第 51 卷第 03 期
2024 年 3 月
湖南大学学报（自然科学版）
Journal of Hunan University（Natural Sciences）
Vol. 51. No. 03.
March 2024

Open Access Article

AMBIENT AIR QUALITY AND COAL MINING COMPLEX INDEX ARE ASSESSED USING CHEMOMETRIC METHODS

Munneswari Durga#[1], G. Siva Nageswara Rao#[2], V. Gopi#[2], M. Harsha Vardhan#[2] ,V. Harisainath Reddy#[2]

[1]Assistant Professor, [2]UG Scholar, Department of Mining Engineering, Godavari Institute of Engineering & Technology (A), Chaitanya Knowledge City, Rajahmundry, Andhra Pradesh, India.

ABSTRACT: This study aims to analyze the regional variation in the source of air pollution, identify the percentage contribution of each pollutant, and distribute the mass contribution of each source category using multivariate analysis. The nine air monitoring sites were successfully divided into three groups using hierarchical agglomerative cluster analysis (HACA) (clusters 1, 2, and 3). The collected meteorological data is non-parametric data for the years 2020–2021 which includes PM2.5, PM10, SO2, NO2, NO, NOx, CO, wind speed, humidity, wind direction, temperature, cloud cover, and surface radiation. The most major air pollution sources were identified using Factor Analysis (FA). Multiple linear regression (MLR) and principal component regression (PCR) were utilized to create an equation model explaining the contaminants' impact in each cluster. However, it was shown that the most important pollutants impacting the value of the air pollutant index (API) are gaseous pollutants (NOx and SO2) and particulate matter (PM10 and PM2.5). Gas and non-gas pollutants have a 65% influence on cluster 1 and meteorological conditions have a 35% effect. Cluster 3 is influenced by 65% particle and non-gas pollutants and 35% weather conditions, compared to Cluster 2 which is 100% affected by gas and particulate pollutants because of its spatial location. This study shows the value of the multivariate modeling technique in minimizing the time and expense associated with monitoring redundant stations and parameters.

1. INTRODUCTION

Open-pit coal mines pose considerable issues for air pollution management and control. The extraction of coal from the earth harms the environment, people, and ecology (Agathokleous et al., 2022; Yang et al., 2022; Zipper and Skousen et al., 2021). Currently, most coal mines have continuous ambient air quality monitoring systems (CAAQMS) installed within 1.5 kilometers. These technologies constantly monitor air quality and create massive amounts of data. To eliminate ambiguity, clarify problems, and demonstrate regional differences, vast, complicated data sets from stations monitoring air quality in the atmosphere must be combined with current, reliable statistical approaches. The Air Quality Index (AQI) is a critical tool for determining air quality in a specific location (Kumar, 2022; Wang et al., 2022; Wu et al., 2013). It works by transforming the amount of trash to dimensionless numbers.

Several research have used chemometric modeling of the AQI to estimate the principal sources of air pollutants and their regional distribution (Barjoee et al., 2023; Diana et al., 2022; Galán-Madruga et al., 2023). Chemometrics is the use of statistical or mathematical approaches to create a correlation between measurable values derived from chemical principles or data and the essential parameter. The

Received: February 04, 2024 / Revised: February 27, 2024 / Accepted: March 06, 2024 / Published: March 11, 2024

About the authors：Munneswari Durga

Corresponding author- Email:

bulk of chemometric analysis applications take place in industrial and urban settings (Azid et al., 2015; Nunes et al., 2019; Rani et al., 2017. Certain published research use industrial chemometric analysis to identify the path of pollution flow inside a system (Grabowski et al., 2021; Vakarelska et al., 2021). Numerous studies have assessed the amount of respirable silica released by coal plants. Further research is needed on the use of chemometric approaches to monitor the dispersion of pollutants resulting from diverse mining operations. In this study, the AQI was linked to different concentrations of air contaminants and meteorological data using mathematical techniques.

The AQI is calculated with contaminant and cutoff concentrations as its base. However, weather conditions have the largest influence on local air quality. The majority of mining businesses have built a climate and air quality management system (CAAQMS) to monitor air pollution and temperature. As a result, chemometric approaches can be used to create models for calculating the AQI and conducting complete air quality evaluations. This study looks into the effects of weather and contaminants on air quality in a 330 km2 area occupied by nine open-pit mines. This study aims to create model equations for calculating AQI using Factor Analysis (FA), Cluster Analysis (CA), and classification models (Dragović and Mihailović, 2009; Hooper and Peters, 1989; Huang et al., 2009; Wold et al., 1987). The research hope to use principle component regression (PCR) and multiple linear regression (MLR) to build an equation that integrates meteorological variables, which have a substantial impact on air quality, in order to determine the air quality index (AQI). Furthermore, it shows how various statistical techniques have helped us understand how chemometric analysis has increased our awareness of air pollution in coal mines.

2. STUDY AREA

The study's research location is Singrauli, a region in central India known for its number of abandoned coal mines. Northern Coalfield Ltd manages both this mine and the coalfield. The research area's longitude and latitude coordinates are 82°30'54.71" E to 82° 47'56.13" E and 24°14'06.24" N to 24°05'02.63" N, respectively. Approximately 1.2 million people live in the vicinity of the coalfield. The region is made up of two primary basins. The Moher Sub-Basin has an estimated 6.83 billion tons (BT) of coal reserves, whereas the Singrauli Main Basin possesses 3.23 billion tons (BT) (Javed et al., 2021). Singrauli, Madhya Pradesh's coal mine region receives an average of 1119.65 millimeters of rain each year. The ambient temperature ranges from 47.2 to 4 degrees Celsius. The figure depicts a significant chunk of the mining sector that stretches into Uttar Pradesh's Sonbhadra district. 1. The latitude and longitude coordinates of the nine NCL mines involved in the study are as follows. The related values are shown in Table 1. Within Figure. 1. These mines are denoted with an exclamation sign (*). Four huge power plants located on the outskirts of the mining sector provide the state with a significant amount of energy.

Table 1. The exact position of the CAAQMS Stations in the Singrauli coal mine complex.

