Paper No.368

GIS-BASED HEALTH RISK ASSESSMENT OF AMBIENT AIR QUALITY DUE TO RAPID URBANIZATION IN LAHORE, PAKISTAN

Najma Iftikhar,
Muhammad Muzammil,
Arbab Jamal,
Zill-e-Huma Munir
Muhammad Irfan Jalees

GIS-BASED HEALTH RISK ASSESSMENT OF AMBIENT AIR QUALITY DUE TO RAPID URBANIZATION IN LAHORE, PAKISTAN

By

Najma Iftikhar[1], Muhammad Muzammil[1], Arbab Jamal[1], Zill-E-Huma Munir[1], and Muhmmad Irfan Jalees[1]*

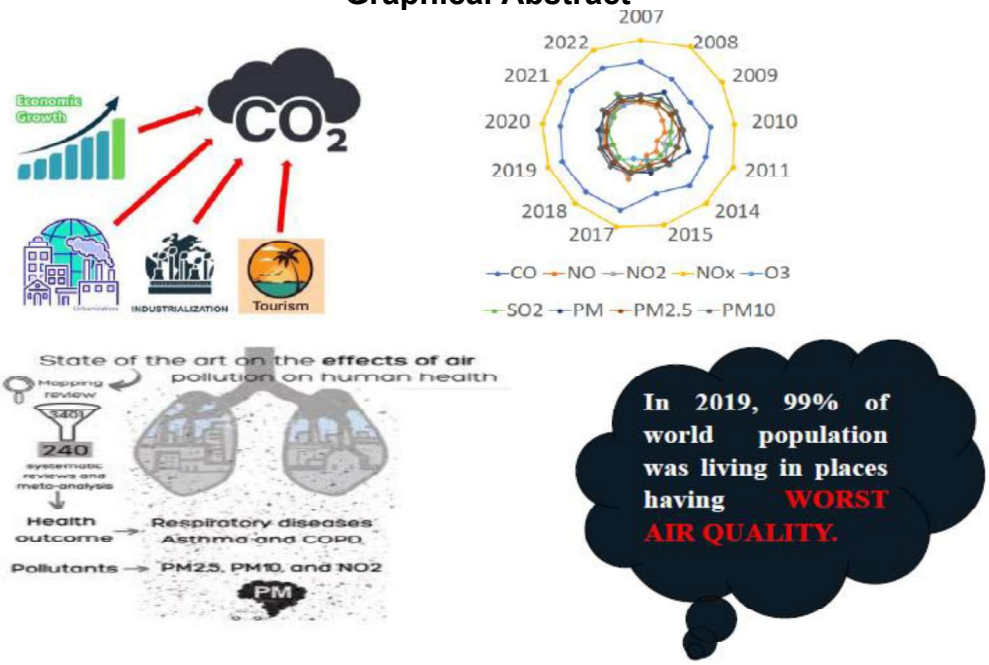
ABSTRACT

The increasing trend of urbanization in Lahore increases the demand for industries and modes of transportation, eventually becoming the center for industrial and vehicular gaseous emissions, i.e., NO_x, SO_x, CO, PM, $PM_{2.5}$, PM_{10}, and O_3. These pollutants are now becoming the cause of climate change. These climate changes have severe health effects. This study, therefore, focused on the Spatial variation of contaminants and their health impacts on the population of Lahore. The air quality data from 2007-2022, monitored through Aeroqual ® by the Environmental Protection Department (EPD), Punjab, was quantified. The risk assessment was estimated by AQI using US EPA guidelines as a standard. The result showed that almost all the pollutants exceed the PEQs. The increasing level of the Air Quality Index (AQI) showed that air quality deteriorated by 69% from 2007 to 2022. Also, the inhabitants of the studied have a cancer risk of 1 person out of 1000. This study can help understand the air pollutants and spatial and temporal patterns in Lahore city and present valuable implications for policymakers to tackle air pollution and adopt strategies for air pollution reduction in the future.

Keywords: GIS mapping; Air Pollution, Health Risk, Lahore, Windrose

[1] Institute of Environmental Engineering and Research, University of Engineering and Technology, Lahore, Pakistan 54890: *Corresponding Author: jalees@uet.edu.pk

Graphical Abstract



1. Introduction:

Air is the name for the mixture of gases that makes up the Earth's atmosphere. Generally, Air consists primarily of nitrogen (78%), oxygen (21%), argon (0.9%), carbon dioxide (0.04%), and trace gases. Pure Air has no discernible scent and no color. Without Air, life would not exist. Essentially, nitrogen and oxygen are the essential components that allow us to sustain ourselves on the planet. Air is an important energy supplier; every living plant and animal depends on oxygen to produce the energy necessary for them through respiration (Madronich et al., 2023).

Nevertheless, the Air is getting polluted due to the daily emissions from natural and anthropogenic activities. Air pollution, both indoor and outdoor, is a source of air pollution. Ambient air pollution is a significant issue worldwide and causes severe environmental pollution. Various pollutants are emitted from industries and automobiles and pollute the environment. It is necessary to deal with this issue, causing severe conditions (Rohde & Muller, 2015). Pollutants emitted from automobiles include carbon monoxide, carbon dioxide, carbon black, oxides of nitrogen, oxides of sulfur, ozone, VOCs, and most importantly, particulate matter. Particulate matter includes fine particles that are also dangerous for human health. PM_{10} and $PM_{2.5}$ are the primary particulates that harm our health and environment. Air pollution levels are increasing day by day (Cepeda et al., 2017). In 2019,

99% of the world's population was living in places where the WHO air quality guidelines levels were not met (Faridi et al., 2023).

Air pollution has adverse effects on human health. Increasing air pollution causes significant health effects like stroke, heart disease, lung cancer, and chronic and acute respiratory diseases, including asthma. According to WHO, the combined effects of ambient and household air pollution are associated with 6.7 million premature deaths annually. Ambient (outdoor) air pollution is estimated to have caused 4.2 million premature deaths worldwide in 2019 (Faridi et al., 2023). During the last 15 years, however, consistent results, mainly from epidemiological studies, have provided evidence that current air pollutant levels have been associated with adverse carcinogenic and non-carcinogenic health effects, including increased mortality (Faridi et al., 2023).

According to the Pakistan Health Survey of 2009 to 2010, Pakistan is sixth on the list of South Asian countries that create ambient air pollution. However, according to IQAIR AIRVISUAL'S ranking, Pakistan's air pollution caused the country to be ranked second most polluted globally, with an annual $PM_{2.5}$ average of 74.3 μg/m³ (Rehan et al., 2023). These countries' use of unclean fuels and inefficient technologies are the primary cause of ambient air pollution (Schikowski et al., 2014). Lahore is Pakistan's second-largest city after Karachi, so it is of significant concern. According to the World Air Quality Report, Lahore ranked No. 1 in IQAIR AirVisual (Rehan et al., 2023).

