Multivariate and Geostatistical Analyses of Groundwater Quality for Acid Rock Drainage at Waste Rock and Tailings Storage Site

Patrick Adadzi[1*], Amy Allwright[1], Francois Fourie[1]

[1] Institute for Groundwater Studies, Faculty of Natural and Agricultural Sciences, University of the Free State Bloemfontein, 205 Nelson Mandela Dr, Park West, Bloemfontein, 9301, South Africa
* Corresponding author's e-mail: 2018425096@ufs4life.ac.za

ABSTRACT

A multi-disciplinary approach is indispensable for adequate acid rock drainage (ARD), mineral leaching impact, and groundwater management. Groundwater is a valuable resource, and it is critical to protect as well as mitigate the effects of pollution such as ARD in the mining environment. Mine waste storage facilities (waste rocks and tailings) are potential ARD sources capable of degrading groundwater reserves. This research investigated and reported the application of a case study of multivariate statistical and spatial variability of selected parameters associated with ARD in groundwater around WRD and TSF at mine sites. Water quality analysis data of seventy water samples from 10 boreholes located at the WRD and TSF mine were utilised in this study. The correlation matrix and principal components analysis was applied to the data set to determine the associated variability in groundwater in relation to ARD. Geostatistical analysis was used to produce contour maps to ARD principal components of the study site, using ordinary kriging of the best fit models. The application of multivariate statistical and geospatial analysis in groundwater quality assessment with coupled soil and groundwater modelling of flow and transport at waste rock dump and tailings storage sites provides an essential tool for exploratory data analysis, and spatial extent determination of the relationship between various data sets significant to acid rock drainage.

Keywords: acid rock drainage, groundwater, correlation matrix, principal component analysis, geostatistical analysis.

INTRODUCTION

Waste rock dumps (WRDs) and tailings storage facilities (TSFs) at a mine site are significant sources of acid rock drainage and potential groundwater contaminants. The central element associated with acid rock drainage in mining activities is sulphide, which can severely deteriorate the groundwater quality. The oxidation of sulphide minerals within the waste rock pile materials ensues in the generation of acidic solutions with a high level of toxic metals, which are potential groundwater contaminants (Sracek et al., 2004; Molson et al., 2005; Akcil and Koldas, 2006; Bao et al., 2020; Gomez-Arias et al., 2021).

The magnitude of groundwater contamination during the operational phase of mining depends on factors such as groundwater-surface water interaction, soil and waste rock erosion intensity, levels of present toxic minerals and trace metals, sulphide mineral levels, available buffering agents, hydrogeological and climatic conditions (Wahsha et al., 2016 Agboola et al., 2020; Vriens et al., 2020; Ram et al., 2021). Groundwater hydrochemistry comprises variable characteristics, and therefore, the spatial dependence of water quality parameters is required to characterise the relationship between the hydrochemical parameters for effective mapping of groundwater attributes of a region (Allwright et al., 2013; Emenike et al., 2018; Jiang et al., 2020).

Statistical applications linked with spatial correlation are a helpful tool and modelling technique for creating groundwater quality monitoring maps. The application of geostatistical methods considers the spatial correlation of the regionalised variabilities by computing and modelling the semivariogram. Geostatistical methods can predict the concentration of regionalised variables, such as groundwater

parameters in non-sampled locations (Karami et al., 2018 Adadzi, 2020; Asghar et al., 2020).

The Pearson correlation method helps analyse the relationships between various physicochemical water quality parameters (Harlicek and Perterson, 1976; Benesty et al., 2009). Pearson's correlation is used to investigate the relationship or association between two variables and shows that the changes in one variable are related to the changes in another. The Pearson correlation coefficient is expressed as:

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2] \cdot [n \sum y^2 - (\sum y)^2]}} \quad (1)$$

where: r – Pearson r correlation coefficient;
 N – number of values in each data set;
 $\sum xy$ – sum of the products of paired scores;
 $\sum x$ – sum of x scores;
 $\sum y$ – sum of y scores;
 $\sum x^2$ – sum of squared x scores;
 $\sum y^2$ – sum of squared y scores.

The correlation coefficient can be described as a small correlation for $-0.10 \leq r \leq 0.29$, a medium correlation for $-0.30 \leq r \leq 0.49$, and a significant correlation for $-0.50 \leq r \leq 1.0$ (Pallant, 2011). Positive and negative indicate the direction of the relationship; positive suggests the increase in the other variable associated with the rise in one variable, and negative correlation indicates the decrease in the other variable related to the increase in one variable. The coefficient of determination (r^2), which explains the change in one variable as explained by the difference in the other variable, was calculated.

Principal component analysis (PCA) minimises information loss and reduces large data sets (King and Jackson, 1999; Jolliffe and Cadima, 2016). This study used principal component analysis to obtain the correlation between groundwater quality parameters. The first principal component explains much of the total variance in a data set, while the other members then explain the remainder of the variance. In PCA, only the components with an eigenvalue >1 were often selected and then subjected to the varimax rotation before being used for interpretation (Patil et al., 2008).

The PCA method consists of five main steps:
1. Original data matrix:

$$x = (x_{ij})_{np} = \begin{bmatrix} x_1 & \cdots & x_{1p} \\ \vdots & \vdots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \quad (2)$$

where: x_{ij} – in the matrix is the original measurement data;
 n – the monitoring station;
 p – each water quality parameter.

2. Standardisation of the original data with the Z-score standardisation formula to remove the impact of dimensionality (Belkhiri and Narany, 2015) as:

$$x_{ij}^* = \frac{(x_{ij} - \bar{x}_j)}{s_{j'}} \quad (3)$$

where: x_{ij}^* – the standard variable;
 x_j – the mean value for the jth indicator;
 s_j – the standard deviation of the *jth* indicator.

3. Calculation of the correlation coefficient matrix R using standardised data to establish the correlation between the indicators (Equation 4) (Chen et al., 2015):

$$R = (r_{ij})_{pp} = \frac{1}{n-1} \sum_{t=1}^{n} x_{ti}^* x_{tj} \quad (4)$$

$$(i, j = 1,2, \ldots, p)$$

4. Calculation of the eigenvalues and eigenvectors of the correlation coefficient matrix R to establish the number of principal components.

The eigenvalues of the correlation coefficient matrix R are expressed by λ_i ($i = 1, 2 \cdots n$) and the eigenvectors are $u_i (u_i = u_{i1}, u_{i2}, \cdots u_{in})$, ($i = 1, 2 \cdots n$). The λ value corresponds to the variance of the principal component. Also, the variance value is positively correlated with the contribution of the principal components. The cumulated contribution rate of the first m principal components should be more than 80%, which means: $\sum_{j=1}^{m} \lambda j / \sum_{j=1}^{n} \lambda j \geq 0.80$ (Uddameri et al., 2014). The principal component is expressed by Equation (5).

$$F_i = u_{i1} x_1^* + u_{i2} x_2^* + \cdots + u_{in} x_n^*$$

$$(i = 1,2, \ldots, n) \quad (5)$$

where: x_i^* is the standardised indicator variable $x_i^* = (x_i - \bar{x}_i)/s_i$.

