



# Multivariate analysis applied to X-ray fluorescence to assess soil contamination pathways: case studies of mass magnetic susceptibility in soils near abandoned coal and W/Sn mines

Jelena Milinovic · Patrícia Santos · Helena Sant'Ovaia · Aurora Futuro · Carlos M. Pereira · Bramley J. Murton · Deolinda Flores · Manuel Azenha

Received: 2 January 2024 / Accepted: 6 April 2024  
© The Author(s) 2024

**Abstract** Determining the origin and pathways of contaminants in the natural environment is key to informing any mitigation process. The mass magnetic susceptibility of soils allows a rapid method to measure the concentration of magnetic minerals, derived from anthropogenic activities such as mining or industrial processes, i.e., smelting metals (technogenic origin), or from the local bedrock (of geogenic origin). This is especially effective when combined with rapid geochemical analyses of soils. The use

of multivariate analysis (MVA) elucidates complex multiple-component relationships between soil geochemistry and magnetic susceptibility. In the case of soil mining sites, X-ray fluorescence (XRF) spectroscopic data of soils contaminated by mine waste shows statistically significant relationships between magnetic susceptibility and some base metal species (e.g., Fe, Pb, Zn, etc.). Here, we show how qualitative and quantitative MVA methodologies can be used to assess soil contamination pathways using mass magnetic susceptibility and XRF spectra of soils near abandoned coal and W/Sn mines (NW Portugal). Principal component analysis (PCA) showed how the first two primary components (PC-1 + PC-2) explained 94% of the sample variability, grouped them according to their geochemistry and magnetic susceptibility in to geogenic and technogenic groups. Regression analyses showed a strong positive correlation ( $R^2 > 0.95$ ) between soil geochemistry and magnetic properties at the local scale. These parameters provided an insight into the multi-element variables that control magnetic susceptibility and indicated the possibility of efficient assessment of potentially contaminated sites through mass-specific soil magnetism.

---

J. Milinovic (✉) · C. M. Pereira · M. Azenha  
Chemistry and Biochemistry Department, Faculty of Sciences, CIQ-UP, Institute of Molecular Sciences (IMS), University of Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal  
e-mail: jelena.milinovic@fc.up.pt

P. Santos · H. Sant'Ovaia · D. Flores  
Institute of Earth Sciences, Pole of University of Porto, 4169-007 Porto, Portugal

P. Santos · H. Sant'Ovaia · D. Flores  
Department of Geosciences, Environment and Spatial Planning FCUP, University of Porto, 4169-007 Porto, Portugal

A. Futuro  
CERENA, Faculdade de Engenharia da Universidade do Porto, Rua Dr Roberto Frias s/n, 4200-465 Porto, Portugal

B. J. Murton  
NOC, National Oceanography Centre, European Way, Southampton SO14 3ZH, UK

**Keywords** Metal contamination · Magnetic minerals · Geochemistry · XRF spectra · PCA and cluster analysis · Regression analyses

## Introduction

Soil magnetic susceptibility is a physico-chemical property that describes the magnetic response of soil to a uniform external magnetic field (Evans & Heller, 2003). The volume magnetic susceptibility ( $\kappa$ ) is the ratio of the magnetization ( $M$ ) to the magnetic field ( $H$ ). Since both parameters,  $M$  and  $H$ , have the same units ( $A\ m^{-1}$ ), their ratio, *i.e.*, the value obtained for the  $\kappa$  is dimensionless and difficult to compare with the values reported in the literature. Therefore, it is more convenient to use the mass or specific magnetic susceptibility ( $\chi$ ), which represents the ratio between the  $\kappa$  and density ( $\rho$ ) of soil samples. Results of  $\chi$ , which refer to the mass of soil samples (expressed in units of reciprocal density,  $m^3\ kg^{-1}$ ), are more readily available in the literature and more useful for comparison (Mullins, 2006). Conventional measurements of soil mass magnetic susceptibility ( $\chi$ ) usually performed under a weak magnetic field, give more reliable results because at lower levels of  $H$  ( $< 1000\ A\ m^{-1}$ ), the magnetization is directly proportional to the applied magnetic field (Evans & Heller, 2003).

The magnetic properties of soil are primarily due to the presence of ferromagnetic minerals (magnetite, hematite, and hematite) and some other minerals with paramagnetic behavior (e.g., biotite and ilmenite) (Dekkers, 1997; Magiera et al., 2006; Mello et al., 2020). Other minerals, such as silicates, have less influence on the soil magnetic susceptibility, while their presence and composition can indicate the lithology of the soil's parent rock. In addition to the bedrock contribution, anthropogenic contamination can also affect soil magnetic susceptibility through pathways such as atmospheric deposition and precipitation, road and rail traffic, fires, disposal of waste materials, cement and metallurgical industries, as well as particulates that can be very magnetic and which could indicate a potential risk for human health (Dellibecque et al., 2022; Lu et al., 2012). Soil magnetic susceptibility is, therefore, a sensitive parameter that depends on numerous factors which, if they can be separated and identified, can be used as an indicator of environmental contamination (Pan et al., 2019; Wang et al., 2018).

In the case of mining sites and their surrounding environments, the increase in magnetic susceptibility is positively correlated with the presence of

high contents of certain metal species. For example, Attoucheik et al. (2017) found a positive correlation between magnetic susceptibility and Pb (at concentrations lower than  $500\ mg\ kg^{-1}$ ) in the topsoils near a Pb–Zn mine (in Algeria); Nahan et al. (2023) observed that magnetic susceptibility was correlated with Fe and Sc concentrations in gold mining sites (Indonesia); Ma et al. (2023) concluded that Pb and Co pollution, found in dust from the extremely arid city of Dunhuang (NW China) and mainly derived from coal mining and coal combustion, coincided with mass specific magnetic susceptibility of  $135 \times 10^{-8}\ m^3\ kg^{-1}$ .

The underlying processes behind the correlations between magnetic susceptibility and the presence of these contaminants are not straightforward. Recent studies showed that correlations between magnetic susceptibility and metal contaminants were largely related to geogenic or anthropogenic (technogenic) compositional specifics of the respective mining site (Bourliva et al., 2017; Chakraborty et al., 2023; Magiera et al., 2023; Shaheen & Iqbal, 2018; Szuszkiewicz et al., 2021). For example, influx of paramagnetic mineral particles containing the correlated elements, or the adsorption of correlated elements onto paramagnetic dust particles.

In this research, we report how multivariate analysis (MVA) identifies complex univariate data correlations between the mass magnetic susceptibility and the concentration of individual elements, for soils from one or more mining sites of similar type. Our study uses X-ray fluorescence (XRF) spectroscopy data acquired by portable XRF instruments, that are widely used among geologists and environmental scientists to characterize a wide range of elemental concentrations (Agyeman et al., 2021; Bosco, 2013; Jang, 2010). Given the previous studies, it is reasonable to expect that XRF spectroscopic data encrypt information directly related to magnetic susceptibility, which could be deconvolved using regression MVA. Detailed explanations of the advantages of applying MVA to XRF spectroscopic data in soil survey (e.g., for rapid soil characterization), are given elsewhere (dos Santos et al., 2020; Milinovic et al., 2023; Morona et al., 2017; Panchuk et al., 2018; Rocha et al., 2019). If successful, MVA applied to both XRF and magnetic susceptibility data from soils could provide insights into the pathways that

enable contaminants to enter the environment. Furthermore, the MVA approach would allow field scientists to estimate potentially contaminated sites via mass magnetic susceptibility ( $\chi$ ) alone, enabling rapid scanning for the contamination.

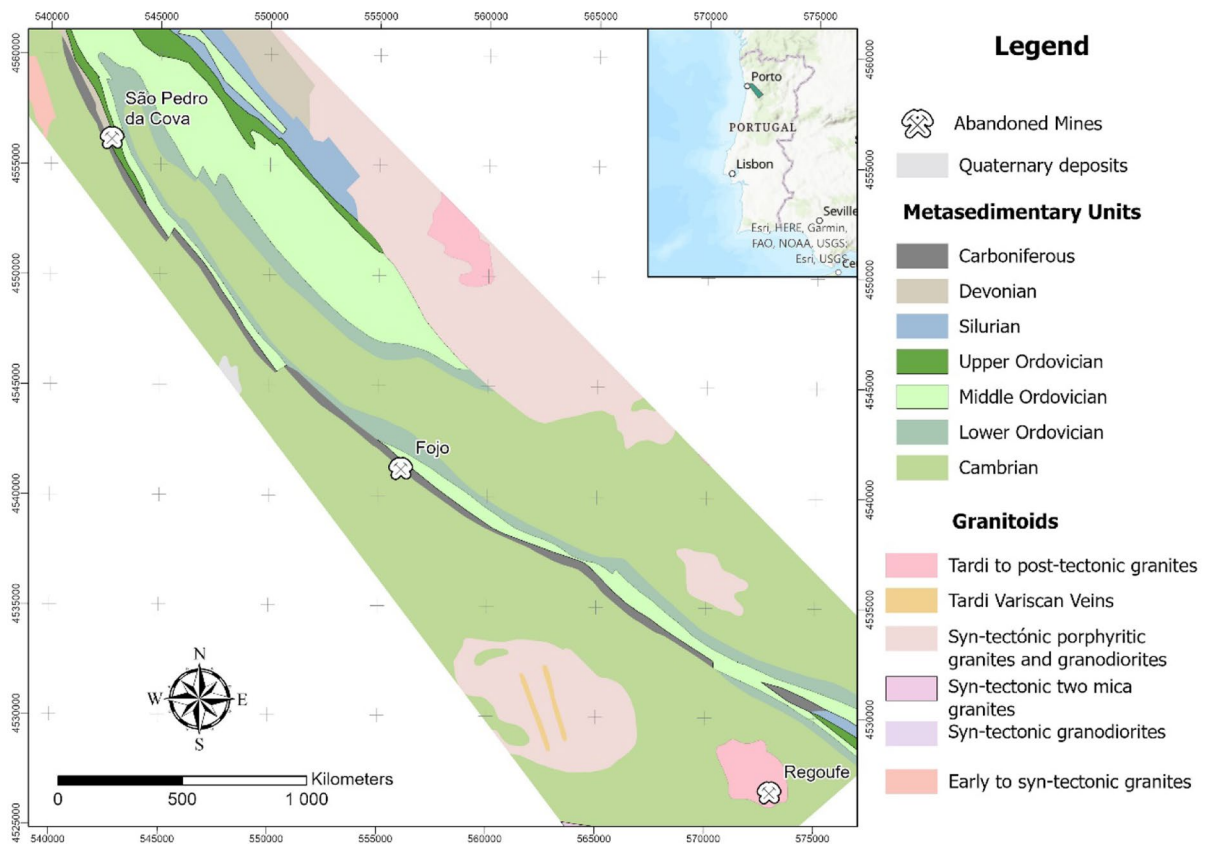
The main objectives of this work were the following: (i) to obtain qualitative geochemical signatures encrypted in soils using XRF spectroscopic data that can be correlated with magnetic properties of the soil; (ii) grouping of different soil types according to geochemical signatures by classification MVA; (iii) application of regression MVA to obtain XRF–MVA models, the interpretation of which can clarify the soil contamination pathways. These objectives were achieved by testing soils from three abandoned mining areas (in NW Portugal): two coal mines (São Pedro da Cova and Fojo) and a W/Sn mine (Regoufe).

### Materials and methods

#### Geological setting of study areas and soil sampling

Soil samples used in this study were taken from three different study areas in NW Portugal (Fig. 1). These sites comprise of three abandoned mines: São Pedro da Cova (41° 09' 25" N; 08° 30' 06" W), a coal mine located in Gondomar; the Fojo mine (41° 09' 54" N; 08° 20' 28" W), which is part of the Pejão coal mining complex, located in Castelo de Paiva, and the Regoufe (40° 52' 44" N; 08° 08' 06" W), a wolframite/cassiterite mine located in Arouca.

