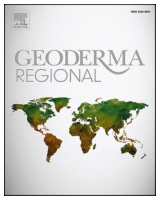
Geoderma Regional 38 (2024) e00845

Contents lists available at ScienceDirect

Geoderma Regional

journal homepage: www.elsevier.com/locate/geodrs



Spatial distribution of soil rare earth elements in Sicily (Italy)

Filippo Saiano [a], Gabriele Buttafuoco [b,*], Maria Fantappié [c], Simone Priori [d], Edoardo A.C. Costantini [e], Riccardo Scalenghe [a]

[a] Università degli Studi di Palermo, Palermo, Italy
[b] National Research Council of Italy, Institute for Agriculture and Forestry Systems in the Mediterranean, Rende CS, Italy
[c] CREA Centro di ricerca Agricoltura e Ambiente, Florence, Italy
[d] University of Tuscia, Department of Agriculture and Forest Sciences, Viterbo, Italy
[e] International Union of Soil Science, National Research Council of Italy, Institute of BioEconomy, Sesto Fiorentino, Italy

ARTICLE INFO

Keywords:
Cerium anomaly
Light and heavy REEs ratio
Compositional data approach
Geostatistics
Stochastic simulation
Uncertainty

ABSTRACT

Rare earth elements (REEs) are becoming increasingly interesting as indicators of soil processes, and modelling their spatial distribution has become a fundamental requirement for this purpose. This study aims to model and quantify the spatial distribution of soil REEs taking into account their compositional nature in their mapping, and to assess the associated spatial uncertainty. In particular, the cerium anomaly (Ce*) and the ratio of the sum (Σ) of light to heavy REEs (ΣLREEs/ΣHREEs) were calculated and analysed. The study was carried out in Sicily (Italy), one of the largest (25,832 km^2) and most populous island in the Mediterranean Sea. Soil REE data were obtained from the soil database of Sicily region. Soil REEs were transformed into a vector of isometric log-ratio (ilr) coordinates and analysed using a geostatistical approach for predicting their values at unsampled locations and generating 100 REEs realizations using turning bands simulation. Each REE element and Ce* were mapped as well as the ΣLREEs/ΣHREEs ratio. The joint variability of the investigated REEs and the ΣLREEs/ΣHREEs ratio provided insights into their abundance and distribution patterns. From the simulated realizations, standard deviation maps were calculated for the Ce* and the ΣLREEs/ΣHREEs ratio, providing a measure of their spatial uncertainty. Although with some limitations, the study established a first baseline of the Ce* and the ΣLREEs/ΣHREEs ratio. The maps of spatial uncertainty may be a useful tool for planning future soil sampling and optimise the choice of new sampling locations.

1. Introduction

There is an increasing interest in soil rare earth elements (REEs) that can be attributed to several factors, including environmental and agricultural implications (Hu et al., 2004; Ramos et al., 2016; Tyler, 2004). This means there is an urgent need to better understand the spatial distribution and abundance of REEs.

Soil REEs are a group of chemical elements that have the peculiarity of always being found together and can be locally differentiated by their mobilisation and redistribution. They include lanthanides, yttrium (Y), and scandium (Sc) according to the International Union of Pure and Applied Chemistry (IUPAC) (Aide and Aide, 2012; Saiano et al., 2019; Weng et al., 2013). REEs can be classified into two groups: light rare earth elements (LREEs) from ^{57}La to ^{64}Gd, and heavy rare earth elements (HREEs) from ^{65}Tb to ^{71}Lu, including yttrium (^{39}Y). Total REEs concentrations in soils generally range from tens to tens of thousands of μg kg^{-1}, with cerium (Ce) being the most abundant, followed by lanthanum (La) and LREEs are typically more abundant than HREEs (Jiménez-Ballesta et al., 2024). The REEs generally exhibit a trivalent oxidation state, although europium (Eu) and cerium (Ce) can also be found in divalent or tetravalent oxidation states, respectively (Adeel et al., 2019; Dushyantha et al., 2020; Leybourne et al., 2000; Reimann et al., 2014). Despite these small variations, as all 16 REEs generally exhibit homogeneous geochemical behaviour, REEs are useful in soil science and geochemistry as tracers to study pedogenetic transformations. The assessment of variation and their relative distributions in REEs abundance (i.e. fractionation, anomalies) involves normalization to a reference, which is done for different geochemical references such as chondrite, Upper Continental Crust (UCC), Post Archean Australian Shales (PAAS), etc. (Wedepohl, 1995), and the successive plot of ratios of elements on a logarithmic scale against their atomic numbers, called the 'REEs pattern' (Braun et al., 1993; Saiano and Scalenghe, 2009).

* Corresponding author.
E-mail address: gabriele.buttafuoco@cnr.it (G. Buttafuoco).

https://doi.org/10.1016/j.geodrs.2024.e00845
Received 5 April 2024; Received in revised form 26 July 2024; Accepted 27 July 2024
Available online 30 July 2024

Fractionation, in which one or more REEs are depleted or enriched relative to the other REEs, can also be evidenced by ratios such as LREEs/HREEs, or by calculating the anomalies of elements such as Ce or Eu, due to their two oxidation states (Barrat et al., 2023; Lawrence et al., 2006; Lawrence and Kamber, 2006). The occurrence of REEs and their fractionation, as with other elements, depends on weathering processes driven by hydrogeological, geochemical, and biological factors, including human activities (Saiano and Scalenghe, 2019), as well as pH changes, where increased pH leads to stronger binding sites and reduced ionic strength (Sonke, 2006) or to the presence of carbonates that can induce the reduction of LREEs relative to HREEs (Kasper-Zubillaga et al., 2017). However, unless there are abrupt changes in pH, the dynamics of REEs during weathering allow the pattern of REEs in soils to reflect the geochemistry of the parent materials, allowing subtle variations to be mapped (Ren et al., 2015).

In general, the analysis of REE data is based on classical statistics, which does not take into account the spatial position of the data and, therefore, does not allow modelling of their spatial distribution. In the case of REEs, the interest is often in their spatial distribution and the relationship of their concentrations to geomorphological and environmental attributes that allow some insight into their origin and source as well as to make some form of inference. Indeed, the parent material is a crucial factor in pedogenesis, especially for understanding the spatial distribution of REEs in soils (Bačeva Andonovska et al., 2023). For example, the impact of a Cretaceous volcanic-sedimentary lithosequence from Northeastern Brazil has been analysed (Brilhante et al., 2024) and it has been shown that, despite significant chemical weathering, soils derived from these lithologies retain distinct REE geochemical signatures. It has also been shown that soil texture plays a key role with clays acting as sinks for REEs during the pedogenic processes, where the capacity of sorption sites for REEs increases with decreasing soil particle size (Brilhante et al., 2024; Han et al., 2022; Jo et al., 2023). Several studies focused on human and environmental exposure to REEs, mapping them in topsoil, and assessing human health risks (da Pereira et al., 2023; Shomar et al., 2023). However, due to the complexities of considering thirteen elements together, most studies did not map REEs (Ion and Cosac, 2023).

Geostatistical methods (Matheron, 1973) are well-established and widely applied in a variety of fields, including mining, geochemistry, agriculture, forestry, climatology and meteorology, image analysis, and environmental remediation, etc. (Chilès and Delfiner, 2012). These methods also include the stochastic simulation, which allows to assess the uncertainty associated with maps of soil properties (Goovaerts, 1997; McBratney et al., 2003). In addition, the geostatistical approach can be easily extended to the multivariate case (Wackernagel, 2003) and then considering all REEs together enhancing the information contained in each element and filtering out redundant information.

Another important issue that has rarely been addressed is that REEs are parts of a whole, and each of its element cannot vary independently from the others. The REEs constitute a composition and requires the use of appropriate methods (Pawlowsky-Glahn et al., 2015a; Pawlowsky-Glahn and Egozcue, 2016a). Compositional data analysis is based on Aitchison's idea (Aitchison, 1986) and subsequent developments (Pawlowsky-Glahn et al., 2015b; Tolosana-Delgado et al., 2011), which considers the sum of concentrations of all chemical elements in a soil sample as a constant. This constraint leads to conditioned relationships among variables, with at least one covariance and correlation coefficient being negative, thereby inducing spurious correlations. Standard statistical and geostatistical techniques for unconstrained random variables cannot be directly applied to the analysis of compositional data (Tolosana-Delgado and Mueller, 2021).

Sicily is an Italian region and the largest island in the Mediterranean Sea where, despite the availability of numerous remote sensing tools and increased computational capacity, there is a lack of information on the spatial distribution of REEs. In Sicily, only very few local studies on REEs are available. Therefore, modelling soil REEs and their uncertainty at the regional scale will provide a relevant reference for studies focused on investigating the relationships between rocks, plants, animals, and humans (Aceto et al., 2017; Censi et al., 2013; Drivelos et al., 2016; Fantappiè et al., 2015, 2016; Pastor et al., 2016; Pisciotta et al., 2017). The aim of this study is to model and quantify the spatial distribution of soil REEs at regional scale. In particular, the cerium anomaly (Ce*) and the ratio of the sum (Σ) of light to heavy REEs (ΣLREEs/ΣHREEs) were calculated and analysed. Moreover, the compositional nature of REEs was taken into account and the spatial uncertainty associated with the mapping of soil REEs was assessed.

