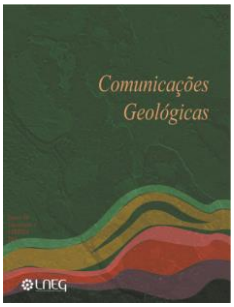
Versão online: https://www.lneg.pt/wp-content/uploads/2020/05/Volume_107_CIG.pdf

Comunicações Geológicas

Comunicações Geológicas (2020) 107, Especial II, 55-62
ISSN: 0873-948X; **e-ISSN:** 1647-581X

High-resolution geochemical mapping in the Mociços mine (Ossa-Morena Zone, Portugal). Contributions from machine learning methods



Mapeamento geoquímico de alta resolução na mina de Mociços (Zona de Ossa-Morena, Portugal). Contributos de métodos de aprendizagem automática

P. Nogueira[1,2*], S. Vicente[1], M. Maia[1,2], J. Roseiro[2], N. Moreira[1,2], J. X. Matos[3]

Recebido em 20/01/2020 / Aceite em 11/02/2020
Publicado online em julho de 2020

Artigo original
Original article

Resumo: A mina de Mociços consiste num filão de quartzo e carbonatos rico em Cu, hospedado em rochas metassedimentares detríticas de idade paleozoica da Zona de Ossa-Morena (SW Portugal), tendo sido explorado no início do século XX. Foi realizada uma campanha de geoquímica de solos, com auxílio de uma fluorescência de raios-X portátil, de modo a estimar a distribuição dos metais nas imediações da mina de Mociços, seguida de estudos de estatística descritiva, mapeamento de interpolação espacial elementar e diferentes métodos de aprendizagem automática (PCA, hclust e k-média) dos dados obtidos. Os resultados evidenciam que a geoquímica dos solos da mina de Mociços é espacialmente variável, sendo reconhecidas três populações principais (distinguidas por todos os métodos de aprendizagem automática, mas melhor definidas por PCA e k-média), correlacionáveis com aspectos geológicos: i) o filão mineralizado, o "gossan", escombreira e rochas félsicas, ii) o fundo geoquímico regional e iii) uma zona de alto contraste geoquímico (Zn acima do valor médio regional). Esta abordagem metodológica é plausível com os mecanismos de acumulação-dispersão de metais nestes ambientes e confirma a importância da geoquímica multivariada na prospeção mineral, assim como dos métodos de aprendizagem automática como ferramenta complementar.

Palavras-chave: Análise de componentes principais, aglomeração hierárquica, k-média, prospeção mineral.

Abstract: The Mociços mine is a Cu-rich quartz-carbonate vein-type deposit hosted in the Palaeozoic detrital metasediments of the Ossa-Morena Zone (SW Portugal), exploited during the beginning of the XX century. A soil geochemistry campaign was carried out, with a portable X-Ray Fluorescence, in order to estimate metal distribution throughout the Mociços mine, followed by summary statistics, interpolation mapping and different machine learning methodological studies (PCA, hclust and k-means) of the obtained data set. Overall results showed that soil geochemistry in the Mociços mine is spatially variable, in which three main populations were recognized (distinguished by all machine learning methods, but better defined by PCA and k-means) and correlated with geological features: i) the mineralized vein, gossan structures, mine tailings and felsic rocks, ii) the geochemical background and iii) a chemically contrasting (Zn higher than regional average) zone. This methodological approach is plausible with typical metal accumulation-dispersion mechanisms in these environments and confirms the relevance of multivariate geochemistry studies for mineral exploration purposes, and of machine learning methods applied to mineral exploration.

Keywords: Principal components analysis, hierarquical clustering, k-means, Mineral exploration.

[1] Departamento de Geociências da Escola de Ciências e Tecnologia da Universidade de Évora. Colégio Luís António Verney, Rua Romão Ramalho, 59, 7000-671 Évora, Portugal.
[2] Instituto de Ciências da Terra - Pólo de Évora. Rua Romão Ramalho, 59, 7000-671 Évora, Portugal.
[3] Laboratório Nacional de Energia e Geologia (LNEG), Campus de Aljustrel; Bairro da Vale d´oca 14, 7601-909 Aljustrel, Portugal.
*Autor correspondente/Corresponding author: pmn@uevora.pt

1. Introduction

Soil geochemistry is a methodological approach accomplished in the strategic phase of any mineral exploration campaign. This technique implies field sampling followed by laboratorial sample preparation, milling the soils to a pulp and, finally, sending the samples to a certified geochemical laboratory to obtain the elemental composition of the sample. Often, the chemical analysis is expensive and time consuming, and the number of elements analysed is restricted to minimize costs, implying a limited number of variables held to define anomalies and pathfinders, with the undesirable possibility of missing relevant exploration targets.

In the last decade, the development of portable X-Ray Fluorescence (pXRF) analysers changed this approach, allowing "in situ" semi-quantitative chemical analysis without sample preparation and expensive analytical costs. This technique is subject to some constraints (*e.g.* Hall *et al*. 2013, Lemière 2018) but the results are perfectly fitted for most of the mineral exploration projects. Common pXRF can identify nineteen or more elements (Lemière, 2018), including the majority of the transition metals group, essential for mineral exploration purposes.

The present work used the pXRF data for high-resolution geochemical soil mapping around the Mociços ancient copper mine in the Ossa-Morena Zone (OMZ), namely in the Sousel-Barrancos Metallogenic Belt (SBMB) (Oliveira, 1986; Mateus *et al*., 2013). The approach starts with the initial exploratory data analysis and extends to the employment of unsupervised machine learning techniques, attempting to unravel the patterns created by the surficial dispersion of the elements around the ancient mine and mine tailings. These advanced techniques include principal components analysis (PCA), hierarchical clustering (hclust) and

k-means clustering (k-means), to study the multi-element behaviour in this type of deposits.

