DOI: https://doi.org/10.52756/ijerr.2023.v36.018a *Int. J. Exp. Res. Rev.*, Vol. 36: 433-466(2023)

Original Article | Peer Reviewed | Open Access

International Journal of Experimental Research and Review (IJERR)

ISSN: 2455-4855 (Online)

www.iaph.in

Chemometric Techniques in the Assessment of Ambient Air Quality and Development of Air Quality Index of Coal Mining Complex: A Statistical Approach

Navin Prasad, Tanushree Bhattacharya* and Bindhu Lal

Department of Civil and Environmental Engineering, Birla Institute of Technology, Mesra, Ranchi-835215, Jharkhand, India

E-mail/Orcid Id:

NP, phdcee10001.21@bitmesra.ac.in, https://orcid.org/0000-0003-2120-0258; *TB*, tbhattacharya@bitmesra.ac.in, https://orcid.org/0000-0003-3547-2529; *BL*, bindhu@bitmesra.ac.in, https://orcid.org/0000-0001-8200-8468

Article History:
Received: 18th Oct., 2023
Accepted: 12th Dec., 2023
Published: 30th Dec., 2023

Keywords:
Multivariate technique, Opencast Coal Mine, Cluster Analysis, Factor Analysis, Multiple Linear Regression, Air Pollution Index

How to cite this Article:
Navin Prasad, Tanushree Bhattacharya and Bindhu Lal (2023). Chemometric techniques in the assessment of ambient air quality, and development of air quality index of coal mining complex: A statistical approach. *International Journal of Experimental Research and Review, 36*, 433-446. **DOI:** https://doi.org/10.52756/ijerr.2023.v36.018a

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 $PM_{2.5}$, PM_{10}, SO_2, NO_2, 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 SO_2) and particulate matter (PM_{10} and $PM_{2.5}$). 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.

Introduction

Air pollution control and management in open-cast coal mines is challenging. Mining coal impacts the environment and influences the life of the people and ecosystem near the mining area (Agathokleous et al., 2022; Yang et al., 2022; Zipper and Skousen et al., 2021). Most of the coal mines are now equipped with continuous ambient air quality monitoring systems (CAAQMS) installed within 1.5 km of the mining site, which monitors air quality continuously and generates huge data sets. Massive, complicated data sets from atmospheric air quality monitoring stations must be combined with contemporary, reliable statistical approaches to simplify, minimize ambiguity, and display spatial variation. The Air Quality Index (AQI) is important in determining the ambient air quality for any location (Kumar, 2022; Wang et al., 2022; Wu et al., 2013). It is based on the conversion of the concentration of pollutants in non-dimensional numbers.

Many studies have used chemometric techniques to model the AQI to find the major contributors of air pollutants and their spatial variation (Barjoee et al., 2023; Diana et al., 2022; Galán-Madruga et al., 2023). Chemometrics is the science of linking measured values based on chemical measurements or principles with the parameter of interest by statistical or mathematical applications. The chemometric analysis is done mainly for industrial areas and urban cities (Azid et al., 2015; Nunes et al., 2019; Rani et al., 2017). In some reported studies, industrial chemometric analysis is used to

***Corresponding Author:** tbhattacharya@bitmesra.ac.in

understand the movement of pollutants (Grabowski et al., 2021; Vakarelska et al., 2021). Various research has been reported in quantifying the respirable silica emerging from coal mines (Stacey et al., 2022). Still, there needs to be more research on spatial analysis of pollutants emerging from the various activities of the mining operations using chemometric techniques. In this study, various concentrations of air pollutants and meteorological parameters were linked to AQI by statistical techniques.

AQI calculation is based on the concentration of pollutants and breakpoint concentrations. However ambient air quality is mostly influenced by meteorological parameters. Most mining industries have CAAQMS, which includes monitoring the concentration of air pollutants and meteorological parameters. So chemometric techniques can be used to develop models to calculate AQI and to analyze the air quality in detail.

This paper focuses on the combined effect of pollutants and meteorological parameters on the air quality of a location consisting of nine open-cast mines working simultaneously under a 330 km^2 area. This paper also aims to develop model equations to calculate AQI using cluster analysis (CA) and classification model techniques and Principal Component Method (PCM) under Factor Analysis (FA) (Dragović and Mihailović, 2009; Hooper and Peters, 1989; Huang et al., 2009; Wold et al., 1987). Moreover, this study aims to develop an equation using multiple linear regression (MLR) and principal component regression (PCR) for the calculation of AQI, including meteorological parameters that influence air quality to a great extent. It also signifies the contribution of various statistical techniques in chemometric analysis for understanding air pollution in coal mines.

Study Area

The research site chosen for this study is situated in Singrauli, a region in central India known for its major open coal mining area. This mining operation is overseen by Northern Coalfield Ltd. (NCL). The study area lies between the latitude of 24°14' 06.24" N to 24°05'02.63" N and the longitude of 82°30' 54.71" E to 82° 47'56.13" E. A population of 1.2 million people resides in the vicinity of the coalfield. The area comprises two main basins: the Moher Sub-Basin, with a total coal reserve of approximately 6.83 Billion Tons (BT), and the Singrauli Main Basin, containing around 3.23 BT of coal (Javed et al., 2021). Situated in Singrauli Madhya Pradesh, the coal mine area experiences an average annual rainfall of 1119.65 mm, while the temperature varies between extremes, ranging from 47.2 degrees Celsius to 4 degrees Celsius. Notably, some portions of the mining area extend into the Sonbhadra district of Uttar Pradesh, as illustrated in Fig. 1. The study includes nine NCL mines, their respective Latitude and Longitude coordinates are listed in Table 1. In Fig. 1, these mines are represented by an asterisk sign. In the mining area's periphery, four major power plants play a significant role in the energy supply to the state.

Table 1. Locations of CAAQMS Station in Singrauli coal mining 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

Methods

Data Collection

The data have been collected from nine continuous monitoring networks from the central control room ambient air quality and Management Stations (CAAQMS) installed at every mine of the Singrauli coalfield complex, as shown in Table 1. The data include the gaseous and non-gaseous pollutants and meteorological data taken annually from January 1st,

DOI: https://doi.org/10.52756/ijerr.2023.v36.018

2020, to December 31, 2020. The daily 24-hour average data was calculated from the CAAQMS data taken every 15 minutes for each day from 00:00 hrs. to 24:00 hrs. for the year 2020.

