ORIGINAL PAPER

Stream sediment pollution: a compositional baseline assessment

Teresa Albuquerque[1,2,3] · Rita Fonseca[3] · Joana Araújo[3] · Natália Silva[3] · António Araújo[3]

Abstract

A high concentration of potentially toxic elements (PTEs) can affect ecosystem health in many ways. It is therefore essential that spatial trends in pollutants are assessed and monitored. Two questions must be addressed when quantifying pollution: how to define a non-polluted sample and how to reduce the problem's dimensionality. A geochemical dataset is a composition of variables (chemical elements), where the components represent the relative importance of each part of the whole. Therefore, to comply with the compositional constraints, a compositional approach was used. A novel compositional pollution indicator (CPI) based on compositional data (CoDa) principles such as the properties of sparsity and simplicity was computed. A dataset of 12 chemical elements in 33 stream-sediment samples were collected from depths of 0–10 cm in a grid of 1 km × 1 km and analyzed. Maximum concentrations of 3.8% Pb, 750 µg g^{-1} As, and 340 µg g^{-1} Hg were obtained near the mine tailings. The methodological approach involved geological background selection in terms of a trimmed subsample that could be assumed to contain only non-pollutants (Al and Fe) and the selection of a list of pollutants (As, Zn, Pb, and Hg) based on expert knowledge criteria and previous studies. Finally, a stochastic sequential Gaussian simulation of the new CPI was performed. The results of the hundred simulations performed were summarized through the mean image map and maps of the probability of exceeding a given statistical threshold, allowing the characterization of the spatial distribution and the associated variability of the CPI. A high risk of contamination along the Grândola River was observed. As the main economic activities in this area are agricultural and involve animal stocks, it is crucial to establish two lines of intervention: the installation of a surveillance network for continuous control in all areas and the definition of mitigation actions for the northern area with high levels of contamination.

✉ Teresa Albuquerque
   teresal@ipcb.pt

1   Instituto Politécnico de Castelo Branco, Av. Pedro Álvares Cabral, n° 12, 6000-084 Castelo Branco, Portugal

2   Centro de Estudos de Recursos Naturais, Ambiente E Sociedade (CERNAS)—Instituto Politécnico de Castelo Branco, Av. Pedro Álvares Cabral, n° 12, 6000-084 Castelo Branco, Portugal

3   ICT| Universidade de Évora, Largo Dos Colegiais 2, 7000 Évora, Portugal

🙋 Springer

Graphical Abstract
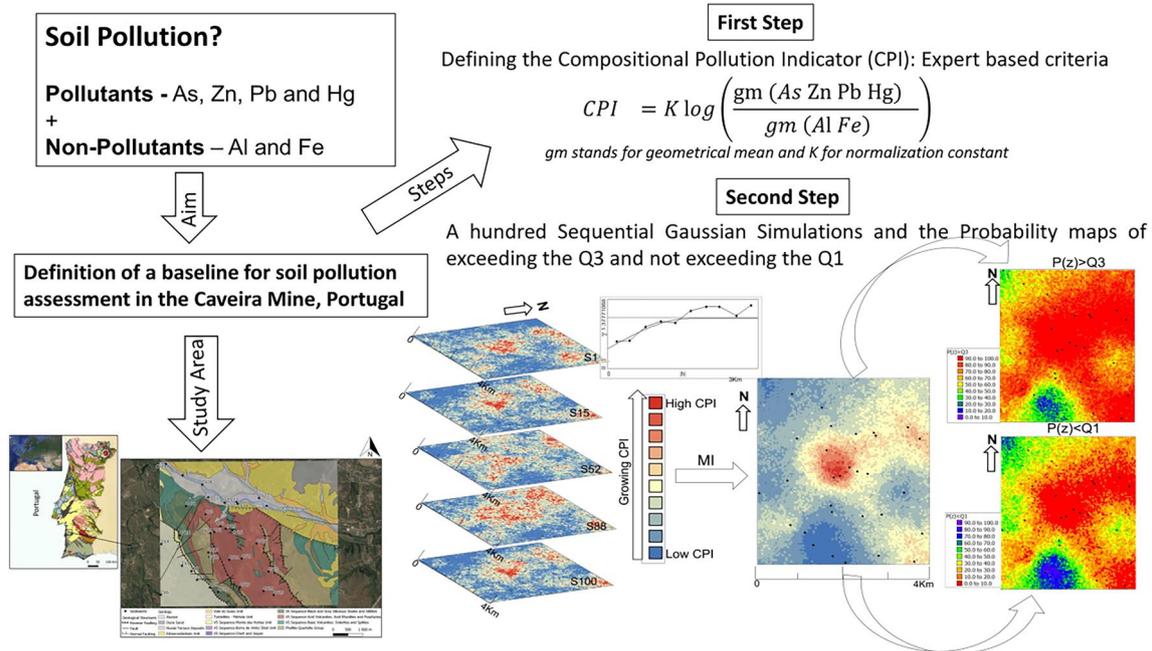
Introduction

The presence of potentially toxic elements (PTEs) poses a significant risk in various environmental sectors (Papamichaela et al. 2023), Boumaza et al. 2023, Ebru Yeşim Özkan et al. 2024, Hadzi et al. 2024). Over the last few decades, the cumulative impact of these elements on the environment has been considerable. During that time, there has been an exponential increase in the concentrations of PTEs, thus enhancing the risk to humans and the environment (Antoniadis et al. 2017; Kumudunis et al. 2020). Mining and heavy industrial activities may potentiate these high observed levels of PTEs and may be the origin of numerous sources of contamination (Boente et al. 2022, 2018; Carvalho et al. 2022). Thus, in recent decades, researchers have invested in the development of new techniques that offer accurate scenarios of the spatial distribution of PETs (Özkan et al. 2024, Sulemana et al. 2024, Petryshen 2023, Zhang et al. 2023). The definition of geochemical backgrounds and the identification of enrichment sources are key to the accomplishment of this objective (Wang et al. 2021; McKinley et al. 2016). The visualization and depiction of pollutants requires the use of simulated maps to visualize spatial–temporal distribution models. The definition of vulnerability and risk hot clusters

