Geogenic versus anthropogenic behaviour and geochemical footprint of Al, Na, K and P in the Campania region (Southern Italy) soils through compositional data analysis and enrichment factor

Matar Thiombane[1*], Marcello Di Bonito[2], Stefano Albanese[1], Daniela Zuzolo[3], Annamaria Lima[1], Benedetto De Vivo[4]

[1]. Department of Earth, Environment and Resources Sciences (DiSTAR), University of Napoli "Federico II", Complesso Universitario di Monte Sant'Angelo, Via Cintia snc, 80125 Napoli, Italy.

[2]. School of Animal, Rural and Environmental Sciences, Brackenhurst Campus Southwell NG25 0QF Nottingham Trent University, Unites Kingdom.

[3]. Department of Science and Technology, University of Sannio, via dei Mulini 59/A, 82100 Benevento, Italy.

[4]. Pegaso University, Piazza Trieste e Trento 48, 80132 Naples & Benecon Scarl, Dip. Ambiente e Territorio, Via S. Maria di Costantinopoli 104, 80138 Naples, Italy.

*First author, email:thiombane.matar@unina.it

Abstract

Geochemical studies that focus on environmental applications tend to approach the chemical elements as individual entities and may therefore offer only partial and sometimes biased interpretations of their distributions and behaviour. A potential alternative approach is to consider a compositional data analysis, where every element is part of a whole. In this study, an integrated methodology, which included compositional data analysis, multifractal data transformations and interpolation, as well as enrichment factor analysis, was applied to a geochemical dataset for the Campania region, in the south of Italy, focusing in particular on the behaviour, footprints and sources of a smaller pool of elements: Al, Na, K and P. The initial dataset included 3669 topsoil samples, collected at an average sampling density of 1 site per 2.3 km^2, and analysed (after an aqua regia extraction) by a combination of ICP-AES and ICP-MS for 53 elements. Frequency based methods (Clr biplot, Enrichment Factor computation) and frequency spatial-method (fractal and multifractal plots) allowed identifying the relationships between the elements and their possible source patterns in Campania soils in relation to a natural occurring concentrations in geogenic material (rocks, soils and sediments) or human input. Results showed how the interpretation of concentration and behaviour of Al, Na, K and P was enhanced thanks to the application of data log-ratio transformation in univariate and multivariate analysis compared to the use of raw or log-normal data. Multivariate analyses with compositional biplot allowed the identification of four element associations and their potential association with the underling geology and/or human activities. When focusing on the smaller pool of elements (Al, P, K and Na), these relationships with the unique geology of the region were largely confirmed by multifractal interpolated maps. However, when the local background was used for the calculation of the enrichment factor, the resulting interpolated maps allowed to identify smaller areas where the greater concentrations of P could not be possibly associated to a mineralization (e.g., ultrapotassic rocks) but were more likely to be associated to anthropogenic input such as agriculture activities with potentially extensive use of phosphate fertilizers. The integrated approach of this study allowed a more robust qualitative and quantitative evaluation of elemental concentration, providing in particular new and vital information on the distribution and patterns of P in soils of the Campania region, but also a viable, more robust, methodological approach to regional environmental geochemistry studies.

Keywords: Compositional data; multi-fractal computation; robust biplot; background concentration; enrichment factor

1. Introduction

Element distributions in soils are generally related to a variety of factors such as geology, chemical reactivity, mineralogy, hydrology, vegetation, and anthropogenic activities. In recent years environmental geochemistry has allowed to gain much insight on the relationship between chemical elements in soils and their sources by means of various tools. In particular, univariate and multivariate analysis (Otero et al., 2005; Reimann et al., 2008;; Zuo, 2011; Thiombane et al., 2018a) as well as spatial analysis (frequency space-method) have helped to discriminate between anthropogenic contribution (and contamination) compared to natural or geogenic sources. More recently, different approaches have also been used to treat compositional geochemical data in a more comprehensive way. Geochemical data are closed number systems (parts of a whole) and they should be treated as that to avoid spurious correlations and misleading interpretations. By taking only the absolute values or the log-normal data, the proportional nature of the geometry of the simplex may be not fully captured (Han and Kamber, 2001). Compositional data instead, consider each element as part of a whole which carry relative information (Aitchison, 1986; Egozcue et al., 2003). For this reason, and to minimize and/or eliminate the presence of outliers and spurious correlation (Pawlowsky-Glahn and Buccianti, 2011), they have increasingly been approached in alternative ways. In particular, the use of log-ratio transformations such as additive log-ratio (alr), centered log-ratio (clr) (Aitchison, 1986) and isometric log-ratio (ilr) (Egozcue et al., 2003) were found to be effective approaches to deal with these complex datasets and better explain their significance. Furthermore, mapping remains a useful tool to display data distribution through fractal and multifractal analysis (Mandelbrot, 1983; Cheng et al., 1994) and through geostatistics (Oleo et al., 2018; Thiombane et al., 2018b). These have been demonstrated to be a powerful means for identifying geochemical anomalies (Cheng, 1999-, 2007; Cheng et al., 2000, 2010; Lima et al., 2003a; Cicchella et al., 2005) and enhance elements patterns.