The results of this study might be the basis for planning new soil REEs monitoring, mapping, and soil sampling in areas having low sampling density and characterized by greater uncertainty.

2. Materials and methods

2.1. Soil data and laboratory analysis

The soil samples used in this study (Fig. 1) are part of a more comprehensive soil survey of Sicily carried out by CREA-AA (Centre of Agriculture and Environment), which included soil profiles, soil augers, and mini pits of the Italian 1:1,000,000 scale soil databases. Their details and full datasets can be found at https://zenodo.org/records/7085005 (Italian Soil Information System) and https://zenodo.org/records/7081223 (The soil province feature of Italy at the 1:1,000,000 scale).

Among the available soil samples, 199 (Fig. 1) were collected to be submitted to laboratory analysis. The laboratory analyses refer to concentrated nitric acid (65%), hydrochloric acid (37%) hydrofluoric acid (40%) and boric acid (30%) of ultrapure grade were from Baker (Milano, Italy). Ultrapure water 18.2 MΩ cm, produced with an EASYpureII (Thermo, Italy), was used for all standard solution and sample preparations. Working standard solutions for each element (Y, lanthanides and rhenium, Re) were prepared through successive dilution of BDH, Merck or CPI International, 1000 ± 5 μg mL^{-1} elemental standard solution in a HNO_3 1 mol L^{-1} medium. Laboratory equipment was in polyethylene, polypropylene or Teflon. The soil samples were dried in an oven at 105 °C overnight, sieved (0.5 mm) and homogenized. Aliquots of 200 mg (dry weight) were digested using HNO_3/HCL/HF 3/2/1 *v*/v (EPA Method 3052) in Teflon® liners in a microwave system (Mars Xpress, CEM, Milano, Italy). Successively, 500 mg of H_3BO_3 was added to each sample, to complex the HF excess, and diluted to 10 mL. Each analytical sequence included a blank (ultrapure water digested as the other samples). An Agilent 7500ce ICP-MS instrument was used to analyse all the investigated trace elements. All parameters were optimized for the studied elements. Each solution was measured three times and ICP-MS analyses were carried out with a classical external calibration approach. It was carried out from 2.5 to 10,000 pg mL^{-1} for each investigated element using ^{187}Re (1000 pg mL^{-1}) as the internal standard to compensate for any signal instability, and all data in cps were normalized to the internal standard. To quantify the accuracy and precision of the laboratory analyses a validation of the whole procedure (sample preparation and quantitative measurement) was carried out using certified soil reference materials: NCS DC 77301 and NCS DC 77303 (China National Analysis Center for Iron and Steel, Beijing, China). Experimental data, evaluated by reproducible assessed analyses, were in the range reported for all elements.

2.2. Ratios and anomalies

The REEs concentration data (available at https://zenodo.org/records/10370549) were compared with those of the GEMAS project (Reimann et al., 2018) and found to be consistent in terms of values. In addition, as is common and more appropriate in geochemistry studies, data were normalized to the Upper Continental Crust (UCC). Normalized UCC data can provide more immediate and unique information on soil biogeochemical processes, such as the determination of the ratio

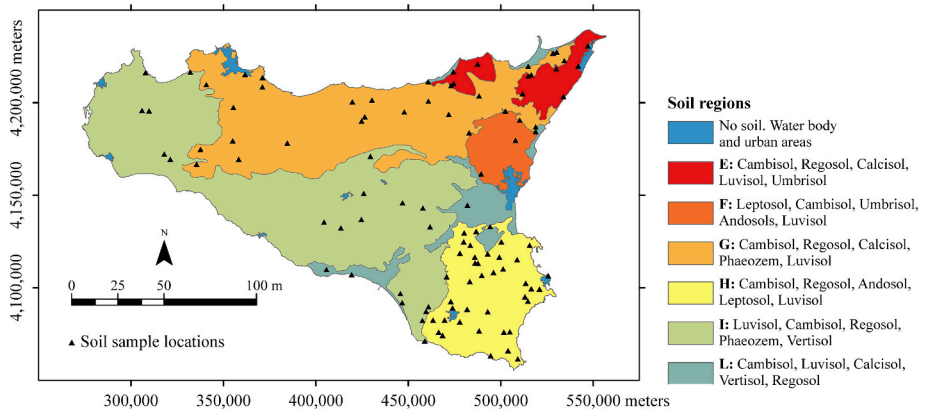


Fig. 1. Study area and soil sampling locations (filled black triangles). Soil regions of Sicily, according to the 'Soil map of Italy' [from Costantini et al., 2013, modified] are also reported. E: soils of Apennine of Calabria and Sicily on igneous and metamorphic rocks; F: soils of Etna volcano; G: soils of the hills of Calabria and Sicily on Tertiary calcareous rocks and sediments, with included alluvial and coastal plains; H: soils of the hills and mountains on limestone and igneous rocks of Sicily; I: soils of the hills of Calabria and Sicily on Tertiary clayey flysch, limestone, sandstone, gypsum and coastal plains; L: soils of the alluvial and coastal plains of central and southern Italy.

between the sum of light REEs and the sum of heavy REEs (ΣLREEs/ΣHREEs) or the REEs anomalies (Laveuf and Cornu, 2009). In literature, different approaches have been used to calculate Eu/Eu*, Ce/Ce*, La/La* based on the comparison with measured and calculated abundances derived from neighbouring REE concentrations normalized to reference values. In this study, the Ce anomaly was calculated as (Lawrence et al., 2006):

$$Ce_n^* = Pr_n \times \left(\frac{Pr_n}{Nd_n}\right) \quad (1)$$

where Ce_n^* is the cerium anomaly, Pr_n is the normalized praseodymium (Pr) and Nd_n, the normalized neodymium (Nd) (Lawrence et al., 2006).

2.3. Statistical and geostatistical analysis

A principal component analysis (Jackson, 2003) was performed as a preliminary step to check the REEs associations inherent in the structure of the correlation matrix into some groups of elements which together account for the greater part of the observed variability in the original data (Jimenez-Espinosa et al., 1993). Principal component analysis is a method of analysis for multivariate data which is widely used because of the simplicity of its algebra and straightforward interpretation (Wackernagel, 2003). It consists of transforming a set of linearly correlated variables $Z_i (i = 1, \ldots, N)$ into uncorrelated principal components (PCs) $Y_i (p = 1, \ldots, N)$ which partition optimally the total variance and ordering the PCs by decreasing explained variance. Principal components (PCs) are nothing more than the eigenvectors of a variance-covariance matrix or a correlation matrix (Davis, 2002), and each PC extracts a maximal share of the total variance. The correlation c_{ij} (Eq. (2)) between the original variables and a pair of principal components can be shown in a graphic display called the circle of correlations. The correlation c_{ij} can be computed as:

$$c_{ij} = a_{ij}\sqrt{(\lambda_j/\sigma_i^2)} \quad (2)$$

where a_{ij} is the i^{th} element of the j^{th} eigenvector, λ_j is the j^{th} eigenvalue, and σ_i^2 is the variance of the i^{th} original variable. The coordinates of the variables are obtained from the correlation values of the geochemical elements with the first (ordinate axis) and the second (coordinate axis) principal components.

Within the geostatistical approach, the set of REEs concentrations is distributed in space and can be regarded as a particular realization of random processes (Matheron, 1971). The concentration of any REEs i, ($i = 1, \ldots..., N$, is the number of REEs) measured at any location $z_i(\mathbf{x}_\alpha)$ at any place $\mathbf{x}_\alpha$ ($\mathbf{x}$ denotes the coordinates vector in two dimensions and $\alpha = 1, \ldots, n$ the sampling points), is just one of the infinite outcomes of a random variable $Z_i(\mathbf{x}_\alpha)$ that are possible there. The set of concentrations of each REEs for all points $\mathbf{x}_\alpha$ within the study area is the regionalized variable $z_i(\mathbf{x}_\alpha)$ and can be considered as a particular realization of a set of random variables $\{Z_i(\mathbf{x}_\alpha),\ \alpha = 1, \ldots, n;\ \ i = 1, \ldots,\ N\}$, which is called a random function $Z_i(\mathbf{x})$. Interested readers are referred to Journel and Huijbregts (1978), Goovaerts (1997), Wackernagel (2003), Webster and Oliver (2007), and Chilès and Delfiner (2012), among many others.