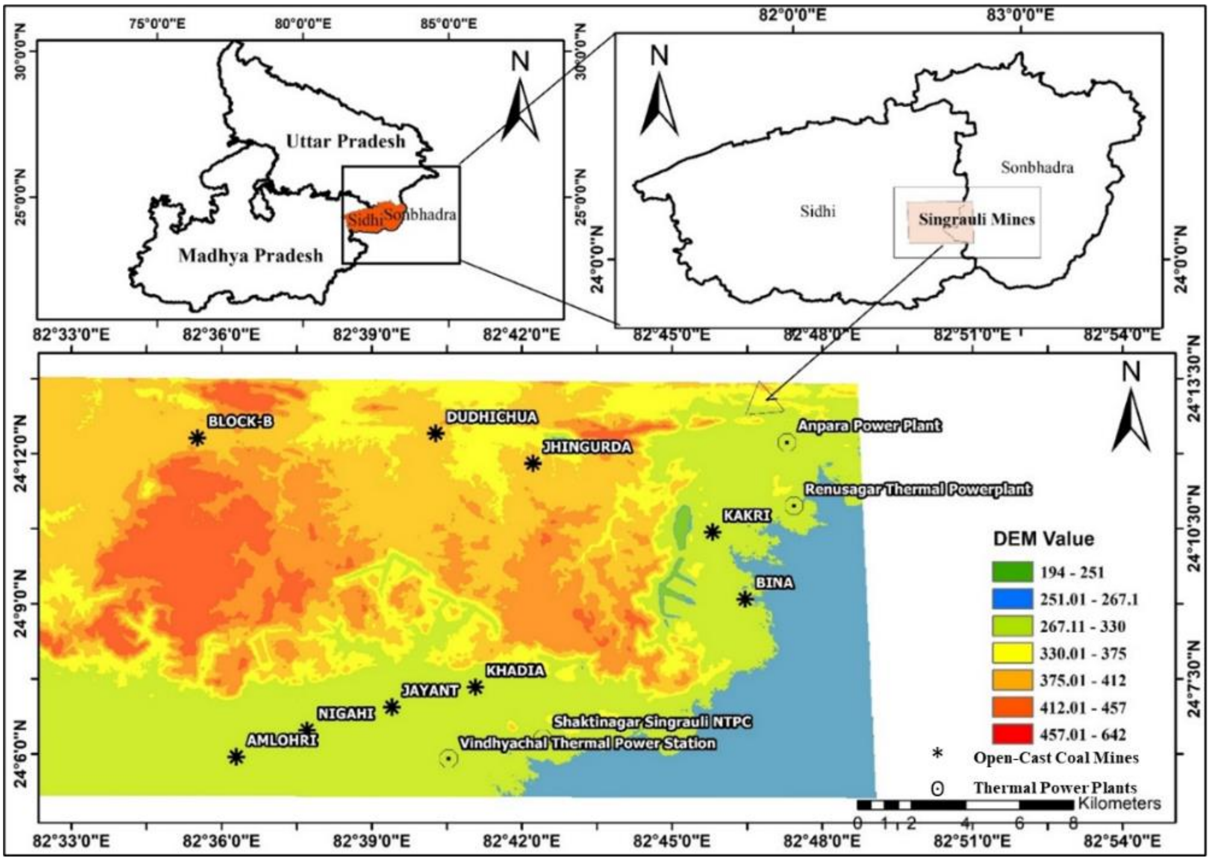

Figure 1. Coal Mining Complex.

Table 2. Availability of Meteorological Parameters.

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 (µg/m^3)	✓	✓	×	✓	✓	✓	✓	×	✓
8	NO_2 (µg/m^3)	✓	✓	✓	✓	✓	✓	✓	✓	✓
9	NO (µg/m^3)	✓	✓	✓	✓	✓	✓	✓	✓	✓
10	SO_2 (µg/m^3)	✓	✓	✓	✓	✓	✓	✓	✓	✓
11	PM_{10} (µg/m^3)	✓	✓	✓	✓	✓	✓	✓	✓	✓
12	$PM_{2.5}$ (µg/m^3)	✓	✓	✓	✓	✓	✓	✓	✓	✓

DOI: https://doi.org/10.52756/ijerr.2023.v36.018

The distance between the stations and the core mining sites was approximately 1.5 km. The data included $PM_{2.5}$, PM_{10}, SO_2, NO_2, NO, NOx, and CO pollutants. The $PM_{2.5}$ & and PM_{10} were estimated based on the beta ray attenuation technique using a beta gauge by absorbing energy passing through the filter tape where the particulate matter is collected. SO_2 measurement was based on the Pulse Fluorescence Analyzer. Chemiluminescence was the measurement technique for NO_2, NO, NOx, and non-dispersive infrared for CO. The meteorological parameters include temperature, humidity, precipitation, wind direction, wind velocity, and solar radiation, which were detected by meteorological instruments.

Where DBT (Average Dry Bulb Temperature), RH (Relative Humidity), WS (Wind Direction), HR (Horizontal Solar Radiation), CC (Cloud Cover), CO (Carbon Monoxide), NO_2 (Nitrogen Dioxide), NO (Nitric oxide), SO_2 (Sulphur Di-oxide), PM (Particulate Matter with diameters 10 and 2.5 µm). * Only six-month data is available for this coal mine.

Organization of Data

Daily average data of all mines was used as input data for cluster Analysis and Factor Analysis. The data obtained have an overall 4.74% of missing values. The missing values are imputed with the help of a hybrid method of multivariate imputation with interpolation (Junninen et al., 2004) using IBM SPSS 26.0.0.0 64-bit edition software.

Chemometric Analysis

Hierarchical Agglomerative Cluster Analysis (HACA)

Cluster analysis is an unsupervised method to handle a large amount of data and reduce it into smaller groups of factors known as clusters based on data similarities or differences. (Isiyaka et al., 2015; Ramson et al., 2016; Too et al., 2011).

The HACA is applied to the daily average data of thirteen parameters, including six meteorological parameters and the concentration of seven pollutants obtained from all nine mines. A dendrogram plot shows the degree of homogeneity through Ward's methodology and Euclidean distance measurements (Lu et al., 2012). This method has been performed with the help of IBM SPSS 26.0.0.0 64-bit edition software.

The Euclidean distance (D_{ij}) is defined by equation (1):

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

Where $x_1, x_2, \ldots\ldots\ldots x_m$ is the number of observations, i and j are the two observed data, and the distance has been calculated. Whereas in Ward's methodology, Analysis of Variance (ANOVA) is used to analyze distance and ensure that the sum of squares between two clusters is minimal (Azid et al., 2015).

Factor Analysis

Factor analysis is performed to find the relation among variables and to reduce the number of factors influencing the overall result of the variables (Mutalib et al.,2013). It is a descriptive method similar to Principal Component Analysis (PCA). In PCA new variables are created based on a linear combination of the observed variable, whereas FA factors are identified, which are linear functions of observed variables.

FA is defined by the equation (2):

$$F_{ij} = \sum_{j=1}^{m} C_{fj} f_{ji} + E_{fi} \quad (2)$$

Where F is the measured values of the variable, C is the factor loading, f is the factor value, E is the error or variation, i is the number of samples, j is the number of variables and m is the total number of factors.

