Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Groundwater Quality Evaluation Using Pollution Indices and Multivariate Analysis: A Case Study of Faridabad, Haryana, India

Eksha Guliani
Manav Rachna University
Vinayak V. Pathak
Manav Rachna University
Sangita Banga
Manav Rachna University
Adarsh Kumar Pandey
adarsh.889@gmail.com
United Arab Emirates University
Amit Awasthi
University of Petroleum & Energy Studies

Article

Keywords: Faridabad, Groundwater, Heavy Metal Pollution Index, Principal Component Analysis, Water Quality Index

Posted Date: August 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10617493/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Groundwater Quality Evaluation Using Pollution Indices and Multivariate Analysis: A Case Study of Faridabad, Haryana, India

Eksha Guliani[1], Vinayak V. Pathak[1], Sangita Banga[1], Adarsh Kumar Pandey[2*], Amit Awasthi[3]

[1]*Department of Basic Sciences (Chemistry), School of Sciences, Manav Rachna University, Faridabad, 121004, Haryana, India*

[2]*Mechanical and Aerospace Engineering Department, College of Engineering, United Arab Emirates University, Al Ain, 15551, United Arab Emirates*

[3]*Department of Applied Sciences, University of Petroleum & Energy Studies, Uttarakhand, 248007, India*

Corresponding Author: Dr. Adarsh Kumar Pandey

Email:- adarsh.889@gmail.com

Abstract

The present investigation is an attempt to evaluate groundwater (GW) quality of Faridabad city by performing multivariate statistical analysis. Key water quality parameters metals were analyzed in GW samples collected from rural (S1, S2), urbanized residential (S3, S4), residential areas near industries (S5, S6) and industrial areas (S7, S8) of Faridabad. The minimum water quality index (WQI) was for S7 and S8 (23.80±1.36–25.27±0.48) followed by S5 and S6 (30.95±2.27–34.78±1.72). Moderate WQI value was for S3 and S4 (59.41±1.53–66.55±4.69) whereas good WQI for S1 and S2 (83.00±2.65–89.20±1.06). Heavy metal pollution index (HPI) was extremely high (>1390) in S7 and S8 along with the highest degree of contamination (C_d) (>16). On the other hand, GW quality for rural areas was good with HPI values (<65) and almost no C_d. Principal Component Analysis and Pearson's correlation matrix determined that mineralization and redox-controlled input drives GW quality. TDS exhibited a perfect correlation with electrical conductivity (r=0.99), hardness (r=0.97) and pH (r=0.89) indicating significant ionic enrichment. Ward's hierarchical clustering method separated sites into clusters whereas Monte Carlo simulation determined robustness of obtained dataset (Mean≈265). The proposed multivariate study provides robust methodology for evaluating GW quality for urban regions like Faridabad.

Keywords: Faridabad; Groundwater; Heavy Metal Pollution Index; Principal Component Analysis; Water Quality Index

1. Introduction

Groundwater (30.1% of total freshwater) provides the source of drinking water (~50%) for both rural (~85%) and urban residents (~50%). The aquifers that make up the Indian subcontinent are experiencing great stress, due to the fact that more than 30% of all groundwater units evaluated are labelled as over-exploited [1]. Its highest levels of exploitation are recorded in Delhi, Haryana, Punjab, Rajasthan and Tamil Nadu due to large amounts of agricultural, industrial and domestic pumping use. In Haryana, ~7.5 km^3 of groundwater is being extracted annually which often exceeds the recharge provided from the natural monsoon season of approximately 4-6 km^3/yr. Thus, resulting in declining water tables of between 1 and 3 m/yr in certain parts of the Yamuna-Gurgaon-Faridabad belt. Groundwater recharge takes place through rainfall, canal seepage and rainwater runoff from impermeable surfaces. The Faridabad district is a part of the Indo-Gangetic alluvial aquifer system with shallow to semi-confined aquifers having static water levels.

The primary sources of heavy metals being introduced into the groundwater include industrial effluents, sewage discharge, leachate from agricultural waste and production of plastics [3]. Even at very low concentrations, many types of heavy metals have shown to be extremely harmful to human health. For example, WHO guidelines recommend that copper concentration in drinking water should not be higher than 2 mg/L [4]. The chronic exposure to higher levels of copper or any other heavy metal like lead has been associated with various harmful disorders like renal failure, delayed neuro-development in children and increased risk of cancer [5]. According to zoning reports for the industrial belt of Faridabad, many different types of electroplating, metal finishing and chemical plants are discharging their wastewater (untreated industrial effluent) into local canals. It percolates into shallow aquifers raising concern of their long term contamination. The presence of heavy metals beyond their permissible limits as designated by the Bureau of Indian Standards are being routinely detected in wells throughout the region. If monitoring and mitigation of heavy metal contamination is not looked upon, the local communities relying on groundwater as a source of drinking water will face serious health issues as well as serious economic consequences due like reduced crop yields [6]. Therefore, the measurement of heavy metal contamination is a part of early warning systems for sustainable use of groundwater resources in Faridabad. Without proper surveillance, the progressive build-up of contamination will go undetected until concentration levels reach those that create irreversible impacts on health and ecology.

A lot of studies have been conducted previously that have evaluated the groundwater conditions in various regions of Faridabad district. Researchers have used many different techniques to perform the characterization of site specific water which resulted in different types of reporting formats. For example, Swain et al. (2022) carried out an extensive geospatial-statistical analysis to characterize and identify key physico-chemical parameters (e.g., pH, electrical conductivity (EC), major ions, nitrate and fluorides) from a wide range of sites in both Faridabad and Gurgaon [7]. Additionally, they demonstrated significant spatial patterns and variations within the groundwater based on the geographical distribution of industry within the National Capital Region (NCR). Similarly, Singh et al. (2024) used the method of chemometrics to assess the potential health risks of groundwater and indicated which wells within the NCR are likely to be unsafe for drinking water [8]. They have then developed groupings of the factors that contributed most to the predictive modelling of health risks for those wells. Furthermore, various researchers have created

a spatially aware map of the intensity of heavy metals in the ground and the areas that would benefit most from remediation efforts. At an organizational level, the Central Ground Water Board (2024) regularly releases bulletins by outlining the changes caused in groundwater quality in Haryana, before and after monsoon season [9]. According to this report, on the basis of districts segregation, a continued increase in the concentration of specific contaminants in the groundwater within Faridabad has been observed. In addition to the documented cases of localized contamination from waste facility mis-management (specifically, leachate drainage from Bandhwari Landfill), it has been determined that the drainage from this type of waste management indeed contaminates the surrounding area's groundwater. Thus, there are clear pathways through which heavy metals and organic compounds (e.g., municipal solid waste) can leach from the landfill and contaminate groundwater aquifers located close by. Some studies have found elevated levels of total dissolved solids (TDS), hardness and elevated concentrations of metals in groundwater [10]. However, these researches used only limited numbers of parameters without comprehensively evaluating the other sources contributing to groundwater contamination and were therefore unable to fully assess uncertainty in findings. The use of limited sets of parameters and not quantifying the uncertainty in the identified parameters, shows the weak connections between the statistical trends in the previous researches.

The present study aims to examine the groundwater quality of Faridabad city by following multivariate statistical analysis. On the basis of previously reported literature, the study eliminates the research gaps that were identified in previous assessments of groundwater quality for Faridabad region. First, a variety of pollution indices like Water Quality Index (WQI), Heavy Metal Pollution Index (HPI) and Degree of Contamination (C_d) were combined with multivariate statistical methods namely Principal Component Analysis (PCA), Monte Carlo Simulation, Hierarchical Cluster Analysis (HCA) and Pearson's Correlation Matrix to simultaneously identify the hydrochemical processes involved. Second, unlike many of the previous deterministic studies, we account for uncertainty in our analysis by using Monte Carlo simulation to provide ranges of WQIs rather than single-value indices. This approach allows the comparison of sites based on both the expected and uncertain aspects of WQI, allowing us to rank the sites as per risk-based approaches. Lastly, we have developed a transferable workflow from our study that can be used to target monitoring and remediation efforts in industrially impacted Faridabad through the mapping of pollutant indices in correlation with the loadings from correlation analysis.

2. Materials and Methods

2.1 Description of Site and Sampling

The study area for the present research is Faridabad city, located in the south-eastern part of Haryana, India (Fig. 1). Faridabad (28°10'50"N to 28°29'04"N latitude and 77°06'49"E to 77°33'23"E longitude) borders Delhi to the north, Gurgaon to the west, Palwal to the south and Uttar Pradesh to the east. It covers about 742.90 km^2 of area and is a part of National Capital Region (NCR). Faridabad has a sub-tropical continental monsoon climate with a satisfactory distribution of rainfall. Also, Faridabad ranks on the top in terms of industrialization throughout the state due to which it has attracted a number of multi-national companies. According to CGWB (Central Ground Water Board), Faridabad area is largely drained by the Yamuna River along with other small tributaries connected to the Aravalli mountains, running into large-sized, well-draining areas. These two characteristics provide fine particulate matter as the primary means through which groundwater obtains a higher salinity or higher EC at pH levels of 7.75-8.62. Then, as water flows down through sand, there will be high concentrations of dissolved minerals that were already present in the silt, kankar and gravel layers. That downward water flow as it drains into the existing shale aquifers, along with existing material and sediment overlying the aquifers and the overall slope/flow characteristics toward the southeast, is how dissolved salts and contaminants are being transmitted through the aquifer system [11].