may provide a basis for environmental policy-making in complex scenarios (Boente et al. 2020; Albuquerque et al. 2017; McKinley et al. 2016). In soil and stream-sediment science, mapping a new variable called an index or an indicator is a common technique for describing the distribution of PTEs. The use of classical characterization methods like statistics and soil/stream sediment pollution indexes (SPIs) can help identify potentially polluted areas. A review by Joanna et al. (2018) provides a detailed and critical assessment of heavy metal soil pollution using various indicators. Unfortunately, however, the compositional nature (Pawlowsky-Glahn et al. 2015; Filzmoser et al. 2009) of the geochemical data is usually not considered. In most cases, the indicators are related to the study of individual elements without considering the interdependence of the concentrations of all the elements in the same set. Non-compositional indices that are often used to study geochemical data include the geoaccumulation index (Muller 1969), the enrichment factor (Sucharova et al. 2012), and the single pollution index (SPI) (Hakanson 1980), as reviewed in Kowalska et al. (2018). Nevertheless, it is well known that a traditional statistical approach using direct raw data can be misleading (Chayes 1962, 1971). Aitchison (1982, 1986) answered these questions in his fundamental work on the logarithmic ratio method. Theories of composition data

(CoDa) have enhanced our understanding of the sampling space of composition data and their corresponding structure (Pawlowsky-Glahn and Egozcue 2001). Representations of data that consider pairwise log ratios (pwlr), isometric log-ratio coordinates (ilr), centered log-ratio coordinates (clr), and additive log-ratio coordinates (alr) are statistically robust approaches to deal with the compositional nature of chemical concentration data (Pawlowsky-Glahn and Egozcue 2001; Egozcue et al. 2003; Buccianti and Grunsky 2014). The compositional approach (CoDa) is well represented in various fields of research in environmental science, such as ecotoxicology (Mullineaux et al. 2021), urban impacts (Cicchella et al. 2020), water quality management (Wei et al. 2018), and human health (Tepanosyan et al. 2020, Pawlowsky-Glahn and Buccianti 2011; Filzmoser et al. 2021). Recently, the adoption of compositional indicators for characterizing PTE pollution of soil has been increasing (Boente et al. 2022; Petrik et al. 2018). Compositional indicators involving the definition of geochemical baselines offer a valuable contribution as they are scale invariant and sub-compositionally coherent, meaning that a change in the concentration unit used will not modify the study's results (Pawlowsky-Glahn et al. 2015).

Compositional indicators are commonly used to measure water and air contamination. In stream sediments and soils, the use of compositional indexes or indicators to address pollution has only recently been explored (Boente et al. 2022). The primary challenge is the highly varied geochemical background, which hinders the distinction between what is polluted and what is natural. In addition, to assess stream-sediment pollution, the compositional baseline needs to account for a set of key issues: (1) the compositional nature of data; (2) spatial changes in the background; (3) the definition of pollution; (4) the indicator as a log-contrast; and (5) that an indicator should be provided for every type of pollution. This research introduces a new compositional pollution indicator (CPI) of riverine sediments, based on expert criteria, to characterize pollution in the Caveira mine in southern Portugal. This indicator corresponds to a balance of elements that respect the CoDa principles (Aitchison 1982).

Material and methods

Characteristics of the study area and the dataset

The studied sector is part of the Portuguese Iberian Pyrite Belt and is an example of a European post-mining area that dates back to the 1990s. Mining activity ceased there mainly because of ore exhaustion and more profitable methods worldwide, which resulted in an ore price reduction and made local mining activities infeasible (Martins and Oliveira

2000). Therefore, major pollution problems related to metal dispersion and mine waste management are present there. The geological sequence at Caveira mine, which closed back in the 1980s, corresponds to (from bottom to top) phyllites and quartzites (PQG) followed by a volcanic sedimentary complex sequence (VSC) unit (Late Famennian) represented by pyroclastics, rhyolitic lavas, tuffs, dark gray and siliceous shales, and rare jaspers. Intruding diabase rocks can be seen in the northern sector (Fig. 1). The massive sulfide deposits that were exploited in the region occurred in the vicinity of felsic volcanic rocks. The Mértola formation, from the Visean age, overlays the CVS and corresponds to a flysch sequence consisting of sandstones alternating with shales and thin-bedded siltstones. From a structural point of view, the whole sequence is part of the South Portuguese Zone, a thin-skinned fold and thrust belt from the Variscan age. Tailings and associated waste rock resulting from 129 years of pyrite and Cu mining are scattered along Grândola Creek. The semi-arid climatic conditions result in high erosion of residues by surface water, primarily during rainfall, causing serious contamination of Grândola Stream and its tributaries due to the degradation of sediments (Ferreira da Silva et al. 2015).

A dataset of 33 bottom-sediment samples distributed across small and narrow creeks—two of them flowing by the mine tailings pile and the larger Grândola Stream, of which they are tributaries—was obtained. These streams belong to the Sado River basin, the second-largest hydrographical basin in Southern Portugal. Samples were collected at 0 to 10 cm depth with an environmental hand soil sampling kit (#209.55, AMS) from within a grid of 1 km × 1 km. These samples were preserved at about 4 °C and later analyzed for 12 chemical elements, including PTEs of variable toxicity (As, Cd, Co, Cr, Hg, Mn, Ni, Pb, Zn, V) and major elements from lithogenic sources (Fe, Al). The most extractable forms of the metals (except for Hg) were obtained by partial digestion with aqua regia (HCl and HNO_3) in a high-pressure microwave digestion unit (Anton Paar Multiwave PRO) following US EPA (2007) method 3051A. The metals and As were analyzed by optical emission spectroscopy with an inductive plasma source (ICP-OES, PerkinElmer Optima 8300), using yttrium as an internal standard. The accuracy and analytical precision of all the analyses were checked through the analysis of reference materials and duplicate samples in each analytical set.