Both São Pedro da Cova and Fojo mines exploited anthracite A (ISO 11760:2005) from the Douro Carboniferous Basin which was later used in energy production. This sedimentary basin, dating from the Upper Pennsylvanian (Correia et al., 2018; Lemos de Sousa & Wagner, 1983) was the most important coal reserve in Portugal, extending



**Fig. 1** Geologic setting of the studied pilot areas: São Pedro da Cova, Fojo, and Regoufe

NW–SE, from São Pedro de Fins to Janarde, over a distance of approximately 53 km and between 30 and 250 m thick (Pinto de Jesus, 2019). The São Pedro da Cova mine was closed in 1972, after working for almost two centuries. The Fojo mine was exploited until 1994, after which the spoil-heaps burned between 2007 and 2019, after being ignited by wild fires in the surrounding vegetation. The combustion of the spoil-heaps was only extinguished after the intervention of a state company that applied a cooling agent combined with the remobilization of the materials affected (Santos et al., 2023).

Regoufe is a wolframite/cassiterite (W/Sn) mine, that started operating in 1915 (motivated by World War 1). During World War 2 it saw a significant increase in production but was closed in the 1970s. The mine is located along the SE border of the Regoufe pluton, a 6 km<sup>2</sup>, NW–SE elongated circular muscovite-albite porphyritic granite batholith, with a whole-rock Rb–Sr age of 280 ± 8 Ma (Fig. 1). Several tungsten-rich deposits occur in this granite body, where they are host in 10–20 cm thick cross-cutting quartz veins (Pereira et al., 2007; Van Gaans et al., 1995) and contain wolframite, cassiterite, arsenopyrite, and minor amounts of pyrite, sphalerite, apatite, and beryl (Sant’Ovaia et al., 2023).

In São Pedro da Cova and Fojo, the soil samples collected are essentially the result of the pedogenesis of underlying metasedimentary rocks of Cambrian to Carboniferous ages, while in Regoufe, the soil is derived from the granitic substrate (Fig. 1).

The mines at the study sites have undergone little intervention or restoration since their abandonment, apart from the extinguishing of the burning spoil-heaps at Fojo. Therefore, the composition and geochemical signature of surrounding soil reflects a complex origin with input from the local bedrock and input from anthropogenic effects including mining activities.

A total of 90 heterogeneous topsoil samples were collected from three mining areas and chosen as a purposeful database that will serve better to test the applicability of MVA (at local vs regional level of tested mining areas). Each sample was collected from a single point at a depth of less than 20 cm (topsoil) from the three abandoned mining areas: São Pedro da Cova (n=50), Fojo (n=25), and Regoufe (n=15) (Fig. 2).

It is not possible to detect all 50 points in this picture above!

There are 24 points and not 25 like expected!

I recommend that you put letters (a), (b), (c) in these pictures.

The soil samples surrounding the waste pile in São Pedro de Cova were collected systematically at approximately 100 m distance (Fig. 2a). The sampling sites covered a complex area including forest, urban, and industrial zones, as well as small landfills.

The soil samples from the Fojo mining area were taken from different forest areas in and around the abandoned coal mine waste pile (Fig. 2b). Some of the soil samples were collected from the areas affected by remobilized burnt spoil-heaps (EA, 1–10), as well as from outside the affected areas (ENA, 1–5). A number of samples were also collected from an area downslope from the spoil areas (SJ, 1–5), and which were affected by the runoff from the waste pile. Reference samples were taken uphill of the waste pile, in the forest area (B, 1–5), thus representing a geochemical baseline for the soils unaffected by the mining deposition. The samples in the Regoufe mining area were taken from a valley next to the old mining facilities (Fig. 2c). Extensive descriptions of the soil sampling, pre-treatments, and homogenization of representative soil samples are provided elsewhere (Sant’Ovaia et al., 2023; Santos et al., 2023).

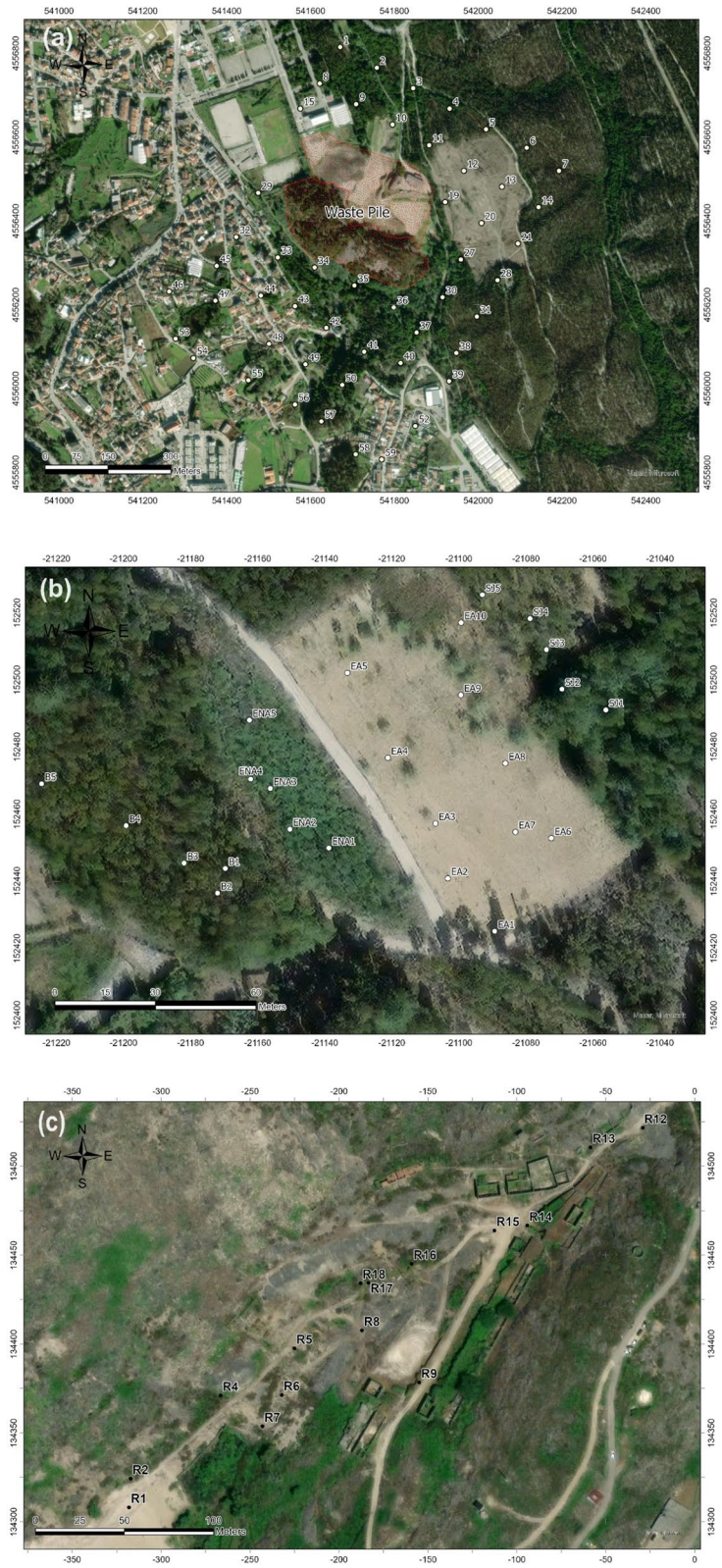
#### Analysis of soil samples

The soil samples from three mining areas were characterized as previously explained (Ribeiro et al., 2011; Sant’Ovaia et al., 2023; Santos et al., 2023). The mass magnetic susceptibility was determined in studied soil samples by their exposure to a magnetic field of 300 A m<sup>-1</sup> on a KLY-4S Kappabridge instrument (Advanced Geoscience Instruments Co., Czech Republic). The KLY-4S Kappabridge is a high-sensitive laboratory instrument, operating at a frequency of 875 Hz and power consumption of 45 VA. Three measurements were performed for each sample and the mean values are presented in the results.

The same soil samples were analysed by energy dispersive X-ray fluorescence spectrometer (X-MET 7500, the handheld XRF Mining Analyzer, Oxford Instruments, UK) equipped with a rhodium anode X-ray tube working at 40 kV voltage and 10 µA current (Figueiredo et al., 2019). Spectroscopic data



**Fig. 2** Maps of the sample locations in studied pilot areas: São Pedro da Cova (a), Fojo (b), and Regoufe (c)



at broad energy level (1–35 keV) were obtained in Mining Mode, and averaged values across the three entire-spectra were extracted for each sample and exported for further treatment by MVA.

Due to XRF spectral interferences (overlapping signals) and the high limit of detection for the elements analysed by XRF, the concentrations of soil elements of geochemical importance in this study (As, Fe, Pb, Rb, Sn, Sr, Ti, Y, Zn, Zr) were independently obtained by inductively coupled plasma mass spectrometry (ICP-MS). These values served as references in the explanation and discussion of the obtained XRF–MVA results. Official ICP-MS analyses were performed in Accredited Laboratory N° 270 in Canada (Bureau Veritas Labs, Vancouver). Samples were digested in HF-HClO<sub>4</sub>-HNO<sub>3</sub> and subsequently the dissolved residue was analyzed by ICP-MS. Blanks, random duplicate samples and three certified reference materials (OREAS: 25A, 45E, and 45H), were used for the quality control of analytical procedure. The limits of detection (LOD) for the analyzed elements were in the following order: 0.000002% for Pb; 0.00002% for As, Zn, and Zr; 0.00001% for Rb, Sn, and Y; 0.0001% for Sr; 0.001% for Ti; and 0.01% for Fe. Full details of the sample preparation, operating conditions and quality control of the ICP-MS analyses are described in detail elsewhere (Sant’Ovaia et al., 2023; Santos et al., 2023).

#### Multivariate analysis applied to X-ray fluorescence spectroscopic data

Quantitative data encrypted in XRF spectra (K-alpha and K-beta peak energies for each of the elements of interest) of studied topsoil samples (n=90) were analyzed by multivariate analysis (MVA) using the software application Unscrambler. Unsupervised classification MVA, principal component analysis (PCA) and cluster analysis were applied to entire XRF spectroscopic data to reduce their dimensionality (by excluding sample outliers), and identifying correlated parameters to arrange the samples into empirical groups (Brereton, 1990). Classification and regression MVA were applied to the raw and transformed XRF spectra of the studied samples.

The XRF spectra were transformed using several techniques, such as normalization by the peak areas (that correlate with each element, background correction, Savitzky-Golay smoothing, and orthogonal

signal correction (OSC). Applied transformation tests showed that only OSC, a mathematical treatment that removes factors unrelated to the sample variable concentration and orthogonal to the reference values of the variable of interest, was suitable for reducing side effects and larger variations in spectra (Blanco et al., 2001; Malmir et al., 2019; Zhang et al., 2013). For soil geochemistry-magnetism correlation, partial least square (PLS) regression was applied to XRF spectroscopic data versus the mass magnetic susceptibility ( $\chi$ ). Calibration was performed with 2/3 of samples while validation of the PLS regression was tested with the remaining 1/3 of the samples, as the most effective method for dealing with heterogeneous samples of smaller size (Kardanpour et al., 2014; Martens & Naes, 1998; Martins et al., 2010; Nengsih et al., 2019). The relevant parameters used to evaluate the regression were the correlation coefficient ( $R^2$ ), which describes the linearity between the XRF-spectra and reference mass magnetic values in the plot, and the root mean square error (RMSE), which is a measure of dispersion for the samples about the regression line. These two parameters are optimal indicators for describing the goodness of fit in linear regression (Engelen et al., 2004). To identify the most important geochemical variables of the best fitted PLS regressions, the magnitudes of the weighted regression coefficients ( $B_w$ ) were examined (Ng et al., 2022). The regression coefficients ( $B_w$ ) were calculated incorporating one factor (or principal component) thus summarizing the relationship between the most important variables and the response.