2. Geological background

The Mociços mine is an ancient copper mine that operated between 1919 and 1934 (see Vicente *et al*., 2018 and references therein). The main exploitation works focused on a mineralized vein with more than 1km length, oriented N15ºW dipping 80º towards ENE (Mateus *et al*., 2013; Moreira *et al*., 2017; 2018), with a width variable from less than 20 cm to more than 2 m. The mineralized structure crosscuts the Palaeozoic metasedimentary successions from Estremoz-Barrancos sector (Oliveira *et al*., 1991; Araújo *et al*., 2013), namely the Colorada (Ordovician), Xistos com Nódulos (Silurian) and Xistos Raiados (Devonian) Formations (Fig. 1). Near the exploited zone several mine tailings are recognized, that mostly extend to the East of the main vein, although minor tailings are also recognized to the West.

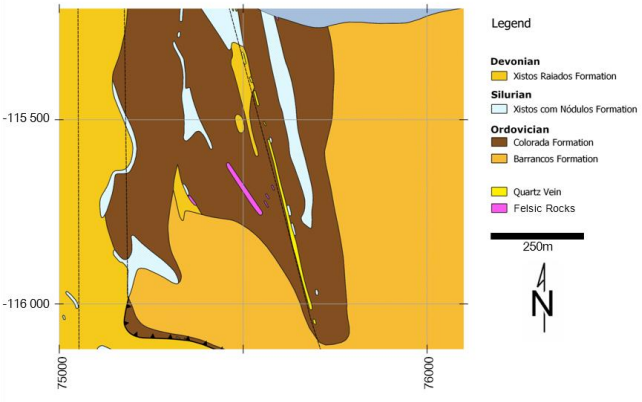


Figure 1. Geological map of the Mociços area. Compiled from different sources (Vicente, in prep.).

Figura 1. Mapa geológico da área de Mociços, compilado apartir de diversas fontes. (Vicente, in prep.).

Mociços is part of a wider cluster of vein-type copper deposits (*e.g.* Bugalho, Zambujeira, Miguel Vacas) which are comprised in the SBMB (Oliveira, 1986; Mateus *et al*. 2013; Matos and Filipe, 2013). These copper occurrences and ancient mines are related with Late Variscan brittle deformation events, which controlled the emplacement of the mineralized veins (Mateus *et al*., 2013; Moreira *et al*., 2017, 2018). The genesis of these copper veins is under discussion: while some authors consider the copper mineralization of SBMB to be related with low to medium salinity fluids that remobilized Cu from Palaeozoic metasedimentary host rocks, later precipitated in these vein structures (Mateus *et al*., 2003, 2013), recent works point towards the presence of magmatic-hydrothermal fluids involved in the ore emplacement conditions (Maia *et al*., 2019a, b), which agree with a hypothesis proposed earlier by Oliveira (1984) that suggest the possibility of porphyry-Cu deposits in the SBMB.

All these ancient mines and occurrences reveal, at surface, the presence of a gossan structure where a secondary mineral paragenesis was developed, mainly composed of copper carbonates and phosphates (*e.g.* malachite, azurite, libethenite), as well as iron and manganese oxides and hydroxides (*e.g.* limonite, goethite, pyrolusite). In the Mociços ancient mine, only the oxidized zone of the vein was exploited. The primary mineralization of SBMB copper deposits is mainly composed of pyrite and chalcopyrite, whereas some of the deposits display variable amounts of arsenopyrite, galena and sphalerite (Mateus *et al*. 2003, 2013; Vicente, in prep.).

3. Methodology and data preparation

To characterize the distribution of the elements in the vicinity of the main vein, eight profiles approximately transversal to the vein strike (N15ºW) were carried out, in a total of 177 points were analysed with the pXRF equipment (c.f. Fig. 4). The analysed soil was previously clean from all the visible vegetation and the surficial (10 cm) layer removed. The soil was compacted and mildly milled to obtain a fine grained and homogeneous spot for analysis. This methodology could be identified as the "mole heap" defined in Lemière (2018), that provides reliable results.

The analyses were conducted "in situ" at the Mociços mine during two surveys, using a pXRF Explorer-9000 analyser (from Skyray Instruments), coupled with an accurate GPS system and four core components, including a miniature X-ray tube with an excitation source of 50 kV/200 µA, a fast SDD detector with 128 eV energy resolution, a digital signal processor and a micro multi-channel analysis module. The internal package "Soil mode" was selected, with an analysis time of 100 seconds; providing good detection for most of the metals (Hall *et al*. 2013) and, in this case, detected a total of fifteen elements: Ti, Mn, Fe, Cd, Sb, V, Cr, Co, Ni, Cu, Zn, As, Se, Pb and Hg. The detection limit of these elements is similar to the ones referred by this author.

Obtained data were preliminarily subjected to a descriptive statistical approach in order to calculate the correlation matrix for all analysed elements. The maps of interpolated results for the selected elements were created using the inverse distance weighted methodology (IDW) with a value of two for the power weight. Principal component analysis was used as an approach to data reduction and to understand the variability of the data. The methods of data clustering (hclust and k-means) are dependent on the absolute values of the data, i.e. values with similar magnitude tend to group together. To avoid this problem all the obtained data were scaled to conform with a mean of zero and a standard deviation of one. Statistics and graphics were done using R 3.6.0 (R Core Team, 2018), the ggplot2 (Wickham, 2016), gplots (Warnes *et al*., 2018) and pheatmap (Kolde, 2015) packages.

4. Data analysis

4.1. Exploratory and spatial data analysis

Table 1 summarizes the main results of the exploratory analysis of the data. It is noteworthy that most of the elements were detected in all the observations, except for Ni (175/177), Zn (171/177) and Hg (17/177). The exploratory data analysis revealed that Hg presents a remarkably high number of observations (n=160) bellow the limit of detection (LOD), therefore, Hg was excluded from interpretation. In some of the detected elements (Cd, Sb and Se) the median is very near or below the LOD of the equipment and, therefore, were not considered.