Air pollution in Lahore is caused by vehicle and industrial emissions, smoke from brick kilns, burning crop residue and general waste, and dust from construction sites (Jalees & Asim, 2016). Other air pollution factors include the large-scale loss of trees when building new roads and buildings. Furthermore, it is to be mentioned that the high pollution level causes different health-related problems, from acute to chronic. It includes increased asthma hospitalization, decreased lung function, and increased prevalence and severity of wheezing and allergic rhinitis. Also, high levels of air pollution may cause perinatal disorders, infant mortality, respiratory disorders, allergies, increases in oxidative stress, and endothelial dysfunction (B.-J. Kim & Hong, 2012). According to the Pakistan Census 2017, Lahore's annual growth rate is 3% (Aslam, Shu, & Yaseen, 2023). The increasing population will increase the trend of urbanization. The increasing trend of urbanization will increase the demand for industries and modes of transportation and eventually become the center for industrial, vehicular, and gaseous emissions. These anthropogenic activities will emit large amounts of NO_x, SO_x, CO, PM, $PM_{2.5}$, PM_{10}, and O_3. These pollutants

cause adverse health effects, disturb the natural environmental cycle, and cause climate change. In this study, data is used to determine the source of pollutants, correlation, health effects, and air quality index (AQI). The meteorological data like wind speed, wind direction, humidity and temperature were also used atmospheric conditions.

2. Methodology:

The Punjab Environmental Protection Department provided the pollutants data, while the Punjab Meteorological Department provided meteorological data. This data was analyzed using different statistical tools like descriptive statistics, Pearson correlation, analysis of variance, principal component analysis, and cluster analysis. In addition, air quality index and health risk assessment (cancer and non-cancer), along with GIS mapping, were also performed. The sampling locations are given in Figure-1.

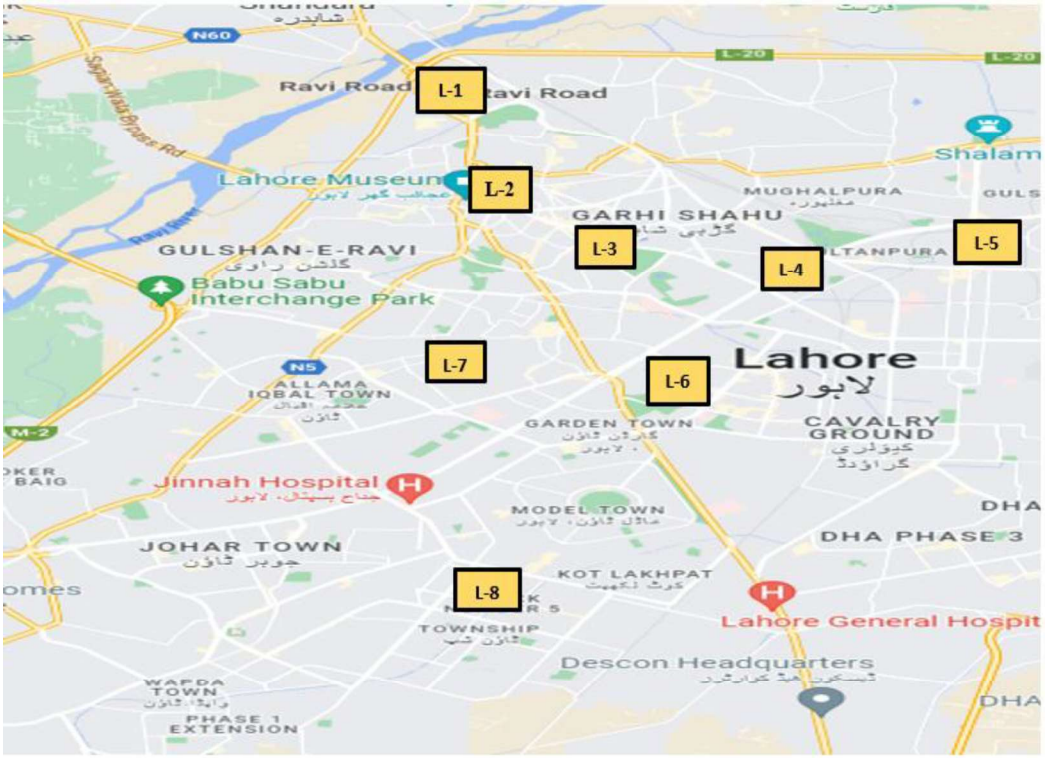


Figure-1: Location of sampling point used for the collection of air quality data

2.1 Air Quality Data Collection

The air quality data from 2007 to 2022 is collected from eight (08) locations within Lahore city. The locations are mentioned in Figure-1. The

concentration of PM, PM_{10}, $PM_{2.5}$, NO_X, NO, NO_2, CO, O_3, and SO_2 was analyzed.

2.2 Statistical Analysis of Data

Statistical tools identify the correlation and sources for PM, $PM_{2.5}$, PM_{10}, NO_X, NO_2, O_3, SO_2, and CO. Statistical tools include Descriptive statistics, Box Whisker plot, Pearson correlation, ANOVA, Cluster Analysis (CA), and Principal Component Analysis (PCA). Descriptive statistics and Box Whisker are used to check the validity of the data. Pearson correlation and ANOVA are used to identify the relationship between sources of pollutants, while PCA and CA are used to classify the sources further.

2.3 Air Quality Index

The Annual Air Quality Index was calculated using Pollutant concentrations. The US EPA method was used to calculate the Air Quality Index for 2007-2022. The calculated Air Quality Index will further help identify the impact of air quality parameters on Air Quality (Gupta & Dhir, 2022).

2.4 Analysis of Meteorological Data

After data collection from the Meteorological Department of Pakistan, the data of the years from 2007 to 2022 was used to determine wind speed, precipitation, and wind direction. The values are compared to analyze the impact of these climatic factors on the concentration of PM, $PM_{2.5}$, PM_{10}, NO_X, NO, NO_2, O_3, SO_2, and CO. The wind rose diagram (Jalees & Asim, 2016) is used to check the flow and direction of air currents.

2.5 Risk Analysis

This step is divided into two steps:

1. Risk Assessment
2. Risk Management.

2.5.1 Risk Assessment

The following steps are followed for risk assessment.

2.5.1.1 Hazard Identification

The presence of industries in the selected area and the smoke emission from the vehicles reflect the fact that the exposed population gets seriously affected. However, the EPA method is used to assess the hazard identifications.

2.5.1.2 Exposure Assessment

The exposure pathway is ambient Air, where pollutants are present, and the level of exposure is chronic or acute because the population breathes every second and is in contact with the pollutants. So, the factor on which population sensitivity is dependent is being exposed. One is the age factor; therefore, our concern is infants having less than 1. The evaluation of exposure pathways and routes following five elements of exposure pathways is considered (Yunesian, Rostami, Zarei, Fazlzadeh, & Janjani, 2019).

2.5.1.3 Source of Exposure

It includes industries, modes of transportation, and urbanization.

2.5.1.4 The Fate of Concerned Pollutants

These pollutants enter the atmosphere, react with other pollutants, and form compounds that can travel a long distance in the atmosphere and cause health impacts on the people of the surrounding areas.

2.5.1.5 Exposure Area

The exposure area is Lahore City.

2.5.1.6 Exposure Rout

The pollutants get into the body through inhalation and can cause respiratory disorders in respiratory tracts.

2.5.1.7 Potentially Exposed Population

People of all ages use Air for breathing, but as mentioned earlier, our concern is with the infant population. After the exposure evaluation, the quantification is done by considering the exposure frequency as 365-days/year.

2.5.1.8 Dose Response Analysis

Mathematical equations were used to determine the number of people at risk (either carcinogenic or non-carcinogenic risk).

$$\boldsymbol{Total\ Risk = Total\ Carcinogenic\ Risk + Total\ Non\ Carcinogenic\ Risk} \ldots (\mathbf{1})$$

2.5.1.9 Risk Characterization

The risk characteristics are the last step of risk analysis. It is concluded who is affected and what their likely impacts are based on health-integrated

outcomes of the exposure assessment and dose-response assessment, and then it is decided whether the outcomes are manageable.