5. Weighing and summing the resulting principal components to obtain a comprehensive score function, as shown in Equation (6):

$$F_i = u_{i1}x_1^* + u_{i2}x_2^* + \cdots + u_{in}x_n^*$$

$$(i = 1,2,\ldots,n)$$

(6)

Applying geostatistical methods, the surfaces are connected if the sampling points are at different locations. Deterministic methods use mathematical functions to interpolate based on surrounding measurements directly. Geostatistical methods rely on both statistical and mathematical functions, including autocorrelation. The geostatistical tool ArcGIS (ESRI, 2019) was used to predict the distribution of groundwater quality within the study catchments. Kriging interpolation techniques and a semivariogram modelling approach were used to analyse the spatial distribution of parameters in groundwater associated with acid rock drainage. Semivariogram model types were considered for the selected groundwater quality parameters: exponential, linear, circular, spherical, stable, Gaussian and quadratic. Each parameter was analysed across various semivariograms. The most appropriate model of the semivariograms was selected by examining the spatial distribution of the dataset along with geostatistical features such as standard deviation, percentage error and skewness. A Kriging interpolation technique was used to interpolate readings without actual data. A spatial analysis of groundwater quality was conducted to identify variations in groundwater quality across vulnerable areas and sites. Kriging interpolation and semivariograms were the main geostatistical methods used in spatial analysis.

Geostatistical applications through either deterministic or stochastic methods have been applied in many studies to model and generate a spatial mapping of groundwater quality dynamics and patterns of major elements and trace metals. Amongst several methodologies, the kriging method has been used successfully to estimate the pollution patterns in groundwater, soil and the environment for designing proper management, remediation and preventive strategies (Amini et al., 2019; ESRI, 2019; Hosseini et al., 2019; Jha et al., 2020; Sideri et al., 2020; Tziritis et al., 2021).

The geostatistical simulation approach by kriging has been used to assess the uncertainty in the distribution of significant groundwater contaminants. The application of the ordinary kriging method in spatial modelling and mapping of groundwater quality distribution produces smoothing effects that emanate to accuracy in the interpretation of produced contaminant distribution maps (Naudet et al., 2004; ; Rabah et al., 2011; Safarbeiranvnd et al., 2018; Rata et al., 2020; Varmaghani et al., 2021).
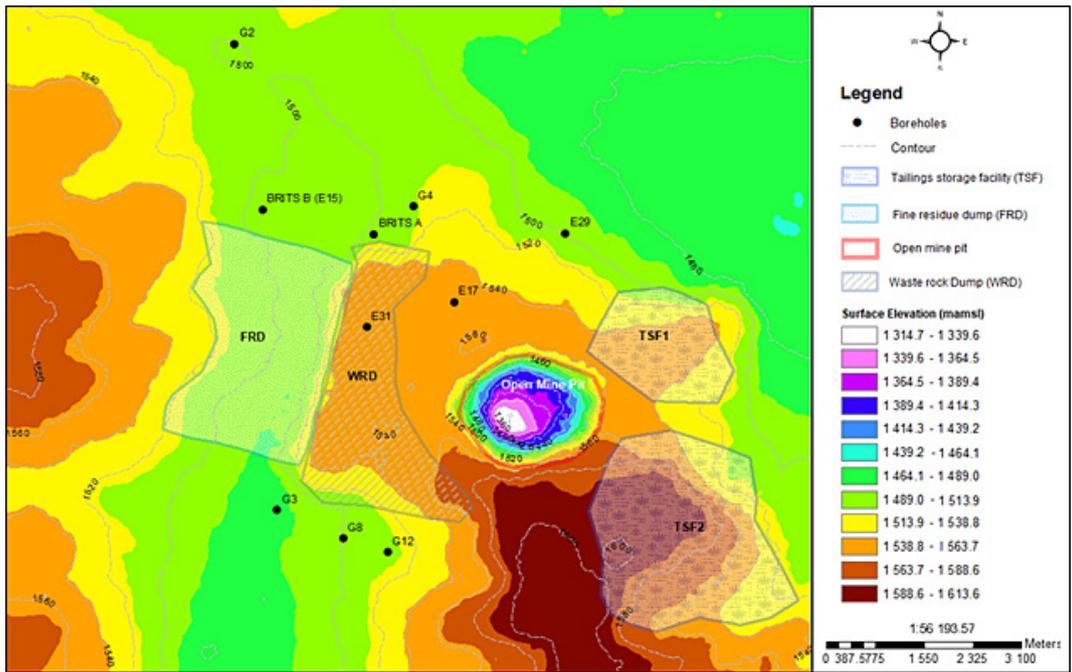
The application of various geostatistical methods in groundwater, soil and environmental studies to locate the potential contamination and uncontaminated regions has, in recent decades, gained extensive comprehension for evaluating uncertainties (Theocharopoulos et al., 2004; Scholz and Schnabel, 2006; Bodrud-Doza et al., 2016).

This study applied an integrated multivariate statistical and geostatistical analysis to investigate and determine the spatial models for major groundwater parameters associated with acid rock drainage at the waste rock dump and tailings storage site at a mine. Despite the availability of monitoring data at mine sites (WRD and TSF), the data alone is insufficient for assessment and the distribution, extent and magnitude of ARD impact at specific locations within the mine. The multivariate spatial distribution assessment ensures that attention is focused on specific locations of concern, resulting in time and cost components of environmental impact assessment and mitigation measures. The spatial distribution maps are a useful management tool for groundwater protection, remediation strategies and optimisation of groundwater-monitoring design at the mining site.

STUDY AREA

The study area is at a mean altitude of 1500 metres above mean sea level (mamsl) and has a temperate Highveld climate with warm to hot summers as well as mild to cold winters. The mean annual precipitation for the area at a 50% percentile is 286.9 mm, with mean yearly evaporation of 2600–2800 mm (Midgley et al., 1997). The mine is situated on a topographic high, mainly sloping east and northeast of the mining area (Figure 1).

The WRD and a fine residue dam (FRD) are located to the west, and the TSFs are situated to the east of the open mine pit and cover a surface area of approximately 7.2 km². Groundwater monitoring boreholes are located near the WRD, FRD and TSF to facilitate the monitoring of water chemistry and groundwater levels. The WRD is composed of rock materials that overlay the ore body and were displaced during mining. These rock materials mainly consist of dolomite from the excavated upper, middle, and lower
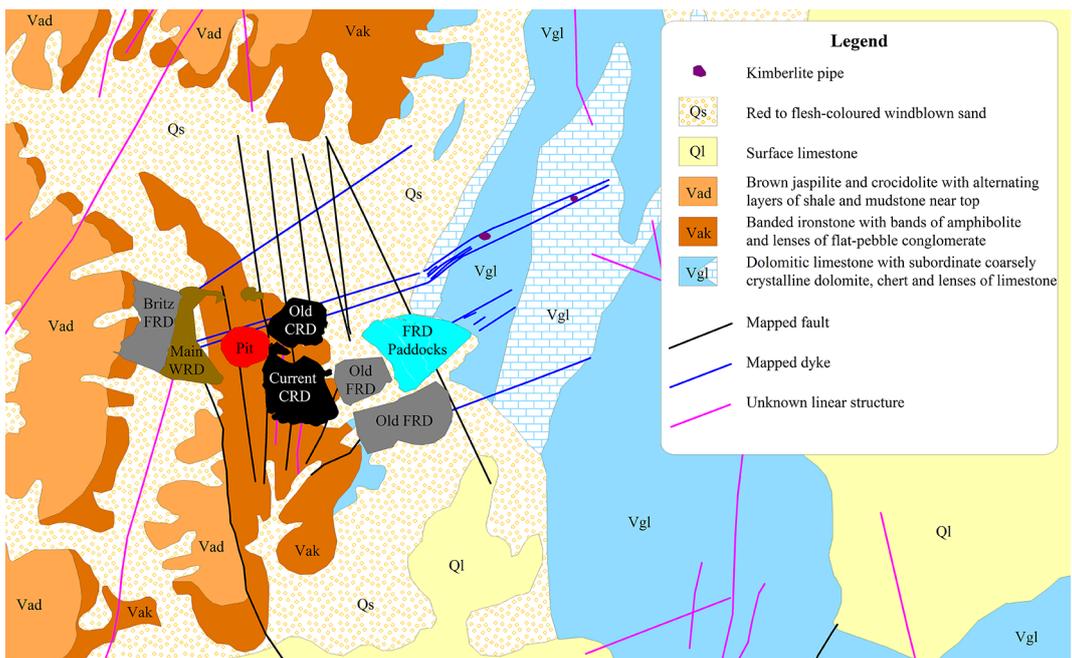
Figure 1. Topographical map of the study area showing the location of the WRD, FRD, TSF and monitoring boreholes relative to the open mine pit location

dolomite members, as well as secondary dolomitisation. The WRD also contains remnants of the volcanoclastic kimberlite (Ekkerd and Stiefenhofer, 2003) discarded after removing economic minerals and fragments of Karoo and Transvaal Supergroup sedimentary rocks, xenoliths of basalt, mudstone, dolerite, and sandstone. The FRD, located adjacent to the WRD,

is delivered in a slurry form to the deposit with about 49% moisture content. The return water from the slurry flows to the deposit pool and is then decanted in a controlled manner.