The evaluation of the obtained mean and median values are within the expected values when compared with the Post-Archean Australian Shale (PAAS; Lentz *et al*. 2003) and Upper Continental Crust (UCC, Taylor and McLennan, 1985) values, with the exception of Cu (11 x PAAS; 23 × UCC), As (19 x UCC), Pb (2 x PAAS; 2 × UCC) and Mn (2 x PAAS; 3 × UCC) values that are distinctively higher (Tab. 1). Outlier values were identified in ten elements (Tab. 1), particularly in Cu (n=18) and Mn (n=17).

Table 1. Exploratory analysis of obtained data in the Mociços area. All values in ppm except Fe in molar %

Tabela 1. Análise exploratória dos dados obtidos na área de Mociços. Todos os valores em ppm, exceto o Fe que está em % molar (* elementos com concentrações superiores aos padrões usados – PAAS e/ou UCC; ** elementos com concentrações mediana abaixo ou próxima dos limites de deteção-LOD).

	LOD	n	Mean	Sd	Median	Min	Max	Skew	Kurtosis	Out
Ti	90	177	4235.28	2076.16	4645.90	859.57	8845.94	-0.18	-1.21	0
Mn*	46	177	1652.53	1416.30	1319.21	270.14	8799.91	2.83	9.26	17
Fe	.003	177	3.45	2.54	3.11	0.57	16.44	2.41	7.63	12
Cd**	-	177	0.49	0.21	0.41	0.08	0.90	0.10	-1.55	0
Sb**	3	177	14.58	12.25	4.52	0.34	31.20	0.07	-1.94	0
V	55	177	107.39	49.59	131.17	0.00	182.32	-0.89	-0.47	0
Cr	53	177	75.46	57.11	72.90	0.00	381.41	1.93	6.53	7
Co	11	177	19.86	20.68	15.50	4.03	138.59	3.23	11.60	14
Ni	10	175	49.64	18.32	47.45	3.44	125.35	1.20	3.18	12
Cu*	9	177	563.31	1213.22	203.06	31.08	8411.82	4.28	19.51	18
Zn	10	171	41.42	30.05	36.43	0.00	166.85	0.95	1.29	6
As*	5	177	28.42	25.34	20.96	6.01	207.98	3.69	19.19	11
Se**	-	177	0.35	0.41	0.12	0.09	2.99	2.65	10.26	10
Pb*	5	177	36.24	22.74	34.08	12.01	271.29	6.58	62.55	9
Hg**	-	17	0.09	0.07	0.06	0.01	0.26	1.00	-0.14	0

*elements with significant higher concentration than used standards – PAAS and/or UC / elementos com concentrações superiores aos padrões usados – PAAS e/ou UCC
** elements with median concentrations below or near the limit of detection-LOD / elementos com concentrações mediana abaixo ou próxima dos limites de deteção-LOD

The analysis of the correlations of the different elements allows to determine the joint behaviour of several elements and, consequently, group them based on their relationships. Figure 2 displays the correlation matrix between all the considered elements, being possible to group distinct elements with similar behaviour:

- Ti and V show very strong positive correlation (r=0.88);
- Pb shows a weak correlation with all the elements, except with As, Cu and Ni, presenting moderate positive correlation (r=0.68, r=0.51 and r=0.43, respectively);
- Cu and As have a strong positive correlation (r=0.79) and moderate positive correlation with Ni (0.66 and 0.58, respectively).
- Cr, Mn, Fe and Co present very strong positive correlations between them (r>0.79);
- Zn has negative correlation with all other elements, displaying a moderate negative correlation with V (r=-0.6) and with the Cr, Mn, Fe and Co cluster (r between -0.43 and -0.50).

Different factors might control the spatial distribution of the elements, for a better understanding of the patterns interpolation maps were created using IDW approach (Fig. 3). The analysis of the patterns obtained suggest that it might be controlled by i) the presence of the high concentration of metals in the quartz-carbonate lode and gossan (*e.g.* Fe, Cu), ii) the surficial mobility of elements (*e.g.* Zn that is concentrated in a topographic low) and iii) the mineralogy/chemistry of the underlying geological units (host rocks).

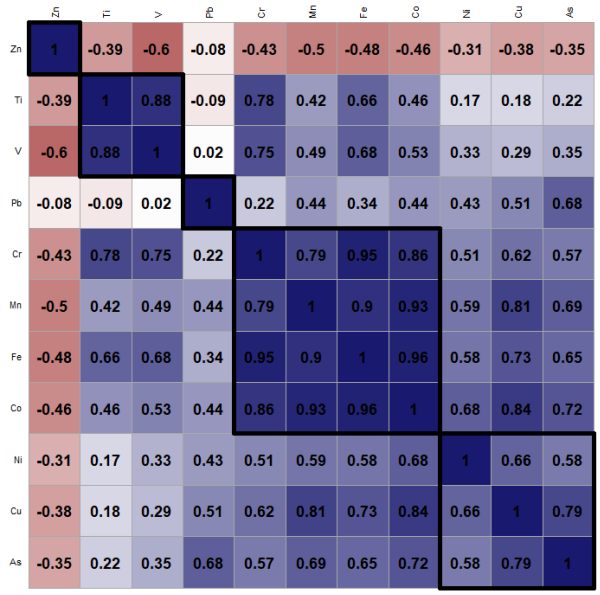


Figure 2. Correlation matrix for the studied elements. Darker tones of colour represent stronger correlations. The black bold line indicates the principal groups of correlated elements. Note that Zn do not correlate significantly with any other element and Pb only shows correlation with As and Cu.

Figura 2. Matriz de correlação para os elementos estudados. Os tons mais escuros representam correlações mais fortes. A linha a negrito representa os principais grupos de correlações. Note-se que o Zn não se correlaciona significativamente com nenhum elemento e o Pb apenas se correlaciona com o As e Cu.