Aside from their inherent complexity, often environmental geochemistry studies still present some discrepancies in the definition of key concepts and therefore interpretation of results. For example, the concept of background and/or baseline concentration are still confused and misused, whilst they should be clearly defined and interpreted: a background value corresponds to the range of concentration of a given element in a given area which is completely dependent on the compositional and mineralogical characteristic of the parent/source geological material (Reimann et al., 2005; Albanese et al., 2007). On the other hand, a baseline value relates the actual most diffused range of concentration of a given element in a specific area depending both on the nature of the parent geological/source material (Salminen and Gregorauskiene, 2000) and on the historic diffuse release into the environment from anthropogenic sources.

In reality, in areas where the anthropogenic impact is very small or not significant, the background and baseline values can overlap, as it is often difficult to distinguish the natural sources compared to the anthropogenic contribution. In areas where the anthropogenic impact is evident or significant, however, the background and baseline values should be interpreted appropriately and in a distinct way. This study focuses on a region in south of Italy, Campania, characterized by complex geological and geomorphological features and by a significant presence of human activities, such as industry, agriculture, tourism and urbanization. The industrial presence is often related to the agricultural activities and is mostly developed in proximity of the agricultural and surrounding urban areas. Many of these industrial activities are devoted to food production and preservation process (e.g., San Marzano tomatoes conserves), but also to clothes productions and tanneries (e.g., Solofra tanneries industries). These activities represent the main livelihood for the local communities and a major economic input not only for the Campania region, but for the Italian economy. Nevertheless, as with any agriculture activity, it is inevitable that chemical fertilizers (phosphates and sulfates based) are used to improve soil productivity and quality. Some of the benefits of using fertilizers, however, carry also 'externalities' that can travel from the soil, to the crop, and the entire food chain. In particular, metallic impurities of P-based fertilizers containing Cu, As and Cd, can affect the quality of the soil and possibly its contamination; phosphate fertilizers are in fact known as a major source of trace metals among all mineral fertilizers (Nziguheba and Smolders, 2008). These sources of contamination can be investigated by means of Enrichment Factor (Chester and Stoner, 1973) which is able to display the degree of contamination related to a mineralisation (geogenic) or anthropogenic activities (Reimann and de Caritat, 2005).

This study proposes an assessment of the potential impacts and footprint of some of the agricultural activities in Campania, by use of compositional data analysis, concentration-area (C-A) and spectrum-areas (S-A) fractal and multifractals models applied to geochemical data. In particular, the study focuses on a smaller pool of elements that can be potentially directly related to either geogenic, anthropogenic or mixed source: Al, Na, K, and P will be investigated to detect their potential origin and their background concentrations in Campania soils. The main objectives of this study were:

(1) to illustrate the importance of compositional log-transformations on geochemical data; (2) to delineate the main sources of elements by using a combination of multivariate analysis and GIS based approach; (3) to use enrichment factor (EF) to assess geogenic or anthropogenic behaviour of P in the study area; (4) to prove the importance and ease of using the local reference elements as opposed to the traditional continental crust references values in EF calculation to determine degree of contamination. The results from this investigation could greatly influence and provide a blueprint to future similar studies.

2. Materials and methods

2.1. Features of the study areas

The Campania region is located in the south of Italy, covering an area of about 13,600 km^2. The region borders the Tyrrhenian Sea and the Lazio region at the western and northern sides, respectively (Fig. 1-A).