The regional geology of the study mine consists of 17.9 ha, 118 ± 2.8 Ma (Smith et al., 1985), and a Group-2 kimberlite (Smith, 1983a; Fraser and Hawkesworth, 1992), as presented in Figure 2.

Figure 2. Geological map of the study area

The geology of the mine at the open pit and sampling level 348 m below ground was described by Clement (1982) as a pipe through the Griqualand West suite, which consists of thick Proterozoic sedimentary rock suites, dolomites, banded iron formations, and shale. Large fragments of Karoo, lava, and dolerite deposits were observed in the tube and subsequently eroded. The study mine is directly underlain by dolomitic limestone with interbedded chert of the Ghaap Plateau Formation and Campbell Group, forming part of the Griqualand West Sequence of rocks.
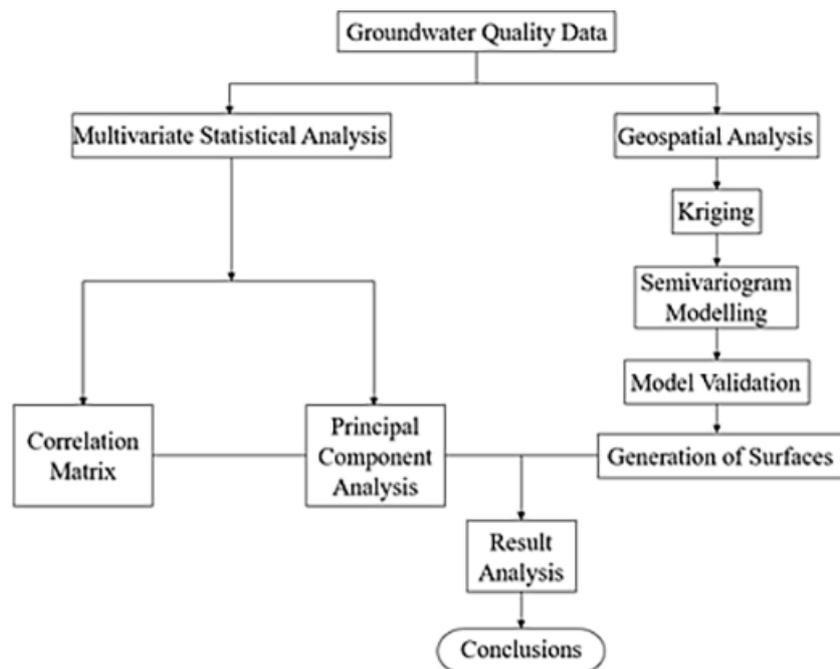
MATERIALS AND METHODS

In this study, correlation matrix, principal component analysis and geospatial analysis were applied on the obtained groundwater quality data from 10 boreholes (70 water samples in total), taken quarterly from December 2018 to December 2020. The multivariate statistical analyses were performed using Origin-Pro statistical tools. The multivariate statistical analysis covers 14 groundwater parameters (EC, pH, Ca, Mg, Na, K, Cl, SO_4, HCO_3, NO_3, Si, T. Hardness, TDS, and Turbidity), while the geospatial analysis focused on the most critical parameters, likely influencing ARD in the groundwater (SO_4, pH, EC, Ca, Mg, Na, Cl,

and HCO_3). Geostatistical modelling used a numerical model optimal for further analysis of each groundwater quality parameter. Semivariogram selection was made by visually reviewing the map and examining the distribution of groundwater throughout the study area, considering statistical characteristics such as standard deviation, percentage error, and skewness. The Kriging interpolation method analysed the selected groundwater quality parameters for all semivariograms. This procedure was repeated for all groundwater quality parameters to determine the optimal semivariogram for all data. The multivariate and spatial distribution assessment of the groundwater parameters associated with ARD ensures that attention is focused on specific locations of concern, resulting in time and cost components of environmental impact assessment and the application of mitigation and management measures.

The methodology was applied to the case study location, as illustrated in the schematic diagram of Figure 3.

The method was applied to the case study because of the distribution of sampling points and the number of samples, which require classification into several homogenous groups to reference the underlying significant hydrogeochemical anomalies related to ARD. The method involves subjecting the groundwater quality data to the following analyses:

Figure 3. Illustration of methodology for multivariate and geostatistical analysis

1. Multivariate statistical analysis techniques made up of:
- Correlation matrix; and
- Principal component analysis.

2. Geostatistical/geospatial application using:
- Ordinary Kriging;
- Semivariogram models;
- Model validation; and
- Generation of surface maps.

The methodology is based on a combination of the correlation matrix, principal component analysis, variogram modelling and kriging. In applying this method, the aim is the interpretation of specific anomalous values in the data, which could indicate the presence of acid rock drainage in groundwater. In my opinion, this can be done best if the interpretation of the anomalies is based upon an explanation of the chemical phenomena described by the data. In ideal situations, the groundwater parameters can be related directly to acid rock drainage influence. In practical applications, looking for the chemical interpretation of the parameters is generally helpful. However, it may not be possible to explain precisely at a particular stage of acid rock drainage development. The multivariate statistical analyses were performed using OriginPro software.

Correlation matrices are usually created using the Pearson method. On the basis of standardised data, R-mode factor analysis (FA) and principal component analysis (PCA) show the interaction of variables. In general, it seeks to simplify the complex and diverse relationships between a set of observed variables by revealing the common and unobserved factors associated with clearly unrelated variables.

The eigenvalues of the correlation/ covariance matrix represent the division of total variation that causes each principal component in groundwater relating to ARD property. The magnitude of the relationships between variables is used as indicators and as a covariance matrix standardised by setting all the variances equal to one. The correlation matrix calculation is a primary tool to identify the dominant correlating parameters/variables before the data is further subjected to principal component analysis.

PCA was mainly used for strongly correlated variables. PCA does not work well to reduce data when the relationships between variables are weak. PCA recommends a correlation coefficient greater than 0.3. Principal component variables were linear combinations of the original variables that produced the extracted eigenvectors. Scree plots and biplots were created to interpret the relationships between observations, while the biplots represented both loadings and scores. The correlation matrix and principal component analysis were used to identify the groundwater quality components associated with acid rock drainage linked to specific sampling locations. The clustering of groundwater parameters was then inferred to relate the potential for acid rock drainage influence on the groundwater quality.

RESULTS AND DISCUSSION

Groundwater composition

The groundwater composition for the study area is based on the existing quarterly monitoring data obtained from the case study mine from December 2018 to June 2020. The results of the statistical representation of groundwater composition from the boreholes at the study location grouped according to the areas of the boreholes relative to the WRD, FRD and TSF are presented in Table 1.