Mercury (Hg) was analyzed by a mercury analyzer (NIC MA-3000) based on thermal decomposition, gold amalgamation, and cold vapor atomic absorption spectroscopy detection. Sampling was followed by immediate readings of pH and redox potential values in wet samples using a portable multi-parameter analyzer (Consort C5020—the SP10T model for pH and the SP50X model for redox potential). In samples with insufficient moisture
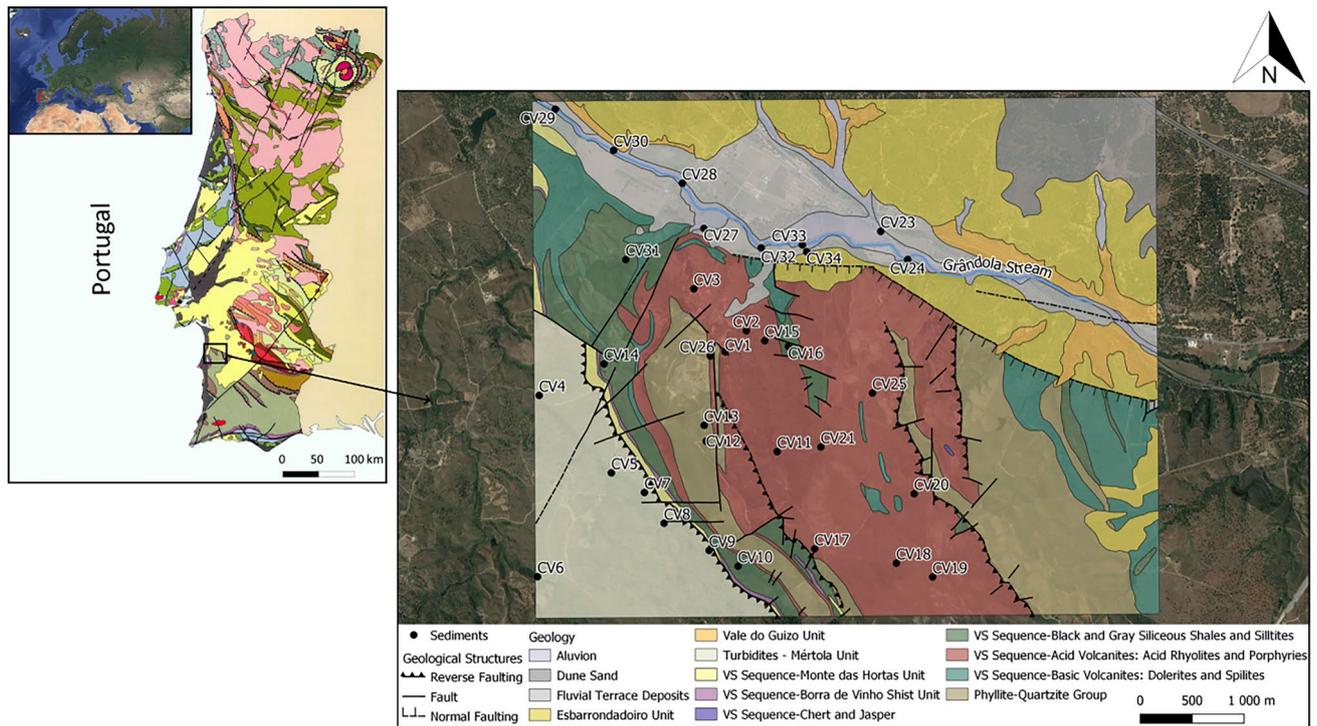
Fig. 1 Study area and the sample collection locations

for direct pH readings, this parameter was measured in a water–sediment suspension (2.5:1) in the laboratory. Concerning to samples' chemistry, the dataset included PTEs of variable toxicity (Fabian et al. 2014). The set of 12 elements was reported for each of the 33 sampling points, resulting in a 12-part composition that was assumed to represent the stream sediments.

Compositional pollution indicator (CPI) construction

The first fundamental principles of composition data are to be found in the founding work of Aitchison (1986). These initial contributions were explained and expanded into general-purpose works such as those of Pawlowsky-Glahn et al. (2015), Boogaart van den and Tolosana-Delgado (2013), Filzmoser and Hron (2011), Pawlowsky-Glahn and Buccianti (2011), and Pawlowsky-Glahn and Serra (2019).

The analysis of a stream-sediment sample based on its chemical composition should be conducted under the assumption that the data are compositional. As a result, when performing data analysis, the functions used to describe the composition should be invariant under multiplication by a positive constant (Boente et al. 2022). Also, any composition can be expressed in proportions (where the components sum to 1) without adding or losing any information, irrespective of the units in which the data were initially represented.

Analysis of the chemical composition of a sample of riverine sediments in units such as mg/kg should be performed assuming that the data are compositional. Moreover, the conversion of units from mg/kg to g/kg, as an example, must not change the information about the sample. This is summed up by one of the principles of CoDa analysis, named the principle of scale invariance. Thus, when analyzing the data, the functions used to describe the composition should be invariantly multiplied by a positive constant. Consequently, any composition can be expressed as proportions (where the components add to 1) without adding or losing information, regardless of the units in which the data were originally reported. A second assumption is known as the sub-compositional coherence principle. The whole periodic table is never presented; only a subset of elements is measured, and this subset may change over time and across the field. The elements observed form a composition, and any subassembly of them is a sub-composition that is again subject to the principle of scale invariance. Analyses of the initial composition or sub-composition should lead to coherent conclusions describing the roles of common elements (Aitchison 1986).

The CPI balance was obtained based on expert criteria attending a selection of elements (Boente et al. 2022), of which some are considered pollutants while others are not. In the case of the Caveira mine, the main contaminants were

selected from typical pollutants, namely As, Zn, Pb, and Hg, while Al and Fe were selected as the main natural-source elements (or non-pollutants). Based on this previous study, the selected balance, the CPI, was constructed as follows:

$$\text{CPI} = \sqrt{\frac{4}{3}} \ln\left(\frac{(As\text{ZnPbHg})^{1/4}}{(Al Fe)^{\frac{1}{2}}}\right). \tag{1}$$

Spatial modeling—a geostatistical approach

The computed compositional pollution indicator (CPI) is unbounded—a real random variable. Therefore, it fulfills the assumptions underlying a conventional geostatistical approach. Its spatial probability maps were computed by following a two-step geostatistical modeling method: (1) a structural analysis and a computation of experimental variograms (Journel and Huijbregts 1978) were performed, followed by (2) sequential Gaussian simulation (SGS), which was used as a stochastic simulation algorithm over a 100×100 km grid mesh.