[Figure 1 about here]

The main geological features of the Campania region are constituted by the Apennines chain which cross the areas as a backbone oriented NW-SE. The hilliest part of the chain forms the Mt. Matese in the north, Mt. Taburno and Picentini in the center, and the Mt. Alburni in the southeast. These mountains are mostly formed by sedimentary rocks such as limestone and dolostone whilst the external domains are constituted by siliceous schist and terrigenous sediments (clays, siltstone, sandstone, and conglomerate) (Bonardi et al., 1998; De Vivo et al., 2016). Igneous rocks are present in the region and are mostly formed by potassic and ultrapotassic volcanic rocks of the Roccamonfina volcano (De Vivo et al., 2001, 2010, 2016; Rolandi et al., 2003; Albanese et al., 2007;) in the northwest, the Mt. Somma-Vesuvius, Campi Flegrei and Ischia volcanoes which are located nearer the coastal areas of the study region. The coastal areas and plains are constituted by alluvial, lacustrine and coastal (mixes of oceanic and terrestrial) sediments (Bonardi et al., 1998; De Vivo et al., 2016).

The hydrography of the Campania region is characterised by three main river catchments (Garigliano, Volturno and Sele) and by numerous minor streams, mostly draining towards the Tyrrhenian Sea. The morphology of the drainage network is irregular and controlled by geological and structural features and rivers are source of water irrigation for agriculture field (Ducci and Tranfaglia, 2005).

Campania is one of the most populated regions of Italy with more than 5.8 million inhabitants (ISTAT, 2016).This high density of population is coupled by the presence of a large number of industrial activities, with the majority being involved in agriculture: vineyards and olive plantations - mostly in hilly areas – seasonal crops, and greenhouse products (tomatoes, potatoes, aubergines, peppers, peas, and citrus fruits) represent major resources for the region and the local economy (Albanese et al., 2007). This intensive agriculture activity occupies more than 50% of the total land and occurs mostly in the coastal and mountainous areas (Fig. 1B), where fertile land and suitable soils are occurring. Unfortunately, such industrial activities are known to have a potential negative impact if not properly managed, contributing to the contamination of natural resources such as superficial and groundwater as well as soils. Campania is not immune to these problems, and some studies have already highlighted their existence and relation to its natural resources (Cicchella et al., 2005; Minolfi et al., 2018).

2.2. Sampling procedures and analyses

From 2013 to 2015, 3669 samples were collected from topsoil of the Campania Region (13,600 km^2) with a nominal density of 1 sample in each 3.2 km^2. Each top soil sample (from 0-20 cm) was made by homogenizing 5 subsamples at the corners and the centre of a 100m^2 square, collecting approximately 1.5 kg in total. The sampling procedure followed the Geochemical Mapping of Agricultural and Grazing Land Soils (GEMAS) sampling procedure described by Reimann et al. (2014). At each sampling site, several physico-chemical parameters of the soil properties were measured, including pH, total water content, conductivity, total organic content and the geographical coordinates system recorded by geospatial positioning systems (GPS).

Chemical analyses were carried out at an international accredited Laboratory, Acme Analytical Laboratories Ltd (now Bureau Veritas, Vancouver, Canada). The samples were analyzed after an aqua regia extraction, by a combination of inductively coupled plasma atomic emission (ICP-AES) and inductively coupled plasma mass spectrometry (ICP/MS) for "pseudototal" concentration of 53 elements (Ag, Al, As, Au, B, Ba, Be, Bi, Ca, Cd, Ce, Co, Cr, Cs, Cu, Fe, Ga, Ge, Hf, Hg, In, K, La, Li, Mg, Mn, Mo, Na, Nb, Ni, P, Pb, Pd, Pt, Rb, Re, S, Sb, Sc, Se, Sn, Sr, Ta, Te, Th, Ti, Tl, U, V, W, Y, Zn, and Zr). A sub-sample of 15 g of the sieved <2 mm soil fraction was digested in 90 ml aqua regia and leached for 1 hour in a 95°C water bath. After cooling, the solution was diluted to a final volume of 300 ml using a solution of 5% HCl. The sample weight to solution volume ratio was 1 g per 20 ml. The solutions were analyzed using a Perkin Elmer Elan 6,000/9,000 inductively coupled plasma emission mass spectrometer (ICP-MS). The accuracy and precision of the data was measured by comparison to known analytical standards. Calibration solutions were included at the beginning and end of each analytical run (a total of 40 solutions). Precision is ±100% at the detection limit, and improves to better than ±10% at concentrations 50 times the detection limit or greater.