Correlation matrix

This study used a Pearson correlation matrix to identify the relationships among the individual groundwater quality parameters/variables. The analysed correlation matrix for the groundwater samples is presented in Table 2.

The results indicate that the strongest correlated variables are SAR and Na (0.981), Mg and HCO_3 (0.961), and SO_4 and Na (0.906). Other groundwater quality variables also showing relatively strong correlations are SO_4 and Cl (0.816), Ca and Cl (0.805), EC and Cl (0.838), TH and HCO_3 (0.884), TH and EC (0.817), SO_4 and Cl (0.78), SO_4 and NO_3 (0.705), SO_4 and EC (0.772), SO_4 and SAR (0.86), EC and Ca (0.722), and pH and Mg (0.714). A moderate correlation was observed between TH and Cl (0.576), SAR and Cl (0.683), Na and NO_3 (0.636), SAR and NO_3 (0.635), EC and HCO_3 (0.568), pH and HCO_3 (0.678), Ca and SO_4 (0.597), K and Mg (0.504), EC and Mg (0.612), K and EC (0.573), K and TH (0.509), EC and SAR (0.581), and pH and TH (0.623). However, weaker correlations were

Table 1. Results of groundwater composition

Location	Borehole ID	Parameter	pH	Na$^+$ (mg/L)	Ca^{2+} (mg/L)	Mg^{2+} (mg/L)	K$^+$ (mg/L)	Cl$^-$ (mg/L)	SO$_4^{2-}$ (mg/L)	HCO$_3^-$ (mg/L)	NO$_3^-$ (mg/L)	Si (mg/L)	TDS (mg/L)	$^\#$TH (mg/L)	EC (mg/L)
North of WRD/FRD	Brits A	Ave	7.3	14.1	26.9	13.9	3.1	31.4	7.0	109.6	22.0	6.2	190.3	124.7	284.9
		Stdev	0.4	1.2	1.5	1.2	0.3	1.6	2.7	22.0	7.4	0.7	13.4	7.4	40.3
	Brits B	Ave	7.1	10.7	13.0	7.0	1.4	15.3	13.0	46.2	15.9	11.7	131.6	61.4	156.0
		Stdev	0.5	1.2	1.8	0.6	0.3	3.5	3.8	7.3	2.8	0.9	14.7	6.4	63.6
	G2	Ave	7.7	8.6	24.5	15.5	2.2	17.7	7.7	139.2	4.2	4.6	162.4	125.0	246.4
		Stdev	0.6	1.3	7.7	4.5	0.2	2.0	2.3	49.7	2.2	0.9	37.2	37.3	37.9
	G4	Ave	7.1	9.7	16.0	12.1	2.5	17.4	1.8	108.9	2.0	5.3	129.7	89.9	246.3
		Stdev	0.6	0.7	2.4	1.1	0.1	1.0	1.3	15.1	0.0	0.9	9.0	10.1	140.8
South of WRD/FRD	G3	Ave	8.0	14.5	26.4	23.0	3.8	34.3	26.1	127.7	18.2	7.7	231.1	160.4	326.3
		Stdev	0.3	3.7	4.4	4.0	0.4	5.0	12.8	21.9	3.6	1.5	35.5	25.9	76.5
	G8	Ave	8.4	17.3	21.7	63.6	6.6	27.0	11.1	370.0	3.3	10.0	362.7	315.9	520.6
		Stdev	0.2	0.8	2.9	1.7	0.1	1.4	3.6	25.4	0.5	0.5	13.3	6.3	177.2
	G12	Ave	8.4	17.3	25.8	57.5	1.8	26.6	17.7	328.5	2.0	8.3	335.6	301.1	506.0
		Stdev	0.2	0.7	4.2	1.7	0.2	1.6	4.3	22.3	0.0	0.5	13.5	7.6	117.2
East of WRD/FRD	E17	Ave	7.8	9.3	70.0	41.2	2.8	49.2	37.2	302.1	2.0	7.2	381.7	344.4	597.7
		Stdev	0.4	0.6	4.8	1.6	0.2	2.9	13.1	11.2	0.0	0.7	21.6	18.5	76.3
	E29*	Ave	7.9	49.3	87.7	37.3	4.7	87.8	133.8	239.4	35.5	6.8	575.8	373.6	879.4
		Stdev	0.2	3.9	3.9	1.6	0.4	1.8	8.3	22.0	5.1	0.4	22.3	13.9	56.7
In WRD	E31	Ave	7.6	97.6	42.8	26.9	4.7	73.4	179.0	147.3	32.2	6.3	547.4	217.6	822.9
		Stdev	0.3	6.2	1.4	1.0	0.3	2.8	30.1	18.8	7.1	0.5	26.7	6.8	50.2

Note: *averages and standard deviations calculated from the last five sets of analytical data; $^\#$TH – total hardness.

Table 2. Pearson's correlation matrix (parameters with correlation coefficients ≥ 0.70 are highlighted in bold)

Location	Borehole ID	Parameter	F (mg/L)	Al (mg/L)	Fe (mg/L)	Mn (mg/L)
North of WRD/ FRD	Brits A	Ave	-0.5	-0.002	-0.009	0.01
		Stdev	5.9E-17	0.007	0	0.02
	Brits B	Ave	-0.5	-0.005	-0.009	0.02
		Stdev	5.9E-17	0	0	0.04
	G2	Ave	-0.5	-0.002	-0.009	0.05
		Stdev	5.9E-17	0.007	0	0.06
	G4	Ave	-0.5	-0.005	0.007	0.1
		Stdev	5.9E-17	0	0.04	0.05
South of WRD/ FRD	G3	Ave	-0.5	-0.005	-0.009	0.08
		Stdev	5.9E-17	0	0	0.1
	G8	Ave	-0.5	-0.005	-0.009	0.001
		Stdev	5.9E-17	0	0	0.002
	G12	Ave	0.8	-0.01	-0.01	0.002
		Stdev	0.04	0	0	0.01
East of WRD/ FRD	E17	Ave	-0.5	-0.005	-0.01	0.05
		Stdev	5.9E-17	0	0	0.05
	E29*	Ave	-0.5	-0.005	-0.01	0.1
		Stdev	0	0	0	0.1
In WRD	E31	Ave	-0.5	-0.002	-0.01	0.01
		Stdev	5.9E-17	0.01	0	0.02

observed amongst relatively few groundwater quality variables at the study site.