The quantitative measure of spatial autocorrelation of each regionalized variable, $z_i(\mathbf{x}_\alpha)$, is the experimental variogram $\gamma_i(\mathbf{h})$(Chilès and Delfiner, 2012). It is a function of the distance vector ($\mathbf{h}$) of data pairs values $[z_i(\mathbf{x}_\alpha), z_i(\mathbf{x}_\alpha + \mathbf{h})]$ and to it, a theoretical function (variogram model) is fitted to allow one to estimate the variogram analytically for any distance $\mathbf{h}$. The variogram model requires two parameters: range and sill. The first is the distance over which pairs of REEs concentrations are spatially correlated, whereas the sill is the variogram value corresponding to the range (Chilès and Delfiner, 2012). Generally, the optimal fitting is chosen based on cross-validation (Webster and Oliver, 2007) and its goodness is evaluated by the Mean Error (ME) and the ratio between the squared errors and the kriging variance (Mean Squared Deviation Ratio, MSDR). The ME proves the unbiasedness of the estimate requires an ME value close to 0, whereas the variogram accuracy of an MSDR is close to 1 (Webster and Oliver, 2007).

The multivariate variation of this study can be represented by a linear model of coregionalization (LMC) (Goovaerts, 1997), which consists of a variogram for each variable (here each REE) and a cross-variogram for each pair of variables. The parameters of these variograms and cross-variograms must be constrained to ensure that no negative prediction variances result. The LMC is then used in one of the kriging types for the later prediction of variables at the nodes of an interpolation grid for mapping. However, when the structure of spatial variability is more important than local estimation accuracy, geostatistical stochastic simulation in comparison with an optimal procedure such as any type of kriging, provides a more realistic means of evaluating the spatial variability of a variable (Castrignanò and Buttafuoco, 2004; Chilès and Delfiner, 2012). Stochastic simulation consists of generating (simulating) a large number of equiprobable images (also called realizations) that honour the sample data and reproduce statistical characteristics and spatial features. In this study, the turning bands simulation (TBS) method was used because it represents a reasonable trade-off between prediction quality and computing time (Lantuéjoul, 2002; Matheron, 1973). The number of realizations was set at 100

because a satisfactory level of accuracy is achieved only when the number of runs is sufficiently high.

The turning bands method is a Gaussian simulation technique that requires a multi-Gaussian framework. Therefore, REEs concentration data were transformed into a normal distribution shaped variable with zero mean and unit variance, using a procedure known as Gaussian anamorphosis (Chilès and Delfiner, 2012; Wackernagel, 2003). It is a mathematical function that transforms a variable Y with a Gaussian distribution into a new variable with any distribution.

The realizations obtained from the stochastic simulation are consistent with the observed data and are equally probable, then it follows that pixel-by-pixel histograms, summarising a large number of simulations, approximate the probability distribution functions corresponding to the locations. Therefore, it is possible to present a map of the 'expected' value at any given location (E-type or Expected-value estimate) and the related standard deviation as a measure of uncertainty of the predictions (Journel, 1983).

However, soil REEs are only a portion of the soil constituents and REEs data have a restricted space in which they can vary only from 0 to total (100% or 1 or any constant values). These data are called compositional data, and their sample space is not the real space R^3, but the simplex (S^D) (Aitchison, 1986; Pawlowsky-Glahn et al., 2015b). The D-part simplex, S^D, is a D-dimensional real space of strictly positive vectors of D parts summing to a constant κ:

$$S^D = \left\{ \mathbf{u} = [u_1, u_2, \ldots, u_D] \,|u_j \geq 0 \text{ and } \sum_{j=1}^{D} u_j = \kappa \right\} \quad (3)$$

where the constant κ can be 1, 100, 10^6, or any other constant (Tolosana-Delgado et al., 2005).

Statistical methods designed for multivariate normal distributions can be used if the data in the original D-part simplex sample space are projected to multivariate real space (Aitchison, 1986; Pawlowsky-Glahn et al., 2015a).

Therefore, soil REEs are compositional data because each of them provides only relative information about a part of a whole. The application of any geostatistical approach to compositional data may result in spurious spatial correlation (Pawlowsky-Glahn and Egozcue, 2016b). This concept of composition also applies to the spatially distributed vectors of random variables $\{Z_i(\mathbf{x}_\alpha); i = 1, \ldots, N\}$ at any point $\mathbf{x}_\alpha$ ($\alpha = 1, \ldots, n$) of a spatial domain R^2. This type of composition is called regionalized composition (r-composition) (Pawlowsky-Glahn and Egozcue, 2016b; Tolosana-Delgado et al., 2018) and it is both a composition and a coregionalization (Pawlowsky-Glahn and Olea, 2004). Further practical issues arise in the context of geostatistical spatial prediction (Chilès and Delfiner, 2012; Tolosana-Delgado et al., 2018) and to deal with these problems, a log-ratio transformation has been proposed (Aitchison, 1986). The transformation lets a composition be a real vector and allows applying the geostatistical methods to these log-ratios. After the geostatistical applications, the predicted log-ratio is back-transformed into the raw data (Tolosana-Delgado et al., 2011). Different log-ratio transformations to deal with compositional data are available (Pawlowsky-Glahn and Egozcue, 2020), and among them, the isometric log-ratio (ilr) transformation (Egozcue et al., 2003) has been selected because allows an effective and practical treatment of compositional data in geostatistical applications (Tolosana-Delgado et al., 2019; Tolosana-Delgado and Mueller, 2021).

Coordinates or coefficients are different ways to term the log-ratio transformations because of their more natural terminology (Pawlowsky-Glahn and Egozcue, 2006).

Since only the REEs were considered in this study, a complementary component (filler, which will be denoted as F) was calculated to ensure the composition sums to a constant at any given location.

Therefore, the REEs were transformed in isometric log-ratio coordinates and simulated using the fitted LMC at the nodes of a 250 m × 250 m grid (Chilès and Delfiner, 2012). The 100 realizations of the Gaussian ilr coordinates obtained by Turning Bands Simulation (TBS) were first back-transformed into the raw distribution using the inverse of the Gaussian anamorphosis and then, back-transformed into the raw REEs for subsequent analyses and for calculating the ratio between HREEs and LREEs.

All geostatistical analyses were performed using the software package Isatis.neo, release 2023.04 (Bleines et al., 2023), whereas the compositional data procedures were performed by using the software package CoDaPack 2.0 (Comas-Cufi and Thió-Henestrosa, 2011).

3. Results and discussion

Traditional soil maps are produced using specific criteria to extrapolate observations made on a three-dimensional space called a 'pedon' (ranging from 1 to a few cubic meters) to a much larger volume (tens of thousands of cubic meters) that encompasses a particular soil type. Only a tiny proportion of soils are observed, sampled, and analysed. This method is effective and widely accepted for soil mapping (Arnold, 2006). However, if these soil samples are later utilized to develop other maps, the sampling criteria originally employed may not be suitable for the new objective. This can lead to an overabundance of observations in certain areas and a scarcity in others. As scientific and technological advances continue, new objectives arise to explore aspects that were previously unimaginable and are now being investigated. These new objectives often necessitate new surveys, which, unfortunately, are typically the limiting factor due to their high costs. Moreover, the ability to chronologically align new observations with previously obtained ones provides additional value by enhancing data integration, known as complementarity. Creating soil maps based on taxonomic terms is fundamentally distinct from generating maps that describe the concentrations of chemical elements found in soils. In the context of rare earths, this study aimed to determine whether a priori knowledge of soil types can facilitate the inference of their spatial variability. In this context, the REEs normalized data, in the form of LREEs/HREEs and Ce*, have been used to describe their distribution in the soils of Sicily and to provide a tool for their mapping. The results of the explanatory data analysis for normalized REEs are summarized in Table 1, where the main statistics are reported. No REEs showed multimodal histograms (not shown), therefore all data of each element were considered to belong to the same population. Apart from La, Ce, and Pr, the different REEs are all characterized by the great asymmetry of their distributions. In fact, the coefficient of skewness varies from a minimum of 0.71 for the Nd to a maximum of 3.11 for the Y (Table 1). However, Fig. 2 shows that this asymmetry mainly concerns the data in the upper and lower quartiles. Despite the high asymmetry values, the mean and median values are very close for many items. (See Fig. 2.)

The results of PCA are summarized in Table 2, where the loadings for each element of the first three principal components as well as the eigenvalues, and the percentages of explained variance for each of the components are reported.

The first component accounts for more than 90% of the total variance, the second one for an additional 5.66%, and the third component for 2.33%. However, no element has a large contribution to the principal components, but it is quite evenly distributed with coefficients ranging from 0.221 (Eu) to 0.270 (Tb) (Table 2). All REEs have positive correlation with the first principal component (PC1) (Table 2). For PC2 only La, Ce, Pr, Nd, Sm, Eu, and Gd have a positive correlation, whereas a negative correlation is shown for Tb, Dy, Y, Ho, Er, Tm, Yb, Lu (Table 2). These correlations are most evident when the first two associate components were converted into correlations between the original variables and the principal components using Eq. (1). The results are then plotted in the unit circle in the plane of the first two principal components (Fig. 3). The circle of correlations shows the proximity of the elements inside the unit circle, allowing for an assessment of their associations. Two groups of elements can be defined, separating the elements belonging to LREEs from those belonging to HREEs (Fig. 3).

Table 1
Basic statistics of normalized soil rare earth elements.