A total of 24 samples of groundwater were collected from the selected 8 locations during pre-monsoon (April-June), monsoon (July-September) and post-monsoon (October-December) season. The groundwater samples were collected within the radius of 2 km by taking Manav Rachna University, Faridabad, India as the center point or origin. Therefore, the groundwater samples were collected from rural/village areas (S1 and S2), urbanized residential areas (S3 and S4), residential areas located near industry (S5 and S6) and industrial areas (S7 and S8). The samples were filtered using a 0.22 μM microfilter and stored at 4 ℃ in polypropylene bottles as per the standard protocol. All the samples were acidified with 10% HNO_3 and analyzed on ICP-AES to determine the presence of heavy metals. Every sample was analyzed in a batch of triplicates in order to ensure the accuracy of the measurements. The robustness of the obtained data was evaluated by causing uncertainty and variability in the dataset and then monitored by performing Monte Carlo simulation using python programming language. The statistical parameters of heavy

metal contamination were carried out by evaluating the water pollution indices: water quality index (WQI), heavy metal pollution index (HPI) and degree of contamination (C_d). The study was validated using various multivariate analysis methods: principal component analysis (PCA), hierarchical cluster analysis (HCA) and Pearson's correlation matrix. The multivariate statistical analysis was performed using SPSS 22 v software along with MS Excel.

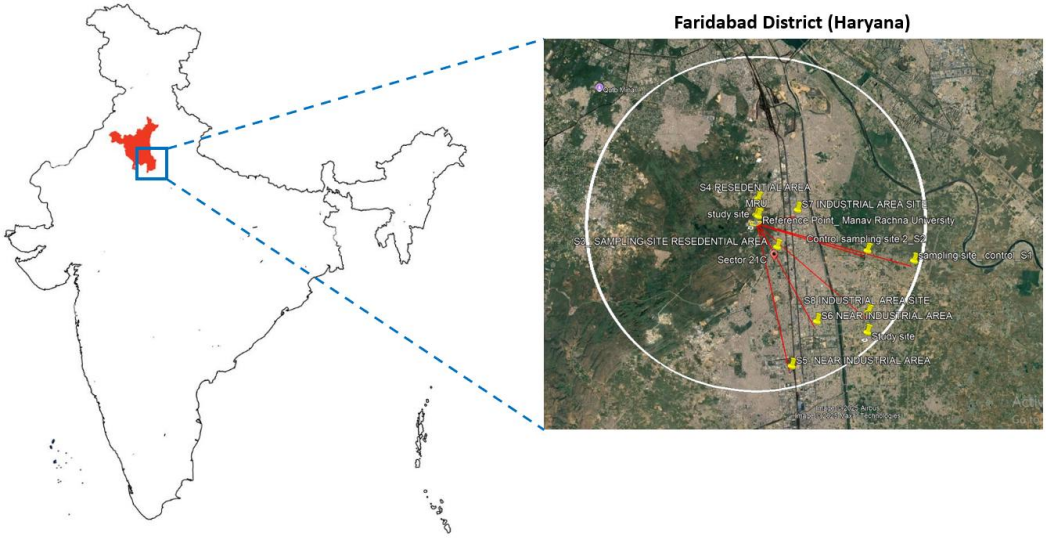


Fig. 1: Location of sampling site chosen is Faridabad district located in the south-eastern direction of Haryana, India.

2.2 Evaluation of Physicochemical Parameters

All analytical measurements of wastewater parameters were conducted using the Indian Standard (IS 3025) series and according to accepted environmental laboratory methods to ensure both reproducibility and regulatory compliance. The instruments used for testing were calibrated using certified reference materials as well as including analytical blanks to detect any potential contamination during the processing of the samples [12].

2.2.1 pH

A digital pH meter with a combined glass electrode was used to measure the pH of the wastewater samples. Prior to measuring, the instrument was calibrated with pH 4.0, 7.0 and 9.0 buffer solutions. The pH value is an essential measurement as it affects microbial activity, the speciation of nutrients, the solubility of metals and the efficiency of treatment in the groundwater samples.

2.2.2 Temperature

To avoid temperature changes during transport, temperature was measured *in situ* with a digital thermometer. Temperature has a significant impact on the rates of biochemical reactions, solubility of oxygen and microbial metabolism in the groundwater samples.

2.2.3 Conductivity

Electrical conductivity of water samples was evaluated using a digital conductivity meter according to the APHA Standard Method 2510B. The conductivity probe was calibrated using standard potassium chloride (KCl) solution prior to obtaining the conductivity reading. The electrode was then immersed in the water sample, and the final reading taken once it was stable. The results of the conductivity measurements were reported as µS/cm.

2.2.4 Total Dissolved Solids (TDS)

Total dissolved solids were determined by evaporating the filtered sample to dryness at 100 °C. The total dissolved solids reflect the amount of dissolved inorganic salts and organic materials present and will influence how much conductivity is produced and how much osmotic stress will be caused in the groundwater samples.

2.2.5 Total Suspended Solids (TSS)

Determining total suspended solids was done using a gravimetric method. A known volume of sample was filtered through pre-weighed glass fiber filter paper and the filtered paper was then dried at 70-80 °C to constant weight. The difference in weight of the dried paper represented the concentration of suspended solids in the sample. The concentration of total suspended solids is directly related to turbidity and potential for sedimentation.

2.2.6 Turbidity

Turbidity in the water samples was analyzed using the Nephelometric Methods of the APHA Standard Method (2130 B). In this method, turbidity values are determined by measuring the amount of light that is scattered by suspended solids, at a pre-determined angle from the vertical, in response to incident light. All turbidity values were recorded using a digital nephelometer (turbidity meter) that was pre-calibrated with standard formalin suspension solutions. A water sample was placed into a clean sample cuvette that was free of air bubbles and fingerprints on the

cuvette surface. The cuvette was positioned in the turbidity meter for reading, and the turbidity was read when the reading reached stability. The turbidity of each sample was recorded in Nephelometric Turbidity Units (NTU).

2.2.7 Alkalinity

The alkalinity of the sample water was determined using the acid titration method (APHA 2320 B). An aliquot of the water sample was titrated with H_2SO_4 using phenolphthalein and methyl orange as indicators to determine the alkalinity of phenolphthalein and total alkalinity, respectively. The end of the titration was determined by a color change: the color of phenolphthalein changed from pink to colorless, while the color of methyl orange changed from yellow to orange. The calculated alkalinity values are presented as mg/L $CaCO_3$.

2.2.8 Dissolved Oxygen (DO)

Dissolved oxygen values for the water samples were determined using the Winkler method (APHA 4500-O). This method utilizes manganese sulfate and alkaline iodide addition to the water sample in order to fix the dissolved oxygen. After the samples were acidified, an equivalent amount of liberated iodine was released as a result of the acidification which will then be titrated with standard sodium thiosulfate using starch as that indicator in order to determine the concentration of dissolved oxygen. The concentration of dissolved oxygen is recorded as mg/L.

2.2.9 Biochemical Oxygen Demand (BOD)

To determine the BOD at a temperature of 27°C, three days of samples were incubated in BOD bottles which contain diluted samples according to Indian Standard Methods (IS: 3025). As a standard practice before an incubation, the concentration of dissolved oxygen in the sample was measured with the use of the Winkler Method. Following the incubation period, the concentration of dissolved oxygen was again measured. The BOD value of the sample provides an indication of how much biodegradable organic material exists in a waste water sample; the BOD value is also an indirect indication of the treatment efficiency of the water sample through biological means.

2.2.10 Chemical Oxygen Demand (COD)

COD was measured using the closed reflux dichromate method according to APHA Standard Method 5220 C. The groundwater samples were digested with potassium dichromate in an acidic

medium under reflux conditions. The dichromate oxidizes the organic matter and residual dichromate is then titrated with ferrous ammonium sulfate. COD is the measure of total oxidizable organic load in groundwater sample and a measure of the strength of the pollution.

2.2.11 Hardness

The hardness of the sample water was determined by the EDTA titrimetric method as outlined by APHA Standard Methods (APHA 2340 C). To measure the hardness, a measured volume of the water sample was buffered to pH 10 using an ammonium buffer, and Eriochrome Black T indicator was added. The sample was then titrated with a standard solution of EDTA until the color of the sample changed from wine red to pure blue; this indicated that the endpoint had been reached. The calculated hardness values are presented as mg/L CaCO3.

2.2.12 Phosphate Content

The level of phosphates was found by the Ascorbic Acids Spectrophotometric Method (APHA 4500-P E). The method involves the testing of phosphates, ammonium molybdate and antimony potassium tartrate in an acidic solution to develop a phosphomolybdenum complex, which is affected by Ascorbic Acid and develops colour. Using a UV-Visible Spectrophotometer the intensity of the colour was measured at 880 nm, as well as using a standard calibration curve to determine phosphates in solution.