The Principal Component Method (PCM) is a widely used Factor Analysis (FA) technique. PCM aims to capture the essential patterns in the data by first identifying the factor with the highest variability and then extracting the maximum variability for each subsequent factor. Varimax rotation is employed to enhance the understanding of the Principal Components (PCs) generated by PCM to improve interpretability. The eigenvalues resulting from the varimax rotation are a preliminary step for Factor Analysis. Factors with eigenvalues exceeding 1 are deemed significant and termed Varimax Factors (VFs). VFs with loading values surpassing 0.75 are considered to exhibit strong factor loadings. In this research, we select factors with factor loadings greater than 0.75 to serve as principal components. (Azid et al., 2015, Juahir et al., 2011). This study applies FA (PCM) to 13 variables independently using IBM SPSS 26.0.0.0 64-bit edition software.

Air Quality Index (AQI)

The Air Quality Index (AQI) is a helpful tool for finding air quality information simply. It takes the pollution data about different pollutants in the air and converts it into a single number, along with labels and colors, to make it easy to understand.

Calculation of AQI for some pollutant p given by National Ambient Air Quality standards (NAAQS) defined by equation (3):

$$I_p = \frac{I_{HI}-I_{LO}}{BP_{HI}-BO_{LO}}(C_P - BP_{LO}) + I_{LO} \quad (3)$$

Where I_p is the pollutant index, C_p is the rounded concentration of pollutant p, and BP_{HI} and BP_{LO} are the breakpoint concentrations that are greater than and less than C_p, respectively. I_{HI} and I_{lo} are the AQI values corresponding to BP_{HI} and BP_{LO}.

This index is computed by looking at the average concentrations of specific pollutants over 24 hours at a monitoring site. To calculate the AQI, data for at least three pollutants must be available, with one of them being either $PM_{2.5}$ or PM_{10}, as per guidelines from the Central Pollution Control Board (CPCB). This ensures that the index accurately reflects the air quality conditions. For this study, the concentration of 5 major pollutants that is SOx, NOx, CO, $PM_{2.5}$, and PM_{10} were considered for the AQI calculation.

Multiple Linear Regressions (MLR)

In atmospheric modeling, MLR is frequently utilized (Azid et al., 2015; Dominick et al., 2012; Aertsen et al., 2010) and is a suitable tool for studying the interaction between independent and dependent variables by forming a linear equation using observed data. The MLR method was used in this study to support the relationship between the AQI data and the meteorological parameters along with pollutant concentration, which was the most important among others. MLR is given by equation (4):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + \varepsilon_i \quad (4)$$

Where i = 1…..n, β_0, β_1, and β_k are regression coefficients, X_1 and X_k are independent variables and ε is the regression error.

The contribution of each parameter in AQI is determined with the help of the coefficient of determination (R^2), the adjusted value of the coefficient of determination (adjusted R^2), and the Root mean square error (RMSE). The variables from varimax rotation have also been taken as independent variables for AQI calculation.

Results and Discussion

Air Quality Status of the Study Area

The annual average data of the pollutants were compared with the National Ambient Air Quality standard. Figure 2 displays the average $PM_{2.5}$, PM_{10}, SO_2, and NO_2 concentrations over 2020–2021. The variation in concentration of CO (8 hrs.) is shown in Figure 3. For $PM_{2.5}$, four of the nine mines are beyond the permissible limit. The maximum concentration was found in Bina and Block-B. In comparison, PM_{10} concentrations exceed the ambient air quality standard. The concentration of the other two pollutants, NO_2 and SO_2, was below the permissible limit for all nine mines. The mean concentration of CO was below the desired level except for Block-B and Jayant. The Vindhyachal Thermal Power Plant has latitude and longitude of 24° 06' 56.00" N and 82° 39' 24.00" E respectively 6 km southeast of Amlohri, and the prevailing wind is from the ESE (East-South-East) direction. Pollutants from thermal power plants may be dragged in this direction. In contrast, Jayant has less pollution than the nearby mine because the prevailing wind direction is NW (North-West). It was observed that the East-South-East (ESE) was the predominant wind direction in Amlohri, Bina, Block-B, and Jayant, while the West-South-West (WSW) was more prevalent in Dudhichua and Khadia. While the wind patterns in Jhingurda and Kakri are the same, Nigahi has experienced WNW winds for longer periods of the year.

However, the trucks, bulldozers, pay loaders, cranes, and heavy earth-moving machinery required for the transportation and mining of coal run on diesel, which is the main producer of gaseous pollutants which include SO_2, NO, NO_2, and CO. Cowherd et al. (2013) established a correlation between emissions from heavy machinery and the deterioration of air quality, particularly in urban areas.

Similarly, Ghose and Majee (2000) identified that diesel combustion engines are prominent sources of particulate emissions, releasing pollutants that adversely affect human health and the environment. Incomplete fuel combustion of the vehicles moving coal and overburden inside the mine, such as trucks, bulldozers, payloaders, cranes, and heavy earth-moving machinery, are the leading causes of CO pollution. Nie et al. (2022) confirm that vehicles moving coal and overburden in mines, like trucks, bulldozers, and more, are a major source of carbon monoxide pollution due to incomplete fuel burning. Machinery operations directly contribute to releasing harmful pollutants like CO. Jayant is one of the most significant open-cast mines with an annual production of 25 million tons of coal. This substantial output involves many vehicles dedicated to hauling coal and overburden, which may contribute to CO emissions within the mine's vicinity. Block-B is positioned in the northwest corner and proximate to a populated region.

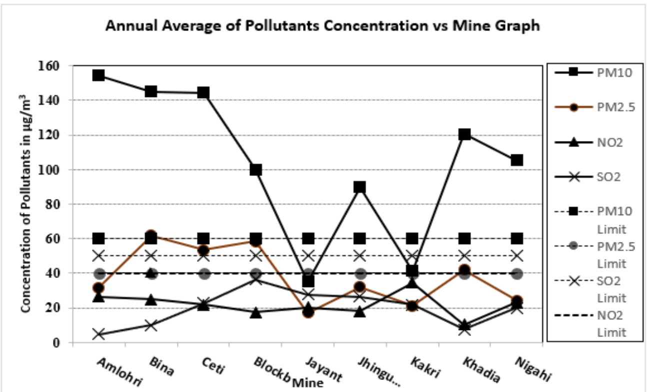

Figure 2. Yearly Mean Pollution concentration of all the nine mines in the Singrauli Coal Complex.

DOI: https://doi.org/10.52756/ijerr.2023.v36.018

The movement of vehicle exhaust in the nearby market area is drawn towards the Continuous Ambient Air Quality Monitoring Station (CAAQMS) due to prevailing east-southeasterly winds (ESE winds), which might be the reason for excess CO emission for this mine.