The presence of Na^+ controls the salinity load in groundwater, and then by Cl^- and HCO_3^-. The strong positive correlation of Mg and HCO_3 (0.961) indicates carbonate buffering resulting in the weak correlation of SO_4 with pH (0.016). However, the strong positive correlation of SO_4 with Na (0.906) suggests acid rock drainage influence under saline conditions (neutral mine drainage). The strong positive correlation of Na^+ and Cl^- (0.78) and the weak positive correlation between Ca^{2+} and Mg^{2+} (0.278) indicate these parameters have mixed sources of origin. Therefore, the acid rock drainage influence in the groundwater is likely due to the leachates from the WRDs and TSFs at the study location.
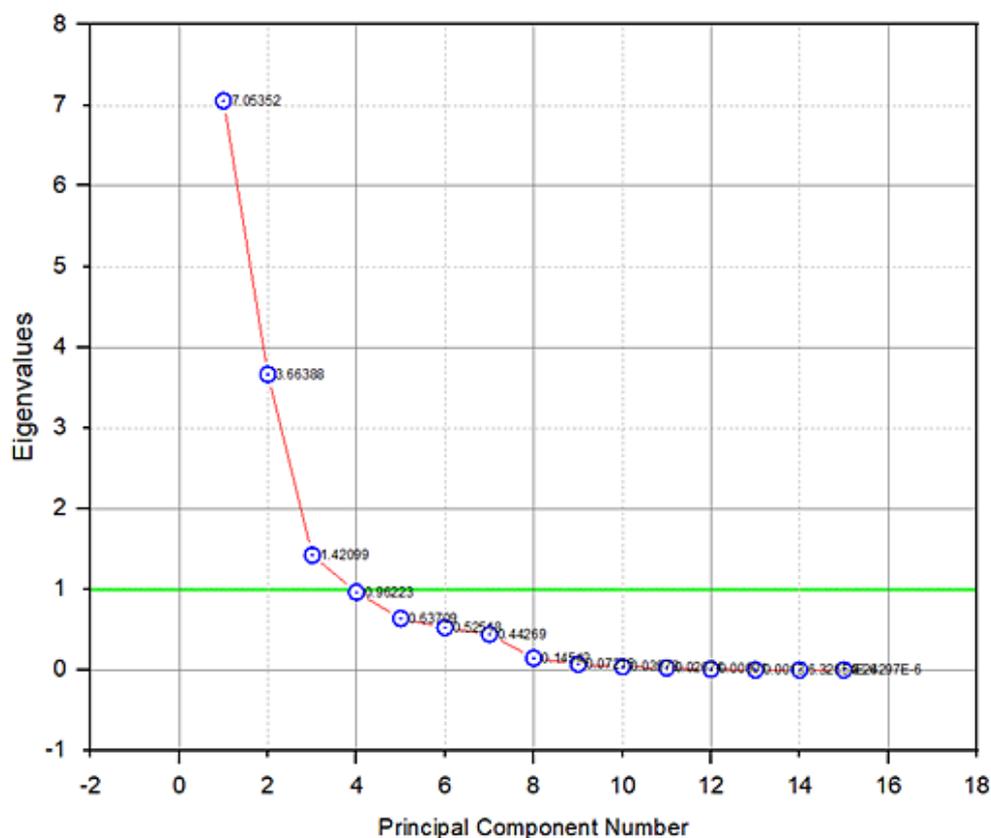
Principal component analysis

Principal component analysis was performed to identify the dominant groundwater component, which is inferred to acid rock drainage characteristics. Kaiser-Meyer-Olkin (KMO) suggests a value of 0.86, showing a perfect sample fit, and p <0.001, indicating a high sphericity significance
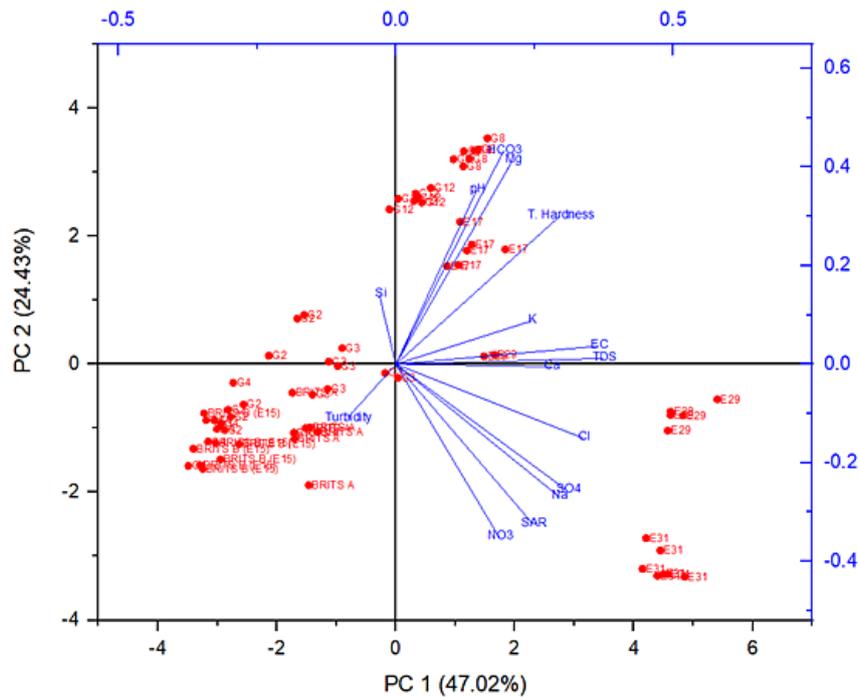
Table 3. Loadings of experimental variables on principal components, Eigenvalues, and variances for groundwater quality data of study site

Parameters	Principal components			
	PC1	PC2	PC3	PC4
Cl	0.338	-0.149	0.158	0.090
NO₃	0.186	-0.351	-0.194	0.243
HCO₃	0.193	0.431	0.062	-0.055
SO₄	0.308	-0.256	-0.054	0.052
Ca	0.278	-0.005	0.308	0.522
Mg	0.210	0.413	-0.079	-0.099
Na	0.294	-0.268	-0.092	-0.282
K	0.244	0.088	-0.051	-0.404
EC	0.365	0.037	-0.009	0.082
pH	0.145	0.352	0.031	-0.232
T. Hardness	0.295	0.300	0.098	0.191
Si	-0.029	0.140	-0.632	0.435
SAR	0.247	-0.324	-0.143	-0.329
Turbidity	-0.088	-0.110	0.623	0.067
TDS	0.374	0.011	-0.003	0.061
Eigen value	7.054	3.664	1.421	0.962
Proportion %	47.02	24.43	9.47	6.41
Cumulative %	47.02	71.45	80.92	87.34

Note: values > 0.7 represent strong positive factor loading.

Figure 4. PCA scree plot of the Eigenvalues

Figure 5. Biplot of PCA loadings scores for the dataset of groundwater quality parameters and sampling boreholes

by the Bartletts test. This means a dominant relationship between variables for the application of PCA. Four principal components (PCs) of the groundwater quality dataset were extracted based on eigenvalues greater than 0.86, representing 87.34% of the total variance for the study area (Table 3). The scree plot of the PCs presented in Figure 4 was used to identify the number of PCs retained to understand the structure of groundwater quality parameters. The calculated component loadings, cumulative percentage, and percentages of variance explained by each component are presented in Figure 4. The results indicate that 47.02% of the total variance represents the first principal component (PC). The PC1, PC2, PC3, and PC4 for groundwater quality data were 47.02%, 24.43%, 9.47% and 6.41%, respectively.

The Biplot of PCA loadings scores for the dataset of groundwater quality parameters and sampling boreholes is presented in Figure 5. The primary principal component (PC1), explaining 47.02% of the data variance, is positively loaded with TDS, Cl, SO_4, EC and Na, which is mainly distributed in boreholes E29 and E31 (Figure 5).

PC1 shows a weak positive loading of SO_4 (0.308) and Na (0.295), indicating a moderate influence of secondary salts and leachate sulphide-bearing minerals at these locations. The PC2 shows that 24.43% of the total variance is

positively loaded with HCO_3, Mg, pH and TH, significantly distributed in boreholes E17, G8, and G12. The medium positive loading of Mg (0.413) and HCO_3 (0.431) indicates lithological influence and carbonate buffering in the groundwater at these locations. PC3 (9.47%) is positively loaded with Ca (E29) and Tbd (G2, G3, G4, Brits A and Brits B (E15), while PC4 (4.41%) is positively loaded with Ca (E29), Si (G2 and G3) and NO_3 (E29 and E31). PC3 and PC4 suggest significant carbonate weathering and silicate weathering, respectively, at these locations.

