Environmental Research

Stream sediments pollution: a compositional baseline assessment – the Caveira mine, Portugal

--Manuscript Draft--

Manuscript Number:	ER-23-9416
Article Type:	Research paper
Section/Category:	Environmental Health & Risk Assessment
Keywords:	Caveira mine; Pollution; Compositional Pollution Indicator (CPI); Sequential Gaussian Simulation; Probability map.
Corresponding Author:	Teresa Albuquerque, Ph.D Castelo Branco, PORTUGAL
First Author:	Teresa Albuquerque, Ph.D
Order of Authors:	Teresa Albuquerque, Ph.D
	Rita Fonseca
	Joana Araújo
	Natália Silva
	António Araújo
Abstract:	A high concentration of Potentially Toxic Elements (PTEs) can affect ecosystem health in many ways. It is therefore essential that spatial trends of pollutants are assessed and controlled. Two questions must be addressed when quantifying pollution. How to define a non-polluted sample? and, how to reduce the problems' dimensionality. Since the concentration of chemical elements is compositional, a compositional approach was used as the attributes vary together. A novel Compositional Pollution Indicator (CPI) based on Compositional Data (CoDa) principles such as sparsity and simplicity as properties, was computed. A dataset of 33 stream-sediment samples was collected from within 0 to 10 cm depth, in a grid of 1Km x 1Km, and twelve chemical elements were analyzed. Concentrations, reaching 3.8% Pb, 750µgg-1 As, and 340 µgg-1 Hg, were obtained near the mine tailings. The methodological approach implied the geological background selection in terms of a trimmed subsample that can be assumed as non-pollutant (Al and Fe) and the selection of a list of pollutants based on expert knowledge criteria and previous studies (As, Zn, Pb, and Hg). Finally, a sequential stochastic Sequential Gaussian Simulation was performed on the new CPI. The results of the performed hundred simulations are summarized through the mean image maps and the probability maps of exceeding a given statistical threshold, thus, allowing the characterization of the spatial distribution and associated variability of the CPI. A better understanding of the trends of relative enrichment and PTEs' fate is discussed.
Suggested Reviewers:	Carlos Boente, Ph.D Assistant Professor, Autonomous University of Madrid carboente@gmail.com Dr. Carlos Boente has been working on pollution and risk assessment, namely related to old mining and industrial areas.
	Catarina Silva, Ph.D Assistant Professor, University of Lisbon csilva@fc.ul.pt Dr. Catarina Silva has been working in water pollution and water supply focusing in urban and industry contributions.
	José Luís Gallego, Ph.D Full Professor, University of Oviedo jgallego@uniovi.es Dr. Gallego is an expert in Geochemistry and Environmental Engineering, namely, soil remediation.

	Pedro Proença e Cunaha, Ph.D Full Professor, University of Coimbra pcunha@dct.uc.pt Dr. Proença e Cunha has been working in hazard monitoring related to old mining areas.
	Carlos Sierra, Ph.D University of León sierrafernandezcarlos@gmail.com Dr. Sierra is an expert in Environmental Engineering and Chemical Engineering, namely related to old mining and industrial areas.
	Alexandre Tavares, Ph.D Associate Professor, University of Coimbra atavares@ci.uc.pt Dr. Tavares has been working in environmental risk assessment.

Castelo Branco, 25 August 2023.

Dear Prof. Dr., José L. Domingo

I hereby submit to the journal Environment Research the manuscript entitled “Stream sediments pollution: a compositional baseline assessment – the Caveira mine, Portugal”. The manuscript tackles an important topic regarding natural hazards namely what concerns a better understanding of the trends of relative enrichment and Potentially Toxic Elements (PTE) fate related to old mining activities. The methodological approach developed can be easily applied to other areas under acid waters’ drainage and is a key support tool for decision-makers.

All authors have seen and accepted the manuscript's contents. The submitted manuscript describes the original work and we all confirm that results have not been published elsewhere and are not under consideration by another journal.

Yours sincerely,

Teresa Albuquerque
(Correspondent author)

Highlights

- A high concentration of Potentially Toxic Elements (PTEs) can affect ecosystem health in many ways
- Two questions must be addressed when quantifying pollution. How to define a non-polluted sample? and, how to reduce the problems' dimensionality
- A novel Compositional Pollution Indicator (CPI) based on Compositional Data (CoDa) was computed
- Sequential Gaussian Simulation was performed on the new CPI and probability maps of exceeding a given statistical threshold computed

Stream sediments pollution: a compositional baseline assessment – the Caveira mine, Portugal

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

[1] Instituto Politécnico de Castelo Branco, Av. Pedro Álvares Cabral, nº 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, nº 12, 6000-084 Castelo Branco, Portugal.
[3] ICT| Universidade de Évora, Largo dos Colegiais 2, 7000 Évora, Portugal.

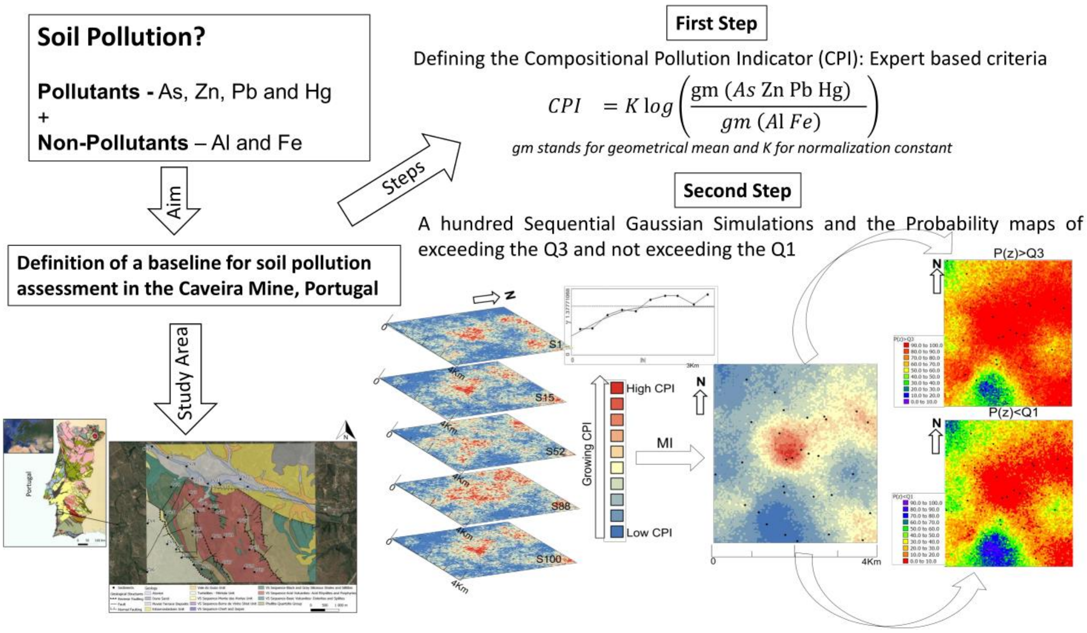
Abstract

A high concentration of Potentially Toxic Elements (PTEs) can affect ecosystem health in many ways. It is therefore essential that spatial trends of pollutants are assessed and controlled. Two questions must be addressed when quantifying pollution. How to define a non-polluted sample? and, how to reduce the problems' dimensionality. Since the concentration of chemical elements is compositional, a compositional approach was used as the attributes vary together. A novel Compositional Pollution Indicator (CPI) based on Compositional Data (CoDa) principles such as sparsity and simplicity as properties, was computed. A dataset of 33 stream-sediment samples was collected from within 0 to 10 cm depth, in a grid of 1Km x 1Km, and twelve chemical elements were analyzed. Concentrations, reaching 3.8% Pb, 750µgg-1 As, and 340 µgg-1 Hg, were obtained near the mine tailings. The methodological approach implied the geological background selection in terms of a trimmed subsample that can be assumed as non-pollutant (Al and Fe) and the selection of a list of pollutants based on expert knowledge criteria and previous studies (As, Zn, Pb, and Hg). Finally, a sequential stochastic Sequential Gaussian Simulation was performed on the new CPI. The results of the performed hundred simulations are summarized through the mean image maps and the probability maps of exceeding a given statistical threshold, thus, allowing the characterization of the spatial distribution and associated variability of the CPI. A better understanding of the trends of relative enrichment and PTEs' fate is discussed.

Key Words: Caveira mine; Pollution; Compositional Pollution Indicator (CPI); Sequential Gaussian Simulation; Probability map.

