Understanding Air Pollution trends in various Part of Maharashtra, India

maxim rohit

Thesis Report

December 2019

ABSTRACT

Indian Government released a first of its kind report on air pollution pattern & related health impact for the country on 8th December 2018[1].

Few of the disturbing facts presented were that 12.5% of death in India are attributable to air pollution along with the average life going down by 1.7 years owing to health loss due to high concentration level of various pollutant including suspended particulate matter(SPM).  
In the same report Prof. Balram Bhargava, Secretary of the GOI, Department of Health Research, said ‘It is important to have robust estimates of the health impact of air pollution in every state of India to have a reference for improving the situation’.

This Novel exploratory research is aimed at answering the question from Pollutant’s point of view for the State of Maharashtra, India, for the first time. This Research will try to establish the relationship between four major pollutants SO₂, NO₂, SPM2.5, and SPM10. Additionally, finding the natural grouping of cities concerning pollutants and other novel & non-novel factors impacting them. A few examples of novel factors under consideration are Elevation from Sea level, Forest & Industrial Area Distribution among others. Few non-novel features like population density, seasonal patterns are also evaluated for their impact.

We found that Mumbai although being a commercial hub of India still maintains low levels of NO₂ and SO₂ as per Indian pollutant acceptable standards, whereas Pune being the 2nd most progressive city in Maharashtra is doing much worse but still within standards. Other major upcoming cities like Nagpur, Nashik & Amravati has the lowest level of pollutant concentration, signifying that the progression of the cities does have an impact on them. Hence proper consideration needs to be put in place while planning new and upcoming cities, as to not to repeat the pattern.

LIST OF TABLES

[Table 1. Indian Government Permissible Concentration Level.[6] 19](#_Toc25837566)

LIST OF FIGURES

[Figure 1. Residential pollutant levels for years 2004-2010 21](#_Toc26034321)

[Figure 2. Residential pollutant levels for years 2011-2015 22](#_Toc26034322)

LIST OF ABBREVIATIONS

|  |  |
| --- | --- |
| SPM | Suspended particle concentration in air that can be inhaled is considered as air pollutants |
| SPM2.5 | Suspended particulate matter of diameter below or equal to 2.5 μm. Our nasal hair cannot prevent their inhalation and they reach our lungs and blood circulation directly. |
| SPM10 | Suspended particulate matter of diameter between 2.5 μm and 10 μm.  Inhalation is prevented by our nasal hair |
| SO₂ | Sulphur dioxide, its concentration in air is representative of the sulphur oxide family’s concentration. |
| NO₂ | Nitrogen dioxide, its concentration in air is representative of the nitrous oxide family’s concentration. |
| P10 | Monthly percentile 10 value for the pollutant. |
| P50 | Monthly percentile 50 values for the pollutant. |
| P90 | Monthly percentile 90 values for the pollutant. |
|  |  |

ACKNOWLEDGEMENT

I would like to thank the faculty of LJMU for explaining the nuance of report writing.

Also, faculty of IIIT-B for in-depth teaching of data science concepts.

Rutuja a colleague for proofreading.

Last but not the least my wife Morien for going through endless non-user-friendly government sites in search of relevant data.

TABLE OF CONTENTS

[CHAPTER 1 : INTRODUCTION 17](#_Toc25836939)

[CHAPTER 2 : BACKGROUND AND LITERATURE REVIEW 18](#_Toc25836941)

[CHAPTER 3 : ANALYSIS 19](#_Toc25836943)

[CHAPTER 4 : RESULTS AND DISCUSSION 20](#_Toc25836946)

[CHAPTER 5 : CONCLUSION 23](#_Toc25836948)

CHAPTER 1

INTRODUCTION

The research aims to understand pollutant spread, correlation and growth pattern over the region of Maharashtra, India.  
Indian government’s first comprehensive report[1] released stated ‘1.24 million deaths in India are attributable to air pollution in 2017, of which 50+% were in individuals younger than 70 years’. A more recent example of the seriousness of this issue is the government school’s remaining closed for 2 days in Nov’2019 due to unbreathable living conditions post-Diwali celebration.  
Building an understanding of the pattern behind Air Pollutants is the first step towards being able to prepare for controlling/reducing these pollutants.

**Past and Current**

1. Part I: Natural grouping of cities will be explained via clustering in terms of pollutants and properties of cities like elevation from sea level, population-density, total area, industrial area, number and type of industries among others.  
   This will enable us to compare the progression of the cities through the years and answer questions like how do major industrial and IT hub based cities - Mumbai and Pune compare against the smaller town of Kolhapur, Nashik, and Nagpur among others.
2. PART II: We will use PCA to understand the interaction between the pollutants, by quantifying their combined loading on individual components.  
   With the knowledge of the impact of SO₂ on NO₂ concentration, we will be able to plan better control measures for both.

**Future**

1. PART III: Time Series analysis to predict when the pollutant concentration will go beyond an acceptable government standard.  
   This exercise may give us a time frame for our action items.

