International Journal of Earth, Energy and Environmental Sciences
ISSN: 2517-942X
Vol:15, No:9, 2021

Hydrochemical Contamination Profiling and Spatial-Temporal Mapping with the Support of Multivariate and Cluster Statistical Analysis

S. Barbosa, M. Pinto, J. A. Almeida, E. Carvalho, C. Diamantino

***Abstract*—**The aim of this work was to test a methodology able to generate spatial-temporal maps that can synthesize simultaneously the trends of distinct hydrochemical indicators in an old radium-uranium tailings dam deposit. Multidimensionality reduction derived from principal component analysis and subsequent data aggregation derived from clustering analysis allow to identify distinct hydrochemical behavioral profiles and generate synthetic evolutionary hydrochemical maps.

***Keywords*—**Contamination plume migration, K-means of PCA scores, groundwater and mine water monitoring, spatial-temporal hydrochemical trends.

I. Introduction

THE case study relates to an old radium-uranium tailings dam located in the Central Region of Portugal. The deposition of tailings started in early 1910s and finished in 1988. In the initial years mining activity at the site was related to the production of radium concentrate. After the Second World War, the production changed to uranium concentrate. The tailings deposited at this mining site have therefore high heterogeneity regarding its mineralogical and geochemical composition. Its granulometric composition varies spatially ranging from clays, silts, silty-clay, clay loam and fine to medium sands [1], [2]. The tailings dam is located at the Hesperian (Iberian) Massif and was constructed over a monzonitic, two-mica and predominantly biotitic Hercynian Granite of late carbonic age. According to local geological and geophysical studies promoted by the operator [3], the underlying granite has two major areas in depth of distinct weathering and joint degrees [4]. Major faults in the granite are possibly the main conduits for water circulation beneath the tailings dam and are linked to a local riverside. Old mine shafts installed within the tailings have acted as conduits for air and water circulation in the deposit. For several years, before the remediation works, these shafts promoted the connectivity with atmosphere and oxidation phenomena within the tailings deposit.

According with previous characterization studies [3], the area of the tailings deposit comprises two main aquifer systems. The first, and most superficial, is of hypodermic nature and three to six meters thin. It is composed of residual soil that results from the granite weathering. The groundwater flow is performed by a “porous” type media and develops accordingly with local porosity and permeability. The second aquifer is semi-confined and underlies the first. It has a very distinct anisotropic hydraulic behavior once it is formed by the granite rock matrix. In this case, groundwater flows are primarily directly dependent on fracture and on its spatial density, interconnectivity, filing and aperture characteristics. Interconnectivity between the two aquifers is expected to be difficult because of the clayed nature of the residual topsoil of the first layer which may act in some places as a geological barrier. Infiltration and leakage are mainly present in areas where fractures or faults facilitate hydraulic connectivity between the two aquifers.

Fig. 1 presents a groundwater flow conceptual model for the study area that includes the water potential surface of the complex system composed by the tailings and the hypodermic aquifer that underlays the tailings dam. It includes the main water inflow and outflow in the dam and the preferential water flows inside and beneath de tailings deposit. It is to be referred that a local inflow stream that surrounds the west side of the tailings dam has a significant importance in local groundwater flow percolation. Also, riverside embankments on the east side work as main receptors of the seepage that is generated in the tailings dam and percolate through the preferential flow streamlines.

The priority objectives of the environmental remediation work fields at the tailings dam deposit were to confine the waste, circumscribing the dispersion of sources of contamination and the levels of radiation, as well as establishing the safety conditions associated with the mechanical stability of slopes. Landfill was modelled and re-profiled with the purpose of slope stabilization and a drainage system and a multi-layer covering system were installed.

Fig. 2 presents the best groundwater flow modelling results achieved for pre- and post-remediation scenarios. These groundwater models were performed with the freeware version of the Processing Modflow for Windows, PMWIN 5.3.1 [6].

S. Barbosa and J. A. Almeida are with FCT-NOVA Nova School of Science and Technology & GeoBioTec, Campus da Caparica, 2829-516 Caparica, Portugal (e-mail: svtb@fct.unl.pt, ja@fct.unl.pt).

M Pinto was with FCT-NOVA Nova School of Science and Technology, Campus da Caparica, 2829-516 Caparica, Portugal. He is now with the IST, Instituto superior Técnico, Lisbon (e-mail: mariana.mateus.pinto@gmail.com).

E. Carvalho and C. Diamantino are with EDM - Empresa de Desenvolvimento Mineiro, S.A., R. Sampaio e Pina, 1, 5º Esqº1070-248 Lisboa, Portugal.