The new CPI can be considered a regionalized variable (Matheron 1971), as it depends on the spatial location determined by the coordinates and is additive by construction. Indeed, the mean value within a given observed support is equal to the arithmetic average of the sample values independently of the associated statistical distribution (Albuquerque et al. 2017; Rivoirard 2005). Thus, the vector function used to calculate the spatial variation structure was the semi-variogram (Journel and Huijbregts 1978).

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left[Z(x_i) - Z(x_i + h)\right]^2 \tag{2}$$

The arguments taken into consideration are h (the distance) where $Z(x_i)$ and $Z(x_i + h)$ are the numerical values of the variables assigned to x_i and $x_i + h$. The total number of couples at a specified distance of h is $N(h)$. Therefore, it is the average value of the square of the differences between all couples of points in the geometric field spaced at a distance h (Journel and Huijbregts 1978). Plotting the behavior of the variogram gives an overall view of the spatial structure of the variable. One of the parameters that provide this information is the nugget effect (C_o), which supplies the behavior at the origin. The two other parameters are the sill (C_1) and the amplitude (a), which define the inertia used in the subsequent interpolation process and the influence radius of the variable, respectively.

The SGS starts by computing the univariate experimental distribution of values and performing a normal score transformation of the original values to a standard normal distribution (Goovaerts 1997). Normal scores at grid node locations are then simulated sequentially with simple kriging (SK) using the normal score data and a zero mean. Once all normal scores have been simulated, they are back-transformed to their original units. The outcome of a simulation is always a random version of the estimation process that reproduces the statistics of the known data and builds a realistic picture of reality. The associated spatial uncertainty is visualized through the construction of probability maps. If multiple sequences of simulation are computed, it is possible to obtain reliable probabilistic maps. The mean image map and maps of the probability of exceeding the third quartile (Q_3) and the probability of not exceeding the first quartile (Q_1) were computed.

Results and discussion

Geochemical data

The analyses of physicochemical parameters and the determination of the levels of PTEs of variable toxicity (As, Cd, Co, Cr, Hg, Mn, Ni, Pb, Zn, V) as well as the selected elements from lithogenic sources (Fe, Al), accounting for their capacity for solubilization and mobilization, were performed with the aim of achieving contamination mapping. The evaluation of each metal's mobility was based on partial digestion analysis (using aqua regia), considering the pH values. The element concentrations and pH values in the stream sediment samples are reported in Table 1. The values of the physical–chemical parameter that most affects the solubility, mobility, and precipitation of potentially toxic metals in the sediments from shallow streams, i.e., the pH, range from 2.06 and 7.39. The lower values (2.06–4.57) are found for the sediments from the two creeks that flow through the mine tailings pile. As would be expected, those sediments (Cv1, Cv2, Cv3, Cv26, Cv33, Cv34) contain the highest values of Pb, As, and Hg—the main contaminants in the mine tailings: the levels of those contaminants rise above the critical levels that require immediate intervention according to European regulations (based on the Netherlands legislation—Soil Quality Regulation, 2006). Zn, another element with levels of concern, and one which presents high contents in the massive sulfides that have been exploited in this mine, shows slight contamination levels in all the streams flowing from the tailings pile—mostly in locations that do not coincide with those where the other elements exceed critical levels. The highest values of this element also do not coincide with the most acidic conditions in the environment. Although the ores that were exploited in this mining area contained all of these elements, Zn has a high chemical mobility that is mostly influenced by the oxidation conditions that occur in all the sediments (240–650 mV), so its distribution is more diffuse.

Table 1 Element concentrations (mg/kg) in the stream sediment samples (spring season) from Grândola and its tributary streams. These waterways belong to the Sado watershed