For PCA, cluster analysis, OSC transformation, and PLS regression analyses, the nonlinear iterative partial least square (NIPALS) algorithm was chosen, as the most appropriate algorithm for the smaller sample size (Nengsih et al., 2019). Outliers were filtered according to Hotelling’s  $t^2$  statistic ( $P < 0.05$ ).

## Results and discussion

### Variability of mass magnetic susceptibility in studied mine soil samples

The studied soil samples show a wide range of mass magnetic susceptibility ( $\chi$ ), *i.e.*, from 2.15 to 1003, from 13.7 to 749, and from 2.17 to 477 ( $\times 10^{-8} \text{ m}^3 \text{ kg}^{-1}$ ) in the São Pedro da Cova (Santos

**Table 1** Obtained values of the mass magnetic susceptibility ( $\chi$ ) in studied soils from São Pedro da Cova, Fojo, and Regoufe mining areas

Soil samples origin	n	Range of $\chi$ ( $\times 10^{-8} \text{ m}^3 \text{ kg}^{-1}$ )
São Pedro da Cova	50	2.15–1003
Fojo (label)	25	13.7–749
Burned waste piles (EA)	10	285–749
Unburned waste piles (ENA)	5	20.4–98.4
Downhill area (SJ)	5	134–244
Forest (B)	5	13.7–65.8
Regoufe	15	2.17–477

et al., 2023), Fojo, and Regoufe (Sant’Ovaia et al., 2023) mining areas, respectively (Table 1). These results show that samples in São Pedro da Cova (n=50) had the highest variability in magnetic susceptibility, which differed by four orders of magnitude. For the samples surrounding the self-burning coal waste pile, the highest values were observed upstream of the waste pile pointing to possible origins for the ferrimagnetic minerals not related to the mine waste (Santos et al., 2023).

The values of mass magnetic susceptibility in the Fojo samples from the burned mine waste area (EA, Fig. 2b), showed the highest values, ranging from 285 to 749 ( $\times 10^{-8} \text{ m}^3 \text{ kg}^{-1}$ ). The next highest values, in a range from 134 to 244 ( $\times 10^{-8} \text{ m}^3 \text{ kg}^{-1}$ ) are from samples (SJ) taken downhill and in the runoff affected area from the burned mine waste. Samples from unburned waste piles (ENA), have values that are an order of magnitude smaller than the burned mine waste samples, i.e., from 20.4 to 98.4 ( $\times 10^{-8} \text{ m}^3 \text{ kg}^{-1}$ ) (Table 1). The reference soil samples from the forest uphill and away from the mined area (B) showed the lowest values, ranging from 13.7 to 65.8 ( $\times 10^{-8} \text{ m}^3 \text{ kg}^{-1}$ ), and overlapping with the unburned waste pile samples (ENA).

The high variability of magnetic susceptibility in the studied soil samples from the three abandoned mining areas (São Pedro da Cova, Fojo, and Regoufe) indicated that ferromagnetic minerals were present in different concentrations in the studied soil samples. Variability appears to relate to location with respect to sources of contamination, especially for the Fojo area, where there is better control on the location of the contaminated and uncontaminated sites (Flanders, 1994; Santos et al., 2023). To better explain the

variability of the soil mass magnetic susceptibility, and hence the source of potential contamination, we made a MVA of the qualitative geochemical information encoded in the XRF spectra and compared them to the reference ICP-MS data.

### Soil geochemistry near abandoned coal and W/Sn mines

In our study, the reference ICP-MS results showed wide ranges of element concentrations in the studied soil samples (Table 2). Arsenic was found at similar concentrations in samples from both the coal mine regions (São Pedro da Cova and Fojo), while in the W/Sn (Regoufe) mine As contents reached a value of up to 1%. In continental crust, As is found primarily in an inorganic form (arsenic compounds) associated with igneous and sedimentary rocks (Biney et al., 2022), at average concentrations of 0.0002% and 0.0013%, respectively (Turekian & Wedepohl, 1961). Hence, the elevated As concentrations are likely to reflect contamination from mineralisation and mining. Besides the geogenic contribution, in non-mining areas, As can originate from several anthropogenic sources such as fertilizers applied to agricultural soils. This may explain the As levels in the São Pedro da Cova urban zone (Fig. 2). In these samples, As was found at levels up to 0.006% (Table 2).

Iron has very high concentrations in the coal and W/Sn mine areas, reaching maximum concentrations

**Table 2** Obtained ICP-MS concentrations of elements in studied soils from São Pedro da Cova, Fojo, and Regoufe mining areas

Element	Range of element concentrations (%)		
	São Pedro da Cova n=50	Fojo n=25	Regoufe n=15
As	0.001–0.006	0.001–0.008	0.075–1
Fe	1.08–9.39	1.17–7.09	0.85–5.99
Pb	0.002–0.016	0.002–0.012	0.001–0.198
Rb	0.007–0.029	0.004–0.017	0.043–0.082
Sn	0.0001–0.002	0.0002–0.001	0.005–0.010
Sr	0.003–0.069	0.005–0.014	0.002–0.005
Ti	0.077–0.520	0.154–0.375	0.042–0.066
Y	0.0005–0.002	0.001–0.002	0.0005–0.001
Zn	0.002–0.030	0.001–0.015	0.008–0.050
Zr	0.004–0.017	0.006–0.012	0.002–0.007



of 9.39%, 7.09%, and 5.99% in samples from the São Pedro da Cova, Fojo, and Regoufe mines, respectively. Although the raw coal has relatively low concentrations of Fe, combustion (e.g., of the self-burning waste piles) can concentrate non-volatile elements, such as iron oxide (Santos et al., 2023). Other metallic elements (Pb, Rb, Sn, Sr, Y, Zn and Zr) were found in similar concentrations in the coal mine samples, reaching a maximum level of up to 0.07%. Similar concentrations were observed in the Regoufe W/Sn mine area (with the exception of Pb, which was found in concentrations up to 0.2%). The maximum Ti content in the soil was equal to 0.52% and 0.37%, in São Pedro a Cova and Fojo, respectively, while in the Regoufe soils its concentration was less than 0.07%.

#### Qualitative geochemical information encrypted in soil X-ray fluorescence spectra

Iron and some other elements, typical in the studied mining areas (As, Pb, etc.) were detected by XRF spectral analysis, at characteristic energy levels (Fig. 3a).

The raw XRF-spectra showed the most intense peaks at 6.4 ( $K_{\alpha}$ ) and 7.1 keV ( $K_{\beta}$ ), corresponding to Fe, while Rb, Sr, Zr, and Y at 13.4, 14.1, 14.9, and 15.7 keV, showed respective smaller peaks due to their lower concentrations in these samples (Table 2). The intense peaks of As and Pb at 10.5 ( $K_{\alpha}$ ) and 11.7 keV ( $K_{\beta}$ ) were typical for the samples collected in the Regoufe (W/Sn) mining area, where these elements were found in high concentrations (e.g., As up to 1%). Broad peaks at 19.2, 20.2, 21.1, and 25.2 keV correspond respectively to Mo, Rh, Pd, and Sn.

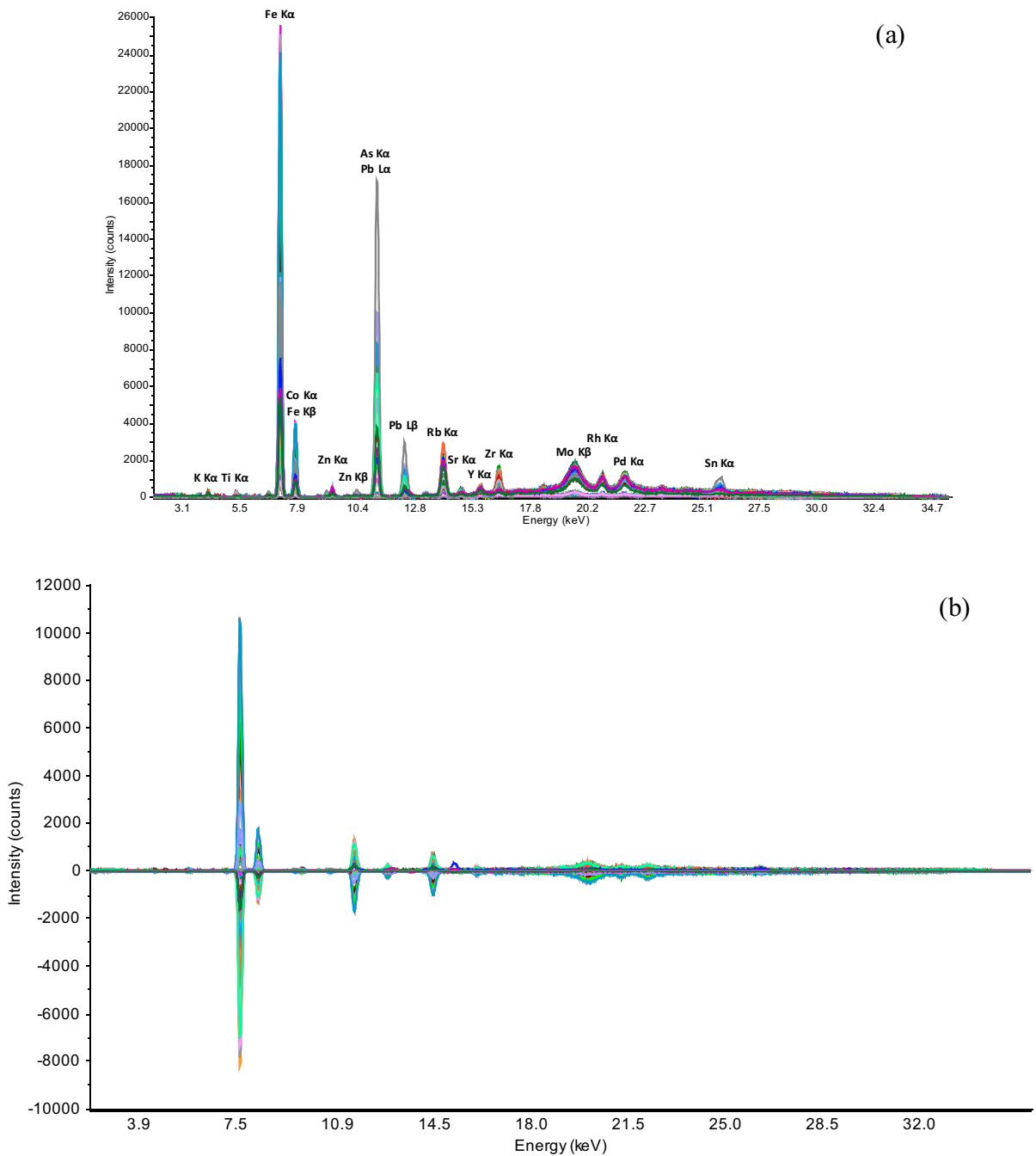
After applying the OSC transformation (relative to the reference soil mass magnetic susceptibility values), the spectral differences have been minimized (Fig. 3b). The OSC pretreatment reduced the noise level and acted as a filter by eliminating all the information irrelevant to the target variable, *i.e.*, mass magnetic susceptibility ( $\chi$ ) and interfering features, improving the quality of spectroscopic data in the entire energy-range. As a result, peaks unrelated to mass magnetic susceptibility became much smaller (compared to their original intensities), but still present in the spectra (Blanco et al., 2001; Wu et al., 2006).