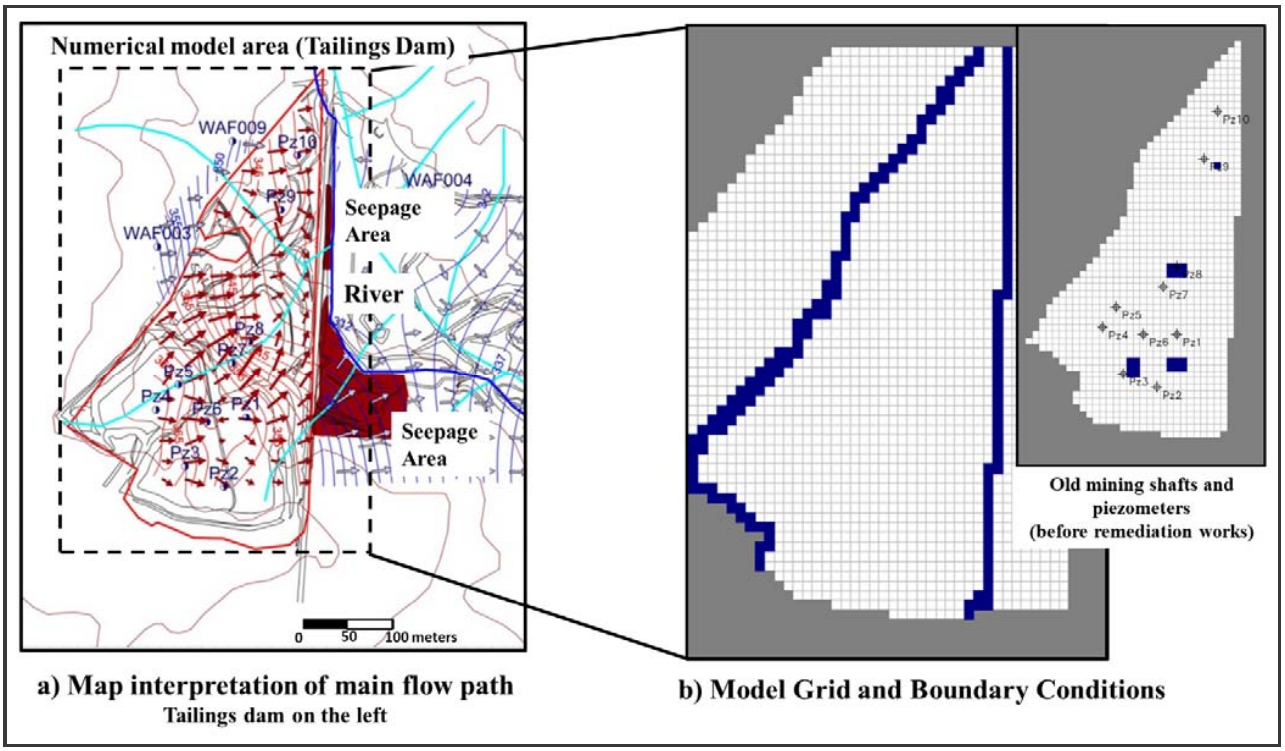

Fig. 1 (a) Conceptual model of water potential surface and flow path in the tailing dam and in the surrounding area. Main seepage in dark red color is related with the location of old streamlines (in light blue color) circulating beneath the tailings dam and in the surrounding granite. The old streams and the actual location of the main river are indicated (adapted from [5]); (b) Grid of the numerical flow model and boundary conditions (2D planar view, XY plan). Location of old mining shafts and piezometers are also indicated

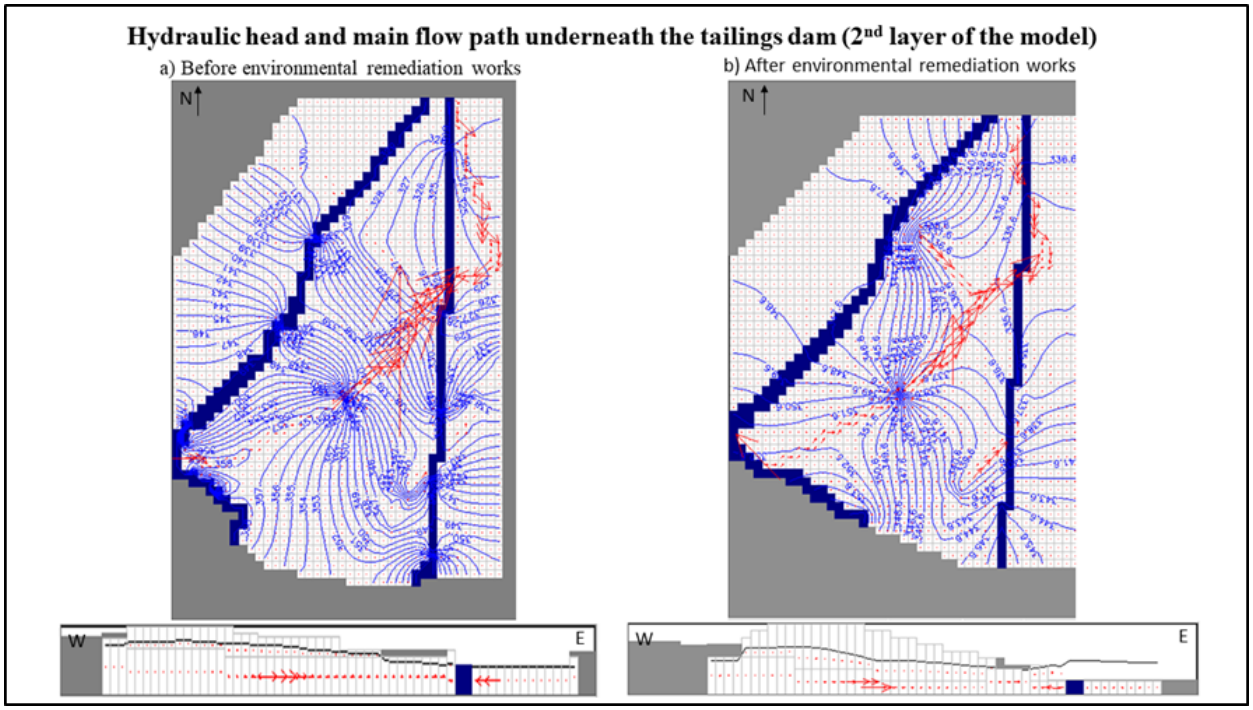

Fig. 2 Results of the simulated numerical groundwater models for the cases of (a) before and (b) after the environmental remediation works. 2D planar XY and YZ (or W-E) sections show the flow for the case of the second layer, that is, for the underlying granite (adapted from [5])

It is relevant to emphasize that the main hydraulic gradient and flow paths are consistent with the local old preferential streamlines that are located beneath the tailings dam (Figs. 1 and 2). It is also possible to verify that seepage locations were respected in the numerical simulated model. The tailings started to be deposited directly on the ground without any previous preparation or drainage work. It is therefore expectable that these old streams exist and act as preferential conducts for groundwater circulation beneath the dam, being possibly responsible for the main hydrochemical exchanges due to hydrodynamic effects. The location of outflow derived from these old stream lines coincides with the main seepage areas (Figs. 1 and 2). The numerical simulation results demonstrate the existence of a main flow with direction from SW to NE and another smaller flow with NW to SE direction. These flows remained even after the dam was requalified. Their existence and remaining presence are related with the fact that they are established in the underling granite, beneath the old tailings deposit. In the granites, the water fluxes will receive contributions through local faults and high-density fracture areas which coincide with main seepage areas that appears on the riverbanks.

Before the mentioned environmental remediation works, the superficial water and groundwater at the tailings site offered a very poor quality. Besides the radiological effects, significant local impacts in hydrological and groundwater systems were also present in the consequence of this old mining structure.

In contamination problems, detection and estimation of statistical tendencies and patterns may be a complex process especially in situations where environmental conditions change significantly in time, as in this case. Seasonality factors, data asymmetry, moderate and severe outliers, missing data, are examples of situations that often occur and that may impede interpretations, statistic processing and data trend detection [7]-[17]. In some cases, uncertainty generated by noise in the observed data complicates the detection of possible trends and behaviors in average terms which are necessary for accurate statistical methods [13]. These difficulties constituted the reason of inspiration to test the statistical and mapping methodology presented in this study.

II. Materials and Methods

A. Hydrochemical Data

Since 2002, a Water Quality Monitoring Plan (WQMP) has been in place at this site by EDM, a Portuguese state-owned company [3]. From 2005 to 2007, EDM carried out environmental remediation works at this old tailing dam. The site in question was subject to a relevant environmental rehabilitation project to mitigate the radiological impacts, leachate generation control and chemical contamination dispersion, as well as to create the necessary security conditions for the mechanical stability of slopes. This study considers the time-series results of the WQMP for the period between 2002 and 2016, which comprises data from pre to post environmental remediation works. It is important to mention that for this period of time, the developed WQMP allowed the identification by the operator and subsequent acceptance by the responsible authorities of the main hydrochemical indicators to be considered for the control of water contamination derived from the tailings dam. These indicators are: pH, Conductivity (μS/cm), eH, total uranium (p.p.b.), radium-226 (mBq/l), sulfate (mg/l), chloride (mg/l), manganese (mg/l) and calcium (mg/l). The developed statistical tests consider therefore these hydrochemical indicators. Statistical interpretations of hydrochemical data of the piezometers and water shafts installed in the dam and in its surroundings were performed. Monitoring of these piezometers is included in the WQMP that was initiated in 2002 by the operator and has continued to the present days [3], [18], [19].