CHAPTER 2

BACKGROUND AND LITERATURE REVIEW

Most of the studies of similar nature are targeted at explaining these variable concentrations at City level, observing the pattern over the Season (New York [2]) and Nature of Area (Madrid’s Metropolitan vs Rural part [3]). Another study based on the city of Kolkata, India [4], attempts to quantify pollutants in different parts of the city and during different time frames within the day.  
Our study uses approaches applied in these analyses on a broader level for a set of neighbouring cities i.e. State Level; while comparing the cities distinguishable attributes impact on pollutant concentrations  
One of the innovative attempts towards explaining the pollutant behaviour via visual representation for Hong Kong city[5] also elaborates on the Time Series nature of pollutant concentrations. While this study concentrated on exploring the interaction between the pollutants over time, we will be concentrating on forecasting their concentration in the future.

Seasonal impact of natural factors like temperature, humidity, wind speed/direction on the coastal city of Chennai [7] showed that SO₂ & NO₂ were negatively co-related to summer. Same study showed that both SPM types has positive correlation with all seasons except post-monsoon.

CHAPTER 3

ANALYSIS

Data Preparation

Pollutants data for SO₂, NO₂, SPM10 & SPM2.5 was aggregated for each city at monthly levels and percentile 10, 50 & 90 were calculated to represent the range. We calculated month on month increase for these percentile values for tracking their seasonal pattern. Percentile 50 values were used for the final comparison.

Other features that are merged with the data are:

1. Population & Population Density
2. Elevation from sea level
3. Total/Forest/Industrial area spread and respective percentages
4. Roads
   1. national/state/district/rural
   2. A representative of automobile circulations & count
5. Rail line
6. Industrial
   1. services/manufacturing
   2. micro/mini/medium/large

Few interaction variables were calculated like

1. Rainfall per Area
2. Industrial area per Forest area among others.

**Clustering**

We performed two-levels of clustering, first was all variables from 2004-10 and the next set was limited to pollutant information from 1987-2015, purely due to lack of availability of data. The natural grouping found in the first exercise was validated against the longer range of the second cluster. All variables were scaled for consistency and equal weightage.

Silhouette Analysis and Elbow curves were used to decide the ideal number of clusters.

**PCA**

PCA analysis was used to understand the underlying themes of the data. With 4 principal components inline with four pollutant types, we were able to cover only 87% of variation of data. This signifies that the percentile values that signified the monthly range of pollutant are following different pattern. We needed all 12 PCA (3 each) component to get 100% variance coverage.

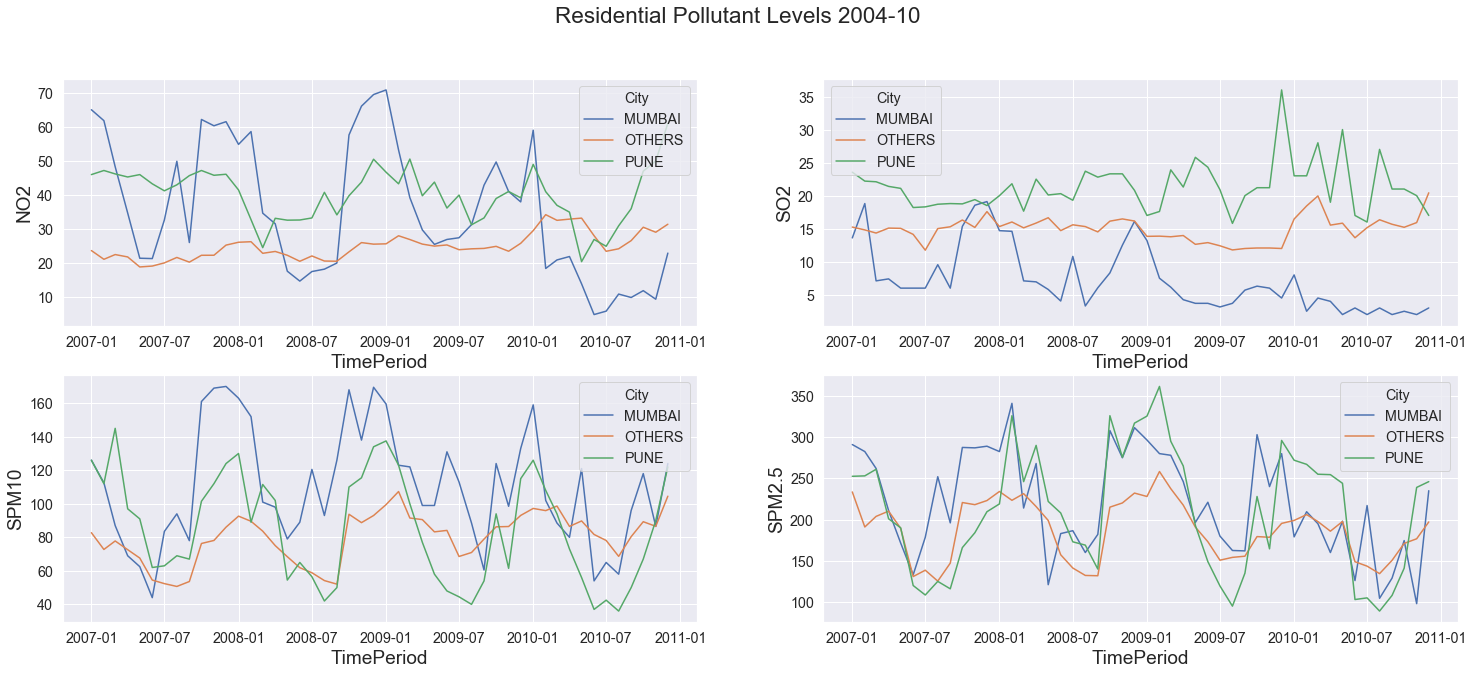
Timeseries analysis is pending and will be carried out in the 2nd half of the term.