Element	Minimum (–)	Maximum (–)	Mean (–)	Median (–)	Lower quartile (–)	Upper quartile (–)	Stand. Deviation (–)	Skewness (–)	Kurtosis (–)
La	0.132	1.811	0.752	0.726	0.549	0.947	0.300	0.62	3.52
Ce	0.131	1.349	0.682	0.669	0.464	0.844	0.272	0.48	2.68
Pr	0.175	1.675	0.727	0.707	0.523	0.902	0.284	0.54	3.18
Nd	0.172	1.958	0.771	0.760	0.567	0.950	0.306	0.71	4.12
Sm	0.248	2.237	0.861	0.841	0.607	1.084	0.345	0.78	4.40
Eu	0.252	3.098	1.131	1.076	0.732	1.381	0.522	1.22	5.65
Gd	0.275	3.129	1.037	1.029	0.737	1.257	0.436	1.25	6.97
Tb	0.217	2.671	0.811	0.810	0.544	1.009	0.360	1.52	8.43
Dy	0.199	2.867	0.806	0.773	0.584	0.952	0.366	1.88	10.93
Y	0.117	3.201	0.655	0.604	0.457	0.755	0.382	3.11	19.69
Ho	0.158	2.769	0.683	0.669	0.453	0.811	0.340	2.34	14.43
Er	0.147	2.856	0.674	0.668	0.473	0.802	0.340	2.61	16.78
Tm	0.139	2.617	0.606	0.602	0.399	0.730	0.319	2.48	15.68
Yb	0.148	2.526	0.630	0.602	0.431	0.756	0.310	2.33	14.38
Lu	0.142	2.600	0.617	0.609	0.390	0.771	0.321	2.33	14.52

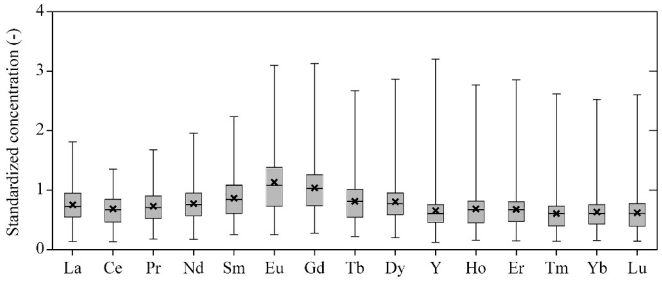


Fig. 2. Box plots of the rare earth elements in soil.

Table 2
First three eigenvectors from the PCA of the standardized elements data.

Element	PC 1 (–)	PC 2 (–)	PC 3 (–)
La	0.255	0.292	0.231
Ce	0.248	0.323	0.445
Pr	0.261	0.244	0.225
Nd	0.264	0.241	0.085
Sm	0.267	0.167	0.010
Eu	0.221	0.451	−0.617
Gd	0.268	0.101	−0.161
Tb	0.270	−0.030	−0.091
Dy	0.268	−0.138	−0.022
Y	0.246	−0.206	−0.472
Ho	0.266	−0.200	−0.066
Er	0.263	−0.253	−0.004
Tm	0.260	−0.287	0.053
Yb	0.258	−0.311	0.140
Lu	0.255	−0.326	0.161
Eigenvalue	**13.576**	0.85	0.35
Ratio %	**90.50**	5.66	2.33
Cumulative %	90.50	96.16	98.49

The REEs and the filler variable (F) were transformed into isometric 15 log-ratios (ilr) and used for the following approach steps. Moreover, since the turning bands simulation requires a Gaussian framework, all ilr coordinates were transformed to normality and standardized to zero mean and unit variance using the Gaussian anamorphosis and such transformed data were used for all subsequent analyses.

The linear coregionalization of the 15 ilr coordinates was modelled calculating 120 (for $N = 15$; $15\times(15 + 1)/2$) direct (or auto) and cross variograms. This included a direct variogram for each REE and a cross variogram for each pair of variables. The variation in the REEs was isotropic because no anisotropy was evident from the directional variograms (not shown) and a nested isotropic LMC (not shown) was fitted to the experimental variograms. Therefore, the variation of REEs data

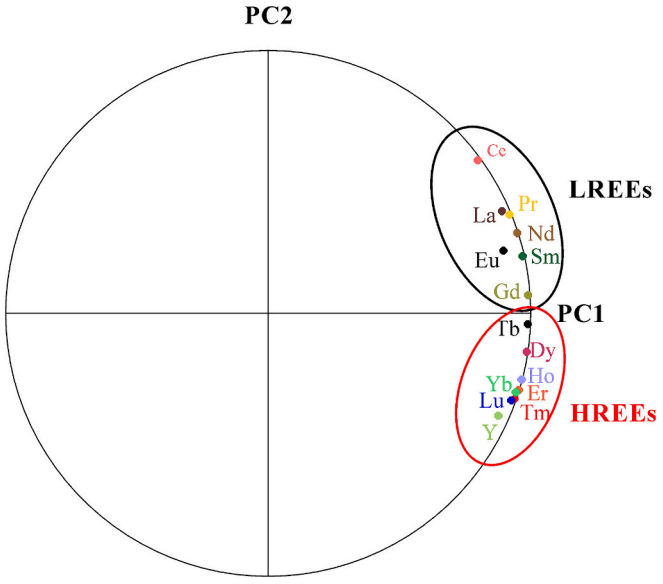


Fig. 3. Correlation circle of REYs concentrations. Light (L) and heavy (H) REEs are delineated by ellipses.

only depends on the module h of the vector distance. The LMC included two basic structures (Table 3): a nugget effect, and an exponential model (Webster and Oliver, 2007) with a practical range equal to 23,864.49 m.

The exponential model (Webster and Oliver, 2007) is:

$$\gamma(h) = c\left\{1 - exp\left(-\frac{h}{r}\right)\right. \quad (4)$$

where c is the *sill* (a constant or asymptote variogram value at a finite distance called *range* (a), h is the distance and r is a distance parameter defining the spatial extent of the model. The exponential variogram model has no finite range (a), and usually the practical range is reported. It is the distance at which the variogram value equals 95% of the sill variance, which approximately amounts to $3r$ (Webster and Oliver, 2007).

The results of cross-validation showed good reproducibility of REEs data with mean errors (ME) very close to 0 and the mean squared deviation ratios (MSDR) to 1.

Table 3 reports the decomposition of the coregionalization matrix by principal component analysis in the regionalized factors (F) for the Gaussian values of the 15 isometric log-ratio (ilr) coordinates for the nugget effect and the exponential model. For each regionalized factor (F), the loadings for each element are reported as well as the eigenvalues

Table 3
Decomposition of the coregionalization matrix in the regionalized factors (F) for the Gaussian values of isometric log-ratio (ilr) coordinates. EV stands for eigenvalue and Var. % for percentage of explained variance.