*Corresponding author.
E-mail address: teresal@ (M-T.D. Albuquerque)

Graphical Abstract



1. Introduction

Potentially toxic elements (PTE) constitute a significant liability in different environmental sectors. Over the past decades, the cumulative environmental impact has been enormous. The problem of PTEs in stream sediments has led during the time to an exponential concentration increase and, therefore, excelled the human and environmental risks (Antoniadis et al., 2017; Kumudunis et al., 2020). Mining and heavy industrial activities may potentiate these high observed levels of PTEs and be the origin of numerous sources of contamination (Boente et al., 2022, Carvalho et al. 2022, Boente et al., 2018). Thus, in recent decades, researchers have invested in the development of new techniques to offer accurate scenarios of the spatial distribution of PETs. 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). Visualization of spatial-temporal distribution models using simulated maps is an important tool for the visualization and depiction of pollutants. The definition of vulnerability and risk hot clusters may act as the basis for supporting environmental policy-making in complex scenarios (Boente et al., 2020, Albuquerque et al., 2017, McKinley et al., 2016). In the area of soil sediment science, a common strategy for describing the distribution of PTEs is to map a single-component synthesis new variable called indices or indicators. Unfortunately, usually, the compositional nature of the geochemical data is considered (Pawlowsky-Glahn et al., 2015, Filzmoser et al., 2009,). In most cases, these indicators are related to the study of individual elements, without considering the existing dependence between the concentrations of all elements in the same set. usually use The non-compositional indices often used to study geochemical data are the Geoaccumulation Index (Muller, 1969), the Enrichment Factor (Sucharova et al., 2012), or the Single Pollution Index (SPI) (Hakanson, 1980), 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 and his fundamental work on the method of logarithmic ratio (1982, 1986) answered these questions. Theories of composition data (CoDa) have enhanced the understanding of the sampling space of composition data and their corresponding structure (Pawlowsky-Glahn and Egozcue 2001). Representations of data considering pairwise log ratios (pwlr), isometric lg-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 PTEs soil pollution characterization is increasing (Boente et al., 2022, Petrik et al., 2018). Compositional indicators involving geochemical baselines definition offer a valuable contribution as they are scale-invariant and sub-compositionally coherent, meaning that a change in the concentration's unit will not modify the study's results (Pawlowsky-Glahn et al., 2015).

This research introduces a new Compositional Pollution Indicator (CPI) of riverine sediments, built to characterize pollution in the Caveira mine in southern Portugal, using the approach developed by Boente et al. (2022) corresponding to a balance of elements chosen through expert criteria although respecting the same CoDa principles.

2. Material and Methods

2.1 Characteristics of the study area and the data set

The studied sector is part of the Portuguese Iberian Pyrite Belt and is an example of post-mining European areas back to the 1990s. Mining activity ceased mainly because of ore exhaustion and more profitable methods worldwide which resulted in 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 noteworthy. The geological sequence at Caveira mine which closed back in the 1980s, corresponds from bottom to top to phyllites and quartzites (PQG), followed by a volcanic sedimentary complex sequence (VSC) unit (Late Famennian) represented by pyroclastics, rhyolitic lavas, tuffs, dark grey, and siliceous shales, and rare jaspers. Intruding diabase rocks are spotted 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 with 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, with Variscan age. Tailings, and associated waste rock, resulting from 129 years of pyrite and Cu mining, are scattered along the Grândola Creek. The semi-arid climatic conditions encompass high erosion of residues by surface water, primarily during rainfall, causing serious contamination of the Grândola stream and its tributaries, conducted to the degradation of sediments (Ferreira da Silva et al, 2015).

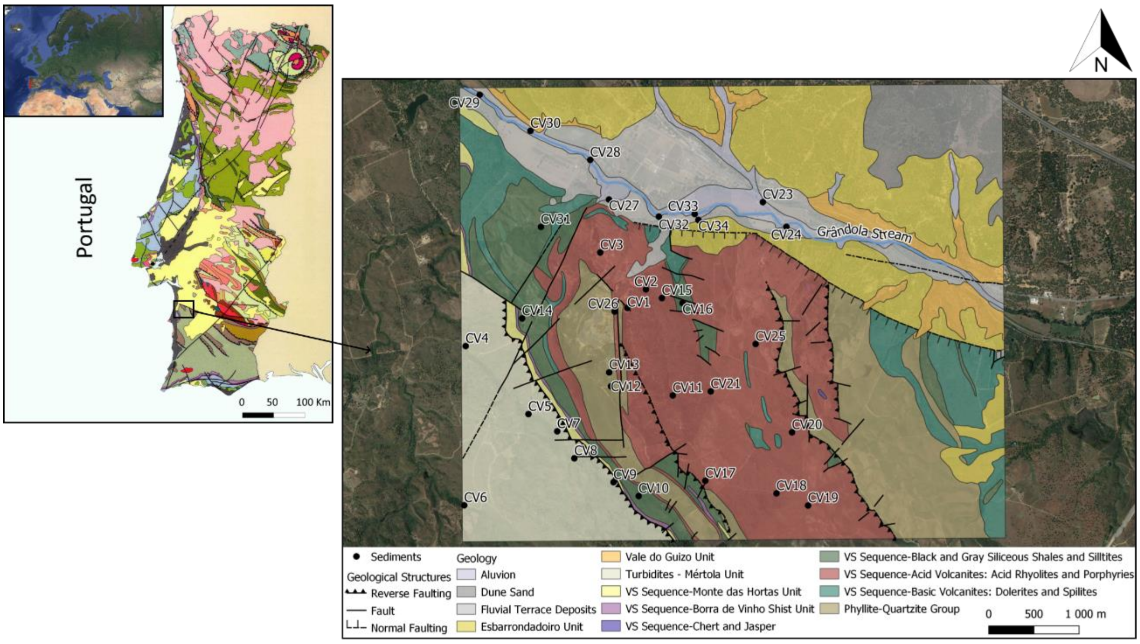


Fig 1. Study area and collected sample's location.

A dataset of 33 bottom sediment samples distributed over small and narrow creeks two of them flowing by the mine tailings pile and the larger Grândola stream, of which they are tributaries. These streams belong to the Sado River Basin, the second-largest hydrographical basin in Southern Portugal. Samples were collected from within 0 to 10 cm depth with an environmental hand soil sampling kit (#209.55, AMS), in a grid of 1Km x 1Km and twelve 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), were analyzed in preserved samples at about 4°C. The most extractable forms of metals (except for Hg) were obtained by partial digestion with aqua regia (HCl and HNO3) in a high-pressure microwave digestion unit (Anton Paar Multiwave PRO) following the US EPA (2007) Method 3051A. Metals and As were analyzed by optical emission spectroscopy with an inductive plasma source (ICP-OES, Perkin-Elmer OPTIMA 8300), using yttrium as an internal standard. The accuracy and analytical precision of all the analyses have been checked by 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 Consort, C5020 (SP10T model for pH, SP50X model for redox potential). In samples with insufficient moisture for direct pH readings, this parameter was measured in water–sediment suspension (2.5:1), in the laboratory. Respecting to samples' chemistry, the dataset includes PTEs of variable toxicity (Fabian et al., 2014). The set of 12 elements was reported across the 33 sampling points, resulting in a 12-part composition that is assumed to represent the stream sediments.

2.2 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 are explained and expanded into general-purpose works such as 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, given by its chemical composition should be conducted under the assumption that these 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 (components adding to 1) without adding or losing any information, irrespective of the units in which the data were initially represented.

The chemical composition of a sample of riverine sediments in units such as mg/kg should be performed assuming that these data are compositional. Moreover, the conversion of units from mg/kg to g/kg, as an example, must not change the information in the sample. This is summed up in 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 in proportions (components adding 1) without adding or losing information regardless of the units in which the data were originally reported. A second assumption is known as Sub compositional Coherence Principle. The whole periodic table is never presented, only a subset of elements is measured, and this subset may change in time and the field. The elements observed form a composition and any subassembly of the same is a sub-composition, again subject to the Principle of Scale Invariance. Analyses of initial composition or sub-composition should lead to coherent conclusions describing the role 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, CPI, was constructed as follows:

$$\mathrm{CPI} = \sqrt{\frac{4}{3}} \ln\left(\frac{(As\ \mathrm{Zn\ Pb\ Hg})^{1/4}}{(Al\ Fe)^{\frac{1}{2}}}\right) \qquad (1)$$

2.3 Spatial modeling – geostatistical approach

The computed Compositional Pollution Indicator (CPI) is unbounded, a real random variable. Therefore, it fulfills the assumptions underlying a conventional geostatistical approach. Their spatial probability patterns were computed following a two-step geostatistical modeling method: 1. Structural analysis and experimental variograms computation (Journel and Huijbregts,1978) followed by 2. Sequential Gaussian Simulation (SGS) is 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 & Huijbregts, 1978).