2.5.2 Risk Management

The risk management for these pollutants suggested considering EPA recommendations.

2.5.3 Carcinogenic Risk Analysis

The carcinogenic risk that may be incurred due to exposure to PM, PM10, PM2.5, NOX, NO, SO2, O3, and CO in ambient Air is calculated in the equation below. (Yunesian et al., 2019)

$$\textbf{Carcinogenic Risk} = \textbf{CDI}\left(\frac{\textbf{mg}}{\textbf{kg}}.\textbf{day}\right) * \textbf{SF}\left(\textbf{kg}.\frac{\textbf{day}}{\textbf{mg}}\right) \ldots (2)$$

To calculate CDI, each value of the concentration of each year is multiplied by the constant factor that is calculated as

$$\boldsymbol{CDI = FxC} \ldots (3)$$

Where C is the concentration of pollutants, and F is the Constant Factor that is calculated as

$$\boldsymbol{Constant\ Factor} = \frac{\textbf{CR}\left(\frac{\textbf{m3}}{\textbf{day}}\right) * \textbf{EF}\left(\frac{\textbf{days}}{\textbf{year}}\right) * \textbf{ED(years)}}{\textbf{BW(kg)} * \textbf{AT(days)}} \ldots (4)$$

All the parameters were obtained from literature reviews or authentic websites. So, to calculate SF, use the following formula.

$$\boldsymbol{SF} = \frac{\textbf{Unit Risk}/(\textbf{mg}/\textbf{m3}) * \textbf{BW(kg)}}{\textbf{Constant Rate}\ (\frac{\textbf{m3}}{\textbf{day}})} \ldots . (5)$$

2.5.4 Non-Carcinogenic Risk Analysis

Non-carcinogenic risk is determined in terms of HI, which is the ratio of CDI to RfD. For estimating HI. So, HQ is calculated with the help of the following equations (Yunesian et al., 2019).

$$\boldsymbol{HQ} = \frac{\textbf{Chronic Daily Intake}_{\textbf{CDI}\left(\frac{\textbf{mg}}{\textbf{kg.day}}\right)}}{\textbf{Reference Dose}_{\textbf{RfD}}\left(\frac{\textbf{mg}}{\textbf{kg. day}}\right)} \ldots (6)$$

After calculating HQ, each non-carcinogenic Risk (HI) is calculated by summing up all HQ.

$$HI = \sum HQ \ldots (7)$$

For the calculation of HQ, a reference oral dose (RfD) is required. It is calculated as

$$\boldsymbol{R_fD} = \frac{\mathbf{C}\left(\frac{\mathbf{mg}}{\mathbf{m3}}\right) * \boldsymbol{CR}(\frac{\boldsymbol{m3}}{\boldsymbol{day}})}{\mathbf{BW(kg)}} \ldots (8)$$

2.5.5 Spatial Distribution

The spatial distribution method is used to identify the distribution of pollutants in Lahore. It is conducive to estimating the area's affected population. The size of circles at the specified locations represents the dispersion of concentration of pollutants at those locations. Different color coding is used to represent the concentration ranges at that location (Shen et al., 2020).

3. Results and Discussions

This chapter contains results obtained by performing different statistical tests on our data. The results, their interpretation, and the discussion are given in this chapter. The tool used for these tests is Minitab Software and Risk Analysis Excel sheet.

3.1 Statistical Analysis

Air quality data collected from EPD, Punjab, from 2007 to 2022 is summarized, and average values were used for the determination of various parameters (Table-1). The average value of each pollutant in each year is calculated. Table-1 shows the average concentration of each pollutant in each year, which is further used to assess the risk.

3.1.1 Descriptive Statistics

The mean values of the pollutants show that the maximum value is for NO_x, which means that NO_x has the highest emission value from sources (Table-1). The mean value after NO_x is for CO. These two pollutants have very high values, showing that most of the pollution is due to the transport sector. Vehicular emission is the primary source of these pollutants. The third number value of the mean is for PM (2.5 and 10), which is also experienced daily and increased by heart and lung diseases. The rest of the pollutants have relatively fewer mean values than these three. The standard error for different pollutant concentrations is also calculated. The standard error values show that most of the pollutants have precise values, and it also shows that the concentration values of these pollutants are centered

around the mean value. The variation in the concentration's dataset is less for NO, NO_2, O_3, SO_2, and PM. However, the error value for CO and NO_x is relatively higher, which shows a variation in the emission concentrations for these pollutants, and the values are not centered around the mean value. It gives us a clue that the concentrations for these pollutants are less precise (Jalees, Farooq, & Shah, 2021).

Similarly, the values of standard deviations also have the same trend, i.e., low for NO, NO_2, O_3, SO_2, and PM, while high for CO and NO_x. The sample variance also shows that the concentration values for NO, NO_2, O_3, SO_2, and PM are centred around the mean values, while for CO and NO_x, the sample variance is high. It shows a constant emission source for NO, NO_2, O_3, SO_2, and PM, while CO and NO_x vary in their sources.

Table-1: Descriptive Statistics for Pollutant Concentration Under Study

Parameters	Mean	Standard Error	Median	Mode	Standard Deviation	Sample Variance
CO	4513.99	829.66	4510	8211.97	2991.39	8948425
NO	49.6643	15.04	40.21	50.78	54.23	2941.68
NO_2	57.15	6.75	59.66	64.77	24.36	593.69
NO_x	49300.8	4726.7	41820	41820	17042.37	$2.9X10^8$
O_3	43.18	4.09	45.10	45.10	14.76	217.92
SO_2	55.65	14.51	33.06	33.06	52.34	2739.65
PM	116.73	16.67	92.11	92.11	60.11	3614.31
$PM_{2.5}$	61.47	4.03	52.85	52.85	14.54	211.44
PM_{10}	103.73	7.93	105.71	105.71	28.6	818.41

3.1.2 Box Whisker Plot

The box whisker plot (Figure-2) is used to check the validity of the data. The unequal sizing of these plots shows that the data is valid. The dark color shows the average of maximum mean values, while the light color shows the average of minimum mean values. The box plot shows that the values for CO are below the line and minimum most of the time. For NO, the values of concentrations are majorly below the line, and minor values are maximum. The box plot for the rest of the pollutants shows relatively less variation in the maximum and minimum values.