#### Soil grouping based on geochemical signatures by principal component analysis

After applying PCA to raw XRF spectroscopic data of the studied soil samples ( $n=90$ ) three groups, each corresponding to a different mining area, have merged (Fig. 4a). Two-dimensional PCA also performed well, with PC-1 explaining quantitatively (79%) the variability of the spectroscopic data. The São Pedro da Cova and Fojo samples are distributed along the PC-1 axis, while the Regoufe samples are distributed along PC-2. The homologous distribution of samples from São Pedro da Cova and Fojo can be explained by the similarity in the geochemical signatures of these coal mining areas, *i.e.*, the soils of both pilots originated from similar metasedimentary units (Fig. 1) and were influenced by similar coal mining effects. In contrast, the Regoufe samples (W/Sn mine) had a distinctive signature because the soils derived from the Regoufe granite, and were later affected by tungsten mining activities, which form different rock type to the sedimentary coal sequences. The PCA also discriminates between a subset of samples from the Fojo mining area, designated as the self-burning waste pile (EA), and bounded by a dashed line in Fig. 4a, and other samples from the same mining area. The separate distribution of Fojo samples from the self-burning mine waste pile can be explained by the samples having a higher magnetic susceptibility (Ribeiro et al., 2014; Santos et al., 2023). Multivariable geochemical information, encrypted in the XRF spectra of the studied soil samples, can distinguish the different origins of the soil samples including those with distinctive and magnetic susceptibility.

The PCA applied to the raw XRF spectra (Fig. 4a) shows better discrimination of different mine samples, explaining 94% of the sample variability (PC-1 = 79% + PC-2 = 15%). The results of the PCA applied to OSC-XRF spectra (Fig. 4b) showed that the first component explained quantitatively the variability of samples (PC-1 = 100%) and all soil samples are essentially distributed along the PC-1 axis, less separated into discrete groups of mine origin, but rather in relation to the magnetic susceptibility values. The self-burning waste pile soil samples (bounded by a dashed line) from the Fojo were more anti-correlated with other samples from the same mining area groups (Fig. 4a, b). This also coincides with the higher magnetic



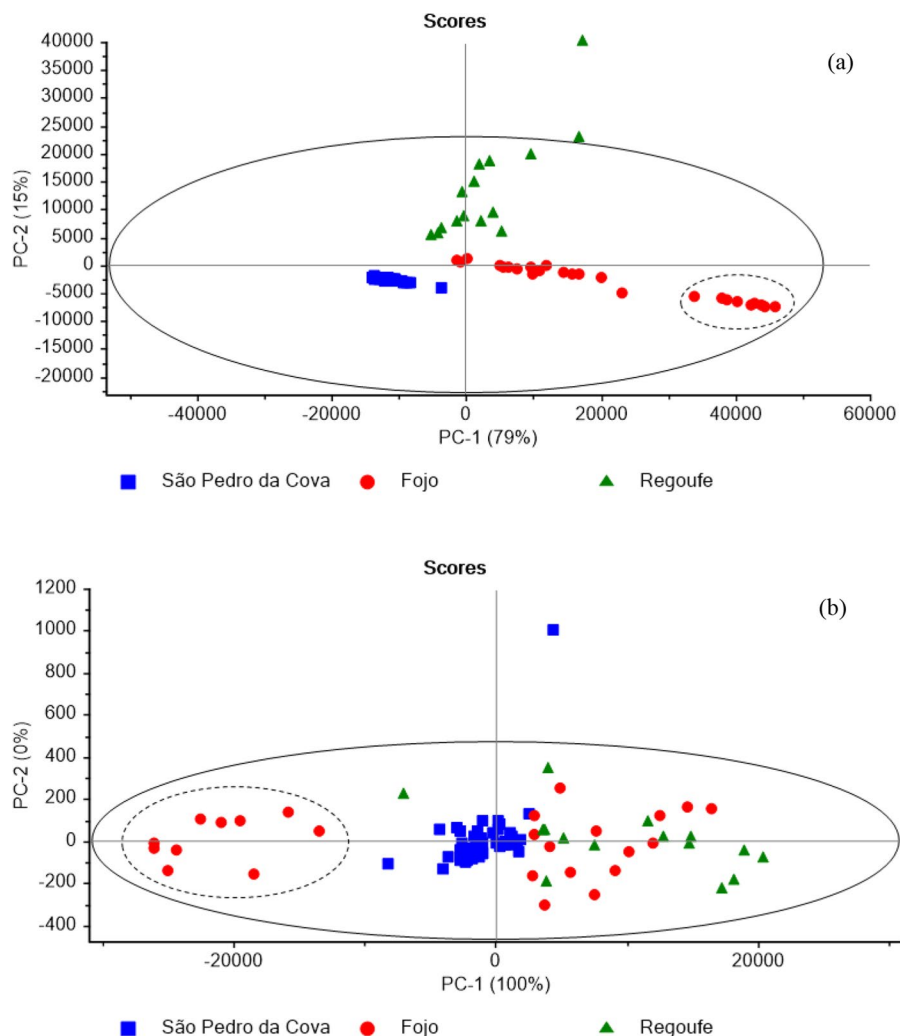


**Fig. 3** Mean-plot of spectra from studied soils (n=90) around three mining areas: raw XRF (a) and transformed OSC-XRF spectra (b)

susceptibility of the burnt waste-pile soils and hence indicates their unique origin. One sample from the São Pedro da Cova (19) was considered a significant outlier (visualized as the point in the top

far away from the scores plot, Fig. 4b), due to its very high Fe content, equal to 9.39% (Table 2), and hence was excluded before PLS regression analysis was run.

**Fig. 4** PCA applied to: raw XRF (a) and OSC-XRF (b) spectra of studied soils (n=90) around three mining areas (São Pedro da Cova, Fojo, and Regoufe). The scores plot of the first two principal components shows that ten self-burning subset of samples of Fojo (bounded by a dashed line) were separated into an independent group

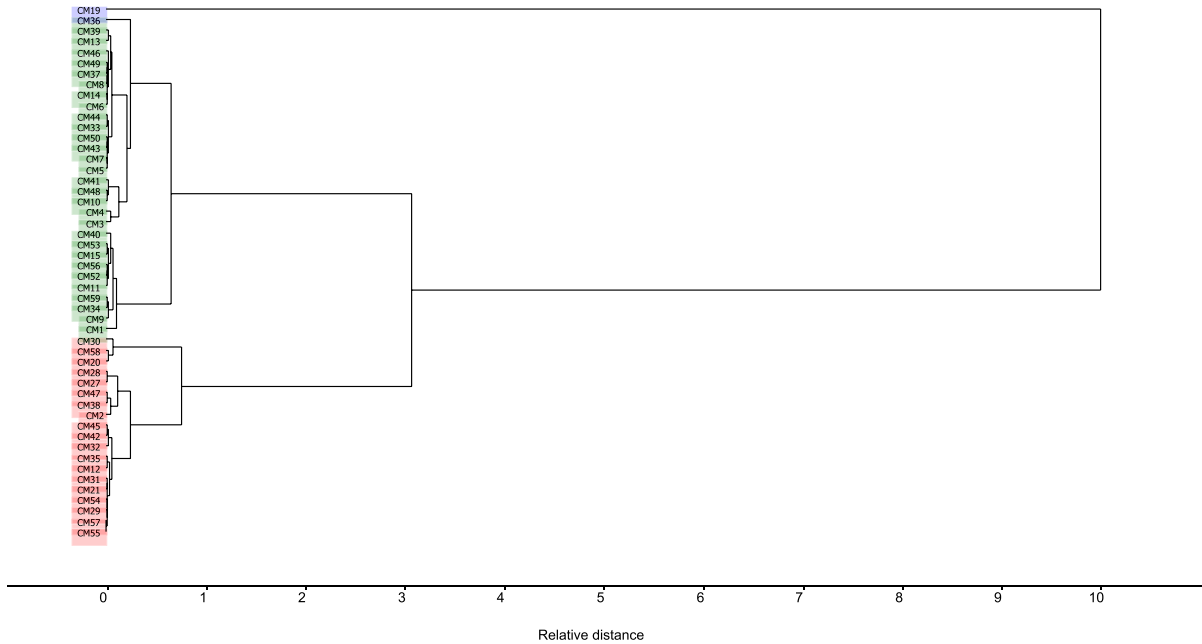


#### Cluster analysis of soils from São Pedro da Cova mining area

The interpretation of geochemical signatures in soil samples from São Pedro da Cova mine area (including forest, urban and industrial zones and small landfills), was facilitated by cluster analysis directly applied to raw XRF spectroscopic data of the entire set of soil samples. The procedure of cluster analysis is designed to discover natural clusters within a complex data set that would not otherwise be apparent. In this case, the dendrogram (produced using squared Euclidean distance) confirmed that one sample (19) was grouped as an independent cluster (i.e., an outlier, Fig. 5) as previously shown by PCA (Fig. 4b). Despite its maximum Fe content, an outlier

sample showed a very low mass magnetic susceptibility ( $\chi = 11 \times 10^{-8} \text{ m}^3 \text{ kg}^{-1}$ ), which can be explained by the presence of antiferromagnetic mineral, for. ex., goethite (Guyodo et al., 2003; Mello et al., 2020).

The two remaining statistically meaningful clusters, labelled as clusters 1 and 2 are shown in different colors (green and red, respectively) in Fig. 5. Cluster 1 (n=30) was composed of samples from the forest and small landfill areas of the São Pedro da Cova mine sites (e.g., 3–11, 13–14, 39–41, 48–52, etc., Fig. 2a), while cluster 2 (n=19) contained samples mostly around the waste pile, and from urban and industrial areas (e.g., samples 2, 21–22, 27–32, 54–55, etc., Fig. 2a). The cluster 1 (n=30) was characterized with lower Fe (1.08–4.05%) while the cluster 2 (n=19) had samples with similar or



**Fig. 5** Cluster analysis using squared Euclidean distance, applied to samples (n=50) of São Pedro da Cova complex area. After excluding one sample (19), as an outlier, a dendro-

gram illustrates the arrangement of two major clusters based on the relative distance of the corresponding XRF spectra: a cluster 1 in green (n=30) and the cluster 2 in red (n=19)

higher levels of Fe (3.45–5.77%). Both clusters showed a very broad magnetic susceptibility ranges (cluster 1:  $2.15\text{--}1003 \times 10^{-8} \text{ m}^3 \text{ kg}^{-1}$ ; cluster 2:  $20.8\text{--}893 \times 10^{-8} \text{ m}^3 \text{ kg}^{-1}$ ), which suggested that there were differences in the Fe-mineralogy and geochemistry, hence in corresponding contributions to the absolute values of magnetic susceptibility in these samples. Cluster analysis could deconvolve the hidden correlations between the most important variables (of XRF signatures) and the two representative

clusters will be used in further regression analyses to test the correlation with magnetic susceptibility.