In our work, distinct subsets of quarterly frequency hydrochemical data were considered for each monitoring piezometer (that is, piezometer numbers 1, 2, 3, 5, 6, 7, 9, 10 and 11) in three distinct periods of time T1, T2 and T3. T1 aggregates hydrochemical data for piezometers before environmental remediation works (from 2002 to 2005), T2 aggregates the data immediately after environmental remediation works (from 2008 until 2010), and T3 aggregates data after the remediation works from 2015 to 2016. Principal Component Analysis (PCA) and K-means clustering method (KMC) were selected to develop statistical analysis of the space-time data, [17], [20]-[33]. An interpretation of the hydrochemical quality evolution for each piezometer (1, 2, 3, 5, 6, 7, 9, 10 and 11) along time (T1, T2 and T3) was possible to be developed with the adopted statistical methodology. According to the available data for each piezometer at each time-period, 24 subsets were established and considered for statistical interpretations: 1T1, 1T2, 2T1, 2T2, 2T3, 3T1, 3T2, 3T3, 5T1, 5T2, 5T3, 6T1, 6T2, 6T3, 7T2, 7T3, 9T1, 9T2, 9T3, 10T1, 10T3, 11T1, 11T2 and 11T3. For each of these subsets, the hydrochemical indicators pH, uranium total mass concentration (referred as Utotal or total uranium) [18], [34], [35], radium-226 (Ra226), sulfate (SO4), chloride (Cl), manganese (Mn) and calcium (Ca) were considered for PCA and KMC statistical analysis.

B. PCA and KMC Analyses

The multivariate statistical analyses were performed using R software “R Project for Statistical Computing”, [36]-[39]. PCA and KMC were applied to the referred 24 hydrochemical data subsets. Quantitative variables of the statistical analysis were the hydrochemical indicators of contamination an considered in the monitoring plan (WQMP) of the tailings dam; pH, Utotal, Ra226, SO4, Cl, Mn and Ca.

PCA can be done by eigenvalue decomposition of a data covariance or correlation matrix or singular value decomposition of a data matrix. The number of the newly generated variables, the principal components, is equal to the smaller of the number of original variables minus one. With this procedure, the dimensionality of the data is reduced, and (multivariable) interpretations are facilitated. PCA is mostly used as a tool in multivariate statistical analysis and also to visualize genetic distance and relatedness between data

observations and quantitative or qualitative variables [27], [40]-[43]. The results of a PCA are usually discussed in terms of component scores, sometimes called factor scores, that is, the transformed variable values corresponding to a particular data point and loadings which are the weight by which each standardized original variable should be multiplied to get the component score. PCA aims to construct a low-dimensional subspace based on a set of principal components (PCs) to approximate all the observed samples in the least-square sense [44], [45]. Due to the quadratic loss used, PCA is notoriously sensitive to corrupted observations (outliers) and the quality of its outputs can suffer severely in the face of even a few outliers [45]. For this, and to ensure a better statistical significance of the selected data series, it was decided during the pre-processing data stage to eliminate the strongest outliers. In our study, PCA was then used to the 24 data sets for unsupervised dimension reduction of data. Package FactoMineR was used [46]. PCA for the hydrochemical variables pH, Utotal, Ra226, SO4, Cl, Mn and Ca was performed. The data were reduced into three main components which describe differences in the concentrations of Utotal, Ra226, SO4, Cl, Mn and Ca according with pH conditions at the different piezometers (1, 2, 3, 5, 6, 7, 9, 10 and 11) and at distinct time intervals T1, T2 and T3. The three PCs describe differences in groundwater compositions, in accordance with different pH conditions that result from modifications in tailings composition and/or from changes in drawdown and/or hydrodynamics conditions in groundwater flow.

The PCs scores were then considered for a subsequent KMC analysis. KMC is a method that is popular for cluster analysis. KMC aims to partition n observations into k clusters, in which, each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. It is an unsupervised clustering technique commonly used. This procedure results in partitioning of the data space into Voronoi cells. Dissimilar and contrasting expert opinions may be found in related literature [20], [22], [27] regarding the order of analysis that was chosen between PCA and KMC, and the effectiveness of using PCA scores to subsequent KMC ("PCA before KMC" or "KMC before PCA"). In [22] the authors consider PCA variables as the continuous solutions to the discrete cluster membership indicators for KMC. For the case of "PCA before KMC", PCA is applied before clustering analysis to reduce dimensionality and to facilitate the visualization of the main relevant clusters. PCA variables are the continuous solutions to the discrete cluster membership indicators for KMC. Oppositely, for instance, in [27] the authors showed that clustering with the PCA scores instead of the original variables may not necessarily improve cluster quality. These authors concluded inclusively that the first PCA variables (which contain most of the variation in the data) may not necessarily capture most of the cluster structure. Therefore, the application of PCA and KMC, and the effectiveness of using PCA scores to subsequent KMC must be carefully verified, being its effectiveness directly depend on the intrinsic characteristics of the data. In our work, we compute KMC after PCA. The achieved results were verified variable by variable, time step by time step, location by location step in order to ensure the efficiency of the results of data series that were subsequently mapped and presented in a synthetic way. KMC was performed to group the data observations into distinct major clusters. Tests were performed for clusters 4, 5, 6 and 7. Distinct cluster groups represent different hydrochemical profiles. Subsequently, the space-time data subsets (1T1, 1T2, 2T1, 2T2, 2T3, 3T1, 3T2, 3T3, 5T1, 5T2, 5T3, 6T1, 6T2, 6T3, 7T2, 7T3, 9T1, 9T2, 9T3, 10T1, 10T3, 11T1, 11T2 and 11T3), considered as qualitative variables, were classified according with these hydrochemical profiles. Also, with this procedure, it was also possible to make an interpretation of the hydrochemical quality evolution for each piezometer (1, 2, 3, 5, 6, 7, 9, 10 and 11) along time (T1, T2 and T3). For better visualization and spatial interpretation, PCA scores were mapped using the geostatistical procedure of kriging. Environmental quality evolution of the groundwater that percolates within the tailings dam was mapped considering a multivariate interpretational approach where all the hydrochemical variables are considered simultaneously. This is quite unique and interesting because in monitoring groundwater plans it is always necessary to take into consideration the time and space evolutions of distinct (and sometimes several) hydrochemical indicators. With this methodology, it is possible to make multivariate data interpretations in simplified synthetic way without losing information.

