), CO (Carbon Monoxide) and Temperature. The dataset is collected on hourly basis and is then aggregated on monthly, seasonally and yearly basis for purpose of analysis.

4.2 Data Preprocessing

The dataset is collected for four Tier I cities namely, New Delhi, Mumbai, Chennai and Bengaluru are from 1-Jan-2006 to 14-Apr-2016. The dataset collected contained some missing values so in order to fill in the missing values three methods were considered. First method being deleting the records with missing values but this method leads to distortion of data so this method was not used. Second method being using mean substitution i.e. replace missing values with nearby mean values or the whole mean value of the particular parameter ,this method preserves the mean but at the same time leads to a low standard deviation hence this was not used. Third method being imputation ,several tests were made so that it doesn't lower the standard deviation and also the mean value is not disturbed , hence interpolation has been used to fill in the missing values so as to get a usable dataset to do analysis. After imputing the missing values using interp1d interpolation method. After that every parameter is normalized using **z-score** method as calculated in eq. (4.2.1). For a value x of parameter X, z-score is as below:

$$\mathrm{z-score(x)} = \frac{\mathrm{x-mean(X)}}{\mathrm{stdev(X)}} \qquad (4.2.1)$$

where mean(X) is calculated as per eq. (4.2.2) is the average value of parameter X over N number of records being considered and stdev(X) is the standard deviation of X.

$$mean(X) = \sum_{i=1}^{N} \frac{X(i)}{N} \qquad (4.2.2)$$

4.3 Investigation into relationship among various pollutants

The yearly variations of Oxides of Nitrogen (NO_x), Ozone (OZONE), Nitrogen Dioxide(NO_2), Carbon Monoxide (CO), Nitric Oxide (NO) and Sulphur Dioxide (SO_2) for Delhi are shown in Figure 4.3.1. Similarly for Mumbai, Bengaluru and Chennai the variations are in Figure 4.3.2, Figure 4.3.3 and Figure 4.3.4. Table 4.3.1 summarizes the yearly variation of various pollutants for Delhi, Mumbai, Bengaluru and Chennai. Below figures shows the seasonal variation of four Tier I cities for years 2006-15 summarized for NO, NO_2, NO_x, OZONE, SO_2 and CO.

Table 4.3.1: Yearly variation of pollutants

Pollutant	Delhi	Mumbai	Bengaluru	Chennai
NO	Shows seasonal variation over years	Shows seasonal variation from 2014-16	Shows seasonal variation	NO concentrations were lower than NO_x during the initial years 2009-2013 but after that NO concentrations are greater than NO_x due to more photochemical activity.
NO_2	Shows seasonal variation over years	Shows seasonal variation from 2014-16	Shows seasonal variation	NO_2 concentrations are higher than NO_x due to NO titration effect.
NO_x	Shows seasonal variation over years	Shows seasonal variation from 2014-16	. Concentration of NO_x is greater than NO and less than NO_2 except when NO_x achieved its highest in 2014. This behavior is attributed to changing environmental conditions which favor photochemical reactions leading to NO titration effect.	Shows seasonal variation.
Ozone	It has been almost stable from 2008-14 but there is steady decline after 2014 owing to increased emission of chlorofluorocarbons (CFCs).	Ozone increased from 2011 to 2013, decreased from 2013 to 2014 and after 2014 it is almost constant.	In the available dataset of Bengaluru Ozone is not present.	Not present in the available dataset.
SO_2	SO_2 has gradually decreased during the study period, an	Not much variation in concentration over the	There are 277 instances of SO2 that has exceeded the	There are 169 instances of SO2 in 2012,2013 and 2014

	estimate made based on data from ground-based monitoring stations. Whereas, majority of the stations are located in urban areas, where environmental regulations have reduced pollution levels locally. There are only some stations in India that collect measurements near power plant emissions.	study period.	prescribed standard of 80.00 µg/ m^3 and those instances appear in year 2013, 2014 and 2015, the reason being increasing vehicular pollution, as mentioned more in Alarming situation in Bengaluru for other pollutants increasing concentration in Bengaluru.	that has exceeded the prescribed standard of 80.00 µg/ m^3.
CO	CO and Ozone are positively correlated	CO increased in 2010 and achieved highest concentration in 2015 and again declined in 2016.Ozone and CO are negatively correlated during the course of time.	Not much variation over the years.	For CO adequate numbers of observations are not present in the dataset.

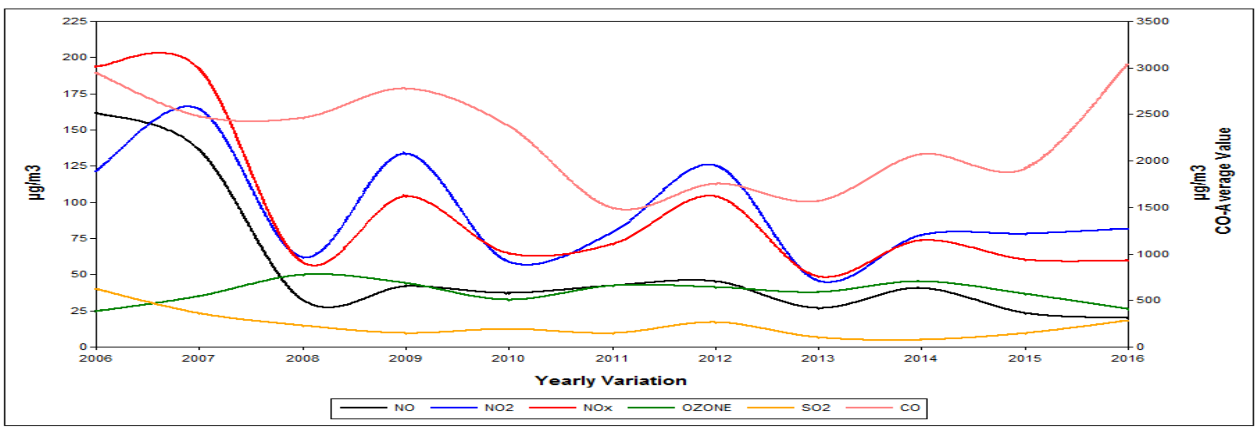

Figure 4.3.1: Yearly variation of NO, NO_2, NO_x, OZONE, SO_2 and CO from 2006 to 2016 in Delhi

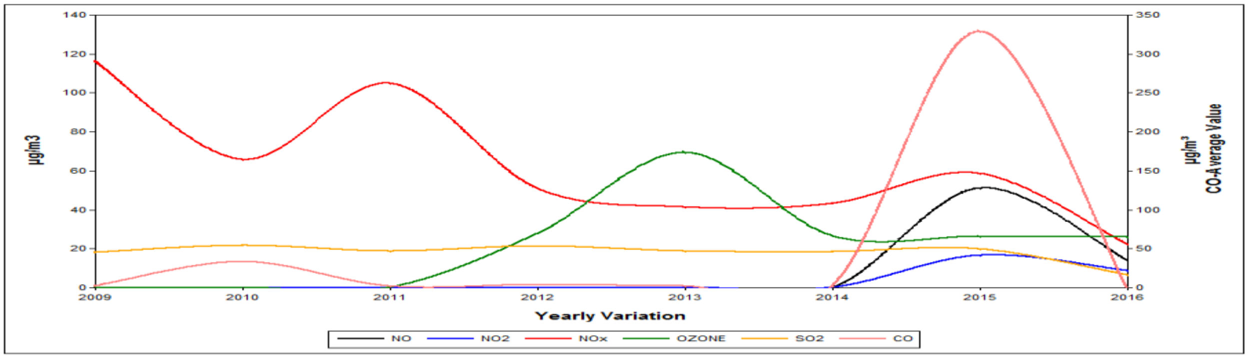

Figure 4.3.2: Yearly variation of NO, NO_2, NO_x, OZONE, SO_2 and CO from 2006 to 2016 in Mumbai

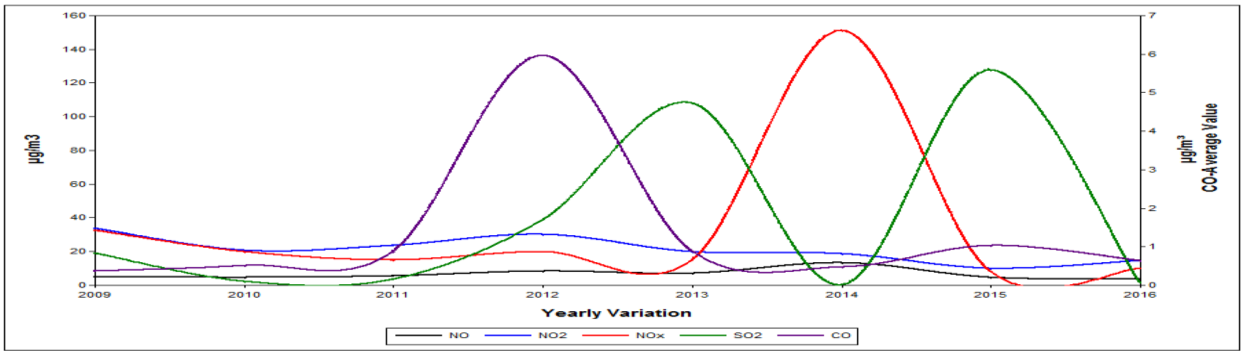

Figure 4.3.3: Yearly variation of NO, NO_2, NO_x, SO_2 and CO from 2006 to 2016 in Bengaluru

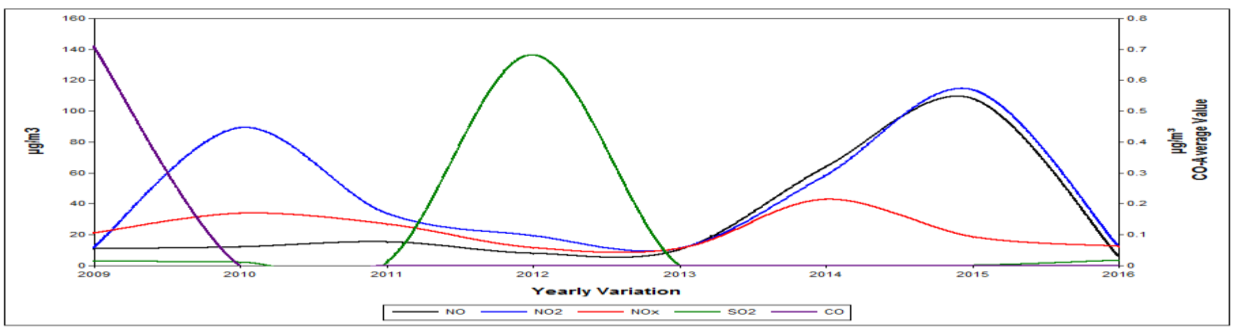

Figure 4.3.4: Yearly variation of NO, NO_2, NO_x, OZONE, SO_2 and CO from 2006 to 2016 in Chennai

With the semi-arid climate in Delhi four seasons have been identified as: Jan-Mar, Apr-Jun, Jul-Sep, Oct-Dec and the seasonal variations of NO, NO_2, NO_x, Ozone, SO_2 and CO for the year 2006-15 are shown in Figure 4.3.5. For Mumbai, Bengaluru and Chennai seasonal variations are in Figure 4.3.6, Figure 4.3.7 and Figure 4.3.8.