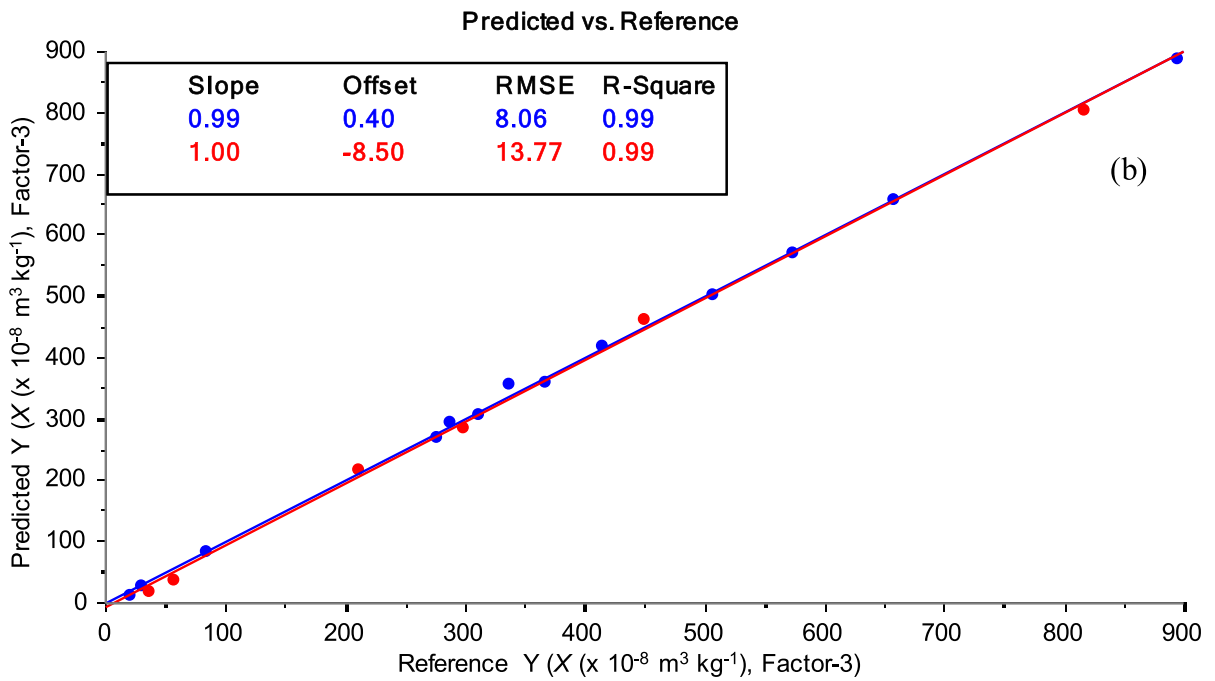
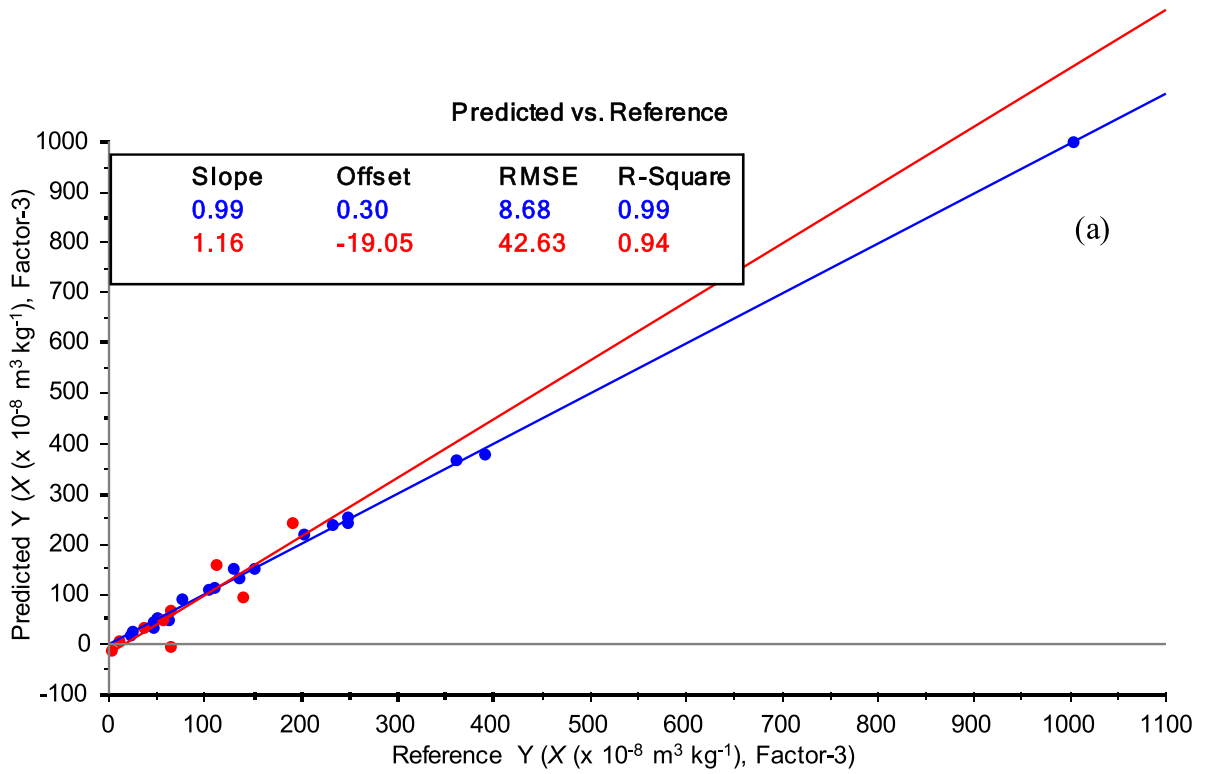
Geochemistry versus mass magnetic susceptibility of the soils from mining areas

Unlike univariate correlation which failed to yield any meaningful relationship, (e.g., Fe concentration vs soil mass magnetic susceptibility), multivariate regression (using a set of explanatory variables

**Table 3** PLS regression parameters for the mass magnetic susceptibility ( $\chi$ ) in the soils around studied mining areas: São Pedro da Cova, Fojo, and Regoufe

Mining area	n	LV	$R^2$		RMSE	
			cal	val	cal	val
<i>Individual test</i>						
São Pedro da Cova	49	6	0.97	0.62	44	147
Cluster 1	30	3	0.99	0.95	8.7	43
Cluster 2	19	3	0.99	0.99	8.1	14
Fojo	25	4	0.99	0.99	5.8	19
Regoufe	15	3	0.99	0.99	2.3	7.0
<i>Paired two-sample test</i>						
São Pedro da Cova—cluster 1 + Fojo	55	6	0.96	0.47	44	186
São Pedro da Cova—cluster 2 + Fojo	44	6	0.99	0.58	15	164

n, the total number of soil samples used for calibration and validation PLS regressions



from the entire XRF spectra vs soil mass magnetic susceptibility) has a better chance at identifying



◀**Fig. 6** PLS regression (calibration—blue, external validation—red) between OSC-XRF spectra and mass magnetic susceptibility ( $\chi$ ) content in soil samples from São Pedro da Cova obtained after cluster analysis: cluster 1 (a) and cluster 2 (b). The model used three latent variables (factor=3) to optimize covariance between data sets. The inserted boxes contain statistical parameters that explain the goodness of regression with high values of  $R^2$  for calibration ( $>0.99$ ) and validation ( $>0.95$ ), and low values of RMSE of calibration (equal to  $8.1\text{--}8.7 \times 10^{-8} \text{ m}^3 \text{ kg}^{-1}$ ) and validation (equal to  $14\text{--}43 \times 10^{-8} \text{ m}^3 \text{ kg}^{-1}$ ). Slope and offset (intercept) define the linear relationship between two variables of the regression line. The closer the slope is to 1, the data are better correlated. The offset is the intercept of the line when the X-axis is set to 0 and its value is nearly 0

hidden correlations between significant variables (Golia & Diakouloukas, 2022; Morrissey & Ruxton, 2018). To test for multivariate regression between XRF spectroscopic data and the mass magnetic susceptibility ( $\chi$ ) in different soils, samples originating from the three different mining areas were tested individually and as paired two-sample (coal mine) tests by PLS regression. Table 3, shows the PLS results and the main parameters obtained for both the calibration and validation datasets for each of the combinations tested. In cases where the multivariate regressions were good for both calibration and validation data sets ( $R^2 > 0.90$ ), XRF spectroscopic data were used alone to predict the origin of magnetic susceptibility, i.e., identifying the potential pathway for soil contamination.

Considering the individual mining areas, it can be noted that the PLS regressions gave satisfactory results for the calibration datasets in all three localities ( $R^2 > 0.97$ ), but the validation results were less satisfactory in samples ( $n=49$ ) from São Pedro da Cova ( $R^2=0.62$ ). This means that the calibration can be well matched within individual similar soils, but if the soil type is significantly different, the calibration may show poor results. This can be solved by making the selection of samples as similar as possible to the calibration data, or by having calibration for as many types of soil as possible. The high difference in values for calibration and validation data, can be explained mainly due to the heterogeneity of these samples, i.e., their high variability in the mass magnetic susceptibility ( $\chi$ ) caused by different factors (Ribeiro et al., 2014; Santos et al., 2023). As previously mentioned, these samples had a wide range of mass magnetic susceptibility ( $\chi=2.15\text{--}1003 \times 10^{-8} \text{ m}^3 \text{ kg}^{-1}$ )

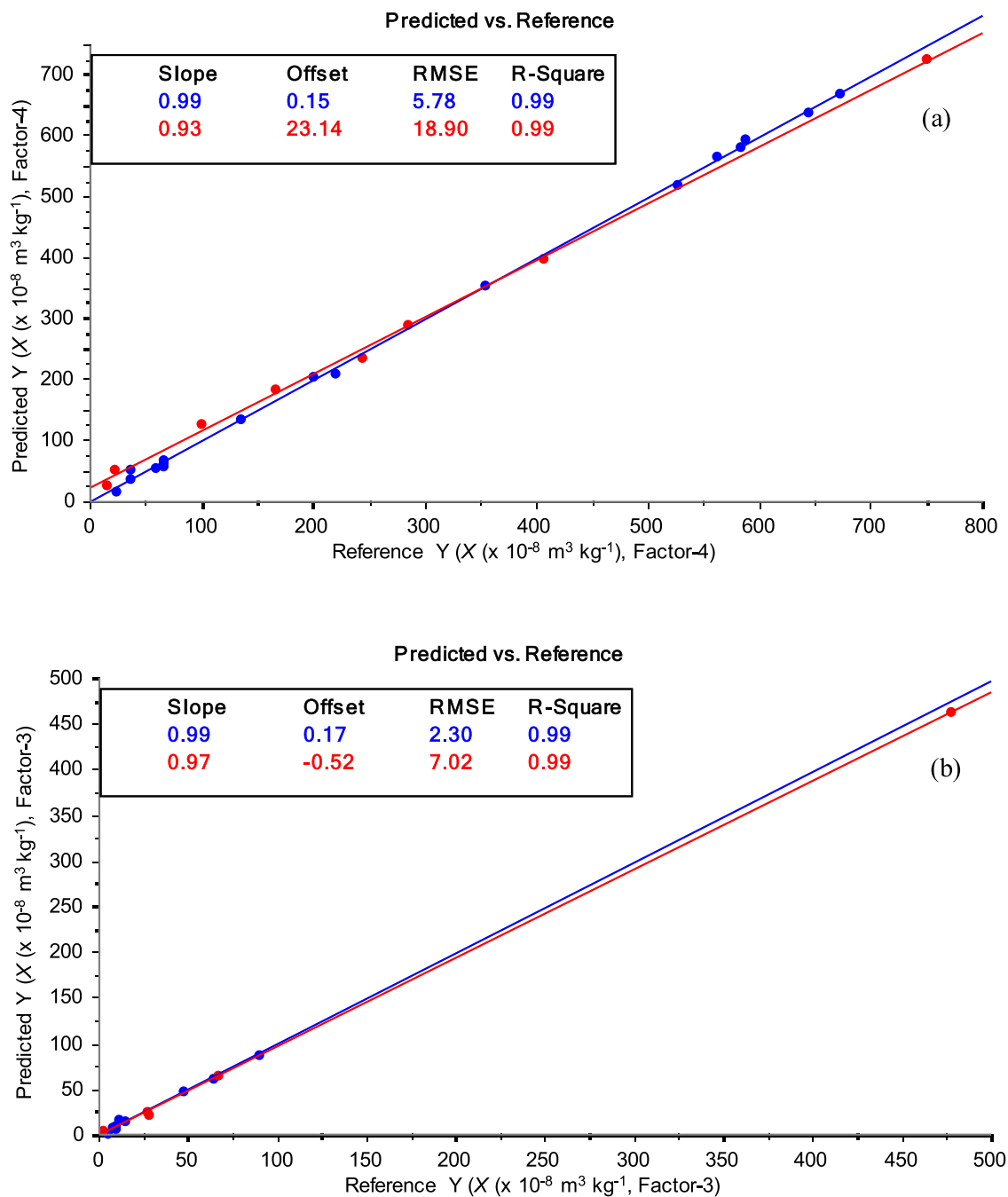
not only due to particular pedogenetic factors (i.e., possible higher content of ferromagnetic minerals in the host rocks) but also due to anthropogenic activities, through release of industrial waste, incineration of fossil fuel, and utilization of pesticides and fertilizers (a regular agricultural practice in urban and industrial zones of São Pedro da Cova (Santos et al., 2023). To improve the regression, a subset of samples from São Pedro da Cova: cluster 1 ( $n=30$ ) and cluster 2 ( $n=19$ ) were used (Fig. 6a, b, respectively) and regression analysis statistics gave satisfactory results (Table 3). These results confirm the previous qualitative soil clustering in which two independent groups are identified, each characterized by homogenous (similar) geochemistry. The significantly improved regressions, with higher  $R^2$  and lower RMSE for validation datasets, confirm the two subsets (cluster 1 and cluster 2) which are used for the following paired two-sample tests.