4.2. Principal components analysis (PCA)

PCA is a common tool for dimension reduction of data, providing insights on the behaviour of the different variables and observations (Grunsky, 2001; 2010, Chen *et al.* 2015). Commonly, a biplot diagram (Fig. 4A) is used to characterize the simultaneous behaviour of variables (arrows) and observations (dots).

For the Mociços data, the first three principal components (PC) explain 84.9% of the variability in the data (Tab. 2). The ellipse in figure 4 contains 99.5% of the observed points.

Fe, Co, Mn, Cr, Cu and As are the most significant elements in the PC1 axis, each one explains more than 8% of the variability of this component (PC1 explains 60.4% of data variability), and all plot in the positive part of this axis. Zn is the only element that plots in the negative part of this component.

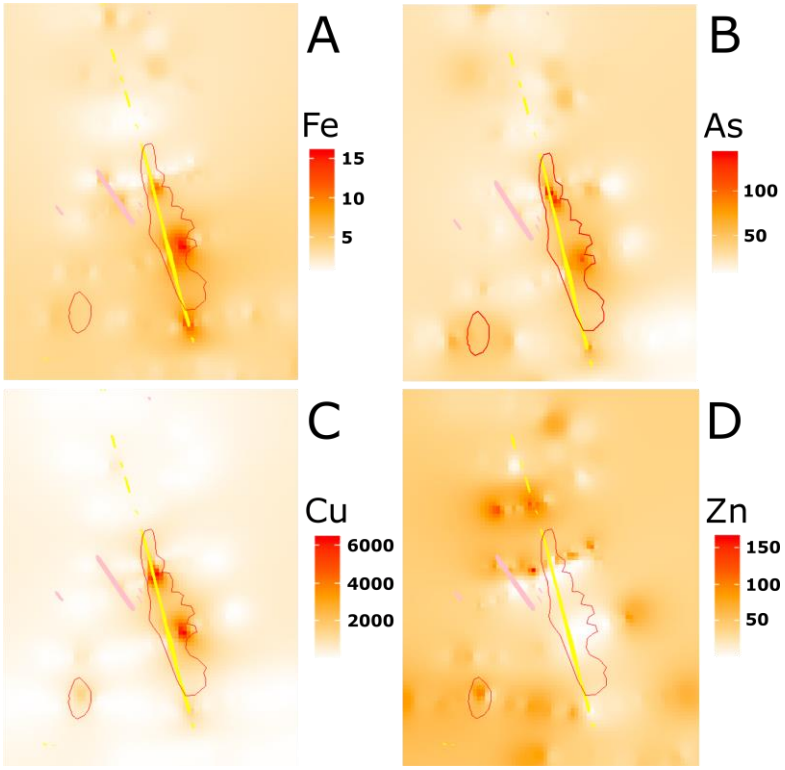


Figure 3. Maps with the distribution of selected elements Fe (%), As (ppm), Cu (ppm) and Zn (ppm). The line demarks the areas where the mine tailings are located. Quartz veins and felsic rocks as in figure 1.

Figura 3. Mapas de distribuição de elementos selecionados Fe (%), As (ppm), Cu (ppm) e Zn (ppm). A linha representa a área onde se localizam as escombreiras da mina. O filão de quartzo e as rochas félsicas estão representados da mesma forma que a figura 1.

The elements in the positive part of the PC2 axis are Ti, V, Cr and Fe, but only Ti and V are significant components (*i.e.* each one explains more than 10% of the variability in this axis). Pb is also significant (explains 22% of variability for the PC2) and plots in their negative part (Fig. 4); Ni, As, Cu, Mn and Co also plot in the negative quadrant, though no relevant variability (given their relationship with PC1). PC3 (not plotted) is dominated by the behaviour of Zn explaining 6.6% of the variability, therefore not considered in this analysis.

The analysis of PC1-PC2 biplot diagram (Fig. 4A) clearly shows a similar elemental relationship, when compared with the correlation matrix (Fig. 3):

1) the Zn has a diametrically opposite behaviour in relation with the other elements in PC1 component, which agrees with the negative correlation with all the elements;
2) Ti and V have similar behaviour with the positive quadrants of PC1 and PC2 (the most significant);
3) PC1 component is mostly influenced by the variability of Cr, Fe, Co and Mn, indicating that those elements present similar geochemical behaviour;
4) Ni, As and Cu, have positive PC1 and negative PC2, having less influence in the principal components;
5) although with a particular behaviour, Pb has similarities with Ni, As and Cu, which is also in agreement with the correlation values previously presented.

PC1 was mapped (Fig. 4B), dividing observations in scores according to their distinct variability: Zn related population (PC1 < -2), low variability population (-2 ≤ PC1 < 5) and Cr. Fe Mn, and Co influenced population (PC1 ≥ 5). The pattern of the scores of PC1 reflects with good precision the dependence of the variables on the local geological features, namely the quartz-carbonate rock and gossan, the mine tailings and the felsic rock (PC1 ≥ 5), the host rocks (-2≤ PC1 <5) and a zone (PC1 < -2) coincident with the Zn anomalous values (c.f. Fig. 3D).

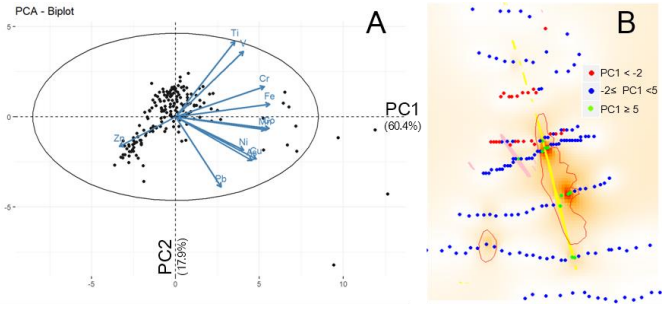


Figure 4. A) Principal components PC1-PC2 biplot. Variables as arrows and observations as dots; B) map of PC1 scores. The background map is the idw map for Cu (Fig. 3C).