Sl.No.	Project	District	State	Latitude	Longitude
1	AMLOHRI PROJECT	Singrauli	MP	24° 05' 56.24" N	82° 36' 17.50" E
2	BINA PROJECT	Sonbhadra	MP/UP	24° 09' 05.20" N	82° 46' 27.40" E
3	BLOCK-B PROJECT	Singrauli	MP	24° 12' 18.68" N	82° 35' 30.88" E
4	CETI (DUDHICHUA)	Singrauli	MP	24° 12' 24.18" N	82° 40' 16.59" E
5	JAYANT PROJECT	Singrauli	MP	24° 06' 56.00" N	82° 39' 24.00" E
6	JHINGURDA PROJECT	Singrauli	MP	24° 11' 48.10" N	82° 42' 13.00" E
7	KAKRI PROJECT	Sonbhadra	UP	24° 10' 25.83" N	82° 45' 48.55" E
8	KHADIA PROJECT	Sonbhadra	MP/UP	24° 07' 20.00" N	82° 41' 04.20" E
9	NIGAHI PROJECT	Singrauli	MP	24° 06' 28.23" N	82° 37' 42.44" E

3. METHODS

Data Collection

As stated in Table 1, it came from nine continuous monitoring networks installed in the central control centers of the ambient air quality and management stations (CAAQMS) of the nine Singrauli coalfield complex mines. The dataset includes gaseous and non-gaseous contaminants, as well as meteorological data gathered annually between January 1, 2020 and December 31, 2020. Every day at 0 hours and 15 minutes, CAAQMS data was gathered. Using this data, the daily 24-hour average was computed. Prior until 24:00. in 2020.

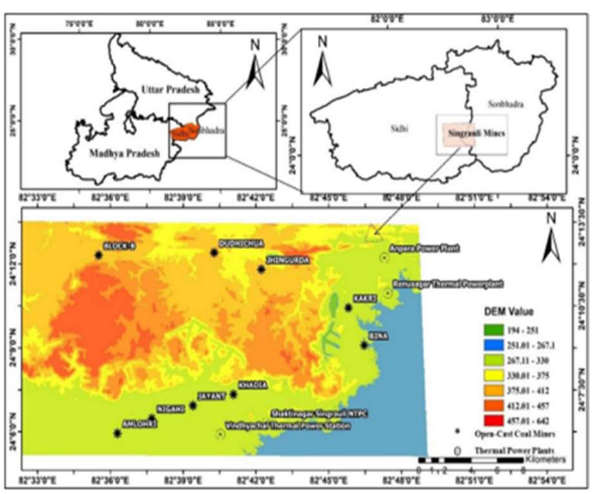

Figure 1. Mining coal is a difficult process.

Table 2. The table in question is Table 2. Meteorological parameters are recognized. Sl. No. Amlohri Bina Dudhichua Bloc borders

Sl. No.	Parameters	Amlohri	Bina*	Dudhichua	Block-B	Jayant	Jhingurda	Kakri*	Khadia	Nigahi
1	DBT(C)	✓	✓	✓	✓	✓	✓	✓	✓	✓
2	RH (%)	✓	✓	✓	✓	✓	✓	✓	✓	✓
3	WS(m/s)	✓	✓	✓	✓	✓	✓	✓	✓	✓
4	WD ()	✗	✓	✓	✓	✓	✓	✗	✓	✓
5	HR (kWh/m^2)	✓	✓	✓	✗	✓	✓	✓	✓	✓
6	Rainfall (mm)	✓	✓	✓	✓	✓	✓	✗	✓	✓
7	CO ($\mu g/m^3$)	✓	✓	✗	✓	✓	✓	✓	✗	✓
8	NO_2 ($\mu g/m^3$)	✓	✓	✓	✓	✓	✓	✓	✓	✓
9	NO ($\mu g/m^3$)	✓	✓	✓	✓	✓	✓	✓	✓	✓
10	SO_2 ($\mu g/m^3$)	✓	✓	✓	✓	✓	✓	✓	✓	✓
11	PM_{10} ($\mu g/m^3$)	✓	✓	✓	✓	✓	✓	✓	✓	✓
12	$PM_{2.5}$ ($\mu g/m^3$)	✓	✓	✓	✓	✓	✓	✓	✓	✓

The stations were around 1.5 kilometers from the central excavation site. The results revealed pollutants such as PM2.5, PM10, SO2, NO2, NO, NOx, and CO. The quantification of PM2.5 and PM10 was accomplished by absorbing energy that passed through the particle-collecting filter membrane. This was performed using a beta gauge and the beta-ray attenuation method. The Pulse Fluorescence Analyzer served as the foundation for quantifying SO2. Non-dispersive infrared and chemiluminescence were used to detect NO2, NO, NOx, and CO, respectively. The following meteorological variables were measured using instruments: temperature, humidity, precipitation, wind velocity, wind direction, and sun radiation.

Pollutants in the atmosphere include CO (carbon monoxide), NO2 (nitrogen dioxide), NO (nitrogen oxide), SO2 (sulfur dioxide), WS (wind direction), HR (horizontal solar radiation), CC (cloud cover), and PM (particulate particles with a size of 10 to 2.5 µm). The earliest evidence for this coal mine dates back only six months.

Organization of Data

Factor and cluster analyses were performed using each miner's daily average data. 4.74 percent of the collected data remain unidentified. Missing data are supplied using IBM SPSS 26.0.0.0 64-bit software and a mixed technique of multivariate imputation with interpolation (Junninen et al., 2004).

Chemometric Analysis

Hierarchical Agglomerative Cluster Analysis (HACA)

Cluster analysis is an independent method for organizing large amounts of data by categorizing them into smaller groups known as clusters based on their similarities or differences. Recent study in this topic includes Isiyaka et al. (2015), Ramson et al. (2016), and Too et al. (2011). On a daily basis, the HACA examines an average of thirteen factors, including seven pollution levels and six meteorological variables gathered from all nine mines. The degree of homogeneity is depicted in a dendrogram plot using Ward's technique and Euclidean distance values. The technique was carried out using the 64-bit version of IBM SPSS 26.0.0.0. Equation (1) gives the following formula for Euclidean distance (Dij):

$$D_{ij}$$
$$\sqrt{(x_{i1}-x_{j1})^2+(x_{i2}-x_{j2})^2+\cdots+(x_{im}-x_{jm})^2} \quad (1)$$

For example, if i and j are two observations and x1, x2,..., xm is the number of observations, the distance can be calculated. Ward's method, on the other hand, uses ANOVA to separate two groups and verify that the sum of their squares is minimal (Azid et al. 2015).