CHAPTER 4

RESULTS AND DISCUSSION

**Part I Clustering:**

**Figure 1. Residential pollutant levels for years 2004-2010**



The above figure represents the monthly level of pollutants over the time frame 2004-2010 which was used for the first level of clustering with all variables.

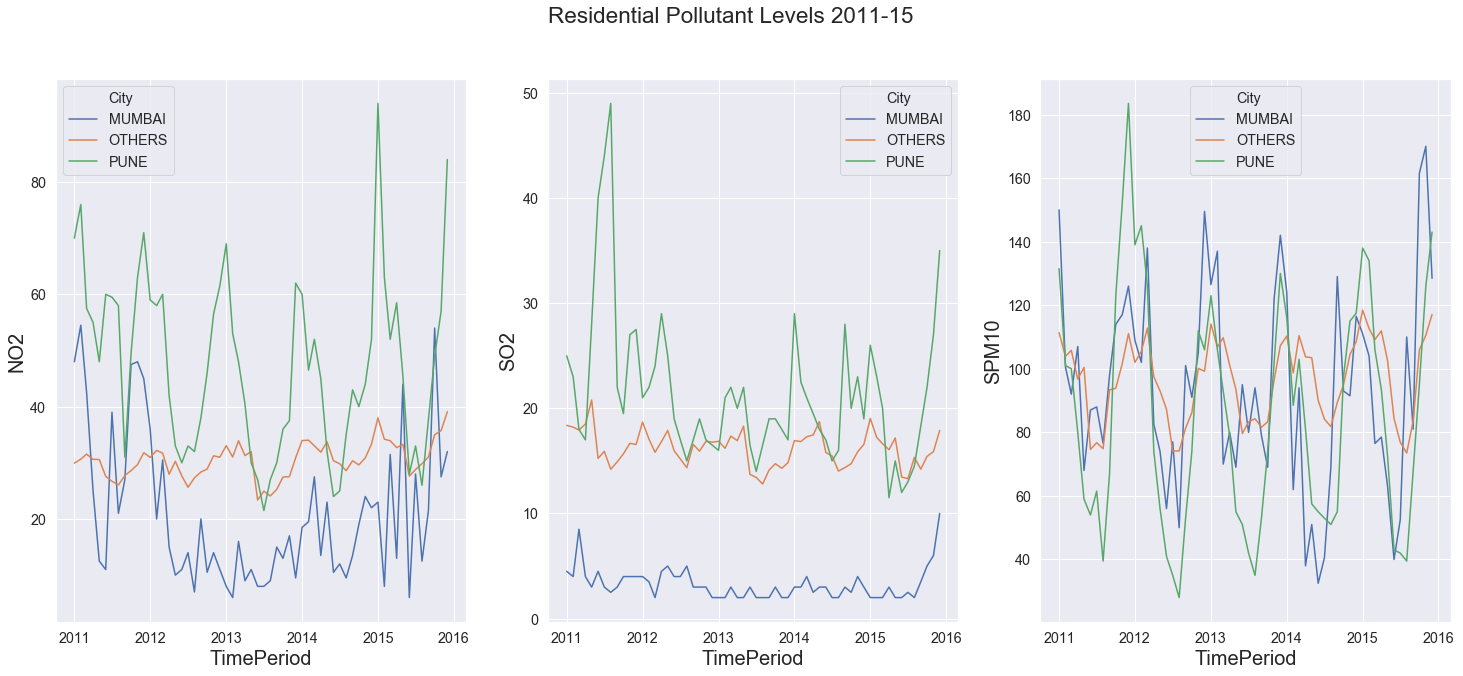
**Figure 2. Residential pollutant levels for years 2011-2015**

Figure 2. represents the state of the pollutants post-2010. SPM2.5 data is sparingly present for this period hence dropped from the graph.

