Application of Multivariate Statistical Analysis in Mine Water Hydrogeochemical Studies of the Outcropped Upper Carboniferous, Ruhr Area, Germany

Tuan Quang Tran[1, 2], Andre Banning[1], Stefan Wohnlich[1]

[1]*Hydrogeology Department, Faculty of Geosciences, Ruhr University Bochum, Universitätsstr. 150, 44801 Bochum, Germany, quang.tran@rub.de or tranquangtuan@humg.edu.vn*

[2]*Hydrogeology Department, Faculty of Geosciences and Geoengineering, Hanoi University of Mining and Geology, Hanoi, Vietnam*

Abstract

This study aimed to identify the processes controlling the geochemical characteristics of adit mine waters using correlation analysis and multivariate statistical techniques. The Hierarchical cluster analysis (HCA) classified water samples into 5 geochemically distinct clusters of similar characteristics. HCO_3^- and Ca^{2+} were the dominant ions in the mine water (clusters C1, C2, C3, and C4), indicating the processes of carbonate dissolution, while Na^+ was the most abundant of the cations in cluster C5, suggesting processes of rock weathering. Samples of C2, C4, and C5 were of Ca-Na-HCO_3-SO_4, Ca-Mg-Na-HCO_3-SO_4-Cl, and Na-Ca-HCO_3-SO_4 types, respectively, while those from C1 and C3 were of Ca-Mg-HCO_3-SO_4 type. A total of three principal component analysis (PCA) components accounted for 82.95% of the total variance. PCA 1 accounted for a variance of 58.18%, characterized by high positive loadings for HCO_3^-, SO_4^{2-}, Ca^{2+}, Mg^{2+}, Na^+, K^+, TDS, and EC, suggesting carbonate dissolution, rock weathering, and water salinity. The other components may be related to more local and geological effects. This study demonstrated the usefulness of multivariate statistical analysis in hydrogeochemistry.

Keywords: Multivariate Statistics, Cluster Analysis, Principal Component Analysis, Mine Water

Introduction

The last remaining German active hard coal mine was closed in December 2018. However, after mining, the mine water quality is still one of the concern issues nowadays. The study area is a part of the Ruhr coalfield, where the coal-bearing Upper Carboniferous strata crop out directly on the surface of the Ruhr river valley or underneath very thin overlying sediments of Quaternary age (Frankenhoff *et al*. 2017; Zieger *et al*. 2018), with about 50 km in East-West length and 15 km North-South width (Drozdzewski *et al*. 2008). It covers an area of about 754 km^2 (Fig. 1), spreading over the cities Hattingen, Sprockhövel, Witten, Wetter and Schwerte, and including parts of the cities Essen, Bochum, Dortmund, Holzwickede and Herdecke. The topography of the study area is generally shallow hills and undulating terrain, the elevation ranges between 20 and 333 m above sea level. The highest elevations are in Sprockhövel (about 330 m in Holthausen). The annual average precipitation in the area is 891 mm (DWD 2019). The geology is characterized by Upper Carboniferous cyclothems: sandstone, silt- , mud-, and claystone layers, interspersed with hard coal seams. The layers contain clay minerals and mica (kaolinite, chlorite, biotite), feldspar (albite, anorthite), siderite, dolomite, calcite, quartz, and pyrite (Strehlau 1990; Wisotzky 2017).

It is not easy to identify the processes controlling the geochemical characteristics of mine water, especially in abandoned coal mines, through complex hydrogeochemical data. To protect water resources, it is important to understand processes controlling geochemical variability in such post-mining systems.