Factor Analysis

The goal of factor analysis is to uncover correlations between variables while minimizing the impact of several factors on the overall outcome of the variables (mutalib et al., 2013). This method, like Principal Component Analysis (PCA), is also thorough. PCA creates new variables by linearly combining observed variables, whereas FA factors identify linear measurements of observed variables. The approach determines FA for us.

$$F_{ij} = \sum_{j=1}^{m} C_{fj} f_{ji} + E_{fi}$$

In this equation, i denotes the number of samples, j the number of variables, m the total number of factors, F the variation or error, C the factor loading, f the factor value, and E the factor value.

The principal component method (PCM) is the most widely used factor analysis (FA) technique. PCM seeks to find fundamental patterns in data by first locating the element with the most variation and then determining which factor closely follows in terms of variance. For the purpose of explanation, the Principal Components (PCs) created by PCM are rotated using Varimax. The varimax spin's eigenvalues are used as fundamental components in factor analysis. An eigenvalue greater than one implies that the factor is a varimax factor (VF). Values greater than 0.75 imply that VFs have significant factor loadings. Key components in this study are those with factor loadings greater than 0.75. Juahir et al. (2011) and Azid et al. (2015) both do this. This study uses IBM SPSS 26.0.0.0 64-bit edition software to successively apply FA (PCM) to thirteen factors. The AQI stands for Quality of Air Index.

The Air Quality Index (AQI) This website provides timely information about air quality. It simplifies the relationship between pollution data for different air contaminants, labels, and hues by consolidating them into a single number. Equation (3) calculates the National Ambient Air Quality Standards (NAAQS). The NAAQS is used to determine the AQI for a specific pollution level.

$$I_p = \frac{I_{HI}-I_{LO}}{BP_{HI}-BO_{LO}}\,(C_P - BP_{LO}) + I_{LO}$$

where BPHI and BPLO are the concentrations at which the reaction happens when BPHI > Cp and BPLO < Cp. Cp is the normalized concentration of pollutant p, and Ip is the pollutant index. The AQI values IHI and Ilo are equivalent to BPHI and BPLO.

This index is calculated at a monitoring station by average the amounts of specific pollutants throughout the next twenty-four hours. According to the Central Pollution Control Board (CPCB), the AQI calculation requires data for at least three contaminants. At least one of the pollutants must be PM2.5 or PM10. This guarantees that the measurement accurately reflects the air quality. The AQI for this inquiry was calculated using five important pollutants: SOx, NOx, CO, PM2.5, and PM10.

Multiple Linear Regressions (MLR)

Many scientists use MLR to analyze the atmosphere. Using observed data, one can design a linear equation to investigate the interplay of independent and dependent components (Aertsen et al., 2010; Azid et al., 2015; Dominick et al., 2012). This study used the MLR approach to demonstrate the most important relationship between AQI data, weather conditions, and pollution levels. MLR is determined by using the following equation:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + \varepsilon_i$$

Where i = 1….n, β_0, β_1, and β_k

symbolize the regression coefficients, ε denotes the regression error, and X1 and Xk are independent factors.

Three metrics can be used to determine the extent to which each measure influences the AQI: adjusted R2, root mean square error (RMSE), and coefficient of determination (R2). In addition, the varimax rotation variables have been included as separate elements in the AQI calculations.

4. RESULTS AND DISCUSSION

Air Quality Status of the Study Area

The National Ambient Air Quality standard was validated using yearly average data on contaminants collected. Figure 2 depicts the ambient PM2.5, PM10, SO2, and NO2 concentrations for 2020 and 2021. Figure 3 depicts the change of CO concentrations over an eight-hour period. PM2.5 concentrations surpassed acceptable limits in four of the nine mines tested. Bina and Block B were the most populous areas. In contrast, PM10 concentrations exceed the ambient air quality level. The remaining two pollutants, SO2 and NO2, were too low to be considered harmful in any of the nine mines. Except for Block-B and Jayant, all of the average CO concentrations were below the allowed levels. Vindhyachal Thermal Power Plant is six kilometers southeast of Amlohri, at latitude 24° 06' 56.00" N and longitude 82° 39' 24.00" E. The wind blows east-southeast. Thus, contaminants produced by thermal power plants can be removed. Regardless, Jayant is cleaner than the neighboring mine due to the prevalent northwest breeze. Amlohri, Bina, Block-B, and Jayant had east-southeast winds. In contrast, Dudhichua and Khadia had more West-South-West (WSW) breezes. Wind patterns between Kakri and Jhingurda are similar, although Nigahi sees WNW winds for long stretches of the year.

However, the vehicles, bulldozers, payloaders, cranes, and other large earthmoving equipment used in coal mining and transportation are all driven by diesel fuel. Diesel is the principal source of gaseous pollutants, including CO, SO2, NO, and NO2. Cowherd et al. (2013) link heavy machinery emissions to poor air quality, particularly in urban areas.

Similarly, diesel engines are known for generating large amounts of dangerous particles and pollutants, which harm both human health and the environment (Ghose and Majee, 2000). Vehicles, bulldozers, payloaders, cranes, and other heavy equipment used in mines to carry coal and overburden emit the most CO pollution due to incomplete combustion of their fuel. According to Nie et al. (2022), a major amount of the carbon monoxide emissions from trucks, bulldozers, and other mining equipment used for coal and waste transportation are caused by incomplete fuel combustion. When machines work, they emit CO and other pollutants straight into the atmosphere. Jayant, one of the largest open-pit mines, generates 25 million tons of coal per year. The mine's high output involves the employment of numerous automobiles to transport overburden and coal, potentially increasing carbon dioxide emissions into the environment. Block B is located in the northwest quarter, in a densely populated neighborhood.