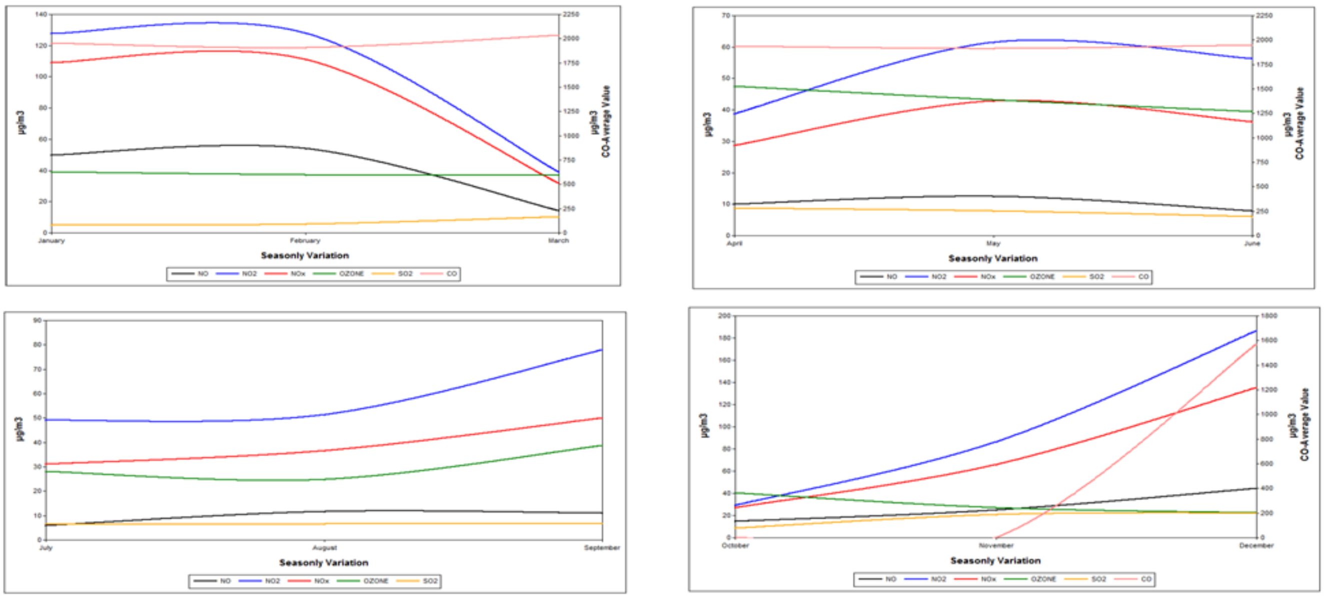

Figure 4.3.5: Seasonal variation of NO, NO_2, NO_x, OZONE, SO_2 and CO in Delhi

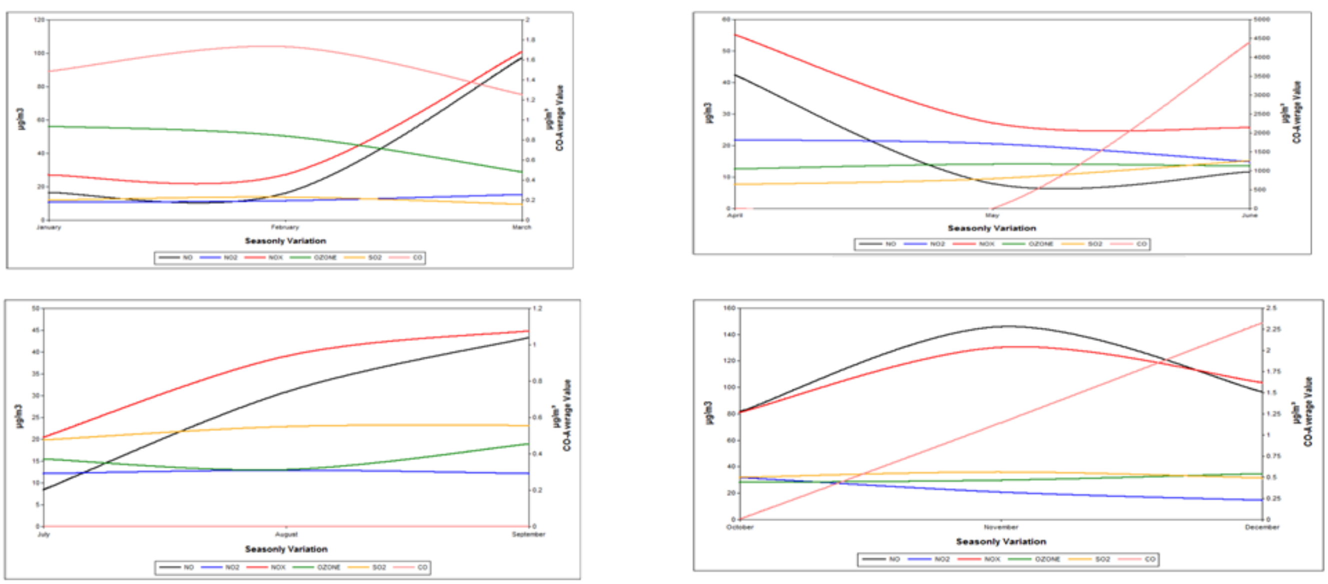

Figure 4.3.6: Seasonal variation of NO, NO_2, NO_x, OZONE, SO_2 and CO in Mumbai

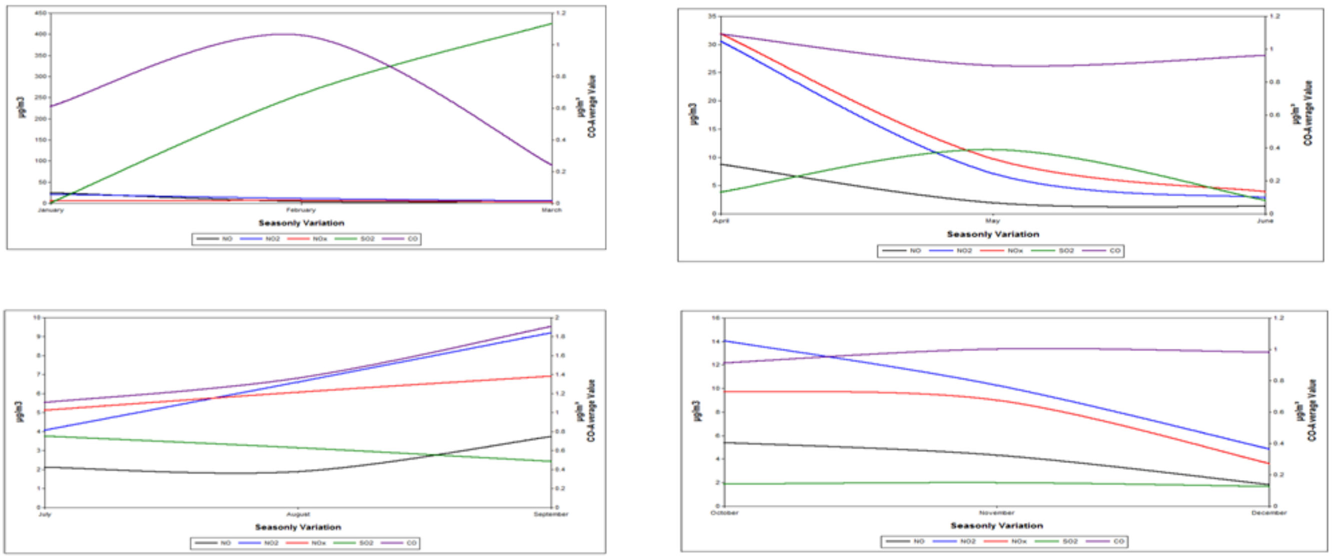

Figure 4.3.7: Seasonal variation of NO, NO_2, NO_x, OZONE, SO_2 and CO in Bengaluru

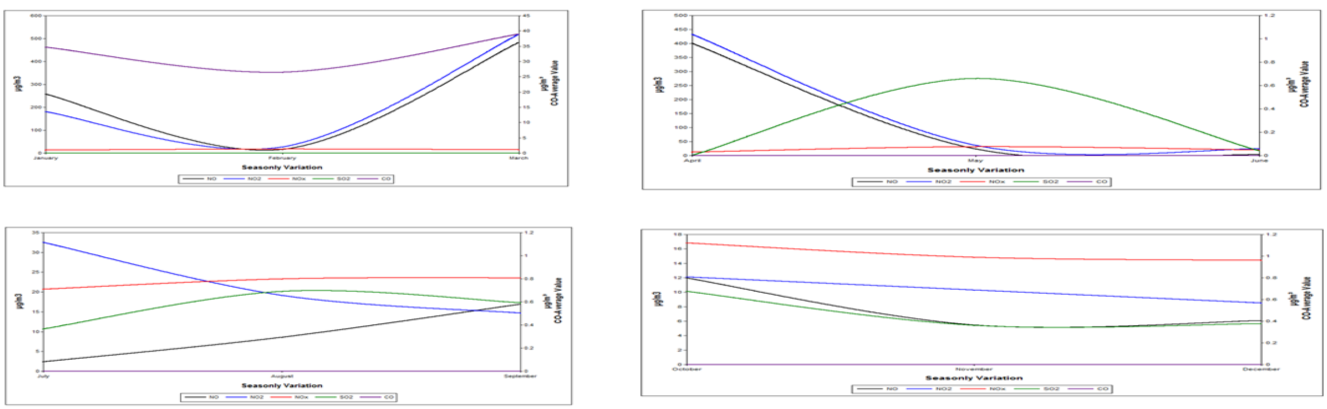

Figure 4.3.8: Seasonal variation of NO, NO_2, NO_x, OZONE, SO_2 and CO in Chennai

Table 4.3.2: Seasonal highest and lowest contribution of various pollutants of Tier I cities

Quarter	NO		NO_2		NO_x		Ozone		SO_2		CO	
	Highest	Lowest	Highest	Lowest	Highest	Lowest	Highest	Lowest	Highest	Lowest	Highest	Lowest
Jan-Mar	Chennai	Bengaluru	Delhi	Mumbai	Delhi	Bengaluru	Mumbai	Delhi	Bengaluru, Chennai	Delhi	Delhi	Chennai
Apr-Jun	Chennai	Bengaluru	Delhi	Mumbai	Delhi, Mumbai	Bengaluru	Delhi	Mumbai	Chennai	Delhi	Delhi	Bengaluru
Jul-Sep	Chennai	Bengaluru	Delhi	Mumbai	Delhi	Bengaluru	Delhi	Mumbai	Mumbai	Bengaluru	Bengaluru	Chennai
Oct-Dec	Delhi	Bengaluru	Delhi	Chennai	Mumbai	Bengaluru	Chennai	Mumbai	Mumbai	Bengaluru	Delhi	Chennai

From Table 4.3.2 it is found that Delhi is the worst city to stay and Bengaluru is the best city to stay in all the four quarters. From the above four seasonal variation figures in Delhi it is deduced that there is a strong seasonality between NO, NO_x and NO_2. Their concentrations are considerably higher during winter. Also, the Ozone levels are highest during summer as the Ozone production is enhanced by the suitable meteorological conditions for the chemical production of Ozone. CO and Ozone show higher concentrations during summer and lower concentrations during winter. The aggregated monthly variations of NO, NO_2, NO_x, OZONE, SO_2 and CO are shown in Figure 4.3.9 below.

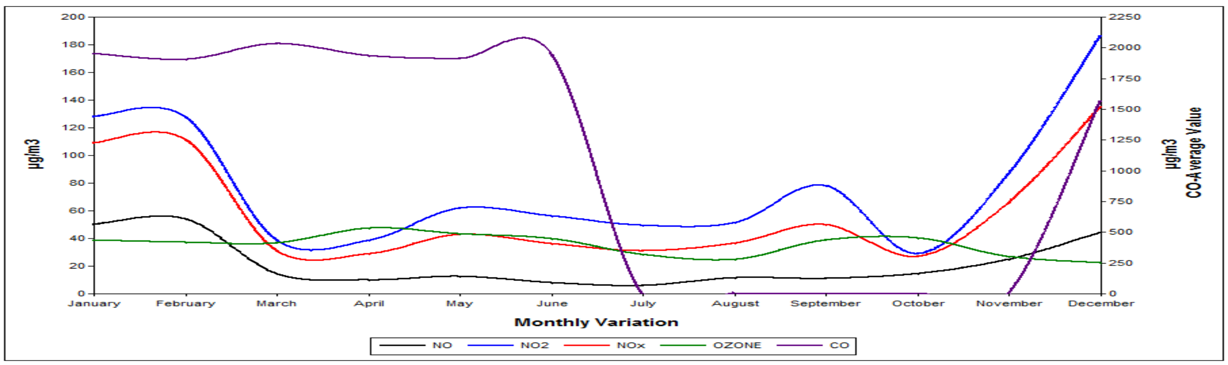

Figure 4.3.9: Monthly variation of NO, NO_2, NO_x, OZONE and CO in year 2015 in Delhi

From data available variations of temperature and humidity in Delhi show that temperature reaches its maximum of 38^{o}C during summer and minimum during winters. Also a negative correlation is seen between temperature and humidity of Delhi.