Figura 4. A) Mapa das duas primeiras componentes principais (PC1-PC2); B) mapa dos valores de PC1. O mapa de fundo é o mapa de interpolação idw do Cu (ver Fig. 3C).

4.3. Hierarchical clustering

Hierarchical clustering (hclust) is an unsupervised learning method to classify and agglomerate data from observations and variables. This method uses distances between measured values to assign each observation/variable to a certain cluster (Kassambara, 2017). For this study the "Canberra" method to calculate the distance, was used, combined with the "ward.d2" agglomeration to separate the groups (Murtagh and Legendre, 2014). Hclust was applied to the same set of elements used in

PCA analysis. The values were used in a scaled mode removing the different magnitude biases.

Table 2. A) Variability and elements assigned to the first three principal components (PC1, PC2 and PC3); B) factor loadings for each principal component.

Tabela 2. A) Variabilidade e respetivos elementos químicos atribuídos aos três primeiros componentes principais (PC1, PC2 e PC3); B) pesos dos fatores para cada componente principal.

(A)

Component	Variability (%)	Elements
PC1	60.4	Fe, Co, Mn, Cr, Cu
PC2	17.9	Ti, Pb, V
PC3	6.6	Zn

(B)

Elements	PC1	PC2	PC3
Ti	0.2342584	0.51050496	0.24991828
Mn	0.3550499		
Fe	0.3726983		0.17962851
V	0.2686686	0.44087710	
Cr	0.3510982	0.20346900	0.27928922
Co	0.3706444		0.10732247
Ni	0.2690596	-0.22478934	
Cu	0.3192635	-0.28495968	
Zn	-0.2220279	-0.20009777	0.89061164
As	0.3023914	-0.29595767	
Pb	0.1796366	-0.47268535	

To synthesize the analysis from hclust results, a heatmap was created (Fig. 5A). The heatmap comprises two dendrograms: the top dendrogram corresponds to clustering based on the variables (geochemical elements), whereas, the left dendrogram represents the clustering based on the observations. The cells are coloured based on the Z-score value and the vertical broken lines represent the Z-scores.

The top dendrogram (Fig. 5A) divided three distinct clusters: C1) Ti, V, Fe and Cr; C2) Zn; and C3) Mn, Co, Cu, As, Ni, Pb.

The left dendrogram, constructed from observations (Fig. 5A), also allowed the division of three clusters, where:

- K1 (n=49) - dominated by samples with negative Z-score values for C1, generally with very high Z-score values of C2 and intermediate positive Z-score values for C3, namely for Cu, As and Pb;
- K2 (n=50) - dominated mostly by intermediate to high Z-score values of C1, very low negative Z-score values for most of C2 and variable intermediate Z-score values for C3;
- K3 (n=78) - characterized by high positive Z-score values for C1 elements (namely Ti and V), Z-score values for C2 are highly variable and dominant negative Z-score values for C3 namely in Cu, As, Ni and Pb.

The spatial distribution of these clusters (Fig. 5B) reveals a not very sharp distinction between K1 and K2, thus being unable to clearly separate between the recognized geological features, contrarily to what was previously observed using PCA. Nevertheless, the K1 defines the alignment of the mineralized vein (including the northern part of mineralized vein), the majority of the mine tailings and the felsics, it is also expressed randomly in the country rocks. K3 is well represented in the already defined Zn rich zone.

4.4. K-means clustering

An alternative method of unsupervised learning was tested. K-means finds the centroids of multidimensional data, assigning each observation to the nearest centroid (Hartigan and Wong, 1979; Mohamad and Usman, 2013). The number of centroids is pre-determined and corresponds to the anticipated number of clusters as defined in the. In the unsupervised learning techniques, the number of clusters (k) that group the data must be predefined. This work applies the function proposed by Malika *et al.* (2014) for determining the optimal number of clusters

In this case each observation is attributed to a cluster and therefore the heatmap rows divide each group (Fig.6A). The variables (chemical elements) are organized by groups with the same approach. The cells are colour coded based on the Z-score and its values are within each cell.

K-means was applied to the same data set of the previous methods. The created K-means method heatmap (Fig. 6A) separates the observations (rows) in three clusters:

- Ka (n=12) show positive high Z-score values for Fe, Cu, Mn and Co and somewhat lower, but significant, positive Z-score values for As, Ni and Cr and negative Z-score values for Zn;
- Kb (n=118) shows more averaged values (close to Z-score = 0), but with positive values for Ti and V (0.42 and 0.45, respectively);
- Kc (n=47) is characterized by positive high Zn Z-score values and negative to low values of the remaining elements, highlighting for the significant low values of Ti and V;
- The resulting map (Fig. 6B) highlights the relation of Ka with the quartz-carbonate mineralized zone and gossan, presenting one observation in the vicinity of the felsic rock. The most represented cluster (Kb) maps essentially the country rocks. Kc maps highlight the higher values of Zn.

5. Discussion

5.1. Feasibility of machine learning methods in the Mociços soils

Machine learning techniques regarding Mociços soil geochemistry data were tested to determine their usefulness for geochemical exploration. The combined behaviour of the chemical elements in these types of environment are susceptible to distinguish cryptic anomalies, not previously found within the studied area.

PCA shows that Fe, Co, Mn and Cr (and, to a lesser extent, Cu and As), mainly explains the first principal component (60.4% variability), representing their opposing behaviour to Zn (Fig. 4A). The PC1 elements (Fig. 4A) are interpreted as the contribution of the vein and gossan structures, as well as the mine tailings (Fig. 4B). The other observations ($-2 \leq PC1 < 5$) should represent the host rocks geochemical background.