III. Results and Discussion

The developed PCA analysis has reduced the data set from seven quantitative variables to three new variables, the principal components PC1, PC2 and PC3. These three PCs explain 72.16% of the total variance (Table I) which is a reasonable result considering the high variability of to hydrochemical data set.

According to the results in Table I, it is obvious the closer relation between the parameters SO4, Cl and Mn, and, at a second correlation level, with Utotal and Ca. pH has an inverse correlation for all these mentioned parameters. This is because, in acidic environments, as the pH increases to a value of 6-7, the environmental conditions turn less acidic, and the concentration of contaminants drops. The hydrochemical profile of the contaminated waters generated in the tailings deposit is of acidic nature (low pH) and has a high level of contamination in some characteristic anions, cations, base-metals, and semi-metals, including, in this case, Utotal and Ra226 because of the nature of the exploited ore in the past. One aspect of relevance is that, according to the PCA1 scores, Ra226 does not have the same behavior profile of the indicators Utotal, SO4, Mn, Cl and Ca. Its behavior is better explained by the variables PCA2 and PCA3 from which it is possible to conclude that Ra226 will have some direct correlation with pH, that is, as pH increases, Ra226 will have some tendency to increase as well. Ra226 has therefore an opposite behavior comparatively to the other hydrochemical indicators. A subsequent KMC analysis was developed for all the qualitative samples 1T1, 1T2, 2T1, 2T2, 2T3, 3T1, 3T2,

3T3, 5T1, 5T2, 5T3, 6T1, 6T2, 6T3, 7T2, 7T3, 9T1, 9T2, 9T3, 10T1, 10T3, 11T1, 11T2 and 11T3 according with its scores in PCA1, PCA2 and PCA3. Tests with clusters 4, 5 and 6 were performed allowing the conclusion that better results were achieved when 6 clusters are considered. For computing cluster analysis, the "maximum interaction number" was 100, the "n start number" was 25 and "Hartigan-Wong" was the algorithm considered. Representativeness of the results (between SS/total SS) was 90.1%. Clustering vectors for all the samples were selected for interpretation, allowing the association of each one of the six clusters to its respective group of qualitative samples. That is, all qualitative samples were classified in one of the six clusters. The results of KMC are presented in Table II. As expected, KMC results confirm the distinct hydrochemical profiles that results from the previous PCA analysis (Fig. 3).

TABLE I
SCORES, EIGENVALUES, AND CUMULATIVE VARIANCES FOR VARIABLES PC1, PC2, AND PC3.

Variables	PC1 (Dim.1)	PC2 (Dim.2)	PC3 (Dim.3)
Quantitative			
SO4	0.81	0.14	0.08
Cl	0.78	0.03	0.35
Mn	0.78	-0.07	0.35
Utotal	0.68	0.36	0.05
Ca	0.55	-0.56	-0.35
Ra226	0.34	0.66	-0.58
pH	-0.59	0.41	0.42
Qualitative			
5T1	3.64	2.13	-0.51
7T3	3.02	-0.99	1.01
7T2	2.62	0.43	0.76
5T2	2.38	-0.76	0.37
6T1	1.75	0.41	-0.51
5T3	0.98	-0.72	0.73
2T2	0.89	-0.79	0.37
2T1	0.49	-0.42	-0.29
3T1	0.36	-0.65	-0.35
6T2	0.34	2.19	-3.39
1T1	0.17	-0.60	0.12
3T2	-0.11	0.46	-1.30
7T1	-0.12	-1.11	-0.52
6T3	-0.14	-0.73	0.50
9T2	-0.15	0.04	-0.78
2T3	-0.23	-1.14	-0.23
11T1	-0.29	-0.58	0.50
9T3	-1.00	-0.54	-0.47
1T2	-1.12	0.55	-0.75
11T3	-1.48	0.29	-0.40
10T1	-1.48	-0.16	-0.06
9T1	-1.49	-0.11	-0.01
3T3	-1.56	-0.12	0.09
11T2	-1.60	0.19	0.03
10T3	-1.80	0.56	0.05
10T2	-2.47	1.26	1.40
Eigenvalue	3.09	1.07	0.89
Variance (%)	44.15%	15.25%	12.76%
Variance (Cumul. %)	44.15%	59.40%	72.16%

Each cluster represents a certain hydrochemical profile which, in turn, corresponds to the considered qualitative variables (of location and time stage). Also, each established hydrochemical profile corresponds to a certain specific contamination condition. From the results and interpretations (see Table II), it is possible to verify the existence of a trend from the most contaminated locations, in cluster 1 (C1), to the less ones, represented by locations and time steps of cluster 5 and cluster 6 (C5 and C6). Following this sequence, it is possible to identify the locations within the tailings dam with positive evolutions in the consequence of environmental remediation works – piezometers 1, 3, 5, 6 and 11 –, and those who have a low but also positive evolution or which have stabilized in time - piezometers 2, 9, 10 and 11 (Table II and Fig. 4). Afterwards, the first three PC scores (PC1, PC2 and PC3) for all sample locations and time intervals, 1T1, 1T2, 2T1, 2T2, 2T3, 3T1, 3T2, 3T3, 5T1, 5T2, 5T3, 6T1, 6T2, 6T3, 7T2, 7T3, 9T1, 9T2, 9T3, 10T1, 10T3, 11T1, 11T2 and 11T3 were mapped at 2D, in the tailings dam, through its spatial interpolation with kriging. The interpretative maps of PCA and KMC results for the case of PCA1 variable at each time interval T1, T2 and T3 are presented in Fig. 4.

From Fig. 4 and considering previous results presented in Tables I and II, it is possible to consider the following conclusions:

- PCA1-T1:

a) Cluster 1 represents the most contaminated hydrochemical profile and is associated to variable 5T1 (that is, to the location of piezometer 5 at the pre-remediation stage). The profile with the highest contamination detected in the monitoring plan is only present at time-period T1 (that is, before the environmental remediation). It represents an acid environment with all parameters having high to very high concentrations. Utotal, Ra226 and SO4 have extremely high to very high concentrations. Cl, Mn, and Ca have high concentrations.

b) At this same time-period (T1), piezometer 6 is the other one that also has a much-demarcated contamination pattern, and it is represented by cluster 2. This cluster also represents a profile of an acidic environment, like the environment of cluster 1 but with lower Ra266 concentrations.

c) In the remaining area of the tailings dam the profiles are of type C4 and C5. Environmental conditions are better with less effects of contamination, groundwater is less contaminated (cluster 4) or only slightly contaminated or not contaminated (cluster 5).