2.3. Compositional data analysis

Nowadays it appears necessary to reconsider geochemical data under a compositional data analysis perspective (Aitchison, 1986; Buccianti et al., 2006, 2014, 2017; Pawlowsky-Glahn et al., 2011). A composition is defined as a sample space of the regular unit D-simplex, S^D that is a vector of *D* positive components summing up to a given constant *k*, set typically equal to 1 (proportions), 100 (percentages), or 10^6 (ppm) by closure. It relates parts of some whole that carry relative information (ratios of components) whose sample space is the simplex (Pawlowsky-Glahn and Egozcue, 2001).

$$S^D = \{X=[x_1, x_2 \dots x_D] \mid x_i>0; \sum_{i=1}^{D} xi = k\} \qquad (1)$$

As explained in the introduction, working with log-ratio transformation such as additive log-ratio (alr), centered log-ratio (clr) and isometric log-ratio (ilr) allows to overcome some of the issues related to the complexity of geochemical data, helping to highlight the relative magnitudes and variations of the components of a composition rather than their absolute values (Buccianti and Magli, 2011). In this study, due to its orthonormal character, ilr was applied on the datasets and compared to raw and log-normal data to show how it allows to normalize the data distribution and its closure effects of geochemical data prior to statistical analyses (Fig. 2-4).

[Figure 2-4 about here]

This log-ratio transformation was applied on data taking into account the compositional vectors of n parts partitioned into groups of parts presenting a certain affinity (Egozcue et al., 2003; Filzmoser and Hron, 2008, Filzmoser et al., 2009a, 2009b; Thiombane et al., 2018a).

To better visualize the element distributions and possible natural and anthropogenic behaviour of the variables into the survey area, a compositional biplot was created. This is a powerful statistical tool that displays both samples and the variables of a data matrix in terms of the resulting scores and loading (Gabriel, 1971). Thus, the scores represent the structure of the compositional data into a Euclidian space based on variance and covariance matrix; moreover, they display the association structure of the dataset. The biplots present rays (or vectors) defined from the center of the plot, where their length is proportional to the amount of explained variance (communality) of the variables it represents. The interpretation of the graphic depends on the loading (rays) structures and in more details on the approximate links between rays and samples, the distances between vertexes and their directions (Otero et al., 2005). For a full description of compositional biplots and an appreciation of their utility, several examples are available in the literature (e.g., Pison et al., 2003; Maronna et al., 2006; Filzmoser and Hron, 2008, Filzmoser et al., 2009a, 2009b; Hron et al., 2010).

When, however, biplots are used with raw data, these can be substantially influenced by the occurrence of outliers which can mislead the compositional nature of the data matrix and affect the principal components when interpreting results (Aitchison, 1986; Filzmoser et al., 2009a, 2009b). For these and others reasons log-transformed data are recommended to be used in multivariate analysis, and strengthened compositional biplots (Egozcue et al., 2003; Filzmoser et al., 2009a; Hron et al., 2010). Taking into account the singularity of the clr transformed data (Aitchison, 1986), these should be computed in orthonormal coordinates such as ilr transformed data, and back-transformed to the clr space for further interpretation. This back transformation allows preserving the linear relationship between clr and ilr coordinates (Egozcue et al., 2003). Furthermore, the application of the minimum covariance determinant (MCD) estimator (Roosseeuw and Van Driessen, 1999) allows to display the observations to the smallest determinant of their sample covariance matrix which tend to hold the variables into a normal distribution. In this study, a classical compositional biplot (CCB) and a robust compositional biplot (RCB) were used to identify the relationships between variables using compositional raw data and log-transformed data, respectively (Fig. 5).

From the total of 53 elements analysed for the soils, only eighteen elements were considered to test this approach, with the aim of better representing the correlation between variables and investigate more robustly their main sources in the study area. Their main descriptive statistics are shown in Table 1.