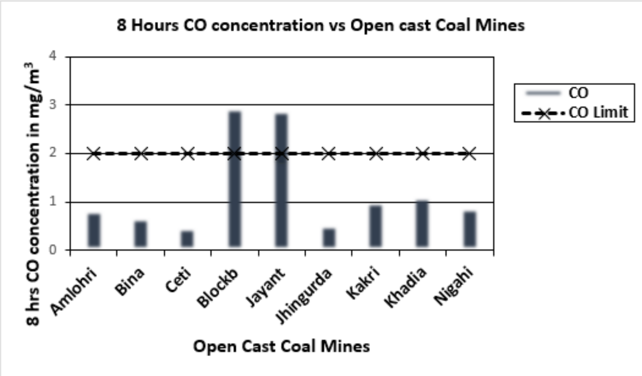

Figure 3. The 8 hrs. Concentration of CO for all the nine mines.

The AQI values for the mining complex are shown in Fig. 4. The average AQI for the first three coal mines (Amlohri, Nigahi, and Khadia) in Cluster-1 is 153. These mines produce about 16 million tons of coal per year and cover the largest area for mining. In Cluster-2, in the northern part of the study area and including mines like Block-B, Dudhichua, Jhingurda, and Bina, the air quality index is the highest among all clusters at 224. Cluster-3 has the lowest air quality index among the three clusters, with a value of 120. This cluster includes the Jayant and Kakri mines, which produce the least amount of Over Burden (OB), about 44 million cubic meters per year. The variation of AQI is maximum in Dudhichua and least in Jayant.

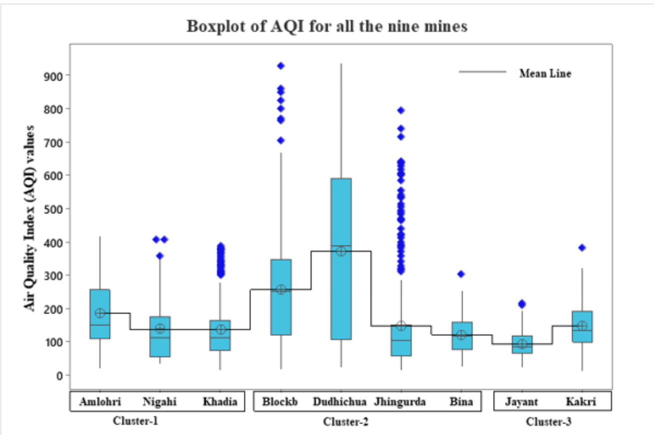

Figure 4. Boxplot of AQI values for the year 2019-2020.

Spatial Classification of Mines Based on Air Quality Parameters

The similarities and variations in the air quality characteristics and meteorological parameters were identified using HACA. Those that exhibited a high degree of spatial similarity were put together in one cluster. Three clusters were created because of this process, as shown in Figure 5.

Cluster 1 is classified as a moderately Polluted site because the average value of AQI is 153 during the entire year. These areas comprise Amlohri, Khadia, and Nigahi located in the southwest corner of the mining complex, and share the same elevation level of approximately 194-251m approximately. Cluster 2 is classified as a severely polluted site as the yearly average value of AQI is 224, which includes Jhingurda, Bina, Dudhichua, and Block-B which lies in the northern part of the mining complex and lies at a higher altitude concerning other mines. Cluster 3 is the least polluted site among the other two as the value of AQI is 120 during the entire year. This cluster includes the most significant mines, Jayant and Kakri. The mean values of different attributes of all three clusters and their associated open-cast coal mines are defined in Table 3. Similar results were reported by Wang et al. (2018) by identification of redundant stations in air quality networks. Gouveia et al. (2015) employ wavelet-based clustering techniques, offering a potential approach for efficient spatial grouping of stations, which resonates with the current research's exploratory methodologies.

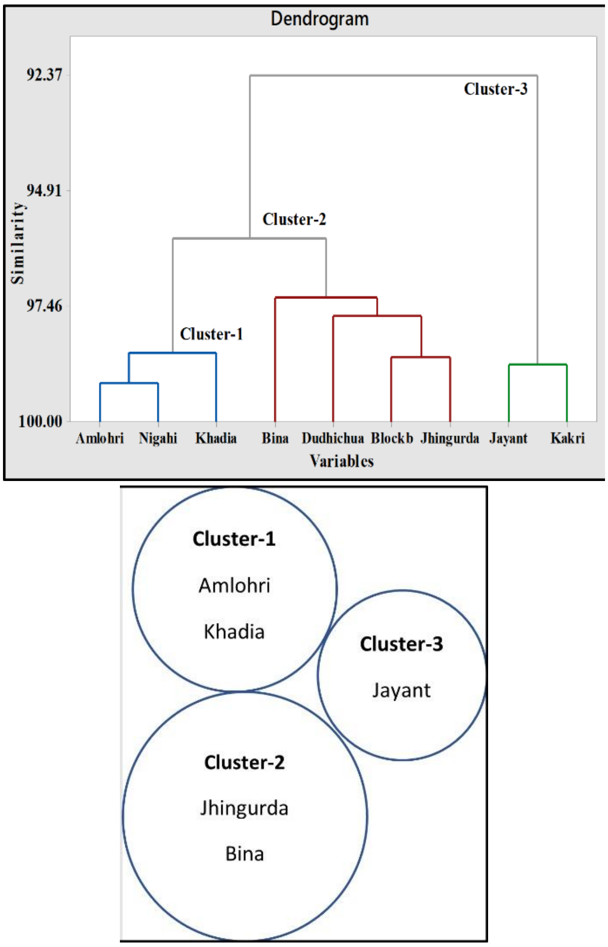

Figure 5. Dendrogram Plot showing spatial classification for all the monitoring stations.

DOI: https://doi.org/10.52756/ijerr.2023.v36.018

Ignaccolo et al. (2008) analyzed air quality monitoring networks using functional clustering, paralleling the current investigation's pursuit of extracting meaningful patterns from station data. Furthermore, Lizuka's (2014) cluster analysis of air monitoring data from the Kanto Region of Japan provides a practical contextualization and shows the implications of station clustering on real-world data.

Table 3. Mean Values of different attributes for all the three clusters.

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

Mt/Year (Million Tons per year), Mm^3/Year (Million Cubic Meters per Year), Km^2 (Square Kilometer)

Principal Component Analysis

The factor analysis is mainly used to find the eigenvalues. Each of these eigenvalues is related to eigenvectors, which correspond to a group of air quality parameters that are mostly correlated. The Principal Component method was used to visualize patterns and correlations between the data and subsequently identify potential emission sources. The implementation of Principal Component Analysis (PCA) as a multivariate analysis technique was reported by Alonso (2019). In his study, Alonso examines statistical tools for air pollution assessment, emphasizing the role of multivariate and spatial analysis in the Madrid Region.

Similarly, Yadav et al. (2022) employ multivariate statistical methods to investigate air quality in an industrially polluted city. Their focus on assessing air quality's sustainability aligns with the broader objectives of the present study. The PCA for the current data gives eigenvalues for all three clusters as shown in scree plots in Fig. 6. The factors are shown in Table 4.