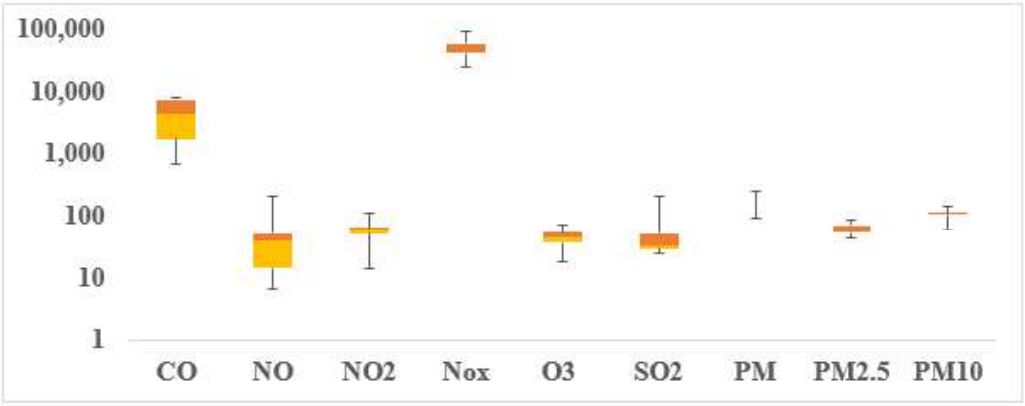


Figure-2: Box Whisker Plot for Pollutants Under Study

3.1.3 Pearson Correlation

Pearson's correlation analysis is a statistical technique used to measure the strength and direction of the linear relationship between two variables. The Pearson correlation coefficient, denoted by r, ranges from -1 to +1. A value of +1 indicates a perfect positive linear relationship between the two variables, while a value of -1 indicates a perfect negative linear relationship. A value of 0 indicates no linear relationship between the two variables (Jalees, Farooq, & Shah, 2021). Table-2 shows the correlation between the pollutants. There is a strong relationship between CO and CO, while it is moderately associated with NO, NO2, PM, PM2.5, and PM10 and weakly associated with NO_x, O_3, and SO_2. NO is strong with NO_2, moderate with $PM_{2.5}$ and O_3, and weak with NO_x, SO_2, PM, and PM_{10}. NO_2 is moderate with PM, $PM_{2.5}$, and PM_{10} and weak with NO_x, O_3, and SO_2. The table shows a moderate relation of NO_x with SO_2 and PM while a weak relation with O_3, $PM_{2.5}$, and PM_{10}. Table-2 indicates that the relation for O_3 is moderate with only PM_{10} while weak with SO_2, PM, and $PM_{2.5}$. The relation for SO_2 is moderate with $PM_{2.5}$ and PM_{10} while weak with PM. PM has moderate relation with $PM_{2.5}$ and PM_{10}, while $PM_{2.5}$ has moderate relation with PM_{10}. The overall relation between the sources of pollutants is identified to be weak to moderate. It can be said that the pollutants have the exact source of pollutants, which is an alternative hypothesis.

Table-2: Pearson Correlation Matrix showing weak to moderate correlation among pollutants under study

	CO	NO	NO_2	NO_x	O_3	SO_2	PM	$PM_{2.5}$	PM_{10}
CO	1.00								
NO	0.65	1.00							
NO_2	0.67	0.82	1.00						
NO_x	-0.09	0.21	0.25	1.00					
O_3	-0.18	-0.47	-0.17	-0.15	1.00				
SO_2	0.03	0.05	0.15	0.47	0.06	1.00			
PM	-0.43	-0.23	-0.43	0.52	-0.20	0.05	1.00		
$PM_{2.5}$	0.40	0.50	0.36	-0.04	-0.12	0.31	0.02	1.00	
PM_{10}	0.50	0.18	0.51	0.11	0.44	0.40	-0.32	0.52	1.00



3.1.4 Analysis of Variance (ANOVA)

Analysis of variance (ANOVA) is one of the most frequently used statistical methods in research. The need for ANOVA arises from the error of alpha-level inflation, which increases Type-1 error probability (false positive) and is caused by multiple comparisons. ANOVA uses the statistic Frequency (F), the ratio between and within group variances. The main interest of analysis is focused on the differences in group means. However, ANOVA focuses on the difference in variances (T. K. Kim, 2017). The maximum allowable error range that can claim "differences in means exist" can be defined as the significance level (α). It is the maximum probability of Type-I error that can reject the null hypothesis of "differences in means do not exist" in the comparison between two mutually independent groups obtained from one experiment. When the null hypothesis is true, the probability of accepting it becomes 1-α. Although various methods have been used to avoid hypothesis testing error due to significance level inflation, such as adjusting the significance level by the number of comparisons, the ideal method for resolving this problem as a single statistic is using ANOVA, an acronym for analysis of variance, and as the name itself implies, it is variance analysis. Table-3 shows that the mean values of pollutants are not the same. It does not support the null hypothesis, as the mean values of all the pollutants are not the same.

Table-3: Summary Results of ANOVA for the Pollutants under Study

Groups	Count	Sum	Average	Variance
CO	13	58681.95	4513.996	8948425
NO	13	645.637	49.66438	2941.683
NO_2	13	742.95	57.15	593.6931
NO_x	13	640910	49300.77	2.9E+08
O_3	13	561.38	43.18308	217.9284
SO_2	13	723.544	55.65723	2739.651
PM	13	1517.61	116.7392	3614.319
$PM_{2.5}$	13	799.2253	61.47887	211.4438
PM_{10}	13	1348.59	103.7377	818.4164

However, these variations can be due to measuring errors (Jalees, Irfan, Ali, & Batool, 2023). The p-value (Table-4) calculated in the ANOVA test is $4.19x10^{-47}$, less than the significance value (0.05). The value of F between the groups is 103.72 times the group variation. Table-4 shows that the null hypothesis is not genuine, as the value of F is not equal to 1. The comparison of the value of F with F_{crit} showed that the value of F is greater than F_{crit}. Thus, the result obtained from ANOVA proves the alternative hypothesis. Hence, the Null Hypothesis is rejected. It also shows that the difference in mean values is not due to a testing error.

Table-4: ANOVA Results for the Testing of Null Hypothesis

Source of Variation	SS	Df	MS	F	P-value	F_{crit}
Between Groups	$2.76x10^{10}$	8	$3.45x10^{9}$	103.7197	$4.19x10^{-47}$	2.02
Within Groups	$3.59x10^{9}$	108	33266893			
Total	$3.12x10^{10}$	116				

3.1.5 Principle Component Analysis

Principal component analysis, or PCA, is a dimensionality reduction method that is often used to reduce the dimensionality of large data sets by transforming a large set of variables into a smaller one that still contains most of the information in the large set. PCA should be used mainly for

strongly correlated variables. If the relationship is weak between variables, PCA does not work well to reduce data. Table-5 shows that the first three components have Eigenvalues> 1 and can be retained. The data is divided into three groups. The cumulative proportion of 1st group (PC-1) shows a 37% variation, while for PC-2 and PC-3, the cumulative is 58% and 76%, respectively.

The principal component matrix (Table-6) shows that PC-1 has a negative association with NO, NO_2, NO_x, SO2, PM, and $PM_{2.5}$ with unequal magnitudes while having a strong positive association with CO and PM_{10}. PC-1 has a strong, weak association with O_3. PC-2 has a positive association with CO, NO, and NO_2, whereas it has a negative strong association with NO_x, O_3, SO_2, and PM and a negative weak association with $PM_{2.5}$ and PM_{10}. The principal component, PC-3, has a positive relationship with every pollutant. The PCA shows that all the principal components have a value less than 0.7, which indicates that the pollutants may have the same source of origin (Jalees et al., 2023).

Table-5: Eigenvalue test Results of PCA for the pollutants under study

Eigenvalue	3.33	1.86	1.68	0.89	0.61	0.37	0.18	0.06	0.01
Proportion	0.37	0.21	0.19	0.1	0.1	0.04	0.02	0.01	0.01
Cumulative	0.37	0.58	0.76	0.86	0.93	0.97	0.99	0.99	0.1

Table-6: Principal Component Matrix for Pollutant Under Study

Variable	PC1	PC2	PC3
CO	0.452	0.153	0.142
NO	-0.452	0.158	0.331
NO_2	-0.49	0.025	0.087
NO_x	-0.073	-0.617	0.19
O_3	0.089	-0.36	0.574
SO_2	-0.162	-0.328	0.481
PM	-0.237	-0.553	0.035
$PM_{2.5}$	-0.352	-0.1	0.118
PM_{10}	0.364	-0.134	0.501

3.1.6 Cluster Analysis (CA)

A dendrogram is a graphical representation of hierarchical relationships between objects or variables. The technique used to create this structure is called cluster analysis, where objects or variables are grouped into clusters based on their similarities and differences. The dendrogram consists of

horizontal and vertical lines that connect the clusters. The horizontal lines represent the linkage between clusters, and shorter lines indicate stronger linkages. The vertical line distance between two parameters or clusters indicates the degree of dissimilarity, which tells us how weak or strong their linkage is. Dendrograms help us identify patterns and relationships within a large dataset by visually representing the hierarchical relationships between objects or variables.