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

The arguments taken into consideration are *h* (distance) where $Z\ (x_i)$ and $Z\ (x_i+h)$ are the numerical values of the variables assigned to xi and xi 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 existing in the geometric field spaced at an h distance (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 (C_o) effect, which supplies the behavior at the origin. The two other parameters are the sill (C_1) and the amplitude *(a)* which define correspondingly the inertia used in the subsequent interpolation process and the influence radius of the variable.

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, reproducing the statistics of the known data and building 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 and the Probability maps of exceeding the third quartile (Q_3) and not exceeding the first quartile (Q_1) were computed.

3. Results and Discussion

3.1 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), were evaluated considering their capacity of solubilization and mobilization aiming contamination mapping. The evaluation of the 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. Considering the physical-chemical parameter that most affects the solubility, mobility, and precipitation of potentially toxic metals in the sediments from shallow streams, pH, values range from 2.06 and 7.39, corresponding to the lower values (2.06-4.57) to the sediments from the 2 creeks flowing through the mining tailings pile. As would be expected, these sediments (Cv1, Cv2, Cv3, Cv26, Cv 33, Cv34) are those with the highest values of Pb, As, and Hg, the main contaminants in the mine tailings that reach levels above those considered critical and which require immediate intervention, according to the European Regulations (based on the Netherlands legislation – Soil Quality Regulation, 2006). Zn, another element with levels of concern, and which represents one of the elements with high contents in the massive sulphides that have been exploited in this mine, presents slightly contaminating levels in all the diffused streams flowing from the tailings pile, mostly in locations that do not coincide with the locations where the other elements have exceeded critical levels. The highest values of this element also do not coincide with the most acidic conditions of the environment. Although any of these elements originate from the ores that were exploited in this mining area, Zn is an element with higher chemical mobility, and because this mobility is mostly influenced by the oxidation conditions that occurred in all sediments (240 – 650 mV), its distribution is more diffuse.

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

Samples * (mg/Kg)	As	Cd	Co	Cr	Mn	Ni	V	Zn	Al	Fe	Pb	Hg	pH (H_2O)
CV1	234,9	1,5	0,4	2,3	16,9	1,6	3,3	94,2	1441,6	10217,0	38458,7	127,9	2,30
CV2	238,6	< 0,5	0,4	1,6	24,5	1,0	2,0	96,4	2168,4	17815,8	29837,2	78,2	2,06
CV3	517,6	< 0,5	0,8	1,5	99,6	0,9	2,9	122,3	16128,4	65825,8	13775,3	125,8	4,57
CV4	< 1,5	< 0,5	3,9	13,3	172,7	10,3	14,3	36,4	18321,3	31828,4	51,2	1,7	6,61
CV5	14,7	< 0,5	3,5	15,6	261,9	12,7	15,5	47,2	13099,4	27183,4	2881,5	13,7	6,00
CV6	< 1,5	< 0,5	3,4	17,6	208,4	9,3	24,3	30,0	26272,6	21172,3	221,4	0,2	6,54
CV7	< 1,5	< 0,5	4,2	18,6	399,1	11,2	29,2	36,4	29376,6	28193,2	97,2	0,6	6,64
CV8	< 1,5	< 0,5	4,6	19,5	455,6	12,3	25,7	37,5	24811,4	28361,9	15,0	0,3	7,17
CV9	< 1,5	< 0,5	5,8	13,4	414,1	12,4	13,8	34,7	13007,8	27207,6	16,4	0,2	4,99
CV10	< 1,5	< 0,5	5,6	15	315	21,6	22,9	46,3	23727,5	23919,7	18,2	0,6	5,12
CV11	3,4	< 0,5	1,0	1,4	148	1,2	2,6	15,8	8807,1	12058,1	21,5	0,3	5,05
CV12	31,2	< 0,5	4,2	9,6	138,6	10,6	15,9	140,4	11176,5	29630,2	92,7	2,6	5,30
CV13	1,6	< 0,5	34,4	23,8	903	21,8	36,4	165,8	44359,5	42371,5	36,3	1,6	7,29
CV14	1,6	< 0,5	4,1	16,1	434,9	15,9	24,3	47,5	18683,1	21463,1	36,6	0,6	6,03
CV15	1,7	< 0,5	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	< 0,5	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	< 0,5	1,6	4,3	412,9	4,0	5,2	15,8	10396,1	16746,8	27,1	0,1	6,35
CV18	< 1,5	< 0,5	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	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	< 0,5	1,0	4,5	169,6	1,5	6,8	10,6	3677	5856,6	9,8	0,2	5,82
CV21	< 1,5	< 0,5	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	< 0,5	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	< 0,5	1,5	7,0	79,2	5,8	12,1	122,4	9728,8	13370,1	585,3	5,7	6,74
CV25	< 1,5	< 0,5	0,4	1,5	84,8	0	3,7	6,3	14787,2	7836,1	4,4	0,1	6,63
CV26	140,3	< 0,5	0,4	1,6	34,4	2,0	2,6	123,4	1193,7	10730,9	44540,5	46,8	2,34
CV27	32,3	< 0,5	3,6	11,6	177,5	12,5	17,9	129,0	13208,0	25540,5	82,8	7,3	6,39
CV28	< 1,5	< 0,5	3,7	16	385,5	10,1	24,6	36,0	22281,1	17596,5	29,1	1,3	5,34
CV29	< 1,5	< 0,5	3,9	18,4	394,6	10,7	25,6	36,2	22665,1	18556,6	14,5	0,2	6,82
CV30	< 1,5	< 0,5	4,1	13,8	258,4	11,8	15,9	39,8	11984,7	19370,5	21,5	0,1	6,16
CV31	9,0	< 0,5	2,1	14,2	106,4	10,5	15,7	30,2	6909,5	23545,1	31,0	0,2	6,21
CV32	30,8	< 0,5	3,4	15,7	291,3	8,9	24,6	88,6	24220,0	19152,6	107,7	2,3	6,57
CV33	265,2	< 0,5	3,9	9,3	422,5	8,6	11,6	108,2	11998,7	30136,3	88,8	1,6	4,26
CV34	748,2	< 0,5	0,4	8,3	77,6	3,9	10,8	161,8	6854,4	44363,4	23162,4	381,4	2,13

*Sample Cv22 was eliminated

Furthermore, a heat map was used for data exploratory analysis of the geochemical composition and sample clustering simultaneously in a synthetic way (Fig.2) ((Wilkinson and Friendly, 2009; Langella et al, 2013). The heat map observation shows the elements divided into two groups (upper dendrogram). The first one corresponds to Al and Fe (non-pollutants) and the second one corresponds to As, Cd, Co, Cr, Mn, Ni, V, Zn, Hg, and Pb. This last element shows a distance within the major group. Based on an expert-driven approach As Zn Pb Hg were selected as pollutants. The samples dendrogram (left one). is divided into three groups The central group corresponds to samples Cv6; Cv8; Cv10; Cv13; C14; Cv16; Cv17; Cv18; Cv21; Cv23; Cv25; Cv28; Cv29 and Cv32; the left group to samples Cv1; Cv2; Cv15 and Cv26 and the right group to samples CV3; Cv4; Cv5; Cv9; Cv11; Cv12; Cv19; Cv20; Cv24 and Cv27; Cv30; Cv31; Cv33 and Cv34. The map represents values in the dataset re-arranged according to the dendrograms. Focusing on rectangle/square patterns (from red to blue through a white increasing level of significance) inside the map it is possible to see, concerning the bottom group of samples Cv1; Cv2; Cv15, and Cv26, together with samples Cv34 and Cv3 of the upper group a lower significance cluster for Al and a higher significance cluster for Pb, relatively to the other samples. In future work, the relationship between the samples' geochemical print and the associated geology will be explored.