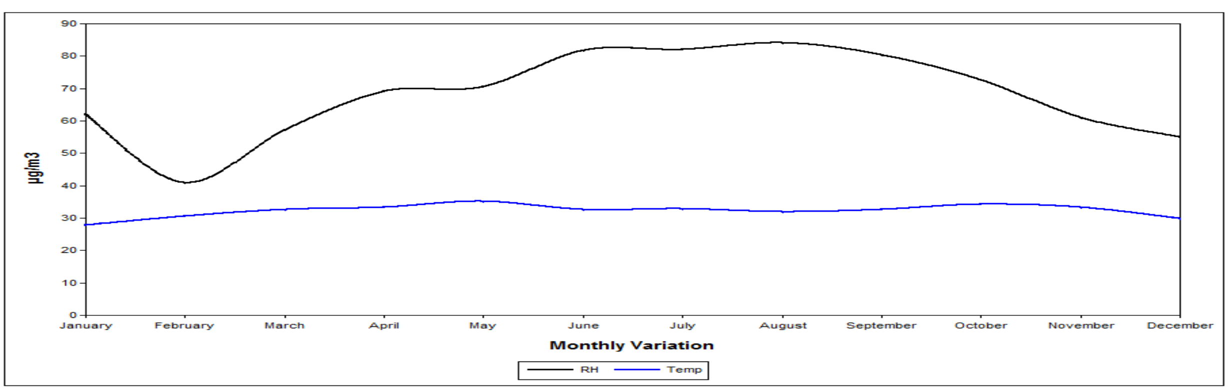

Figure 4.3.10: Monthly Variation of Temperature and Relative Humidity in Mumbai in 2015

Above Figure 4.3.10 for monthly variation of temperature and relative humidity in Mumbai shows that the temperature is almost constant whereas humidity increases from May-Sep. No correlation between temperature and humidity as temperature is almost constant and humidity varies independent of it.

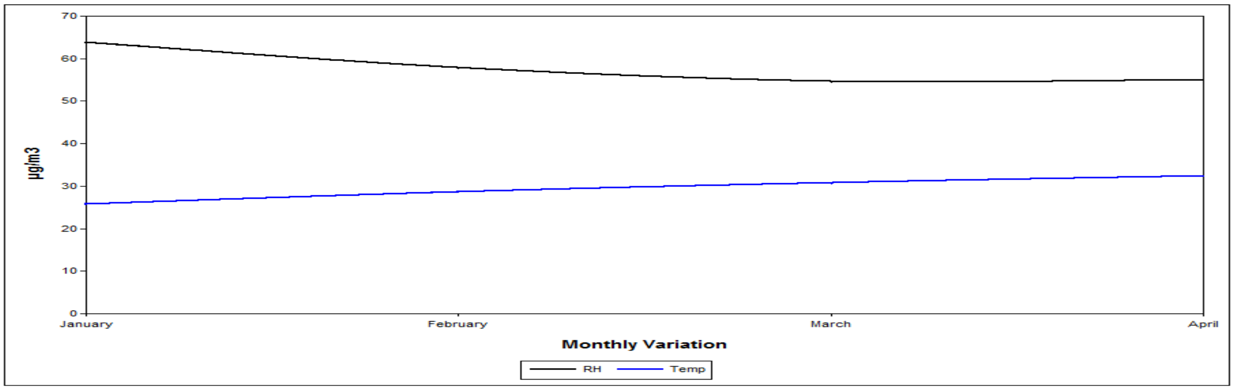

Figure 4.3.11: Monthly Variation of Temperature and Relative Humidity in Bengaluru in 2015

Above Figure 4.3.11 for monthly variation of temperature and relative humidity in Bengaluru shows that temperature and humidity are almost constant. Below Figure 4.3.12 for Temperature Variations of Delhi, Mumbai, Bengaluru and Chennai shows that Delhi has maximum temperature during summer and lowest temperature during winters. While Mumbai has almost the same temperature throughout the year between the range of 28-34^{o}C.For Bengaluru the temperature varies between 26-32^{o}C and for Chennai the temperature ranges between 25-36^{o}C. From Oct 2015-Mar 2016 the temperature of Chennai falls down.

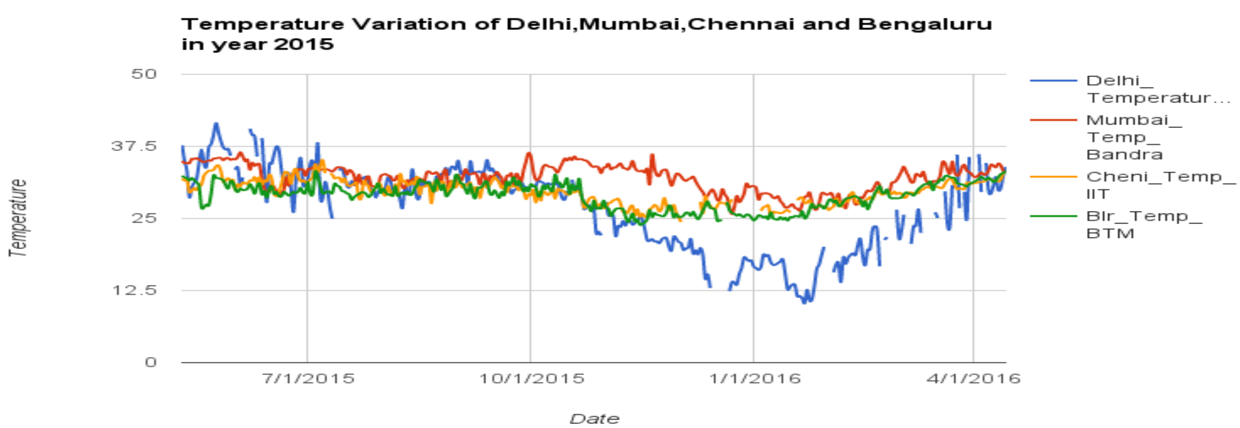

Figure 4.3.12: Temperature Variations of Delhi, Mumbai, Bengaluru and Chennai in 2015

Below Figure 4.3.13 for Humidity Variations of Delhi, Mumbai, Bengaluru and Chennai shows that Chennai exhibits high humidity during Apr-Jul 2015 and has moderate humidity from Jul-Oct 2015 and then the humidity increases from Nov 2015-Mar 2016.

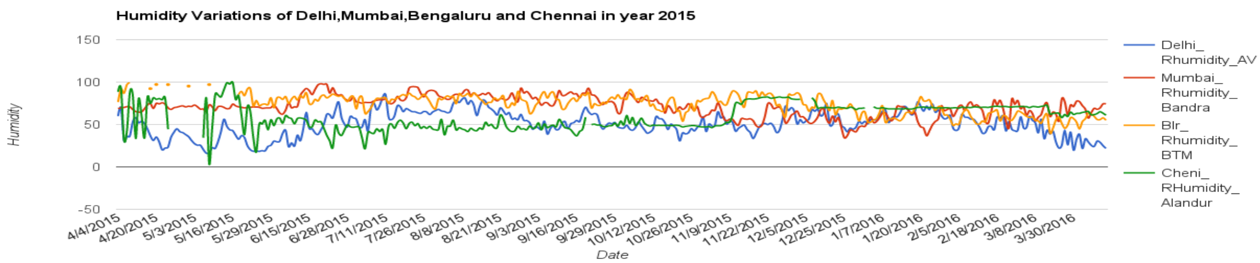

Figure 4.3.13: Humidity Variations of Delhi, Mumbai, Bengaluru and Chennai in 2015

Below figure shows the NO_x and CO contributions of Tier I cities during 2009-16. Mumbai has been the main contributor for NO_x and Delhi being the highest contributor for CO during the time range.

Figure 4.3.14: NO_x and CO contributions of Tier I cities during 2009-16

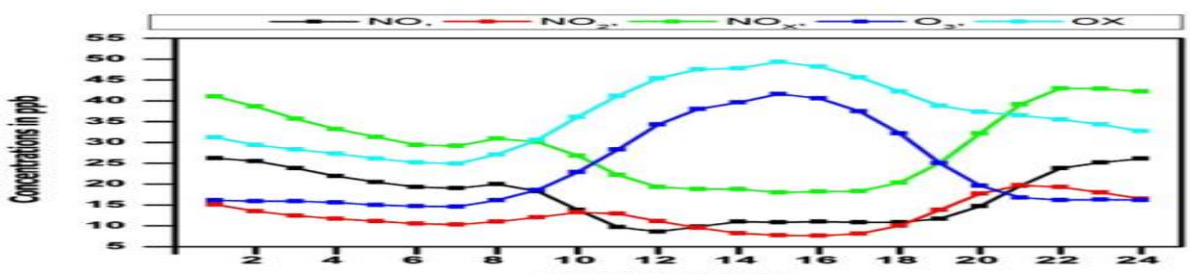

Figure 4.3.15: Hourly variability of annual averaged hourly concentrations of NO, NO_2, NO_x, OZONE and OX (sum of Ozone and NO_2) for the period from 2006 to 2016 in Delhi

Figure 4.3.15 shows the hourly variability of annual averaged hourly concentrations of NO, NO_2, NO_x, OZONE and OX (sum of Ozone and NO_2) for the period from 2009 to 2016 in Delhi. From Figure 4.3.15 it is deduced that the concentration of NO is highest during midnight to 4:00a.m.in the morning due to increased level of traffic emissions. Concentration of NO decreases after 8:00a.m due to its reaction with O_3 which leads it to the oxidation to NO_2. Concentration of NO is lowest at 11:00a.m which conforms to the highest photolytic activity [2] during the time. Concentration of NO_2 decreases till 7:00a.m in the morning due to below reaction that eventually leads to the production of O_3.

$$NO_2 + hvr ==> NO + O \qquad (4.3.1)$$

$$O + O_2 + E ==> O_3 + E \qquad (4.3.2)$$

$$NO + O_3 ==> NO_2 + O \qquad (4.3.3)$$

In eq. (4.3.1) there is no O_3 production as the O_3 produced during eq. (4.3.2) is then destroyed by its reaction with NO in eq. (4.3.3) which leads to its oxidation to NO_2. Concentration of NO_2 start increasing after 4:00p.m and is highest at around 1:00 in night. Concentration of O_3 increases after 7:00a.m due to increased photochemical activity during the time and peaks around 2:00p.m after which it starts declining. During the same time NO_2 decreases as it is a preliminary in the formation of O_3.

Thus, in summarized way, Ozone concentrations vary negatively with NO_x i.e. the concentration of Ozone falls rapidly for higher values of NO_x. NO and NO_2 vary positively with NO_x i.e. the concentrations of NO and NO_2 increase with the increasing values of NO_x. NO is higher than NO_2 and Ozone when the concentration of NO_x is higher. These observations represent the complex chemical reactions wherein NO reacts with O_3 to form NO_2 which is called as NO **titration effect**.

From Figure 4.3.16 nitrogen-dependent and nitrogen-independent contributions are identified [2]. It shows the plot of OX (NO_2+Ozone) and NO_x to be fitting a linear regression model and nitrogen-dependent and nitrogen-independent contributions. NO_x-dependent region can be contributed to local pollution sources and photochemical reactions as the level of NO_x level’s increase/decrease leads to a simultaneous increase/decrease in the total oxidants level. NO_x-independent region can be contributed to level of ozone at background level as the concentration of NO_x is unaffected by the change in concentration of oxidants.

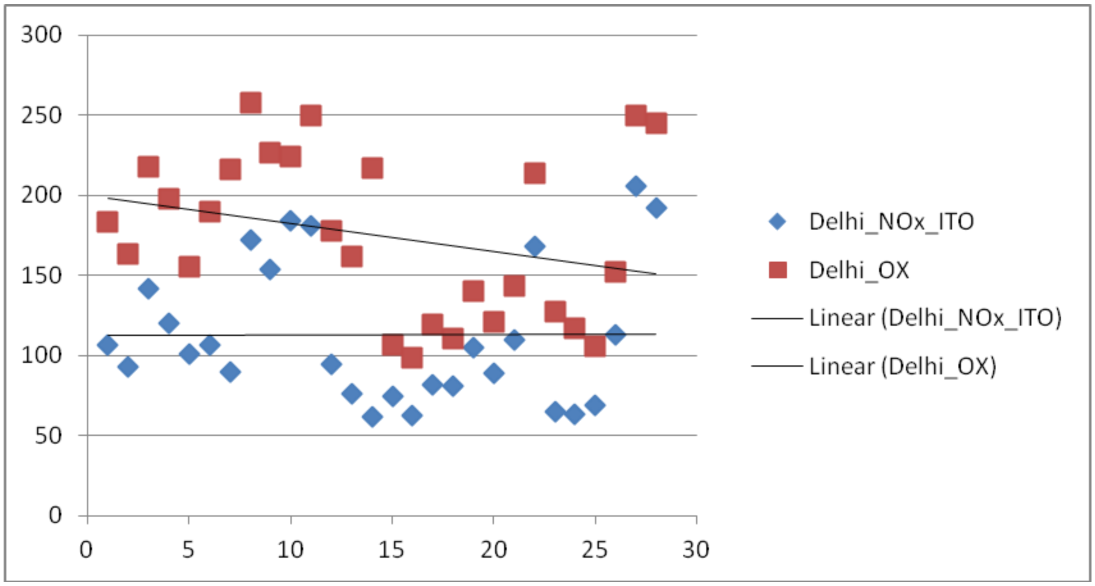

Figure 4.3.16: Nitrogen-dependent and Nitrogen-independent contributions

4.4 Application of PCA on Tier I cities Air Quality Data

Principal Component Analysis (PCA) has been applied to the four Tier I cities and the PCA results have been used to identify associations between various parameters of the dataset[1]. The factors being considered for checking and validating pollutant emiisions include seasonal and time-of-day factors alongwith photochemical activity leading to transformations producing Ozone and nitrogen dioxide .