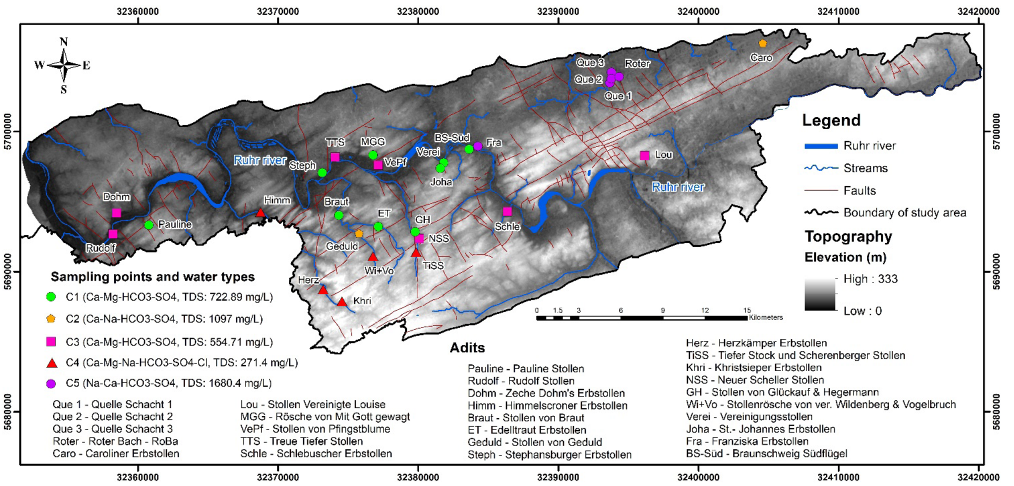

Fig. 1 *Map of the study area, including sampling points and spatial distribution of the mine water types.*

For decades, besides hydrogeochemical and isotopic methods, multivariate analysis techniques have also been used widely in mine water studies (e.g. Li *et al.* 2018). These methods are performed as useful tools to reduce and organize a large hydrogeochemical data set into a small number of factors, grouping samples with similar characteristics (Reghunath *et al.* 2002). Multivariate techniques also were used to study the environmental impacts of medieval mining (Horák and Hejcman 2016) or have been extensively applied to groundwater quality data (Khelif and Boudoukha 2018). Moreover, multivariate analysis has been applied to hydrogeochemical data by many researchers to analyze hydrogeochemical parameters to classify groundwater of varied geochemical features and to identify primary processes controlling groundwater chemical components (e.g. Qian *et al.* 2016).

The purpose of this study was to characterize the chemistry of mine water and to develop hypotheses about processes controlling mine water composition and the origin of the chemical constituents present in mine water of abandoned coal mines in the outcropped Upper Carboniferous of the Ruhr area, Germany. The present study also contributed and demonstrated the usefulness of multivariate statistical analysis in mine water assessment and established a better understanding of hydrogeochemical characteristics of mine water in the study area.

Sample collection and methods

Sampling and analysis

Water samples were collected from the 28 different drainage sites of abandoned coal mines in the study area from January to March 2018. The sampling locations are given in Fig. 1. Physicochemical parameters of water (pH, EC) were measured in situ using portable meters (WTW, Weilheim, Germany). These probes were calibrated using standard solutions, and bottles were washed and rinsed thoroughly with sample water. Samples for cation analysis were acidified using concentrated HNO_3. Samples were stored in the refrigerator at 4 °C until completed analyses. HCO_3^- concentrations were measured using HCl titration. Major anions (SO_4^{2-}, Cl^-, NO_3^-, F^-) were measured using the DIONEX Ion Chromatography System model ICS - 1000, with an AD20 Absorbance Detector system. Major cations (Ca^{2+}, Mg^{2+}, Na^+, and K^+) were analyzed by CD25 Conductivity Detector, IP 20 Isocratic Pump DIONEX.

Multivariate statistical techniques

The correlation matrix, Hierarchical cluster analysis (HCA), and Principal component analysis (PCA) methods were used to evaluate the physicochemical parameters of water samples. Spearman correlation was used to evaluate correlations between different variables, using the statistical software SPSS 23, combined with the XL-Stat software to check and compare results. These methods were applied to a subgroup of the dataset to evaluate their usefulness for classifying water samples and identifying relevant geochemical processes. The dataset consisted of 28 hydrogeochemical analyses.

The Q-mode HCA technique was performed on the water chemistry data to classify samples into distinct hydrogeochemical groups using values for 12 physicochemical parameters (HCO_3^-, SO_4^{2-}, Cl^-, NO_3^-, F^-, Ca^{2+}, Mg^{2+}, Na^+, K^+, pH, EC, and TDS). The Ward's linkage method with Euclidean distance was used for the measurement of similarity between the water quality variables. The results were represented in a dendrogram. PCA was used to identify major factors governing mine water geochemistry. The varimax rotated factor analyses were carried out for 12 parameters and factor loading was calculated using eigenvalue greater than 1 and sorted by the results having values greater than 0.3. Most mine water physicochemical parameters in this study are non-normally distributed, the Spearman method and Kaiser-Meyer-Olkin (KMO) test were used to indicate the proportion of common variance.