[Table 1 about here]

The number of elements was reduced to 18 variables based on 3 main criteria: 1) the removal of elements with more than 40 % of values below the detection limit (LOD), 2) choosing arbitrary mostly two representative Rare Earth elements which are geochemically congruent and 3) choosing elements with a communality of extraction higher than 0.5 (50%) and/or common variances <0.5 (e.g. Reimann et al., 2002).

Based on the robust biplot, sequential binary partition was implemented using the same 18 variables by dividing them into specific groups of non-overlapping elements (Table 2).

[Table 2 about here]

Balances are particular ilr-coordinates (isometric-logratio) having orthonormal bases which can be interpreted in the D-1 (D: dimension) real space as ratios of elemental associations (Egozcue et al., 2003). Balances can be calculated using the following formula:

$$\mathrm{Zi} = \sqrt{\frac{\mathrm{rs}}{\mathrm{r+s}}}\ \ln \frac{(\prod_{+} xj)^{1/r}}{(\prod_{-} xk)^{1/s}} \quad \text{for i= 1,...., D-1,} \qquad (2)$$

where the products $\prod_+$ and $\prod_-$ only include parts coded with + and -, and r and s are the numbers of positive and negative signs (parts) in the i-th order partition, respectively (Egozcue and Pawlowsky-Glahn, 2005). From the established sequential binary partition and eq. (2), Z_1 (Al/P) and Z_2 (Na/K) were calculated and ilr coordinates displayed through geospatial mapping. Balances were back-transformed based on the sequential binary partition matrix and the bijection between the original space of the parts and that of the log-ratios (Egozcue et al.; 2003; Egozcue and Pawlowsky-Glahn 2005; Olea et al., 2018). The back-transformed results in the original part space for Al, P, Na and K elements concentrations were computed before applying geostatistical computations.

2.4. Interpolated and Background/baseline maps

Geographical Information Systems (GIS) and technology was used to map and display data distribution and characterize the footprint, possible main sources, and the behaviour of the elements considered. For this study, one of the aims was to determine the background concentration of major elements Al, Na, K, and P in the Campania soils. ArcGIS (ESRI, 2012) and GeoDAS (Cheng et al., 2001) were used as the main GIS tools. In particular, GeoDAS™ was used to produce interpolated geochemical maps by means of the multifractal inverse distance weighted (MIDW) algorithm (Lima et al., 2003a). In previous geochemical studies of the Campania region (De Vivo et al., 2001; Lima et al., 2003a; Cicchella et al., 2005; Albanese et al., 2007), the MIDW was chosen as an interpolation method as it preserves high frequency information (anomalies), while taking into account both spatial associations and local singularity in geochemical data (Cheng, 1999). The concentration–area (C–A) fractal method (Cheng et al., 1994) was applied to set the concentration intervals of the interpolated surfaces generated by the MIDW method, and ArcGIS™ software was used for the graphical presentation of the results (Figure 6).

[Figure 6 about here]

Different tools are used to determine the background and baseline concentration of elements (EPA, 2001; Reimann et al., 2005; APAT-ISS, 2006; Tarvainen and Jarva, 2011; Cave et al., 2012; Ander et al, 2013). They are called "traditional approaches" due to the fact that most of them do not take into account both spatial association and the data distribution local singularity. In this study, maps showing geochemical background/baseline concentrations have been obtained using the S-A (spectrum-areas) method which preserves high frequency information. The S–A method is a fractal filtering technique, based on a Fourier spectral analysis (Cheng, 1999; Cheng et al., 2001), and is used to separate anomalies from background values starting from a geochemical interpolated concentrations map. It also uses both frequency and spatial information for geochemical map and image processing. Fourier transformation can convert geochemical values into a frequency domain in which different patterns of frequencies can be identified. The signals with certain ranges of frequencies can be converted back to the spatial domain by inverse Fourier transformation (Zuo et al., 2015; Zuo and Wang, 2016). The interpolated maps generated from geochemical data have been transformed into the frequency domain in which a spatial concentration–area fractal method has been applied to distinguish the patterns on the basis of the power-spectrum distribution. A log–log plot (Fig. 8; Fig. 10 under plots) was used to show the relationship between the area and the power spectrum values on the Fourier transformed map of the power spectrum. The values on the log–log plot were modelled by fitting straight lines using least squares. Distinct classes can be generated, such as lower, intermediate, and high power spectrum values approximately corresponding to baseline values, anomalies, and noise of geochemical values in the spatial domain (Fig.9; Fig. 11, under plots), respectively.