**Table 1. Indian Government Permissible Concentration Level.[6]**

|  |  |  |
| --- | --- | --- |
| **Pollutant** | **Time Weighted Average** | **Concentration in Ambient Air: Industrial, Residential, Rural and Other Areas (in** µg/m3**)** |
| **Sulphur Dioxide (SO₂), µg/m3** | Annual\* | 50 |
| **Nitrogen Dioxide (NO₂), µg/m3** | Annual\* | 40 |
| **Particulate Matter (size less than 10 µm) or SPM10 µg/m3** | Annual\* | 60 |
| **Particulate Matter (size less than 2.5 µm) or SPM2.5µg/m3** | Annual\* | 40 |
| Note: \* Annual arithmetic mean of minimum 104 measurements in a year at a particular site taken twice a week 24 hours at uniform intervals. | | |

**Cluster 2004-2010**

Three major clusters were observed one for Mumbai & Pune each and the third one composed of all the remaining cities.

1. **Pune cluster** has the highest level of average monthly levels of SO₂ @29µg/m3 & NO₂ @39µg/m3 over the year, still lower than acceptable levels of NO₂ 40µg/m3 & SO₂ 50µg/m3. The **Residential** **area** of Pune has shown an increasing pattern for months Oct, Dec, Jan post the year 2006 till 2009 for SPM2.5/10. The **Industrial** **area** replicates the same behaviour and decreases during the remainder of the year. Even monthly Percentile 10 values for SPM2.5 is above the acceptable level of 40µg/m3.

Attribute wise Pune cluster stands out with the highest elevation i.e. 34% higher than others. Highest Roads length including national, state highways and internal district and rural roads also stand out. Micro-industries are also highest for this cluster in terms of manufacturing & services both, questions can be asked about them following effective measures to control their share of pollutants. Rainfall per area is also lowest for Pune along with the low population density and lowest percentage of forest area.

Pune’s behaviour over the clustering exercise for years 1987-2015 saw that the residential area consistently had higher values as compared to other cities whereas industrial area has seen a decline in NO₂, SO₂, and SPM2.5 levels.

1. **Mumbai cluster** stands out for its lowest SO₂ level.

For **Industrial** **area** level of NO₂, SPM2.5&10 decreases during Feb till Aug and is within the acceptable standards accompanied by higher rainfall levels. Also, the range of all the pollutants except SPM2.5 remains under acceptable levels. During the remainder of the year the level of the three pollutants increases, specifically SPM2.5 level rises dangerously to @229 µg/m3 and SPM10 goes beyond acceptable level @124 µg/m3. **Residential** **area** observes the increasing phase during Aug, Oct & Dec and decreases during the remainder of the year.

Attribute wise Mumbai site on the opposite end of the spectrum with the lowest elevation & highest rainfall per area when compared to the Pune cluster. Another standout feature of the cluster is the largest industrial area and the highest number of large manufacturing firms all that is packed in the lowest overall area and forest portion-wise among the cities. 1987-2015 clustering also shows similar results w.r.t pollutant concentration behaviour as compared to the other clusters. The only exception being NO₂ whose levels starts decreasing from the year 2010 and continues dropping into the next decade.

1. **Other Clusters** display a similar pattern as in Mumbai cluster’s decreasing phase but at 20-30% lower intensity, except for SO₂ concentration which is 230% higher.

Attribute wise these cities are placed in between Mumbai & Pune in terms of elevation, rainfall (per total area). These cities stand out in terms of the highest forest area (per total area), the lowest population density, industrial area and count of large industries.

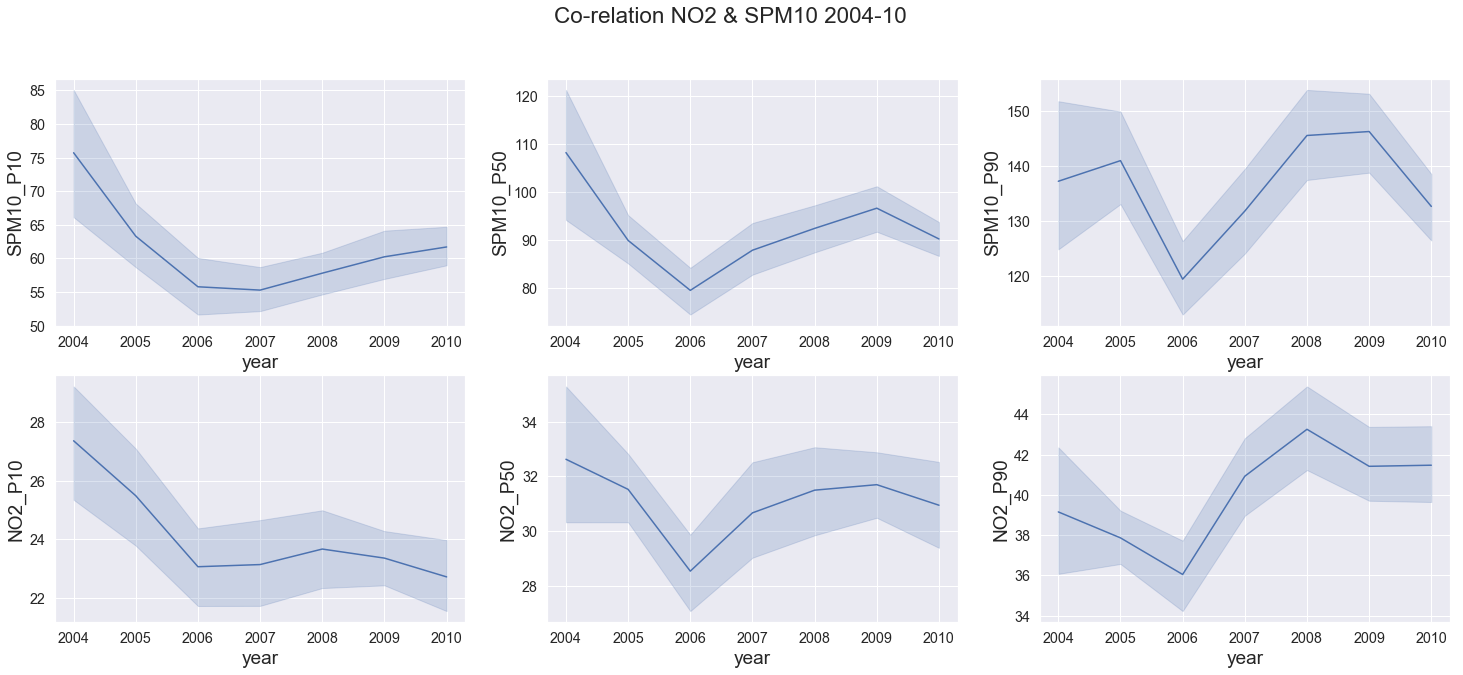
While validating the behaviour over 1987-2015 few exceptional observations were made:

