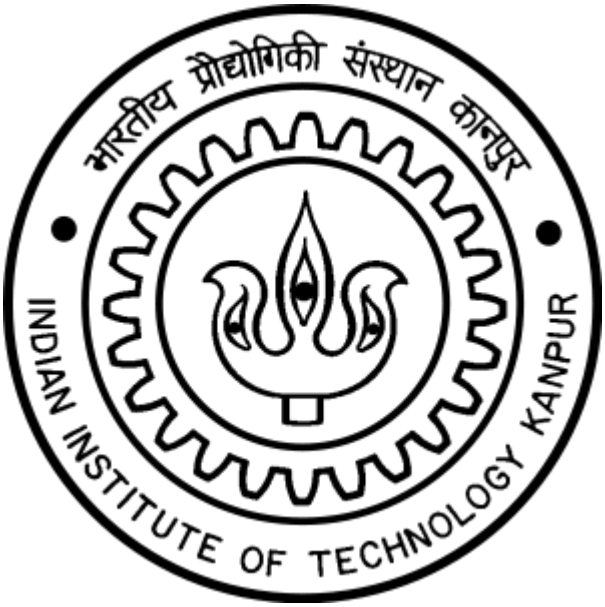
UNDERGRADUATE PROJECT

UG - 492 A



Group Members:

Prateek Roy (14490)

Vinay Duli (14231)

Kuthadi Suresh(14347)

Ajeet Kumar (13058)

UGP Report

Objective – To plot various pollutants along with different meteorological parameters to observe any correlation between them.

Cities – Delhi, Kanpur, Mumbai and Kolkata.

Pollutant Parameters – $PM_{2.5}$, CO, NO_2 and O_3.

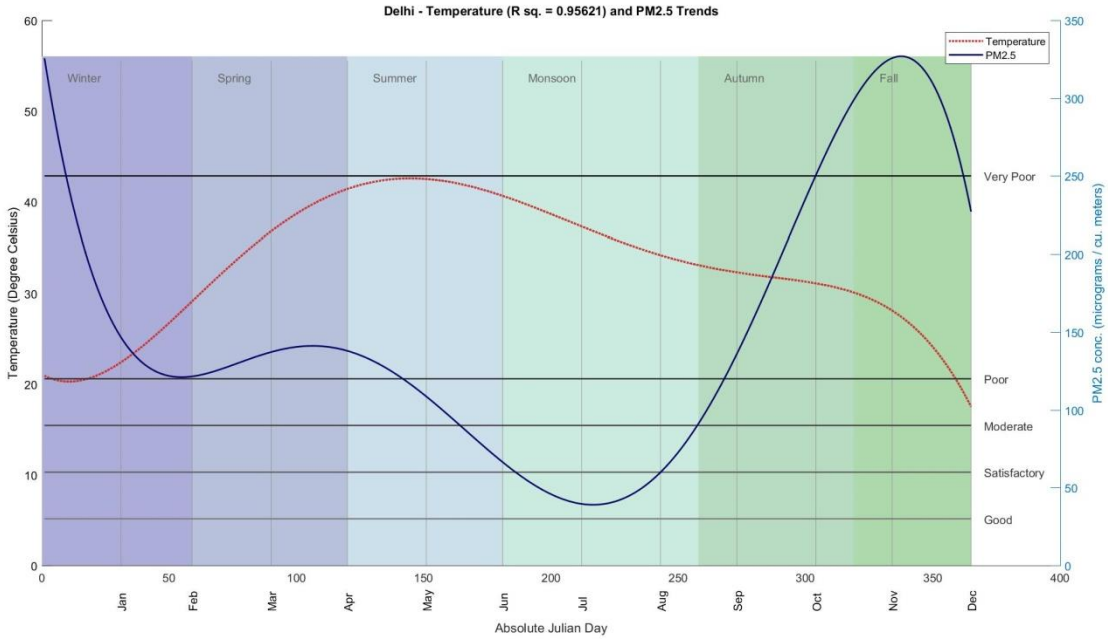
Meteorological Parameters – Temperature, Solar Radiation, Relative Humidity and Precipitation.

Method –

1. For the given 4 cities, latest (available) 3 years of pollutant data and latest (available) 10 years of meteorological data was collected.
2. Based on the data an annual trend curve was plotted for each meteorological parameter and pollutant parameter.