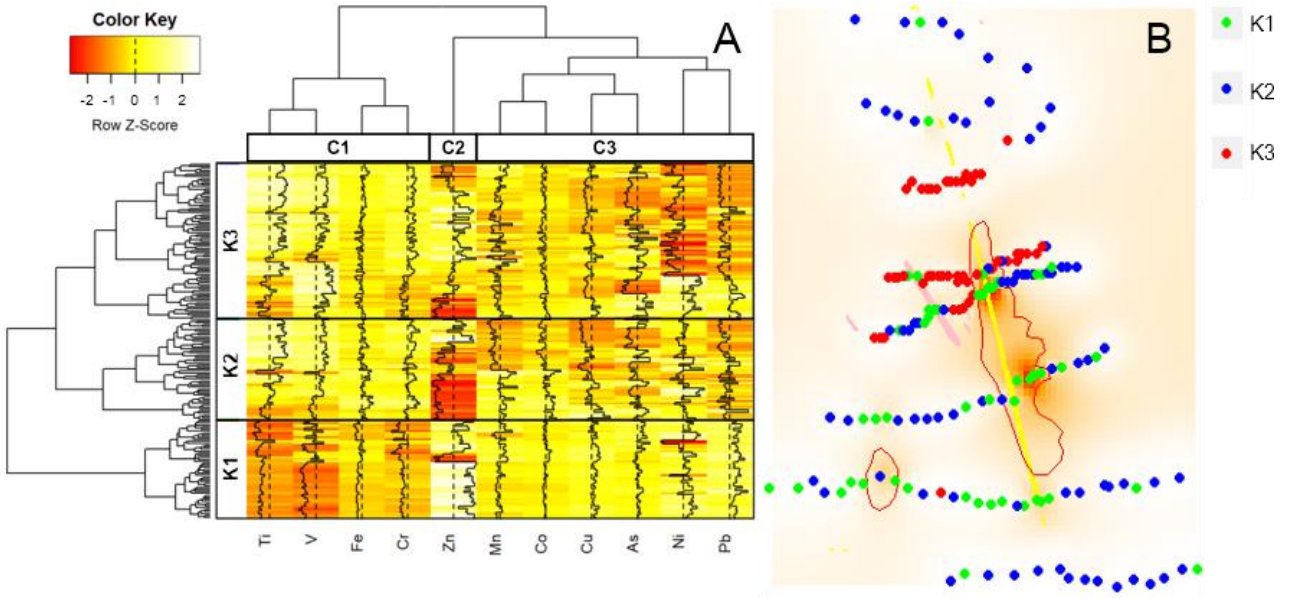


Figure 5. A) Heatmap for the data using hclust for three clusters (K1, K2 and K3): left dendrogram with the observations ($\mathcal{V}$). Top dendrogram represent the studied variables (C1, C2 and C3; chemical elements). $Z-score=\frac{v-\bar{s}}{\sigma}$ calculated for each analysis from K1, K2 and K3; B) map of hclust for three observations-dependent clusters (K1, K2, K3).

Figura 5. A) *Heatmap* dos dados utilizando hclust para 3 clusters (K1, K2 e K3): dendrograma da esquerda com as observações ($\mathcal{V}$). Dendrograma de topo representa as variáveis (C1, C2 e C3; elementos químicos). $Z-score=\frac{v-\bar{s}}{\sigma}$ calculado para cada análise de K1, K2 e K3; B) mapa do *hclust* para três clusters dependentes das observações (K1, K2 e K3).

Hclust method grouped variable-dependent (C1, C2, C3; analysed elements) and observation-dependent clusters (K1, K2, and K3; soil analysis). The outcome of the variable-dependent clusters revealed that Ti, V, Fe and Cr group together, as well as Mn, Co, Cu, As, Ni and Pb, whereas Zn do not tend to group with any element. Observation-related clusters were mapped (Fig. 5B), well-defining the geochemical contrasting soil sector (K3) of the northern Mociços area, as also shown by PCA. The same observations that define the mineralized vein, gossan, mine tailings and felsics (K1) are also located unsystematically in the host rocks (K2), with no apparent geological meaning. This way, it is assumed that hclust method based on observations is not as sharp as PCA, merely revealing the more contrasting clusters. However, hclust is the only clustering method that distinguished the northern part of the mineralized vein and well-defined the felsic rocks, although presenting some randomly distributed values.

The K-means clustering method for the variables, when divided in three clusters (Fig. 6A), separates similar behaviours for: Ka) Fe, Cr, Cu, Mn, Co, Pb, Ni and As; Kb) Ti and V; and Kc) Zn. The corresponding map for this division (Fig. 6B) clearly identifies the difference between (Ka) the mineralized structure, gossan and mine tailings, as well as the felsic rocks from the (Kb) surrounding soils. The same distinct sector in the northern area (Kc) is also demarked, as corroborated by previous machine learning methods.

All the unsupervised methods tested in this work (PCA, hclust and K-means) can separate the observations related to the mineralized structure, gossan and mine tailings from those related to the geochemical background and host rocks, although with some differences related to each method limitations. Observations also corroborate summary statistics study and single element interpolation maps, determined with IDW (Fig. 3).

Three groups can thus be defined in the Mociços area, as determined by the PCA, hclust and K-means methods. Despite slightly different results, the integrated study of the different methods was clear in the distinction of the Mociços soils in:

1) The mineralized vein, gossan, adjacent mine tailings and felsic rocks (particular geological features, previously recognized);
2) The surrounding country rocks corresponding to the majority of the representative area (in the central-south and northernmost sectors of the area);
3) A geochemically contrasting soil sector in the northern part of the Mociços mine.

5.2. Geochemical considerations

In the Mociços area, the Cu anomaly is well-constrained to the quartz-carbonate vein, gossan structure and the mine tailings, as shown by the geochemical maps and the machine learning methods presented. Besides Cu, As is the only other element that is considerably enriched in the mineralized area, which might be due to the presence of primary As-bearing minerals, as described for other SBMB similar deposits (*e.g.* Mateus *et al*., 2003; 2013). Despite the Pb and Ni correlation with these two elements, no high values were detected.