Sample*	As	Co	Cr	Mn	Ni	V	Zn	Al	Fe	Pb	Hg	pH (H$_2$O)
CV1	234.9	0.4	2.3	16.9	1.6	3.3	94.2	1441.6	10,217.0	38,458.7	127.9	2.30
CV2	238.6	0.4	1.6	24.5	1.0	2.0	96.4	2168.4	17,815.8	29,837.2	78.2	2.06
CV3	517.6	0.8	1.5	99.6	0.9	2.9	122.3	16,128.4	65,825.8	13,775.3	125.8	4.57
CV4	<1.5	3.9	13.3	172.7	10.3	14.3	36.4	18,321.3	31,828.4	51.2	1.7	6.61
CV5	14.7	3.5	15.6	261.9	12.7	15.5	47.2	13,099.4	27,183.4	2881.5	13.7	6.00
CV6	<1.5	3.4	17.6	208.4	9.3	24.3	30.0	26,272.6	21,172.3	221.4	0.2	6.54
CV7	<1.5	4.2	18.6	399.1	11.2	29.2	36.4	29,376.6	28,193.2	97.2	0.6	6.64
CV8	<1.5	4.6	19.5	455.6	12.3	25.7	37.5	24,811.4	28,361.9	15.0	0.3	7.17
CV9	<1.5	5.8	13.4	414.1	12.4	13.8	34.7	13,007.8	27,207.6	16.4	0.2	4.99
CV10	<1.5	5.6	15	315	21.6	22.9	46.3	23,727.5	23,919.7	18.2	0.6	5.12
CV11	3.4	1.0	1.4	148	1.2	2.6	15.8	8807.1	12,058.1	21.5	0.3	5.05
CV12	31.2	4.2	9.6	138.6	10.6	15.9	140.4	11,176.5	29,630.2	92.7	2.6	5.30
CV13	1.6	34.4	23.8	903	21.8	36.4	165.8	44,359.5	42,371.5	36.3	1.6	7.29
CV14	1.6	4.1	16.1	434.9	15.9	24.3	47.5	18,683.1	21,463.1	36.6	0.6	6.03
CV15	1.7	0.5	1.9	82.2	0.6	3.6	5.3	8882.1	3720.9	2.5	0.2	5.48
CV16	<1.5	0.6	2.1	198.2	1.3	4.6	17.5	4821.9	5337.6	25.9	0.2	5.98
CV17	4.0	1.6	4.3	412.9	4.0	5.2	15.8	10,396.1	16,746.8	27.1	0.1	6.35
CV18	<1.5	0.6	2.5	117.3	1.5	2.7	8.9	3475.2	3540.5	8.1	0.3	5.51
CV19	1.8	0.5	1.4	66.3	0.7	2.3	8.1	3044.8	3855.6	7.6	0.1	5.55
CV20	2.1	1.0	4.5	169.6	1.5	6.8	10.6	3677	5856.6	9.8	0.2	5.82
CV21	<1.5	0.8	2.0	194.6	0.5	3.6	9.9	3574.9	3936.7	7.9	0.1	6.04
CV23	1.9	1.3	11.9	98.5	6.2	14.1	136.3	9503.4	9550.3	38.1	3.0	7.39
CV24	44.2	1.5	7.0	79.2	5.8	12.1	122.4	9728.8	13,370.1	585.3	5.7	6.74
CV25	<1.5	0.4	1.5	84.8	0	3.7	6.3	14,787.2	7836.1	4.4	0.1	6.63
CV26	140.3	0.4	1.6	34.4	2.0	2.6	123.4	1193.7	10,730.9	44,540.5	46.8	2.34
CV27	32.3	3.6	11.6	177.5	12.5	17.9	129.0	13,208.0	25,540.5	82.8	7.3	6.39
CV28	<1.5	3.7	16	385.5	10.1	24.6	36.0	22,281.1	17,596.5	29.1	1.3	5.34
CV29	<1.5	3.9	18.4	394.6	10.7	25.6	36.2	22,665.1	18,556.6	14.5	0.2	6.82
CV30	<1.5	4.1	13.8	258.4	11.8	15.9	39.8	11,984.7	19,370.5	21.5	0.1	6.16
CV31	9.0	2.1	14.2	106.4	10.5	15.7	30.2	6909.5	23,545.1	31.0	0.2	6.21
CV32	30.8	3.4	15.7	291.3	8.9	24.6	88.6	24,220.0	19,152.6	107.7	2.3	6.57
CV33	265.2	3.9	9.3	422.5	8.6	11.6	108.2	11,998.7	30,136.3	88.8	1.6	4.26
CV34	748.2	0.4	8.3	77.6	3.9	10.8	161.8	6854.4	44,363.4	23,162.4	381.4	2.13

*Sample Cv22 was eliminated

Descriptive statistical analysis

A preliminary descriptive analysis was conducted to gain a comprehensive overview of the dataset in terms of the statistical distribution of the elements chosen for the CPI balance. This crucial step offers a preliminary glimpse into the central tendencies, dispersion, and distribution of the variables under scrutiny and at the same time provides a succinct summary of the main features of the studied dataset (Table 2).

Pollution assessment relies heavily on the presence of severe outliers for Pb, which clearly demonstrates a consistently high concentration of this element (Fig. 2).

Descriptive statistics of the main selected contaminants, As, Zn, Pb, and Hg as pollutants and Al and Fe as the main natural-source elements or non-pollutants. Furthermore, a heat map was used for exploratory data analysis of the geochemical composition and sample clustering simultaneously in a synthetic way (Fig. 3) (Wilkinson and Friendly 2009; Langella et al. 2013). The heat map shows that the elements are divided into two groups (upper dendrogram). The first group corresponds to Al and Fe (non-pollutants) and the second group corresponds to As, Cd, Co, Cr, Mn, Ni, V, Zn, Hg, and Pb. However, Pb is significantly separated from the other covariates in the major group. Based on an expert-driven approach, As, Zn, Pb, and Hg were selected as pollutants.

Table 2 Descriptive statistics of the main selected contaminants: As, Zn, Pb, and Hg (pollutants) as well as Al and Fe (the main natural-source elements or non-pollutants)

Descriptive statistic	Al (mg/L)*	Fe (mg/L)*	Pb (mg/L)**	Zn (mg/L)**	Hg (mg/L)**	As (mg/L)**
Mean	13,407.283	20,465.320	4398.473	61.898	0.024	69.377
Median	11,386.020	20,718.660	36.558	37.255	0.001	1.815
Standard deviation	9754.315	13,372.587	11,032.719	51.415	0.072	160.541
Kurtosis	1.996	1.831	6.473	−0.970	19.616	10.037
Asymmetry	1.185	1.048	2.654	0.702	4.203	3.075
Min	1228.151	3441.193	2.235	5.254	0.000	1.500
Max	45,202.780	62,891.626	45,480.308	163.293	0.381	739.024

*Non-pollutant element

**Pollutant element

Fig. 2 Box-plot diagrams of the CPI's elements
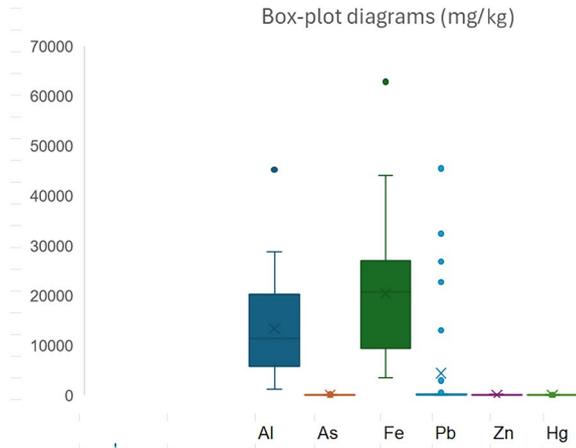
The dendrogram of samples (the left dendrogram) is divided into three groups. The central group corresponds to samples Cv6, Cv8, Cv10, Cv13, C14, Cv16, Cv17, Cv18, Cv21, Cv23, Cv25, Cv28, Cv29, CV7 and Cv32; the left group corresponds to samples Cv1, Cv2, Cv15, and Cv26; and the right group corresponds to samples Cv3, Cv4, Cv5, Cv9, Cv11, Cv12, Cv19, Cv20, Cv24, Cv27, Cv30, Cv31, Cv33, and Cv34. The map presents values in the dataset rearranged according to the dendrograms. Focusing on the rectangle/square patterns (note that the level of significance increases from red to blue through white) in the map, in particular those for the bottom group of samples Cv1, Cv2, Cv15, and Cv26 together with samples Cv34 and Cv3 of the upper group, it is possible to discern a lower-significance cluster