2.2.13 Nitrate Content

The level of Nitrates in the water samples was obtained using the UV spectrophotometric method (APHA 4500-NO_3 B). The absorbance was measured at 220 nm using a UV-Visible Spectrophotometer, and absorbance for organics was corrected using 275 nm values. The Nitrate level in the sample was then determined using a standard calibration curve, and reported as mg/L NO_3^-.

2.2.14 Heavy metals

Heavy metals including copper and iron were measured using APHA 3111 B, atomic absorption spectrophotometry (AAS). Multi-element calibration curves were prepared for the metals; detection limits were established to ensure analytical reliability and all metals were monitored

based on industrial contributions of heavy metals to the sampling site as well as their respective ecological toxicity.

2.3 Evaluation of Water Pollution Indices

The water pollution level is determined by estimating the values of water pollution indices [13]. In the present investigation, WQI, HPI and C_d are calculated to define the contamination level in water samples and compared with the guidelines (Table 1) as laid by National Sanitation Foundation (NSF). Therefore, this section summarizes the suitability of water quality by evaluating the combined effect of all the parameters.

2.3.1 Estimation of Water Quality by Water Quality Index (WQI)

The water quality index is a comprehensive tool that helps in assessing the overall groundwater quality and is calculated by arithmetic weighted method [14]. This method integrates all the values of physico-chemical parameters by converting it into a sub-index score known as Quality Rating (Q_i) as shown in Eq. 1:

$$Q_i = \frac{(V_n - V_i)}{(V_s - V_i)} \times 100 \qquad \text{(Eq. 1)}$$

In the above equation, (V_n) is the n^{th} parameter amount, V_i is the ideal value (V_i=0) of given parameter which is 7 in case of pH, V_s is the standard value of the parameter as set by NSF guidelines. The next step involves the calculation of relative weight (W_i) by dividing the recommended standard value (S_i) of the assigned parameter as given in Eq. 2.

$$W_i = \frac{1}{S_i} \qquad \text{(Eq. 2)}$$

All the sub-indices (Q_i) are multiplied with their standard weight factors (W_i). These are then added together and divided by the summation of weight factors to obtain the desired Water Quality Index (WQI) as shown in Eq. 3.

$$WQI = \frac{\sum_{i=1}^{n} (Qi \times Wi)}{\sum_{i=1}^{n} Wi} \qquad \text{(Eq. 3)}$$

The calculated WQI values reflects a cumulative impact of all the identified physico-chemical parameters on water quality. Water can be divided into five categories based upon the WQI value obtained as compared according to NSF WQI values: excellent (90-100), good (70-89), medium (50-69), bad (25-49) and very bad (0-24).

2.3.2 Estimation of Water Pollution by Heavy Metal Pollution Index (HPI)

The heavy metal pollution index is one of the most valuable metrics that assess the heavy metal pollution level in the water [15]. It helps in determining the water suitability to consume in the presence of various heavy metals. It is calculated using weighted arithmetic mean calculation as performed to calculate the WQI. According to the relative importance of the metal, their relative weighted values are assigned between 0 and 1 which can be calculated by Eq. 2. The sub-index factor (Q_i) in this case is calculated as follows:

$$Q_i = \sum_{i=1}^{n} \frac{|M_i - I_i|}{S_i - I_i} \times 100 \quad \text{(Eq. 4)}$$

Here, M_i is the monitored value of the heavy metal, I_i is the ideal value and S_i is the standard value of the i^{th} heavy metal. The required HPI is determined from the following formula (Eq. 5):

$$HPI = \frac{\sum_{i=1}^{n} (Qi \times Wi)}{\sum_{i=1}^{n} Wi} \quad \text{(Eq. 5)}$$

The calculated HPI values are compared with the accepted standards for drinking water. If HPI<100, the water sample contains less contamination of heavy metals which is considered suitable for drinking purpose. On the other hand, if HPI>100, the water sample contains high contamination of heavy metals which can cause adverse effect on health if consumed (Table 1). Therefore, this kind of water is unsuitable for drinking and hence, HPI rating method evaluates the influence of heavy metals on the overall water quality.

2.3.3 Estimation of Water Pollution Level by Degree of Contamination (C_d)

The Degree of Contamination (C_d) is another pollution index that quantitatively evaluates the extent of contamination of heavy metals in the water sample [16]. It is calculated by taking the

summation of all the contamination fit factors (C_{fi}) for respective heavy metals as shown in Eq. 6 and Eq. 7.

$$C_{fi} = \frac{C_{Ai}}{C_{Ni}} - 1 \quad \text{(Eq. 6)}$$

$$C_d = \sum_{i=1}^{n} C_{fi} \quad \text{(Eq. 7)}$$

Here, C_{Ai} is the measured concentration and C_{Ni} is the maximum permissible concentration for the i^{th} parameter. The obtained C_d values provide a direct cumulative measure of the severity of heavy metal contamination in water sample. Based upon the calculated values of C_d, water can be categorized into three classes: low contamination ($C_d < 1$), medium contamination ($C_d = 1$-3) and high contamination ($C_d > 3$) (Table 1).

Table 1: The standard range of pollution indices as set by NSF guidelines depicting groundwater quality.

Pollution Indices	Standard Range	Remark
WQI	90-100	Excellent
	70-89	Good
	50-69	Medium
	26-49	Bad
	0-25	Very Bad
HPI	>100	Bad
	<100	Good
C_d	<1	Low
	1-3	Medium
	>3	High

2.4 Multivariate Statistical Methods

In order to determine the most influencing variable affecting the water quality, multivariate statistical analysis has been performed. The multivariate statistical analysis used in the present

study are Principal Component Analysis (PCA), Monte Carlo Simulation, Hierarchical Cluster Analysis (HCA) and Pearson's Correlation Matrix.

2.4.1 Identification of Dominant Factors by Principal Component Analysis (PCA)

The PCA is performed using Kaiser Normalization method (24 samples analyzed from 8 locations) to establish a homologous behavior among the dominant physico-chemical parameters of procured water samples [17]. Before applying PCA, the values obtained were first converted into *Z*-score normalization (Eq. 8) in order to eliminate any sort of dimensional bias due to different measurement units.

$$Z = \frac{X - \mu}{\sigma} \qquad \text{(Eq. 8)}$$

Here, X is the measured value, μ is the calculated mean and σ is the standard deviation of the considered dataset. In this study, the eigen values of the physico-chemical factors signify the variance in the data which is represented in the form of factor scores. The Principal Component (PC) values are represented as positive and negative values based upon the impact of relationship among the parameters (strong, medium or weak).

2.4.2 Estimation of Uncertainty and Variability by Monte Carlo Simulation (MCS)

Monte Carlo Simulation is another statistical model that utilizes random sampling process to understand the uncertainty and variability associated with the collated dataset [18]. In order to ensure the reliability of MCS, 10,000 iterations were carried out using Python and simulated values were used to calibrate the efficiency of model. This model provides insights of all the possible outcomes associated with the considered dataset.

2.4.3 Hierarchical Cluster Analysis (HCA) by Ward's Linkage Method

Hierarchical cluster analysis (HCA) is conducted in order to group the groundwater sampling sites based on the extent to which the physico-chemical characteristics of each site were similar to one another [19]. The Euclidean distance is used in conjunction with Ward's linkage method, which minimizes the amount of variability between samples within each group and creates compact (easily interpretable) clusters. Origin software was used to create a cluster analysis and dendrogram. The resulting dendrogram provided the basis for grouping sampling locations with

similar hydrochemical signatures and for assessing how different land use patterns affect groundwater quality geographically.

2.4.4 Statistical Significance Test by Pearson's Correlation Matrix

The Pearson's Correlation analysis is performed in order to quantify the strength and direction of linear relationships among groundwater quality indicators [20]. It is formulated by calculating the correlation coefficient (r) which ranges between -1 and +1. The correlation strength was assessed of the basis of their magnitude: those with $|r| \geq 0.70$ being classified as strong; $0.50 \leq |r| < 0.70$ being moderate and $|r| < 0.50$ weak. This classification follows by identifying those groundwater quality parameters that may be impacted by similar hydrochemical processes or share the same source. The correlation analysis also assessed the interrelationship between salinity indicators, nutrient indicators, metals contamination and organic pollution parameters. The correlation matrix formed as a result of the Pearson's correlation analysis is also used to validate and reduce the number of dimensions of the PCA allowing for the most accurate interpretation of multivariate pattern of groundwater quality.