Nugget effect																	
	ilr.1	ilr.2	ilr.3	ilr.4	ilr.5	ilr.6	ilr.7	ilr.8	ilr.9	ilr.10	ilr.11	ilr.12	ilr.13	ilr.14	ilr.15	EV	Var. %
F 1	0.272	0.287	0.285	0.275	0.276	0.289	0.256	0.294	0.185	0.128	0.225	0.239	0.263	0.271	0.273	3.792	95.83
F 2	0.462	0.204	0.303	0.274	0.076	−0.603	−0.064	−0.297	−0.100	−0.057	−0.195	−0.165	−0.007	−0.093	0.172	0.129	3.26
F 3	0.299	−0.006	0.076	0.016	−0.423	0.419	−0.602	−0.291	−0.087	−0.070	0.047	0.161	0.047	0.151	0.191	0.026	0.67
F 4	−0.438	0.372	−0.193	0.303	0.240	0.071	−0.097	−0.466	−0.273	0.161	0.126	0.094	0.328	−0.135	−0.042	0.005	0.13
F 5	0.030	−0.412	0.295	0.059	0.355	0.025	−0.206	−0.133	0.017	−0.083	0.117	−0.253	0.282	0.371	−0.501	0.003	0.07
F 6	0.005	0.013	−0.211	0.325	−0.075	−0.108	−0.334	0.419	−0.061	0.081	0.512	−0.494	−0.042	−0.126	0.107	0.002	0.05
F 7	−0.071	−0.058	0.509	−0.397	0.049	0.020	0.129	−0.177	−0.280	0.037	0.567	0.012	−0.182	−0.268	0.136	0.000	0
F 8	−0.457	0.073	0.470	0.334	−0.283	0.108	0.016	−0.104	0.507	−0.124	−0.088	−0.198	−0.148	−0.096	0.017	0	0
F 9	−0.307	−0.264	−0.008	0.137	0.064	−0.454	−0.241	0.036	0.070	0.057	0.176	0.527	−0.221	0.359	0.232	0	0
F 10	−0.265	0.007	0.227	−0.343	0.185	−0.069	−0.370	0.322	−0.043	−0.037	−0.340	−0.086	0.441	−0.149	0.373	0	0
F 11	0.160	0.124	−0.074	−0.239	0.313	0.009	−0.317	−0.144	0.553	0.523	0.033	0.021	−0.196	−0.213	−0.111	0	0
F 12	−0.036	0.071	0.311	0.242	0.064	0.158	−0.174	0.326	−0.445	0.332	−0.310	0.154	−0.373	−0.092	−0.311	0	0
F 13	0.118	0.044	−0.019	0.122	0.210	0.010	−0.191	0.152	0.137	−0.611	0.127	0.396	0.007	−0.485	−0.257	0	0
F 14	−0.065	0.105	−0.118	−0.019	0.508	0.264	−0.077	−0.119	−0.073	−0.356	−0.117	−0.260	−0.511	0.224	0.313	0	0
F 15	−0.090	0.676	0.066	−0.336	−0.157	−0.206	−0.140	0.137	0.010	−0.175	0.123	−0.022	−0.054	0.392	−0.333	0	0
Exponential model – Practical range = 23,864.49 m																	
	ilr.1	ilr.2	ilr.3	ilr.4	ilr.5	ilr.6	ilr.7	ilr.8	ilr.9	ilr.10	ilr.11	ilr.12	ilr.13	ilr.14	ilr.15	EV	Var. %
F 1	0.252	0.240	0.261	0.269	0.270	0.238	0.276	0.251	0.295	0.274	0.267	0.258	0.241	0.241	0.232	9.399	91.11
F 2	0.341	0.402	0.279	0.256	0.153	0.226	0.083	−0.033	−0.047	−0.270	−0.170	−0.248	−0.312	−0.318	−0.365	0.728	7.05
F 3	0.292	−0.066	0.013	−0.045	0.037	−0.231	0.072	−0.054	−0.263	0.826	−0.150	−0.103	−0.116	−0.068	−0.207	0.089	0.86
F 4	−0.022	0.613	0.155	0.030	−0.017	−0.629	−0.230	−0.215	−0.011	−0.059	−0.073	0.065	0.026	0.258	0.153	0.056	0.55
F 5	0.099	−0.515	0.250	0.285	0.296	−0.193	−0.054	−0.096	−0.009	−0.172	−0.143	0.251	−0.400	0.409	−0.073	0.018	0.18
F 6	−0.679	0.055	0.111	0.151	0.478	0.085	0.075	−0.065	0.071	0.204	−0.283	−0.320	−0.019	−0.058	0.148	0.015	0.15
F 7	−0.095	−0.177	0.393	0.172	0.096	−0.240	−0.465	0.339	−0.108	0.020	0.294	0.079	0.189	−0.486	−0.037	0.006	0.05
F 8	0.218	0.052	−0.238	−0.532	0.685	−0.029	−0.141	−0.094	0.057	−0.074	0.162	0.126	−0.132	−0.179	0.107	0.003	0.03
F 9	−0.179	−0.017	0.430	−0.241	−0.053	0.063	0.276	−0.368	−0.311	−0.010	0.602	−0.144	−0.124	0.105	−0.032	0.002	0.02
F 10	0.038	−0.102	0.369	−0.357	0.082	0.058	−0.037	−0.168	0.171	−0.058	−0.315	0.039	0.578	0.142	−0.444	0.001	0
F 11	−0.034	−0.039	0.058	−0.156	−0.143	−0.117	−0.168	0.128	0.729	0.138	0.236	−0.356	−0.318	0.113	−0.212	0	0
F 12	−0.092	0.035	0.108	0.075	−0.193	0.257	−0.208	−0.559	0.285	0.214	−0.067	0.491	−0.166	−0.324	0.095	0	0
F 13	0.179	−0.204	0.284	−0.177	−0.121	−0.324	0.471	0.006	0.210	−0.124	−0.246	−0.093	−0.016	−0.361	0.463	0	0
F 14	0.150	0.023	0.328	−0.284	−0.134	0.385	−0.441	0.175	−0.185	0.071	−0.235	−0.189	−0.164	0.227	0.434	0	0
F 15	−0.330	0.208	0.141	−0.340	−0.093	−0.017	0.223	0.477	−0.073	0.029	−0.146	0.485	−0.334	−0.013	−0.228	0	0

(EV) and the percentage of explained variance (Var. %). The elements of the regionalized factors represent the contribution of the ilr coordinates to each regionalized factor. An element with a value close to 1 means that the ilr coordinate makes a large contribution to that regionalized factor. Conversely, if an element of an eigenvector is near 0 the contribution to that regionalized factor is small. Omitting the nugget effect, as it is mostly affected by experimental errors and spatial variation within the shortest sampling interval, the first factor corresponding to the exponential model can be analysed. Initially, the first regionalized factor produced an eigenvalue (EV in Table 3) greater than one (9.399) and accounted for more than 91% of the spatial variability at the scale of the exponential model having a practical range of 23,864.49 m. The first regionalized factor may be considered as a new variable that allow aggregate and summarise the joint variability in the 15 ilr coordinates. The elements of the first regionalized factor make a similar contribution to it because their values are similar and range between 0.232 and 0.276. This means that all ilr coordinates contribute similarly to the first regionalised variable without any of them having a dominant weight.

The Gaussian data of ilr coordinates and the LMC were used to generate 100 realizations of ilr coordinates using TBS at the nodes of a 250 m × 250 m grid. Then, the 100 realizations of ilr coordinates were transformed into their raw REEs values, and the different statistical values for each REEs were calculated: the estimated expected values (E-type estimate) and the standard deviations. As explained above, the 100 simulations are to be viewed as statistically indistinguishable alternative realizations of the underlying, but unknown, real world. The differences among the 100 simulated realizations of each REEs can provide a measurement of the spatial uncertainty.

Moreover, the 100 simulated realizations of Pr_n and Nd_n were used with Eq. (1) for the Ce_n^* anomaly calculation. Therefore, 100 realizations of Ce_n^* were obtained and the Ce_n^* expected values (Fig. 4) and standard deviations (Fig. 5) were calculated.

The map of the Ce_n^* expected values (Fig. 4) shows its spatial distribution whereas the map of standard deviation of the 100 realizations for the anomaly Ce_n^* provides a measure of spatial uncertainty and confidence in their expected values. The highest values of negative Ce* are concentrated in the central-western part of Sicily with some small areas in the central-eastern part (Fig. 4). According to the geo-morphological studies (Fantappiè et al., 2016), these areas mainly correspond Neogene marine deposits, characterized by clay, silty-clay and marls with evaporites (presence of abundant gypsum). The standard deviation map of the 100 realizations for the anomaly Ce_n^* (Fig. 5), which provides insights into the past redox conditions, shows a similar behaviour of the one of the Ce_n^* expected values (Fig. 4) although in a fuzzier pattern and less defined boundaries.

To model and map the spatial distribution of the LREEs/HREEs ratio, the 100 realizations of each REE belonging to light REEs were summed. The same process was applied to the REEs in the heavy REEs group. From these sums of LREEs and HREEs, the ratios for all 100 realizations were calculated. From these 100 LREEs/HREEs ratios, the estimated expected values (E-type estimate) (Fig. 6) and the standard deviations of realizations (Fig. 7) were also calculated. This approach, as previously explained, avoids propagating any errors in the modelling and simulation process of each element. The map of standard deviations provides a measure of spatial uncertainty and confidence in the expected values of the LREEs/HREEs ratio.

The map of the estimated expected values (E-type estimate) (Fig. 6) shows two well-defined areas of high and low values of the LREEs/HREEs ratio. In the north-eastern part of Sicily, an area of low ratio values is clearly visible, fading southwards with a few hot spots in the central-southern eastern part (Fig. 6). These areas mainly correspond to the presence of volcanic (Etna volcano, Iblei mounts) and metamorphic rocks (Peloritani mounts, north-east of Sicily), namely phyllites, micaschists, and gneiss.

A well-defined area of high values is found in the south-eastern part of Sicily (Fig. 6). This area is mainly characterized by red soils on marine biocalcarenites and limestones (Fantappiè et al., 2016). The central-western part of Sicily is characterized by high values, shifting towards values close to the median of the distribution of expected ratio values. This area is mainly characterized by marine terraces and calcareous rocks. A few hot spots with median values are also visible in this area (Fig. 6).

At the resolution of 250 k, interpreting soil REEs unambiguous, using solely taxonomic types, is challenging. However, a spatial correlation is evident between soils with the most similar characteristics, particularly regarding their parent material. The LREEs to HREEs ratio shows a remarkable relationship, especially as these more mobile elements transition from primary to sedimentary rocks. This influence on the LRREs/HREEs ratio is both generalized and notable. Consequently, using these findings as a baseline, consistent with the results of the GEMAS network, more detailed studies can derive significant insights by utilising soil information as a covariate.

The standard deviations of LREEs/HREEs ratio (Fig. 7) reproduce the same pattern as the expected value map of the ratio, with the difference being the occurrence of high and low values appearing next to each other.