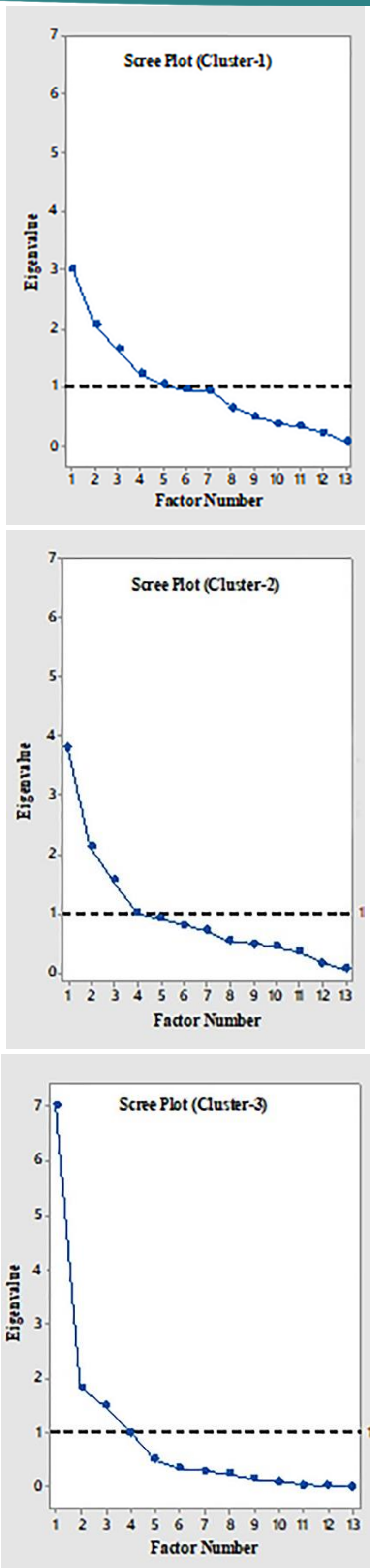

Figure 6. Scree Plot for all the three Clusters formed.

Cluster 1:

The first variable factor VF1, clarifies 42.2% of the differences in cluster 1. It shows strong positive loadings with WS (0.919), Temp (0.915), CO (0.918), PM_{10} (0.861), $PM_{2.5}$ (0.814), SO_2 (0.771), and NOx (0.95) as depicted in Fig. 7. Similarly, the factor having maximum factor loading signifies the accumulation of PM_{10} and $PM_{2.5}$ in these mines which is largely due to mining activities. This cluster falls into the category of moderate pollution. In Cluster 1, Amlohri, Khadia, and Nigahi mines handle more OB than Cluster 2 but less than Cluster 3, which is around 66 million cubic meters per year (Mm3/year) (Table 3).

Moreover, burning fossil fuels generates gaseous pollutants and serves as the primary pollution source in open coal mining areas. Coal and overburden are moved using rear-loading dumpers. In cluster 1, 88 dumpers with a capacity of 190 tons, 63 dumpers with a capacity of 120 tons, and 17 dumpers with a capacity of 100 tons are used for OB transportation. For coal transport from the quarry to the coal stockyard, 93 dumpers with a capacity of 100 tons and 14 dumpers with a capacity of 85 tons. On average, these dumpers handle a combined total of 12157 tons of OB and Coal every day. The application of factor analysis is substantiated by the study conducted by Gocheva et al. (2014) this study explored the utilization of factor analysis to enhance the understanding of air pollution dynamics in a small urban area. Their investigation employed a combination of SARIMA (Seasonal Autoregressive Integrated Moving Average) and factor analysis techniques, showcasing the potential of factor analysis in identifying underlying patterns and contributors to air pollution. A similar study was done by Keresztes (2017) for an in-depth exploration of air pollution dynamics, centered around factor analysis, conducted within the context of the Ciuc Basin region. The findings of this study were to understand the complexities of air quality patterns and their underlying contributors.

Table 4. Factor Analysis of Different Clusters.

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

The presence of a nearby water body influences the fluctuations in meteorological conditions and their impact on overall pollution. The research conducted by Gang et al. (2016) examines the impacts of land use on air quality from a spatiotemporal perspective in Wuhan, China; the study investigates the important relationship between land use patterns and air quality dynamics, offering a comprehensive spatiotemporal analysis of these environmental interactions. It also highlighted that water bodies exhibit a notable mitigating influence on SO_2 and PM_{10} pollution. Similarly, in the study area, Govind Ballabh Pant Sagar is situated within a 3 km radius of these mines, which can influence the local meteorological parameters.

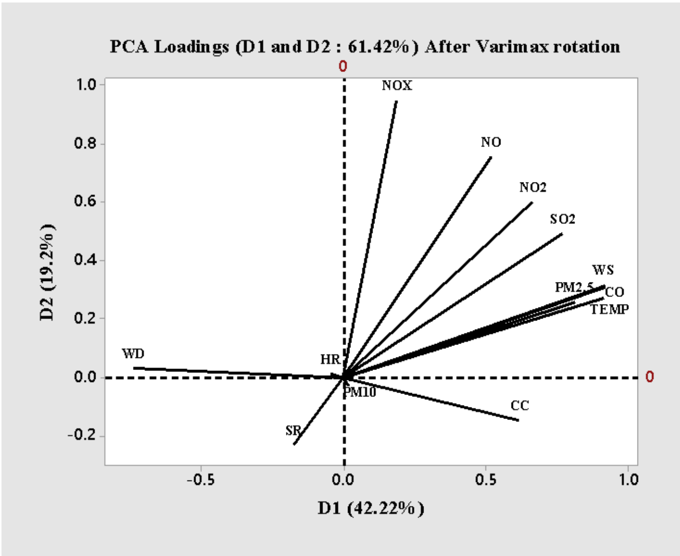

Figure 7. PCA loading for Cluster-1

Cluster 2:

In Varimax Factor One (VF1), significant positive loadings are observed for NOx (0.872), PM_{10} (0.822), and NO (0.95), accounting for a considerable 33.2% of the total variation. Meanwhile, Varimax Factor Two (VF2) reveals that $PM_{2.5}$ and SO_2 have the highest loading factors, as shown in Figure 8. These mines handle the greatest amount of overburden (OB) at 76 million cubic meters per year (Mm3/year), which is the highest among all three clusters. The average leasehold area is comparatively smaller, approximately 17 km^2, compared to the other clusters, and this factor contributes significantly to the overall pollution levels. The vehicles transporting soil and coal via haul roads are the primary sources of PM_{10} and $PM_{2.5}$ emissions (Aneja et al., 2012). Within Cluster 2, the mines have the longest average haul roads for OB transportation, with a maximum capacity of 15,419 tons per day from the OB Bench to the OB dump. This is achieved using 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 transportation involves 84 trucks with a capacity of 100 tons and three dumpers with a rear capacity of 85 tons. The movement of these heavy vehicles is a significant contributor to gaseous pollution in these mines.