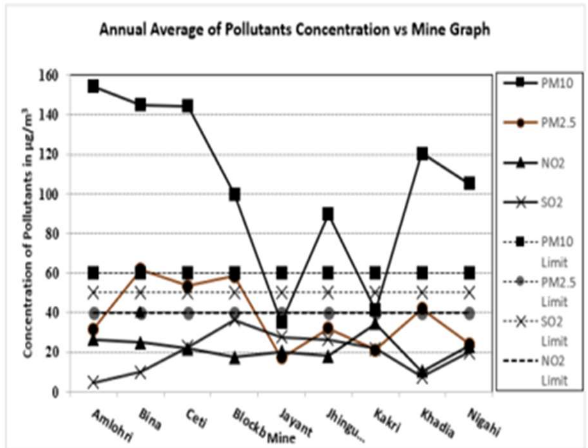

Figure 2. Visual 2. The Singrauli Coal Complex's nine mines generate an average amount of trash per year.

East-southeasterly (ESE) gusts in the market area surrounding the Continuous Ambient Air Quality Monitoring Station (CAAQMS) transport automobile emissions to the station. This may explain the mine's high CO2 emissions.

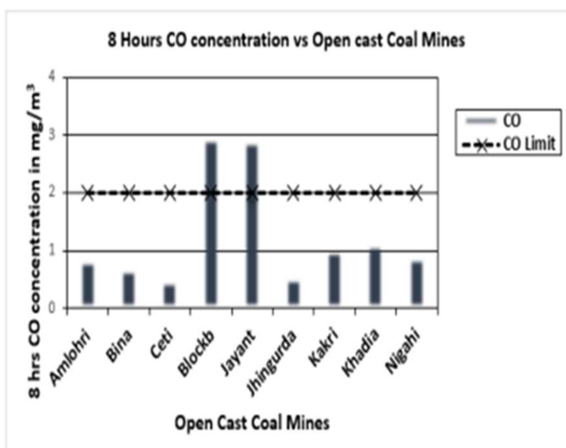

Figure 3. Forty-eight hours. CO content for each of the nine mines.

Fig. The AQI scores for the mine complex are displayed. 4. The first three coal mines in Cluster 1, Amlohri, Nigahi, and Khadia, had an average AQI of 153. These mines produce around 16 million tons of coal annually and occupy the greatest land. Cluster-2, located in the northernmost portion of the research area and home to mines such as Block-B, Dudhichua, Jhingurda, and Bina, has the highest air quality rating of 224. Cluster-3 has the poorest air quality of the three categories, with a score of 120. This group includes the Jayant and Kakri mines, which generate only 44 million cubic meters of overburden (OB) each year. Dudhichua has the most AQI fluctuation, whereas Jayant has the lowest.

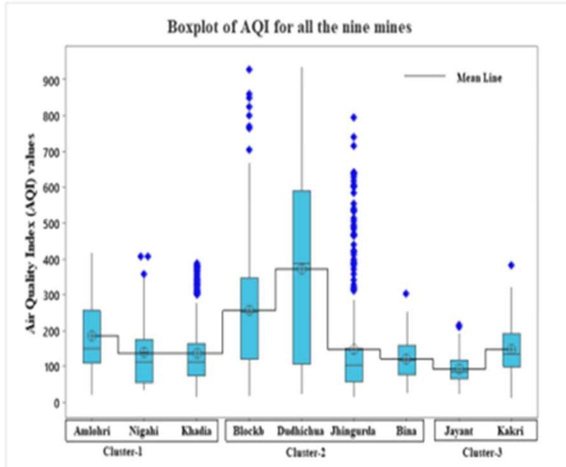

Figure 4. A box plot showing the AQI readings in 2019 and 2020.

Spatial Classification of Mines Based on Air Quality Parameters

Air quality parameters are used to categorize mine sites. The similarities and differences between air quality parameters and meteorological conditions were determined using HACA. Individuals who were very similar in space were placed together. As illustrated in Figure 5, this method resulted in the formation of three separate categories.

Cluster 1's annual average AQI is 153, suggesting that the area is highly polluted. Amlohri, Khadia, and Nigahi are located in the southwestern section of the mine complex and have elevations ranging from 194 to 251 meters above mean sea level. Cluster 2, which includes Jhingurda, Bina, Dudhichua, and Block B, is located at a higher elevation than the other mines in the mining complex. With an annual average AQI score of 224, the site is classified as very polluted. Cluster 3, with an annual AQI rating of 120, is the least polluted of the three sites. This group includes the two most prominent mines, Jayant and Kakri. Table 3 shows the mean values of several parameters for each of the three groups, as well as the corresponding open-pit coal mines. Wang et al. (2018) found additional air quality locations in networks and reached the same finding. Gouveia et al. (2015) use wavelet-based clustering approaches, which are successful for grouping geographical stations in a manner congruent with the exploratory methodologies used in this work.

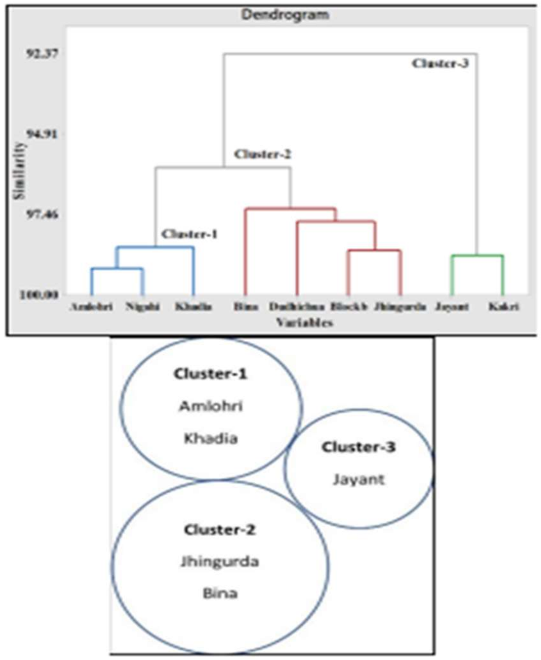

Figure 5 depicts the location-based classification of each monitoring station as a dendrogram plot. Ignaccolo et al. (2008) used functional clustering to analyze air quality monitoring networks. This is consistent with the study's purpose of identifying notable trends in station data. Furthermore, Lizuka's (2014) cluster analysis gives useful contextual information about air monitoring data from Japan's Kanto Region, as well as an example of how station clustering affects actual data.