The various steps involved in the principal component analysis [1] are described below. First step is to find the covariance of normalized table (it is the rotated version of COEFF from princomp method of MATLAB). If the resultant matrix S is symmetrical, it implies that eigen vectors are orthogonal. PCA identifies the principal components that are perpendicular to each other, hence knowing about orthogonality is important

Table 4.4.1: Delhi Covariance Matrix

Delhi Covariance matrix										
	Delhi_PM25_ITO	Delhi_Rhumidity_AV	Delhi_SO2_ITO	Delhi_PM10_ITO	Delhi_O3_ITO	Delhi_NOx_ITO	Delhi_NO2_ITO	Delhi_NO_ITO	Delhi_CO_ITO	Delhi_Temperature_AV
Delhi_PM25_ITO	1	0.18	0.12	0.09	-0.14	0.26	0.17	0.26	0.2	0.06
Delhi_Rhumidity_AV	0.18	1	0.16	0.18	-0.02	0.13	0.02	0.2	-0.06	0.07
Delhi_SO2_ITO	0.12	0.16	1	0.86	-0.07	0.27	0.16	0.22	0.16	-0.12
Delhi_PM10_ITO	0.09	0.18	0.86	1	-0.15	0.25	0.16	0.2	0.11	-0.09
Delhi_O3_ITO	-0.14	-0.02	-0.07	-0.15	1	-0.14	-0.09	-0.14	-0.12	-0.07
Delhi_NOx_ITO	0.26	0.13	0.27	0.25	-0.14	1	0.76	0.7	0.36	-0.18
Delhi_NO2_ITO	0.17	0.02	0.16	0.16	-0.09	0.76	1	0.42	0.27	-0.15
Delhi_NO_ITO	0.26	0.2	0.22	0.2	-0.14	0.7	0.42	1	0.29	-0.17
Delhi_CO_ITO	0.2	-0.06	0.16	0.11	-0.12	0.36	0.27	0.29	1	-0.06
Delhi_Temperature_AV	0.06	0.07	-0.12	-0.09	-0.07	-0.18	-0.15	-0.17	-0.06	1

Table 4.4.2: Mumbai Covariance Matrix

	Mumbai Covariance Matrix								
	Mumbai_PM25_Bandra	Mumbai_Rhumidity_Bandra	Mumbai_SO2_Bandra	Mumbai_PM10_Bandra	Mumbai_O3_Bandra	Mumbai_NOx_Bandra	Mumbai_NO2_Bandra	Mumbai_CO_Bandra	Mumbai_Temp_Bandra
Mumbai_PM25_Bandra	1.0003	0.0707	-0.0159	0.5451	-0.0905	0.1388	0.2263	-0.0548	-0.0969
Mumbai_Rhumidity_Bandra	0.0707	1	-0.0119	-0.2797	-0.0036	-0.1318	0.1496	-0.0495	-0.014
Mumbai_SO2_Bandra	-0.0159	-0.0119	1.0005	0.0891	0.0254	0.165	0.0622	0.0545	-0.0718
Mumbai_PM10_Bandra	0.5451	-0.2797	0.0891	0.9997	0.1048	0.0762	0.0329	0.1121	-0.325
Mumbai_O3_Bandra	-0.0905	-0.0036	0.0254	0.1048	1.0005	-0.1467	0.0975	0.0988	-0.169
Mumbai_NOx_Bandra	0.1388	-0.1318	0.165	0.0762	-0.1467	1.0002	-0.0666	-0.0404	0.0241
Mumbai_NO2_Bandra	0.2263	0.1496	0.0622	0.0329	0.0975	-0.0666	0.9995	0.0518	-0.1966
Mumbai_CO_Bandra	-0.0548	-0.0495	0.0545	0.1121	0.0988	-0.0404	0.0518	1.0001	-0.1166
Mumbai_Temp_Bandra	-0.0969	-0.014	-0.0718	-0.325	-0.169	0.0241	-0.1966	-0.1166	1.0001

Table 4.4.3: Bengaluru Covariance Matrix

	Bengaluru Covariance Matrix								
	Blr_PM25_BTM	Blr_Rhumidity_BTM	Blr_SO2_BTM	Blr_PM10_CRS	Blr_NOx_BTM	Blr_NO2_BTM	Blr_NO_BTM	Blr_CO_BTM	Blr_Temp_BTM
Blr_PM25_BTM	0.9999	-0.1432	0.0125	0.0325	0.086	0.1195	0.3156	0.1894	0.1636
Blr_Rhumidity_BTM	-0.1432	0.9996	0.008	-0.0999	-0.1994	-0.1192	-0.3804	0.1585	-0.5179
Blr_SO2_BTM	0.0125	0.008	0.9998	0.0087	-0.0116	0.018	0.003	-0.0256	0.0042
Blr_PM10_CRS	0.0325	-0.0999	0.0087	1.0001	-0.0094	-0.0175	0.0351	-0.0161	0.009
Blr_NOx_BTM	0.086	-0.1994	-0.0116	-0.0094	0.9998	0.1735	0.1289	-0.0366	0.1752
Blr_NO2_BTM	0.1195	-0.1192	0.018	-0.0175	0.1735	1.0004	0.3597	0.0505	-0.0102
Blr_NO_BTM	0.3156	-0.3804	0.003	0.0351	0.1289	0.3597	0.9998	0.0054	0.2238
Blr_CO_BTM	0.1894	0.1585	-0.0256	-0.0161	-0.0366	0.0505	0.0054	0.9961	0.0015
Blr_Temp_BTM	0.1636	-0.5179	0.0042	0.009	0.1752	-0.0102	0.2238	0.0015	1.0019

Table 4.4.4: Chennai Covariance Matrix

	Chennai Covariance Matrix								
	Cheni_PM25_IIT	Cheni_RHumidity_Alandur	Cheni_SO2_IIT	Cheni_O3_Alandur	Cheni_NOx_IIT	Cheni_NO2_IIT	Cheni_NO_IIT	Cheni_CO_IIT	Cheni_Temp_IIT
Cheni_PM25_IIT	0.9999	0.1182	0.0404	-0.0829	-0.0528	-0.1568	-0.1436	0.092	0.112
Cheni_RHumidity_Alandur	0.1182	1.0002	-0.0057	-0.0907	0.0053	-0.0169	-0.0486	-0.0069	-0.049
Cheni_SO2_IIT	0.0404	-0.0057	1.0001	0.1242	-0.1103	0.0289	0.0871	-0.3004	0.6878
Cheni_O3_Alandur	-0.0829	-0.0907	0.1242	1.0002	-0.0263	0.0184	-0.0048	-0.0638	0.0937
Cheni_NOx_IIT	-0.0528	0.0053	-0.1103	-0.0263	1.0001	0.5476	0.4929	0.0526	-0.1443
Cheni_NO2_IIT	-0.1568	-0.0169	0.0289	0.0184	0.5476	1	0.7224	0.0696	-0.0546
Cheni_NO_IIT	-0.1436	-0.0486	0.0871	-0.0048	0.4929	0.7224	0.9999	-0.1058	-0.006
Cheni_CO_IIT	0.092	-0.0069	-0.3004	-0.0638	0.0526	0.0696	-0.1058	1	-0.2309
Cheni_Temp_IIT	0.112	-0.049	0.6878	0.0937	-0.1443	-0.0546	-0.006	-0.2309	0.9999

As can be seen from Tables above, covariance matrices of all Tier I cities are symmetrical, hence their eigen vectors are orthogonal. The correlation matrices of four Tier I cities below shows the parameters with correlation greater than 0.5 and less than -0.5 shaded in red. Table 4.4.9 enlists the various positive and negative correlation of various pollutants across Tier I cities as obtained from above tables of correlation matrices. Second step is to determine the Eigen values to identify the number of principal components to be considered. After performing pca on data COEFF matrix, loadings matrix, is obtained. Suppose there are n numbers of variables then COEFF is an nXn matrix.

Table 4.4.5: Delhi Correlation Matrix

Delhi Correlation Matrix										
	Delhi_PM25_ITO	Delhi_Rhumidity_AV	Delhi_SO2_ITO	Delhi_PM10_ITO	Delhi_O3_ITO	Delhi_NOx_ITO	Delhi_NO2_ITO	Delhi_NO_ITO	Delhi_CO_ITO	Delhi_Temperature_AV
Delhi_PM25_ITO	1	0.1784	0.1243	0.0906	-0.1399	0.2587	0.1687	0.2606	0.1958	0.0578
Delhi_Rhumidity_AV	0.1784	1	0.164	0.1828	-0.0228	0.1321	0.0192	0.1978	-0.0607	0.0749
Delhi_SO2_ITO	0.1243	0.164	1	0.8592	-0.075	0.2677	0.1605	0.2153	0.1606	-0.1216
Delhi_PM10_ITO	0.0906	0.1828	0.8592	1	-0.1492	0.2541	0.1621	0.2023	0.1118	-0.0859
Delhi_O3_ITO	-0.1399	-0.0228	-0.075	-0.1492	1	-0.137	-0.0852	-0.135	-0.1182	-0.0658
Delhi_NOx_ITO	0.2587	0.1321	0.2677	0.2541	-0.137	1	0.7573	0.6986	0.359	-0.1846
Delhi_NO2_ITO	0.1687	0.0192	0.1605	0.1621	-0.0852	0.7573	1	0.4201	0.2738	-0.1529
Delhi_NO_ITO	0.2606	0.1978	0.2153	0.2023	-0.135	0.6986	0.4201	1	0.2948	-0.1739
Delhi_CO_ITO	0.1958	-0.0607	0.1606	0.1118	-0.1182	0.359	0.2738	0.2948	1	-0.0644
Delhi_Temperature_AV	0.0578	0.0749	-0.1216	-0.0859	-0.0658	-0.1846	-0.1529	-0.1739	-0.0644	1

Table 4.4.6: Mumbai Correlation Matrix

Mumbai Correlation Matrix									
	Mumbai_PM25_Bandra	Mumbai_Rhumidity_Bandra	Mumbai_SO2_Bandra	Mumbai_PM10_Bandra	Mumbai_O3_Bandra	Mumbai_NOx_Bandra	Mumbai_NO2_Bandra	Mumbai_CO_Bandra	Mumbai_Temp_Bandra
Mumbai_PM25_Bandra	1	0.0707	-0.0159	0.5451	-0.0905	0.1387	0.2264	-0.0547	-0.0969
Mumbai_Rhumidity_Bandra	0.0707	1	-0.0119	-0.2797	-0.0036	-0.1318	0.1497	-0.0495	-0.014
Mumbai_SO2_Bandra	-0.0159	-0.0119	1	0.089	0.0254	0.1649	0.0622	0.0545	-0.0718
Mumbai_PM10_Bandra	0.5451	-0.2797	0.089	1	0.1048	0.0762	0.033	0.1121	-0.3251
Mumbai_O3_Bandra	-0.0905	-0.0036	0.0254	0.1048	1	-0.1467	0.0975	0.0987	-0.169
Mumbai_NOx_Bandra	0.1387	-0.1318	0.1649	0.0762	-0.1467	1	-0.0666	-0.0404	0.0241
Mumbai_NO2_Bandra	0.2264	0.1497	0.0622	0.033	0.0975	-0.0666	1	0.0518	-0.1966
Mumbai_CO_Bandra	-0.0547	-0.0495	0.0545	0.1121	0.0987	-0.0404	0.0518	1	-0.1166
Mumbai_Temp_Bandra	-0.0969	-0.014	-0.0718	-0.3251	-0.169	0.0241	-0.1966	-0.1166	1