3. Results and Discussion

3.1 Estimation of the Values of Physicochemical Parameters

The groundwater analyses indicated seasonal variation in physical and chemical parameter data (Table 2). Generally, groundwater quality during the pre-monsoon and monsoon seasons is dominated by slightly alkaline pH level (6.03 to 7.9) than the average range of pH based on BIS permissible limits (6.5-8.5). The higher combined values observed for electrical conductivity (389 to 1695 µS/cm) and total dissolved solids (213 to 1188 mg/L) indicates groundwater mineralization processes, particularly during the pre-monsoon season. Therefore, there appears to be some anthropogenic influence. Total alkalinity (306 to 771 mg/L) and total hardness (292.5 to 764.7 mg) were frequently above BIS desirable limits (200 mg/L), often as the result of carbonate weathering processes. Water turbidity was consistently low (0 to 0.7 NTU), however, dissolved oxygen concentrations tended to vary depending on seasonal recharge from rainfall or biological (microbial) activities, with dissolved oxygen levels ranging from 0.6 to 7.7 mg/L.

Overall, BOD (0-3.8 mg/L) and COD (0-9.05 mg/L) are the main organic pollution indicators. Regardless of this, most sites had minimal organic contamination since these values were found to

be either within the permissible limit. Similarly, nutrient concentrations were also determined to be low overall; ortho-phosphate was recorded in the range from 0 to 0.12 mg/L and NO_3^- was only present in the range from 0.15 to 3.7 mg/L. Both of these parameter concentrations are within the limits set by the BIS. However, concentrations of some heavy metals had elevated levels in various locations; Cu (0.94 mg/L) and Fe (6.96 mg/L) were found to consistently exceed permissible limits, particularly in areas positioned near potential anthropogenic or industrial sources. The majority of parameters presented have reasonably acceptable values, although elevated EC, TDS, alkalinity, hardness, and levels of selected metals indicated local groundwater quality degradation as a result of both natural geochemical processes and anthropogenic activities.

Table 2: The values of physicochemical parameters that determine the groundwater quality of all the identified sampling sites (S1-S8) showing seasonal variation.

Site	Season	pH	Temperature (^{0}C)	Conductivity (µS/cm)	TDS (NTU)	TSS (mg/L)	Turbidity	Alkalinity (mg/L)	DO (mg/L)	BOD (mg/L)	COD (mg/L)	Hardness (mg/L)	P (mg/L)	NO_3^- (mg/L)	Cu (mg/L)	Fe (mg/L)
S1	Pre_Monsoon	7.9	28.9	480	253	52.5	0	470	7.05	1.6	5.75	428.75	0.08	0.79	0	0.24
	Monsoon	7.8	29.6	389	231	47.5	0.7	445	7.7	2.3	5.24	398.4	0	1.25	0.05	0.18
	Post_Monsoon	7.78	25.5	412	270	57.5	0.7	478.5	7.25	1.4	5.85	325.7	0.09	0.39	0.1	0.28
S2	Pre_Monsoon	7.8	29.3	398	225	50.5	0	459.5	6.9	1.4	5.85	322.3	0.08	0.52	0	0.28
	Monsoon	7.52	28.4	410	213	45	0.2	447.5	6.82	1.65	5.45	340.9	0.02	1.07	0.28	0.19
	Post_Monsoon	7.75	24.8	490	264	55	0.6	487.5	6.55	1.25	67	378.7	0	0.67	0.32	0.23
S3	Pre_Monsoon	7.68	24.55	613.5	364	69	0	372	6.7	1.2	3.35	318	0.05	0.15	0.28	1.14
	Monsoon	7.62	27.6	661	383.5	88	0.5	364	6.1	1.3	3.15	336.5	0	0.26	0.29	0.56
	Post_Monsoon	7.4	25.65	622.5	355.5	64.5	0	320	6.8	0.95	3.5	387.5	0.09	0.37	0.31	0.83
S4	Pre_Monsoon	7.54	25.5	634	339	56	0	306	7.15	1.8	3.8	349.5	0.08	0.32	0.29	1.06
	Monsoon	7.34	28.9	674	387.5	53	0.2	380	6.5	1.25	3.1	322	0.083	0.23	0.45	0.82

	Post_Monsoon	7.52	24.7	650	362	62.5	0.5	322	7.5	1.05	2.2	292.5	0.12	0.46	0.33	0.71
S5	Pre_Monsoon	7.05	24.85	1360	877	83	0	777	1.6	2.55	7.2	663.6	0.062	1.07	0.9	1.45
	Monsoon	7.21	28.3	1335.5	771	79	0.7	698	1.75	1.75	5.4	711.8	0.079	0.52	0.94	1.12
	Post_Monsoon	7.36	29.95	1438	993	106	0	705.5	1.05	2	9.05	695.9	0.07	1.09	0.79	1.29
S6	Pre_Monsoon	7.27	25.05	1298	854	102	0	754	0.9	1.9	5.05	764.7	0.04	1.49	0.58	1.99
	Monsoon	7.1	29.6	1307.5	793.5	94.5	0.2	694	1.1	1.1	5.5	758.7	0.07	0.79	0.6	1.01
	Post_Monsoon	7.16	28.05	1404	898	126	0	771	0.6	1.55	8.7	719.6	0.056	1.25	0.65	1.56
S7	Pre_Monsoon	6.9	23.85	1590.1	1178	183	0	624	5.2	2.97	6.8	453	0.095	3.1	0.23	5.53
	Monsoon	6.03	28.05	1463.6	1048	172	0	580	6.89	2.5	7.6	413	0.098	3.7	0.23	4.65
	Post_Monsoon	6.64	27.75	1695.2	1134	126	0	690	6	2.8	5.4	464	0.097	3.5	0.14	6.04
S8	Pre_Monsoon	7.04	24.6	1511.8	1038	106	0	570	4.1	3.8	5.1	420	0.091	3.01	0.25	6.96
	Monsoon	6.02	28.55	1595.9	1188	185.5	0	530	7.2	2.37	6.3	493	0.097	3.25	0.29	4.29
	Post_Monsoon	6.9	26.95	1640.7	1178	154.5	0	740	6.5	3.68	5.3	468	0.095	3.2	0.38	5.62
	BIS LIMITS	**8.5**	**NA**	**1500**	**500**	**NA**	**5**	**200**	**8.26**	**3**	**<20**	**200**	**0.1**	**45**	**0.05**	**0.3**

380

3.2 Result Analysis of Water Pollution Indices

The results of WQI, HPI and C_d provides a comprehensive and multi-level approach for evaluating the quality of water. WQI is a general indicator of most of the physico-chemical characteristics of water which makes it suitable for consumption. HPI is an indicator of the toxicity caused by heavy metals whereas C_d estimates its quantification.

3.2.1 Water Quality Index (WQI) Results

The water quality index has been widely utilized to study a variety of physico-chemical characteristics of water samples depicting its overall suitability for human utilization. According to previous researches, there exists a broad range of WQI values for various water samples. It depends upon their specific land use, hydrogeological characteristics and level of human impact. According to the data analysis of this present study, village (S1 and S2) had an average WQI score between 83 and 89 (Table 3). These results are similar to the outcomes of previous studies, which indicated that rural waters are subject to low human impact [21, 22]. At the residential locations (S3 and S4), the WQI range drops to 59 to 67 which corresponds with the levels of moderately polluted urban groundwater. The most drastic drop in WQI values occurs at residential sites S5 and S6 located next to industrial areas, where WQI values were recorded at approximately 31 to 35. This WQI range is identical to what was found in investigations conducted in similar areas [23, 24]. The WQI values of water samples collected from industrial sites (S7 and S8) were the lowest, recorded with a range of 23 to 26. Therefore, the water is a severely polluted water which should never be used for either drinking or domestic use. Collectively stated, the gradual reduction of WQI scores from village to industrial sites represents an intensive land usage contributing to water pollution.

3.2.2 Heavy Metal Pollution Index (HPI) Results

The Heavy Metal Pollution Index (HPI) is a widely accepted method for evaluating cumulative contamination of water with respect to heavy metals by taking 100 as the critical threshold. In this study, low HPI values (~61 to 64) were recorded at the village sites (S1-S2) and are well below the critical threshold value and similar to the HPI values reported for rural low-metal contaminated sites (Table 3). The higher HPI values (~220 to 230) for the residential sites (S3-S4) are consistent with HPI results from urban areas affected by domestic wastewater, plumbing corrosion and

emissions from automobiles. The highest HPI values for residential-industrial sites (S5-S6) were between 350 and 400 and are comparable to the HPI values recorded for mixed urban-industrial locations [25, 26]. A striking increase in HPI was observed for the industrial sites (S7-S8) (greater than 1,400) which is consistent with the literature reports of extensively industrialized regions where untreated or insufficiently treated wastewater is dominant.