1. Aurangabad is seeing a rise in all four pollutant levels over the years.
2. Solapur and Mahad are the only cities reducing the level of SPM2.5, the pollutant with the most extreme values.

**Exceptional Observations** - Chandrapur is an exception to these clusters and shows the highest value for SO₂ @24.3µg/m3 (still acceptable), RSPM2.5 & 10 (extremely high) from all the clusters with a monthly increase in RSPM2.5 & 10 values over the years 2004-2010.

S02 being high signals towards Diesel based instrumentation not being optimally levelled. For the years 2006-2010 and months Sep, Oct & Dec Nagpur displayed similar behaviour.

**PART II PCA:**

**Figure 3. NO₂ & SPM10 co-relation for years 2004-2010**

First principle component has a high loading for SPM10 & NO₂, figure 3 displays the annual tread across Maharashtra, the relation is highlighted at al1 three level P10, P50 & P90 with co-relation being 0.49, 0.52 & 0.56 increasing with percentile.

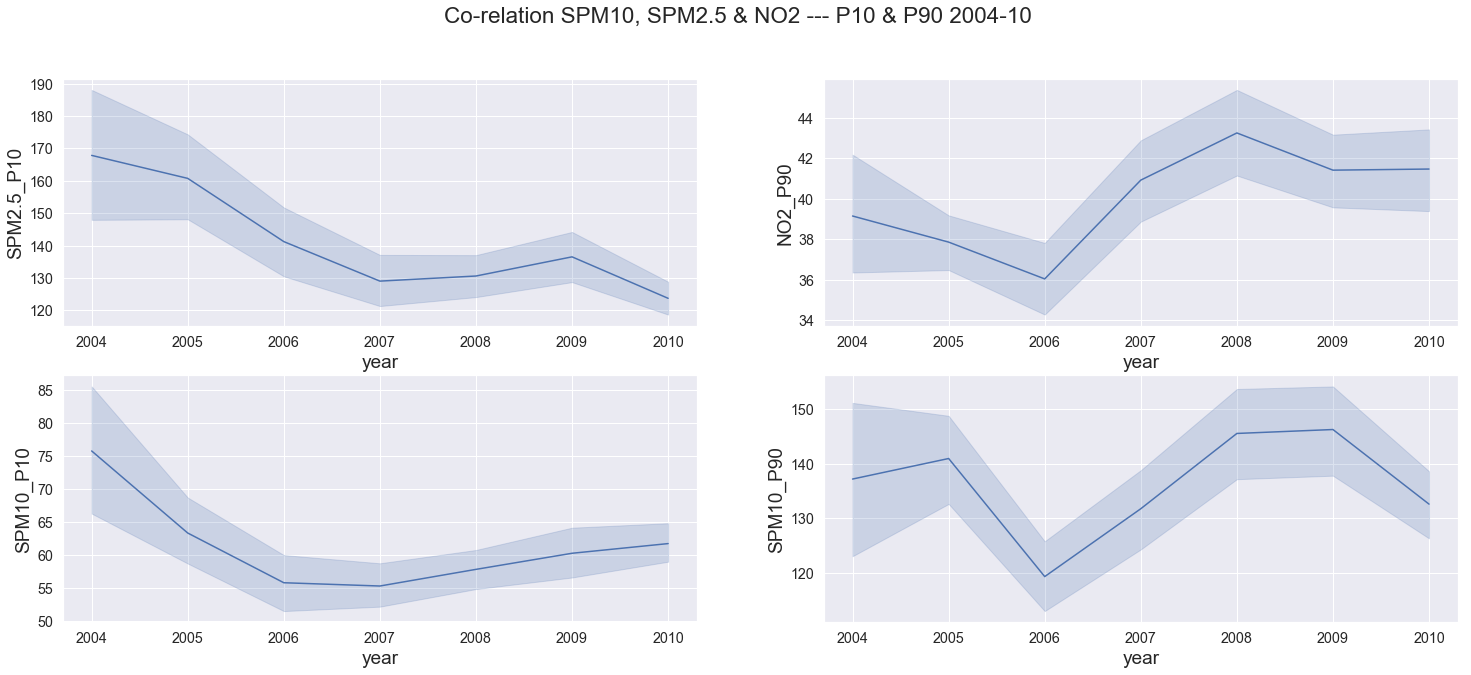
Except for SMP10 P90 growth vs drop of NO₂ P90 between 2004-05, the trend matches for all other years. The matching trends signify a common source mostly fossils fuels combustions (commercial & individual) and industrial activities, Maharashtra is also known as the industrial capital of India.