Furthermore, these mines are situated farther from the lake than others, resulting in a lesser impact of meteorological conditions on overall pollution levels. This causes pollutants to be more concentrated in these areas due to their elevated geographical location.

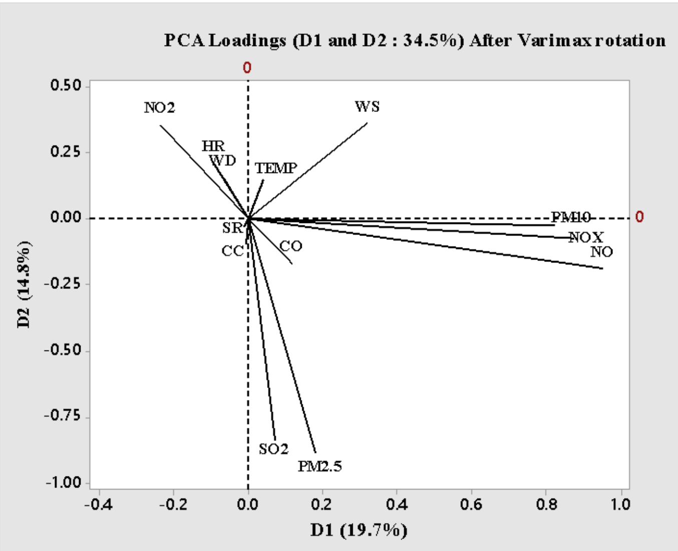

Figure 8. PCA loading for Cluster-2.

Cluster 3:

Fig. 9 shows strong positive loadings for SR (0.877), WS (0.87), NO (0.859), and NOx (0.927), which are associated with two varimax factors, VF1 and VF2, at this specific location. These factors together explain 42% of the overall variance.

This cluster of mines has the lowest Air Quality Index (AQI) compared to the other clusters. The mines in this group handle the smallest amount of overburden (OB), specifically 44 million cubic meters per year (Mm^3/year), and possess the largest lease area.

The transportation of OB and coal via the haul road is also minimal, with only 11,983 tons per day. Consequently, fewer trucks and dumpers are involved in this cluster. For OB transport, 85 dumpers with capacities of 190 tons and 17 with capacities of 85 tons are used. For coal transportation, 58 trucks with a capacity of 100 tons each are employed.

The movement of wind, particularly from the southeast direction originating from Govind Ballabh Pant Sagar, plays a significant role in dispersing particle pollution. This is particularly relevant to these mines as they are situated closest to the water body, so the loading of WS is higher.

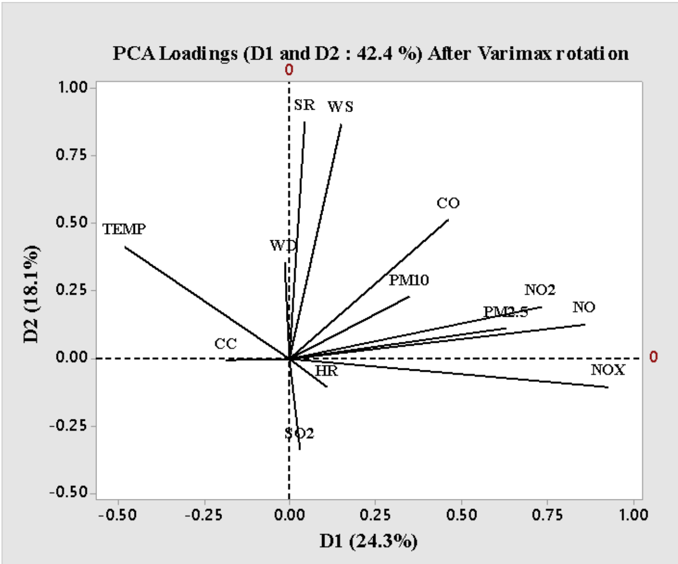

Figure 9. PCA loading for Cluster-3.

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

To establish the percentage contribution of each pollutant and meteorological parameter in calculating the air quality index for each of the three clusters, a Multiple linear equation model was developed using MLR and PCR. Instead of using all 13 parameters, only the principal components from the varimax rotation whose factor loading is more than 0.75 have been considered. Nazif et al. (2019) focus on multivariate analysis to understand monsoon seasonal variations and predict particulate matter emission using regression and hybrid models. Similarly, Ausati et al. (2016) assess the

DOI: https://doi.org/10.52756/ijerr.2023.v36.018

predictive accuracy of various models, including PCR and MLR, for $PM_{2.5}$ levels.

Proposed equations for the model and comparison:

Table 5. Model Equation for all three clusters using MLR and PCR.

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

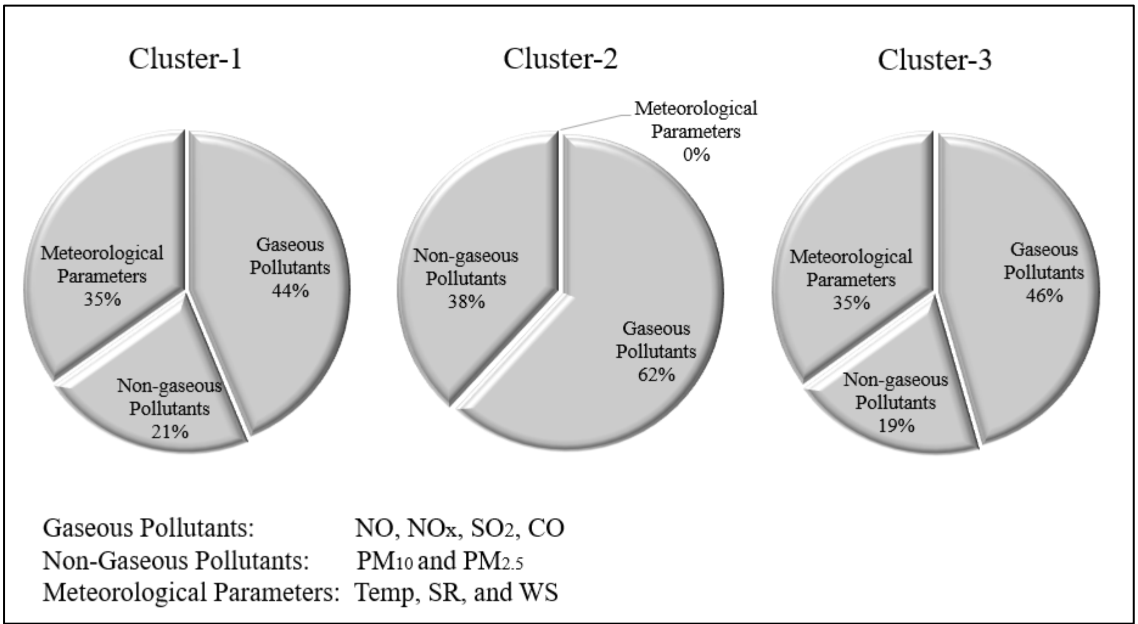

Figure 10. Percentage Contribution of gaseous, non-gaseous, and meteorological parameters.