The strong association of Fe, Mn, Co and Cr, elements that are, generally, immobile in surficial conditions, are probably related with the presence of gossan-forming minerals, such as Fe(-Mn) oxides (and hydroxides) and other residual mineral phases. Ti and V strong correlation can be assigned to their identical geochemical characteristics (ionic radii and cationic charge), thus likely being incorporated in the same mineral phases and tend to display immobile behaviour during weathering and oxidising events.

The spread Zn distribution can be assigned to its mobile behaviour in surficial (low temperature, high oxygenation) conditions, therefore depleted in the vicinity of the quartz-carbonate lodes, nevertheless it will be further investigated.

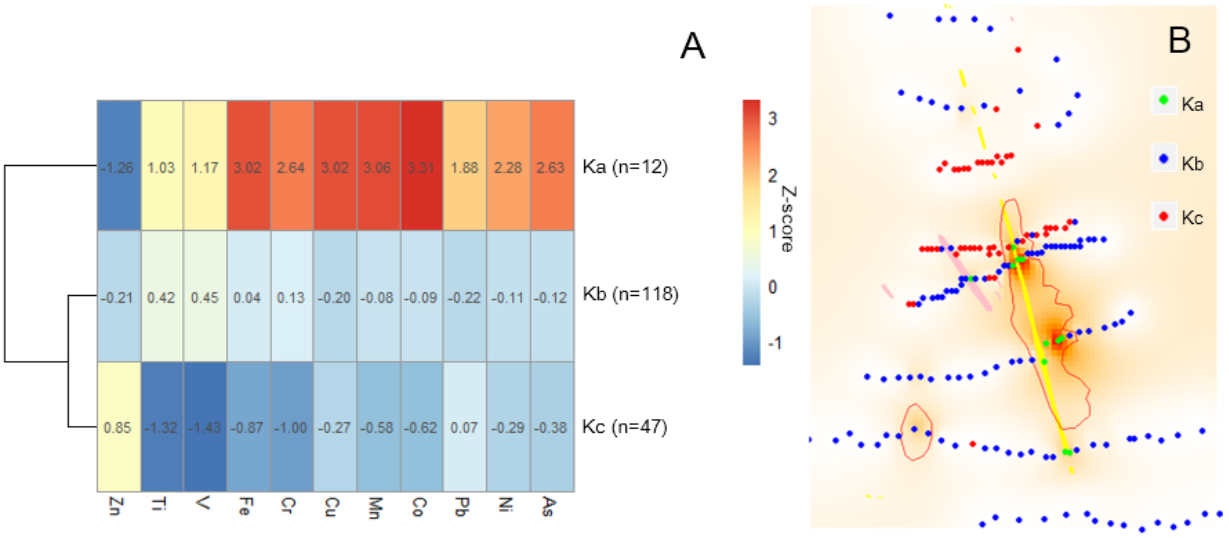


Figure 6. K-means clustering for three clusters, emphasizing the mineralization elemental association: A) Heatmap and B) spatial distribution of k-means.

Figura 6. Aglomeração por K-média para três clusters. A) heatmap e B) distribuição espacial dos três clusters K-média.

6. Final remarks

The pXRF provided fast and low-cost soil geochemistry analytical results in the Mociços mine. The combined use of summary statistics, data interpolation mapping and three different machine learning techniques allowed to distinguish the regional geochemical background from other geological features, including 1) the previously exploited mineralized vein, gossan structures, mine tailings and mapped felsic rocks, and 2) a chemically different sector in the northern region. The obtained results are plausible with known geological information and current interpreted metallogenic model for the SBMB.

The occurrence of anomalous values of the mineralization-related clusters in the felsic rocks can provide some perceptions on the genesis of the Mociços mineralization and should be further investigated.

In addition, the highly contrasting (Zn enriched) soil analyses in the northern Mociços area also requires better geological explanations, thus a thoroughly chemical and mineralogical characterization should be performed.

Acknowledgments

This work is a contribution to the project ZOM3D: ALT20-03-0145-FEDER-000028, funded by Alentejo 2020 through the FEDER/FSE/FEEI. The authors also acknowledge the funding provided to the Institute of Earth Sciences through the COMPETE 2020 project (UIDB/GEO/04683/2020) under the reference POCI-01-0145-FEDER-007690. The authors acknowledge the two revisors for its valuable insights, Amélia Paula Marinho e Maria João Batista.

References

Araújo, A., Piçarra de Almeida, J., Borrego, J., Pedro, J., Oliveira, T., 2013. As regiões central e sul da Zona de Ossa-Morena. In: Dias, R., Araújo, A., Terrinha, P., Kullberg, J. C. (Eds.), *Geologia de Portugal*, Escolar Editora, Lisboa, **1**: 509-549.

Chen, S., Grunsky, E.C., Hattori, K., Liu, Y., 2015. Principal Component Analysis of Geochemical Data from the REE-rich Maw Zone, Athabasca Basin, Canada. *Geological Survey of Canada, Open File* 7689, 24. DOI: 10.4095/295615.

Grunsky, E.C., 2001. A program for computing RQ-mode principal components analysis for S-plus and R. *Computers & Geosciences*, **27**: 229-235. DOI: 10.1016/S0098-3004(00)00111-4.

Grunsky, E.C., 2010. The interpretation of geochemical survey data. *Geochemistry-Exploration Environment Analysis*, **10**(1): 27-74. DOI: 10.1144/1467-7873/09-210.

Hall, G., Page, L., Bonham-Carter, G., 2013. *Quality Control Assessment of Portable XRF Analysers: Development of Standard Operating Procedures, Performance on Variable Media and Recommended Uses*. Phase II. Canadian Mining Industry Research Organization (Camiro) Exploration Division, Project 10E01 Phase I Report.