Table 4.4.7: Bengaluru Correlation Matrix

Bengaluru Correlation Matrix									
	Blr_PM25_BTM	Blr_Rhumidity_BTM	Blr_SO2_BTM	Blr_PM10_CRS	Blr_NOx_BTM	Blr_NO2_BTM	Blr_NO_BTM	Blr_CO_BTM	Blr_Temp_BTM
Blr_PM25_BTM	1	-0.1433	0.0125	0.0325	0.086	0.1195	0.3156	0.1898	0.1634
Blr_Rhumidity_BTM	-0.1433	1	0.008	-0.0999	-0.1994	-0.1192	-0.3805	0.1588	-0.5175
Blr_SO2_BTM	0.0125	0.008	1	0.0087	-0.0116	0.018	0.003	-0.0257	0.0042
Blr_PM10_CRS	0.0325	-0.0999	0.0087	1	-0.0094	-0.0175	0.0351	-0.0162	0.0089
Blr_NOx_BTM	0.086	-0.1994	-0.0116	-0.0094	1	0.1735	0.129	-0.0366	0.175
Blr_NO2_BTM	0.1195	-0.1192	0.018	-0.0175	0.1735	1	0.3597	0.0506	-0.0102
Blr_NO_BTM	0.3156	-0.3805	0.003	0.0351	0.129	0.3597	1	0.0054	0.2236
Blr_CO_BTM	0.1898	0.1588	-0.0257	-0.0162	-0.0366	0.0506	0.0054	1	0.0015
Blr_Temp_BTM	0.1634	-0.5175	0.0042	0.0089	0.175	-0.0102	0.2236	0.0015	1

Table 4.4.8: Chennai Correlation Matrix

Chennai Correlation Matrix									
	Cheni_PM25_IIT	Cheni_RHumidity_Alandur	Cheni_SO2_IIT	Cheni_O3_Alandur	Cheni_NOx_IIT	Cheni_NO2_IIT	Cheni_NO_IIT	Cheni_CO_IIT	Cheni_Temp_IIT
Cheni_PM25_IIT	1	0.1182	0.0404	-0.0829	-0.0528	-0.1568	-0.1436	0.092	0.112
Cheni_RHumidity_Alandur	0.1182	1	-0.0057	-0.0907	0.0053	-0.0169	-0.0486	-0.0069	-0.049
Cheni_SO2_IIT	0.0404	-0.0057	1	0.1242	-0.1103	0.0289	0.0871	-0.3004	0.6878
Cheni_O3_Alandur	-0.0829	-0.0907	0.1242	1	-0.0263	0.0184	-0.0048	-0.0637	0.0937
Cheni_NOx_IIT	-0.0528	0.0053	-0.1103	-0.0263	1	0.5476	0.4929	0.0526	-0.1443
Cheni_NO2_IIT	-0.1568	-0.0169	0.0289	0.0184	0.5476	1	0.7224	0.0696	-0.0546
Cheni_NO_IIT	-0.1436	-0.0486	0.0871	-0.0048	0.4929	0.7224	1	-0.1058	-0.006
Cheni_CO_IIT	0.092	-0.0069	-0.3004	-0.0637	0.0526	0.0696	-0.1058	1	-0.2309
Cheni_Temp_IIT	0.112	-0.049	0.6878	0.0937	-0.1443	-0.0546	-0.006	-0.2309	1

Table 4.4.9: Positive and Negative correlation of various pollutants across cities

CITY	Positive Correlation	References for Positive Correlation	Negative Correlation	References for Negative Correlation
Delhi	SO_2, PM10	[6]	None	None
	NO_2, NO_x	[2]		
	NO, NO_x	[2]		
	NO, NO_2	[2]		
	CO, NO_x	[2]		
Mumbai	PM10, PM2.5	[17], [18]	None	None
Bengaluru	NO, PM2.5	[18]	Temperature, Relative Humidity	[19]
			NO, Relative Humidity	[19]
Chennai	Temperature, SO_2	[2]	CO, SO_2	[2]
	NO_2, NO_x	[2]		
	NO, NO_x	[2]		
	NO, NO_2	[2]		

It is used to interpret the principal components. In order to analyze the closeness of two variables v1 and v2 , analyze the loadings matrix column wise and club together the values with high absolute value (e.g. criteria could be absolute value>0.5). For Delhi, Mumbai, Bengaluru and Chennai COEFF matrix is shown in Table 4.4.10, Table 4.4.11, Table 4.4.12 and Table 4.4.13.

Table 4.4.10: Delhi PCA Table

Components	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Delhi_PM25_ITO	0.2309	-0.0806	0.5001	0.0181	0.3452	-0.5296	-0.5351	-0.0483	-0.0153	-0.0339
Delhi_Rhumidity_AV	0.1431	0.2154	0.4444	0.6195	-0.1732	-0.0643	0.4475	-0.3419	-0.0122	0.0192
Delhi_SO2_ITO	0.343	0.576	-0.1486	-0.079	0.1395	0.0308	-0.0712	0.0361	0.1033	0.6963
Delhi_PM10_ITO	0.3346	0.5949	-0.112	-0.1047	0.0288	0.0863	-0.0741	0.0185	-0.1476	-0.689
Delhi_O3_ITO	-0.1441	-0.0246	-0.3563	0.5995	0.6769	0.131	-0.0676	0.0877	-0.0075	-0.0685
Delhi_NOx_ITO	0.4907	-0.2695	-0.0508	0.1023	-0.08	0.2161	-0.073	0.0242	0.7722	-0.1349
Delhi_NO2_ITO	0.3982	-0.3116	-0.1345	0.0467	-0.0711	0.3841	-0.2791	-0.4983	-0.4841	0.1046
Delhi_NO_ITO	0.4251	-0.2129	0.0527	0.1846	-0.1251	-0.0332	0.1626	0.7456	-0.367	0.0578
Delhi_CO_ITO	0.2771	-0.2024	-0.0085	-0.4093	0.5191	-0.1406	0.6226	-0.1864	-0.0415	-0.0356
Delhi_Temperature_AV	-0.1378	0.0692	0.606	-0.1584	0.2753	0.689	-0.0309	0.1784	0.0039	0.0234

Table 4.4.11: Mumbai PCA Table

Components	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Mumbai_PM25_Bandra	0.5093	0.1761	0.4594	-0.1813	0.0932	0.2129	-0.2209	0.1154	-0.5913
Mumbai_Rhumidity_Bandra	-0.144	-0.3862	0.5497	0.2292	0.1036	-0.2248	-0.4192	0.3951	0.2868
Mumbai_SO2_Bandra	0.1551	0.0836	-0.1493	0.7785	-0.1917	0.1282	-0.403	-0.3402	-0.0962
Mumbai_PM10_Bandra	0.6229	0.1667	-0.106	-0.2449	-0.0353	0.025	-0.2525	-0.0491	0.667
Mumbai_O3_Bandra	0.1403	-0.4586	-0.3561	-0.0429	-0.4833	0.3789	-0.0392	0.5012	-0.1066
Mumbai_NOx_Bandra	0.1298	0.5273	0.0247	0.4129	-0.0107	-0.0765	0.3992	0.5989	0.1
Mumbai_NO2_Bandra	0.2636	-0.3935	0.3687	0.2328	0.0697	0.3536	0.5907	-0.2654	0.1907
Mumbai_CO_Bandra	0.1547	-0.2232	-0.4185	0.1332	0.8277	0.0963	-0.0628	0.1704	-0.0774
Mumbai_Temp_Bandra	-0.4217	0.3033	0.1215	-0.0655	0.1376	0.7768	-0.1888	0.0546	0.2241

Table 4.4.12: Bengaluru PCA Table

Components	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Blr_PM25_BTM	-0.33	0.42	-0.35	0.05	0.09	-0.05	-0.68	0.32	-0.09
Blr_Rhumidity_BTM	0.52	0.36	0.09	0.01	0	0.12	0.16	-0.04	0.74
Blr_SO2_BTM	-0.01	0	0.14	0.69	0.69	0.13	-0.05	0.04	-0.02
Blr_PM10_CRS	-0.07	-0.13	-0.31	0.62	-0.62	0.31	-0.06	-0.04	0.1
Blr_NOx_BTM	-0.3	-0.06	0.33	-0.24	0.02	0.81	0.22	0.18	0.02
Blr_NO2_BTM	-0.3	0.4	0.56	0.1	-0.19	-0.06	-0.34	-0.52	0
Blr_NO_BTM	-0.5	0.21	0.13	0.12	-0.11	-0.35	0.03	0.67	0.31
Blr_CO_BTM	0.01	0.6	-0.45	-0.14	0.12	0.29	-0.51	0.2	-0.17
Blr_Temp_BTM	-0.44	-0.33	-0.34	-0.18	0.26	0.03	-0.28	-0.32	0.55

Table 4.4.13: Chennai PCA Table

Components	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Cheni_PM25_IIT	-0.18	-0.02	0.61	0.49	0.06	-0.56	-0.18	0.05	0.04
Cheni_RHumidity_Alandur	-0.04	-0.07	0.57	-0.48	0.59	0.28	0	-0.06	-0.04
Cheni_SO2_IIT	-0.11	0.63	0.12	0.07	-0.03	0.23	0.05	0.72	-0.04
Cheni_O3_Alandur	-0.02	0.19	-0.47	0.31	0.79	-0.15	-0.03	-0.03	-0.03
Cheni_NOx_IIT	0.52	0	0.15	0.09	0.06	-0.19	0.81	0.06	-0.09
Cheni_NO2_IIT	0.58	0.13	0.08	0.09	0.04	0.13	-0.29	-0.03	0.72
Cheni_NO_IIT	0.56	0.21	0.06	-0.03	-0.05	-0.06	-0.44	-0.1	-0.66
Cheni_CO_IIT	0.06	-0.39	0.11	0.61	0.04	0.65	0	0.06	-0.18
Cheni_Temp_IIT	-0.18	0.59	0.15	0.19	-0.11	0.23	0.18	-0.68	0.01

Third step is to interpret the coefficient for principal components. As the data is standardized, the relative magnitude of those coefficients within a column can be directly assessed. Each column here corresponds with a column in the output of the program labeled Eigen vectors. Significant values (i.e. the values greater than 0.4 and less than -0.4) are considered important.

In PCA first need to find the components with largest variances as these components contain the most information and hence can help to summarize the air pollution data. Principal component analysis results depend on the scale of input data [11], therefore, need to decide whether there is a need for rescaling. As mentioned in the data preprocessing section z-score scaling has been used on the input data.

The main objective of applying PCA on air quality data is to provide usable description in components understandable in terms of sources of air pollution. The parameters are normalized using z-score method before PCA application. In each, principal component is a linear combination of standardized concentrations. The vector $a_i=\{a_{i,1},a_{i,2}\ldots a_{i,10}\}$, for i, j in$\{1,2\ldots10\}$ are real constants and the constant $a_{i,j}$ are called loadings. The loadings are chosen such that they satisfy the following conditions. Firstly, the length of the vectors a_i equals 1. Secondly, the principal component with the highest variance is chosen as the first principal component. Thirdly, the component with the second maximum variance is chosen as the second principal component such that first and second principal components are uncorrelated. In general, the i_{th} principal component is chosen such that Var (PC_i) is maximal on the condition that PC_k and PC_i are uncorrelated for k in $\{1,2,\ldots i\text{-}1\}$.

Another output obtained from PCA of data is SCORE matrix. SCORE gives the actual principal components as calculated by princomp method. It gives the observation in terms of latent variables. The magnitudes of the coefficients give the contributions of each variable to that component. As the data has been standardized, they do not depend on the variances of the corresponding variables. Another output obtained from PCA is the LATENT matrix. It gives the Eigen value of covariance of principal components.