Parameter	Plot	Duration of Data
Temperature	5^{th} order polynomial regression	10 years
Solar Radiation	3^{rd} order polynomial regression	10 years
Relative Humidity	5^{th} order polynomial regression	10 years
Precipitation	Bar Graph	10 years
$PM_{2.5}$	5^{th} order polynomial regression	2-3 years
CO	5^{th} order polynomial regression	2-3 years
NO_2	5^{th} order polynomial regression	2-3 years
O_3	5^{th} order polynomial regression	2-3 years

3. Then for each city every pollutant parameter was plotted along with each meteorological parameter. Here is an example -



Specific features –

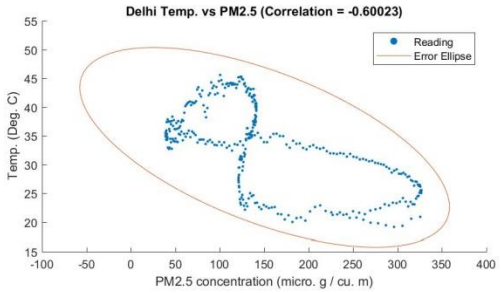
a. The meteorological parameter is plotted in reference to the left y axis.
b. The pollutant concentration is plotted along the right y axis.
c. AQI indexes are labelled for reference.
d. X – Axis represents the timeline in Absolute Julian Days. There are labels for different months and seasons.

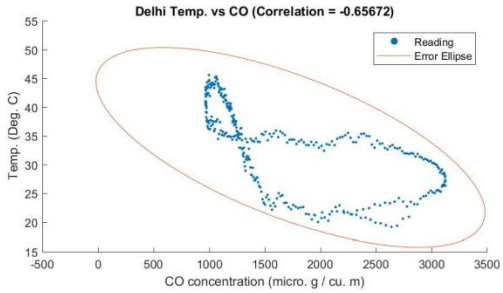
4. Once the pollutants are plotted along each meteorological parameter, all it remains is to find any significant correlation between them. To find correlation, each pollutant is plotted against a meteorological parameter. Such that -

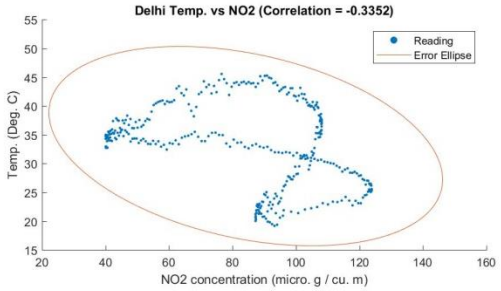
   Y – axis – meteorological parameter.

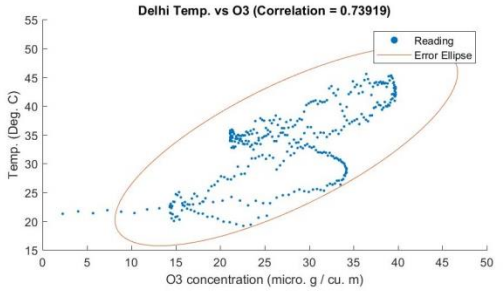
   X – axis – pollutant concentration.

   This gives some sort of data distribution. To calculation correlation between X and Y we approximated this distribution as a Gaussian distribution. Then it remains to calculate the mean vector and the covariance matrix.



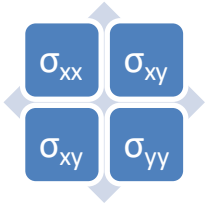






The co-variance matrix is calculated as

Cov(X,Y) =



Using the eigenvectors of the covariance matrix, the direction of maximum and minimum variance is found. Using this information along with the mean, a 95% confidence error ellipse is plotted to give an insight of the spread of the data in the approximate Gaussian Model.

5. Finally, correlation is calculated as -

Correlation(X,Y) = $\sigma_{xy}/sqrt(\sigma_{xx}\sigma_{yy})$

The value of correlation lies between -1 and 1.