**Figure 4. Co-relation SO2 & SPM2.5 2004-10**

On the 2nd principle component, SO₂ has high positive loading and inversely SPM2.5 has negative loading. Their behaviour across the year as depicted in figure 4 confirm similar movement before 2007 (SPM2.5 has greater quantum) and 2007 onwards both variables start to move in inverse directions.

One exception prominent was between 2004-05 for P90 values of SPM2.5 which saw an increase while all other values decreased. This signifying that sources of SPM alone were at play like construction work, natural wind-based debris.

As depicted by their corresponding negative and positive loading in the component the co-relation is non-existence below 0.07.

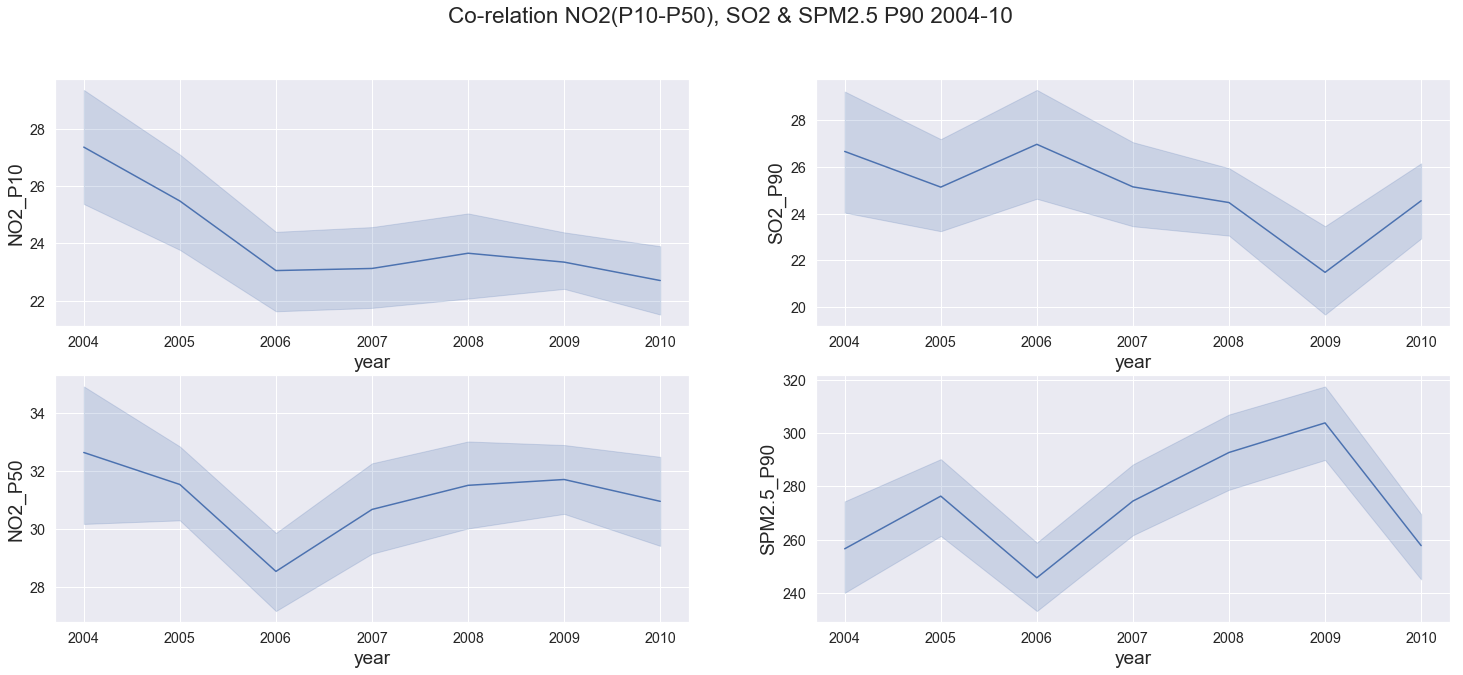
**FIGURE 5. Co-relation SPM10, SPM2.5 & NO2 --- P10 & P90 2004-10**

SPM2.5 & SPM10 are expected to be co-related, their P10 values loading on the 3rd principal confirm that relationship. An exception is observed for year 2009-10 where both variable move in opposite direction. Their correlation is above .66 at all percentile levels.