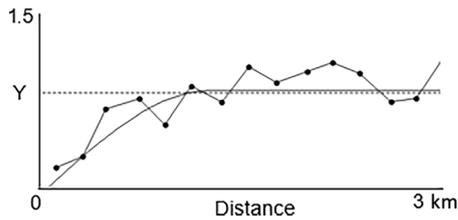**Fig. 3** Heat map and simultaneous sample/geochemical print dendrograms

Fig. 4 Experimental spherical omnidirectional variogram and the fit to it

for Al and a higher-significance cluster for Pb relative to the other samples. In future work, the relationship between the samples' geochemical print and the associated geology will be explored.

The compositional pollution indicator

The compositional balance of the CPI was obtained according to expert criteria. These criteria account for a selection of factors, some of which are considered pollutants while others are not. In the case of the Caveira mine, the identification of the main pollutants was addressed in previous studies (e.g., Ferreira da Silva et al. 2015), where typical pollutants such as As, Zn, Pb, and Hg were identified as being related to the activities of the old Iberian Pyrite Belt mines. The main natural-source elements (or non-pollutants) were two major elements (i.e., Al and Fe). Spatial modeling of the CPI was performed to identify hazardous clusters. A two-step geostatistical approach was used. As no clear evidence of anisotropy was found, the experimental isotropic variogram was computed, and the corresponding fitted model is shown in Fig. 4. The cross-validation correlation index for the observed and estimated CPI values is 0.70 and is therefore considered satisfactory for the selected models. Furthermore, a hundred simulations were performed using SGS as a conditional stochastic simulation of the CPI value distribution, and a hundred equiprobable scenarios were computed.

Probability maps corresponding to different thresholds allowed the visualization of spatial variability while setting aside the discussion of local accuracy, and they allowed the identification of hotspot clusters of pollution in the subject area. Realization numbers 1, 15, 32, 52, 67, and 99 are shown in Fig. 5.

The problem is that all representations (scenarios) have the same reliability, which means that a single achievement cannot be seen as a better representation of reality. Therefore, the mean spatial image (MI)—the average map—was computed and used as the CPI spatial distribution (Fig. 6a). The presentation of the probability of exceeding the third quartile (Q3) and the probability of not exceeding the first quartile (Q1) allows broad discussion of the CPI spatial distribution and the identification of hazard clustering (Fig. 6b,
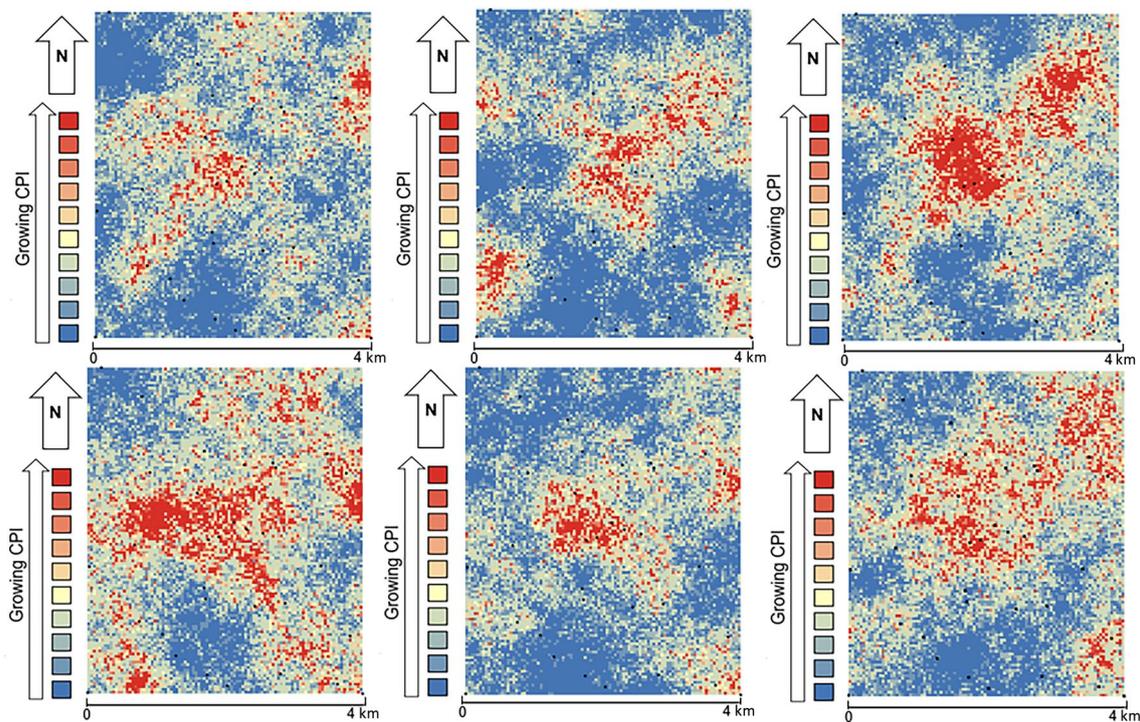