The PCA method uses an overall score to represent the critical groundwater quality parameters associated with each site within the study area. The study identified the groundwater quality components associated with acid rock drainage linked to specific sampling locations. The clustering of groundwater parameters indicates that the potential for acid rock drainage influence on groundwater quality is significant at boreholes E29 (located near TSF) and E31 (located in WRD). This is conspicuous in the biplot (Figure 5), showing a strong positive correlation and close clustering of SO_4 and Na within PC1 (47.02%). Sulphate and halite are strong indicators of anthropogenic influences of ARD in the groundwater at these locations. The biplot also indicated the evidence of carbonate buffering reactions at boreholes G8,

G12 and E17 because of the close correlation and clustering of HCO$_3$ and Mg with a medium positive correlation.

Geostatistical analysis

Geostatistical analysis was carried out with ArcGIS 10.8 to perform ordinary kriging for predictive surface output with no transformation. Predictive performance was assessed by cross-validation, examining the accuracy of the selected fitting models and parameters, as shown in Table 4.
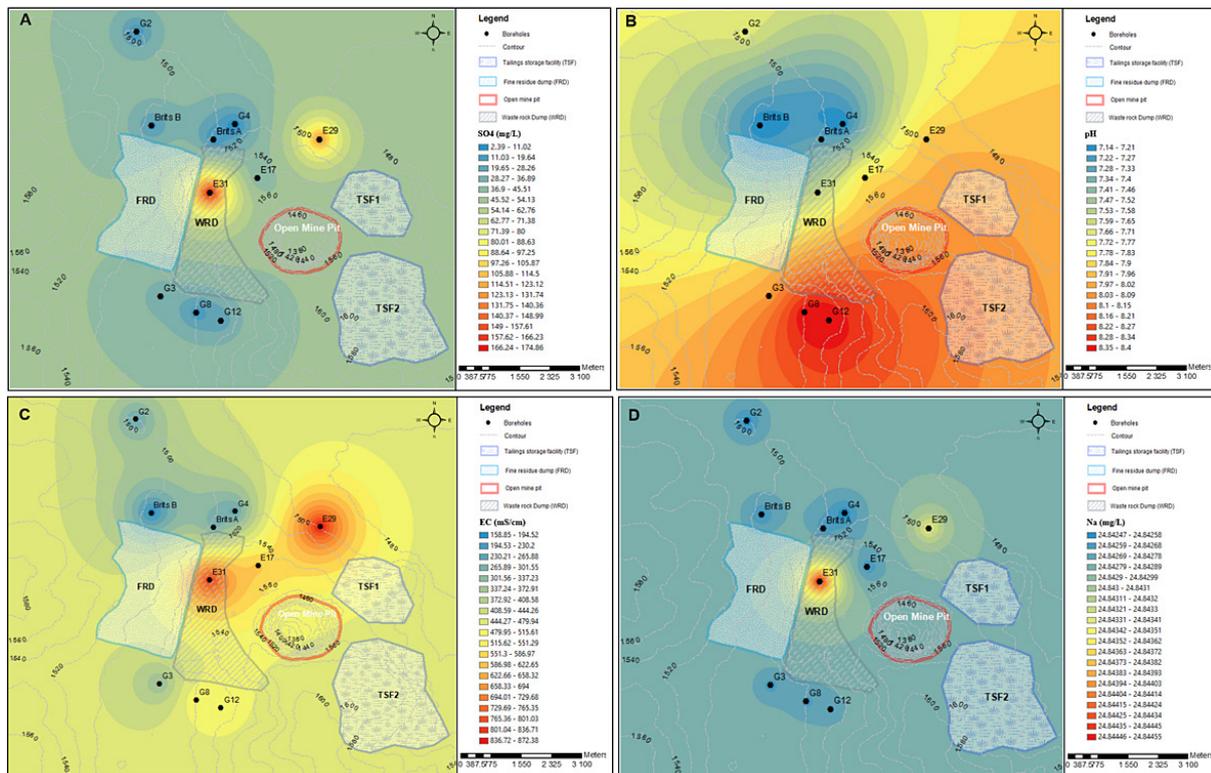
All the groundwater quality parameters were evaluated by cross-validation as follows:

- The Standardised Mean Error range closest to zero,
- The Root Mean Square Standardised Error (RMSSE) range closest to 1, and
- Standard errors are close to the root-mean-square prediction errors.

The results in Table 4 indicate the best fit semivariogram models based on the defined error analysis schemes selected for kriging interpolation to generate surface maps for each groundwater parameter associated with ARD at the study location. Semivariogram models (circular, spherical, tetraspherical, pentaspherical, exponential,

Table 4. Cross-validation calculated statistics for the groundwater quality parameters

Parameters	Best fit models	Prediction Errors				
		Mean	Root mean square	Average standard	Mean standardised	Root mean square standardised
SO$_4$	Exponential	-2.300	60.48	55.58	-0.016	1.100
pH	Rational quadratic	-0.002	0.253	0.308	0.039	0.919
EC	Stable	0.596	22.43	24.38	0.023	0.911
Na	Exponential	-0.439	33.41	27.64	-0.001	1.202
Cl	Exponential	-0.958	23.48	22.56	-0.017	0.963
HCO3	Circular	9.064	84.54	89.15	0.068	1.085
Ca	Rational quadratic	-1.426	16.98	18.49	-0.037	0.881
Mg	Circular	1.546	12.57	15.01	0.068	0.974

Figure 6. Spatial distribution map of (A) SO4, (B) pH, (C) EC, and (D) Na in the study area
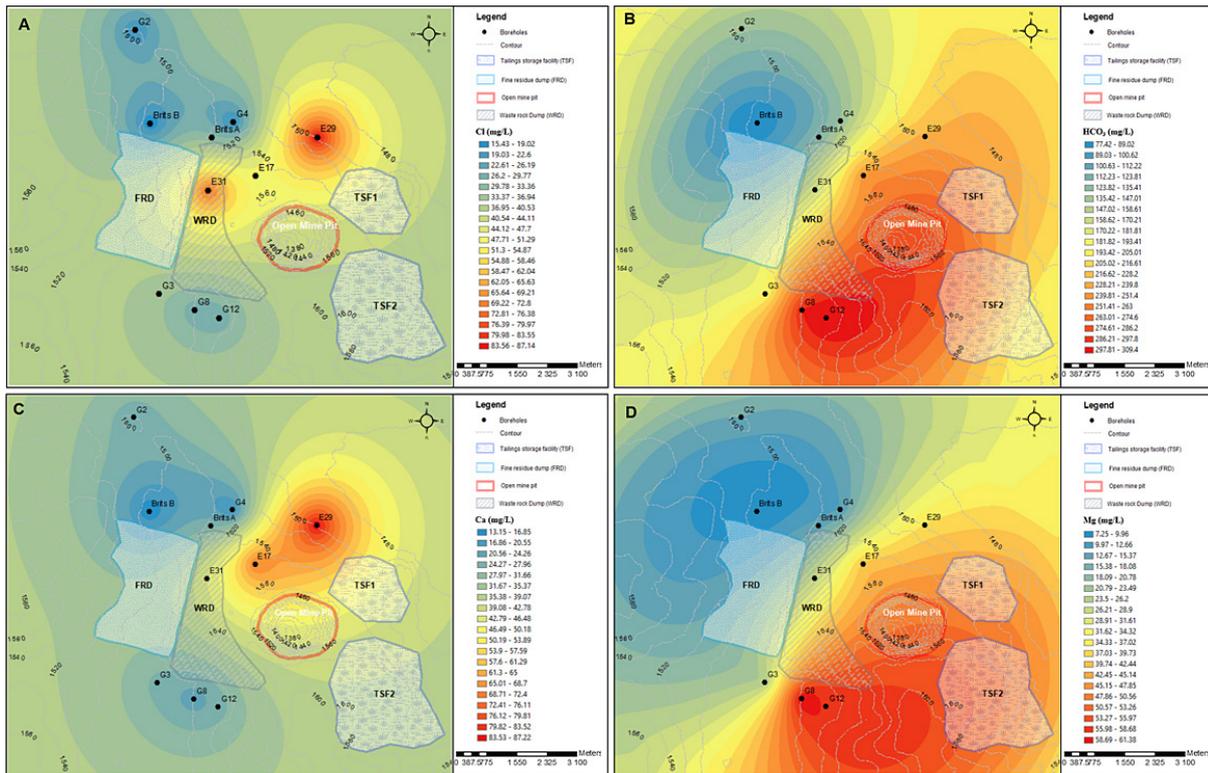
Figure 7. Spatial distribution map of (A) Cl, (B) HCO3, (C) Ca, and (D) Mg in the study area