P90 values for SPM10 & NO₂ have negative loading on this component and drastically increase between 2006-08. SPM10 two percentile ranges depicted by P10 & P90 show very different behaviour, highlighting fact that the range is increasing owing to more days with higher pollutant concentration during the years between 2006-2008.

A consistent observation in all this variable expect SO₂ is a negative change in trend between 2008-10 coinciding with impact of Great Recession [ref- <https://en.wikipedia.org/wiki/Great_Recession>] that affected US between Dec07-June09 and whose after effect were felt in India from Sep08-Sep09 .

**FIGURE 6. Co-relation NO2(P10-P50), SO2 & SPM2.5 P90 2004-10**



Fourth principle component has NO₂ positively loaded and SPM2.5 & SO₂ negatively loaded.

The raise in NO₂ P50 levels (while P10 remained constant) post year 2006 stretches the range of the pollutant signifying that its level went up overall and are persistence.

On the other hand, the movement of SPM2.5, SPM10 & NO₂ are similar and SO₂ seems to be following a different pattern observed here and overall.

CHAPTER 5

CONCLUSION

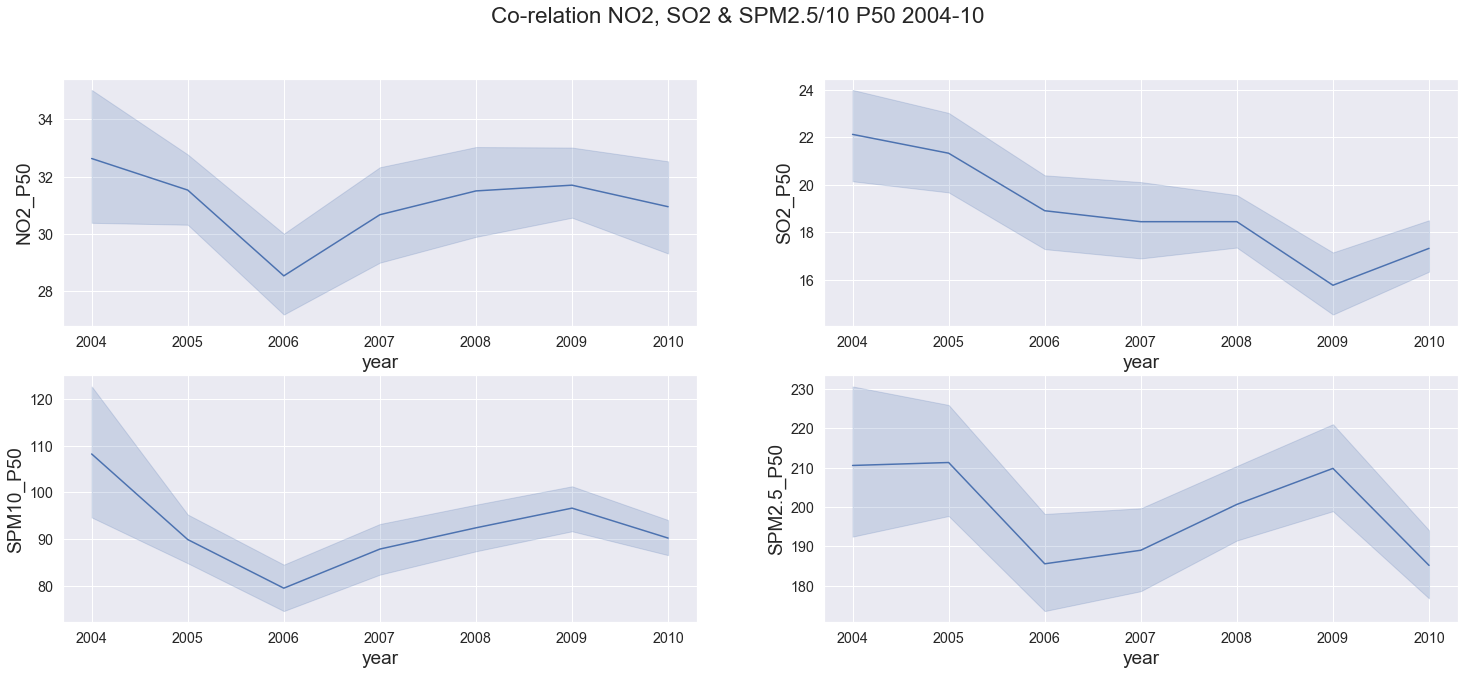
**Clustering**

Mumbai although heavy on the industrial area and large industries count still registered smaller pollutants concentration numbers, this can be co-related to its properties of lower elevation and high rainfall. Pune is on the other end of the spectrum with higher elevation and lower rainfall and shows the highest pollutant concertation levels.