Hartigan, J. A., Wong, M. A., 1979. Algorithm AS 136: A K-means clustering algorithm. *Applied Statistics*, **28:** 100-108. DOI: 10.2307/2346830.

Kassambara A., 2017. *A pratical guide to cluster analysis in R. Unsupervised machine learning*. STDHA editions, 187.

Kolde, R., 2015. *pheatmap: Pretty heatmaps [Software]*. https://CRAN.R-project.org/package = pheatmap

Lemière, B., 2018. *A Review of pXRF (Field* Portable X-ray Fluorescence) Applications for Applied Geochemistry. Journal of *Geochemical Exploration*, **188**: 350-363. DOI: 10.1016/j.gexplo.2018.02.006.

Lentz, D. R., 2003. Geochemistry of sediments and sedimentary rocks: historical to research perspectives. Inorganic geochemistry of sediments and sedimentary rocks: evolutionary considerations to mineral deposit-forming environments. Edited by DR Lentz. Geological Association of Canada, *GeoText*, **4**: 1-6.

Maia, M., Moreira, N., Mirão, J., Noronha, F., Nogueira, P., 2019a. Fluid inclusions study of Cu-rich deposits from Sousel-Barrancos metallogenic belt (Ossa-Morena Zone, Portugal). *Acta Mineralogica-Petrographica Abstract Series (ECROFI 2019)*, **10**: 79.

Maia, M., Moreira, N., Vicente, S., Mirão, J., Noronha, F., Nogueira, P., 2019b. Fluid constraints in the Mociços Cu deposit (Ossa-Morena Zone, Portugal). *Acta Mineralogica-Petrographica Abstract Series (ECROFI 2019)*, **10**: 78.

Malika, C., Ghazzali, N., Boiteau, V., Niknafs, A., 2014. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *J. Stat. Softw*., **61**: 1-36.

Mateus A., Matos J.X., Rosa C., Oliveira V., 2003. Cu-ores in quartz-carbonate veins at Estremoz-Alandroal and Barrancos-Sto Aleixo regions (Ossa Morena Zone): a result of Late-Variscan hydrothermal activity. VI Congresso Nacional de Geologia, Lisboa (Portugal), *Ciências da Terra (UNL)*, Lisboa, CD-ROM, **V:** F90-F93.

Mateus, A., Munhá, J., Inverno, C., Matos, J. X., Martins, L., Oliveira, D., Jesus, A., Salgueiro, R., 2013. *Mineralizações no sector português da Zona de Ossa-Morena*. *In*: Dias, R., Araújo, A., Terrinha, P., Kullberg, J. C. (Eds.), Geologia de Portugal, Escolar Editora, Lisboa, **1**: 577-619.

Matos, J.X., Filipe, A. (Coord.), 2013. *Carta de Ocorrências Mineiras do Alentejo e Algarve, Escala 1:400 000*. Laboratório Nacional de Energia e Geologia, 1st Edition. ISBN: 978-989-675-029-9.

Mohamad, I. B., Usman D., 2013. Standardization and Its Effects on K-Means Clustering Algorithm. *Research Journal of Applied Sciences, Engineering and Technology*, **6**(17): 3299-3303. DOI: 10.19026/rjaset.6.3638.

Moreira, N., Vicente, S., Maia, M., Nogueira, P., Araújo, A., 2017. Controlo estrutural de mineralizações de Cobre na Mina dos Mociços (Zona de Ossa-Morena); dados preliminares. In: Moreira, N., Silva, V., Paiva, A. (Eds.), *Livro de actas do VII Congresso Jovens Investigadores em Geociências*, LEG 2017, Estremoz, 45-48.

Moreira, N., Vicente, S., Maia, M., Oliveira, R., Nogueira, P., Borges, F., Caldeira, B., Araújo, A., 2018. *Mineralização de Cobre na Mina dos Mociços (Zona de Ossa-Morena); uma abordagem transversal para a sua caracterização*. *Vulcânica, Volume Especial do X Congresso Nacional de Geologia*, **2018**(II): 237-240.

Murtagh, F., Legendre, P., 2014. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *Journal of Classification*, **31**: 274-295. DOI: 10.1007/s00357-014-9161-z.Oliveira, V., 1984. Contribuição para o conhecimento geológico-mineiro da região de Alandroal-Juromenha (Alto Alentejo). *Estudos Notas e Trabalhos, Serviço de Fomento Mineiro*, **XXVI**(1-4): 102-126.

Oliveira V., 1986. Prospecção de minérios metálicos a sul do Tejo. *Geociências*, **1**(1-2): 15-22.

Oliveira, J. T., Oliveira, V., Piçarra, J. M., 1991. Traços gerais da evolução tectono-estratigráfica da Zona de Ossa-Morena. *Cadernos do Laboratorio Xeolóxico de Laxe*, **16**: 221-250.

R Core Team, 2014. *R: A language and environment for statistical computing. R Foundation for Statistical*. Computing, Vienna, Austria. URL http://www.R-project.org/.

Taylor, S. R., McLennan, S. H., 1985. *The Continental Crust: Its Composition and Evolution*. Blackwell, Oxford, 312.

Vicente, S. (in prep.). *"Caracterização Geoquímica, Mineralógica e Petrográfica da Mina de Mociços".* MSc Thesis, Universidade de Évora.

Vicente, S., Maia, M., Moreira, N., Matos, J. X., Nogueira, P., 2018. Aplicação de FRX portátil na prospeção. Caso de estudo na Mina de Mociços (Cu), Zona Ossa Morena, Portugal. *Livro de Atas de XIV Congresso de Geoquímica dos Países de Língua Portuguesa e XIX Semana de Geoquímica*, 321-324.

Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W. H. A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., Venables, B., 2016. *gplots: Various R Programming Tools for Plotting Data*. R package version 2.17.0.

Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 224.