Table 5. Summary of spatial analysis of groundwater quality parameters associated with acid rock drainage at the study location

Parameter	Spatial distribution	Highest Locations	Lowest Locations	Analysis of results
SO_4	Figure 6a	E29 and E31	G2, G4, G8, G12, Brits A and Brits B(E15)	E29 and E31 are potential acid rock drainage zone with the mixing of groundwater with leachates of sulphide-bearing minerals from WRD and TFS at low pH compared to the other locations with low SO_4 concentrations
pH	Figure 6b	G2 and G8	Brits A and Brits B(E15)	pH generally between 7 to 8.4 at the study location, with near neutral pH values observed at G2 and G8. Saline pH condition observed near WRD and towards TSF
EC	Figure 6c	E29 and E31	Brits B(E15), G2 and G4	Sulphate and EC are usually closely correlated, with the degree of correlation enhanced with increasing ARD contamination of groundwater (Heikkinen et al., 2002). E29 and E31 elevated EC likely related to acid rock drainage influences
Na	Figure 6d	E29 and E31	G2, G3, G4, G8, G12, Brits A and Brits B(E15)	Na likely anthropogenic salt load component associated with acid rock drainage and other pollution sources
Cl	Figure 7a	E29 and E31	G2, G4 and Brits B(E15)	Cl likely anthropogenic salt component associated with acid rock drainage and other pollution sources
HCO_3	Figure 7b	G2 and G8	Brits B(E15)	Carbonated neutralises pyrites and the dissolution of calcite/dolomite in the carbonate bedrock releases calcium and magnesium (Webb and Sasowsky, 1994), contributing to alkalinity. The alkalinity buffers the groundwater against rapid pH changes
Ca	Figure 7c	E17 and E29	G4 and Brits B(E15)	Relatively higher Ca release by carbonate weathering at E17 and E29 compared to other locations
Mg	Figure 7d	G8 and G12	G4 and Brits B(E15)	Relatively higher Mg release by carbonate weathering and contributes to hardness at G8 and G12 compared to other locations

Gaussian rational quadratic, hole effect, K Bessel, J Bessel and stable) were tested for each parameter. The best fit semivariogram model for each parameter-based cross-validation is presented in Table.

By utilising the best model determined from the cross-validation process, the predicted groundwater quality parameters were used to produce the spatial distribution maps for SO_4, pH, EC, Na, HCO_3, Ca and Mg for the study, as presented in Figures 6 and 7.

The discussion on the spatial distribution of the groundwater quality parameters of the study area associated with acid rock drainage impact is discussed in Table 5. The generated map for each parameter indicates that the influence of ARD has a high magnitude at boreholes E31 and E29, located in the WRD and close to the TSF. This is because of the generally elevated concentrations of SO_4, Na, Cl and EC, which are characteristics of potential ARD influence in the groundwater. Sulphate and EC are usually closely correlated, with the degree of correlation enhanced with increasing ARD contamination of groundwater (Heikkinen et al., 2002). Carbonated neutralisation of pyrites and the dissolution of calcite/dolomite were also observed in boreholes G2 and G8, contributing to alkalinity.

CONCLUSIONS

The application of multivariate and geostatistical analysis for identifying ARD in groundwater at mine sites provides a practical, cost-effective framework capable of pinpointing the areas of critical ARD impacts for implementing mitigation and remediation strategies. The multivariate statistical methods of the correlation matrix (CM) and principal component analysis (PCA) were integrated with geostatistical spatial analysis (GPA) to determine the potential localities of acid rock drainage influence at the study site. On the basis of the correlation matrix analysis, the strongest correlated groundwater quality parameters at the study location were Na, Cl, SO_4, and HCO_3, indicating the influence of secondary salts with leachate sulphide-bearing minerals accompanied by carbonate weathering. PCA shows that the primary principal component in the groundwater accounting for 47.02% of the groundwater parameter variance is SO_4, Cl, Na, EC and TDS, which are mainly distributed at boreholes E29 and E31. The application of geostatistical analysis resulted

in the generation of spatial distribution maps of the groundwater quality parameters for the study area. The spatial distribution maps indicate that the maximum SO_4, Na, Cl and EC concentrations occurred at boreholes E29 and E31. The integrated CM-PCA-GPA showed elevated sulphate concentrations at circumneutral pH (Neutral Mine Drainage) at E29 near a tailings storage facility and E31 in a waste rock dump. The integrated CM-PCA-GSA benefits the ability to identify the patterns within multivariate groundwater hydrochemical data that are often difficult to characterise for acid mine drainage. The integrated CM-PCA-GSA application validates the proposed integrated hydrogeochemical analysis method by identifying the correlations and spatial distribution of groundwater parameters according to ARD influence on groundwater hydrogeochemistry. Multi-disciplinary research on acid rock drainage and mineral leaching from tailings and waste rock facilities at mining sites facilitates diverse prevention, mitigation, and management options.

REFERENCES

1. Adadzi P.C. 2020. Geostatistics and Spatial Analysis of Groundwater Hydrochemistry near Leliefontein in the Northern Cape, South Africa. Journal of Ecological Engineering, 21(8), 243–260. https://doi.org/10.12911/22998993/126877

2. Agboola O., Damilola E., Fayomi B.O.S.I., Sadiku E.R., Popoola P., Moropeng L., Yahaya A., Mamudu, O.A. 2020. A review on the impact of mining operation: Monitoring, assessment and management, Results in Engineering, 8, 100181. https://doi.org/10.1016/j.rineng.2020.100181

3. Akcil A., Koldas S. 2006. Acid Mine Drainage (AMD): causes, treatment and case studies. Journal of cleaner production, 14(12–13), 1139–1145.

4. Allwright A., Witthueser K., Cobbing J., Mallory S., Sawunyama T. 2013. Development of a Groundwater Resource Assessment Methodology for South Africa: Towards a Holistic Approach. WRC Report No. 2048/1/13 ISBN 978-104312-0458-8

5. Amini M.A., Torkan G., Eslamian S., Zareian M.J., Adamowski J.F. 2019. Analysis of deterministic and geostatistical interpolation techniques for mapping meteorological variables at large watershed scales. Acta Geophysica, 67(1), 191–203.

6. Asghar F., Sun Z., Chen G., Zhou Y., Li G., Liu H., Zhao K. 2020. Geochemical characteristics and uranium neutral leaching through a CO2+ O2 system—An example from Uranium Ore of the ELZPA Ore Deposit in Pakistan. Metals, 10(12), 1616.

7. Bao Z., Bain J., Holland S.P., Wilson D., MacKenzie P., Ptacek C.J., Blowes D.W. 2020. Faro Waste Rock Project: Characterizing geochemical heterogeneity in sulfide-and carbonate-rich waste rock. Applied Geochemistry, 121, 104691.

8. Belkhiri L., Narany T.S. 2015. Using multivariate statistical analysis, geostatistical techniques and structural equation modeling to identify spatial variability of groundwater quality. Water Resources Management, 29(6), 2073–2089.