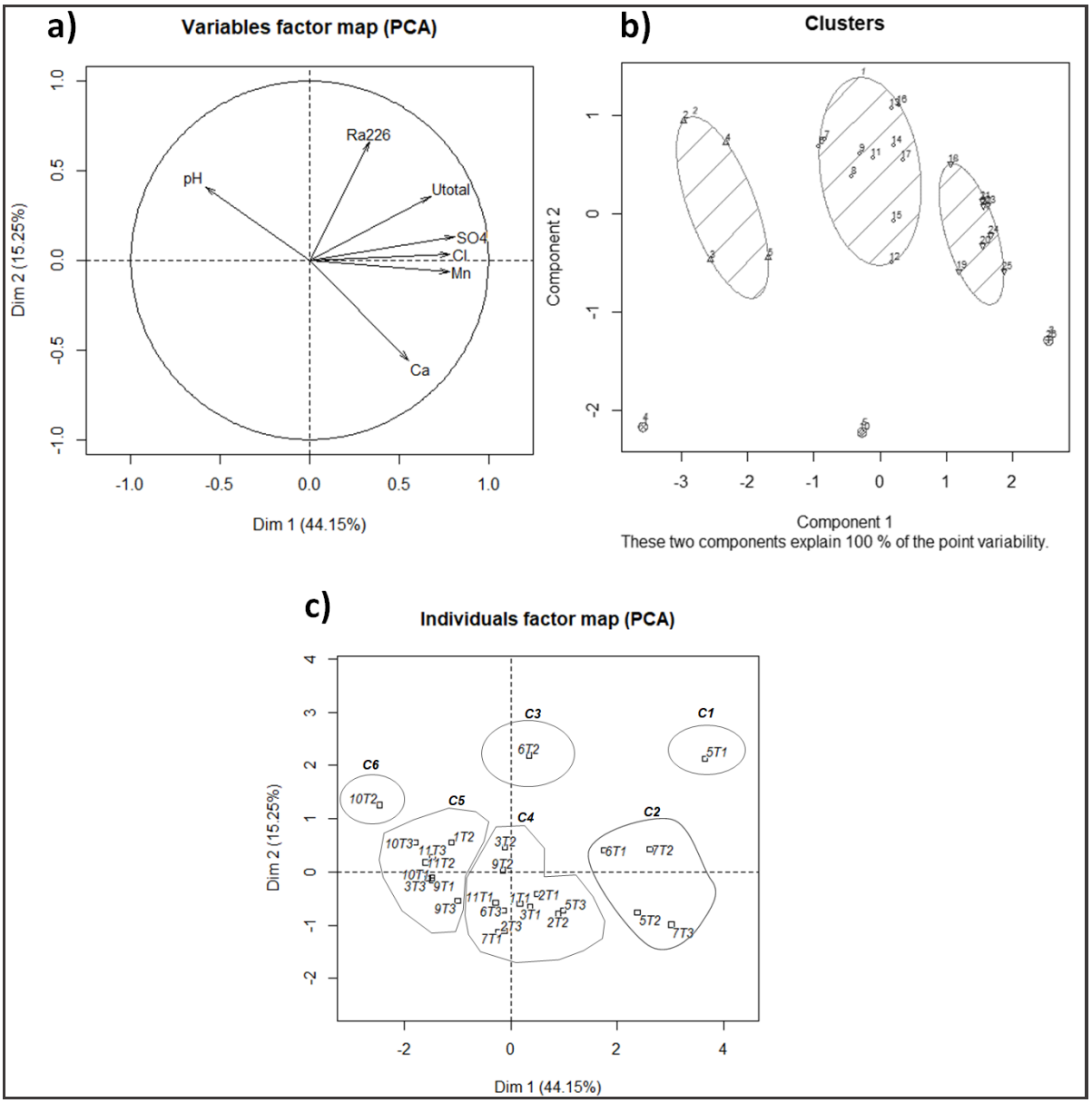

Fig. 3 Interrelation between results of PCA and KMC: (a) Projection of quantitative variables in PCA1 and PCA2; (b) KMC considering 6 clusters; (c) Adequation of the clustering results to PCA results for the case of the qualitative variables

- PCA1-T2 and PCA1-T3: at T2 the profile of piezometer 5 is represented by cluster 2. During the remediation works, piezometer 5 evolved to a less extremely high contaminated profile although it remains with high contamination. It is also possible to conclude that this contamination profile spreads upward from piezometer 5 to piezometer 7. These evolutions during period T2 are spatially close to the location of the main old stream water underneath the tailings with main direction from SW to NE which is highlighted from numerical groundwater modelling. It seems that, somehow, and probably in the dependence of local hydrodynamic flow, the main contamination plume has migrated northwards from site 5 to site 7. Another interesting hydrochemical behavior is detected in piezometer 6 which has, in the period T2, a hydrochemical profile represented by cluster 3. In this case, local environmental conditions are less acidic to closer to neutral. All the concentrations decrease except for Ra226, SO4 and Ca. This behavior is reported by PCA2 scores and is well reflected in map PCA2-T2 (Fig. 5); however, it is not well represented statistically (Table II, cluster 3, n = 1) despite conditioning the analysis. Also, locally, at piezometer 10, during the period T2, an alkaline profile is detected. This fact is probably related with some earth movements in the northern part of the tailings dam, during remediation works for the re-

profiling of slopes, where old tailings materials were either moved or rotated, inducing influence on the materials which have a more carbonated composition. Here, the alkaline or close to neutral waters and the presence of some ions in freshwaters that inflow through the northwest part of the dam, like chloride (and sulfate), may have also helped facilitate the temporary remobilization and transportation of Ra226. This possible phenomenon may be observed in Fig. 5. In T2 and T3 periods, all the other sites at the tailings dam present profiles of very less contamination (cluster 4) or of environmental conditions without or with very low contamination (cluster 5). In periods T2 and T3, slightly contaminated profiles maybe present in some locations that are represented by cluster 4. This behavior may be related with mixing groundwater effects from fresh waters that inflow laterally at western and north-western and from SW to NE direction, which result in pH increments and general reduction in hydrochemical concentrations. After the remediation works, that is, in T3 period, contamination is much more circumscribed and only remains in the location of piezometer 7 (PCA1-T3 map). In T3 the only location with remaining strong contamination is given by piezometer 7. All the other studied sites present a significant improvement regarding groundwater quality.