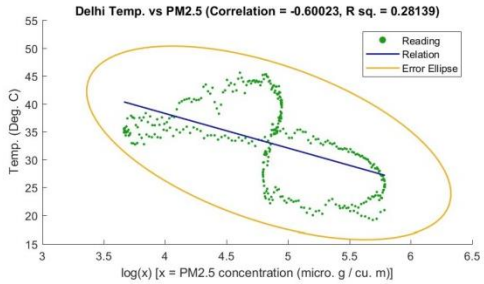
1 means strongly correlated

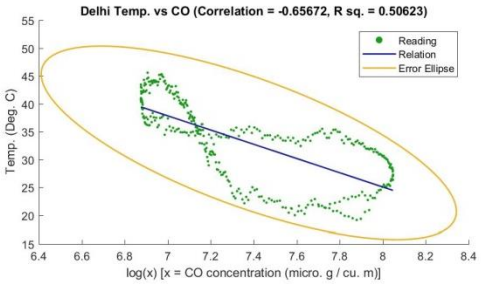
0 means no correlation

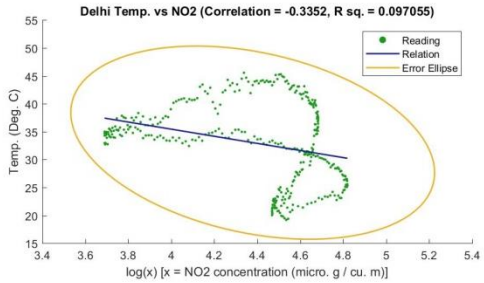
-1 means strong negative correlation

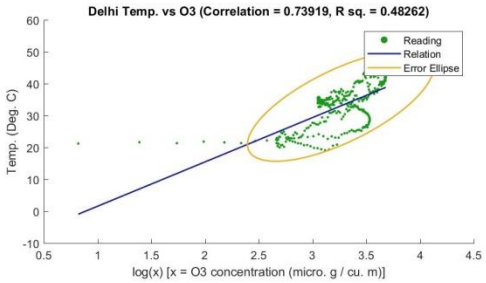
6. This process is repeated again but this time log normal distribution is used instead of normal distribution. A line was fitted using least squares through the dataset to give a rough estimate of how the pollutant parameter varies with temperature. The correlation and the

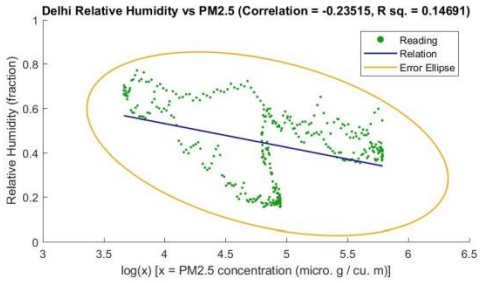
R^2 values were calculated for each of the plots.