Table 3 shows the average values for numerous qualities across all three groups.

Characteristics (unit)	Cluster 1	Cluster 2	Cluster 3
Production (Mt/year)	16	11	15
Over Burden (Mm^3/year)	66	76	44
Lease Hold Area (km^2)	23	17	18
Green Cover (km^2)	8	7	7
Mining Operation (km^2)	11	8	8
Haul Road (OB) km	9	12	12
Haul road (Coal) km	10	8	10
Transportation of OB and Coal (tons/day)	12157	15419	11983

Annual volume in thousands of tons, one million cubic meters, and square kilometers (Km2)

Principal Component Analysis

Determining the eigenvalues is the major goal of factor analysis. All of these eigenvalues are coupled with eigenvectors, which are a collection of largely interconnected air quality parameters. Principal Component Analysis was used to detect links and trends in the data, assisting in the identification of likely emission sources. Alonso (2019) examined the use of Principal Component Analysis (PCA) as a multivariate research tool. Alonso's study investigates the use of statistics in the assessment of air pollution, with a focus on the Madrid Region and the importance of spatial and multivariate analysis. Similarly, Yadav et al. (2022) use multivariate statistics to assess air quality in an industrially contaminated area. Their emphasis on determining the duration of exceptional air quality aligns with the key goals of this study. Figure 1 shows sieve plots. The PCA for this data set produces eigenvalues for all three groups (6). The component lists are shown in Table 4.

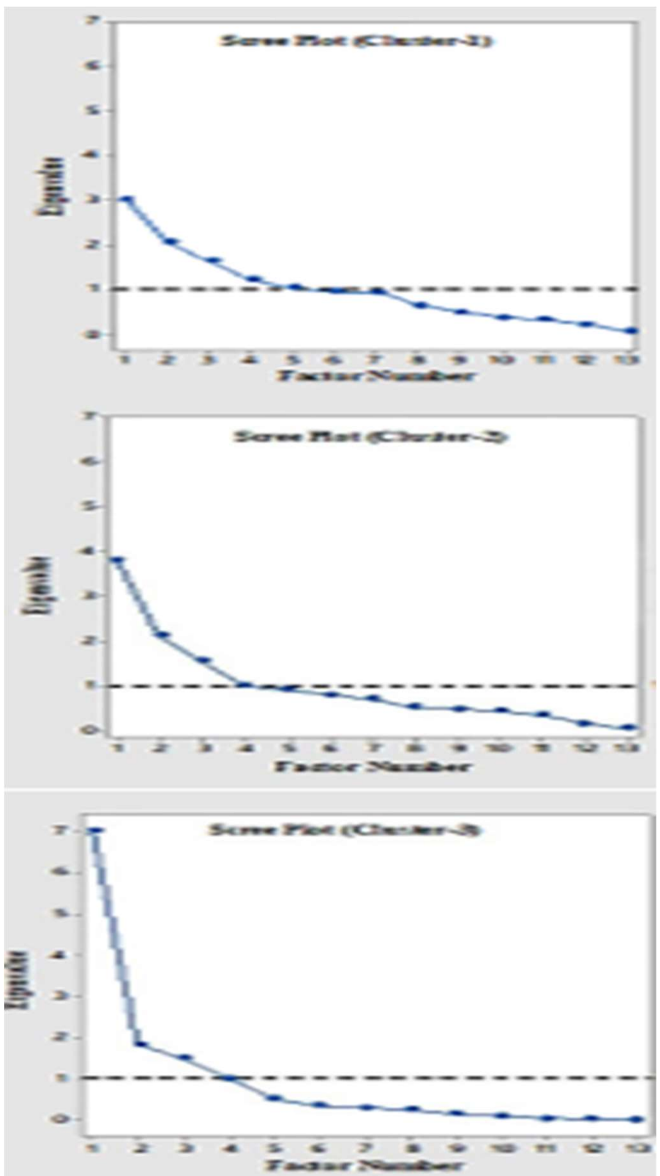

Figure 6 shows the skew plots for each of the three groups that were generated.

Cluster 1:

The first variable factor (VF1) explains 42.2% of the changes observed in Cluster 1. Figure 1 illustrates this. NOx (0.95), WS (0.919), temperature (0.915), CO (0.918), PM10 (0.861), PM2.5 (0.814), and SO2. Number 7 has strong positive loadings of these pollutants. Similarly, the component with the largest factor loading shows that PM10 and PM2.5 levels are increasing in these mines, mostly as a result of mining activities. This category is characterized as having a low level of contamination. Cluster 1 mines Amlohri, Khadia, and Nigahi produce roughly 66 million cubic meters (Mm3/year) of OB. This is larger than the second cluster, but smaller than the third. This is Table 3.

Gaseous pollutants emitted by the combustion of fossil fuels are the principal cause of pollution in areas with open coal mines. Coal and overburden are carried using rear-loading dumpers. Cluster 1 consists of 88 dumpers with a capacity of 190 tons, 63 dumpers with a capacity of 120 tons, and 17 dumpers with a capacity of 100 tonnes. Each of these helps to transfer OB. To transport coal from the coal mine to the coal stockyard, 93 dumpers with a capacity of 100 tons and 14 dumpers with a capacity of 85 tons each are used.

Table 4 shows the analysis of factors for various groups.