2.5. Enrichment factor

The Enrichment Factor (EF) approach, which was historically introduced to identify the level of (economically viable) mineralisation and origin of elements in atmosphere, precipitation or seawater (Goldberg, 1972; Chester and Stoner, 1973; Peirson et al., 1974; Duce et al., 1975; Rahn, 1976; Buat-Ménard and Chesselet, 1979) was used in this study to ascertain soil contamination on a long term scale (see for example Hakanson, 1980; Sutherland et al., 2000; Abrahim and Parker, 2008;; Wu et al., 2011; Saeedi et al., 2012;;). EF is computed using the equation described below (Eq. 2) which was first introduced by Chester and Stoner (1973):

$$\mathbf{EF} = \frac{\left(\frac{\mathbf{Cx}}{\mathbf{Cref}}\right)\mathbf{sample}}{\left(\frac{\mathbf{Cx}}{\mathbf{Cref}}\right)\mathbf{background}} \qquad (3)$$

were C_x is the concentration of the element under consideration and C_{ref} is the concentration of a reference element. Here, the reference element is an element that is particularly stable in soil. In fact, the stability of the element is demonstrated by a vertical immobility and/or his chemical stability (non-degradability) (Reimann et al., 2008). Aluminium, Sc, Zr and Ti are the main elements considered in the literature to be stable, and naturally occurring in soils. In this study, the choice of the most stable element in EF computation was based on a robust statistical estimation called coefficient of variation (CV) which allows a more extensive interpretation of the variability of distribution of reference elements (Al, Zr, Sc, and Ti) using the equation:

$$\mathbf{CV} = \frac{\mathbf{MAD}}{\mathbf{MD}} \times \mathbf{100\%} \qquad (4)$$

Where CV displays the variability of distribution in percentage (%), MAD corresponds to the median absolute deviation that is the median value (50th percentile) of the deviations of all concentrations from the median value of concentration and MD is the median concentration. This is a robust, nonparametric estimate that is not affected by the presence of outliers (Reimann and de Caritat, 2005). The lower the CV value, the more the element is stable (Table 3).

[Table 3 about here]

This study intended to investigate the most effective EF calculation by comparing the use of the reference element in continental crust (Martin and Whitfield, 1983; Peirson et al., 1974;; Taylor and McLennan, 1995; Wedepohl, 1995; Loska et al., 1997) versus the use of the reference element of the local background area as advised by Reimann and de Caritat. (2005) and Sutherland et al. (2000). The variation of P EFs in the study area was then displayed by means of interpolated maps where range of EF scores were calculated taking into account the contamination factor in accordance with Sutherland et al. (2010) (Figure 12, see legend).

3. Results and discussion

3.1. Univariate and multivariate analysis

Results for Na, K and P elemental distribution have been presented by combining Edaplots (top) and CP plots (bottom) with three different data type: raw data concentration (left), log-normal data (middle) and ilr transformed data (right) (Figs. 2, 3 and 4).

The Edaplots for Na (Fig. 2) show different 'shapes' depending on their type of data: the raw data distribution is right-skewed while the log-normal data is left-skewed. This highlights how both raw and log-normal data representation do not match well the real data distribution compared to the ilr transformation. By using the ilr transformed data, the distribution (as shown by histogram and density plot) tends to a normal data distribution. A similar result can be observed for K and P data distribution (in Figs 3 and 4). The strength of ilr transformation in data distribution is shown in the CP plot, which displays the cumulative curve distribution where the straight-line symbolizes the most adequate model of a normal data distribution; by using the ilr transformed data, the elemental distribution fits very closely the straight line compared to raw and log-normal data, which are affected by the occurrence of outliers. The compositional biplot (Fig. 5), based on principal component analysis, displays the correlation and relationship between 18 analytical variables in 3669 sample points, from which the first two principal components were extracted. The principal components are presented in a compositional biplot using raw data (Fig. 5, left) and ilr coordinates clr back-transformed (Fig. 5, right).