Figure-3 gives us information about the strength of similarity between the groups. The cluster analysis is done to identify the naturally occurring groups with the data. The short lines indicate a high similarity level between the pollutants. This means there is a strong relationship between the sources of NO and NO_2, $PM_{2.5}$ and PM_{10}, and NO_x and PM, and the pollutants must be from the same source.

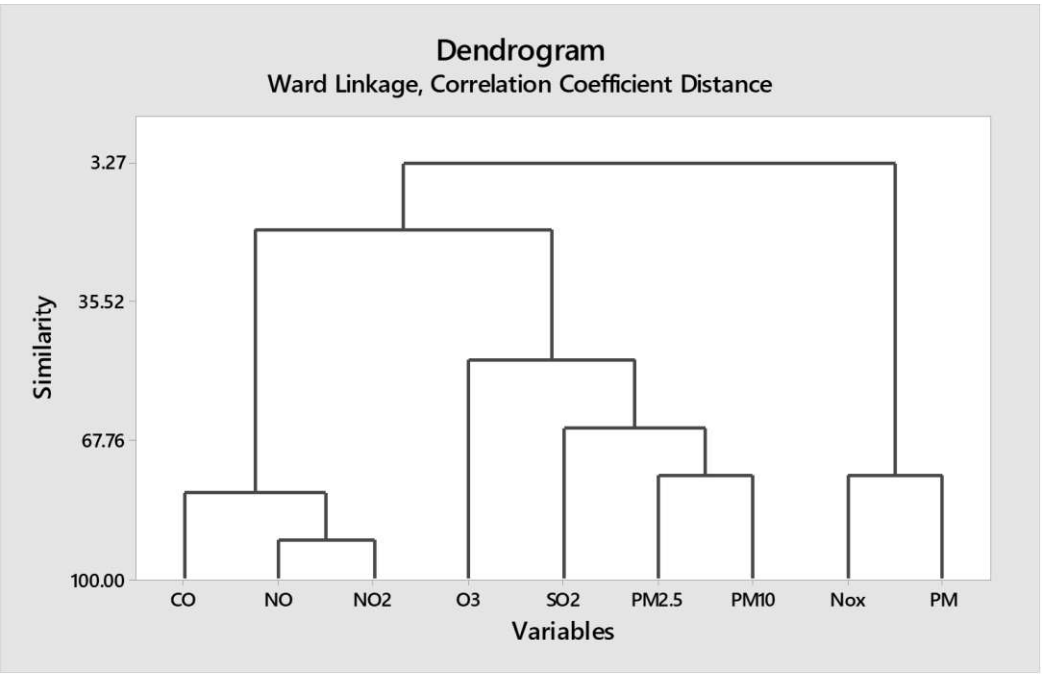


Figure-3: Dendrogram for Cluster Analysis for Pollutant Under Study

3.2 Air Quality Index

The air quality index (AQI) measures the quality of the Air we breathe in a specific location. It is a numerical value used to indicate how polluted the Air is and how it can affect human health. The AQI is based on the concentrations of major air pollutants, such as particulate matter (PM), ozone (O_3), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), and carbon monoxide (CO), in the Air.

The AQI is calculated using a scale ranging from 0 to 500, where higher values indicate greater pollution levels and potential health risks. The AQI is divided into six categories, each indicating a different level of health concern. These categories are:

- Good (0-50): Air quality is satisfactory and poses little or no risk to health.
- (51-100): Air quality is acceptable, but some pollutants may cause minor health problems for sensitive individuals.
- Unhealthy for Sensitive Groups (101-150): Sensitive groups, such as children, adults, and people with respiratory or heart conditions, may experience health effects.
- Unhealthy (151-200): Everyone may begin to experience health effects, and sensitive groups may experience more severe health effects.
- Very Unhealthy (201-300): Health warnings of emergency conditions. The entire population is likely to be affected.
- Hazardous (301-500): Health alert: everyone may experience more severe health effects.

3.2.1 Calculation of AQI

The formula used for the calculation of AQI is:

$$AQI = \frac{(I_1 - I_2)(CP - BP_2)}{BP_1 - BP_2} + I_2 \ldots (9)$$

CP = Concentration of Pollutant

I_1= High Index Value

I_2= Low Index Value

BP_1= High Break Point Value

BP_2= Low Break Point Value

The AQI values (Table-7) showed an increasing trend in the AQI values till 2019. In 2020, the AQI value decreased, and air quality improved slightly due to the pandemic when all the activities were paused. After the pandemic, the AQI value again increased, and air quality again deteriorated.

Table-7: Air Quality Index for Pollutants Under Study

Years	CO	NO_2	O_3	SO_2	$PM_{2.5}$	PM_{10}	AQI
2007	51	57	51	66	144	76	144
2008	17	49	44	109	144	76	144
2009	8	52	48	99	144	76	144
2010	22	52	102	38	159	76	159
2011	20	26	44	42	159	54	159
2014	32	68	43	36	144	76	144
2015	13	66	44	47	117	54	117
2017	87	116	44	71	168	78	168
2018	80	62	44	47	157	94	157
2019	88	68	44	75	144	76	144
2020	67	66	44	47	117	54	117
2021	88	62	50	35	159	94	159
2022	69	62	65	163	163	94	163



3.3 Wind Rose Diagram

A windrose diagram was used to determine the direction and speed of Air. The wind rose diagram is formed using data available in the supplementary material. The windrose diagram shows that most of the current has blown in the north and north-west directions, where the air quality monitoring devices are located. The location of industrial estates also confirms that the flow currents are from industrial areas to these locations. It shows that the people in these directions will be affected more than in the other directions. The spatial data also confirmed that the busiest and main roads of Lahore city are also located in the direction that is pointed out in the windrose diagram. Therefore, vehicular emission also has a dominant contribution to the rising pollution in these directions (Jalees & Asim, 2016).

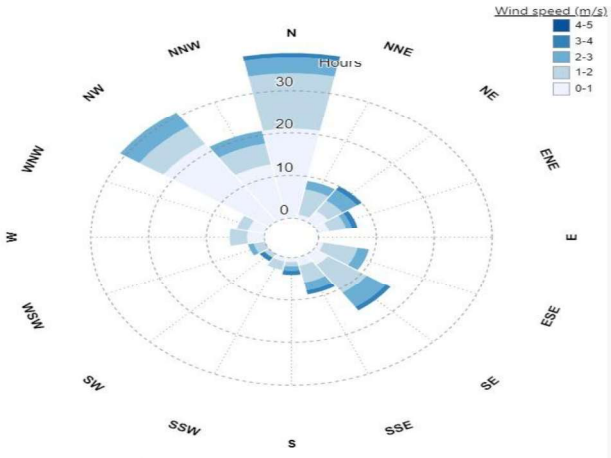


Figure-4: Windrose Diagram

3.4 Spatial Distribution Maps

The spatial distribution of pollutant concentration is crucial for assessing air quality, identifying areas of high risk, and implementing targeted measures to reduce pollution levels. Environmental agencies employ various techniques to study and monitor the spatial distribution of pollutant concentration, including air quality monitoring networks, remote sensing technologies, and computer models. These tools help understand the patterns and trends of pollutant dispersion, identify pollution sources, and formulate effective mitigation strategies.