3.2.3 Degree of Contamination (C_d) Results

The degree of contamination (C_d) is directly proportional with the land-use pattern in the study area, confirming that human activities have successively impacted the quality of groundwater. Samples of groundwater from the S1 and S2 villages show C_d values that approach zero (Table 3). These values correspond to a low contamination level ($C_d < 1$) indicating that the levels of heavy metals at these locations were nearly all within the permissible limits. This is due to the relatively low levels of industrial activity and dominance of natural geochemical influences (such as water–rock interaction and dissolution of minerals) [27]. Samples of groundwater from the S3 and S4 residential areas indicated moderate levels of contamination (C_d values between 1 and 3). This suggests that anthropogenic metal inputs have begun to appear. The moderate level of C_d values in these areas is attributed to urban-related sources such as domestic sewage infiltrating the ground, corrosion of plumbing materials, surface runoff containing vehicle-related contaminants and leachate produced from landfills [28]. Even though the levels of contamination were moderate, the results confirmed that the groundwater quality is deviating from its natural baseline. Contamination in the vicinity of industrial land uses has significantly risen in residential areas near industry (S5 and S6) with C_d values above 3. This has led to high contamination of heavy metals and strong anthropogenic influences from these sites. As these residential areas are situated close to industrial activities, it is likely that effluent containing metals has migrated to the sub-surface through a combination of unlined surface drains, leaking from industrial and hazardous waste containers and surface water contamination. Water samples collected from the industrial areas (S7 and S8) had the highest reported C_d concentrations which indicates extreme and high contamination levels. The high values of C_d indicate excessive metal stress in that particular site which is beyond the permissible limits. This is due to a lot of industrial wastewater discharges, industrial metal manufacturing or extraction processes, atmosphere deposition from industrial emissions and inadequate waste management. The total concentration of C_d found at these two

sites indicates considerable degradation of groundwater within the vicinity which poses potential health hazards if consumed.

Table 3: The calculated average values of WQI, HPI and C_d along with their standard deviations representing the groundwater quality of every site (S1-S8).

Site	WQI (Mean ± SD)	Remark	HPI (Mean ± SD)	Remark	C_d (Mean ± SD)	Remark
S1	83.00 ± 2.65	Good	60.98 ± 13.40	Good	~0 ± 0.19	Low
S2	89.20 ± 1.06	Good	64.44 ± 8.47	Good	~0 ± 0.10	Low
S3	66.55 ± 4.69	Medium	223.01 ± 74.29	Bad	1.11 ± 0.96	Medium
S4	59.41 ± 1.53	Medium	229.60 ± 45.08	Bad	1.23 ± 0.56	Medium
S5	30.95 ± 2.27	Bad	350.14 ± 41.86	Bad	3.16 ± 0.53	High
S6	34.78 ± 1.72	Bad	403.82 ±125.78	Bad	3.68 ± 1.63	High
S7	25.27 ± 0.48	Very Bad	1390.94 ±179.36	Bad	16.22 ± 2.30	High
S8	23.80 ± 1.36	Very Bad	1448.95 ±341.85	Bad	17.05 ± 4.43	High

The consistent results of the current study with respect to WQI, HPI and C_d for all three indices with previously published results further supports its reliability and reinforces the understanding that industrial activity is the primary cause of water quality degradation.

3.3 Multivariate Statistical Analysis Results

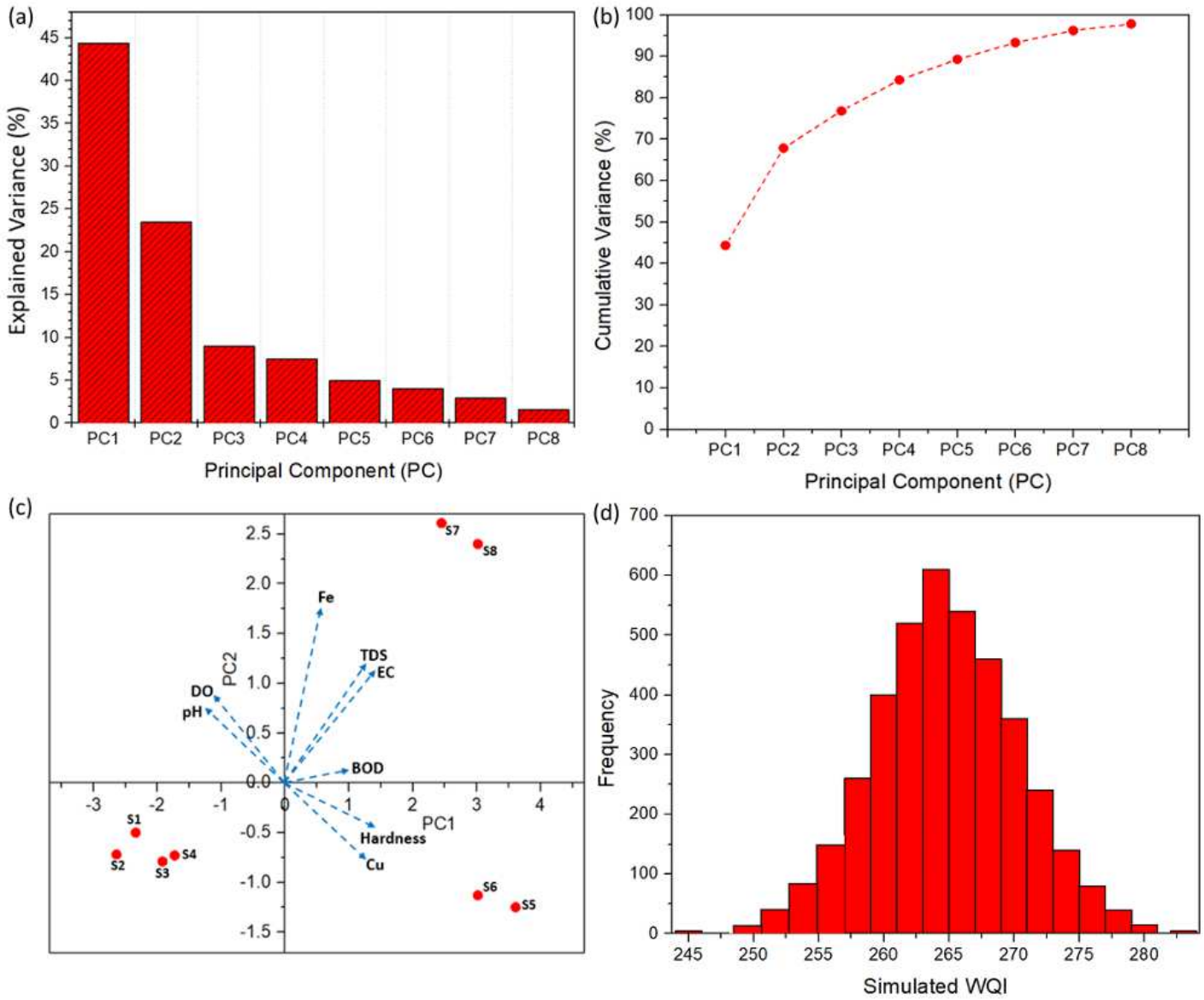
3.*3.1 Results of Principal Component Analysis (PCA)*

Principal Component Analysis (PCA) was performed on the obtained water quality data collected from 8 sampling sites, comprising a total of 24 samples across different seasons (pre-monsoon, monsoon and post-monsoon). Through the use of PCA, the original dataset of physico-chemical parameters was decomposed into 8 orthogonal components (PC1-PC8) representing a linear combination of each original parameter and capturing decreasing proportions of total variance. Mathematically, the interpretation of PCs was limited only to those having eigen values > 1 according to the pattern followed by recent researchers [29, 30]. The scree plot in Fig. 2(a) indicates that the standardized water quality dataset has a wide range of explained variance (%). The two principal components (PCs) with the most variance (PC1 and PC2) together account for a significant portion of the total variance (~68% combined). Each subsequent PC (PC3, PC4 and

PC5) explains progressively less variance and after PC5 each of the subsequent components (PC6, PC7 and PC8) explain only minimal amounts of additional variance, resulting in the creation of a curve in the plot. This denotes that the major hydrochemical processes causing variation in water quality are well summarized within the first few PCs and that the higher order PCs represent predominantly noise or very localized minor effects. Therefore, using the scree plot to reduce dimensionality in this dataset simplifies a large amount of data.

The graph of the cumulative variance in Fig. 2(b) supports the scree plot findings on the levels of total variance. The first two PCs provide ~68% of the total variance and by adding PC3, the cumulative amount of variance rises to approximately ~77%. The data indicates that when four PCs are included, the first four PCs provide approximately 83–85% of the total variance identified. This means they fall above the typical level established as appropriate for the use of environmental data in a meaningful way. Increased addition of PCs after PC4 only results in minimal increases in the percent of total variance explained. Therefore, only using the first four PCs in this case represents the best summary of the main physico-chemical parameters that influence the water sample collection from this site.

The PCA biplot in Fig. 2(c) represents a graphical aspect of the relationship between the sample clusters and the variable loadings from PC1 and PC2 to help us interpret the causes behind the water quality. Specifically, EC and TDS show large positive loadings along PC1, indicating that they are the primary drivers behind the salinity and mineralization gradient. In contrast, hardness and Cu also exhibit a positive loading along PC1, suggesting that they may have been influenced or impacted by human-related activities (anthropogenic) or geologically (geogenic). The PC2 loading is largely formed by the variables Fe, DO, and pH, reflecting natural (geochemical) as well as anthropogenic influences. Also, the locations of sampling sites have formed groups (clusters) that have very different hydrochemical characteristics with sites S5–S8 being clustered towards the positive PC1 area (high ionic strength/pollution) and sites S1–S4 being clustered towards the negative PC1 area (low levels of contamination). In addition, both the length and direction of the vector for each variable demonstrates that the redundant nature of these variables is low while the utility of the chosen variables as descriptors of the variance in water quality is strong.