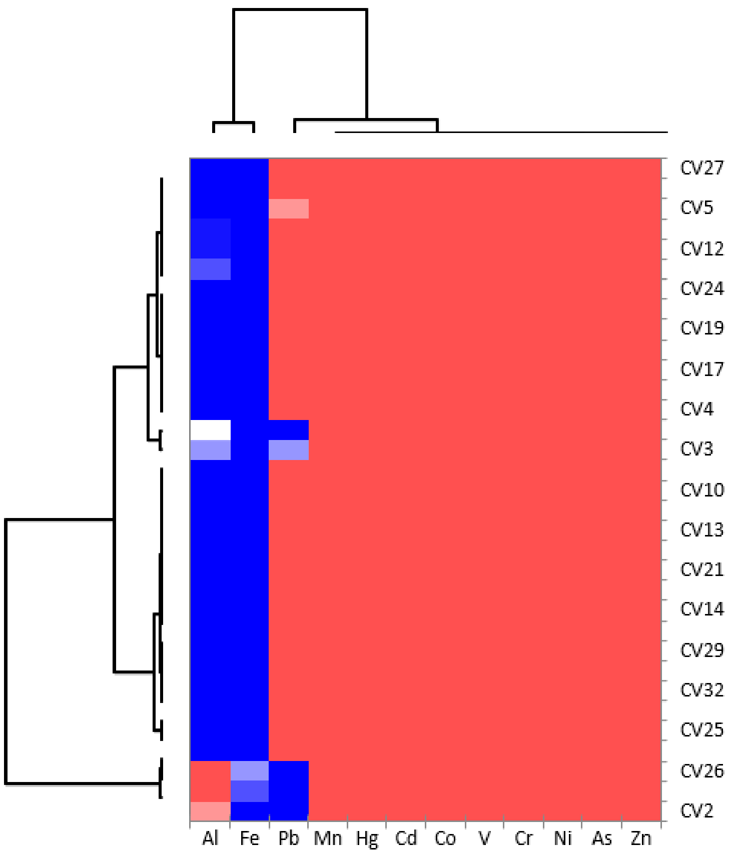


Fig.2. Heat map and simultaneously sample/geochemical print dendrograms.

3.2 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 are identified as related to the Iberian Pyrite Belt old mines activities. While the main natural-source elements (or non-pollutants) were several major elements (i.e., Al and Fe). The CPI spatial modeling aimed at the definition of hazardous clusters. Thus, a two-step geostatistical approach was used. The experimental isotropic variogram was computed as no clear evidence of anisotropy was found and the corresponding fitted model is shown in Fig. 2. Cross-validation correlation index of the observed and estimated CPI values is 0.70 and, 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.

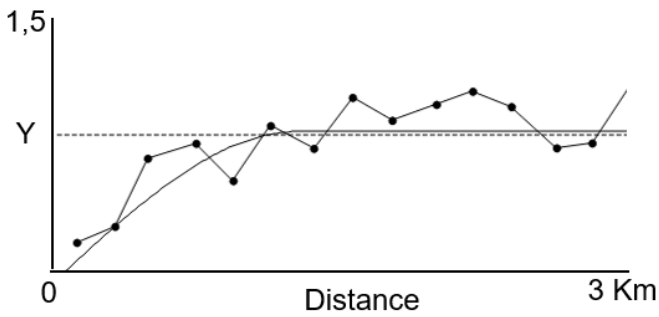


Fig.2. Experimental and fitted spherical omnidirectional variogram.

Probability maps, corresponding to different thresholds allowed the visualization of spatial variability setting aside the discussion of local accuracy and allowing the identification of hot clusters of pollution in the subject area. The realization numbers 1, 15, 32, 52, 67 and 99 are shown in Fig.3.

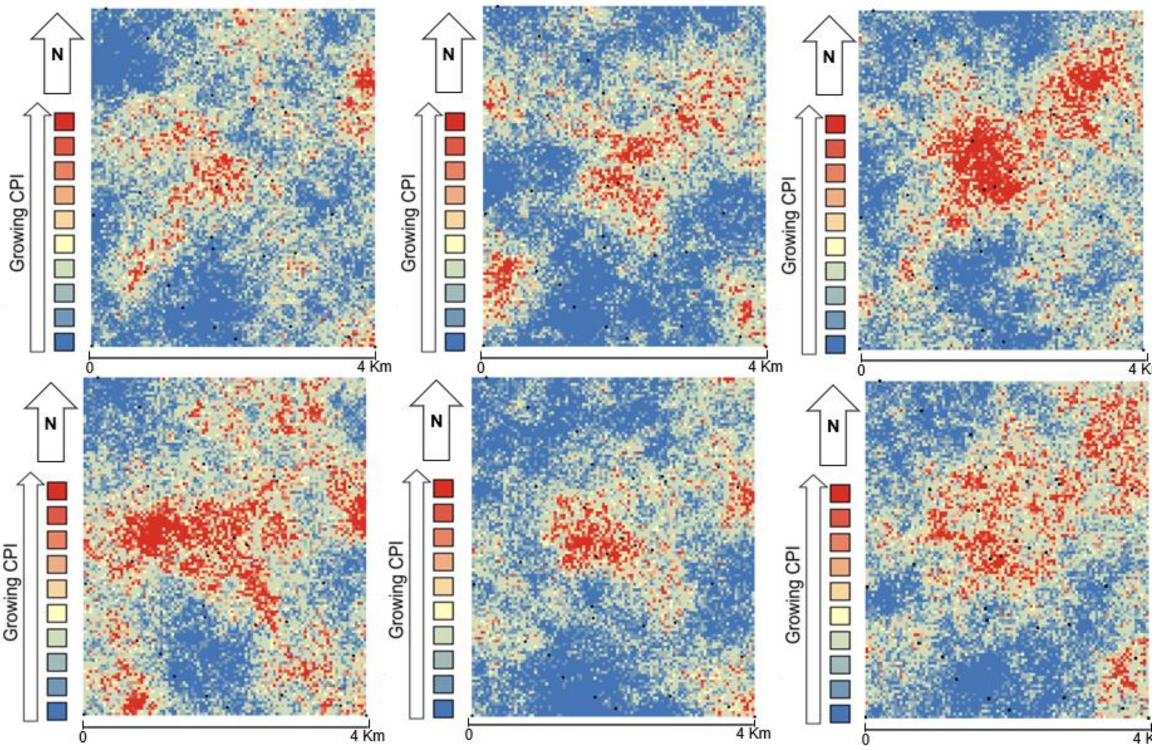


Fig.3. Six different scenarios were obtained by Sequential Gaussian Simulation (SGS).

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 images (MI) - average map – was computed and used as the CPI spatial distribution (Fig.4 a)) The representation of the probability of exceeding the third quartile (Q3) and not exceeding the first quartile (Q1), allows broad discussion of the CPI spatial distribution and the identification of hazard clustering (Fig.4 b) and 4 c)). To create distinct classes, reducing the within classes' variance and maximizing the in-between classes variance, the Jenks natural break classification (Jenks, 1967) was used, allowing the determination of the best arrangement of values. For the computation, the Space-Stat Software V. 4.0.18, Biomedwere, was used (Boente et al., 2022, Albuquerque et al., 2017).

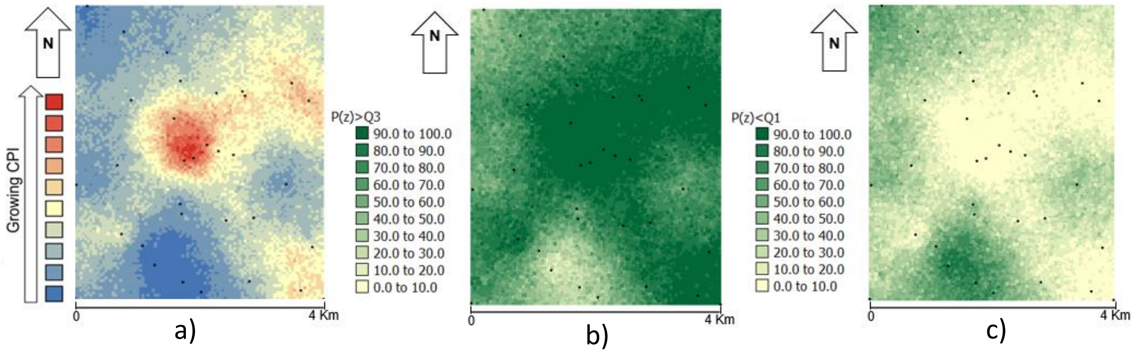


Fig.4. a) SGS average image (MI) b) Probability map of exceeding Q3 and c) Probability map of not exceeding Q1.

The compositional Pollution Indicator (CPI) shows a fair representation of hot spots, especially along the Grândola stream and its tributaries, thereby confirming the larger pollution detected around the old mine tailings and associated waste rock.