Results and discussion

General hydrogeochemical characteristics

The pH values ranged from 6.39 to 7.65, indicating circum-neutral mine water conditions. EC values were in the range of 202 to 1713 µS cm^{-1} and TDS values ranged between 157 and 1806 mg L^{-1}, which indicates that mine water of several adits is highly mineralized, implying fresh and brackish waters. A Durov diagram (Durov 1948) was plotted for 28 water samples to examine the hydrogeochemical types of mine waters (Fig. 2).

This diagram shows the chemical composition of water in the study area that was formed by the main types: Ca-Mg-HCO_3-SO_4, followed by Na-Ca-HCO_3-SO_4, Ca-Mg-Na-HCO_3-SO_4, and Na-Ca-Mg-HCO_3-SO_4 water types. The average ion concentrations followed the order: HCO_3^- > SO_4^{2-} > Cl^- > NO_3^- > F^- and Ca^{2+} > Mg^{2+} > Na^+ > K^+ (% meq L^{-1}). The mine water samples showed the dominance of weak acids (HCO_3^-) over strong acids (SO_4^{2-} + Cl^-), and alkaline earths (Ca^{2+} + Mg^{2+}) over the alkalies (Na^+ + K^+).

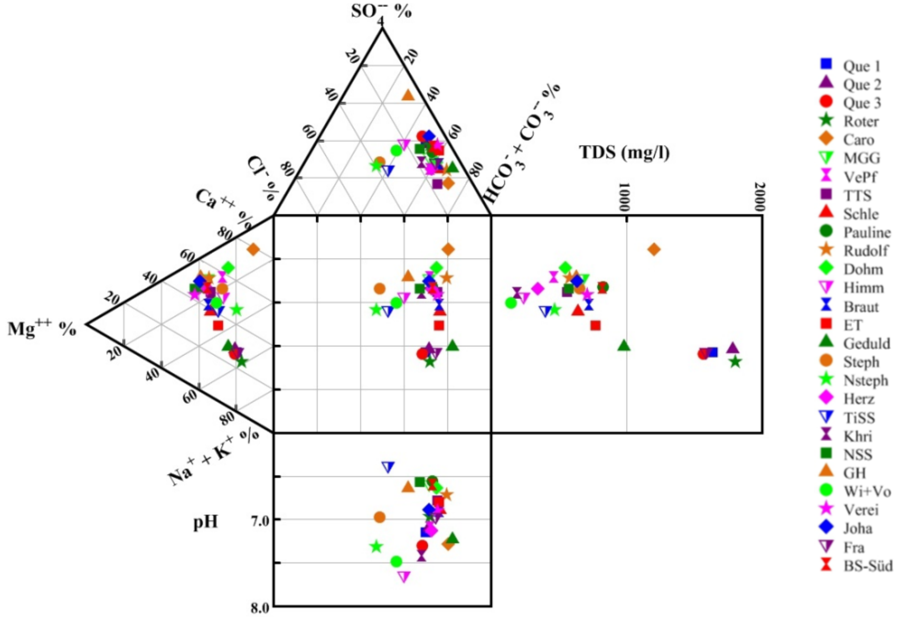

Fig. 2 *Durov diagram of mine water hydrogeochemistry in the study area.*

***Table 1** Spearman correlation matrix of major physicochemical parameters.*

	HCO_3^-	SO_4^{2-}	Cl^-	NO_3^-	F^-	Ca^{2+}	Mg^{2+}	Na^+	K^+	pH	EC	TDS
HCO_3^-	1.000	**0.758**	0.436	**-0.728**	0.210	**0.738**	**0.738**	**0.751**	**0.827**	-0.036	**0.877**	**0.936**
SO_4^{2-}		1.000	0.433	**-0.595**	0.319	**0.796**	**0.928**	**0.675**	**0.759**	-0.140	**0.898**	**0.847**
Cl^-			1.000	-0.172	0.114	**0.639**	0.417	**0.659**	0.476	0.014	**0.573**	0.440
NO_3^-				1.000	-0.332	-0.376	**-0.668**	**-0.557**	**-0.677**	0.170	**-0.537**	**-0.602**
F^-					1.000	-0.014	0.231	0.477	0.380	0.185	0.268	0.299
Ca^{2+}						1.000	**0.718**	**0.536**	**0.646**	-0.063	**0.865**	**0.767**
Mg^{2+}							1.000	**0.630**	**0.733**	-0.268	**0.794**	**0.773**
Na^+								1.000	**0.837**	0.099	**0.804**	**0.773**
K^+									1.000	0.029	**0.870**	**0.851**
pH										1.000	0.038	-0.075
EC											1.000	**0.933**
TDS												1.000

Bold indicates correlation is significant at the 0.01 level (2-tailed),
Underlined indicates correlation is significant at the 0.05 level (2-tailed).

Multivariate Statistical Analysis

Correlation matrix

The Spearman correlation coefficients were calculated, the results of the correlation matrix for 12 variables are given in Table 1. The interpretation of the correlation matrix showed that the pH values are not correlated with ions.