In this study, distribution maps (Figure-5) of pollutants show the average concentrations of each pollutant from all locations over the years, which gives us an idea of the emission of pollutants. These maps help us identify areas with the most emissions. So, mitigation measures could be quickly and effectively implemented. The color coding is assigned to differentiate pollutant concentrations in distribution maps, as some pollutants have high concentrations while others have low concentrations (Jalees, Farooq, Anis, et al., 2021).

3.5 Risk Analysis

3.5.1 Constant Factor

Various equations mentioned in the previous section were used to calculate risk analysis. For the calculation of constant factor (F), the infants' inhalation rate was 8m3/day. The Exposure Frequency (EF) was 365 days per year as per USEPA. The exposure duration for infants is one year. The body weight for infants was 9.2 kg. The average time was taken as 70 years (25550 days). The F value was calculated, which was 0.0124 m^3/kg day. By multiplying the constant factor with concentrations of pollutants, the Chronic Daily Intake CDI for each pollutant was calculated (Table-8).

3.5.2 Slope Factor and Reference Oral Dose

For calculating the slope factor and reference oral dose, Eq. 5 & 8 were used. For Unit Risk, the average concentrations of pollutants are divided by the total flow rate of Air (TFR), which was 0.004998 m^3/sec. Unit Risk for each pollutant is given below in Table-9.

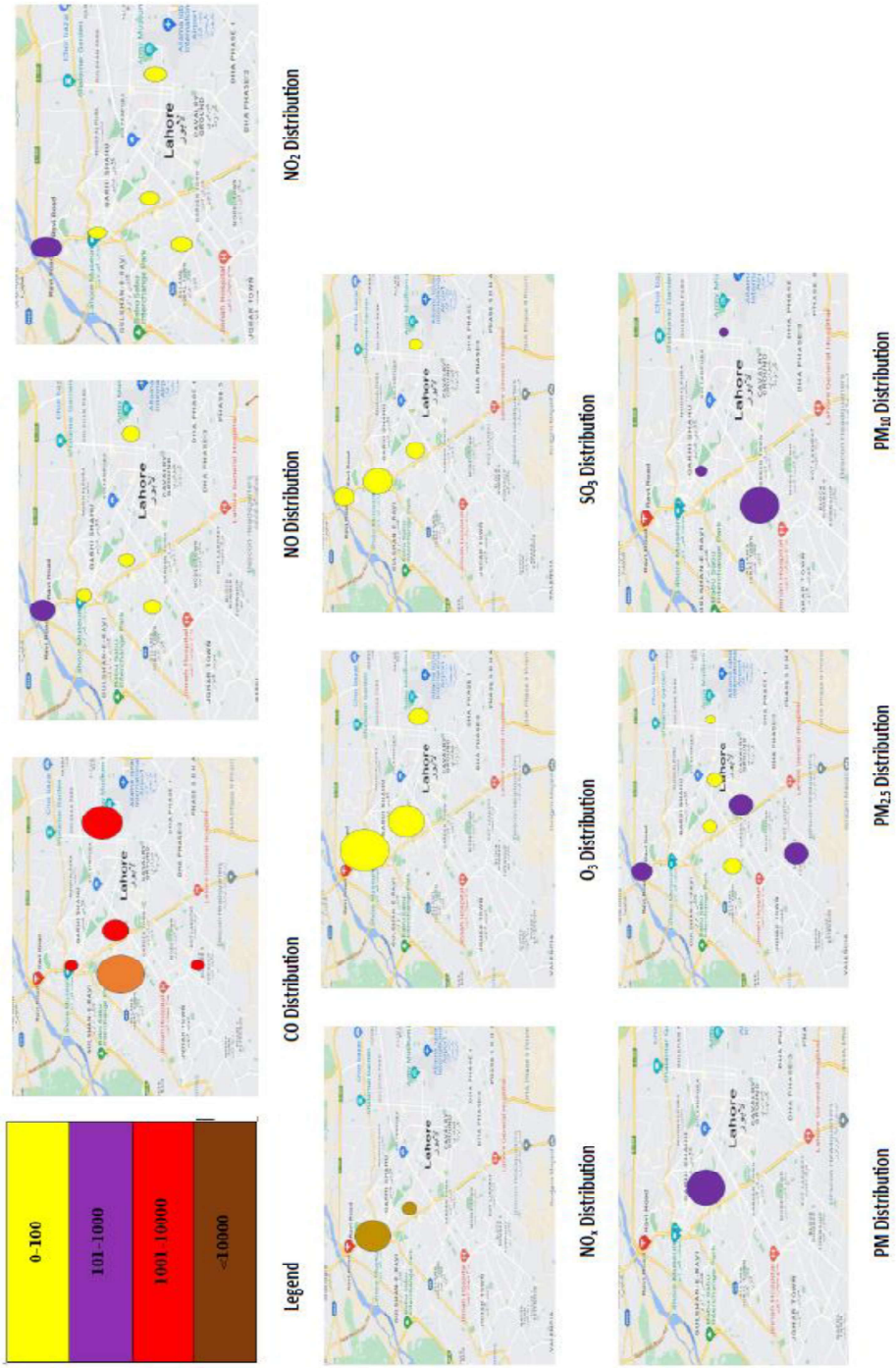


Figure-5: Distribution of Air Pollutants under Study on the Map of Lahore City using GIS

Table-8: Chronic Daily Intake value for Each pollutant under study

Pollutants	Avg. Concentration	Constant Factor	CDI
	ug/m^3	m^3/kg.day	ug/kg.day
CO	4206	0.0124	52.154
NO	49.5	0.0124	0.6138
NO_2	49.42	0.0124	0.6128
NO_x	46626	0.0124	578.16
O_3	42.83	0.0124	0.5311
SO_2	57.5	0.0124	0.713
PM	172.2	0.0124	2.1353
$PM_{2.5}$	67.09	0.0124	0.8319
PM_{10}	104.2	0.0124	1.2921

Table-9: Table showing values of Unit Risk, Slop Factor, and Reference Oral Dose

Pollutants	**Avg. Concentration ug/m3**	**Unit Risk ug/kg. day**	**SF kg. day/ug**	**R_fD for Infants**
CO	4206	841537	$1.37x10^{-06}$	3657.39
NO	49.5	9904	$1.1x10^{-04}$	43.04
NO_2	49.42	9888	$1.1x10^{-04}$	42.97
NO_x	46626	9328932	$1.23x10^{-07}$	40544.4
O_3	42.83	8569.4	$1.3x10^{-04}$	37.24
SO_2	57.5	11505	$1.3x10^{-04}$	50
PM	172.2	34454	$9.9x10^{-5}$	149.73
$PM_{2.5}$	67.09	13423	$3.34x10^{-05}$	58.34
PM_{10}	104.2	20848	$8.57x10^{-05}$	90.61

3.5.3 Carcinogenic Risk

The carcinogenic risk is calculated by multiplying each pollutant's CDI value with SF (Jalees, Farooq, & Shah, 2021). The Cancer Index or Cancer Risk Index is the sum of all Carcinogenic Risk values. In this study, CR and CI

values were calculated for every year (Table-10). The carcinogenic risk has varied a lot over the years. The lowest cancer risk was $2x10^{-5}$ for CO in 2015, and the highest was $3.5x10^{-4}$ for SO_2 in 2022. The cancer risk increases yearly due to increasing pollutant concentrations in the ambient Air. Hence, every year, the number of people exposed to cancer risk increases. The year-wise highest and lowest cancer risk is given in Table-10. There is a variation in the cancer risk values for every year, but still, all the highest risk values exceed the standard or acceptable cancer risk value, i.e., 1×10^{-6}.