4. Conclusions

Geochemical data are compositional data, as the concentrations of elements in any environmental matrix are commonly expressed as parts of a whole and vary together. Once this feature is established, compositional data procedures can be applied to obtain indicators that address pollution, for example, in stream sediment. The method was tested with 33 sediment samples and up to 11 chemical elements from the old Caveira mine in Portugal. Specifically, in the vicinity of the Grândola River and the mine's tailings, the survey revealed a significant risk of contamination. In addition, agriculture is the main focus of economic activities in the region. Two primary courses of action are proposed in light of this:

1. Installation of a surveillance network: The first step is to establish a continuous surveillance and control network in all regions. Setting up a system is necessary to consistently track and assess the levels of contamination in the region. This monitoring would likely involve the installation of sensors for continuous measurement of the PTEs content and a regular collection of in situ samples for validation and to identify any significant changes or deviations.

2. Mitigation strategies for contaminated areas: The second strategy is to focus on the northern region where elevated levels of contamination have been detected. The adverse effects of pollution require the development and implementation of mitigation measures in this region. Efforts to reduce the introduction of contaminants may involve remediation efforts, ecosystem restoration, or changes in local practices.

The survey highlights the interconnectedness of geochemical data and how compositional data procedures can be used to shed light on environmental issues like pollution. It highlights the practical application of these principles through the case study of the Caveira mine area in Portugal, and it emphasizes the importance of addressing contamination risks in a region where agriculture and organic activities are key components of the local economy.

Acknowledgments

The data used in this survey has been presented previously at geoENV2022 in the poster section topic Environmental Pollution and Risk Assessment.

The authors acknowledge the funding provided by the Foundation “LA CAIXA” (Spain) and the Foundation for Science and Technology (FCT) (Portugal), which supported this research through the GeoMatre project: La Caixa/FCT - Project n.º PV20-00006 and, also, the funding provided by the Research Centre for Natural Resources, Environment and Society (CERNAS-IPCB) [project UIDB/00681/2020] funding by Portuguese Foundation for Science and Technology (FCT) and the funding provided by ICT, under contract with FCT (Portuguese Science and Technology Foundation) under the Project FCT—UIDB/04683/2020.

During the preparation of this work, the author(s) used ChatGPT to polish the English, in the conclusions section. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

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 & Hall Ltd., London (UK) ((Reprinted in 2003 with additional material by The Blackburn Press). 416 pp.).

Albuquerque, M.T.D., Gerassis, S., Sierra, C., Taboada, J.,·Martín, J.E., Antunes, I.M.H.R., Gallego, J.R., 2017. Developing a New Bayesian Risk Index for Risk Evaluation of Soil Contamination. Sci of the Total Env., 603-604, 167-177, DOI: 10.1016/j.scitotenv.2017.06.068.

Antoniadis, V., Golia, E.E., Shaheen, S.M., Rinklebe, J., 2017. Bioavailability and health risk assessment of potentially toxic elements in Thriasio Plain, near Athens, Greece.

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, ISSN: 0048-9697Volume: 812, 152383 (12), DOI: 10.1016/j.scitotenv.2021.152383.

Boente, C., Albuquerque, M.T.D., Fernandez-Brana, A., Gerassis, S., Sierra, C., Gallego, J.R., 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, M.T.D., Taboada, J., Gallego, J.R., 2020. Local versus regional soil screening levels to identify potentially polluted areas. Math. Geosci. 52, 381–396.

Boogaart van den, K.G., Tolosana-Delgado, R., 2013. Analysing Compositional Data with R. Springer-Verlag, Berlin (258 pp.).

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, P.C.S., Antunes, I.M.H.R., Albuquerque, M.T.D., Santos, A.C.S., Cunha, P.P., 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, DOI: 0.1007/s12665-022-10275-2.

Cicchella, D., Zuzolo, D., Albanese, S., Fedele, L., Tota, D., Guagliardi, Ilaria, Thiombane, Matar, Vivo, Benedetto De, Lima, Annamaria, 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.

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

Chayes, F., 1971. Ratio Correlation. University of Chicago Press, Chicago, IL (USA) (99 pp.).

Egozcue, J.J., 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 M.R., 2015. An integrative assessment of environmental degradation of Caveira. abandoned mine area (Southern Portugal). Journal of Geochemical Exploration 159 (2015) 33–47, http://dx.doi.org/10.1016/j.gexplo.2015.08.004

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., 2011. Compositional data analysis: theory and applications. In: Pawlowsky-Glahn, V., Buccianti, A. (Eds.), Compositional Data Analysis: Theory And Applications. John Wiley & Sons, pp. 59–72.

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 Publishing.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Applied Geostatistics Series. Oxford University Press, New York, NY (USA) (483 pp.).

Hakanson, L., 1980. An Ecological Risk Index for Aquatic Pollution Control A Sedimentological Approach. Water Research, 14, 975-1001., Bucciantihttp://dx.doi.org/10.1016/0043-1354(80)90143-8

Jenks, G.F., 1967. The data model concept in statistical mapping. International Yearbook of Cartography. 7, pp. 186–190.

Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press, London (UK) (600 pp.).

Kowalska, J.B., Mazurek, R., Gasiorek, M., Zaleski, T., 2018. Pollution indices as useful tools for the comprehensive evaluation of the degree of soil contamination: a review. Environ. Geochem. Health 40, 2395–2420.

Kumuduni, N. P., Sabry M. S., Season S. C., Daniel C.W. T., Yohey H., Deyi H., Nanthi S. B., Jörg R., Yong S. O., 2020. Soil amendments for immobilization of potentially toxic elements in contaminated soils: A critical review. Environment International 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) Management and dissemination of MS proteomic data with PROTICdb: example of a quantitative comparison between methods of protein extraction, Proteomics. 2013 May;13(9):1457-66.

Matheron, G. 1971. The Theory of Regionalized Variables and Its Applications. Les Cahiers du Centre de Morphologie Mathematique in Fontainebleu, Paris.

Martins, L., Oliveira, D., 2000. Exploration and Mining. Instituto Geológico e Mineiro, Lisboa (20 pp.).

McKinley, J.M., Hron, K., Grunsky, E.C., Reimann, C., de Caritat, P., Filzmoser, P., van den Boogaart, K.G., 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, S.T., McKinley, J.M., Marks, N.J., Scantlebury, D.M., Doherty, R., 2021. Heavy metal (pte) ecotoxicology, data review: traditional vs. a compositional approach. Sci. Total Environ. 769 (14524), 6. (378 pp.)

Pawlowsky-Glahn, V., Buccianti, A. (Eds.), 2011. Compositional Data Analysis: Theory and Applications. John Wiley & Sons.

Pawlowsky-Glahn, V., Serra, J. (Eds.), 2019. Matheron's Theory of Regionalised Variables. Oxford University Press (208 pp.).

Pawlowsky-Glahn, V., Egozcue, J.J., Tolosana-Delgado, R., 2015. Modeling and analysis of compositional data. Statistics in Practice. John Wiley & Sons, Chichester UK (272 pp.).

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

Petrik, A., Thiombane, M., Lima, A., Albanese, S., Buscher, J.T., 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.

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, Bilodeau (ISBN: 978-0-387-20331-7).

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.

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. https:// doi. org/ 10. 1198/ tas. 2009. 0033