Varimax Rotation	Cluster-1			Cluster-2		Cluster-3	
Variable	VF1	VF2	VF3	VF1	VF2	VF1	VF2
HR	-0.044	0.014	-0.026	-0.092	0.213	0.109	-0.108
SR	-0.174	-0.231	0.867	-0.011	-0.031	0.043	0.877
TEMP	0.915	0.273	-0.016	0.041	0.152	-0.481	0.416
WD	-0.739	0.034	0.38	-0.067	0.159	-0.014	0.358
WS	0.919	0.311	-0.131	0.322	0.365	0.148	0.87
CC	0.618	-0.145	-0.309	-0.006	-0.095	-0.186	-0.006
NO_2	0.666	0.603	0.066	-0.235	0.358	0.732	0.192
NOX	0.188	0.95	-0.132	0.872	-0.073	0.927	-0.105
NO	0.521	0.757	-0.196	0.95	-0.187	0.859	0.127
PM_{10}	0.019	-0.025	0.861	0.822	-0.022	0.347	0.231
$PM_{2.5}$	0.814	0.26	0.241	0.181	-0.888	0.628	0.113
SO2	0.771	0.497	-0.221	0.071	-0.837	0.029	-0.34
CO	0.918	0.314	-0.137	0.119	-0.167	0.462	0.514
Variance	5.4902	2.5003	1.9357	2.5648	1.924	3.1642	2.3482
% Var	0.422	0.192	0.149	0.197	0.148	0.243	0.181

Each day, these dumpers transport around 12157 tons of OB and coal. Gocheva et al.'s research shows that factor analysis is a useful technique. This 2014 study looked into how factor analysis could be used to obtain more understanding into the dynamics of air pollution in a constrained urban environment. Their research proved that factor analysis, when combined with SARIMA (Seasonal Autoregressive Integrated Moving Average), can be used to detect latent patterns and components contributing to air pollution. Keresztes (2017) conducted an analogous analysis. It stressed factor analysis and gave a thorough evaluation of the Ciuc Basin's shifting air pollution levels. The goal of this study was to identify the factors driving differences in air quality.

The existence of a body of water in close vicinity causes weather fluctuations, which affect pollution levels. Gang et al. (2016) investigated the spatial and temporal aspects of the link between land use and air quality in Wuhan, China. The study examines these natural interactions over time and geography, as well as the critical link between land use and the evolution of air quality. The level of reduction in SO2 and PM10 pollution facilitated by aquatic settings was also explained. Similarly, the study area's Govind Ballabh Pant Sagar is only three kilometers distant from the aforementioned mines, which may cause periodic oscillations in the region.

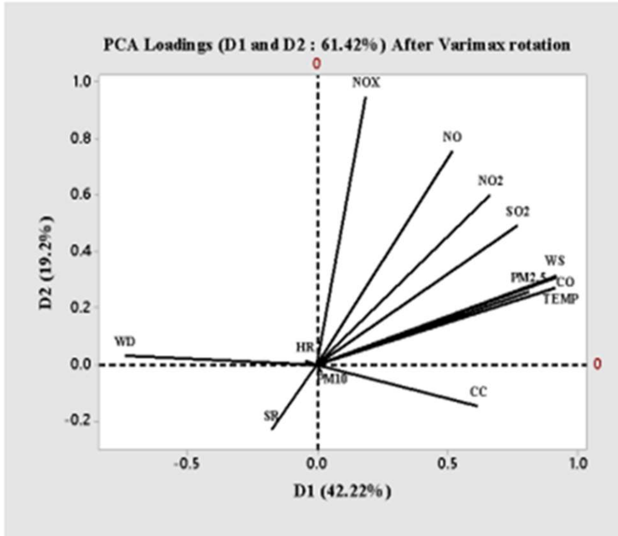

Figure 7: Cluster-1 PCA input.

Cluster 2:

Varimax Factor One has a significant positive loading for NOx (0.872), PM10 (0.822), and NO (0.95). The loadings account for 33.2% of the entire range. As seen in Figure 8, the Varimax Factor Two (VF2) suggests that SO2 and PM2.5 have the highest loading factors. These mines produce the most overflow (OB) among the three groups, at 76 million cubic meters (Mm3/year). The average tenure area is approximately 17 km2, which is significantly less than the other clusters. This considerably raises the overall pollution level. The largest source of PM10 and PM2.5 emissions is the transportation of coal and dirt via haul roads by automobiles (Aneza et al., 2012). Cluster 2 has the longest average trip routes for transferring OB. A maximum of 15,419 tons of OB per day can be moved from the OB Bench to the OB dump. This is accomplished by employing 224 dumpers with a capacity of 190 tons, 63 dumpers with a capacity of 100 tons, and 15 dumpers with a capacity of 85 tons. Coal is transported by three dump trucks with a rear capacity of 85 tons and 84 vehicles with a capacity of 100 tons. Within these mines, the movement of these massive vehicles is a major source of gaseous pollution.

Furthermore, because these mines are farther from the lake than the others, weather fluctuations have a smaller impact on overall pollution levels. Because these areas are at higher elevations, there are more contaminants present.

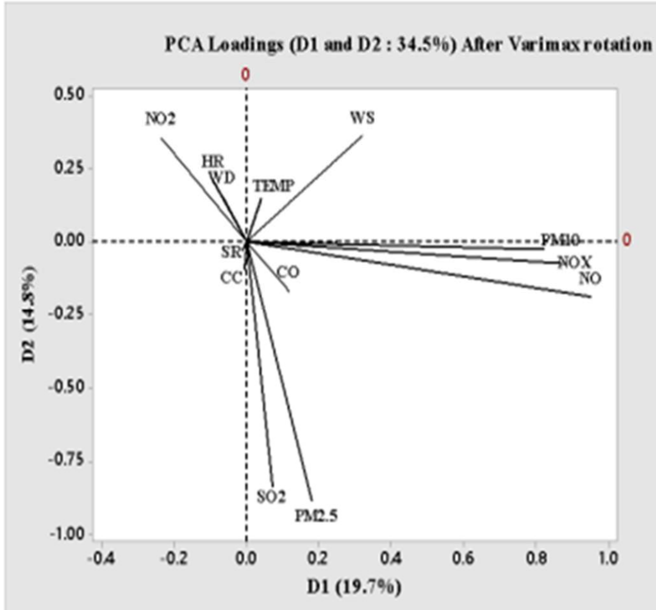

Figure 8 shows the PCA loading for Cluster 2.