Table 4.4.14: Delhi LATENT matrix

Component	Eigenvalue	Proportion	Cumulative
1	3.04	0.3	0.3
2	1.57	0.16	0.46
3	1.22	0.12	0.58
4	1.03	0.1	0.68
5	0.85	0.09	0.77
6	0.8	0.08	0.85
7	0.68	0.07	0.92
8	0.52	0.05	0.97
9	0.15	0.02	0.99
10	0.13	0.01	1
Total	10		

Table 4.4.15: Mumbai LATENT matrix

Component	Eigenvalue	Proportion	Cumulative
1	1.83	0.2	0.2
2	1.4	0.16	0.36
3	1.25	0.14	0.5
4	1.09	0.12	0.62
5	0.9	0.1	0.72
6	0.77	0.09	0.8
7	0.76	0.08	0.89
8	0.74	0.08	0.97
9	0.27	0.03	1
Total	9.01		

Table 4.4.16: Bengaluru LATENT matrix

Component	Eigenvalue	Proportion	Cumulative
1	2.1	0.23	0.23
2	1.27	0.14	0.37
3	1.07	0.12	0.49
4	1.02	0.11	0.61
5	0.99	0.11	0.72
6	0.88	0.1	0.81
7	0.74	0.08	0.9
8	0.52	0.06	0.95
9	0.41	0.05	1
Total	9		

Table 4.4.17: Chennai LATENT matrix

Component	Eigenvalue	Proportion	Cumulative
1	2.24	0.25	0.25
2	1.9	0.21	0.46
3	1.18	0.13	0.59
4	0.97	0.11	0.7
5	0.9	0.1	0.8
6	0.75	0.08	0.88
7	0.51	0.06	0.94
8	0.3	0.03	0.97
9	0.25	0.03	1
Total	9		

The proportion of variation determined by each eigen value is given in the third column. The first five principal components in Table 4.4.14 explain 77% of the variation. This is an acceptably large percentage. An Alternative Method to determine the number of principal components is using Eigen values Plot.

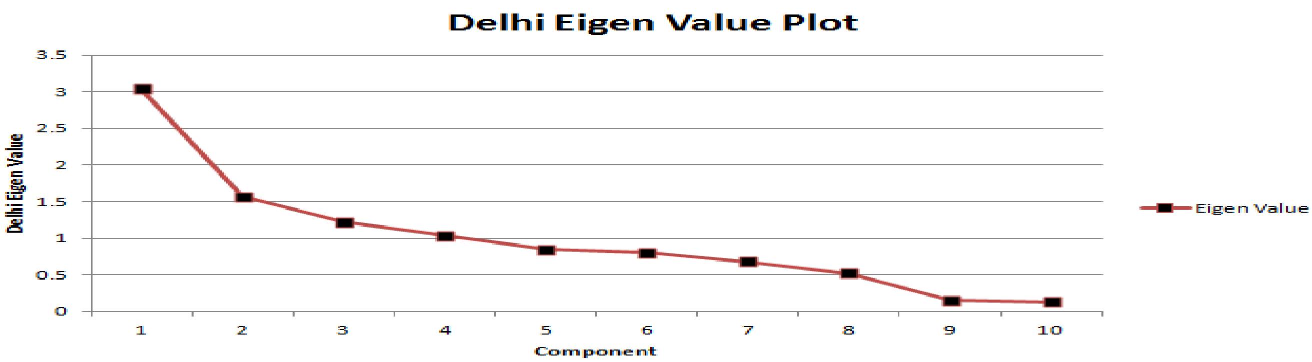

Figure 4.4.1: Delhi Eigen Value Plot

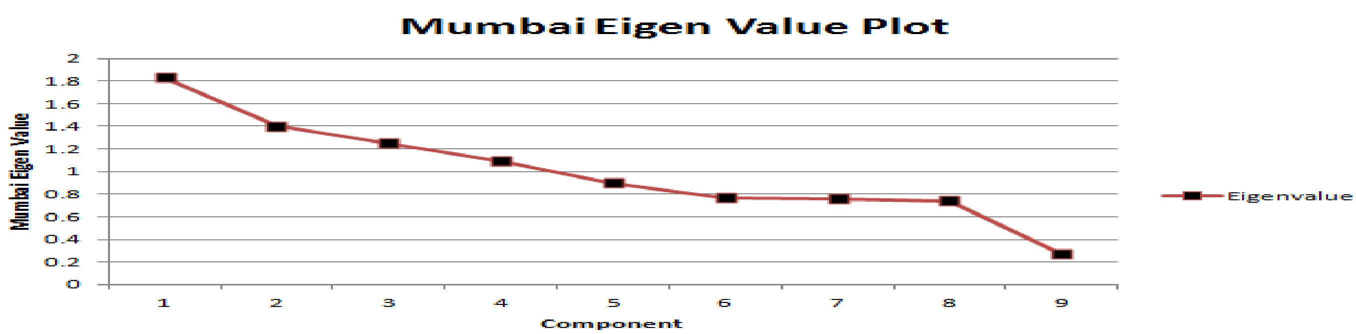

Figure 4.4.2: Mumbai Eigen Value Plot

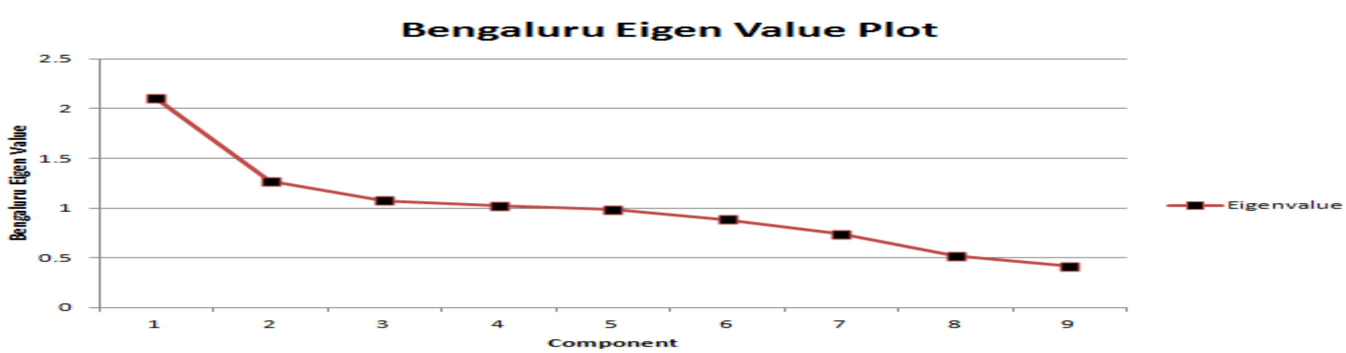

Figure 4.4.3: Bengaluru Eigen Value Plot

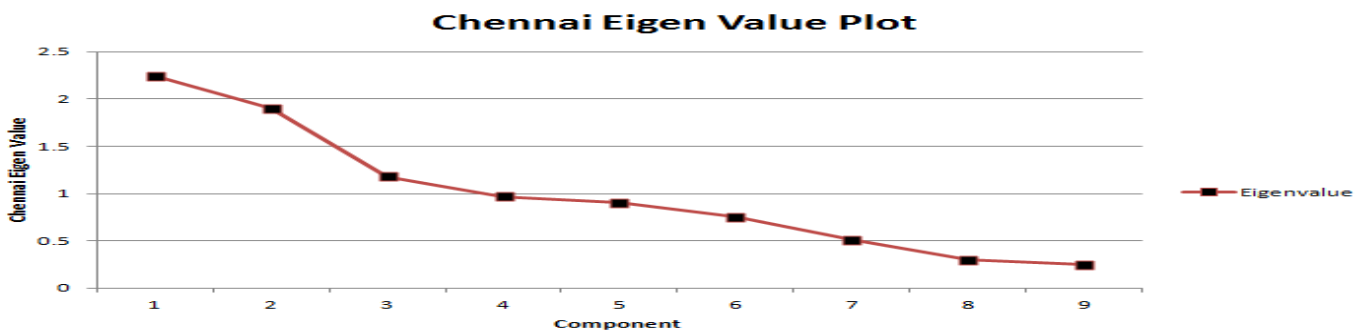

Figure 4.4.4: Chennai Eigen Value Plot

4.5 Factor Analysis of standardized Principal Components

In order to use principal components for the summarization of air pollution data include the first principal components which represent maximum variance in the data and the rotation of the principal components. The purpose of the rotation is to be able to find variables that can represent different sources of air pollution.

For Delhi, by interpreting the eigen values in Table 4.4.14, it has been found that first 5 eigen values represent about 77% of the data, it is considered as a good amount of variance. After examining the correlation loadings of component in Table 4.4.10 need to apply varimax rotation on the first 5 components. The rotation is chosen in a way that sum of new five squared correlation loadings as shown in Table 4.5.1 is maximal.

Table 4.5.1: Varimax Rotations of components of Delhi

Varimax Rotations of first five standardized components of Delhi					
	rComp1	rComp2	rComp3	rComp4	rComp5
Delhi_PM25_ITO	0.139942	0.007467	0.637641	0.009933	0.055625
Delhi_Rhumidity_AV	0.13468	0.074037	0.247658	0.766082	0.079727
Delhi_SO2_ITO	-0.01463	0.702894	-0.00072	-0.02566	0.047687
Delhi_PM10_ITO	-0.02303	0.696347	-0.02767	0.025462	-0.05829
Delhi_O3_ITO	-0.04905	-0.01556	-0.01982	0.005619	0.981208
Delhi_NOx_ITO	0.576393	0.014121	0.005993	-0.0163	-0.01027
Delhi_NO2_ITO	0.513666	-0.04513	-0.06983	-0.10097	0.00173
Delhi_NO_ITO	0.509665	-0.00683	0.047978	0.125737	-0.02416
Delhi_CO_ITO	0.164777	0.094754	0.352706	-0.62059	0.095927
Delhi_Temperature_AV	-0.27593	-0.06324	0.631906	0.016831	-0.11052

rComp1 is mostly correlated with NO_x, NO_2 and NO. rComp2 is related with SO_2 and PM10. rComp3 is mostly related with PM25 and Temperature. rComp4 is mostly related with Relative Humidity. rComp5 is related with Ozone.

For Mumbai, by interpreting the eigen values in Table 4.4.15, it has been observed that the first 5 eigen values represent about 72% of the data, it is a considerable good amount of variance. After examining the correlation loadings of component in Table 4.4.11 need to apply varimax rotation on the first 5 components. The rotation is chosen in a way that sum of new five squared correlation loadings as shown in Table 4.5.2 is maximal.

Table 4.5.2: Varimax Rotations of components of Mumbai

Varimax Rotations of first five standardized components of Mumbai					
	rComp1	rComp2	rComp3	rComp4	rComp5
Mumbai_PM25_Bandra	0.684453	0.147113	0.210275	-0.02786	-0.08856
Mumbai_Rhumidity_Banc	-0.13458	0.107065	0.705586	-0.06426	-0.06147
Mumbai_SO2_Bandra	-0.14823	-0.20855	0.108656	0.786105	0.030962
Mumbai_PM10_Bandra	0.620733	-0.21603	-0.22738	0.015447	0.066091
Mumbai_O3_Bandra	-0.14538	-0.73675	-0.07451	-0.08569	-0.12427
Mumbai_NOx_Bandra	0.139202	0.256374	-0.12533	0.602148	-0.05191
Mumbai_NO2_Bandra	0.152999	-0.16674	0.600174	0.052465	0.071556
Mumbai_CO_Bandra	-0.02666	0.022958	0.003816	-0.02151	0.974704
Mumbai_Temp_Bandra	-0.20371	0.489763	-0.11376	-0.06127	-0.09872

rComp1 is mostly correlated with PM2.5 and PM10. rComp2 is related with Ozone. rComp3 is mostly related with Humidity and NO_2. rComp4 is mostly related with SO_2 and NO_x. rComp5 is related with Carbon Monoxide.

For Bengaluru, by interpreting the eigen values in Table 4.4.16, it is found that the first 5 eigen values represent about 72% of the data, it is considered as a good amount of variance. After examining the correlation loadings of component in Table 4.4.12 need to apply varimax rotation on the first 5 components. The rotation is chosen in a way that sum of new five squared correlation loadings as shown in Table 4.5.3 is maximal.

Table 4.5.3: Varimax Rotations of components of Bengaluru

Varimax Rotations of first five standardized components of Bengaluru					
	rComp1	rComp2	rComp3	rComp4	rComp5
Blr_PM25_BTM	-0.14109	0.60776	0.153319	0.070704	0.057644
Blr_Rhumidity_BTM	0.608764	0.105011	-0.13851	-0.08403	0.009903
Blr_SO2_BTM	0.000291	-0.01063	0.002969	-0.00843	0.988457
Blr_PM10_CRS	-0.02715	-0.03407	0.034947	0.940561	-0.03369
Blr_NOx_BTM	-0.17507	-0.16783	0.32876	-0.27596	-0.11269
Blr_NO2_BTM	0.238383	-0.02763	0.741111	-0.04489	0.011285
Blr_NO_BTM	-0.17025	0.152794	0.523268	0.092492	0.02808
Blr_CO_BTM	0.143941	0.742645	-0.08883	-0.06837	-0.06842
Blr_Temp_BTM	-0.68672	0.120538	-0.13057	-0.10871	0.010029

rComp1 is mostly correlated with Temperature and relative humidity (they are negatively correlated). rComp2 is related with PM2.5 and carbon monoxide. rComp3 is mostly related with NO and NO_2. rComp4 is mostly related with PM10. rComp5 is related with SO_2.