The calibration in Fojo (Fig. 7a) or Regoufe (Fig. 7b) soil samples was confirmed by external validation results with high correlations ( $R^2 > 0.99$ ) and low RMSE ( $< 19 \times 10^{-8} \text{ m}^3 \text{ kg}^{-1}$ ) in these local mining areas (Table 3).

Regarding paired two-sample tests, after including soil samples (cluster 1 or 2) from São Pedro da Cova and Fojo sites, PLS regression could not provide a good regression to satisfy analytical purposes. Both PLS regressions were characterized by low  $R^2$  (0.47–0.58) and high RMSE ( $164\text{--}186 \times 10^{-8} \text{ m}^3 \text{ kg}^{-1}$ ) for external validation. As an example, Fig. 8 shows the regression of the combined test with São Pedro da Cova (cluster 2) and Fojo.

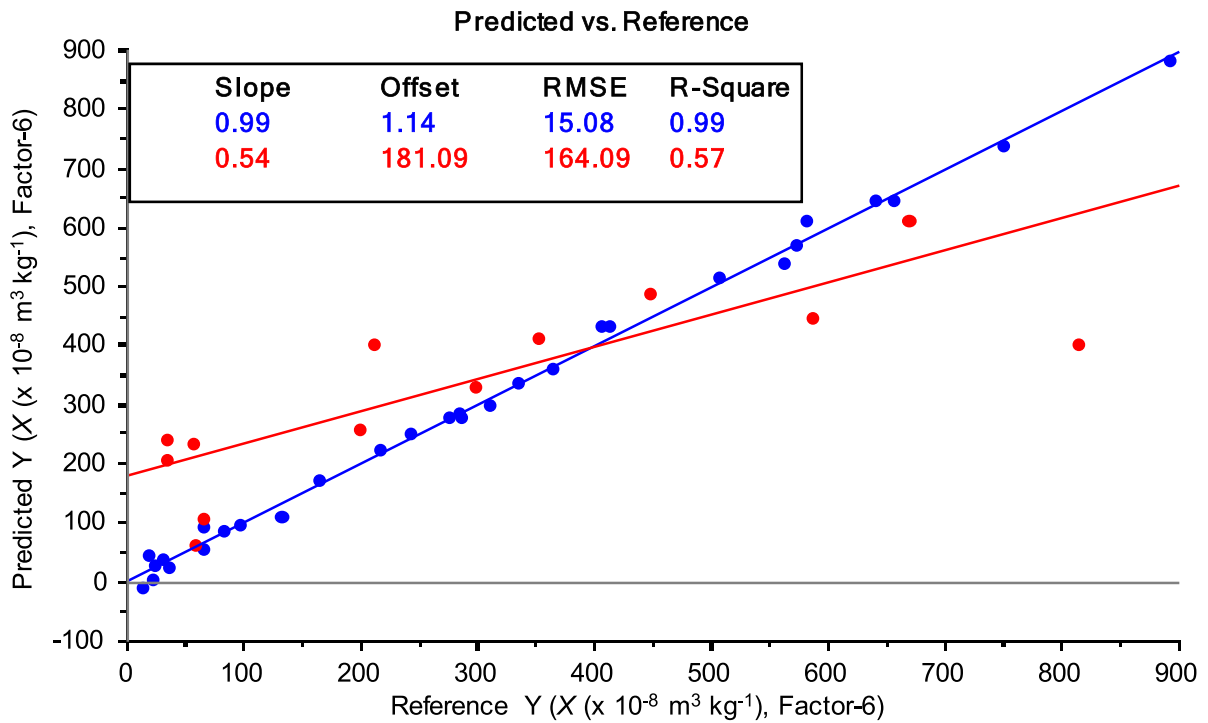
In the paired two-sample tests, the regressions were poor with increased errors due to the greater variety of the mass magnetic susceptibility, which implies a reduced reliability of XRF–MVA at a regional level, including samples from the two coal mining areas of different locations. The obtained PLS regression results show that the applicability of the XRF to monitor soil magnetism origin depends not only on the range of values of soil mass magnetic susceptibility ( $\chi$ ) but also on their source, i.e., the environmental characteristics of each single mining area.

These results support the ability to use mass magnetic susceptibility and XRF spectral data to discrimination between soils of different origins—including influences from geological and anthropogenic



**Fig. 7** PLS regression (calibration—blue, external validation—red) between OSC-XRF spectra and mass magnetic susceptibility ( $\chi$ ) content in soil samples from: Fojo **(a)** and Regoufe **(b)**. 3 latent variables (factor=3) were used to maxi-

mize covariance between data sets. Inserted boxes contain the statistical parameters of the regression (the explanation of the relevant parameters used for statistical analysis of data is given in detail in Fig. 6)



**Fig. 8** PLS regression (calibration—blue, external validation—red) between OSC-XRF and mass magnetic susceptibility ( $\chi$ ) content in soil samples from São Pedro da Cova (cluster 2) and Fojo ( $n=44$ ). 6 latent variables (factor=6) were used to

maximize covariance between data sets. Inserted box contains the statistical parameters of the regression (the explanation of the relevant parameters used for statistical analysis of data is given in detail in Fig. 6)

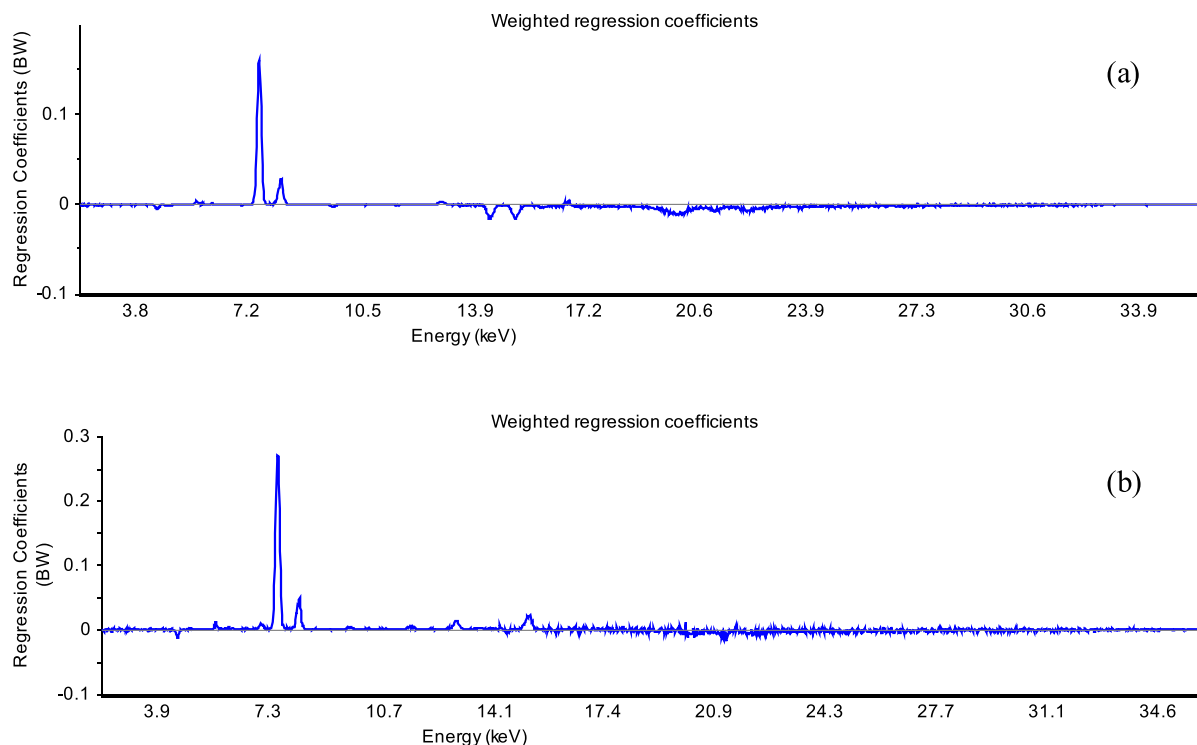
processes, e.g., contamination from agricultural and industrial activities versus natural environmental inputs. The unsupervised classification MVA can successfully distinguish samples, where there are at least two specific clusters with statistically significant regressions at the local scale.

#### Important variables of XRF spectra for soil magnetism assessment

The most important variables of XRF spectra (geochemical features) for the soil magnetism at local mining areas are shown in Figs. 9 and 10. As expected, Fe (6.4 and 7.1 keV) is the variable that dominantly affects the soil magnetism in all studied areas. In addition to Fe, in the regression with cluster 1 ( $n=30$ ; samples from São Pedro da Cova), Rb (13.4 keV) and Sr (14.1 keV) were negatively weighted (with small regression coefficients), showing their influence on magnetic susceptibility (Fig. 9a). For the magnetic properties of soils of

cluster 2 ( $n=19$ ), Ti (4.5 keV) and Sr (14.1 keV) are important positive variables (Fig. 9b). Both clusters 1 and 2 have similar concentrations of Ti, 0.13–0.52% and 0.12–0.46%, respectively, while Sr has a higher content in cluster 1 (up to 0.07%) compared to cluster 2 (up to 0.02%). The ratios between total Fe and Ti concentrations are higher in cluster 2 (Fe/Ti=9–48), almost twice compared to cluster 1 (Fe/Ti 3–25), which might correspond to different iron–titanium oxides, occurring as ferromagnetic minerals in the soils. Similarly, Fe/Sr ratios are as much as 2.5 times higher in cluster 2 (up to 1463) compared to cluster 1 (up to 582). These results indicate different origins for the magnetic properties due to distinct mineralogy in these subsets of soil samples, which were sampled in different zones of the mining area (Fig. 2a).

In the Fojo mining area, the mass magnetic susceptibility was slightly influenced by Zr (15.7 keV) (Fig. 10a), which is found in concentration levels from 0.006 to 0.012% in selected samples (Table 2), probably characterizing a silicate mineral phase



**Fig. 9** The plots of weighted regression coefficients ( $B_w$ ) obtained using 1 factor for the PLS regressions in soil samples from São Pedro da Cova: cluster 1 (a) and cluster 2 (b).