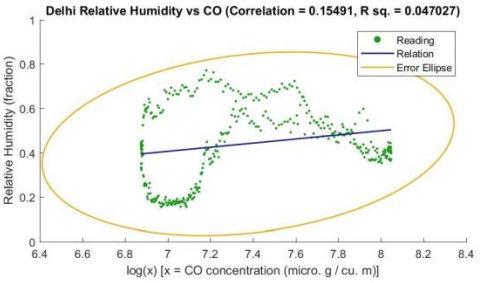


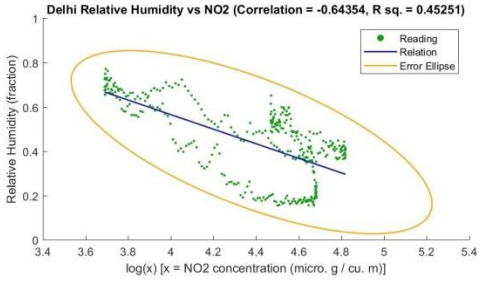


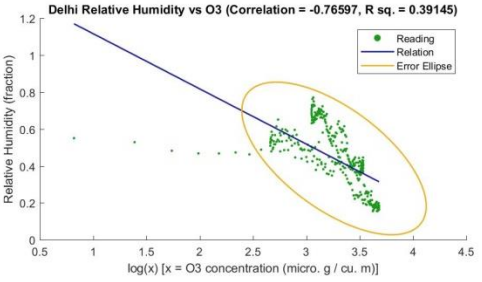




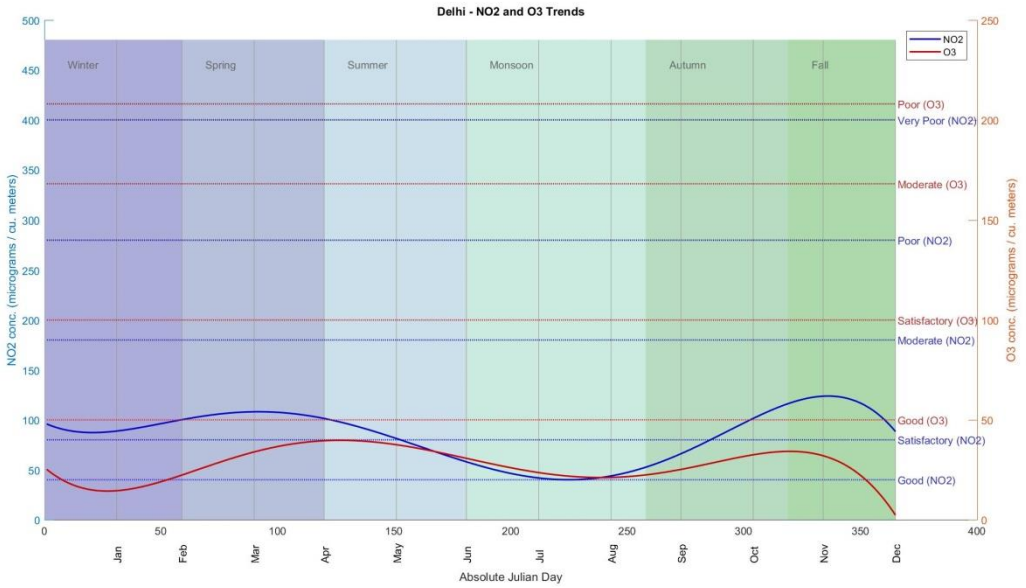


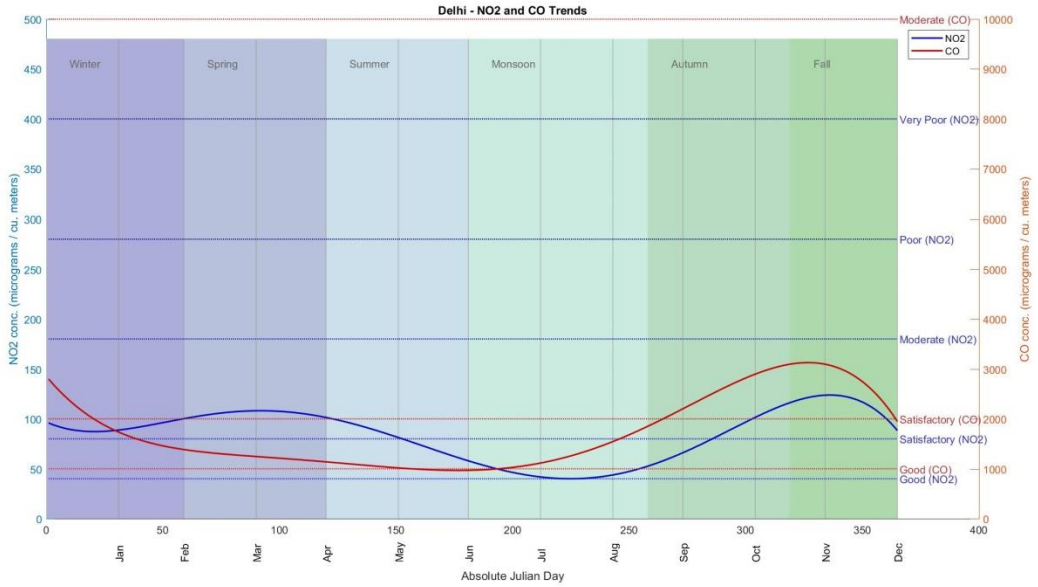






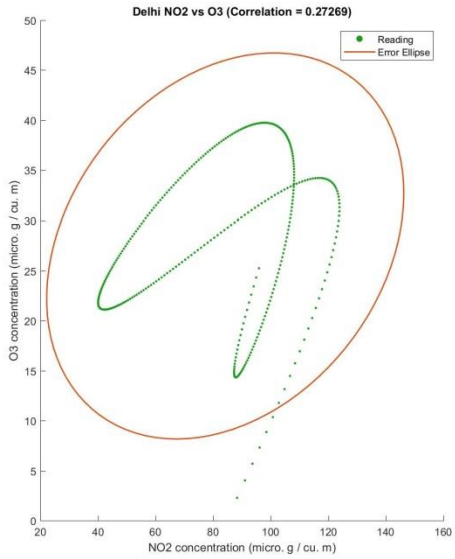
7. Then, to witness any relations that might exist between two pollutant parameter itself, we proceeded to compare the curves of NO_2 with O_3 and NO_2 with CO.

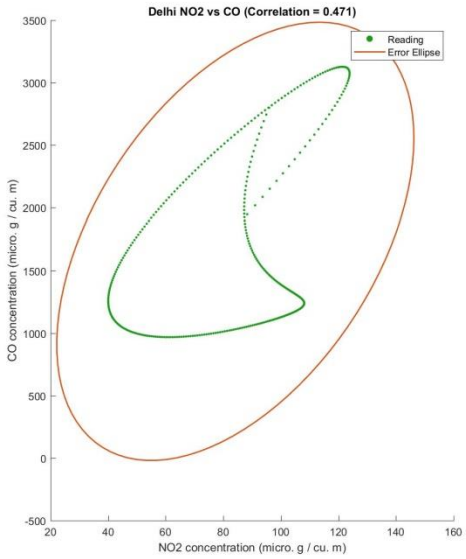




NO_2 was plotted along the left y axis while the other parameter was plotted along the right y axis. Their individual AQI values were also labelled for reference.

8. After plotting the trend curve we require to find correlation. Again using the same method as mentioned earlier we estimate an error ellipse from the covariance matrix and then proceed to calculate correlation values.





Note – The readings mentioned in the shown plot are not the actual average value for a day but the value from the estimated trend curve. This was done to witness any particular trend pattern that is followed.

Observation

Table 1: Correlation Values between met. parameter and pollutant in normal distribution.					
City	**Pollutant**	**Meteorological Parameters**			
		Temperature	**Solar Rad.**	**Rel. Humidity**	**Precipitation**
Delhi	$PM_{2.5}$	-0.6	-0.51	-0.23	-0.54