For Chennai, by interpreting the eigen values in Table 4.4.17, it has been found that the first 5 eigen values represent about 80% of the data which represents a very good amount of variance. After examining the correlation loadings of component in Table 4.4.13 need to apply varimax rotation on the first 5 components. The rotation is chosen in a way that sum of new five squared correlation loadings as shown in Table 4.5.4 is maximal.

Table 4.5.4: Varimax Rotations of components of Chennai

Varimax Rotations of first five standardized components of Chennai					
	rComp1	rComp2	rComp3	rComp4	rComp5
Cheni_PM25_IIT	-0.03526	0.239349	0.188344	0.737241	-0.10404
Cheni_RHumidity_Alandur	0.014915	-0.06267	0.955557	-0.00403	0.022919
Cheni_SO2_IIT	0.034718	0.649753	0.016928	-0.06104	0.033726
Cheni_O3_Alandur	-0.0077	0.038264	0.020236	0.001245	0.985645
Cheni_NOx_IIT	0.532468	-0.08	0.046471	0.097583	-0.00187
Cheni_NO2_IIT	0.609151	0.015629	-0.02057	0.017244	0.04049
Cheni_NO_IIT	0.582694	0.081537	-0.01869	-0.11436	-0.04645
Cheni_CO_IIT	0.051721	-0.24898	-0.20865	0.652518	0.110443
Cheni_Temp_IIT	-0.02178	0.663218	-0.06537	0.063669	-0.0042

rComp1 is mostly correlated with NO_2, NO_x and NO. rComp2 is related with SO_2 and temperature. rComp3 is mostly related with relative humidity. rComp4 is mostly related with PM2.5 and carbon monoxide. rComp5 is related with Ozone. Below Table 4.5.5 summarizes the various components identified after application of PCA on Tier I cities.

Table 4.5.5: Component Identification of Tier I cities

City	Component 1	Component 2	Component 3	Component 4	Component 5
Delhi	NO_x,NO_2, NO	SO_2, PM10	PM25 and Temperature	Relative Humidity	Ozone
Mumbai	PM2.5, PM10	Ozone	Humidity, NO_2	SO_2, NO_x	Carbon Monoxide
Bengaluru	Temperature, Relative Humidity(they are negatively correlated)	PM2.5, Carbon monoxide	NO, NO_2	PM10	SO_2
Chennai	NO_2, NO_x,NO	SO_2, Temperature	Relative Humidity	PM2.5, Carbon Monoxide	Ozone

4.6 ARIMA Model Fitting and Comparison with GARCH Model

ARIMA model and GARCH model has been applied on the data. Stationarity of parameter Delhi_NO2_ITO has been checked and the results are in Table 4.6.1. After a number of iterations, it is found that (1, 1, 2) as (p, d, q) is the combination with least AIC (Akaike information criteria) and BIC (Bayesian information criteria) [14] for the parameter Delhi_NO2_ITO. After that ARIMA model has been fit and the results are in Figure below. The ARIMA model results for Delhi_NO2_ITO , Mumbai_Temp_Bandra, Blr_Temp_BTM and Cheni_Temp_IIT and the results are in Figure 4.6.1, Figure 4.6.2, Figure 4.6.3 and Figure 4.6.4 respectively.

Table 4.6.1: Stationarity Test Results

	Stationary Tests			
	Delhi_NO2_ITO		diff(Delhi_NO2_ITO)	
	h	p-value	h	p-value
KPSS	1	0.01	0	0.1
LMC	1	0.01	0	0.1

ARIMA(1,1,2) Model			
Conditional Probability Distribution: Gaussian			
Parameter	Value	Standard Error	t Statistic
Constant	-0.00821	0.164407	-0.0499528
AR{1}	0.382661	0.0227648	16.8093
MA{1}	-0.53806	0.0225229	-23.8893
MA{2}	-0.18733	0.00935788	-20.0187
Variance	1054	6.60105	159.672

Figure 4.6.1: Delhi ARIMA Model Fitting

ARIMA(1,1,2) Model:			
Conditional Probability Distribution: Gaussian			
Parameter	Value	Standard Error	t Statistic
Constant	0.000409468	0.00284732	0.143808
AR{1}	0.586375	0.028311	20.712
MA{1}	-0.779516	0.0308246	-25.2888
MA{2}	-0.0858079	0.0184744	-4.6447
Variance	1.02167	0.0109988	92.8895

Figure 4.6.2: Mumbai ARIMA Model Fitting

ARIMA(1,1,1) Model:			
Conditional Probability Distribution: Gaussian			
Parameter	Value	Standard Error	t Statistic
Constant	0.000676451	0.00421754	0.16039
AR{1}	0.699356	0.0137095	51.0123
MA{1}	-0.903752	0.0098722	-91.5451
Variance	4.38177	0.0344597	127.156

Figure 4.6.3: Bengaluru ARIMA Model Fitting

ARIMA(1,1,1) Model:			
Conditional Probability Distribution: Gaussian			
Parameter	Value	Standard Error	t Statistic
Constant	-3.89E-07	0.000128916	-0.00302
AR{1}	0.747541	0.0282498	26.4618
MA{1}	-0.846595	0.0283606	-29.851
Variance	0.00156317	9.21E-06	169.651

Figure 4.6.4: Chennai ARIMA Model Fitting

Comparison with GARCH Model

Generalized Autoregressive Conditional Heteroscedasticity (GARCH) Model [10] has also been considered for model fitting of Delhi_NO2_ITO, Mumbai_Temp_Bandra, Blr_Temp_BTM and Cheni_Temp_IIT. In order to apply GARCH model and find the best fit, GARCH model is applied on diff of all parameters considered above for ARIMA model and the results obtained shown in Figure 4.6.5, Figure 4.6.6, Figure 4.6.7 and Figure 4.6.8 below are quite comparable [15] with ARIMA model fitting results.

Mean: ARMAX(1,2,0); Variance: Constant			
Conditional Probability Distribution: Gaussian			
Number of Model Parameters Estimated: 5			
Parameter	Value	Standard Error	t Statistic
C	-0.0083154	0.16443	-0.0506
AR(1)	0.3827	0.022771	16.8063
MA(1)	-0.53809	0.02253	-23.8838
MA(2)	-0.18732	0.0093609	-20.0107
K	1054.3	6.6013	159.7072

Figure 4.6.5: Delhi GARCH Model Fitting

Mean: ARMAX(1,2,0); Variance: Constant			
Conditional Probability Distribution: Gaussian			
Number of Model Parameters Estimated: 5			
Parameter	Value	Standard Error	t Statistic
C	0.000408	0.0028482	0.1433
AR(1)	0.58635	0.028322	20.7031
MA(1)	-0.77949	0.030837	-25.2776
MA(2)	-0.08582	0.018482	-4.6436
K	1.0221	0.011006	92.8602

Figure 4.6.6: Mumbai GARCH Model Fitting

Mean: ARMAX(1,2,0); Variance: Constant			
Conditional Probability Distribution: Gaussian			
Number of Model Parameters Estimated: 5			
Parameter	Value	Standard Error	t Statistic
C	0.00054	0.0032602	0.1656
AR(1)	0.75852	0.021013	36.0976
MA(1)	-0.99392	0.023802	-41.7582
MA(2)	0.068011	0.014913	4.5606
K	4.3709	0.034571	126.4312

Figure 4.6.7: Bengaluru GARCH Model Fitting

Mean: ARMAX(1,1,0); Variance: Constant			
Conditional Probability Distribution: Gaussian			
Number of Model Parameters Estimated: 5			
Parameter	Value	Standard Error	t Statistic
C	-3.51E-07	0.00012873	-0.0027
AR(1)	0.74776	0.028054	26.6538
MA(1)	-0.84677	0.028091	-30.144
K	0.001564	9.18E-06	170.4357

Figure 4.6.8: Chennai GARCH Model Fitting

4.7 Proposed SDA

ARIMA model has been applied on static data in many research works [9] [12] but it has not been simulated for streaming data. In this work ARIMA model has been applied on streaming air quality data and the results obtained are quite comparable with static data ARIMA. The window size selected for the parameters is 350 records as this window size was found to be giving the best results. For Delhi_NO2_ITO the plot of AR and MA coefficients is shown in Figure 4.7.1. It shows the AR and MA parameters vary inversely. Also, the result obtained after applying SDA on Delhi_NO2_ITO are same as the static ARIMA results obtained.

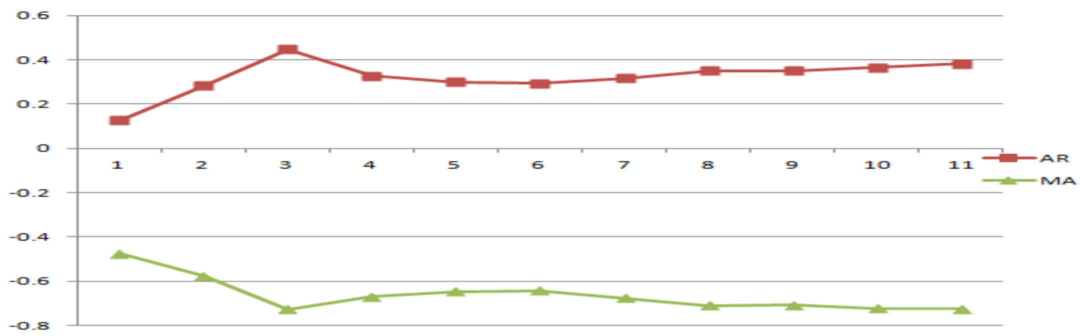

Figure 4.7.1: AR and MA plot for SDA on Delhi NO_2 data

The proposed SDA has also been applied on parameters Delhi_Temperature_AV, Mumbai_Temp_Bandra, Blr_Temp_BTM and Cheni_Temp_IIT of Delhi, Mumbai, Bengaluru and Chennai. For Delhi_Temperature_AV ARIMA(1,1,2) has been found to be the best fitting model for static data of parameter Delhi_Temperature_AV for a total of 2662 records for period ranging from 2009-16 and the window size of 350 records at a time is considered. The plot of AR and MA coefficients of Delhi_Temperature_AV for ARIMA model fitted during the process are shown in Figure 4.7.2 below. The iteration 4 in Figure 4.7.2 shows a ARIMA (0, 1, 0) model.

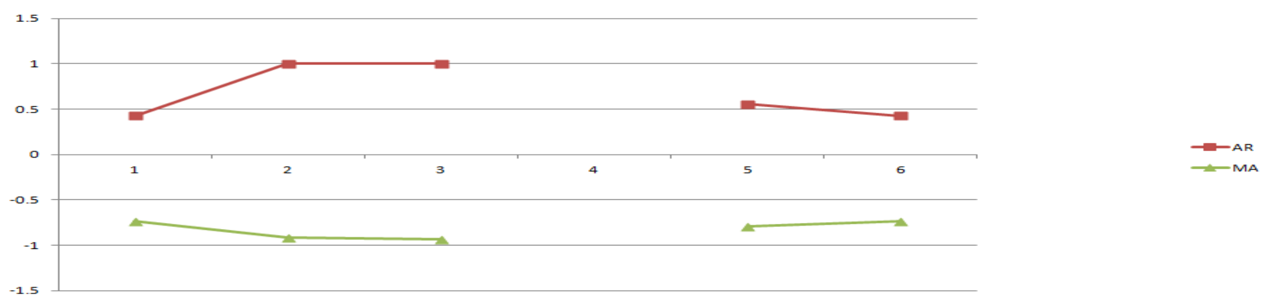

Figure 4.7.2: Plot of AR and MA for SDA on Delhi_Temperature_AV

ARIMA(1,1,2) fits the model for static data of parameter Mumbai_Temp_Bandra for a total of 3758 records for period ranging from 2006-16 and the window size of 350 records at a time is found to be the optimal window size to be considered. The plot of AR and MA coefficients of Mumbai_Temp_Bandra for ARIMA model fitted during the process are shown in Figure 4.7.3 below.