TABLE II
RESULTS OF KMC HYDROCHEMICAL PROFILES WITH ITS RESPECTIVE CLASSIFICATION AND INTERPRETATION FOR EACH CLUSTER

Cluster Piezom./ Time interval	Univariate Statistics	pH	U_{tot}	Ra (226)	Mn	SO4	Cl	Ca	Interpretation Specific site environmental conditions
C1 5T1	Mean	3,2	23667	1447	438	11163	657	331	Very contaminated acid environment. All parameters have high concentrations. U_{tot}, Ra-226 and SO4 with extremely high to very high concentrations while Cl, Mn and Ca have high concentrations.
	St. Dev.	0,1	1528	90	126	8613	34	123	
	Minimum	3,1	22000	1390	309	1490	630	190	
	Maximum	3,3	25000	1550	560	18000	695	415	
C2 7T2, 7T3, 5T2, 6T1	Mean	3,2	5603	622	534	8483	893	395	Very contaminated acid environment. Similar to Cluster 1 but concentrations of U_{tot}, Ra-226 and SO4 tend to be lower.
	St. Dev.	0,2	4906	762	347	6071	411	219	
	Minimum	2,9	10	1	41	1400	391	136	
	Maximum	3,6	13709	2000	994	19647	1527	914	
C3 6T2	Mean	5,4	697	3200	150	2200	38	470	Environment with some specific contaminants. pH slightly acidic; Ra-226 with very high concentrations and Ca with high concentrations; concentrations of U_{tot} and SO4 are slightly high; concentrations of Cl and Mn are low.
	St. Dev.	–	–	–	–	–	–	–	
	Minimum	–	–	–	–	–	–	–	
	Maximum	–	–	–	–	–	–	–	
C4 1T1, 2T1, 2T2, 2T3, 3T1, 3T2, 5T3, 6T3, 9T2, 11T1	Mean	3,8	1261	308	267	3916	310	321	Contaminated acid environment. Concentrations are lower comparatively to Clusters 1 and 2. pH is acidic; concentrations of U_{tot} and SO4 are high; presence of some contamination in Ra-226; all other remaining parameters are present in more low concentrations.
	St. Dev.	0,8	1640	447	158	2828	411	100	
	Minimum	2,9	10	7	10	33	7	43	
	Maximum	5,8	5000	2000	585	10044	1927	490	
C5 1T2, 3T3, 9T1, 9T3, 10T1, 10T3, 11T2, 11T3	Mean	4,7	508	326	18	570	41	134	Environment conditions without or with low contamination, or, locations with significant improvement in environmental conditions. pH neutral to slightly acidic; presence in groundwater of some U_{tot} and SO4 concentrations.
	St. Dev.	1,2	498	397	26	431	28	104	
	Minimum	3,3	53	2	1	231	10	37	
	Maximum	6,8	1800	1700	87	1910	133	440	
C6 10T2	Mean	9,7	333	136	43	668	60	12	Environmental conditions with significant variations in one single temporal stage (T2); In general, strong to medium alkaline environment, without contamination or with low contamination (mainly Ra-226 and SO4).
	St. Dev.	3,0	719	151	49	766	27	23	
	Minimum	3,9	0,9	6,7	6,0	35	34	0,03	
	Maximum	12,1	1800	342	136	2138	100	57	

IV. CONCLUSIONS

The integration of PCA scores in a KMC analysis facilitates multivariate interpretations of PCA results, allowing a better clarification of the sense of PCA grouping relations through the clusters that are obtained. In our study, mapping of PCA scores and its interpretation with KMC allowed a better understanding of spatial-temporal trends of hydrochemical indicators included in a groundwater monitoring plan within a mine tailings dam. KMC results of PCA scores facilitated interpretations and temporal and spatial analysis. Spatial mapping of KMC and PCA scores allows a more synthetic, faster, and easier approach when multivariate indicators need to be considered for an evaluation of contamination profiles. In this case study, clusters of KMC represent distinct groundwater quality hydrochemical profiles. Evolution of environmental conditions at each site location (that is, at each piezometer) in accordance with the identified groundwater quality profiles was performed. KMC enabled a better understanding of the PCA results once observations were classified according to distinct hydrochemical profiles at each piezometer location, according to each time-period T1, T2 and T3. Subsequently, distinct areas within the dam were demarcated according to different degrees of contamination and non-contamination profiles.

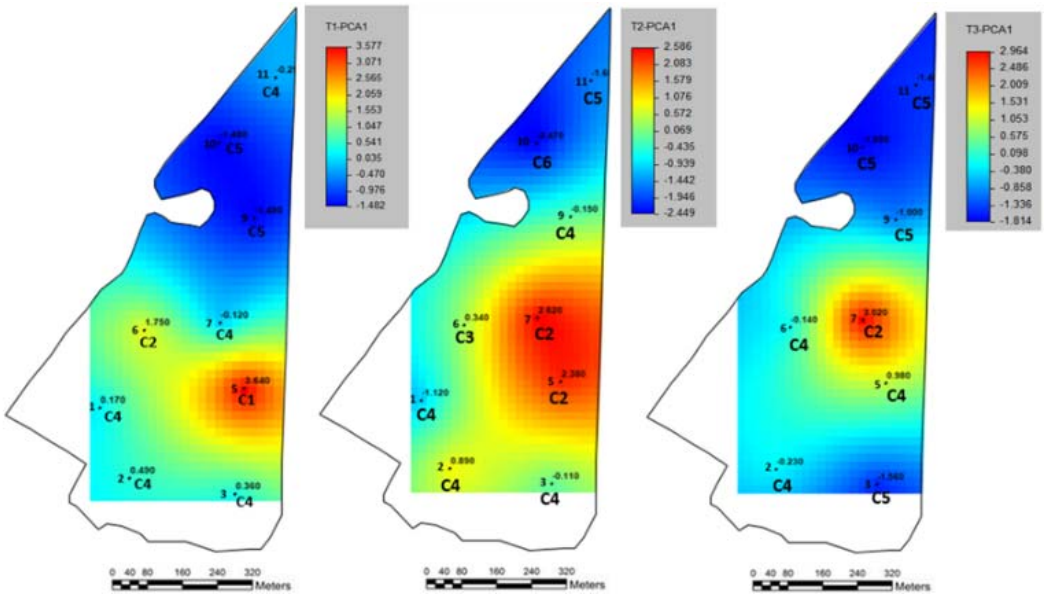

Fig. 4 Mapping of PCA1 scores for time periods T1, T2 and T3. For a better spatial interpretation, KMC results (clusters C1 to C6) are also represented

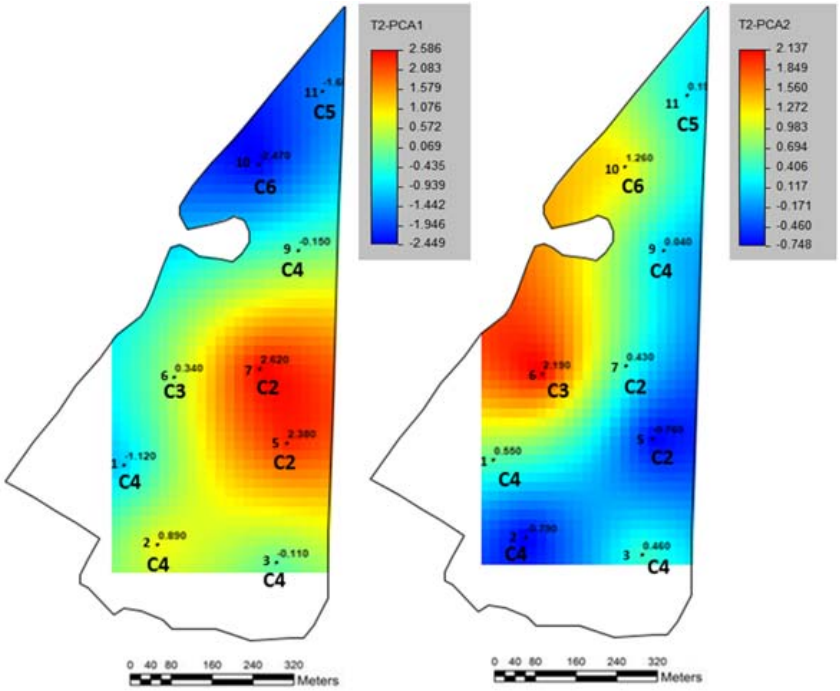

Fig. 5 Mapping of PCA1 and PCA2 scores for the time period T2. For a better spatial interpretation, KMC results (clusters c1 to c6) are also represented