489 **Fig. 2:** Principal Component Analysis results represented in the form of (a) scree plot with
490 individual PC and explained variance, (b) plot of cumulative variance with respect to the principal
491 components, (c) PCA score plot (PC1 vs. PC2) illustrating sample site distribution indicating
492 relationships between groundwater quality parameters and sampling sites, (d) Monte Carlo
493 histogram depicting the robustness of the dataset.

494 *3.3.2 Monte Carlo Simulation Results*

495 Monte Carlo simulation was employed to assess the uncertainty and variability associated with the
496 Water Quality Index (WQI) derived from groundwater quality parameters [31, 32]. The Monte

Carlo histogram [Fig. 2(d)] obtained is shaped similarly to an inverted bell with most simulated values falling within a very tight band (i.e., ≈ 255 to 275) around a mean WQI value (265). Since many of the simulated values clustered together, this indicates a statistically stable calculable WQI value. That also suggests that if a slight variability in individual water quality parameters, it will not disproportionally alter the overall WQI value. All of these factors further support that the integrated water quality assessment framework is reliable as evidenced by the narrow ranges of simulated WQIs and the lack of high extremes on the Monte Carlo Histogram. Thus, the Monte Carlo analysis was successful in determining the robustness of the generated dataset.

3.3.3 Ward's Linkage Method Results

In order to categorize the groundwater sampling sites, Ward's linkage method was used to perform hierarchical cluster analysis (HCA) according to the similarities obtained in their physico-chemical characteristics [33]. The dendrogram produced from this analysis (Fig. 3) illustrates how the groundwater sampling locations can be grouped into two distinct clusters at a greater degree of linkage. This indicates that these sites are in different hydrochemical regimes as a result of human impacts of land use. The first major cluster of sites (S1, S2, S3, and S4) are located in village/residential areas. The short distances between these sites suggest that the water quality between these four sites is similar, likely due to geogenic processes (like mineral dissolution) having a more significant impact than anthropogenic stress. The close relationship between S3 and S4, followed by their connection to S1 and S2 indicates that the hydrochemical conditions of S3 and S4 are likely influenced primarily by natural weathering processes and relatively modest household contributions of wastewater.

The second cluster group of sites consists of sites S5, S6, S7 and S8, which is found in association with the residential areas near industries (S5, S6) and industrial area (S7 and S8). Due to the higher Euclidean distance, the variability of groundwater chemistry within these respective clusters indicates a greater anthropogenic impact on groundwater. S7 and S8 have similar hydrochemical properties (salinity, nutrients, metals and reduced dissolved oxygen) indicative of groundwater that has been subject to an industrial impact. Thus, clustering closely to one another with a very low linkage distance, indicating they are hydro-chemically similar. The other cluster, S5 and S6 forms a distinct sub-cluster as these sites exhibit hydrochemical conditions between the residential area and industrial effluent, so they are transitional in nature.

The significant separation of these two cluster groups supports the theory that the quality of groundwater has been impacted by progressive shifts from geogenic derived conditions to more anthropogenically dominated conditions. This is found in that areas undergoing manufacturing or any other industrial activity. This clustering pattern was found to be consistent with the findings of PCA results and the significant connection between the quality of groundwater is a function of combined analysis associated with salinity, nutrients, metals and organic pollution. In summary, Ward's Linkage Analysis not only illustrated the existence of spatial heterogeneity in groundwater quality but also demonstrated the significant impact on the similarity among groundwater quality of sampling locations.

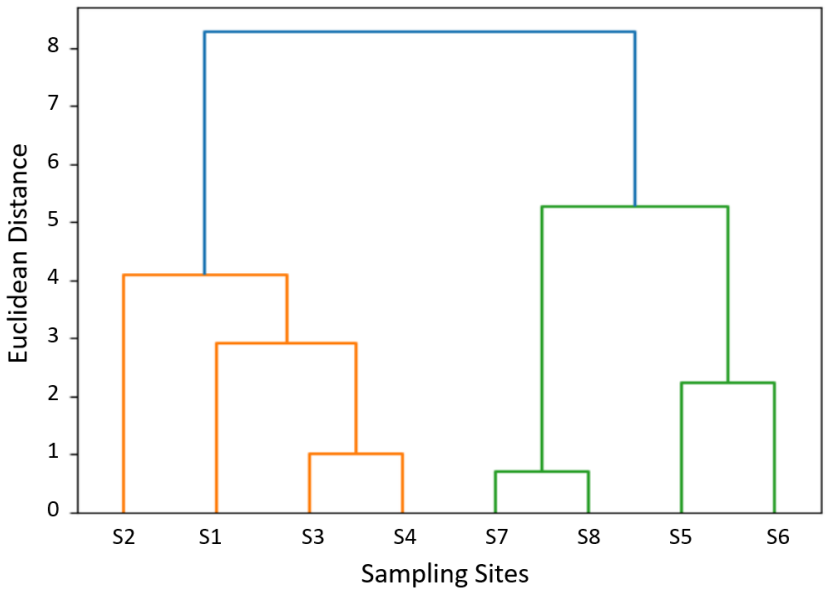


Fig. 3: Ward's dendrogram showing the grouping of groundwater sampling locations into distinct clusters at a greater degree of linkage.

3.3.4 Results of Pearson's Correlation Matrix

The interrelationship among different groundwater parameters is examined using Pearson's correlation analysis [34, 35]. The correlation matrix (Table 4) and heatmap (Fig. 4) reveals a strong relationship between salinity and mineralization related variables. The TDS had an almost perfect correlation (r=0.99) with EC and very strong positive correlations with hardness (r=0.97), alkalinity (r=0.92), TSS (Total Suspended Solids) (r=0.89), and pH (r=0.89). Through these correlations, it is evident that groundwater chemistry is predominantly influenced by ionic

enrichment due to carbonate and silicate mineral dissolution as well as by anthropogenic influences. The close relationship between these parameters supports the concept that there is a common factor controlling them as evidenced by the positive loadings of these parameters on the PC1 of the PCA biplot, which represents the salinity and mineralization gradient. Dissolved oxygen (DO) showed very strong negative correlations with TDS (r=-0.85), EC (r=-0.82), hardness (r=-0.85), alkalinity (r=-0.78) and TSS (r=-0.75). This indicates that as ionic strength and suspended solids increase, the progressive depletion of dissolved oxygen occurs. Therefore, the inverse relationship between DO and salinity is indicative of increased microbial respiration and oxidation of organic matter under highly polluted conditions, particularly along industrial and adjacent zones. The negative orientation of DO in relation to salinity as observed in the PCA biplot provides additional evidence for the role of redox processes in determining the chemical properties of groundwater.

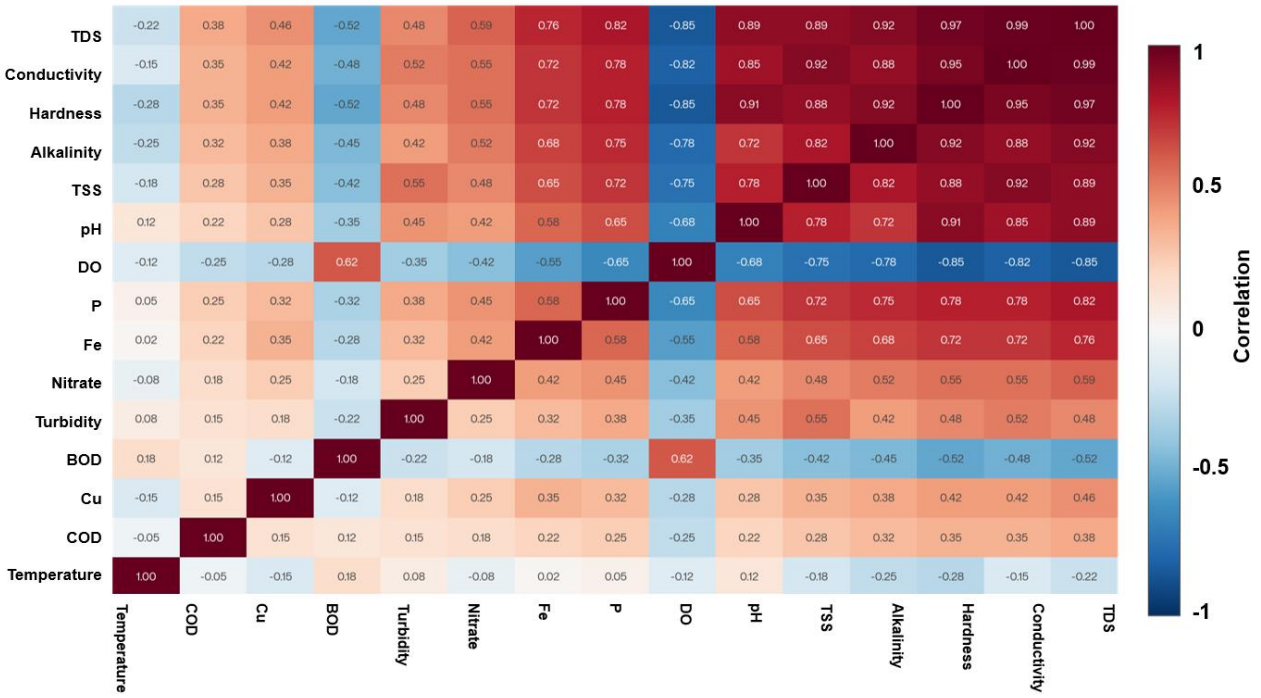
The concentrations of nutrients and metals in groundwater are generally positively correlated with the measures of salinity as they are noticeably affected by human activity or by those processes that naturally change the composition of water. For example, phosphorus is highly correlated with total dissolved solids (r=0.82), hardness (r=0.78), and alkalinity (r=0.75). Thus, it is highly likely that phosphorus is coming from several types of human activity, especially agriculture, human sewage and industrial sewage discharge. Similarly, iron has a positive correlation with TDS (r=0.76) and phosphorus (r=0.58) but has a negative correlation with DO (r=-0.55). The mobilization of iron into the solution is likely to have taken place in reducing conditions and this was seen in the positive relationship of iron with PC2 (factor controlling the release of metals during redox conditions). Nitrate has a moderate correlation with TDS (r=0.59) and EC (r=0.55), which would be consistent with the significant contribution from fertilizer application. Thus, nitrate may have been introduced into the environment both through agricultural and urban discharges or runoff, especially in the rural and urban areas. The weaker correlations of nitrate with the other variables and the negative correlation with DO (r=-0.42) suggest that some of the nitrates may have been reduced through denitrification in areas of low dissolved oxygen. Finally, turbidity is moderately correlated with TSS (r=0.55) and salinity, therefore, it is likely that turbidity is a measure of contaminant transport and was derived from surface contamination.