Cluster 3:

At this location, the varimax factors VF1 and VF2 are related with significant positive loadings of SR (0.877), WS (0.87), NO (0.859), and NOx (0.927). The figures are shown in Figure 9. The combination of these factors accounts for 42% of the overall disparity.

This mine complex has the worst air quality of them all. These mines process the least amount of overburden (OB), 44 million cubic meters (Mm3) per year. They have the greatest leasing territory.

On average, 11,983 cargoes are transported daily using the haul road. Furthermore, its use for OB and coal transportation is uncommon. As a result, the number of vehicles and dumpers in this cluster is reduced. 85 dumpers with a capacity of 190 tons and 17 dumpers with a capacity of 85 tons are used for OB transportation. Coal hauled in each of the 58 lorries can hold more than 100 tons.

Particulate pollution is primarily delivered by wind, with Govind Ballabh Pant Sagar being a source when viewed from the southeast. This is especially important in these mines because they are close to the water, resulting in a higher concentration of WS.

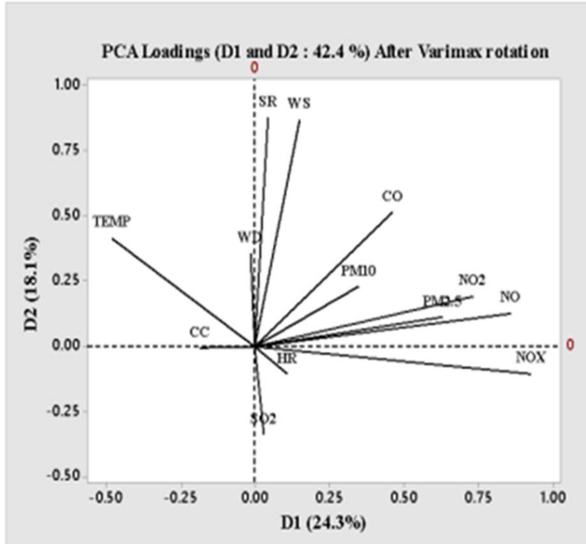

Figure 9 shows the PCA loading for Cluster 3.

Comparison of Multiple Linear Regression and Principal Component Regression for modeling air pollution.

Using MLR and PCR, a multiple linear equation model was created to quantify the proportion of the air quality index that each meteorological component and pollutant accounted for in each of three categories. It was unnecessary to study all thirteen parameters; only the principal components of the varimax rotation with factor loadings greater than 0.75 were investigated. Nazif et al. (2019) use a multivariate analysis with regression and blend models to predict particulate matter emissions and understand monsoon season changes. Similarly, Ausati et al. (2016) evaluate the predictive performance of several models, including PCR and MLR, for the PM2.5 level.

Equations relevant to the model and comparison:

Table 5.The MLR and PCR model formulas for each of the three categories are shown in Table

Model Equation		RMSE	R-sq	R-sq(adj)	Numbers of Parameters
Cluster-1					
MLR	AQI = 114.54 - 0.7546 HR + 0.0092 SR - 1.925 TEMP - 0.0903 WD - 6.25 WS + 0.3302 CC - 0.375 NO_2 + 1.063 NO_X - 0.704 NO + 0.9294 PM_{10} - 0.4194 $PM_{2.5}$ + 0.1898 SO_2 + 14.99 CO	28.31	98.63	98.61	13
PCR	AQI = 76.58 - 0.1138 SR - 2.373 TEMP + 8.999 WS + 0.8558 NO_X - 0.849 NO + 0.9379 PM_{10} - 0.4321 $PM_{2.5}$ + 0.491 SO_2	31.32	98.32	98.30	8
Cluster-2					
MLR	AQI = 73.9 - 2.62 NO_X + 3.36 NO - 2.35 PM_{10} + 1.299 $PM_{2.5}$ + 0.142 SO_2 + 0.0245 NO_2 - 0.278 NO_X + 0.682 NO - 0.505 PM_{10} + 1.2480 $PM_{2.5}$ + 0.1137 SO_2 + 5.9767 CO	43.94	97.22	97.19	12
PCR	AQI = 45.72 - 0.516 NO_x + 0.798 NO - 0.594 PM_{10} + 1.2698 $PM_{2.5}$ + 0.0499 SO_2 + 5.9463 CO	44.99	97.07	97.06	6
Clsuter-3					
MLR	AQI = 26.43 - 0.0185 HR + 0.0042 SR - 0.360 TEMP - 0.0227 WD + 6.74 WS + 0.0153 CC + 0.464 NO_2 - 0.531 NO_X + 0.653 NO + 0.5495 PM_{10} + 0.7766 $PM_{2.5}$ + 0.2454 SO_2 + 3.95 CO	18.92	88.21	87.98	13
PCR	AQI = 21.88 + 8.46 WS - 0.3632 NO_x + 0.533 NO + 0.5510 PM_{10} + 0.7771 $PM_{2.5}$ + 0.2420 SO_2 + 4.83 CO	18.98	88.06	87.90	7

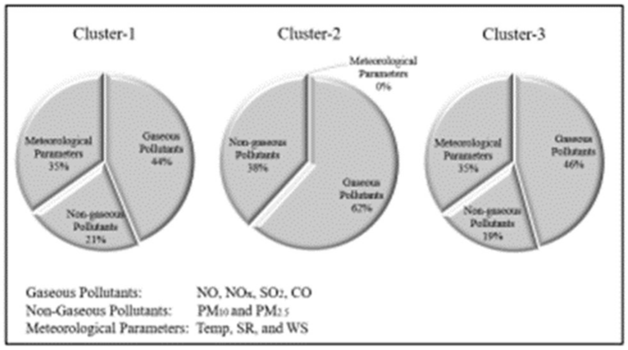

Figure 10 shows the proportion of meteorological, non-gaseous, and gaseous characteristics that are input.