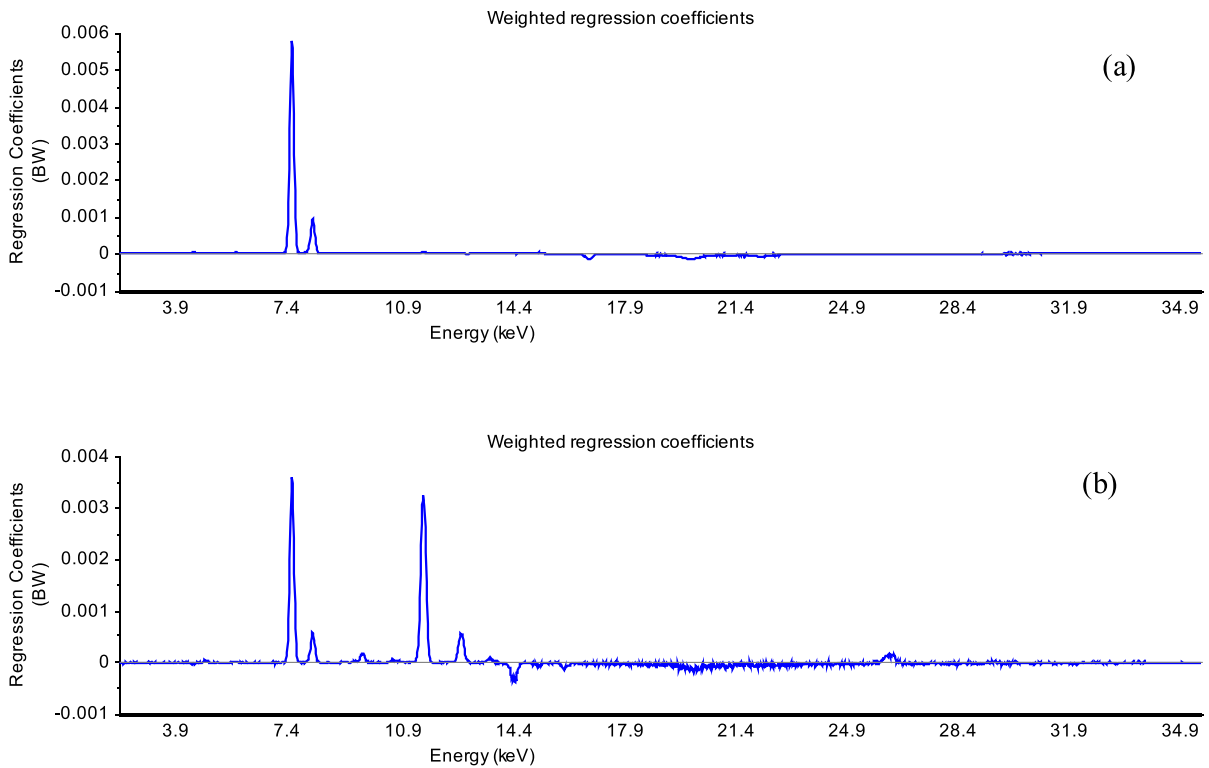
The peaks with the highest regression coefficients (at 6.4 and 7.1 keV) corresponding to Fe, play the most important role for the mass magnetic susceptibility in the studied soils

(Ribeiro et al., 2011). These results differ from the previously identified controlling variables for the two clusters studied around the São Pedro da Cova coal mine (Fig. 9), which explains the different nature of magnetism, as well as complexity of soil magnetic properties in the individual coal mines. Also, the importance of the contribution of geochemical elements to the soil's magnetic properties, identified by MVA, should be emphasized.

In the Regoufe W–Sn mine samples, the profile of the important variables that most affect the mass magnetic susceptibility is different (Fig. 10b). In addition to Fe, the predominant variable is As (10.5 keV and 11.6 keV) with a strong correlation with the mass magnetic susceptibility, which probably indicates the influence of arsenopyrite (FeAsS) in the mine waste soils. In these samples the concentration of As is high, ranging from 0.075% up to 1%

(Table 2), which indicates mine waste contamination. Previous studies have shown that magnetic spherules of magnetite ( $\text{Fe}_3\text{O}_4$ ), of anthropogenic origin, are the main magnetic carrier, heterogeneously distributed in the soil samples (Sant'Ovaia et al., 2023). In addition to Fe and As, the mass magnetic susceptibility is also affected by Zn (8.6 keV), Rb (13.4 keV), and Sn (25.2 keV), which are found in these samples at maximum respective levels up to 0.05%, 0.08%, and 0.01%. This confirmed the influence of pedological processes and lithological differences in the studied samples, as well as the contribution of anthropogenic magnetic spherules. These results are supported by earlier geological studies which showed that the Regoufe pluton was a muscovite-albite porphyritic granite, containing mineralized quartz veins with arsenopyrite (AsFeS) and cassiterite ( $\text{SnO}_2$ ), with





**Fig. 10** The plots of weighted regression coefficients ( $B_w$ ) obtained using 1 factor for the validated PLS regressions in soil samples from: Fojo (a) and Regoufe (b). The peaks with the highest regression coefficients (at 6.4 and 7.1 keV) corresponding to Fe, play the most important role in the regression

for the mass magnetic susceptibility in the studied soils from Fojo and Regoufe. In soils from Regoufe mining area, As (10.5 and 11.7 keV) was also very important variable for the PLS regression of the mass magnetic susceptibility

minor amounts of sphalerite ( $ZnFeS$ ), among other metallic minerals (Sant’Ovaia et al., 2023).

**Conclusions**

The use of MVA to analyse the entire XRF spectra (including all characteristic geochemical features) to discriminate the origin of mass magnetic susceptibility in soil samples shows the relative contribution of different geochemical parameters in determining the soil magnetic properties. The XRF–MVA models were developed and validated for individual mining areas and their application to similar mining areas in other geographical zones should be tested. Importantly, the use of statistical analyses of the type used here demonstrates the linkages between magnetic susceptibility and chemical composition of the soils and

the relationship to contaminant from anthropogenic processes and activities.

Wide ranges of absolute values of the mass magnetic susceptibility ( $2.15\text{--}1003 \times 10^{-8} \text{ m}^3 \text{ kg}^{-1}$ ), obtained by standard laboratory measurements in soils from three abandoned mining areas were comprehensively interpreted by the developed XRF–MVA models. The application of PCA to XRF soil spectroscopic signatures provides significant information to enable sample discrimination (PC-1 + PC-2 = 94%) and identification of soils with specific inputs from both mining activities and the natural environment, with subsequent cluster analysis discriminating between contaminated and non-contaminated soils. The PLS multivariate regression confirmed the strong positive correlation ( $R^2 > 0.95$ ) between soil geochemistry and mass magnetic susceptibility, selecting the most important model variables affecting soil health.

We find the applicability of the XRF–MVA models more useful at the local rather than the regional level due to differences in the sources of soil and contamination between different localities and mine types. Future research should include increased sample sizes at a more diverse range of mining areas (including more and less contaminated soils), to test the regional applicability of the XRF–MVA models with all their advantages in the assessment of specific details of soil magnetic properties. The site-specific model explained here adopts a holistic approach that can be a useful contribution to ecologists and geochemists, as well as other soil science professionals, as a case study for rapid surveys for soil contamination.

**Acknowledgements** The authors are thankful to the Editor of the Journal and to the reviewers.

**Author contributions** All authors contributed to the study conception and design. Conceptualization: J. Milinovic, B. J. Murton, M. Azenha. Methodology: J. Milinovic, A. Futuro, M. Azenha. Formal analysis and investigation: J. Milinovic, M. Azenha. Writing—original draft: J. Milinovic. Writing—review and editing: J. Milinovic, P. Santos, H. Sant’Ovaia, B. J. Murton, M. Azenha. Funding acquisition and resources: C. M. Pereira, D. Flores, M. Azenha.

**Funding** Open access funding provided by FCT/IFCCN (b-on). This research was supported by the project “SHS-Soil health surrounding former mining areas: characterization, risk analysis, and intervention,” which is financed by national funds from the Norte Portugal Regional Coordination and Development Commission (CCDRN), Grant Number: NORTE-01–0145-FEDER-000056. The FCT financially supported partially this work under Research Grants UIDB/00081/2020-CIQUP (<https://doi.org/https://doi.org/10.54499/UIDB/00081/2020>) and LA/P/0056/2020 (IMS).

**Data availability** All data and materials in this study are provided in the manuscript.

## Declarations

**Conflict of interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Ethical approval** The research did not involve human participants or animals.

**Consent to participate** The authors agreed with the participation.

**Consent to publish** The authors agreed with the publishing.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Agyeman, P. C., Ahado, S. K., Kingsley, J., Kebonye, N. M., Biney, J. K. M., Boruvka, L., Vasat, R., & Kocarek, M. (2021). Source apportionment, contamination levels, and spatial prediction of potentially toxic elements in selected soils of the Czech Republic. *Environmental Geochemistry and Health*, 43, 601–620. <https://doi.org/10.1007/s10653-020-00743-8>
- Attoucheik, L., Jordanova, N., Bayou, B., Lagroix, F., Jordanova, D., Maouche, S., Henry, B., & Boutaleb, A. (2017). Soil metal pollution from former Zn–Pb mining assessed by geochemical and magnetic investigations: Case study of the Bou Caid area (Tissemsilt, Algeria). *Environmental Earth Sciences*, 76, 298. <https://doi.org/10.1007/s12665-017-6622-9>
- Biney, J. K. M., Vařat, R., Blöcher, J. R., Borůvka, L., & Němeček, K. (2022). Using an ensemble model coupled with portable X-ray fluorescence and visible near-infrared spectroscopy to explore the viability of mapping and estimating arsenic in agricultural soil. *Science of the Total Environment*, 818, 151805. <https://doi.org/10.1016/j.scitotenv.2021.151805>
- Blanco, M., Coello, J., Montoliu, I., & Romero, M. A. (2001). Orthogonal signal correction in near infrared calibration. *Analytica Chimica Acta*, 434, 125–132. [https://doi.org/10.1016/S0003-2670\(01\)00820-0](https://doi.org/10.1016/S0003-2670(01)00820-0)
- Bosco, G. L. (2013). Development and application of portable, hand-held x-ray fluorescence spectrometers. *Trends in Analytical Chemistry*, 45, 121–134. <https://doi.org/10.1016/j.trac.2013.01.006>
- Bourliva, A., Papadopoulou, L., Aidona, E., & Gouiri, K. (2017). Magnetic signature, geochemistry, and oral bioaccessibility of “technogenic” metals in contaminated industrial soils from Sindos Industrial Area, Northern Greece. *Environmental Science and Pollution Research*, 24, 17041–17055. <https://doi.org/10.1007/s11356-017-9355-0>
- Brereton, R. G. (1990). *Chemometrics: Applications of mathematics and statistics to laboratory systems*. Ellis Horwood.
- Chakraborty, P., Wood, D. A., Singh, S., & Hazra, B. (2023). Trace element contamination in soils surrounding the open-cast coal mines of eastern Raniganj basin, India.