3.5.4 Non-Carcinogenic Risk

The non-carcinogenic risk includes all the other health risks except for cancer. It is also termed a Hazard Quotient (Jalees, Farooq, & Shah, 2021). It is calculated using Eq. 6. The HQ for all pollutants of each year is given in Table-11. After the calculation of HQ, the next step is the calculation of the Hazard Index. The Hazard Index is the sum of all HQ values. Since we have calculated HI for each year, HQ values for all pollutants of a year are summed up (Table-11). The overall lowest HQ is 0.5 for CO, and the highest is 4.5 for NO. However, variation in the pollutant-wise HQ values is observed. The table indicated an overall increasing trend for HQ values over the years. The HQ values decreased in a few years but still have an increasing trend. All HQ's highest values exceed the standard or acceptable value of HQ, i.e., <1. In 2015, the HQ value was 0.99, which is considered 1, but it should be less than 1.

The overall CRI shows that one infant out of 100 is at risk of cancer, which is an alarming situation because the acceptable cancer risk is one person out of 10^6 population. Overall, HI is also exceeding the acceptable limit of 1. Therefore, the ambient Air in Lahore is very poor and is ranked among the top 5 most polluted cities worldwide.

Table-10: Cancer Risk and Cancer Risk Index for each year due to Pollutant Under Study

Cancer Risk	CO	NO	NO_2	NO_x	O_3	SO_2	PM	$PM_{2.5}$	PM_{10}	Cancer Index	Max CR	Min CR
2007	$7.66x10^{-5}$	$5.80x10^{-5}$	$8.62x10^{-5}$	$9.21x10^{-5}$	$9.15x10^{-5}$	$8.01x10^{-5}$	$1.14x10^{-4}$	$2.19x10^{-5}$	$1.13x10^{-4}$	$7.33x10^{-4}$	$1.14x10^{-4}$	$2.19x10^{-5}$
2008	$2.47x10^{-5}$	$5.34x10^{-5}$	$7.58x10^{-5}$	$1.48x10^{-4}$	$7.89x10^{-5}$	$1.57x10^{-4}$	$3.13x10^{-4}$	$2.19x10^{-5}$	$1.13x10^{-4}$	$9.86x10^{-4}$	$3.13x10^{-4}$	$2.19x10^{-5}$
2009	$1.14x10^{-5}$	$2.58x10^{-5}$	$8.00x10^{-5}$	$6.68x10^{-5}$	$9.02x10^{-5}$	$1.24x10^{-4}$	$1.14x10^{-4}$	$2.19x10^{-5}$	$1.13x10^{-4}$	$6.47x10^{-4}$	$1.24x10^{-4}$	$1.14x10^{-5}$
2010	$3.22x10^{-5}$	$2.05x10^{-5}$	$4.04x10^{-5}$	$3.95x10^{-5}$	$1.18x10^{-4}$	$4.44x10^{-5}$	$1.14x10^{-4}$	$2.97x10^{-5}$	$1.13x10^{-4}$	$5.51x10^{-4}$	$1.18x10^{-4}$	$2.05x10^{-5}$
2011	$2.98x10^{-5}$	$9.66x10^{-6}$	$2.08x10^{-5}$	$6.40x10^{-5}$	$4.23x10^{-5}$	$4.93x10^{-5}$	$3.13x10^{-4}$	$2.97x10^{-5}$	$6.45x10^{-5}$	$6.23x10^{-4}$	$3.13x10^{-4}$	$2.08x10^{-5}$
2014	$4.83x10^{-5}$	$1.95x10^{-5}$	$1.01x10^{-4}$	$6.40x10^{-5}$	$8.75x10^{-5}$	$4.21x10^{-5}$	$1.14x10_{4}$	$2.19x10^{-5}$	$1.13x10^{-4}$	$6.11x10^{-4}$	$1.14x10^{-4}$	$1.95x10^{-5}$
2015	$1.96x10^{-5}$	$1.81x10^{-5}$	$4.84x10^{-5}$	$7.05x10^{-5}$	$3.47x10^{-5}$	$5.50x10^{-5}$	$1.14x10^{-4}$	$1.75x10^{-5}$	$6.45x10^{-5}$	$4.43x10^{-4}$	$1.14x10^{-4}$	$1.75x10^{-5}$
2017	$1.38x10^{-4}$	$3.11x10^{-4}$	$1.66x10^{-4}$	$9.00x10^{-5}$	$3.02x10^{-5}$	$8.69x10^{-5}$	$1.14x10^{-4}$	$3.69x10^{-5}$	$1.16x10^{-4}$	$1.09x10^{-3}$	$3.11x10^{-4}$	$3.02x10^{-5}$
2018	$1.26x10^{-4}$	$7.32x10^{-5}$	$9.36x10^{-5}$	$6.40x10^{-5}$	$7.52x10^{-5}$	$5.51x10^{-5}$	$1.14x10^{-4}$	$2.81x10^{-5}$	$1.50x10^{-4}$	$7.80x10^{-4}$	$1.50x10^{-4}$	$2.81x10^{-5}$
2019	$1.39x10^{-4}$	$9.25x10^{-5}$	$1.01x10^{-4}$	$6.40x10^{-5}$	$7.63x10^{-5}$	$4.05x10^{-5}$	$1.14x10^{-4}$	$2.19x10^{-5}$	$1.13x10^{-4}$	$7.62x10^{-4}$	$1.39x10^{-4}$	$2.19x10^{-5}$
2020	$1.04x10^{-4}$	$1.03x10^{-4}$	$7.35x10^{-5}$	$6.40x10^{-5}$	$6.10x10^{-5}$	$5.73x10^{-5}$	$1.14x10^{-4}$	$1.75x10^{-5}$	$6.45x10^{-5}$	$6.59x10^{-4}$	$1.14x10^{-4}$	$1.75x10^{-5}$
2021	$1.39x10^{-4}$	$7.32x10^{-5}$	$9.36x10^{-5}$	$6.40x10^{-5}$	$7.52x10^{-5}$	$5.51x10^{-5}$	$1.14x10^{-4}$	$2.97x10^{-5}$	$1.50x10^{-4}$	$7.95x10^{-4}$	$1.50x10^{-4}$	$2.97x10^{-5}$
2022	$1.06x10^{-4}$	$7.32x10^{-5}$	$9.36x10^{-5}$	$9.00x10^{-5}$	$7.52x10^{-5}$	$3.59x10^{-4}$	$1.14x10^{-4}$	$3.29x10^{-5}$	$1.50x10^{-4}$	$1.09x10^{-3}$	$3.59x10^{-4}$	$3.29x10^{-5}$
Max CR (2007-2022)	$1.39X10^{-4}$	$3.11X10^{-4}$	$1.01X10^{-4}$	$1.48X10^{-4}$	$1.18X10^{-4}$	$3.59X10^{-4}$	$3.13X10^{-4}$	$3.69X10^{-5}$	$1.50X10^{-4}$	----	----	----
Min CR (2007-2022)	$1.14X10^{-5}$	$1.81X10^{-5}$	$2.08X10^{-5}$	$3.95X10^{-5}$	$3.02X10^{-5}$	$4.05X10^{-5}$	$1.14X10^{-4}$	$1.75X10^{-5}$	$6.45X10^{-5}$	----	----	----