TDS values are strongly correlated (r > 0.8) with most of the ions in the water samples, such as TDS - HCO_3^- (r = 0.936) and TDS - SO_4^{2-} (r = 0.847), mostly statistically significant at the 0.01 level ($p < 0.01$), while TDS values show a weaker correlation with Cl- (r = 0.44), significant at the 0.05 level ($p < 0.05$).

Na^+ and K^+ in water samples are correlated positively with Cl^- (r = 0.659, $p < 0.01$) and (r = 0.476, $p < 0.05$), respectively. This may indicate halite/sylvine dissolution influencing mine water chemistry.

Ca^{2+} and Mg^{2+} concentrations are correlated with HCO_3^- with r = 0.738 ($p < 0.01$). This suggests calcite and dolomite dissolution. Ca^{2+} is also correlated significantly with SO_4^{2-} (r = 0.796, ($p < 0.01$)), suggesting possible gypsum dissolution.

Ca^{2+} and Na^+ ions have an average correlation (r = 0.536, $p < 0.01$), indicating that Ca/Na ion exchange also increases the Ca^{2+} concentration in the mine water. EC showed strong to moderate positive and negative correlations with HCO_3^-, SO_4^{2-}, Ca^{2+}, Mg^{2+}, Na^+, K^+, Cl^- and NO_3^-. It reflects the participation of these elements in the brackish mine water environment in the study area.

Hierarchical Cluster Analysis (HCA)

The main result of the HCA performed on the 28 mine water samples is the Dendrogram (Fig. 3). The Q-mode factor analysis, sampling sites with larger similarities are first grouped. Next, groups of samples are joined with a linkage rule and the steps are repeated until all observations have been classified. Based on the desired level of resolution for interpretation and the potential of the data considered, the number of groups is chosen (Ribeiro and Macedo 1995). Ward's method was more successful to form clusters, with the phenomena line of 2.5. Results showed that the HCA classified mine water samples into 5 geochemically distinct clusters (C1-C5) of similar characteristics, represented in Dendrogram to categorize different levels, which the Stiff diagrams show for each group (Fig. 3).

Samples of C2, C4, and C5 are of Ca-Na-HCO_3-SO_4, Ca-Mg-Na-HCO_3-SO_4-Cl, and Na-Ca-HCO_3-SO_4 types, respectively, while those from C1 and C3 are of Ca-Mg-HCO_3-SO_4 type. The C4 cluster accounts for 17.9% of the total samples, including 5 samples, with