The total variance of initial raw data biplot (classical biplot) explains 43.45%, where the first principal component (PC1) accounts for 29.71% and the second principal component (PC2) accounts for 13.74% (Fig. 5, left). On the other hand, the robust biplot based on ilr coordinates and clr back transformed produced significantly different results (Fig. 5, right) with PC1 explaining 47.45% and PC2 explaining 35.12% of the compositional variability.

[Figure 5 about here]

By taking into account the direction and angles formed between the vectors, and the proximity of the rays, it is possible to identify the presence of four groups of elements, which are most likely related to the geogenic features and/or the main human activities in the study areas. The main groups highlighted by the biplots are:

- ✓ Fe, Mn, Co and Ni, a group of element association characterized by their rays closed to one-another and tending to the same direction (classical biplot). This group of element is marked by high communalities (length of the rays) of the vectors. This may be expected given that this association is strongly related to the adsorption and coprecipitation effects operated by Fe and Mn oxides and hydroxides occurring mostly in the sedimentary deposits such as marl-sandstone, conglomerate and silico-clastic flysch deposits outcropping in the surveyed areas (Cicchella et al., 2005; Albanese et al., 2007; Buccianti et al., 2015).
- ✓ Al, Th, As, V, Ti, and Mo form an association of elements where Al, Ti and V dominate the groups with the highest length of their vertexes whilst Mo has a relatively lower communality (classical biplot). This behaviour is possibly related to the fact that these elements are mostly immobile during weathering phenomena of the parental rocks and mostly remaining in the residual fraction of soils. This group could therefore be directly related to the parental geology of the surveyed areas which are dominated by the influence of pyroclastic deposits from different eruptions of nearby volcanoes such as Roccamonfina, Vesuvius, Campi Flegrei (De Vivo et al., 2010).
- ✓ Na, K, P, Cu, Pb and Zn elemental association is dominated by a high communality of Na and P as well as the vicinity of their rays (classical biplot). This elemental association probably reflects the potassic and ultrapotassic rock formations that occur throughout the majority of the

slope of Naples and Benevento areas, associated to the lava and pyroclastic volcanic activity of Mt. Somma–Vesuvius and Roccamonfina (Lima et al., 2003b; Albanese et al., 2013). Zinc and Pb display short vectors which are poorly characterized and seem to be only partially correlated to the others element of this association. These two elements may be related to anthropogenic activities such fossil fuel combustion, as well as industrial and vehicular emissions release (Cicchella et al., 2015; De Vivo et al., 2016).

- ✓ Ca and Mg appear to be correlated and present both a short length of their rays (classical biplot). The communality of Ca is larger and seems to be independent of all others elements due to the fact that the angles formed are greater than 90° compared to Mg. This confirms the high correlation between Ca and Mg, which might be possibly related to the limestones and dolostones of the Mt. Picentini, Mt Lattari and Mt. Cervati.
- ✓ On a closer observation, the elemental association Na, K, P, Cu, Pb and Zn could be reduced to two main subgroups, where the variables K and Na are strongly overlapping on to each other (robust biplot). This highlights the high correlation between these two elements occurring mostly in potassic and ultrapotassic rock formations throughout the surveyed areas. Phosphorous and Cu are also highly correlated both with high communalities, where P seems to be independent of the Na/K vertexes, forming almost an angle of 90° (robust biplot). Interestingly, Cu seems to be poorly correlated to a geogenic origin, as the direction of its ray compared to those of the group of sulfide elements (e.g. Co and Ni) have an angle up to 90°.This may signify that Cu is independent from other naturally occurring sulfide elements, whilst it seems to be highly correlated to P in most soil of the surveyed area. One potential explanation is that P and Cu may be related to agriculture activities, with large areas cultivated as vineyards, where the use of pesticides and phosphate fertilizers is very high (Cicchella et al., 2005).
- ✓ For the Al, Th, As, V, Ti, and Mo association, it is observed the dissociation of Ti variable with a high communality of the ray (robust biplot). Titanium is considered as an immobile and stable element due to low mobility and is mostly found in volcanic materials (Egli et al., 2008).