Table-11: Hazard Quotient (HQ) and Hazard Index (JI) for Pollutants Under Study

Year	CO	NO	NO_2	NO_x	O_3	SO_2	PM	$PM_{2.5}$	PM_{10}	Max HQ	Min HQ	HI
2007	1.07	0.81	1.21	1.29	1.28	0.84	0.53	0.79	1.01	1.29	0.79	8.84
2008	0.35	0.75	1.06	2.08	1.11	1.64	1.46	0.79	1.01	2.08	0.34	10.24
2009	0.16	0.36	1.12	0.93	1.26	1.29	0.53	0.79	1.01	1.29	0.16	7.47
2010	0.45	0.29	0.57	0.55	1.65	0.46	0.53	1.07	1.01	1.65	0.29	6.59
2011	0.42	0.14	0.29	0.90	0.59	0.51	1.46	1.07	0.58	1.46	0.13	5.96
2014	0.68	0.27	1.41	0.90	1.23	0.44	0.53	0.79	1.01	1.41	0.27	7.26
2015	0.27	0.25	0.68	0.99	0.49	0.57	0.53	0.63	0.58	0.99	0.25	5.00
2017	1.94	4.36	2.32	1.26	0.42	0.91	0.53	1.33	1.05	4.36	0.42	14.12
2018	1.77	1.03	1.31	0.90	1.05	0.57	0.53	1.01	1.35	1.77	0.53	9.53
2019	1.95	1.30	1.41	0.90	1.07	0.42	0.53	0.79	1.01	1.95	0.53	9.39
2020	1.46	1.44	1.03	0.90	0.85	0.60	0.53	0.63	0.58	1.46	0.53	8.02
2021	1.95	1.03	1.31	0.90	1.05	0.57	0.53	1.07	1.35	1.95	0.53	9.77
2022	1.49	1.03	1.31	1.26	1.05	3.75	0.53	1.18	1.35	3.75	0.53	12.96
Max HQ (2007-2022)	1.95	4.36	2.32	2.08	1.65	3.75	1.46	1.33	1.35			
Min HQ (2007-2022)	0.16	0.13	0.29	0.55	0.42	0.42	0.53	0.63	0.58			

4 Conclusions:

- The Air Quality data for 2007-2022 showed that most of the pollutant concentrations exceed the PEQs and WHO guideline values. The increasing trend of pollutant concentration is $O_3<NO<SO_2<NO_2<PM_{2.5}<PM_{10}<PM<CO<NO_x$.
- The Statistical Tools results show that data is non-linear and valid. Pearson and ANOVA results show a Weak-Moderate Relation. CA & PCA results represent three significant pollution sources: Industrial emission, traffic, and urbanization.
- The Air Quality Index shows that Air Quality deteriorated by ***69%*** from **2007** to **2022**. Also, most of our air quality is highly polluted and has a high risk to health and the environment.
- The Risk analysis results show a high Carcinogenic and Non-Carcinogenic Risk. HQ shows that *NO* has the highest value (***4.5),*** whereas *CO* has the lowest (***0.5)*** during the study period. HI ranges from ***5-14.1,*** with the highest value of *14.1* was observed in 2017. CR is lowest for *CO,* $1x10^{-5}$, and highest for SO2, $3.5x10^{-4}$.
- CI showed that *one person per 1000* population may have cancer risk.
- Based on the deteriorating air quality in Lahore, the Government must make some legislative measures, implement these measures, and create awareness among people about the deterioration in air quality.

5. References:

Aslam, R. W., Shu, H., & Yaseen, A. (2023). Monitoring the population change and urban growth of four major Pakistan cities through spatial analysis of open source data. *Annals of GIS*, 1-13.

Cepeda, M., Schoufour, J., Freak-Poli, R., Koolhaas, C. M., Dhana, K., Bramer, W. M., & Franco, O. H. (2017). Levels of ambient air pollution according to mode of transport: a systematic review. *The Lancet Public Health, 2*(1), e23-e34.

Faridi, S., Krzyzanowski, M., Cohen, A. J., Malkawi, M., Moh'd Safi, H. A., Yousefian, F., . . . Niazi, S. (2023). Ambient air quality standards and policies in eastern mediterranean countries: a review. *International journal of public health, 68*, 1605352.

Gupta, A., & Dhir, A. (2022). Spatial and temporal variations of air pollutants in urban agglomeration areas in Gujarat, India during 2004–2018. *MAPAN, 37*(1), 215-226.

Jalees, M. I., & Asim, Z. (2016). Statistical modeling of atmospheric trace metals in Lahore, Pakistan for correlation and source identification. *Environmental Earth Sciences, 75*(9), 842.

Jalees, M. I., Farooq, M. U., Anis, M., Hussain, G., Iqbal, A., & Saleem, S. (2021). Hydrochemistry modelling: evaluation of groundwater quality deterioration due to anthropogenic activities in Lahore, Pakistan. *Environment, Development and Sustainability, 23*, 3062-3076.

Jalees, M. I., Farooq, M. U., & Shah, A. T. (2021). Cancer risk assessment and modeling of groundwater contamination near industrial estate, Lahore, Pakistan. *Journal of Water, Sanitation and Hygiene for Development, 11*(2), 314-326.

Jalees, M. I., Irfan, I., Ali, A., & Batool, M. (2023). Spatial distribution and water quality index of Lahore Canal, Pakistan. *Water Practice and Technology*.

Kim, B.-J., & Hong, S.-J. (2012). Ambient air pollution and allergic diseases in children. *Korean journal of pediatrics, 55*(6), 185.

Kim, T. K. (2017). Understanding one-way ANOVA using conceptual figures. *Korean journal of anesthesiology, 70*(1), 22-26.

Madronich, S., Sulzberger, B., Longstreth, J., Schikowski, T., Andersen, M. S., Solomon, K., & Wilson, S. (2023). Changes in tropospheric air quality related to the protection of stratospheric ozone in a changing climate. *Photochemical & Photobiological Sciences*, 1-48.

Rehan, S. T., Iftikhar, A., Khan, S., ul Hussain, H., Ullah, I., Naseer, U., & Asghar, M. S. (2023). Air pollution a hidden scourge to cardiovascular health: insight from top polluted cities of Pakistan. *International Journal of Surgery, 109*(3), 553-554.

Rohde, R. A., & Muller, R. A. (2015). Air pollution in China: mapping of concentrations and sources. *PloS one, 10*(8), e0135749.

Schikowski, T., Mills, I. C., Anderson, H. R., Cohen, A., Hansell, A., Kauffmann, F., . . . Sunyer, J. (2014). Ambient air pollution: a cause of COPD? *European Respiratory Journal, 43*(1), 250-263.

Shen, F., Zhang, L., Jiang, L., Tang, M., Gai, X., Chen, M., & Ge, X. (2020). Temporal variations of six ambient criteria air pollutants from 2015 to

2018, their spatial distributions, health risks and relationships with socioeconomic factors during 2018 in China. *Environment International, 137*, 105556.

Yunesian, M., Rostami, R., Zarei, A., Fazlzadeh, M., & Janjani, H. (2019). Exposure to high levels of PM2. 5 and PM10 in the metropolis of Tehran and the associated health risks during 2016–2017. *Microchemical Journal, 150*, 104174.