In the light of the observations and discussions, it can be said that the main limitation of this study is that it is based on samples collected for other purposes and not for modelling REEs. The samples were originally

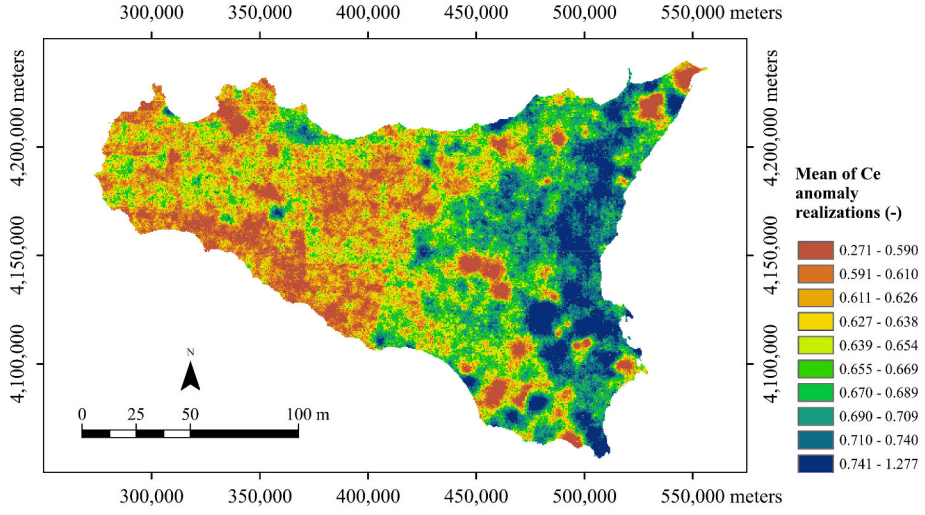


Fig. 4. Map of the means (expected values) of Ce anomaly (Ce*) realizations obtained by turning bands simulation (TBS).

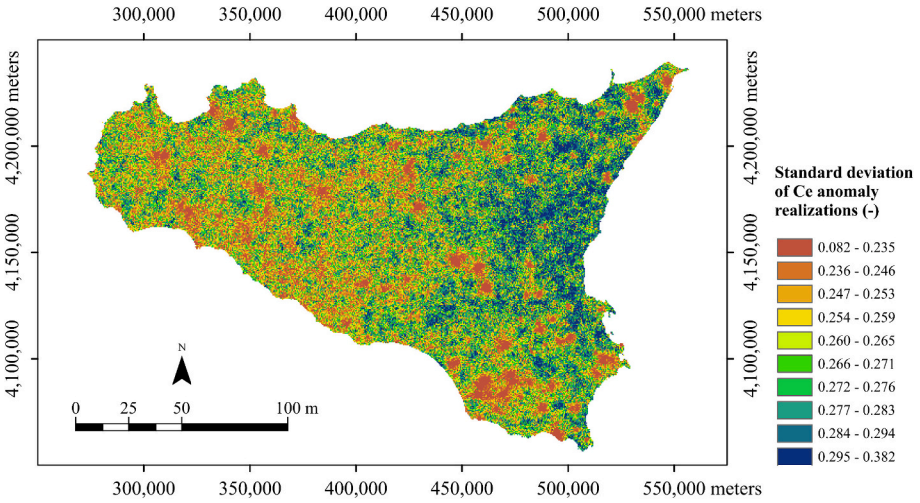


Fig. 5. Map of the standard deviations of Ce anomaly (Ce*) realizations obtained by turning bands simulation (TBS).

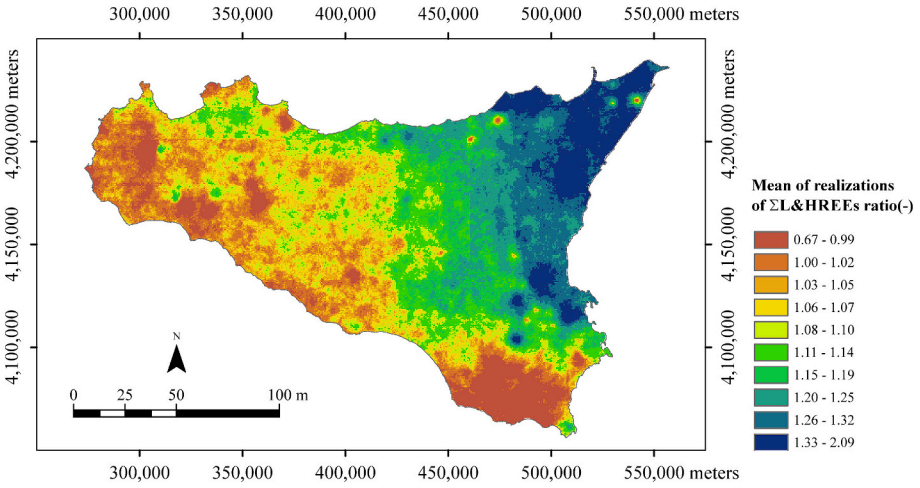


Fig. 6. Map of the means (expected values) of ΣLREEs/ΣHREEs ratio calculated from the REEs realizations obtained by turning bands simulation (TBS).

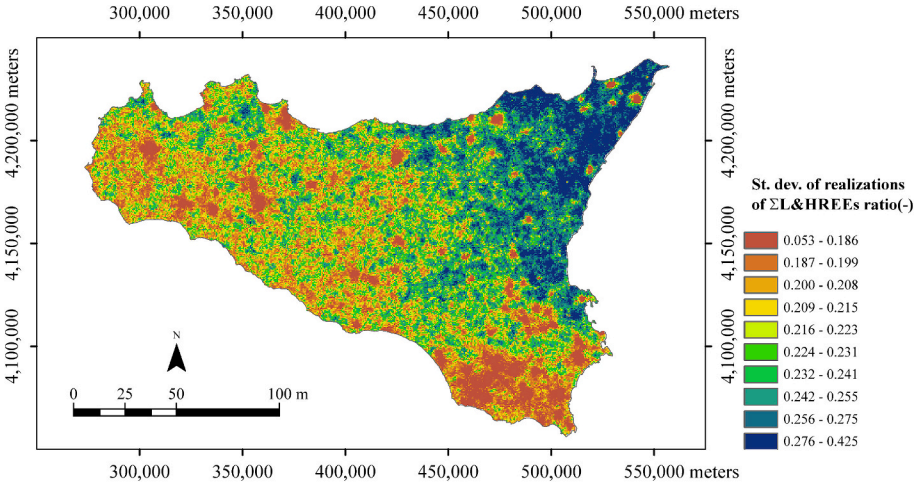


Fig. 7. Map of the standard deviations of ΣLREEs and ΣHREEs ratio calculated from the REEs realizations obtained by turning bands simulation (TBS).

taken to produce a soil map of the region of Sicily. The spatial distribution of the soil samples was based on observations and modelling of the spatial distribution of the different soil types. Therefore, the number of soil samples was small in areas considered homogeneous, because sampling is a very significant cost component in the production of soil maps. This weakness was turned into a strength by using a sub-optimal

dataset for modelling REEs in Sicily, to produce useful findings that establish a baseline of the occurrence of REEs and LREEs/HREEs ratios. In addition, complementing the REEs map with the spatial uncertainty map provides a useful tool for planning future soil sampling. Decisions on increasing sampling density will be made through optimization procedures based on budget availability.

4. Conclusions

Given the distinctive behaviour of REE in soils, establishing a background baseline becomes increasingly important when presenting geochemical patterns, whether characterized by the ratio of light and heavy elements (LREEs/HREEs) or specific element ratios, such as gadolinium to lanthanum. Additionally, anomalies in key elements such as cerium can be effectively represented. Therefore, it is valuable to establish a background baseline for the REEs content in the soils of a given region to link diverse studies to these maps, from the analysis of plant fingerprints to the study of soil weathering processes. This paper reports on the REE values of 199 soil samples analysed in Sicily, an Italian region of approximately 25,000 km^2. These values are fully consistent with the results of the previous European-scale survey called GEMAS. Uncertainty estimates have been provided alongside the maps, to complement the maps and to inform potential users about the quality of soil REEs predictions. However, the real scientific breakthrough of this study lies in the approach to processing georeferenced samples. Instead of relying on purpose-specific sampling, georeferenced soil samples from previous sampling campaigns were enhanced.

Declaration of competing 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.

Data availability

The data that support the findings of this study are openly available in Zenodo at doi:https://doi.org/10.5281/zenodo.10370548 under the Creative Commons Attribution 4.0 International (CC-BY) license.

Acknowledgment

The authors express their gratitude to the DISAFA department at the University of Turin for providing access to laboratories for the extraction of lanthanides from soils. The authors thank the two anonymous reviewers of this paper for providing constructive comments, which have contributed to the improvement of the published version.

References

Aceto, M., Musso, D., Calà, E., Arieri, F., Oddone, M., 2017. Role of lanthanides in the traceability of the Milk production chain. J. Agric. Food Chem. 65, 4200–4208. https://doi.org/10.1021/acs.jafc.7b00916.

Adeel, M., Lee, J.Y., Zain, M., Rizwan, M., Nawab, A., Ahmad, M.A., Shafiq, M., Yi, H., Jilani, G., Javed, R., Horton, R., Rui, Y., Tsang, D.C.W., Xing, B., 2019. Cryptic footprints of rare earth elements on natural resources and living organisms. Environ. Int. 127, 785–800. https://doi.org/10.1016/j.envint.2019.03.022.