Organic pollution indicators show that human activities have had an impact on groundwater quality. There is a strong negative correlation between biochemical oxygen demand (BOD) and DO (r=-0.62). The correlation indicates that BOD is a measure of the amount of oxygen consumed by the biological breakdown of organic materials. Chemical oxygen demand (COD) has positive correlations with ionic components; however, those correlations are weaker than those between the salinity components, BOD and DO. There is a moderate correlation between copper (Cu) and both iron and phosphorus, which suggests that Cu is being transported along with iron and phosphorus through the same sources. Therefore, based on the results of this study, it can be inferred that the decrease in groundwater quality Faridabad district is being driven by the combined effects of mineral dissolution, redox processes and increased anthropogenic contributions from land use activities.

586 **Table 4:** Pearson correlation coefficient values (r) of groundwater quality variables.

	TDS	Conductivity	Hardness	Alkalinity	TSS	pH	DO	P	Fe	Nitrate	Turbidity	BOD	Cu	COD	Temperature
TDS	1														
Conductivity	0.99	1													
Hardness	0.97	0.95	1												
Alkalinity	0.92	0.88	0.92	1											
TSS	0.89	0.92	0.88	0.82	1										
pH	0.89	0.85	0.91	0.72	0.78	1									
DO	-0.85	-0.82	-0.85	-0.78	-0.75	-0.68	1								
P	0.82	0.78	0.78	0.75	0.72	0.65	-0.65	1							
Fe	0.76	0.72	0.72	0.68	0.65	0.58	-0.55	0.58	1						
Nitrate	0.59	0.55	0.55	0.52	0.48	0.42	-0.42	0.45	0.42	1					
Turbidity	0.48	0.52	0.48	0.42	0.55	0.45	-0.35	0.38	0.32	0.25	1				
BOD	-0.52	-0.48	-0.52	-0.45	-0.42	-0.35	-0.62	-0.32	-0.28	-0.18	-0.22	1			
Cu	0.46	0.42	0.42	0.38	0.35	0.28	-0.28	0.32	0.35	0.25	0.18	-0.12	1		
COD	0.38	0.35	0.35	0.32	0.28	0.22	-0.25	0.25	0.22	0.18	0.15	0.12	0.15	1	
Temperature	-0.22	-0.15	-0.28	-0.25	-0.18	0.12	-0.12	0.05	0.02	-0.08	0.08	0.18	-0.15	-0.05	1

587



588

589 **Fig. 4:** Pearson's correlation heatmap illustrating the strength and direction of identified
590 parameters among water-quality data in the study area.

591 **4. Alignment of the Research with United Nations' Sustainable Development Goals (SDGs)**

592 The findings of this research are mapped with several United Nation's Sustainable Development
593 Goals (SDGs) (Fig. 5) [36, 37]. The correlations established between the presence of heavy metal
594 contamination in groundwater sources and their potential health effects support the goal of Good
595 Health and Well-Being (SDG 3). The emphasis on water quality and water pollution is evidently
596 mapped with Clean Water and Sanitation (SDG 6). The study also provides scientific information
597 on the impact of industrial activity and land use affecting groundwater quality which supports
598 Sustainable Cities and Communities (SDG 11). It has provided valuable information to assist in
599 planning for protection of groundwater from degradation and planning for improved sustainable
600 development of industry. The monitoring of industrial pollution should be mandatory and must
601 occur in a responsible manner is consistent with Responsible Consumption and Production (SDG
602 12). Additionally, the authors support the concept of sustainable management of groundwater
603 resources that face increasing anthropogenic stress which will aid in achieving Climate Action
604 (SDG 13). In summary, the study provides useful scientific insights that can support the

development of evidence-based decision making and policy formulation that aims to achieve long-term security in respect of water resources as well as the progressive development of urban areas.

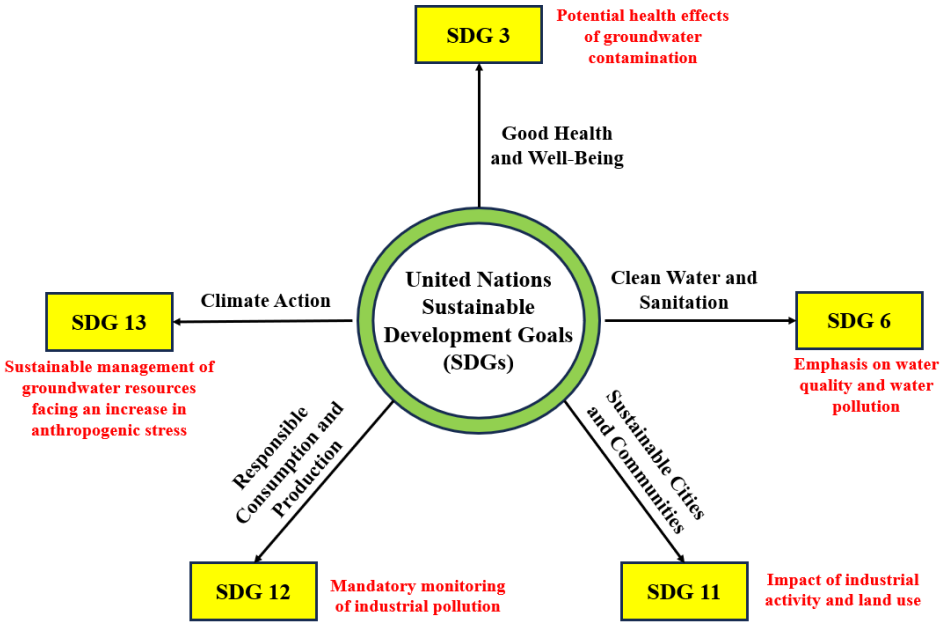


Fig. 5: The outcomes of the research are mapped with the specified United Nation's Sustainable Development Goals (SDGs).

5. Conclusion

The present investigation successfully provides an integrated and a comprehensive approach for groundwater quality status in Faridabad, India by using multivariate statistical analysis. In contrast to the earlier works, this study has incorporated the combined use of pollution indices (WQI, HPI, C_d), together with Pearson's correlation analysis, PCA, HCA via Ward's method and Monte Carlo simulation method to create a highly reliable and consistent assessment of groundwater quality. The groundwater quality was found to be healthy (WQI > 83, HPI < 65, Cd ~ 0) in rural areas; while on the other hand, industrial areas showed severely degraded groundwater quality (WQI < 26, HPI > 1390, Cd > 16). Multivariate analysis helped to clarify salinity-mineralization processes from redox-controlled metal mobilization as demonstrated by highly correlated TDS and EC values (r=0.99) and the negative correlations between dissolved oxygen and salinity measurements (r=−0.85). The PCA also confirmed that salinity-mineralization and redox-controlled metal mobilization were the most influential factors affecting groundwater quality variability, while the

hierarchical clustering of sampling site groupings indicated differences between land use types. Therefore, the consistency between the current study's findings and those from the previous studies strengthen the acceptance of the current assessment as a valid assessment of water quality degradation across all three indices of the study.