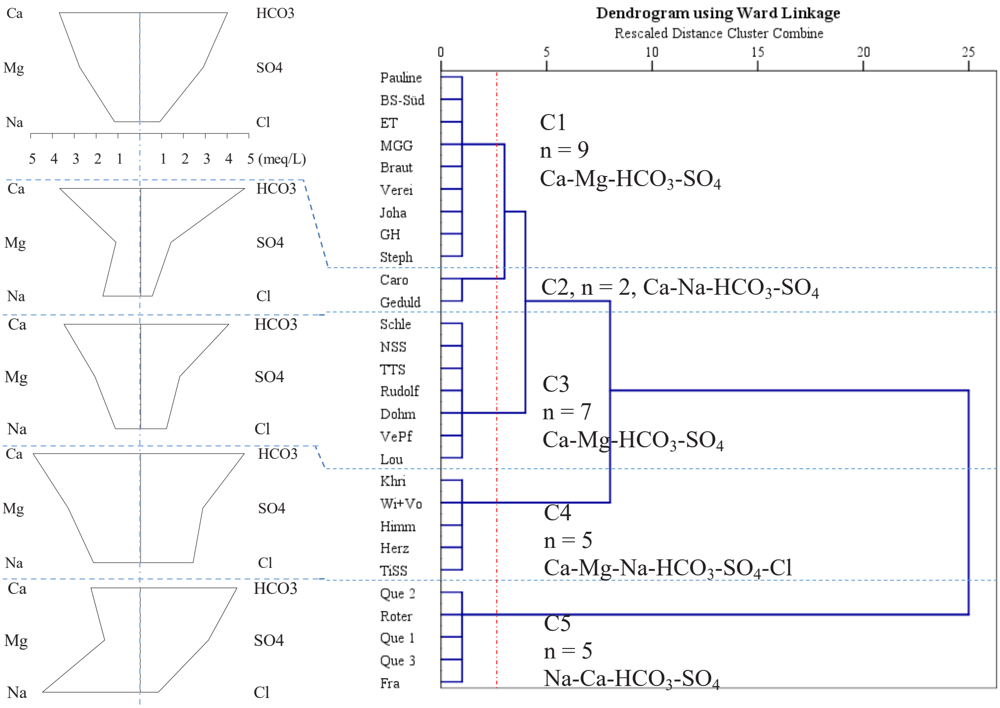

Fig. 3 Dendrogram of the major ions in mine water and corresponding Stiff diagrams.

the average EC and TDS values of 301.6 µS cm^{-1} and 271.4 mg L^{-1}, respectively, mostly distributed in the southern part (the highest elevation). C3 consists of 7 samples located at a lower elevation, accounts for 25% of the total samples (average EC: 579.71 µS cm^{-1} and mean TDS: 554.71 mg L^{-1}). The common feature of these water types is low salinity. C1 is composed of 9 samples and represents 32.1% of the total samples. This water type is relatively fresh, with an average EC of 746.89 µS cm^{-1} (mean TDS = 722.89 mg L^{-1}), mostly distributed in the central study area (Fig. 1). C2 included by two water samples (mean EC and TDS values of 897.5 µS cm^{-1} and 1097 mg L^{-1}, respectively), characteristic for mixed water. C5 is represented by 5 samples, with an average EC of 1609.8 µS cm^{-1} (mean TDS = 1680.4 mg L^{-1}), interpreting the character of blended water and characterized by high salinity, mostly distributed in the east-northern part (Fig. 1). HCO_3^- and Ca^{2+} were the dominant ions in the mine water of the study area (clusters C1, C2, C3, and C4), while Na^+ was the most abundant of the cations (cluster C5). The Que 2, Roter, Que 1, Que 3 and Fra water samples presented the highest average values of HCO_3^-, Ca^{2+}, SO_4^{2-} and Na^+ ions.

Principal Component Analysis (PCA)

The PCA was used to identify major factors governing groundwater geochemistry by many researchers (e.g. Qian *et al.* 2016). The value of the KMO test is 0.752 > 0.5, indicating the total number of significant factors. The loadings of different elements, including the total eigenvalues of different factors, the percentage of variances, and the percentage of cumulative variances are summarized in Table 2. A system of orthogonal axes is created, which helped to visualize the projections of the data matrix elements in a plan diagram (Fig. 4).

From the hydrogeochemical data, a total of three PCAs were extracted with total eigenvalues higher than 1 which account for 82.95% of the total variance (Table 2). Three components reflect the major effective factors controlling the mine water chemistry.