Aide, M.T., Aide, C., 2012. Rare earth elements: their importance in understanding soil genesis. ISRN Soil Sci. 2012, 1–11. https://doi.org/10.5402/2012/783876.

Aitchison, J., 1986. The Statistical Analysis of Compositional Data, Journal of the Royal Statistical Society. Series B. Chapman and Hall, London - New York. https://doi.org/10.1007/978-94-009-4109-0.

Arnold, R.W., 2006. Concepts of soils. In: Soils: Basic Concepts and Future Challenges. Cambridge University Press, pp. 1–10. https://doi.org/10.1017/CBO9780511535802.002.

Bačeva Andonovska, K., Šajn, R., Tănăselia, C., Stafilov, T., 2023. Lithological distribution of rare earth elements in soils in the as-Sb-Tl Allchar mining area, North Macedonia. J. Environ. Sci. Health A 58, 139–151. https://doi.org/10.1080/10934529.2023.2178787.

Barrat, J.-A., Bayon, G., Lalonde, S., 2023. Calculation of cerium and lanthanum anomalies in geological and environmental samples. Chem. Geol. 615, 121202 https://doi.org/10.1016/j.chemgeo.2022.121202.

Bleines, C., Deraisme, J., Geffroy, F., Jeannée, N., Perseval, S., Rambert, F., 2023. Isatis.neo technical references (No. 2017.2). Geovariances, Avon Cedex.

Braun, J.-J., Pagel, M., Herbilln, A., Rosin, C., 1993. Mobilization and redistribution of REEs and thorium in a syenitic lateritic profile: a mass balance study. Geochim. Cosmochim. Acta 57, 4419–4434. https://doi.org/10.1016/0016-7037(93)90492-F.

Brilhante, S.A., da Silva, Ygor Jacques Agra Bezerra, de Medeiros, P.L., do Nascimento, C. W.A., da Silva, Yuri Jacques Agra Bezerra, Ferreira, T.O., Otero, X.L., Silva, A.H.N., Sousa, M.G., Alcantara, V.C., Araújo, J.K.S., Junior, V.S.S., 2024. Geochemistry of rare Earth elements in rocks and soils along a Cretaceous volcano-sedimentary Basin in Northeastern Brazil. Geoderma Reg. 36, e00756 https://doi.org/10.1016/j.geodrs.2024.e00756.

Castrignanò, A., Buttafuoco, G., 2004. Geostatistical stochastic simulation of soil water content in a forested area of South Italy. Biosyst. Eng. 87, 257–266. https://doi.org/10.1016/j.biosystemseng.2003.11.002.

Censi, P., Randazzo, L.A., D'Angelo, S., Saiano, F., Zuddas, P., Mazzola, S., Cuttitta, A., 2013. Relationship between lanthanide contents in aquatic turtles and environmental exposures. Chemosphere 91, 1130–1135. https://doi.org/10.1016/j.chemosphere.2013.01.017.

Chilès, J.-P., Delfiner, P., 2012. Geostatistics: Modeling spatial uncertainty. In: Wiley Series in Probability and Statistics, 2nd Edition, 2nd ed. John Wiley & Sons, Inc, Hoboken, NJ, USA. https://doi.org/10.1002/9781118136188.

Comas-Cufi, M., Thió-Henestrosa, S., 2011. CoDaPack 2.0: A stand-alone, multi-platform compositional software. In: Egozcue, J.J., Tolosana-Delgado, R., Ortego, M.I. (Eds.), CoDaWork'11: 4th International Workshop on Compositional Data Analysis. Sant Feliu de Guixols.

Costantini, E.A.C., Barbetti, R., Fantappiè, M., L'Abate, G., Lorenzetti, R., Magini, S., 2013. Pedodiversity, pp. 105–178. https://doi.org/10.1007/978-94-007-5642-7_6.

da Pereira, W.V.S., Ramos, S.J., Melo, L.C.A., Dias, Y.N., Martins, G.C., Ferreira, L.C.G., Fernandes, A.R., 2023. Human and environmental exposure to rare earth elements in gold mining areas in the northeastern Amazon. Chemosphere 340, 139824. https://doi.org/10.1016/j.chemosphere.2023.139824.

Davis, J.C., 2002. Statistics and Data Analysis in Geology, 3rd ed. John Wiley & Sons.

Drivelos, S.A., Danezis, G.P., Haroutounian, S.A., Georgiou, C.A., 2016. Rare earth elements minimal harvest year variation facilitates robust geographical origin discrimination: the case of PDO "Fava Santorinis.". Food Chem. 213, 238–245. https://doi.org/10.1016/j.foodchem.2016.06.088.

Dushyantha, N., Batapola, N., Ilankoon, I.M.S.K., Rohitha, S., Premasiri, R., Abeysinghe, B., Ratnayake, N., Dissanayake, K., 2020. The story of rare earth elements (REEs): occurrences, global distribution, genesis, geology, mineralogy and global production. Ore Geol. Rev. 122, 103521 https://doi.org/10.1016/j.oregeorev.2020.103521.

Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Barceló-Vidal, C., 2003. Isometric Logratio transformations for compositional data analysis. Math. Geol. 35, 279–300. https://doi.org/10.1023/A:1023818214614.

Fantappiè, M., Priori, S., Costantini, E.A.C., 2015. Soil erosion risk, Sicilian Region (1: 250,000 scale). J. Maps 11, 323–341. https://doi.org/10.1080/17445647.2014.956349.

Fantappiè, M., Priori, S., Costantini, E., 2016. Physiography of the Sicilian region (1: 250,000 scale). J. Maps 12, 111–122. https://doi.org/10.1080/17445647.2014.984785.

Goovaerts, P., 1997. Geostatistics for natural resources evaluation. In: Applied Geostatistics Series. Oxford University Press, New York.

Han, X., Wang, L., Wang, Y., Yang, J., Wan, X., Liang, T., Song, H., Elbana, T.A., Rinklebe, J., 2022. Mechanisms and influencing factors of yttrium sorption on paddy soil: experiments and modeling. Chemosphere 307, 135688. https://doi.org/10.1016/j.chemosphere.2022.135688.

Hu, Z., Richter, H., Sparovek, G., Schnug, E., 2004. Physiological and biochemical effects of rare earth elements on plants and their agricultural significance: a review. J. Plant Nutr. 27, 183–220. https://doi.org/10.1081/PLN-120027555.

Ion, A., Cosac, A., 2023. Rare earth elements distribution in topsoil from Ditrău Alkaline Massif area, eastern Carpathians, Romania. Heliyon 9, e13976. https://doi.org/10.1016/j.heliyon.2023.e13976.

Jackson, J.E., 2003. User's Guide to Principal Components. John Wiley & Sons, New York.

Jiménez-Ballesta, R., Higueras, P.L., García Navarro, F.J., 2024. Rare earths in soils. In: Núñez-Delgado, A. (Ed.), Frontier Studies in Soil Science. Springer International Publishing, Cham, pp. 43–77. https://doi.org/10.1007/978-3-031-50503-4_3.

Jimenez-Espinosa, R., Sousa, A.J., Chica-Olmo, M., 1993. Identification of geochemical anomalies using principal component analysis and factorial kriging analysis. J. Geochem. Explor. 46, 245–256. https://doi.org/10.1016/0375-6742(93)90024-G.

Jo, J., Yamanaka, T., Shin, D., 2023. Nitrogen isotope geochemistry of ion adsorption-type REE mineralization: insights from the weathered granitoid rocks in the Sancheong district, South Korea. Ore Geol. Rev. 157, 105429 https://doi.org/10.1016/j.oregeorev.2023.105429.

Journel, A.G., 1983. Nonparametric estimation of spatial distributions. J. Int. Assoc. Math. Geol. 15, 445–468. https://doi.org/10.1007/BF01031292.

Journel, A.G., Huijbregts, Ch.J., 1978. Mining Geostatistics. Academic Press, London.

Kasper-Zubillaga, J.J., Arellano-Torres, E., Armstrong-Altrin, J.S., Sial, A.N., 2017. A study of carbonate beach sands from the Yucatan peninsula, Mexico: a case study. Carbonates Evaporites 32, 1–12. https://doi.org/10.1007/s13146-015-0283-0.

Lantuéjoul, C., 2002. Geostatistical Simulation. Springer Berlin Heidelberg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-04808-5.

Laveuf, C., Cornu, S., 2009. A review on the potentiality of rare earth elements to trace pedogenetic processes. Geoderma 154, 1–12. https://doi.org/10.1016/j.geoderma.2009.10.002.

Lawrence, M.G., Kamber, B.S., 2006. The behaviour of the rare earth elements during estuarine mixing—revisited. Mar. Chem. 100, 147–161. https://doi.org/10.1016/j.marchem.2005.11.007.

Lawrence, M.G., Greig, A., Collerson, K.D., Kamber, B.S., 2006. Rare earth element and yttrium variability in south East Queensland waterways. Aquat. Geochem. 12, 39–72. https://doi.org/10.1007/s10498-005-4471-8.