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

Samples * (mg/Kg)	As	Cd	Co	Cr	Mn	Ni	V	Zn	Al	Fe	Pb	Hg	pH (H_2O)
CV1	234,9	1,5	0,4	2,3	16,9	1,6	3,3	94,2	1441,6	10217,0	38458,7	127,9	2,30
CV2	238,6	< 0,5	0,4	1,6	24,5	1,0	2,0	96,4	2168,4	17815,8	29837,2	78,2	2,06
CV3	517,6	< 0,5	0,8	1,5	99,6	0,9	2,9	122,3	16128,4	65825,8	13775,3	125,8	4,57
CV4	< 1,5	< 0,5	3,9	13,3	172,7	10,3	14,3	36,4	18321,3	31828,4	51,2	1,7	6,61
CV5	14,7	< 0,5	3,5	15,6	261,9	12,7	15,5	47,2	13099,4	27183,4	2881,5	13,7	6,00
CV6	< 1,5	< 0,5	3,4	17,6	208,4	9,3	24,3	30,0	26272,6	21172,3	221,4	0,2	6,54
CV7	< 1,5	< 0,5	4,2	18,6	399,1	11,2	29,2	36,4	29376,6	28193,2	97,2	0,6	6,64
CV8	< 1,5	< 0,5	4,6	19,5	455,6	12,3	25,7	37,5	24811,4	28361,9	15,0	0,3	7,17
CV9	< 1,5	< 0,5	5,8	13,4	414,1	12,4	13,8	34,7	13007,8	27207,6	16,4	0,2	4,99
CV10	< 1,5	< 0,5	5,6	15	315	21,6	22,9	46,3	23727,5	23919,7	18,2	0,6	5,12
CV11	3,4	< 0,5	1,0	1,4	148	1,2	2,6	15,8	8807,1	12058,1	21,5	0,3	5,05
CV12	31,2	< 0,5	4,2	9,6	138,6	10,6	15,9	140,4	11176,5	29630,2	92,7	2,6	5,30
CV13	1,6	< 0,5	34,4	23,8	903	21,8	36,4	165,8	44359,5	42371,5	36,3	1,6	7,29
CV14	1,6	< 0,5	4,1	16,1	434,9	15,9	24,3	47,5	18683,1	21463,1	36,6	0,6	6,03
CV15	1,7	< 0,5	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	< 0,5	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	< 0,5	1,6	4,3	412,9	4,0	5,2	15,8	10396,1	16746,8	27,1	0,1	6,35
CV18	< 1,5	< 0,5	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	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	< 0,5	1,0	4,5	169,6	1,5	6,8	10,6	3677	5856,6	9,8	0,2	5,82
CV21	< 1,5	< 0,5	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	< 0,5	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	< 0,5	1,5	7,0	79,2	5,8	12,1	122,4	9728,8	13370,1	585,3	5,7	6,74
CV25	< 1,5	< 0,5	0,4	1,5	84,8	0	3,7	6,3	14787,2	7836,1	4,4	0,1	6,63
CV26	140,3	< 0,5	0,4	1,6	34,4	2,0	2,6	123,4	1193,7	10730,9	44540,5	46,8	2,34
CV27	32,3	< 0,5	3,6	11,6	177,5	12,5	17,9	129,0	13208,0	25540,5	82,8	7,3	6,39
CV28	< 1,5	< 0,5	3,7	16	385,5	10,1	24,6	36,0	22281,1	17596,5	29,1	1,3	5,34
CV29	< 1,5	< 0,5	3,9	18,4	394,6	10,7	25,6	36,2	22665,1	18556,6	14,5	0,2	6,82
CV30	< 1,5	< 0,5	4,1	13,8	258,4	11,8	15,9	39,8	11984,7	19370,5	21,5	0,1	6,16
CV31	9,0	< 0,5	2,1	14,2	106,4	10,5	15,7	30,2	6909,5	23545,1	31,0	0,2	6,21
CV32	30,8	< 0,5	3,4	15,7	291,3	8,9	24,6	88,6	24220,0	19152,6	107,7	2,3	6,57
CV33	265,2	< 0,5	3,9	9,3	422,5	8,6	11,6	108,2	11998,7	30136,3	88,8	1,6	4,26
CV34	748,2	< 0,5	0,4	8,3	77,6	3,9	10,8	161,8	6854,4	44363,4	23162,4	381,4	2,13