DOI: https://doi.org/10.52756/ijerr.2023.v36.018

The equations created by MLR and PCR for the three clusters in mentioned Tabler 5 appear to have the highest R2 value of 0.98. PCR for the same cluster employing 8 parameters has an R2 value of 0.98. Cluster 2 has the second-highest correlation coefficient value of 0.97, whereas PCR with 6 parameters offers the same R2 value, these monitored parameters have a significant impact on the level of the air pollution index. The PCR and MLR for cluster 3 with 13 and 7 parameters respectively have the lowest R2 value, which is 0.88. PCR with a lesser number of variables performed well with significant values of coefficient of Correlation and RMSE. The outcomes signify that some of the variables are redundant due to multicollinearity. Hence AQI can be calculated with the least parameters effectively for all these mining sites.

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

Models for the nine mines were created using the Principal Component Regression technique. The most important factor influencing the value of AQI is shown in Fig. 7, 8, and 9. The AQI value is most significantly influenced by particulate pollutants, nitrogen oxides, and carbon monoxide, followed by temperature, wind speed, and surface radiation for clusters 1 and 3. All five pollutants have a dominant impact on Cluster 2, but meteorological parameters do not affect these mines as shown by the AQI equation in table 5. Fig. 10 depicts the results as pie charts for gaseous pollutants, particle pollutants, and weather conditions. 44% of gaseous pollutants, 21% of non-gaseous pollutants, and 35% of meteorological conditions impact Cluster 1, respectively. In cluster 2, gaseous and non-gaseous pollutants have a 100% influence on the air pollution index, with no contribution from the weather. 65% of secondary gas and non-gas pollutants and 35% of climatic conditions impact Cluster 3.

The influence of weather conditions on the air quality index (AQI) of mines in clusters 1 and 3 may be mostly explained by the distance from water bodies and the elevation of the mining regions. The mines in Cluster 2 are more distant from the water body and at higher elevations.

Currently, AQI is calculated based on the concentration of the pollutants and their breakpoints. The AQI equations developed in this study involve a concentration of air pollutants and meteorological parameters, which influence ambient air quality to a large extent.

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.

Acknowledgments

"The Authors wish to thank Coal India Limited for funding this project (project code: CIL/R&D/05/02/2021) and special thanks to Northern Coalfield Limited and Central Mine Planning & Design Institute for their technical and logistic help."

Conflict of interest

The authors declare that there is no conflict of interest.

Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available as the data is collected from the concerned industry with their permission and the dataset is part of the Ph. D thesis but are available from the corresponding author on reasonable request.

References

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

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

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., & 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

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

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

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

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).

Diana, A., Bertinetti, S., Abollino, O., Giacomino, A., Buoso, S., Favilli, L., ... & Malandrino, M. (2022). PM10 Element Distribution and Environmental-Sanitary Risk Analysis in Two Italian Industrial Cities. *Atmosphere*, *14*(1), 48. https://doi.org/10.3390/atmos14010048

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

Dragović S, Mihailović N (2009) Analysis of mosses and topsoils for detecting sources of heavy metal pollution: multivariate and enrichment factor analysis. *Environmental Monitoring and Assessment* *157*(1), 383-390. https://doi.org/10.1007/s10661-008-0543-8

Galan-Madruga, D., Cardenas-Escudero, J., Broomandi, P., Oleniacz, R., & Cáceres, J. O. (2023). Performance assessment of air quality monitoring networks. A specific case study and methodological approach. Air Quality, Atmosphere & Health, 16(1), 113-126. https://doi.org/10.1007/s11869-022-01254-4

Gang, X., Limin, J., Suli, Z., Man, Y., Xiaoming, L., Yuyao, H., ... & Ting, D. (2016). Examining the Impacts of Land Use on Air Quality from a Spatio-Temporal Perspective in Wuhan. China. *Journal of Atmosphere*, *7*(62), 2-18. https://doi.org/10.3390/atmos7050062

Ghose, M. K., & Majee, S. R. (2000). Assessment of dust generation due to opencast coal mining–an Indian case study. *Environmental Monitoring and Assessment*, *61*, 257-265. https://doi.org/10.1023/A:1006127407401

Gocheva-Ilieva, S. G., Ivanov, A. V., Voynikova, D. S., & Boyadzhiev, D. T. (2014). Time series analysis and forecasting for air pollution in a small urban area: a SARIMA and factor analysis approach. *Stochastic environmental Research and Risk Assessment*, *28*, 1045-1060. https://doi.org/10.1007/s00477-013-0800-4

Gouveia, N., Kephart, J. L., Dronova, I., McClure, L., Granados, J. T., Betancourt, R. M., ... & Diez-Roux, A. V. (2021). Ambient fine particulate matter in Latin American cities: Levels, population exposure, and associated urban factors. *Science of the Total Environment*, *772*, 145035. https://doi.org/10.1186/s43088-022-00305-0

Grabowski, J., & Smoliński, A. (2021). The application of hierarchical clustering to analyzing ashes from the combustion of wood pellets mixed with waste materials. *Environmental Pollution*, *276*, 116766. https://doi.org/10.1016/j.envpol.2021.116766

Hooper, R.P., & Norman, E.P. (1989) Use of multivariate analysis for determining sources of solutes found in wet atmospheric deposition in the United States. *Environmental Science & Technology*, *23*(10), 1263-1268. https://doi.org/10.1021/es00068a013

Huang, W., Tan, J., Kan, H., Zhao, N., Song, W., Chen, G., Jiang, L., Jiang, C., Cheng, R., & Chen, B. (2009). Visibility, air quality and daily mortality in Shanghai, China. *Science of the Total Environment*, *407*(10), 3295-3300. https://doi.org/10.1016/j.scitotenv.2009.02.019

Ignaccolo, R., Ghigo, S., & Giovenali, E. (2008). Analysis of air quality monitoring networks by functional clustering. *Environmetrics*, *19*(7), 672-686. https://doi.org/10.1002/env.946

Iizuka, A., Shirato, S., Mizukoshi, A., Noguchi, M., Yamasaki, A., & Yanagisawa, Y. (2014). A cluster analysis of constant ambient air monitoring data from the Kanto Region of Japan. *International Journal of Environmental Research and Public Health*, *11*(7), 6844-6855. https://doi.org/10.3390/ijerph110706844

Isiyaka, H.A., & Azid, A. (2015). Air quality pattern assessment in Malaysia using multivariate techniques. *Malaysian Journal of Analytical Sciences*, *19*(5), 966-978