The concentrations of total uranium, sulfate, chloride, manganese, and calcium decrease in consequence of pH increment. In turn, and temporarily, at an intermediate stage, radium-226 concentrations increased in some circumscribed locations, probably in consequence of pH increase, of more oxidizing environment conditions and mixing phenomena with fresh waters derived from preferential inflow path. In the post-remediation period, a general improvement in groundwater quality is verified.

Finally, it is to enhance the possibilities that may be derived from crossing the spatial-trend maps with groundwater flow maps. For this specific case-study, the superposition of the locations of the distinct hydrochemical profiles (clusters) with groundwater flow modelling results will certainly result in a

better understanding of the inner hydrodynamics of the tailings dam deposit and its direct implications on contamination plume changes over time. Relations between different flow paths located inside and beneath the tailings dam and its spatial correspondence to hydrochemical profiles evolution through time (T1, T2 and T3) are possible to be found. The application potentialities of spatial-trend maps derived from PCA, and K-Means analysis are, therefore, quite unique, and interesting in the context of groundwater quality and monitoring of multi-variable complex systems.

ACKNOWLEDGMENT

The authors are grateful to EDM (Empresa de Desenvolvimento Mineiro) S.A. for providing the original hydrochemical data and for supporting the publication of this work.

This work is a contribution to Project UIDB/04035/2020 funded by FCT-Fundação para a Ciência e a Tecnologia, Portugal.

REFERENCES

[1] Unified Soil Classification System (USCS) ASTM D2487-11 Directive: Standard Practice for Classification of Soils for Engineering Purposes.

[2] A. K. Howard, Soil Classification Handbook: Unified Soil Classification System. Denver, Colorado: Geotechnical Branch, Division of Research and Laboratory Services, Engineering and Research Center, Bureau of Reclamation, 1986.

[3] EDM, The legacy of abandoned mines – The context and the action in Portugal. EDM and DGEG (eds.), 2011. http://ec.europa.eu/environment/waste/mining/pdf/Appendix_III_to_Annex3.pdf.

[4] ISRM, International Society for Rock Mechanics, Rock Characterization, Testing and Monitoring - ISRM Suggested Methods, Pergamon Press, Oxford, UK, 1981.

[5] M. Pinto, Spatio-temporal evolution of hydrodynamic and hydrochemical conditions in a former tailings dam as a result of its environmental remediation (Evolução espácio-temporal das alterações hidrodinâmicas e hidroquímicas numa antiga barragem de rejeitados mineiros em resultado de obras de recuperação ambiental). MSc Dissertation, Faculty of Science and Technology, NOVA University of Lisbon, 2016.

[6] W. H. Chiang, W. Kinzelbach, "PMWIN Processing Modflow, A Simulation System for Modeling Groundwater Flow and Pollution", Hamburg, Zürich, 334 p, 1998.

[7] R. M. Hirsch, J. R. Slack, R. A. Smith, Techniques of trend analysis for monthly water quality data. Water Resour. Res., 1982, 18-1, 107-121. https://doi.org/10.1029/2009WR008071.

[8] R.M. Hirsch, J.R. Slack, A Nonparametric Trend Test for Seasonal Data with Serial Dependence. Water Resources Research, 1984, 20, 727-732. https://doi.org/10.1029/WR020i006p00727.

[9] Z.W. Kundzewicz and A.J. Robson. Change Detection in Hydrological Records - A Review of the Methodology. Hydrological Sciences Journal, 2004, 49, 7-19.

[10] H. Boyacioglu and H. Boyacioglu, Investigation of Temporal Trends in Hydrochemical Quality of Surface Water in Western Turkey. Bull Environ Contam Toxicol, 2008, 80, 469–474. https:doi.org/10.1007/s00128-008-9439-0.

[11] K. Wahlin and A. Grimvall, Uncertainty in water quality data and its implications for trend detection: lessons from Swedish environmental data. Env. Science & Policy, 2008, 11, 2, 115-124. ISSN 1462-9011, https://doi.org/10.1016/j.envsci.2007.12.001.

[12] R. E. Chandler, E. M. Scott (eds.), Statistical Methods for Trend Detection and Analysis in the Environmental Sciences, John Wiley & Sons, Ltd, 2011. https:doi.org/10.1002/9781119991571.

[13] J. Mozejko, Detecting and Estimating Trends of Water Quality Parameters in Water Quality Monitoring and Assessment, Dr. Voudouris (Ed.), 2012. ISBN: 978-953-51-0486-5.

[14] D. Anghileri, F. Pianosi and R. Soncini-Sessa, Trend detection in seasonal data: from hydrology to water resources. Journal of Hydrology, 2014, 511, 171-179. ISSN 0022-1694. https://doi.org/10.1016/j.jhydrol.2014.01.022.

[15] D.T. Monteith, C.D. Evans, P.A. Henrys, G.L. Simpson, I.A. Malcolm, Trends in the hydrochemistry of acid-sensitive surface waters in the UK 1988–2008. Ecological Indicators, 2014, 37, Part B, 287-303. ISSN 1470-160X, https://doi.org/10.1016/j.ecolind.2012.08.013.

[16] R. J. Cooper, K. M. Hiscock, A. A. Lovett, S. J. Dugdale, G. Sünnenberg, E. Vrain, Temporal hydrochemical dynamics of the River Wensum, UK: Observations from long-term high-resolution monitoring (2011–2018), Science of The Total Environment, 2020, 724, 138253. ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2020.138253.

[17] K. Voudouris, A.Panagopoulos, J. Koumantakis, Multivariate Statistical Analysis in the Assessment of Hydrochemistry of the Northern Korinthia Prefecture Alluvial Aquifer System (Peloponnese, Greece). Natural Resources Research, 2000, 9, 135–146. https://doi.org/10.1023/A:1010195410646

[18] C. Diamantino, E. Carvalho, R. Pinto, Water resources monitoring and mine water control in Portuguese old uranium mines. Proceeding of IMWA2016 Annual Conference, 2016, July 11–15, Leipzig, Germany.