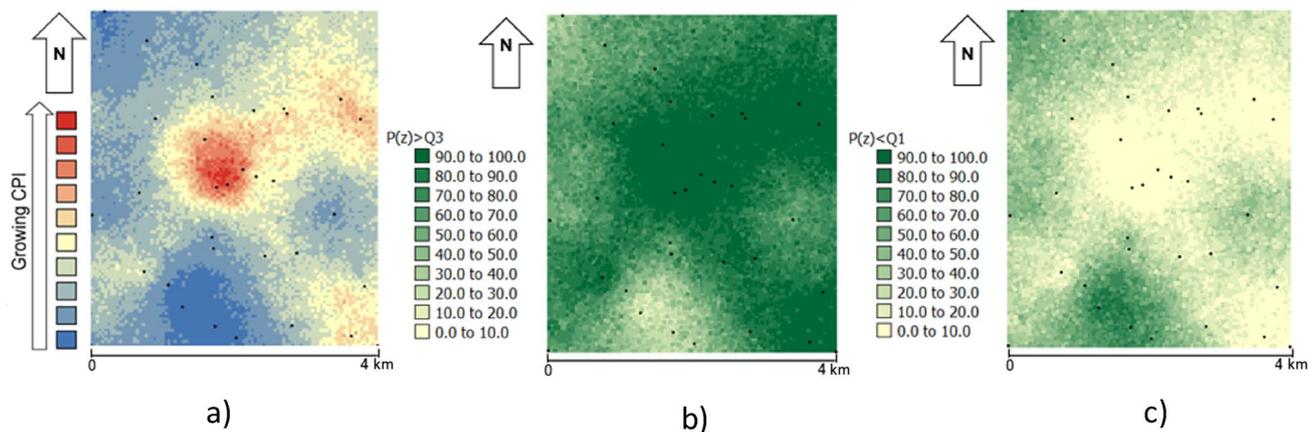**Fig. 5** Six different scenarios obtained by sequential Gaussian simulation (SGS)

Fig. 6 **a** Average SGS image (MI). Maps of the probability of **b** exceeding Q3 and **c** not exceeding Q1

c). To create distinct classes by reducing the within-class variance and maximizing the between-class variance, the Jenks natural break classification (Jenks 1967) was used, which allowed the determination of the best arrangement of values. The software Space-Stat v.4.0.18 (BioMedware) was used for the computation (Boente et al. 2022; Albuquerque et al. 2017).

The compositional pollution indicator (CPI) provides a fair representation of hotspots, especially along Grândola Stream and its tributaries, thereby confirming the high pollution detected around the old mine tailings and associated waste rock.

Conclusions

Geochemical data are compositional data, as the concentrations of the elements in any environmental matrix are commonly expressed as parts of the whole and vary together. Once this feature is accepted, compositional data procedures can be applied to obtain indicators that address pollution in, for example, stream sediment. The method was tested with up to 11 chemical elements in 33 sediment samples from the old Caveira mine in Portugal.

A high risk of contamination is observed along the Grândola River and in the vicinity of the mine tailings. It is important to consider agricultural and organic stocks as the main economic activities when establishing two lines of intervention: (1) the installation of a surveillance network for continuous control in all areas and (2) the definition of mitigation actions for the northern area, where high levels of contamination are observed.

Declarations

References

Aitchison J (1982) The statistical analysis of compositional data (with discussion). J R Stat Soc B 44(2):139–177

Aitchison J (1986) The statistical analysis of compositional data. Chapman and Hall, London, p 416

Albuquerque MT, Gerassis S, Sierra C, Taboada J, Martin J, Antunes IM, Gallego JR (2017) Developing a new Bayesian risk index for risk evaluation of soil contamination. Sci Total Env 603–604:167–177. https://doi.org/10.1016/j.scitotenv.2017.06.068

Antoniadis V, Golia EE, Shaheen SM, Rinklebe J (2017) Bioavailability and health risk assessment of potentially toxic elements in Thriasio Plain, near Athens, Greece. Environ Geochem Health 39:319–330

Boente C, Albuquerque MTD, Fernandez-Brana A, Gerassis S, Sierra C, Gallego JR (2018) Combining raw and compositional data to determine the spatial patterns of potentially toxic elements in soils. Sci Total Environ 632–631:1117–1126

Boente C, Gerassis S, Albuquerque MTD, Taboada J, Gallego JR (2020) Local versus regional soil screening levels to identify potentially polluted areas. Math Geosci 52:381–396

Boente C, Albuquerque MTD, Gallego JR, Pawlowsky Glahn V, Egozcue JJ (2022) Compositional baseline assessments to address soil pollution: an application in Langreo, Spain. Sci Total Environ 812:152383. https://doi.org/10.1016/j.scitotenv.2021.152383

Boumaza B, Kechiched R, Chekushina TV, Benabdeslam N, Senouci K, Hamitouche AE, Merzeg FA, Rezgui W, Rebouh NY, Harizi K (2023) Geochemical distribution and environmental assessment of potentially toxic elements in farmland soils, sediments, and tailings from phosphate industrial area (NE Algeria). J Hazard Mater 465:133110. https://doi.org/10.1016/j.jhazmat.2023.133110

Buccianti A, Grunsky E (2014) Compositional data analysis in geochemistry: are we sure to see what occurs during natural processes? J Geochem Explore 141:1–5

Carvalho PCS, Antunes IMHR, Albuquerque MTD, Santos ACS, Cunha PP (2022) Stream sediments as a repository of U, Th and As around abandoned uranium mines in Central Portugal: implications for water quality management. Env Earth Sci 81:6

Chayes F (1962) Numerical correlation and petrographic variation. J Geol 70(4):440–452

Chayes F (1971) Ratio correlation. University of Chicago Press, Chicago, p 99

Cicchella D, Zuzolo D, Albanese S, Fedele L, Tota D, Guagliardi I, Thiombane M, De Vivo B, Lima A (2020) Urban soil contamination in Salerno (Italy): concentrations and patterns of major, minor, trace, and ultra-trace elements in soils. J Geochem Explor 213:106519

Ebru YO, Serkan K, Şakir F (2024) Temporal change of potential toxic element contamination, ecological and toxicological risk in holocene in southern Black Sea coasts (Turkey). Region Studies Marine Sci. https://doi.org/10.1016/j.rsma.2024.103374

Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, Barceló-Vidal C (2003) Isometric log-ratio transformations for compositional data analysis. Math Geol 35(3):279–300

Fabian C, Reimann C, Fabian K, Birke M, Baritz R, Haslinger E (2014) Gemas: spatial distribution of the pH of European agricultural and grazing land soil. Appl Geochem 48:207–216

Ferreira da Silva E, Durães N, Reis P, Patinha C, Matos J, Costa MR (2015) An integrative assessment of environmental degradation of Caveira. abandoned mine area (Southern Portugal). J Geochem Explor 159:33–47. https://doi.org/10.1016/j.gexplo.2015.08.004