References

[1] Agrawal, S. & Sharma, A. Assessing water quality through correlation matrix analysis and water quality index: a comprehensive review. *J. Water Resour. Eng. Manag.* **10** (2023).

[2] Ali, S. *et al.* Groundwater quality assessment using water quality index and principal component analysis in the Achnera block, Agra district, Uttar Pradesh, Northern India. *Sci. Rep.* **14**, 5381 (2024).

[3] Asim, M. & Nageswara Rao, K. Assessment of heavy metal pollution in Yamuna River, Delhi-NCR, using heavy metal pollution index and GIS. *Environ. Monit. Assess.* **193**, 103 (2021).

[4] Biswas, S., Dandapat, B., Alam, A. & Satpati, L. India's achievement towards Sustainable Development Goal 6 (ensure availability and sustainable management of water and sanitation for all) in the 2030 Agenda. *BMC Public Health* **22**, 2142 (2022).

[5] Bora, M. & Goswami, D. C. Water quality assessment in terms of water quality index (WQI): case study of the Kolong River, Assam, India. *Appl. Water Sci.* **7**, 3125–3135 (2017).

[6] Central Ground Water Board. *Aquifer Mapping and Management Plan: National Capital Region (NCR), Haryana* Vol. I (Ministry of Water Resources, Government of India, 2015).

[7] Central Ground Water Board. *Groundwater Chemical Quality Bulletin: Haryana—Pre-monsoon & Post-monsoon 2024/25* (Government of India, 2024–2025).

[8] Chaurasia, A. K. *et al.* Groundwater quality assessment using water quality index (WQI) in parts of Varanasi District, Uttar Pradesh, India. *J. Geol. Soc. India* **92**, 76–82 (2018).

[9] Edet, A. E. & Offiong, O. E. Evaluation of water quality pollution indices for heavy metal contamination monitoring: a case study from Akpabuyo–Odukpani area, Lower Cross River Basin (southeastern Nigeria). *GeoJournal* **57**, 295–304 (2002).

[10] El-Rawy, M. *et al.* Assessment of groundwater quality in arid regions utilizing principal component analysis, GIS, and machine learning techniques. *Mar. Pollut. Bull.* **205**, 116645 (2024).

[11] Gani, A., Singh, M., Pathak, S. & Hussain, A. Groundwater quality index development using the ANN model of Delhi Metropolitan City, India. *Environ. Sci. Pollut. Res.* **32**, 7269–7284 (2025).

[12] Guliani, E. & Jeyaseelan, C. Bioplastics from microbial and agricultural biomass. In *Green Sustainable Process for Chemical and Environmental Engineering and Science*, 395–412 (Elsevier, 2023).

[13] Guliani, E., Banga, S. & Pathak, V. V. Valorization of citrus fruit waste into Schiff base chromogenic chemosensor for detection of Cu^{2+} ions: an experimental study. *ChemistrySelect* **10**, e04768 (2025).

[14] Islam, R. *et al.* Application of Monte Carlo simulation and artificial neural network model to probabilistic health risk assessment in fluoride-endemic areas. *Heliyon* **10**, e40887 (2024).

[15] Jiang, Y., Nan, Z. & Yang, S. Risk assessment of water quality using Monte Carlo simulation and artificial neural network method. *J. Environ. Manag.* **122**, 130–136 (2013).

[16] Kumari, R., Datta, P. S., Rao, M. S., Mukherjee, S. & Azad, C. Anthropogenic perturbations induced groundwater vulnerability to pollution in the industrial Faridabad District, Haryana, India. *Environ. Earth Sci.* **77**, 187 (2018).

[17] Mahapatra, S. S., Sahu, M., Patel, R. K. & Panda, B. N. Prediction of water quality using principal component analysis. *Water Qual. Expo. Health* **4**, 93–104 (2012).

[18] Masood, A., Aslam, M., Pham, Q. B., Khan, W. & Masood, S. Integrating water quality index, GIS and multivariate statistical techniques towards a better understanding of drinking water quality. *Environ. Sci. Pollut. Res.* **29**, 26860–26876 (2022).

[19] Nathalia, D., Suresh, A. & Panda, B. Assessment of groundwater pollution through metal index analysis and its source identification along the Delhi–Haryana border, India. *Arab. J. Geosci.* **14**, 759 (2021).

[20] Nigam, R. & Srivastava, S. Assessment of groundwater quality in parts of the National Capital Region, India. *Environ. Monit. Assess.* **154**, 421–431 (2009).

[21] Onen, P. *et al.* Health risks from intake and contact with toxic metal-contaminated water from Pager River, Uganda. *J. Xenobiot.* **13**, 544–559 (2023).

[22] Pal, T., Dhar, S. B. & Sen, K. Geochemical insights into water quality index and health risk assessment using Monte Carlo simulation in water-scarce regions of India. *Cleaner Water* **100162** (2025).

[23] Patel, P. S., Pandya, D. M. & Shah, M. A systematic and comparative study of water quality index (WQI) for groundwater quality analysis and assessment. *Environ. Sci. Pollut. Res.* **30**, 54303–54323 (2023).

[24] Patil, P. N., Sawant, D. V. & Deshmukh, R. N. Physico-chemical parameters for testing of water—a review. *Int. J. Environ. Sci.* **3**, 1194–1207 (2012).

[25] Prasad, D., Singh, P. K., Mahato, J. K. & Saw, S. Hydrogeochemical characterization of groundwater in fire and non-fire zones of Jharia Coal Field, Eastern India, using water quality index (WQI), hierarchical cluster analysis (HCA), and human health risk. *Arab. J. Geosci.* **15**, 927 (2022).

[26] Rai, S. & Saha, A. Impact of urban sprawl on groundwater quality: a case study of Faridabad city, National Capital Region of Delhi. *Arab. J. Geosci.* **8**, 8039–8045 (2015).

[27] Sarkar, T., Kannaujiya, S., Taloor, A. K., Ray, P. K. C. & Chauhan, P. Integrated study of GRACE data-derived interannual groundwater storage variability over water-stressed Indian regions. *Groundw. Sustain. Dev.* **10**, 100376 (2020).

[28] Saxena, P. *et al.* Heavy metal contamination in groundwater of Delhi NCR: a study on drinking water safety and risk zones. *Total Chem.* **100025** (2026).

[29] Sharafi, S. & Salehi, F. Comprehensive assessment of heavy metal contamination and associated health risks in agricultural soils and groundwater proximal to industrial sites. *Sci. Rep.* **15**, 91453 (2025).

[30] Sharma, K., Raju, N. J., Singh, N. & Sreekesh, S. Heavy metal pollution in groundwater of urban Delhi environs: pollution indices and health risk assessment. *Urban Clim.* **45**, 101233 (2022).

[31] Singh, M., Wadhwa, V., Batra, L., Khyalia, P. & Mor, V. A chemometric and ingestion hazard prediction study of groundwater in proximity to the Bandhwari landfill site, Gurugram, India. *J. Water Health* **22**, 52–63 (2024).

[32] Singh, S. & Hussain, A. Water quality index development for groundwater quality assessment of Greater Noida sub-basin, Uttar Pradesh, India. *Cogent Eng.* **3**, 1177155 (2016).

[33] Su, K., Zhu, D., Liu, R., Peng, Z. & He, J. Evaluation of spatio-temporal variation of water quality in Qingyi River Basin based on multivariate statistical methods and improved WQI. *Environ. Qual. Manag.* **35**, e70121 (2025).

[34] Swain, S., Sahoo, S. & Taloor, A. K. Groundwater quality assessment using geospatial and statistical approaches over Faridabad and Gurgaon districts of National Capital Region, India. *Appl. Water Sci.* **12**, 75 (2022).

[35] Tiwari, A. K., De Maio, M., Singh, P. K. & Mahato, M. K. Evaluation of surface water quality by using GIS and a heavy metal pollution index (HPI) model in a coal mining area, India. *Bull. Environ. Contam. Toxicol.* **95**, 304–310 (2015).

[36] Upreti, M. & Kumar, A. Landscape modeling for urban growth characterization and its impact on ecological infrastructure in Delhi-NCR: an approach to achieve SDGs. *Phys. Chem. Earth* **131**, 103444 (2023).

[37] Vasistha, P. & Ganguly, R. Water quality characterization for two lakes in Panchkula, Haryana, India. *Water Qual. Res. J.* (2024).

[38] Williams, S. A., Megdal, S. B., Zuniga-Teran, A. A., Quanrud, D. M. & Christopherson, G. Equity assessment of groundwater vulnerability and risk in drinking water supplies in arid regions. *Water* **16**, 3520 (2024).

Author Contributions

E.G.: Writing original manuscript, Data curation, Methodology, Investigation, Visualization; V.V.P.: Formal Analysis, Supervision, Validation, Writing-Review & Editing; S.B.: Project Administration, Supervision; A.K.P.: Supervision, Project Administration; A.A.: Writing-Review & Editing.

Data availability

Data analyzed and recorded during the research is included in this article.

Competing Interests Statement

The authors declare no conflict of interest in this publication.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.