All other less progressive cities of the state fall under the same umbrella. They exhibit the least pollutant concentration (other than SO₂) correlated with the higher forest area percentage, lower industrial area and large industries counts. This behaviour answers our question about the city’s progression resulting in higher pollutant level, which should point us in the direction of better planning for future development keeping measure to control air pollutant in mind.

Lowest SO₂ level observed in Mumbai warrants further research, its relationship to humidity may be a factor here as Mumbai has highest humidity number owing to its closeness to the Arabian Sea.

**PCA**



PCA highlighted that SO₂ concentration level followed a different trend as compared to other pollutants.

**NO₂ & SPM2.5/10 :** We observed that post 2006 the increasing trend of all pollutant reduced in amplitude yearly and became negative by 2009.

SO₂ on the other hand has seen a decreasing trend throughout the years only to increase post 2010**.**

SO₂ decreasing co- relation with NO₂ ranges from 0.55, 0.41 & 0.37 at P10, P50 & P90 level respectively. This highlights that their spikes are caused by uncommon sources, whereas the common sources like industrial/commercial fossil fuels combustion are maintaining the relationship at lower percentile levels.

Time-series analysis for future pollutant levels will be shared in the final report.

REFERENCES

[1] Public Health Foundation of India, New Delhi, 6 Dec 2018, Press Release https://phfi.org/wp-content/uploads/2018/12/first-comprehensive-estimates-of-the-impact-of-air-pollut ion-india.pdf –

[2] Emami, Fereshteh & Masiol, Mauro & Hopke, Philip. (2017). Air pollution at Rochester, NY: Long-term trends and multivariate analysis of upwind SO₂ source impacts. The Science of the total environment. 612. 1506-1515. 10.1016/j.scitotenv.2017.09.026. –

[3] Núñez-Alonso, David & Pérez-Arribas, Luis & Manzoor, Sadia & Caceres, Jorge. (2019). Statistical Tools for Air Pollution Assessment: Multivariate and Spatial Analysis Studies in the Madrid Region. Journal of Analytical Methods in Chemistry. 2019. 1-9. 10.1155/2019/9753927.

[4] Haque, M.S. & Singh, R.B.. (2017). Air Pollution and Human Health in Kolkata, India: A Case Study. Climate. 5. 10.3390/cli5040077.

[5] Qu, Huamin & Chan, Wing-Yi & Xu, Anbang & Chung, Kai-Lun & Lau, Alexis & Guo, Ping. (2007). Visual Analysis of the Air Pollution Problem in Hong Kong. IEEE transactions on visualization and computer graphics. 13. 1408-15. 10.1109/TVCG.2007.70523.

[6] Government of India, Ministry of Environment, Forest & Climate change, Permissible Level for Pollutants: <http://www.indiaenvironmentportal.org.in/files/file/Permissible%20Level%20for%20Pollutants.pdf>

[7] Ramasamy Jayamurugan, B. Kumaravel, S. Palanivelraja, and M. P. Chockalingam, “Influence of Temperature, Relative Humidity and Seasonal Variability on Ambient Air Quality in a Coastal Urban Area,” International Journal of Atmospheric Sciences, vol. 2013, Article ID 264046, 7 pages, 2013. <https://doi.org/10.1155/2013/264046>.

**Additional References**

* Fotopoulou, Eleni & Zafeiropoulos, Anastasios & Papaspyros, Dimitris & Hasapis, Panagiotis & Tsiolis, George & Bouras, Thanassis & Mouzakitis, Spiros & Zanetti, Norma. (2015). Linked Data Analytics in Interdisciplinary Studies: The Health Impact of Air Pollution in Urban Areas. IEEE Access. 4. 1-1. 10.1109/ACCESS.2015.2513439. [10]
* Shaban, Khaled & Kadri, Abdullah & Rezk, Eman. (2016). Urban Air Pollution Monitoring System With Forecasting Models. IEEE Sensors Journal. 16. 1-1. 10.1109/JSEN.2016.2514378.[11]
* Geng, Zhaowei & Chen, Qixin & Xia, Qing & Kirschen, D.s & Kang, Chongqing. (2017). Environmental Generation Scheduling Considering Air Pollution Control Technologies and Weather Effects. IEEE Transactions on Power Systems. 32. 127-136. 10.1109/TPWRS.2016.2544851.[12]

APPENDICES

Data is being collected from the following government repositories: -

* Pollutant:-https://data.gov.in/catalog/historical-daily-ambient-air-quality-data
* Vehicle Registration:-http://mospi.nic.in/statistical-year-book-india/2017/189
* Rainfall:-https://www.indiawaterportal.org/
* Elevation:-https://en.wikipedia.org/wiki/<CityBasedURL>
* Industrial Area:- [http://dcmsme.gov.in/<CityBasedURL](http://dcmsme.gov.in/%3cCityBasedURL)>
* Population:-https://mahasdb.maharashtra.gov.in/population1.do