Javed, W., & Guo, B. (2021). Chemical characterization and source apportionment of fine and coarse atmospheric particulate matter in Doha, Qatar. *Atmospheric Pollution Research*, *12*(2), 122-136. https://doi.org/10.1016/j.apr.2020.10.015

Juahir, H., Zain, S. M., Yusoff, M. K., Hanidza, T. T., Armi, A. M., Toriman, M. E., & Mokhtar, M. (2011). Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques. *Environmental Monitoring and Assessment*, *173*, 625-641. https://doi.org/0.1007/s10661-010-1411-x

Junninen, H., Niska, H., Tuppurainen, K., Ruuskanen, J., & Kolehmainen, M. (2004). Methods for imputation of missing values in air quality data sets. *Atmospheric Environment*, *38*(18), 2895-2907. https://doi.org/10.1016/j.atmosenv.2004.02.026

Keresztes, R., & Rapo, E. (2017). Statistical Analysis of-air pollution with specific regard to factor analysis in the Ciuc basin, Romania. Studia Universitatis Babes-Bolyai. *Chemia*, *62*(3), 283-293. https://doi.org/10.24193/subbchem.2017.3.24

Kumar, P. (2022). A critical evaluation of air quality index models (1960–2021). Environmental Monitoring and Assessment, 194(5), 1-45. https://doi.org/10.1007/s10661-022-09896-8

Lu, P., Casagli, N., Catani, F., & Tofani, V. (2012). Persistent Scatterers Interferometry Hotspot and Cluster Analysis (PSI-HCA) for detection of extremely slow-moving landslides. *International Journal of Remote Sensing*, *33*(2), 466-489. https://doi.org/10.1016/j.isprsjprs.2019.08.004

Mutalib, S. N. S. A., Juahir, H., Azid, A., Sharif, S. M., Latif, M. T., Aris, A. Z., ... & Dominick, D. (2013). Spatial and temporal air quality pattern recognition using environmetric techniques: A case study in Malaysia. *Environmental Science: Processes & Impacts*, *15*(9), 1717-1728. https://doi.org/10.1039/c3em00161j

Nazif, T. M., Chen, S., George, I., Dizon, J. M., Hahn, R. T., Crowley, A., ... & Kodali, S. K. (2019). New-onset left bundle branch block after transcatheter aortic valve replacement is associated with adverse long-term clinical outcomes in intermediate-risk patients: an analysis from the PARTNER II trial. *European Heart Journal*, *40*(27), 2218-2227. https://doi.org/10.1093/eurheartj/ehz227

Nie, W., Liu, X., Liu, C., Guo, L., & Hua, Y. (2022). Prediction of dispersion behavior of typical exhaust pollutants from hydraulic support transporters based on numerical simulation. *Environmental Science and Pollution Research*, *29*(25), 38110-38125. https://doi.org/10.1007/s11356-021-17959-5

Núñez-Alonso, D., Pérez-Arribas, L. V., Manzoor, S., & Cáceres, J. O. (2019). Statistical tools for air pollution assessment: multivariate and spatial analysis studies in the Madrid region. *Journal of Analytical Methods in Chemistry*, *2019*. https://doi.org/10.1155/2019/9753927

Ramson, E., Oluchi, N. E., & John, U. (2016). Multivariate Analysis of Air Quality in Selected Oil Operating Areas in the Niger Delta Region of Nigeria. *In SPE Nigeria Annual International Conference and Exhibition*, pp. SPE-184363. https://doi.org/ 10.5772/16817

Rani, N. L. A., Azid, A., Khalit, S. I., Juahir, H., & Samsudin, M. S. (2018). Air Pollution Index Trend Analysis in Malaysia, 2010-15. *Polish Journal of Environmental Studies*, 27(2). https://doi.org/10.15244/pjoes/75964

Stacey, P., Clegg, F., Rhyder, G., & Sammon, C. (2022). Application of a Fourier Transform Infrared (FTIR) Principal Component Regression (PCR) Chemometric Method for the Quantification of Respirable Crystalline Silica (Quartz), Kaolinite, and Coal in Coal Mine Dusts from Australia, UK, and South Africa. *Annals of Work Exposures and Health*, *66*(6), 781-793. https://doi.org /10.1093/annweh/wxab119

Vakarelska, E., Nedyalkova, M., Nikolova, N., Angelov, C., Tonev, D., Prybilova, P., ... & Simeonov, V. (2021). Tracing the movement of persistent organic pollutants at a high-mountain sampling site by chemometric assessment. *Journal of Environmental Science and Health*, Part A, *56*(9), 1041-1049. https://doi.org/10.1021/es048859u

Wang, P., Tang, J., Wang, S., Dong, X., & Fang, J. (2018). Regional heatwaves in China: a cluster analysis. *Climate Dynamics*, *50*, 1901-1917. https://doi.org/10.1175/JCLI-D-18-0256.1

Wang, Y., Ding, D., Ji, X., Zhang, X., Zhou, P., Dou, Y., ... & Shu, M. (2022). Construction of Multipollutant Air Quality Health Index and Susceptibility Analysis Based on Mortality Risk in Beijing, China. Atmosphere, 13(9), 1370. https://doi.org/10.3390/atmos13091370

Wold, S., Kim, E., & Paul, G. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, *2*(3), 37-52. https://doi.org/10.1016/0169-7439(87)80084-9

Wu, M., Wu, D., Fan, Q., Wang, B. M., Li, H. W., & Fan, S. J. (2013). Observational studies of the meteorological characteristics associated with poor air quality over the Pearl River Delta in China. *Atmospheric*

Chemistry and Physics, *13*(21), 10755-10766. https://doi.org/10.5194/acp-13-10755-2013

Yadav, M., Singh, N. K., Sahu, S. P., & Padhiyar, H. (2022). Investigations on air quality of a critically polluted industrial city using multivariate statistical methods: Way forward for future sustainability. *Chemosphere*, *291*, 133024. https://doi.org/10.1016/j.chemosphere.2021.133024

Yang, Q., Liu, G., Gonella, F., Chen, Y., Liu, C., Zhao, H., & Yang, Z. (2022). Assessing the temporal-spatial dynamic reduction in ecosystem services caused by air pollution: A near-real-time data perspective. Resources, Conservation and Recycling, 180, 106205. https://doi.org/10.1016/j.resconrec.2022.106205

Zipper, C. E., & Skousen, J. (2021). Coal's legacy in Appalachia: Lands, waters, and people. *The Extractive Industries and Society*, *8*(4), 100990. https://doi.org/10.1016/j.exis.2021.

How to cite this Article:

Navin Prasad, Tanushree Bhattacharya and Bindhu Lal (2023). Chemometric techniques in the assessment of ambient air quality, and development of air quality index of coal mining complex: A statistical approach. *International Journal of Experimental Research and Review*, *36*, 433-446.

DOI: ttps://doi.org/10.52756/ijerr.2023.v36.018a