	CO	-0.66	-0.68	0.15	-0.3
	NO_2	-0.34	-0.11	-0.64	-0.72
	O_3	0.74	0.71	-0.77	-0.21
Kanpur	$PM_{2.5}$	-0.7	-0.53	-0.1	-0.53
	CO	-0.35	-0.27	-0.17	-0.47
	NO_2	0.2	0.45	-0.85	-0.63
	O_3	0.45	0.26	-0.24	-0.05
Mumbai	$PM_{2.5}$	-0.02	-0.05	-0.9	-0.62
	CO	-0.18	-0.38	-0.74	-0.47
	NO_2	0.25	-0.01	-0.34	-0.46
	O_3	-0.07	-0.11	-0.8	-0.55
Kolkata	$PM_{2.5}$	-0.54	-0.52	-0.7	-0.7
	CO	-0.55	-0.58	-0.62	-0.68
	NO_2	-0.62	-0.52	-0.52	-0.4
	O_3	0.92	0.51	-0.39	-0.27

Table 2: Correlation Values between met. parameter and pollutant in lognormal distribution.

City	Pollutant	Meteorological Parameters			
		Temperature	Solar Rad.	Rel. Humidity	Precipitation
Delhi	$PM_{2.5}$	-0.60023	-0.51363	-0.23515	-0.5417
	CO	-0.65672	-0.68024	0.15491	-0.29604
	NO_2	-0.3352	-0.10872	0.45251	-0.71539
	O_3	0.73919	0.71705	-0.76597	-0.20931
Kanpur	$PM_{2.5}$	-0.70121	-0.52543	-0.10023	-0.53262
	CO	-0.34934	-0.27186	-0.17156	-0.46854
	NO_2	0.20346	0.45196	-0.84522	-0.62599
	O_3	0.44855	0.26242	-0.24463	-0.048571
Mumbai	$PM_{2.5}$	-0.020027	-0.05403	-0.89764	-0.62499
	CO	-0.18451	-0.37716	-0.74048	-0.46512
	NO_2	0.2527	0.013813	-0.34306	-0.46328
	O_3	-0.069846	0.11265	-0.803	-0.5544
Kolkata	$PM_{2.5}$	-0.53513	-0.52178	-0.6988	-0.69904
	CO	-0.54732	-0.58495	-0.61852	-0.68219
	NO_2	0.43743	-0.52149	-0.52042	-0.40049
	O_3	0.91713	0.51067	-0.39428	-0.27389