9. Benesty J., Chen J., Huang Y., Cohen I. 2009. Pearson correlation coefficient. In Noise reduction in speech processing (1–4). Springer, Berlin, Heidelberg.

10. Bodrud-Doza M., Islam S.D.U., Rume T., Quraishi S.B., Rahman M.S., Bhuiyan M.A.H. 2020. Groundwater quality and human health risk assessment for safe and sustainable water supply of Dhaka City dwellers in Bangladesh. Groundwater for Sustainable Development, 10, 100374.

11. Chen F., Shen F., Song, J. 2015. Robust adaptive beamforming using low-complexity correlation coefficient calculation algorithms. Electronics Letters, 51(6), 443–445.

12. Clement C.R. 1982. A comparative geological study of some major kimberlite pipes in the Northern Cape and Orange Free State.

13. Ekkerd J., Stiefenhofer J., Field M., Lawless P. 2003. The geology of Finsch mine, Northern Cape Province, South Africa. In International Kimberlite Conference: Extended Abstracts, 8.

14. Emenike P.C., Nnaji C.C., Tenebe I.T. 2018. Assessment of geospatial and hydrochemical interactions of groundwater quality, southwestern Nigeria. Environmental monitoring and assessment, 190(7), 440.

15. ESRI. 2019. Introduction to modelling spatial processes using geostatistical analysis.

16. Fraser K.J., Hawkesworth C.J. 1992. The petrogenesis of group 2 ultrapotassic kimberlites from Finsch Mine, South Africa. Lithos, 28(3–6), 327–345.

17. Gómez-Arias A., Yesares L., Caraballo M.A., Maleke M., Vermeulen D., Nieto J.M., Van Heerden E., Castillo J. 2021. Environmental and geochemical characterization of alkaline mine wastes from Phalaborwa (Palabora) Complex, South Africa. Journal of Geochemical Exploration, 224, 106757.

18. Heikkinen P.M., Korkka-Niemi K., Lahti M., Salonen V.P. 2002. Groundwater and surface water contamination in the area of the Hitura nickel mine, Western Finnland. Environmental Geology 42, 313–329.

19. Hosseini E., Gholami R., Hajivand F. 2019. Geostatistical modeling and spatial distribution analysis of porosity and permeability in the Shurijeh-B reservoir of Khangiran gas field in Iran. J Petrol Explor Prod Technol, 9, 1051–1073.

20. Jha M.K., Shekhar A., Jenifer M.A. 2020. Assessing groundwater quality for drinking water supply using hybrid fuzzy-GIS-based water quality index. Water Research, 179, 115867.

21. Jiang Y., Gui H., Yu H., Wang M., Fang H., Wang C., Chen C., Zhang Y., Huang Y. 2020. Hydrochemical characteristics and water quality evaluation of rivers in different regions of cities: a case study of Suzhou City in Northern Anhui Province, China. Water, 12(4), 950.

22. Karami S., Madani H., Katibeh H., Fatehi Marj A. 2018. Assessment and modeling of the groundwater hydrogeochemical quality parameters via geostatistical approaches. Applied water science, 8(1), 1–13.

23. Midgley D.C., Pitman, W.V., Middleton B.J. 1997. Surface water resources of South Africa 1990, Volume VI. Water Research Commission, Pretoria, Report 298/6.1/94.

24. Molson J., Aubertin M., Bussière B. 2012. Reactive transport modelling of acid mine drainage within discretely fractured porous media: Plume evolution from a surface source zone. Environmental Modelling & Software, 38, 259–270.

25. Naudet V., Revil A., Rizzo E., Bottero J.Y., Bégassat P. 2004. Groundwater redox conditions and conductivity in a contaminant plume from geoelectrical investigations. Hydrology and Earth System Sciences, 8(1), 8–22.

26. Patil V.H., Singh S.N., Mishra S., Donavan D.T. 2008. Efficient theory development and factor retention criteria: Abandon the 'eigenvalue greater than one'criterion. Journal of Business Research, 61(2), 162–170.

27. Rabah F.K., Ghabayen S.M., Salha A.A. 2011. Effect of GIS interpolation techniques on the accuracy of the spatial representation of groundwater monitoring data in Gaza Strip. Journal of Environmental Science and Technology, 4(6), 579–589.

28. Ram A., Tiwari S.K., Pandey, H.K., Chaurasia A.K., Singh S., Singh, Y.V. 2021. Groundwater quality assessment using water quality index (WQI) under GIS framework. Applied Water Science, 11(2), 1–20.

29. Rata M., Douaoui A., Larid M., Douaik, A. 2020. Comparison of geostatistical interpolation methods to map annual rainfall in the Chéliff watershed, Algeria. Theoretical and Applied Climatology, 141(3), 1009–1024.

30. Safarbeiranvnd M., Amanipoor H., Battaleb-Looie S. 2018. Quality Evaluation of Groundwater Resources using Geostatistical Methods (Case Study: Central Lorestan Plain, Iran). Water Resour Manage 32, 3611–3628. https://doi.org/10.1007/s11269-018-2009-2

31. Scholz R.W., Schnabel U. 2006. Decision making under uncertainty in case of soil remediation. Journal of Environmental Management, 80(2), 132–147.

32. Sideri D., Roumpos C., Pavloudakis F., Paraskevis N., Modis, K. 2020. Multivariate Geostatistical Modeling of Lower Calorific Value in Multi-Seam Coal Deposits. Applied Sciences, 10(18), 6208.

33. Smith C.B. 1983a. Pb, Sr and Nd isotopic evidence for sources of southern African Cretaceous kimberlites. Nature, 304(5921), 51–54.

34. Smith C.B., Allsop H.L., Kramers J.D., Hutchinson G., Roddick, J.C. 1985. Emplacement ages of Jurassic-Cretaceous South African kimberlites by the Rb-Sr method on phlogopite and whole-rock samples. Transactions of the Geological Society of South Africa, 88(2), 249–266.

35. Sracek O., Choquette M., Gélinas P., Lefebvre R., Nicholson, R.V. 2004. Geochemical characterization of acid mine drainage from a waste rock pile, Mine Doyon, Quebec, Canada. Journal of contaminant hydrology, 69(1–2), 45–71.

36. Theocharopoulos S.P., Mitsios I.K., Arvanitoyannis I. 2004. Traceabilty of environmental soil measurements. TrAC Trends in Analytical Chemistry, 23(3), 237–251.

37. Tziritis E., Pisinaras V., Panagopoulos A., Arampatzis G. 2021. RIVA: a new proposed method for assessing intrinsic groundwater vulnerability. Environmental Science and Pollution Research, 28(6), 7043–7067.

38. Uddameri V., Honnungar V., Hernandez E.A. 2014. Assessment of groundwater water quality in central and southern Gulf Coast aquifer, TX using principal component analysis. Environmental earth sciences, 71(6), 2653–2671.

39. Varmaghani A., Eichinger W.E., Prueger J.H. 2021. A meteorological-based crop coefficient model for estimation of daily evapotranspiration. Hydrological Processes, 35, 14025.

40. Vriens B., Plante B., Seigneur N., Jamieson H. 2020. Mine Waste Rock: Insights for Sustainable Hydrogeochemical Management. Minerals, 10(9), 728.

41. Wahsha M., Nadimi-Goki M., Bini C. 2016. Land contamination by toxic elements in abandoned mine areas in Italy. J Soils Sediments, 16, 1300–1305.

42. Webb J.A., Sasowsky D. 1994, The interaction of acid mine drainage with a carbonate terrane: evidence from the Obey River, north-central Tennessee. Journal of Hydrology, 161, 327–346.