The PCA 1 contributed 58.18% of the total variance and is characterized by high positive loadings on HCO_3^-, SO_4^{2-}, Ca^{2+}, and Mg^{2+} (Table 2, Fig. 4). High scores of this component indicate high mineralization of mine water. This suggested that PCA 1 represented the dissolution or precipitation processes of carbonate, such as calcite and dolomite and sulfate minerals (Qian *et al.* 2016). However, PCA 1 had also high positive loadings for Na^+, K^+, TDS, EC, and Cl^- (Table 2, Fig. 4), suggested that it represented rock weathering and water salinity. It is addressed as the highly mineralized factor, because of exhibited high loadings on TDS (0.944), EC (0.974), SO_4^{2-} (0.873), Cl^- (0.65) and Na^+ (0.852). Therefore, mine water is not only contributed by Ca^{2+} and Mg^{2+} but also contributed by Na^+. It implied that it might be represented the ion exchange. The PCA 2 was 14.18% of the total variance in mine water dataset, with mainly consisted of NO_3^- and pH (Table 2, Fig. 4), implied that it might be the consequence of agricultural activities in the study area. The PCA 3 explained 10.59% of the total variation of the matrix data and mainly consisted of F (Table 2, Fig. 4), suggested that it was contributed from the primary geological environment (Teng *et al.* 2018)management, and development. In this survey of a 10-km-wide area along both sides of the Songhua River, northeast China, the hydrogeochemical responses to different SW–GW interactions were studied. Three types of SW–GW interactions were identified—"recharge", "discharge", and "flow-through"—according to the hydraulic connection between the surface water and groundwater. The single factor index, principal component analysis, and hierarchical cluster analysis of the hydrogeochemistry and pollutant data illuminated the hydrogeochemical response to the various SW–GW interactions. Clear SW–GW interactions along the Songhua River were revealed: (1. Moreover, in Table 2, the PCA 3 had also a low positive loading on Na^+ (*0.419*) and a negative one for Ca^{2+} (*-0.397*), implied that ion exchange processes may occur in the mine water environment, as mentioned above.

***Table 2** Loadings of different parameters for the three PCAs following varimax rotation.*

Parameters	Principal Components		
	PCA 1	PCA 2	PCA 3
HCO_3^-	**0.913**	-0.163	0.184
SO_4^{2-}	**0.873**	-0.192	0.249
Cl^-	**0.650**	0.354	-0.230
NO_3^-	-0.275	**0.839**	-0.277
F^-	0.206	0.014	**0.843**
Ca^{2+}	**0.802**	-0.018	***-0.397***
Mg^{2+}	**0.738**	-0.559	0.097
Na^+	**0.852**	0.035	***0.419***
K^+	**0.785**	-0.240	0.197
pH	0.059	**0.801**	0.360
EC	**0.974**	-0.082	0.184
TDS	**0.944**	-0.118	0.193
Total eigenvalue	6.981	1.702	1.271
% of variance	58.18	14.18	10.59
% of cumulative variance	58.18	72.36	82.95

High loading values are highlighted in **bold**

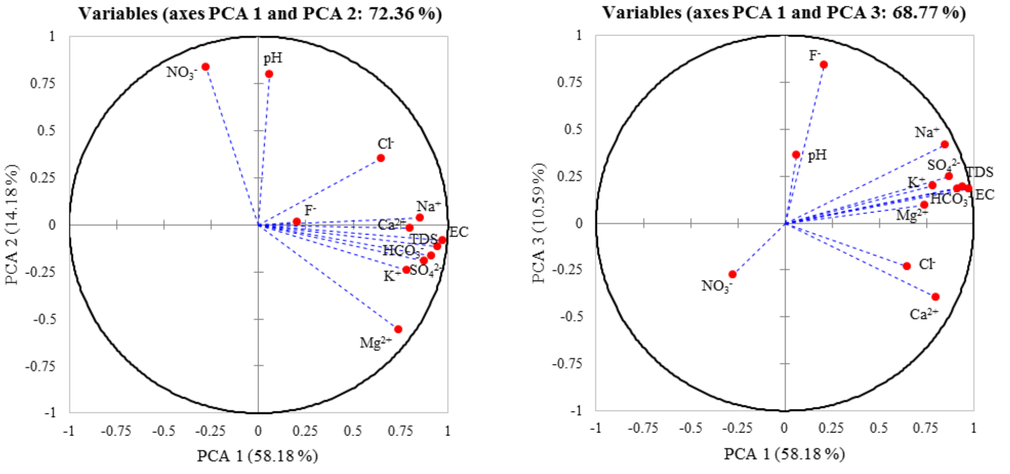

Fig. 4 Representation of the parameters in the factorial plan PCA 1-PCA 2 and PCA 1-PCA 3.