Table 3: R^2 Values for linear regression between met. parameter and pollutant in lognormal distribution.

City	Pollutant	Meteorological Parameters			
		Temperature	Solar Rad.	Rel. Humidity	Precipitation
Delhi	$PM_{2.5}$	0.28139	0.13389	0.14691	0.44859
	CO	0.50623	0.50311	0.047027	0.081514
	NO_2	0.097055	0.0047911	0.45251	0.56454
	O_3	0.48262	040803	0.39145	0.015205
Kanpur	$PM_{2.5}$	0.36727	0.13741	0.062245	0.42712
	CO	0.10698	0.050626	0.049002	0.26281
	NO_2	0.026646	0.1579	0.73507	0.45366
	O_3	0.22867	0.079925	0.04666	1.4898e-05
Mumbai	$PM_{2.5}$	0.014458	0.000721	0.90138	0.52507
	CO	0.02934	0.1502	0.59871	0.25225
	NO_2	0.059329	0.0029828	0.08146	0.169
	O_3	6.5444e-06	0.014283	0.70548	0.39103
Kolkata	$PM_{2.5}$	0.14179	0.18018	0.6017	0.67693
	CO	0.20277	0.266	0.47213	0.60003
	NO_2	0.43743	0.26511	0.29921	0.16433
	O_3	0.76122	0.24412	0.14426	0.065063

Table 4: Correlation Values between pollutant parameters in normal distribution.			
City		**O_3**	**CO**
Delhi	**NO_2**	0.27269	0.4710
Mumbai	**NO_2**	-0.085622	0.17558
Kanpur	**NO_2**	-0.4502	0.81424
Kolkata	**NO_2**	0.33069	0.63117

Table 5: R^2 Values for Meteorological and Pollutant Parameters								
	Meteorological Parameter				**Pollutant Parameter**			
	Temp	**SR**	**RH**	**Rainfall**	**$PM_{2.5}$**	**CO**	**NO_2**	**O_3**
Delhi	0.95621	0.74818	0.90582	-	0.65052	0.4505	0.44894	0.16507
Kanpur	0.94429	0.62754	0.93409	-	0.58523	0.1762	0.40143	0.13532
Mumbai	0.85018	0.60225	0.96881	-	0.47831	0.54137	0.20985	0.41984
Kolkata	0.9206	0.48162	0.96256	-	0.68082	0.48418	0.71515	0.40784

SR - Solar Radiation RH - Relative Humidity $PM_{2.5}$ - Particulate Matter (2.5 micron)

CO - Carbon Monoxide NO_2 - Nitrogen Dioxide O_3 - Ozone

Conclusion –

- Most pollutant has a negative correlation to temperature with exception to ozone. This is due to the fact that diffusion is high during summers. However, ozone is high because its production is increased in abundance of sunlight.
- Solar Radiation has a significant positive correlation with ozone as expected.
- All the pollutants have highly negative correlation with rainfall, with excess rainfall a lot of pollutants dissolve in water droplets and settle.
- Relative humidity shows a negative correlation with pollutant concentration but there is a catch. Since most of the high humidity days there is a good chance of rainfall, the isolated effected of increased humidity is hard to measure. However, looking at the plots it's quite evident that for low humidity days increase in humidity may also increase pollutant concentration.
- NO_2 is expected to show a positive correlation with O_3. Although, it does in few cases, the relation doesn't seem to be very strong for some cities. This might be due to several other unknown factors that might be hindering.
- NO_2 has shown a significant positive correlation with CO. This might be due to the fact that both of these gases are by products of combustion and hence generally always produced together. Moreover, these gases react to temperature and precipitation in similar ways.