- Environmental Geochemistry and Health*, 45, 7275–7302. <https://doi.org/10.1007/s10653-023-01556-1>
- Correia, P., Šimůnek, Z., Sá, A. A., & Flores, D. A. (2018). A new late Pennsylvanian floral assemblage from the Douro basin, Portugal. *Geological Journal*, 53, 2507–2531. <https://doi.org/10.1002/gj.3086>
- Dekkers, M. J. (1997). Environmental magnetism: An introduction. *Netherlands Journal of Geosciences*, 76, 163–182. <https://doi.org/10.1023/A:1003122305503>
- Dellbecque, N., Ranst, E. V., Dondeyne, S., Mouazen, A. M., Vermeir, P., & Verdoodt, A. (2022). Geochemical fingerprinting and magnetic susceptibility to unravel the heterogeneous composition of urban soils. *Science of the Total Environment*, 847, 157502. <https://doi.org/10.1016/j.scitotenv.2022.157502>
- dos Santos, F. R., de Oliveira, J. F., Bona, E., dos Santos, J. V. F., Barboza, G. M. C., & Melquiades, F. L. (2020). EDXRF spectral data combined with PLSR to determine some soil fertility indicators. *Microchemical Journal*, 152, 104275. <https://doi.org/10.1016/j.microc.2019.104275>
- Engelen, S., Hubert, M., Vanden Branden, K., & Verboven, S. (2004). Robust PCR and robust PLSR: A comparative study. In M. Hubert, G. Pison, A. Struyf, & S. van Aelst (Eds.), *Theory and applications of recent robust methods. Statistics for industry and technology* (pp. 105–117). Birkhäuser.
- Evans, M. E., & Heller, F. (2003). *Environmental Magnetism: Principles and applications of enviromagnetics*. Elsevier Science.
- Figueiredo, J., Vila, M. C., Góis, J., Biju, B. P., Futuro, A., Martins, D., Dinis, M. L., & Fiúza, A. (2019). Bi-level depth assessment of an abandoned tailings dam aiming its reprocessing for recovery of valuable metals. *Minerals Engineering*, 133, 1–9. <https://doi.org/10.1016/j.mineng.2018.12.016>
- Flanders, P. J. (1994). Collection, measurement, and analysis of airborne magnetic particulates from pollution in the environment (invited). *Journal of Applied Physics*, 75, 5931–5936. <https://doi.org/10.1063/1.355518>
- Golia, E. E., & Diakouloukas, V. (2022). Soil parameters affecting the levels of potentially harmful metals in Thessaly area, Greece: A robust quadratic regression approach of soil pollution prediction. *Environmental Science and Pollution Research*, 29, 29544–29561. <https://doi.org/10.1007/s11356-021-14673-0>
- Guyodo, Y., Mostrom, A., Penn, R. L., & Banerjee, S. K. (2003). From nanodots to nanorods: Oriented aggregation and magnetic evolution of nanocrystalline goethite. *Geophysical Research Letters*, 30, 1512. <https://doi.org/10.1029/2003GL017021>
- ISO 11760. (2005). Classification of Coals (1st ed). International Organization for Standardization (p. 9). Geneva
- Jang, M. (2010). Application of portable X-ray fluorescence (pXRF) for heavy metal analysis of soils in crop fields near abandoned mine sites. *Environmental Geochemistry and Health*, 32, 207–216. <https://doi.org/10.1007/s10653-009-9276-z>
- Kardanpour, Z., Jacobsen, O. S., & Esbensen, K. H. (2014). Soil heterogeneity characterization using PCA ( $X_{\text{variogram}}$ )—Multivariate analysis of spatial signatures for optimal sampling purposes. *Chemometrics and Intelligent Laboratory Systems*, 136, 24–35. <https://doi.org/10.1016/j.chemolab.2014.04.020>
- Lemos de Sousa, M. J., & Wagner, R. H. (1983). General description of the terrestrial carboniferous basins in Portugal and history of investigations. In M. J. Lemos de Sousa & J. T. Oliveira (Eds.), *The carboniferous of Portugal: Memórias dos serviços geológicos de Portugal* (pp. 117–126). Lisbon.
- Lu, S. G., Wang, H. Y., & Chen, Y. Y. (2012). Enrichment and solubility of trace metals associated with magnetic extracts in industrially derived contaminated soils. *Environmental Geochemistry and Health*, 34, 433–444. <https://doi.org/10.1007/s10653-011-9447-6>
- Ma, X., Xia, D., Chen, P., Yu, Q., & Liu, X. (2023). Heavy metals distribution, magnetic properties, source apportionment, and potential risks in urban street dust of north-west China. *Water, Air, & Soil Pollution*, 234, 133. <https://doi.org/10.1007/s11270-023-06132-y>
- Magiera, T., Górka-Kostrubiec, B., Szumiata, T., & Bučko, M. S. (2023). Technogenic magnetic particles in topsoil: Characteristic features for different emission sources. *Science of the Total Environment*, 865, 161186. <https://doi.org/10.1016/j.scitotenv.2022.161186>
- Magiera, T., Strzyszcz, Z., Kapicka, A., & Petrovsky, E. (2006). Discrimination of lithogenic and anthropogenic influences on topsoil magnetic susceptibility in Central Europe. *Geoderma*, 130, 299–311. <https://doi.org/10.1016/j.geoderma.2005.02.002>
- Malmir, M., Tahmasbian, I., Xu, Z., Farrar, M. B., & Bai, S. H. (2019). Prediction of soil macro- and micro-elements in sieved and ground air-dried soils using laboratory-based hyperspectral imaging technique. *Geoderma*, 340, 70–80. <https://doi.org/10.1016/j.geoderma.2018.12.049>
- Martens, H., & Naes, T. (1998). *Multivariate calibration*. Wiley.
- Martins, J. P. A., Teófilo, R. F., & Ferreira, M. M. C. (2010). Computational performance and cross-validation error precision of five PLS algorithms using designed and real data sets. *Journal of Chemometrics*, 24, 320–332. <https://doi.org/10.1002/cem.1309>
- Mello, D. C., Demattê, J. A. M., Silvero, N. E. Q., Raimo, L. A. D. L., Poppiel, R. R., Mello, F. A. O., Souza, A. B., Safanelli, J. L., Resende, M. E. B., & Rizzo, R. (2020). Soil magnetic susceptibility and its relationship with naturally occurring processes and soil attributes in pedosphere, in a tropical environment. *Geoderma*, 372, 114364. <https://doi.org/10.1016/j.geoderma.2020.114364>
- Milinic, J., Vale, C., & Azenha, M. (2023). Recent advances in multivariate analysis coupled with chemical analysis for soil surveys. *Journal of Soils and Sediments*, 23, 1085–1098. <https://doi.org/10.1007/s11368-022-03377-8>
- Morona, F., dos Santos, F. R., Brinatti, A. M., & Melquiades, F. L. (2017). Quick analysis of organic matter in soil by energy-dispersive X-ray fluorescence and multivariate analysis. *Applied Radiation and Isotopes*, 130, 13–20. <https://doi.org/10.1016/j.apradiso.2017.09.008>
- Morrissey, M. B., & Ruxton, G. D. (2018). Multiple regression is not multiple regressions: The meaning of multiple regression and the non-problem of collinearity.

- Philosophy, Theory, and Practice in Biology*, 10, 3. <https://doi.org/10.3998/ptpbio.16039257.0010.003>
- Mullins, C. E. (2006). Magnetic susceptibility of the soil and its significance in soil science—A review. *European Journal of Soil Science*, 28, 223–246. <https://doi.org/10.1111/j.1365-2389.1977.tb02232.x>
- Nahan, G., Bijaksana, S., Suryanata, P. B., & Ibrahim, K. (2023). Geochemical and magnetic characteristics of placer gold deposits from Central Kalimantan, Indonesia. *The Mining-Geological-Petroleum Engineering Bulletin*, 38, 99–107. <https://doi.org/10.17794/rgn.2023.2.7>
- Nengsih, T. A., Bertrand, F., Maumy-Bertrand, M., & Meyer, N. (2019). Determining the number of components in PLSR regression on incomplete data set. *Statistical Applications in Genetics and Molecular Biology*. <https://doi.org/10.1515/sagmb-2018-0059>
- Ng, W., Minasny, B., Jones, E., & McBratney, A. (2022). To spike or to localize? Strategies to improve the prediction of local soil properties using regional spectral library. *Geoderma*, 406, 115501. <https://doi.org/10.1016/j.geoderma.2021.115501>
- Pan, H., Lu, X., Lei, K., Shi, D., Ren, C., Yang, L., & Wang, L. (2019). Using magnetic susceptibility to evaluate pollution status of the sediment for a typical reservoir in northwestern China. *Environmental Science and Pollution Research*, 26, 3019–3032. <https://doi.org/10.1007/s11356-018-3844-7>
- Panchuk, V., Yaroshenko, I., Legin, A., Semenov, V., & Kirisanov, D. (2018). Application of chemometric methods to XRF-data—A tutorial review. *Analytica Chimica Acta*, 1040, 19–32. <https://doi.org/10.1016/j.aca.2018.05.023>
- Pereira, E., Rodrigues, J., Gonçalves, L. S. M., Moreira, A., Silva, A. F. (2007). Carta Geológica de Portugal Na escala 1:50.000 & Notícia Explicativa da Folha 13D (Oliveira de Azeméis); Instituto Nacional de Engenharia, Tecnologia e Inovação, (p. 55). Lisboa, Portugal
- Pinto de Jesus, A. (2019). Carboniferous intermontane basins of Portugal. In J. Oliveira & C. Quesada (Eds.), *The geology of Iberia: A geodynamic approach* (Vol. 2, pp. 402–408). Springer Nature.
- Ribeiro, J., da Silva, E. F., de Jesus, A. P., & Flores, D. (2011). Petrographic and geochemical characterization of coal waste piles from Douro Coalfield (NW Portugal). *International Journal of Geology*, 87, 226–236. <https://doi.org/10.1016/j.coal.2011.06.014>
- Ribeiro, J., Sant’Ovaia, H., Gomes, C., Ward, C., & Flores, D. (2014). Mineralogy and magnetic parameters of materials resulting from the mining and consumption of coal from the douro coalfield, Northwest Portugal. *Coal and peat fires: A global perspective* (Vol. 3, pp. 494–508). Elsevier Inc.
- Rocha, D. R., Melquiades, F. L., & Thomaz, E. L. (2019). Modeling the soil burnt effect for temperature prediction by energy dispersive X ray fluorescence in an haplic cambisol soil. *Applied Radiation and Isotopes*, 150, 26–30. <https://doi.org/10.1016/j.apradiso.2019.05.012>
- Santos, P., Marques, J. E., Ribeiro, J., Mansilha, C., Melo, A., & Fonseca Sant’Ovaia Flores, R. H. D. (2023). Geochemistry of soils from the surrounding area of a coal mine waste pile affected by self-burning (Northern Portugal). *Minerals*, 13, 28. <https://doi.org/10.3390/min13010028>
- Sant’Ovaia, H., Cruz, C., Guedes, A., Ribeiro, H., Santos, P., Pereira, S., Marques, J. E., Ribeiro, M. A., Mansilha, C., Martins, H. C. B., Valentim, B., Torres, J., Abreu, I., Noronha, F., & Flores, D. (2023). Contamination fingerprints in an inactive W (Sn) mine: The Regoufe mine study case (Northern Portugal). *Minerals*, 13, 497. <https://doi.org/10.3390/min13040497>
- Shaheen, A., & Iqbal, J. (2018). Spatial distribution and mobility assessment of carcinogenic heavy metals in soil profiles using geostatistics and random forest, Boruta Algorithm. *Sustainability*, 10, 799. <https://doi.org/10.3390/su10030799>
- Szuskiewicz, M., Petrovský, E., Lukasiak, A., Gruba, P., Grison, H., & Szuskiewicz, M. M. (2021). Technogenic contamination or geogenic enrichment in Regosols and Leptosols? Magnetic and geochemical imprints on topsoil horizons. *Geoderma*, 381, 114685. <https://doi.org/10.1016/j.geoderma.2020.114685>
- Turekian, K. K., & Wedepohl, K. H. (1961). Distribution of the elements in some major units of the Earth’s Crust. *Geology Society of America Bulletin*, 72, 175–192. [https://doi.org/10.1130/0016-7606\(1961\)72\[175:DOTEIS\]2.0.CO;2](https://doi.org/10.1130/0016-7606(1961)72[175:DOTEIS]2.0.CO;2)
- Van Gaans, P. F. M., Vriend, S. P., & Poorter, R. P. E. (1995). Hydrothermal processes and shifting element association patterns in the W–Sn enriched granite of Regoufe, Portugal. *Journal of Geochemical Exploration*, 55, 203–222. [https://doi.org/10.1016/0375-6742\(95\)00015-1](https://doi.org/10.1016/0375-6742(95)00015-1)
- Wang, G., Liu, Y., Chen, J., Ren, F., Chen, Y., Ye, F., & Zhang, W. (2018). Magnetic evidence for heavy metal pollution of topsoil on Shanghai, China. *Frontiers in Earth Science*, 12, 125–133. <https://doi.org/10.1007/s11707-017-0624-5>
- Wu, Y., Noda, I., Meersman, F., & Ozaki, Y. (2006). Orthogonal signal corrected two-dimensional (OSC 2D) correlation infrared spectroscopy. *Journal of Molecular Structure*, 799, 121–127. <https://doi.org/10.1016/j.molstruc.2006.03.030>
- Zhang, W., Zhang, Y., Chen, D., Zhang, R., Yu, X., Gao, Y., Wang, C., Liu, J., Zhao, N., & Liu, W. (2013). Quantitative analysis of overlapping x-ray fluorescence spectra for Ni, Cu, Zn in soil by orthogonal signal correction and partial least squares algorithm. *Advanced Materials Research*, 705, 70–74. <https://doi.org/10.4028/www.scientific.net/AMR.705.70>

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.