[19] Q. Zhang, H. Wang, Y. Wang, M. Yang, L. Zhu, Groundwater quality assessment and pollution source apportionment in an intensely exploited region of northern China. Environ. Sci. Pollut. Res., 2017, 24, 16639–16650. https://doi.org/10.1007/s11356-017-9114-2.

[20] K. Y. Yeung, D. R. Haynor, W. L. Ruzzo, Validating clustering for gene expression data. Bioinformatics, 2001, Volume 17, Issue 4, 309–318. https://doi.org/10.1093/bioinformatics/17.4.309

[21] S. K. Swanson, J. M. Bahr, M. T. Schwar, K. W Potter, Two-way cluster analysis of geochemical data to constrain spring source waters. Chemical Geology, 2001, 179, 73–91.

[22] C. Ding, X. He, Cluster Structure of K-means Clustering via Principal Component Analysis. In: Dai H., Srikant R., Zhang C. (eds) Advances in Knowledge Discovery and Data Mining. PAKDD 2004. Lecture Notes in Computer Science, vol 3056. Springer, Berlin, Heidelberg, 2004. https://doi.org/10.1007/978-3-540-24775-3_50.

[23] Koonce, J.E., Yu, Z., Farnham, I.M., Stetzenbach, K.J. (2006). Geochemical interpretation of groundwater flow in the southern Great Basin. Geosphere, 2 (2), 88–101. https://doi.org/10.1130/GES00031.1

[24] M. Templ, P. Filzmoser, C. Reimann, Cluster analysis applied to regional geochemical data: Problems and possibilities. Applied Geochemistry, 2008, 23, 2198–2213.

[25] G. Thyne, C. Güler, E. Poeter, Sequential Analysis of Hydrochemical Data for Watershed Characterization. Ground Water, 2008, Vol. 42, 5, 711–723. https://doi.org/10.1111/j.1745-6584.2004.tb02725x

[26] R. Ledesma-Ruiz, E. Pastén-Zapata, R. Parra, T. Harter, J. Mahlknecht, Investigation of the geochemical evolution of groundwater under agricultural land: A case study in Northeastern Mexico. Journal of Hydrology, 2015, 521, 410–423.

[27] D. Machiwala and K. J. Madan, Identifying sources of groundwater contamination in a hard-rock aquifer system using multivariate statistical analyses and GIS-based geostatistical modeling techniques. Journal of Hydrology: Regional Studies, 2015, 4, 80–110.

[28] P. Mandel, M. Maurel, D. Chenu, Better understanding of water quality evolution in water distribution networks using data clustering. Water Research, 2015, 87, 69-78.

[29] K. Peng, X. Li, Z. Wang, Hydrochemical characteristics of groundwater movement and evolution in the Xinli deposit of the Sanshandao gold mine using FCM and PCA methods. Environ. Earth Sci., 2015, 73,7873–7888. https://doi.org/10.1007/s12665-014-3938-6

[30] H. Li and Y. Gao, Multivariate statistical approaches to identify the major factors governing groundwater quality. Appl. Water Sci., 2018, 8, 215. https://doi.org/10.1007/s13201-018-0837-0

[31] G. Sotomayor, H. Hampel, R. F. Vázquez, Water quality assessment with emphasis in parameter optimisation using pattern recognition methods and genetic algorithm. Water Research, 2018, 130, 353-362.

[32] A.E. Marín Celestino, J.A. Ramos Leal, D.A. Martínez Cruz, J. Tuxpan Vargas, J. De Lara Bashulto, J. Morán Ramírez, Identification of the Hydrogeochemical Processes and Assessment of Groundwater Quality, Using Multivariate Statistical Approaches and Water Quality Index in a Wastewater Irrigated Region. Water, 2019, 11, 1702. https://doi.org/10.3390/w11081702.

[33] B. Helena, R. Pardo, M. Veja, E. Barrado, J. M. Fernandez, L. Fernandez, Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis,

2000. Water Research, Vol. 34, Issue 3, 807-816. https://doi.org/10.1016/S0043-1354(99)00225-0
[34] ATSDR (Agency for Toxic Substances and Disease Regestry), Toxicological Profile for Uranium, in Toxic Substances Portal, Agency for Toxic Substances and Disease Registry, 339–357, 2011. https://www.atsdr.cdc.gov/toxprofiles/tp150-c7.pdf, Accessed 5 may 2021.
[35] L.S. Keith, O. M. Faroon, B. A. Fowler, Uranium. In G. F. Nordberg, B. A. Fowler, M. Nordberg (eds.), Handbook on the Toxicology of Metals, pp. 881–900, Academic Press, Fourth Edition, 2014.
[36] T. Hothorn, B. S. Everitt, A Handbook of Statistical Analyses using R. CRAN.R-project.org document, 2015. https://cran.r-project.org/web/packages/HSAUR3. Accessed 15 January 2017.
[37] T. Hothorn, B. S. Everitt, A Handbook of Statistical Analyses using R. Chapman & Hall/CRC Press, Third Edition, 2017.
[38] R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2013. http://www.R-project.org/.
[39] T. Rahlf, Data Visualisation with R – 100 Examples. Springer. e-Book ISBN: 978-3-319-49751-8, 2017.
[40] J. F. Santos, I. Pulido-Calvo, M. M. Portela, Spatial and temporal variability of droughts in Portugal. Water Resour. Res., 2010, 46, W03503.https://doi.org/10.1029/2009WR008071
[41] B. Zhang, X. Song, Y. Zhang, D. Han, C.Tang, Y. Yu, Y. Ma, Hydrochemical characteristics and water quality assessment of surface water and groundwater in Songnen plain, Northeast China, Water Resour. Res., 2012, 46, 2737-2748.
[42] P. Wuttichaikitcharoen and M. S. Babel, Principal Component and Multiple Regression Analyses for the Estimation of Suspended Sediment Yield in Ungauged Basins of Northern Thailand. Water, 2014, 6, 2412-2435; https://doi.org/10.3390/w6082412
[43] S. Selvakumar, N. Chandrasekar, G. Kumar, Hydrogeochemical characteristics and groundwater contamination in the rapid urban development areas of Coimbatore, India. Water Resources and Industry, 2017, 17, 26–33.
[44] K. Pearson, On lines and planes of closest fit to systems of points in space. Philosophical Magazine, 1901, 2:559-572. In <http://pbil.univ-lyon1.fr/R/pearson1901.pdf>
[45] D. Kim and Se-K Kim, Comparing patterns of component loadings: Principal Component Analysis (PCA) versus Independent Component Analysis (ICA) in analysing multivariate non-normal data. Behav. Res., 2012, 44:1239–1243. https://doi.org/10.3758/s13428-012-0193-1.
[46] S. Le, J. Josse, F. Husson, FactoMineR: An R Package for Multivariate Analysis. Journal of Statistical Software, 2008, 25(1), 1-18. https://doi.org/10.18637/jss.v025.i01