Figure1

Click here to access/download;Figure;Fig.1.png

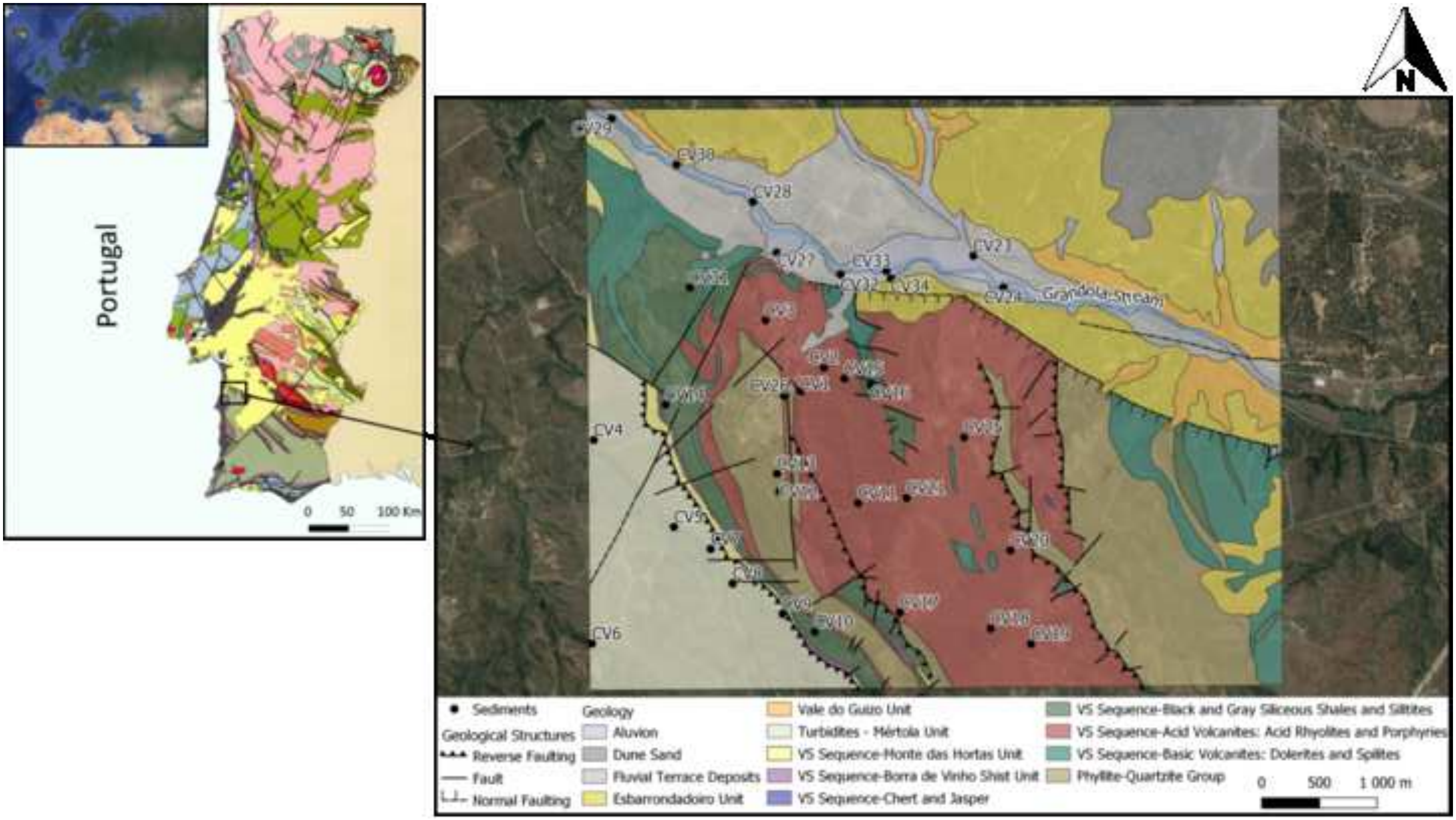


Figure2

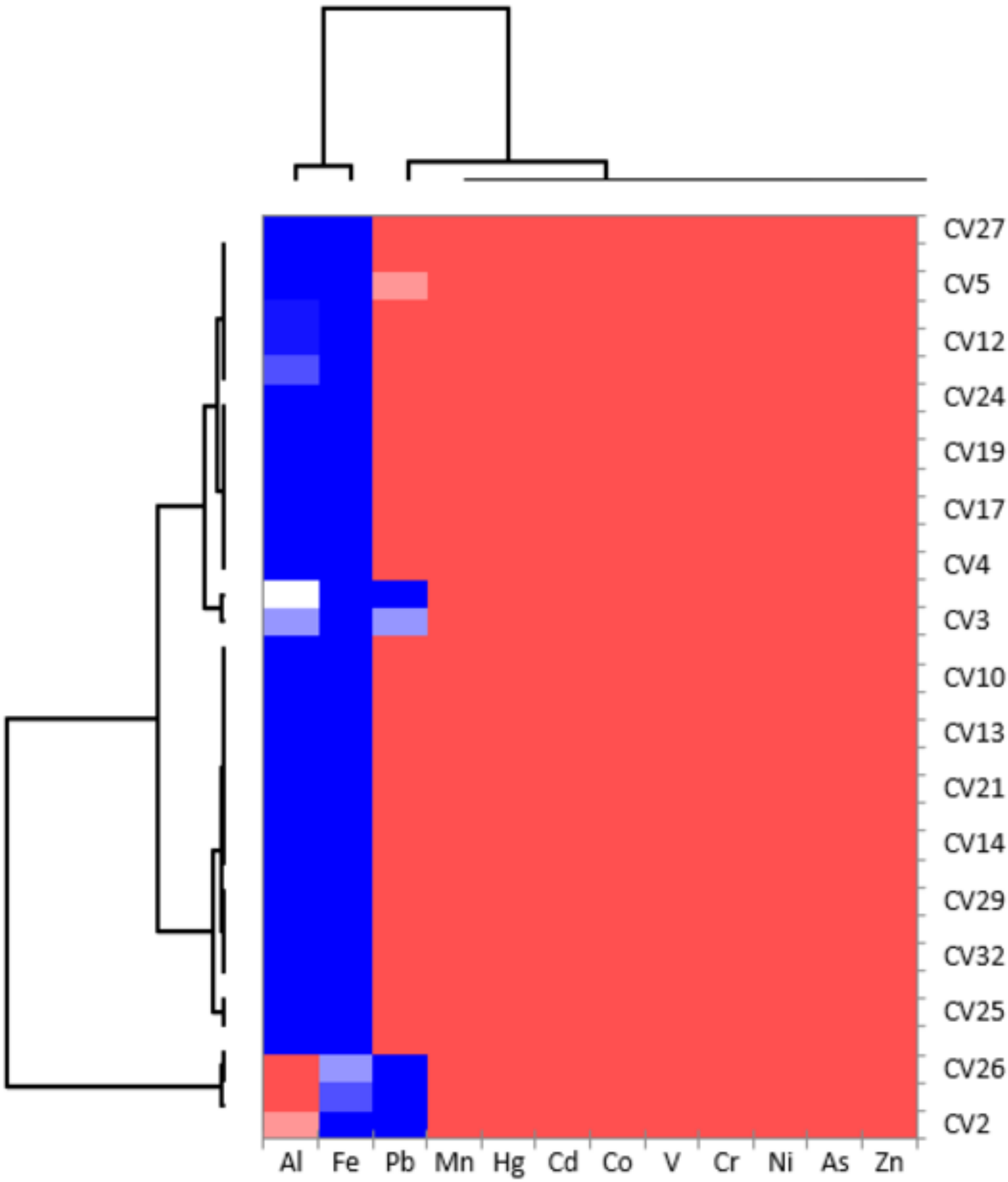

CV27
CV5
CV12
CV24
CV19
CV17
CV4
CV3
CV10
CV13
CV21
CV14
CV29
CV32
CV25
CV26
CV2
Al
Fe
Pb
Mn
Hg
Cd
Co
V
Cr
Ni
As
Zn

Figure3

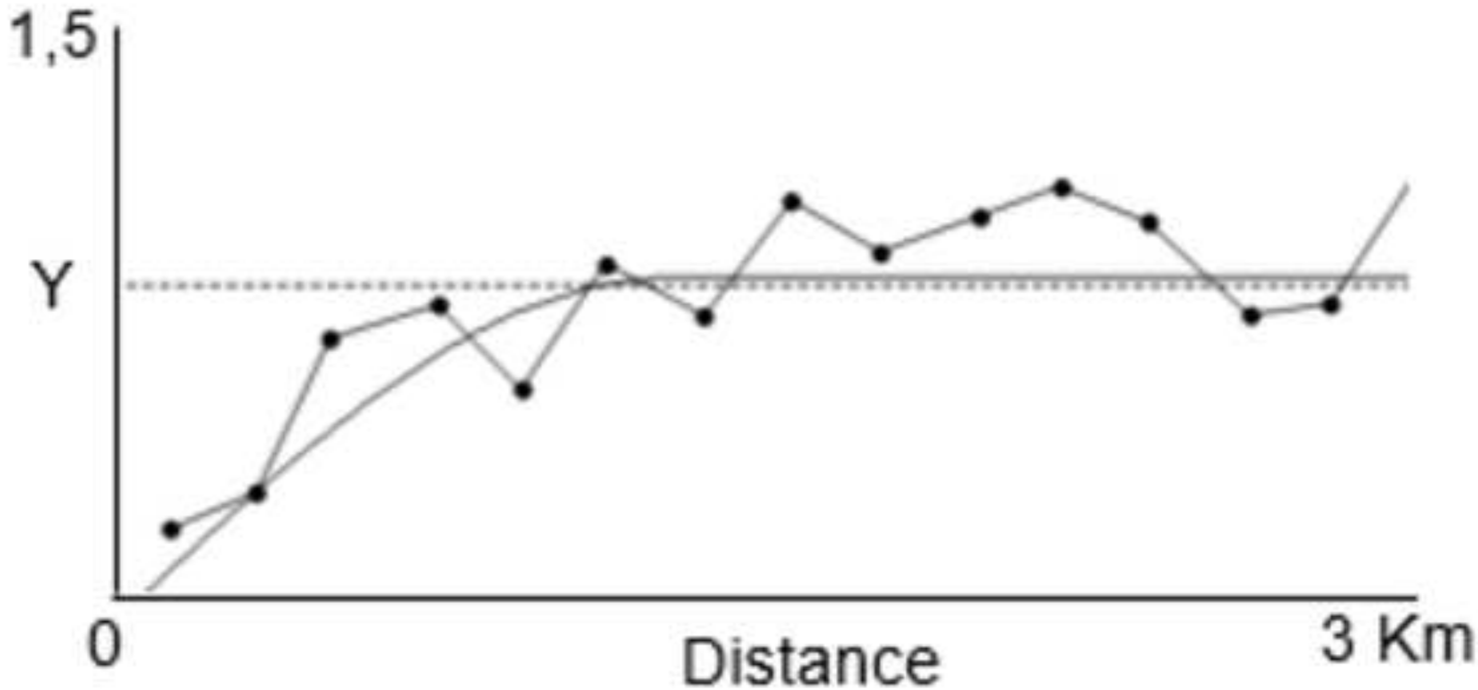


Figure4

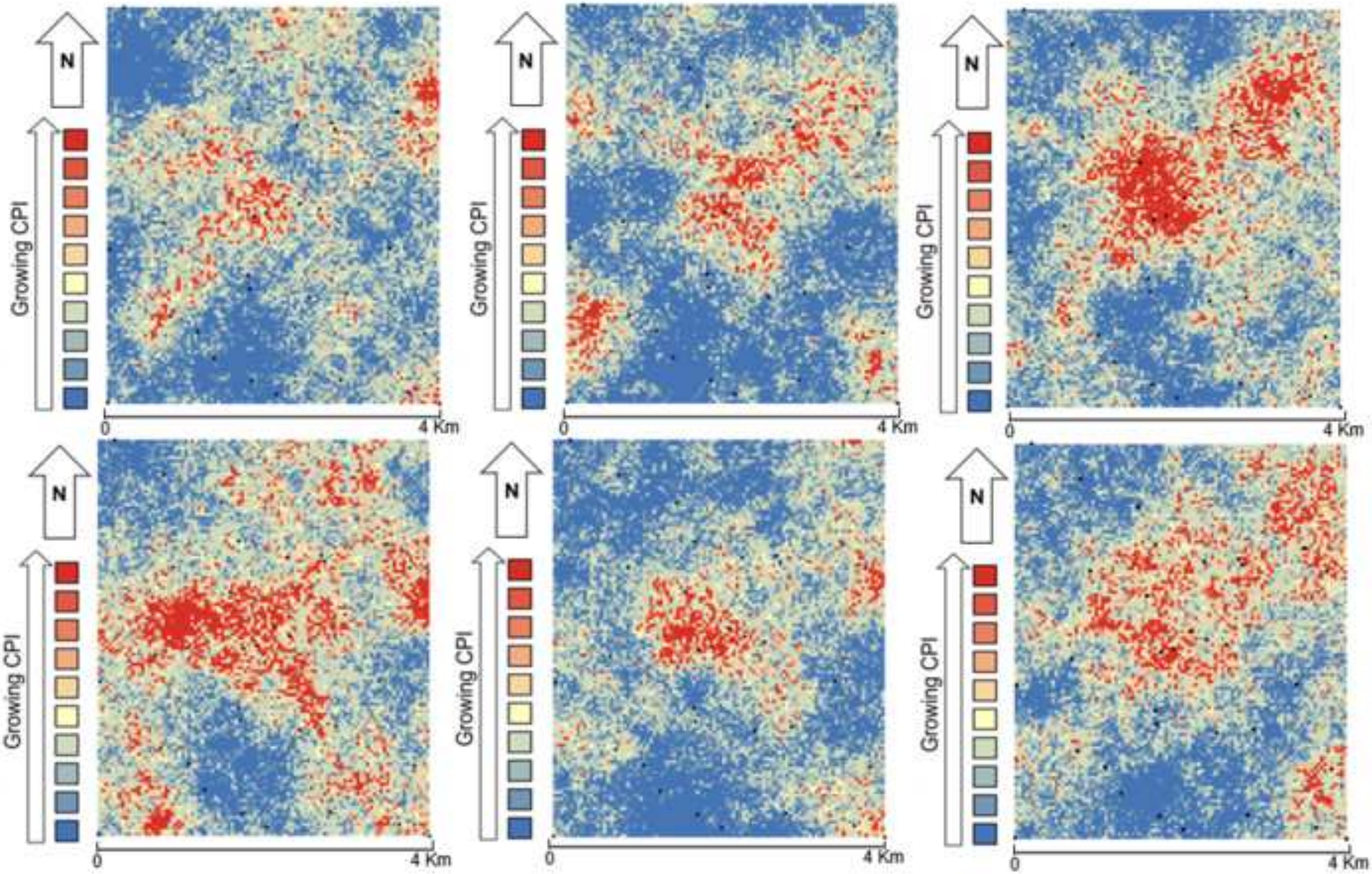
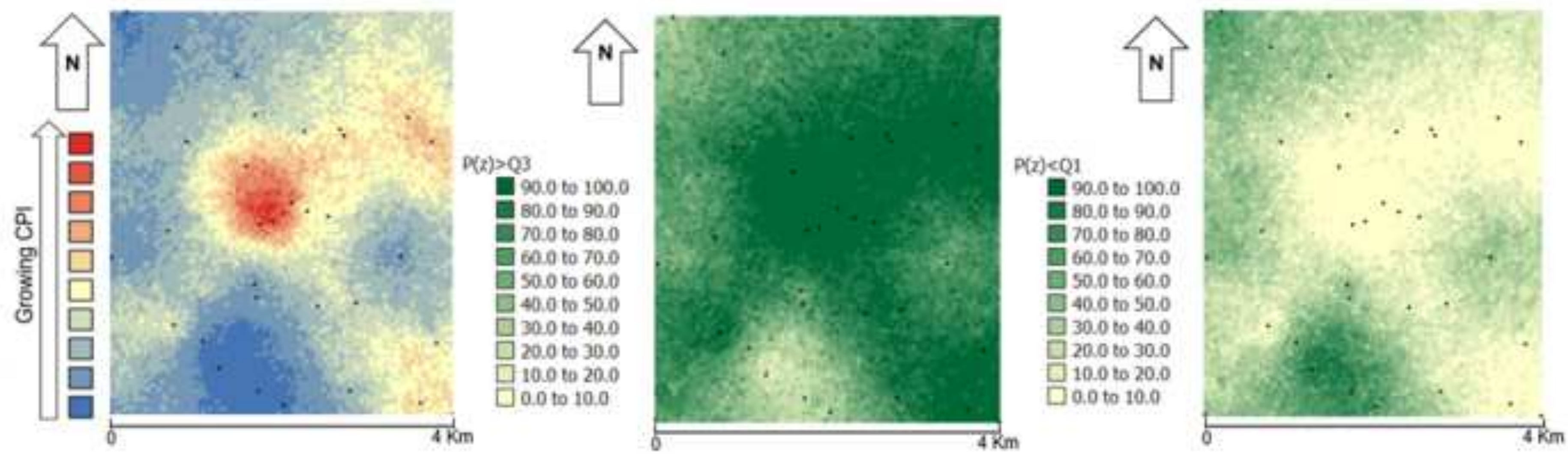

N
Growing CPI
0
4 Km
N
Growing CPI
0
4 Km
N
Growing CPI
0
4 Km
N
Growing CPI
0
4 Km
N
Growing CPI
0
4 Km
N
Growing CPI
0
4 Km

Figure5

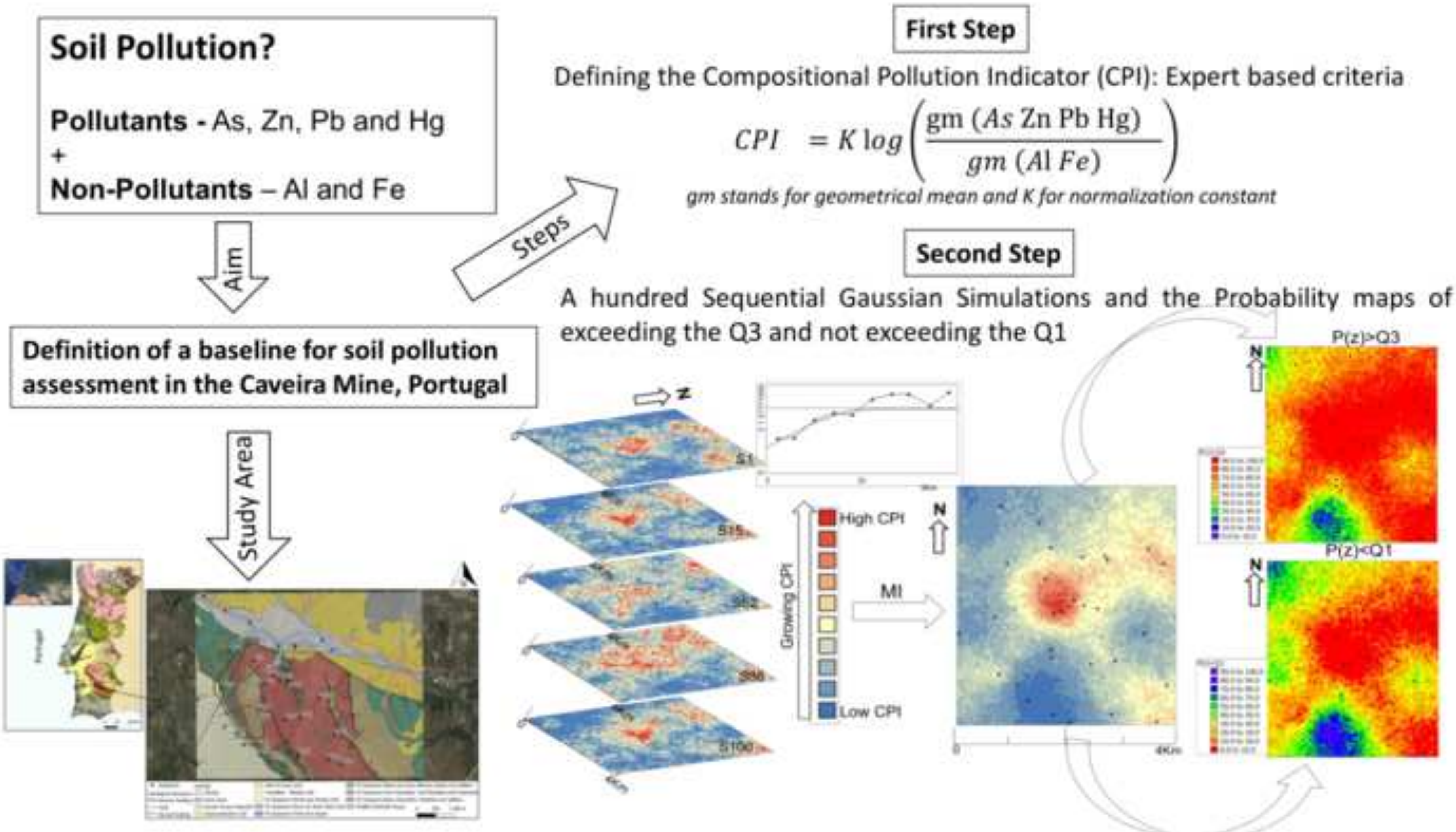
Click here to access/download;Figure;Fig.5.PNG



Declaration of interests

☒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.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:


Soil Pollution?
Pollutants - As, Zn, Pb and Hg
+
Non-Pollutants – Al and Fe
Aim
Definition of a baseline for soil pollution assessment in the Caveira Mine, Portugal
Study Area
Steps
First Step
Defining the Compositional Pollution Indicator (CPI): Expert based criteria
$$CPI = K \log\left(\frac{gm\ (As\ Zn\ Pb\ Hg)}{gm\ (Al\ Fe)}\right)$$
gm stands for geometrical mean and K for normalization constant
Second Step
A hundred Sequential Gaussian Simulations and the Probability maps of exceeding the Q3 and not exceeding the Q1
Z
S1
S15
S52
S86
S100
High CPI
Low CPI
Growing CPI
MI
N
P(z)>Q3
P(z)<Q1
0
4Km