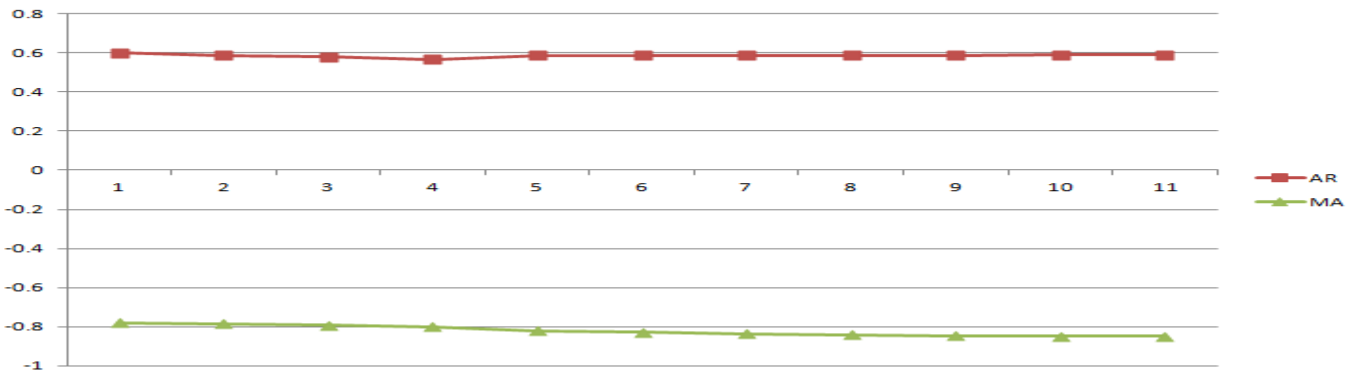

Figure 4.7.3: Plot of AR and MA for SDA on Mumbai_Temp_Bandra

ARIMA (1, 1, 1) fits the model for static data of parameter Blr_Temp_BTM for a total of 2403 records for period ranging from 2009-16 and the window size of 350 records at a time is found to be the optimal window size to be considered. The plot of AR and MA coefficients of Blr_Temp_BTM for ARIMA model fitted during the process are shown in Figure 4.7.4.

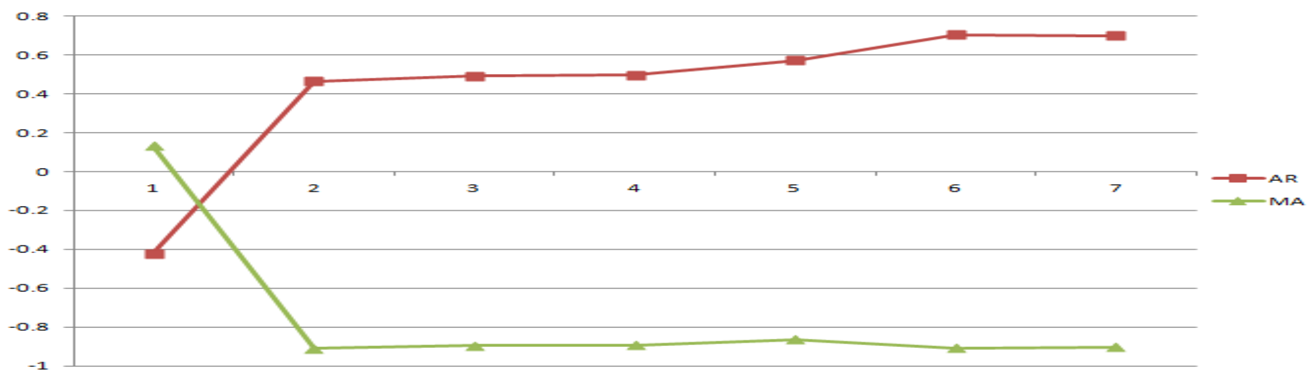

Figure 4.7.4: Plot of AR and MA for SDA on Blr_Temp_BTM

In order to fit the model for Cheni_Temp_IIT the parameter has been stationarized by using log operation on Cheni_Temp_IIT. After that ARIMA (1, 1, 1) fits the model for static data of parameter Blr_Temp_BTM for a total of 2430 records for period ranging from 2009-16 and the window size of 350 records at a time is found to be the optimal window size to be considered. The plot of AR and MA coefficients of Cheni_Temp_IIT for ARIMA model fitted during the process are shown in Figure 4.7.5.

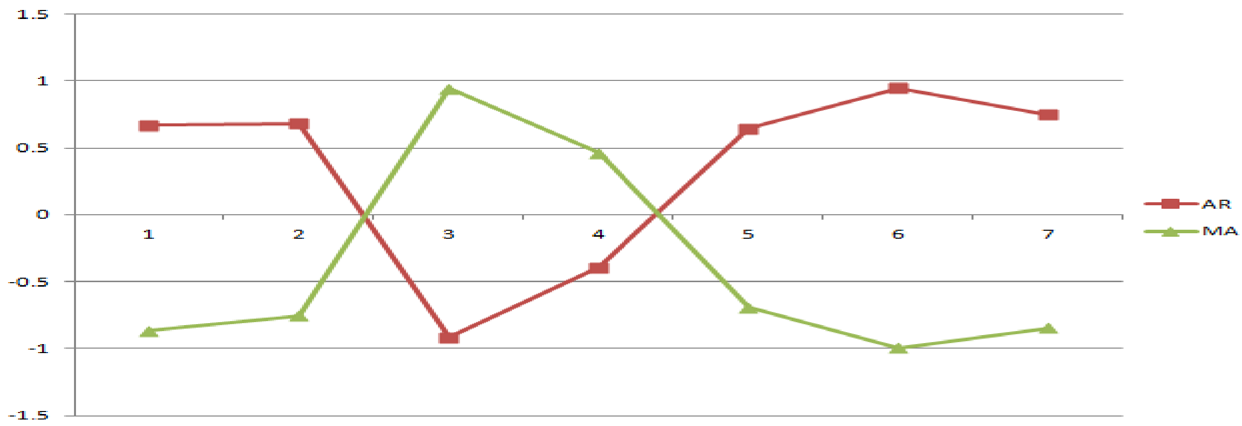

Figure 4.7.5: Plot of AR and MA for SDA on Cheni_Temp_IIT

The above figures show how AR and MA parameters vary over the changing window size for temperature parameter of four Tier I cities. The results obtained after applying SDA and ARIMA o static whole dataset are found to be quite comparable and hence SDA can be used for streaming time series data.

CONCLUSION AND FUTURE WORK

In this report, the air quality data of four Tier I cities (Delhi, Mumbai, Bengaluru and Chennai) in India have been analyzed using both graphical and statistical techniques. The air quality data has a large number of dimensions so in order to reduce dimensions, after experimentation, principal component analysis has been found to be the best technique to reduce dimensionality and identify the hidden associations between various pollutants depending on the region. The components identified from principal component analysis have been applied with varimax rotation to give a more clear description of identified components in terms of sources of air pollution.

Autoregressive Integrated Moving Average (ARIMA), implemented by Box-Jenkins approach, along with GARCH model has been used to predict air quality data. In this work, SDA has been proposed in which ARIMA model has been used on the streaming data of fixed window size and the results obtained are quite satisfactory and hence provide a way to use this model on other kinds of data as well.

In this study few parameters have been used for SDA, so future scope would be to use SDA on more parameters and compare their results with ARIMA on static time sequence. Also, this work has done trend analysis only for Tier I cities in India so future work would be to use discussed techniques for Tier II and Tier III cities as well. Also datasets of other fields like medical, social network, census, etc. can be used for SDA. Also GARCH model can be used for streaming data.

LIST OF PUBLICATIONS

[1] Yadav, Alka.,&Toshniwal, Durga. (2016). *Extracting Patterns and Variations in Air Quality of Four Tier I Cities in India*. IEEE Transactions on Knowledge and Data Engineering. [Communicated]

[2] Yadav, Alka.(2016). *Trend Analysis of Air Quality of Four Tier I Cities in India*. The 4th International Workshop on Big Data. [Communicated]

REFERENCES

[1]Derwent, R., G., & Middleton, D.R.(1995). *Analysis and Interpretation of Air Quality Data from an urban roadside location in Central London over the period from July 1991 to July 1992.* Atmospheric Environment Vol. 29, No. 8, pp. 923-946.

[2]Tiwari, Suresh., Dahiya, Anita., & Kumar, Nandini. (2015). *Investigation into relationships among NO, NO_2, NO_x, O_3, and CO at an urban background site in Delhi, India*. Atmospheric Research Vol. 157, pp. 119-126.

[3]Guttikunda,Sarath., & Jawahar,Puja.(2011). Urban Air Pollution Analysis in India. UrbanEmissions.Info, New Delhi, India.

[4]Mansouri, B., Hoshyari, E., Mansouri,A. (2011). *Study on ambient concentrations of air quality parameters (O_3, SO_2, CO and PM10) in different months in Shiraz city, Iran.* International Journal Of Environmental Sciences Vol. 1, No. 7, pp. 0976 – 4402.

[5]Qin, Jun., Fang, Hong., Chen, Zheng-Hong., Zhai, Hong-Nan., Zhang, Li.,& Chen, Xiao-Wen.(2012). *Impacts of Atmospheric Conditions on Influenza in Southern China. Part I. Taking Shenzhen City for Example.* Open Journal of Air Pollution, Vol. 1, pp. 59-66.

[6] Kar, Sayanti., & Mukherjee, Phalguni.(2012). *Studies on Interrelations among SO_2, NO_2 and PM10 Concentrations and Their Predictions in Ambient Air in Kolkata.* Open Journal of Air Pollution, Vol. 1, pp. 42-50.

[7] Illin, Alexander., & Raiko, Tapani.(2010). *Practical Approaches to Principal Component Analysis in the Presence of Missing Values*. Journal of Machine Learning Research,Vol. 11, pp. 1957-2000.

[8]Kaushik, Inderjeet., & and Melwani, Rinki. (2007). *Time Series Analysis Of Ambient Air Quality at ITO Intersection In Delhi (India).*New Delhi. Journal of Environmental Research And Development. Vol. 2 No. 2, pp. 135-145.

[9] Hisyam Lee , Muhammad., Abd. Rahman,Nur Haizum., Suhartono., Latif, Mohd Talib., Nor, Maria Elena., & Kamisan, Nur Arina Bazilah. (2012) .*Seasonal ARIMA for Forecasting Air Pollution Index: A Case Study*. American Journal of Applied Sciences .Vol. 9 No.4, pp. 570-578.

[10] Wu, Edward Ming-Yang., & Kuo, Shu-Lung.(2012).*Air Quality Time Series Based GARCH Model Analyses of Air Quality Information for a Total Quantity Control District*. Aerosol and Air Quality Research, Vol. 12, pp. 331–343

[11] Quant, CM., Fischer, PH., Buringh, E., Ameling,CB., Houthuijs, DJM., & Cassee,FR.(2003).*Application of principal component analysis to time series of daily air pollution and mortality*. RIVM report 650010035/2003.

[12] Hisyam Lee ,Muhammad., Abd. Rahman,Nur Haizum., Suhartono., Latif, Mohd Talib., Nor, Maria Elena., & Kamisan, 1Nur Arina Bazilah.(2012).*Seasonal ARIMA for Forecasting Air Pollution Index: A Case Study*. American Journal of Applied Sciences .Vol.9, No.4,pp.570-578.

[13] Morley,J.,& Sinclair, Tara.(2009). *Bootstrap Tests of Stationarity*. Research Program on Forecasting Working Paper No. 2008-011 JEL Classifications: C12, C15, C22.

[14] Mondal, Prapanna., Shit, Labani., & Goswami, Saptarsi.(2014*). Study of Effectiveness of Time Series Modelling (ARIMA) in Forecasting Stock Prices*. International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.4, No.2

[15]Yaziz, S.,Ahmad,M., Nian, L., &Muhammad, N.(2011).*A Comparative Study on Box-Jenkins and Garch Models in forecasting Crude Oil Prices*.Journal of Applied Sciences 11(7), pp. 1129-1135.

[16] Valipour, , Mohammad., Banihabib, Mohammad.,& Behbahani, Seyyed.(2013). *Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir*.Journal of Hydrology. Vol. 476,pp.433–441

[17] Kumar, Rakeash.,& Abba,E.(2006).*Air Pollution Concentrations of PM2.5, PM10 and NO2 at Ambient and Kerbsite and Their Correlation in Metro City – Mumbai.* Environmental Monitoring and Assessment.Vol. 119,pp.191-199

[18] Stelson,A., Friedlander ,S.,& Seinfeld,J.(1979). *A note on the equilibrium relationship between ammonia and nitric acid and particulate ammonium nitrate*.Vol 13,pp. 369-371

[19] Lawrence, M.(2005).*The Relationship between Relative Humidity and the Dewpoint Temperature in Moist Air: A Simple Conversion and Applications*. AMS.Vol.86, No.2.