The equations obtained by MLR and PCR for the three groups described in Table 5 appear to be the most effective, as evidenced by an R2 of 0.98. The R2 score for PCR with eight parameters for the same cluster is 0.98. Although the R2 value remains constant for PCR with six parameters, Cluster 2 has the second-highest correlation coefficient (0.97). The observed parameters have a substantial impact on the quantity of air pollution. Cluster 3 contains 13 parameters in the PCR and 7 in the MLR; yet, both have the lowest R2 value of 0.88. Positive findings were obtained using PCR with fewer variables and high RMSE and coefficient of correlation values. The findings show that the presence of multicollinearity makes certain parameters redundant. As a result, the AQI for each of these extraction locations may be calculated with the fewest amount of factors.

Percentage contribution of Gaseous and Non-gaseous Pollutants and meteorological conditions affecting the AQI of the mining complex.

Principal Component Regression was the method used to create models for each of the nine mines. The data show the key factor that influences the AQI score. 7, 8, and 9. Carbon monoxide, nitrogen oxides, and particle pollution have the biggest impact on the air quality index. Surface radiation, wind speed, and temperature are the next most important factors after groups 1 and 3. The AQI equation shows, as indicated in Table 5, that Cluster 2 is strongly damaged by all five contaminants, while meteorological factors have a small impact on these mines. The results for air pollution, particle pollution, and chemical pollution are shown as pie charts in Figure. 10. Cluster 1 is affected by 44% gaseous toxins, 21% nongaseous toxins, and 35% weather-related toxins. Cluster 2 gaseous and particle contaminants both have a 100% impact on the air pollution index; temperature has no effect on them. Meteorological conditions have a 35% impact on Cluster 3, as does secondary gas and non-gas pollution (65%).

The weather may have the greatest influence on the air quality index (AQI) of mines in groups 1 and 3 due to their proximity to bodies of water and altitude. This minefield is higher and located farther from the water.

Currently, the AQI is calculated using the number of pollutants present and the threshold at which their toxicity decreases. The AQI estimates done in this study take into account both the quantity of air pollutants and the weather, both of which have a substantial impact on surrounding air quality.

5. CONCLUSION

The usefulness of the chemometric technique in modeling atmospheric air pollution for a coal mining complex has been demonstrated in this study. Based on the degree of similarity and difference between the monitoring stations, the HACA result correctly divides the nine open-cast coal mines into three clusters. PCM loading through FA helps in finding the most influencing factors. According to MLR's and PCR's explicit equation model for AQI, multicollinearity and the repetition of factors in modeling can be eliminated. AQI can also be influenced by meteorological factors along with pollutants at a particular location. The movement of vehicles on haul roads inside the mining area is the major contributor to gaseous and particulate pollution. Wind Speed and Surface radiation play an important role in the overall pollution dispersion. Additionally, such studies are key in refining the AQI, enabling a more accurate determination of pollution levels in affected regions.

The future scope of this study is to apply these chemometric techniques to various mining regions for broader environmental impact assessments, integrating advanced predictive technologies for enhanced AQI forecasting, and informing policy development for more effective air pollution control in mining areas.

REFERENCES

1. Aertsen, W., Kint, V., Van Orshoven, J., Özkan, K. & Muys, B. 2010. Comparison and ranking of different modeling techniques for prediction of site index in Mediterranean mountain forests. Ecological Modelling, 221(8), 1119-1130. https://doi.org/10.1016/j.atmosenv.2016.08.007

2. Agathokleous, E., De Marco, A., Paoletti, E., Querol, X., & Sicard, P. (2022). Air Pollution and Climate Change Threats to Plant Ecosystems. Environmental research, 113420-113420. https://doi.org/10.1016/j.envres.2022.113420

3. Aneja, V. P., Isherwood, A., & Morgan, P. (2012). Characterization of particulate matter (PM10) related to surface coal mining operations in Appalachia. Atmospheric Environment, 54, 496-501. https://doi.org/10.1016/j.atmosenv.2012.02.063 Ausati, S., &

4. Amanollahi, J. (2016). Assessing the accuracy of ANFIS, EEMD-GRNN, PCR, and MLR models in predicting PM2. 5. Atmospheric Environment, 142, 465-474. https://doi.org/10.1007/s11869-019-00779- 5

5. Azid, A., Juahir, H., Ezani, E., Toriman, M.E., Endut, A., Rahman, M.N.A., Yunus, K., Kamarudin, M.K.A., Hasnam, C.N.C., Saudi, A.S.M. & Umar, R. 2015. Identification of the source of variation on the regional impact of air quality pattern using chemometrics. Aerosol and Air Quality Research, 15(4), 1545-1558. https://doi.org/10.4209/aaqr.2014.04.0073

6. Azid, A., Juahir, H., Toriman, M. E., Endut, A., Kamarudin, M. K. A., & Abd Rahman, M. N. (2015). Source apportionment of air pollution: a case study in Malaysia. Jurnal Teknologi, 72(1), 83-88. https://doi.org/10.11113/jt.v72.2934

7. Barjoee, S. S., Malverdi, E., Kouhkan, M., Alipourfard, I., Rouhani, A., Farokhi, H., & Khaledi, A. (2023). Health assessment of industrial ecosystems of Isfahan (Iran) using phytomonitoring: Chemometric, micromorphology, phytoremediation, air pollution tolerance and anticipated performance indices. Urban Climate, 48, 101394. https://doi.org/10.1016/j.uclim.2022.101394

8. Cowherd, C., Muleski, G. E., & Kinsey, J. S. (1988). Control of open fugitive dust sources. Final report (No. PB-89-103691/XAB). Midwest Research Inst., Kansas City, MO (USA).
9. Diana, A., Bertinetti, S., Abollino, O., Giacomino, A., Buoso, S., Favilli, L., ... & Malandrino, M. (2022). PM10 Element Distribution and EnvironmentalSanitary Risk Analysis in Two Italian Industrial Cities. Atmosphere, 14(1), 48. https://doi.org/10.3390/atmos14010048
10. Dominick, D., Juahir, H., Latif, M.T., Zain, S.M. & Aris, A.Z. 2012. Spatial assessment of air quality patterns in Malaysia using multivariate analysis. Atmospheric Environment, 60, 172-181. https://doi.org/10.1016/j.atmosenv.2012.06.021