Leybourne, M.I., Goodfellow, W.D., Boyle, D.R., Hall, G.M., 2000. Rapid development of negative Ce anomalies in surface waters and contrasting REE patterns in groundwaters associated with Zn–Pb massive sulphide deposits. Appl. Geochem. 15, 695–723. https://doi.org/10.1016/S0883-2927(99)00096-7.

Matheron, G., 1971. The theory of regionalized variables and its applications. In: Les Cahiers du Centre de Morphlogie Mathematique. Ecole Nationale Superieure des Mines de Paris, Fontainebleau.

Matheron, G., 1973. The intrinsic random functions and their applications. Adv. Appl. Probab. 5, 439–468. https://doi.org/10.2307/1425829.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4.

Pastor, A., Gallello, G., Cervera, M.L., de la Guardia, M., 2016. Mineral soil composition interfacing archaeology and chemistry. TrAC Trends Anal. Chem. 78, 48–59. https://doi.org/10.1016/j.trac.2015.07.019.

Pawlowsky-Glahn, V., Egozcue, J.J., 2006. Compositional data and their analysis: An introduction. In: Buccianti, A., Mateu-Figueras, G., Pawlowsky-Glahn, V. (Eds.), Geological Society, London, Special Publications. Geological Society, London, pp. 1–10. https://doi.org/10.1144/GSL.SP.2006.264.01.01.

Pawlowsky-Glahn, V., Egozcue, J.J., 2016a. Spatial analysis of compositional data: a historical review. J. Geochem. Explor. 164, 28–32. https://doi.org/10.1016/j.gexplo.2015.12.010.

Pawlowsky-Glahn, V., Egozcue, J.J., 2016b. Spatial analysis of compositional data: a historical review. J. Geochem. Explor. 164, 28–32. https://doi.org/10.1016/j.gexplo.2015.12.010.

Pawlowsky-Glahn, V., Egozcue, J.J., 2020. Compositional data in Geostatistics: a log-ratio based framework to analyze regionalized compositions. Math. Geosci. 52, 1067–1084. https://doi.org/10.1007/s11004-020-09873-2.

Pawlowsky-Glahn, V., Olea, R.A., 2004. Geostatistical Analysis of Compositional Data. Oxford University Press, New York.

Pawlowsky-Glahn, V., Egozcue, J.J., Lovell, D., 2015a. Tools for compositional data with a total. Stat. Model. 15, 175–190. https://doi.org/10.1177/1471082X14535526.

Pawlowsky-Glahn, V., Egozcue, J.J., Tolosana-Delgado, R., 2015b. Modelling and analysis of compositional data. In: Statistics in practice. John Wiley & Sons, Ltd, Chichester, UK.

Pisciotta, A., Tutone, L., Saiano, F., 2017. Distribution of YLOID in soil-grapevine system (Vitis vinifera L.) as tool for geographical characterization of agro-food products. A two years case study on different grafting combinations. Food Chem. 221, 1214–1220. https://doi.org/10.1016/j.foodchem.2016.11.037.

Ramos, S.J., Dinali, G.S., Oliveira, C., Martins, G.C., Moreira, C.G., Siqueira, J.O., Guilherme, L.R.G., 2016. Rare earth elements in the soil environment. Curr. Pollut. Rep. 2, 28–50. https://doi.org/10.1007/s40726-016-0026-4.

Reimann, C., Birke, M., Demetriades, A., Filzmoser, P., O'Connor, P., 2014. Chemistry of Europe's Agricultural Soils, Part B. Geologisches Jahrbuch (Reihe B), Hannover.

Reimann, C., Fabian, K., Birke, M., Filzmoser, P., Demetriades, A., Négrel, P., Oorts, K., Matschullat, J., de Caritat, P., Albanese, S., Anderson, M., Baritz, R., Batista, M.J., Bel-Ian, A., Cicchella, D., De Vivo, B., De Vos, W., Dinelli, E., Ďuriš, M., Dusza-Dobek, A., Eggen, O.A., Eklund, M., Ernsten, V., Flight, D.M.A., Forrester, S., Fügedi, U., Gilucis, A., Gosar, M., Gregorauskiene, V., De Groot, W., Gulan, A., Halamić, J., Haslinger, E., Hayoz, P., Hoogewerff, J., Hrvatovic, H., Husnjak, S., Jähne-Klingberg, F., Janik, L., Jordan, G., Kaminari, M., Kirby, J., Klos, V., Kwećko, P., Kuti, L., Ladenberger, A., Lima, A., Locutura, J., Lucivjansky, P., Mann, A., Mackovych, D., McLaughlin, M., Malyuk, B.I., Maquil, R., Meuli, R.G., Mol, G., O'Connor, P., Ottesen, R.T., Pasnieczna, A., Petersell, V., Pfleiderer, S., Poňavič, M., Prazeres, C., Radusinović, S., Rauch, U., Salpeteur, I., Scanlon, R., Schedl, A., Scheib, A., Schoeters, I., Šefčik, P., Sellersjö, E., Slaninka, I., Soriano-Disla, J.M., Šorša, A., Svrkota, R., Stafilov, T., Tarvainen, T., Tendavilov, V., Valera, P., Verougstraete, V., Vidojević, D., Zissimos, A., Zomeni, Z., Sadeghi, M., 2018. GEMAS: establishing geochemical background and threshold for 53 chemical elements in European agricultural soil. Appl. Geochem. 88, 302–318. https://doi.org/10.1016/j.apgeochem.2017.01.021.

Ren, L., Cohen, D.R., Rutherford, N.F., Zissimos, A.M., Morisseau, E.G., 2015. Reflections of the geological characteristics of Cyprus in soil rare earth element patterns. Appl. Geochem. 56, 80–93. https://doi.org/10.1016/j.apgeochem.2015.02.011.

Saiano, F., Scalenghe, R., 2009. An anthropic soil transformation fingerprinted by REY patterns. J. Archaeol. Sci. 36, 2502–2506. https://doi.org/10.1016/j.jas.2009.06.025.

Saiano, F., Scalenghe, R., 2019. Soil REE patterns as tracers of the emplacement of metal-rich anthropogenic materials. A case study in Moa (Cuba). J. Soils Sediments 19, 2777–2784. https://doi.org/10.1007/s11368-019-02283-w.

Saiano, F., Pantani, O., Scalenghe, R., 2019. A rapid method of screening ceramic artefacts to reject unlikely hypotheses of provenance. Geoarchaeology 34, 759–767. https://doi.org/10.1002/gea.21749.

Shomar, B., Sankaran, R., Solano, J.R., 2023. Mapping of trace elements in topsoil of arid areas and assessment of ecological and human health risks in Qatar. Environ. Res. 225, 115456 https://doi.org/10.1016/j.envres.2023.115456.

Sonke, J.E., 2006. Lanthanide–humic substances complexation. II. Calibration of humic ion-binding model V. Environ. Sci. Technol. 40, 7481–7487. https://doi.org/10.1021/es060490g.

Tolosana-Delgado, R., Mueller, U., 2021. Geostatistics for Compositional Data with R, Use R! Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-82568-3.

Tolosana-Delgado, R., Otero, N., Pawlowsky-Glahn, V., 2005. Some basic concepts of compositional geometry. Math. Geol. 37, 673–680. https://doi.org/10.1007/s11004-005-7374-8.

Tolosana-Delgado, R., van den Boogaart, K.G., Pawlowsky-Glahn, V., 2011. Geostatistics for compositions. In: Compositional Data Analysis: Theory and Applications. John Wiley & Sons, Ltd, Chichester, UK, pp. 73–86. https://doi.org/10.1002/9781119976462.ch6.

Tolosana-Delgado, R., Mueller, U., van den Boogaart, K.G., 2018. Geostatistics for compositional data: an overview. Math. Geosci. 1–42 https://doi.org/10.1007/s11004-018-9769-3.

Tolosana-Delgado, R., Mueller, U., van den Boogaart, K.G., 2019. Geostatistics for compositional data: an overview. Math. Geosci. 51, 485–526. https://doi.org/10.1007/s11004-018-9769-3.

Tyler, G., 2004. Rare earth elements in soil and plant systems - a review. Plant Soil 267, 191–206. https://doi.org/10.1007/s11104-005-4888-2.

Wackernagel, Hans, 2003. Multivariate Geostatistics: An Introduction with Applications. Springer.

Webster, R., Oliver, M.A., 2007. Geostatistics for Environmental Scientists, Statistics in Practice. John Wiley & Sons, Ltd, Chichester, UK. https://doi.org/10.1002/9780470517277.

Wedepohl, H.K., 1995. The composition of the continental crust. Geochim. Cosmochim. Acta 59, 1217–1232. https://doi.org/10.1016/0016-7037(95)00038-2.

Weng, Z.H., Jowitt, S.M., Mudd, G.M., Haque, N., 2013. Assessing rare earth element mineral deposit types and links to environmental impacts. Appl. Earth Sci. 122, 83–96. https://doi.org/10.1179/1743275813Y.0000000036.