Filzmoser P, Hron K (2011) Compositional data analysis: theory and applications. In: Pawlowsky-Glahn V, Buccianti A (eds) Compositional data analysis: theory and applications. Wiley, London, pp 59–72

Filzmoser P, Hron K, Reimann C (2009) Univariate statistical analysis of environmental (compositional) data: problems and possibilities. Sci Total Environ 407(23):6100–6108

Filzmoser P, Hron K, Martín-Fernández J, Palarea-Albaladejo J (2021) Advances in compositional data analysis: festschrift in honour of Vera Pawlowsky-Glahn. Springer International, Cham

Goovaerts P (1997) Geostatistics for natural resources evaluation. Applied geostatistics series. Oxford University Press, New York, p 483

Hadzi GY, Essumang DG, Ayoko GA (2024) Assessment of contamination and potential ecological risks of heavy metals in riverine sediments from gold mining and pristine areas in Ghana. J Trace Elements Minerals 7(2024):100109. https://doi.org/10.1016/j.jtemin.2023.100109

Hakanson L (1980) An ecological risk index for aquatic pollution control.a sedimentological approach. Water Res 14:975–1001. https://doi.org/10.1016/0043-1354(80)90143-8

Jenks GF (1967) The data model concept in statistical mapping. Int Yearbook Cartogr 7:186–190

Joanna BK, Ryszard M, Michał G, Tomasz Z (2018) Pollution indices as useful tools for the comprehensive evaluation of the degree of soil contamination—a review. Environ Geochem Health 2018(40):2395–2420. https://doi.org/10.1007/s10653-018-0106-z

Journel AG, Huijbregts CJ (1978) Mining geostatistics. Academic, London, p 600

Kumuduni NP, Sabry MS, Season SC, Daniel CWT, Yohey H, Deyi H, Nanthi SB, Jörg R, Yong SO (2020) Soil amendments for immobilization of potentially toxic elements in contaminated soils: a critical review. Environ Int 134(2020):105046. https://doi.org/10.1016/j.envint.2019.105046

Langella O, Valot B, Jacob D, Balliau T, Flores R, Hoogland C, Joets J, Zivy M (2013) 2013 Management and dissemination of MS proteomic data with PROTICdb: example of a quantitative comparison between methods of protein extraction. Proteomics 13(9):1457–1466

Martins L, Oliveira D (2000) Exploration and mining. Instituto Geológico e Mineiro, Lisbon, p 20

G (1971) The theory of regionalized variables and Its applications. Les Cahiers du Centre de Morphologie Mathematique in Fontainebleu, Paris

McKinley JM, Hron K, Grunsky EC, Reimann C, de Caritat P, Filzmoser P, van den Boogaart KG, Tolosana-Delgado R (2016) The single-component geochemical map: fact or fiction? J Geochem Explor 162:16–28

Muller G (1969) Index of geoaccumulation in sediments of the Rhine River. Geol J 2:108–118

Mullineaux ST, McKinley JM, Marks NJ, Scantlebury DM, Doherty R (2021) Heavy metal (PTE) ecotoxicology, data review: traditional vs a compositional approach. Sci Total Environ 769:14524

Papamichaela I, Voukkalia I, Loiziaa P, Pappasb G, Zorpasa AA (2023) Existing tools used in the framework of environmental performance. Sustain Chem Pharmacy 32(2023):101026. https://doi.org/10.1016/j.scp.2023.101026

Pawlowsky-Glahn V, Egozcue JJ (2001) A geometric approach to statistical analysis on the simplex. Stoch Environ Res Risk Assess 15(5):384–398

Pawlowsky-Glahn V, Buccianti A (2011) Compositional data analysis: theory and applications. Wiley, London

Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R (2015) Modeling and analysis of compositional data. Statistics in practice. Wiley, Chichester, p 272

Petrik A, Thiombane M, Lima A, Albanese S, Buscher JT, De Vivo B (2018) Soil contamination compositional index: a new approach to quantify contamination demonstrated by assessing compositional source patterns of potentially toxic elements in the Campania region (Italy). J Appl Geochem 96:264–276

Petryshen W (2023) Spatial distribution of selenium and other potentially toxic elements surrounding mountaintop coal mines in the

Elk Valley, British Columbia. Canada Heliyon 9(2023):e17242. https://doi.org/10.1016/j.heliyon.2023.e17242

Rivoirard J (2005) Concepts and methods of geostatistics. Space, structure, and randomness. In: Meyer F, Schmitt M (eds) Contributions in honor of Georges Matheron in the fields of geostatistics, random sets, and mathematical morphology. Springer, New York

Sucharova J, Suchara I, Hola M, Marikova S, Reimann C, Boyd R, Filzmoser P, Englmaier P (2012) Top-/bottom-soil ratios and enrichment factors: what do they show? J Appl Geochem 27:138–145

Tepanosyan G, Sahakyan L, Maghakyan N, Saghatelyan A (2020) Combination of compositional data analysis and machine learning approaches to identify sources and geochemical associations of potentially toxic elements in soil and assess the associated human health risk in a mining city. Environ Pollut 261:11421

van den Boogaart KG, Tolosana-Delgado R (2013) Analysing compositional data with R. Springer, Berlin, p 258

Wang Z, Chen X, Yu D, Zhang L, Wang J, Lv J (2021) Source apportionment and spatial distribution of potentially toxic elements in soils: a new exploration on receptor and geostatistical models. Sci Total Environ 759(14342):8

Wei Y, Wang Z, Wang H, Yao T, Li Y (2018) Promoting inclusive water governance and forecasting the structure of water consumption based on compositional data: a case study of Beijing. Sci Total Environ 634:407–416

Wilkinson L, Friendly M (2009) The history of the cluster heat map. Am Stat 63(2):179–184

Zhang Y, Li T, Guo Z, Xie H, Hu Z, Ran H, Li C, Jiang Z (2023) Spatial heterogeneity and source apportionment of soil metal(loid)s in an abandoned lead/zinc smelter. J Environ Sci 127:519–529