Conclusions

Multivariate statistical methods were used to characterize the origin of the chemical constituents present in mine water of abandoned coal mines in the study area. The correlation of ions in mine water was assessed. The dissolution of carbonate, sulfate, silicate, and chloride minerals are the primary processes controlling mine water chemical composition. The HCA in Q-mode showed 5 statistical groups based on the major distinguishing factors of EC and TDS values. The PCA identified three principal components, which explained 82.95% of the total variance. The first PCA accounted for 58.18% of the total variance and is related to rock weathering, such as dissolution or precipitation of carbonate and sulfate minerals, ion exchange processes, and water salinity, while the other components are controlled primarily by agricultural activities and local geological environment (14.18% and 10.59%, respectively). The results of this study illustrated the usefulness of multivariate statistical analysis in hydrogeochemical studies.

Acknowledgments

This study is financially supported by the 911 Scholarship of the Vietnamese Government and the German Academic Exchange Service (DAAD). Thanks to M.Sc. Sarah Burscheid who helped with sampling in the field, as well as the helping of Mr. N. Richard and Mr. O. Schübbe (Hydrochemistry Laboratory, Institute of Geology, Mineralogy and Geophysics, Ruhr University of Bochum) for the analyzing assistance the mine water hydrogeochemical samples. Last but not least, the hosts who allowed to access their property.

References

Drozdzewski G, Schäfer A, Brix MR (2008) Excursion Guidebook, 26th Regional meeting of the International Association of Sedimentologists

Durov SA (1948) Classification of natural waters and graphic presentation of their composition. Dokl Akad Nauk SSSR 59(1):87-90

DWD (2019) Weather and Climate-German Weather Service (Deutscher Wetterdienst). Online source, https://www.dwd.de/DE/klimaumwelt/cdc/cdc_node.html. Accessed 2019

Frankenhoff H, Balzer I, Witthaus H (2017) Mine water concept in detail - A case study of closing a German coal mine at Ruhr district. pp 175-182

Horák J, Hejcman M (2016) 800 years of mining and smelting in Kutná Hora region (the Czech Republic)-spatial and multivariate meta-analysis of contamination studies. J Soils Sediments 16:1584-1598

Khelif S, Boudoukha A (2018) Multivariate statistical characterization of groundwater quality in Fesdis, East of Algeria. J Water L Dev 37:65-74

Li P, Tian R, Liu R (2018) Solute geochemistry and multivariate analysis of water quality in the Guohua phosphorite mine, Guizhou province, China. Expo Heal 11:81-94

Qian J, Wang L, Ma L, *et al* (2016) Multivariate statistical analysis of water chemistry in evaluating groundwater geochemical evolution and aquifer connectivity near a large coal mine, Anhui, China. Environ Earth Sci 75:1-10

Reghunath R, Murthy TRS, Raghavan BR (2002) The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India. Water Res 36(10):2437-2442

Ribeiro L, Macedo ME (1995) Application of multivariate statistics, trend- and cluster analysis to groundwater quality in the Tejo and Sado aquifer. In: Groundwater Quality: Remediation and Protection. Proc. conference, Prague. pp 39-47

Strehlau K (1990) Facies and genesis of Carboniferous coal seams of Northwest Germany. Int J Coal Geol 15:245-292

Teng Y, Hu B, Zheng J, *et al* (2018) Water quality responses to the interaction between surface water and groundwater along the Songhua River, NE China. Hydrogeol J 26:1591-1607

Wisotzky F (2017) Water chemistry of the “Erbstollen” waters in the southern Ruhr area. In: Korrespondenz Wasserwirtschaft (KW), p 10 (in German)

Zieger L, Littke R, Schwarzbauer J (2018) Chemical and structural changes in vitrinites and megaspores from Carboniferous coals during maturation. Int J Coal Geol 185:91-102