3.2. Geochemical elemental distribution in the survey areas

Based on the robust biplot, 18 elements have been chosen to perform sequential binary partition and obtain balances (specific ilr-coordinates) (Table 2). In this section, Balances Z_1 (Al/P) and Z_2 (Na/K) will be displayed to reveal the data for the elements of main interest (Al, P, Na and K).

The first balance Z_1 (Al/P) map, ranging from 2.89 to 4.79 reveals a higher proportion of Al in correspondence to large volcanic complexes like Mt. Roccamonfina and Phlegraean fields (Fig.7A). In addition, high proportion of Al is also highlighted in correspondence of part of the Lattari range, along the Apennines and in patches at the southern part of our study area. Scheib et al. (2014) highlighted that the Mt. Roccamonfina volcano is characterized by pyroclastics rocks with high level of elements such Al, Th and Ti, as well as in the Campanian Ignimbrites in the Apennines (De Vivo et al., 2010).

[Figure 7 about here]

In contrast, the higher proportions of P (Fig.7A) in correspondence to lower values of coordinate (ranging from 0.6 to 1.95) are found around Mt. Somma.Vesuvius, and in several areas of the eastern region of our study area, where large agricultural fields (e.g., vineyards and orchards) are located.

The second balance map Z_2 (Na/K) (ranging from -0.65 to 0.41) shows the dominance of Na and K in the study area (Fig. 7B). In fact, higher proportion of Na can be observed in correspondence to Mt. Roccamonfina, Phlegrean fields, and Ischia Island. Sodium may be related to the potassic and ultrapotassic rocks and volcano-sedimentary deposits from major sector collapse of volcanoes in the study area (Scheib et al., 2014; De Vivo et al., 2016). In contrast, the higher abundance of K corresponding to lower balances (ranging from -3.26 to -2.48), can be observed in the southern part of our study area. In fact, Thiombane et al. (2018a) showed a high enrichment of K in silico-clastic deposits dominated by flysch series in southern part of our study area. Furthermore, pyroclastic rocks from different eruptions of nearby volcanoes (Roccamonfina, Vesuvius, Phlegrean Fields - De Vivo et al., 2010; Buccianti et al., 2015; Mt. Vulture and Aeolian islands - Peccerillo, 2005; Scheib et al., 2014) are found in this area. The back-transformation of balances will give the same values of Al, P, Na and P elements concentration because of the bijection between the original space of the parts and that of the ilr-logtransformation (Egozcue et al., 2003; Olea et al., 2018).The concentration of Al, ranging from 2,344 to 94,334 mg/kg with a mean value of 32,918 mg/kg, was separated into five ranges according to C-A fractal plot (Fig. 8A, plot below).

The lowest concentration values roughly ranging from 2,344 to 30,000 mg/kg, are found in the north-eastern and south-western part of the study area in correspondence with the Apennine chain and the Cervati Mt., respectively. The highest concentrations (up to 57,000 mg/kg) are found in soils on the slope of the volcanoes (Mt. Somma.Vesuvius and Roccamonfina), surrounding the Mt. Matese and the Mt. Lattari (Figure 8A). The elevated concentration of Al in soils surrounding the volcanoes is possibly related to the parental pyroclastics which subsequently formed soils (De Vivo et al., 2016). On the other hand, in the Mt. Matese and Mts. Lattari, Al concentrations could result from the occurrence of several imbrications of bauxite minerals which were exploited in the first part of the 20th century (Mondillo et al., 2011). The background/baseline map of Al (Fig. 9A) shows the highest concentration ranging from 51,000 to 61,000 mg/kg in correspondence with soils around the Roccamonfina and Vesuvius volcanoes.

[Figure 8 and 9 about here]

Phosphorus content in Campania soils ranged from 156 to 16,087 mg/kg (Figure 8B) with a mean value of 1,011mg/kg, which corresponds to the mean level of the European soil (Tóth et al., 2014). Average concentrations (ranging from 64,935 to 75,000 mg/kg) increased significantly near Mt. Somma-Vesuvius and the highest values (up to 75,000 mg/kg) were found in Vitulano municipality (Benevento Province), and in the nearby Lioni and Laviano districts (Avellino Province). In particular, the corresponding underlying geology does not seem to be able to account for this P elemental anomaly: the soils of Vitulano district are mostly from sandstone, flysch deposits and limestone whereas Lioni and Laviano districts soils are mostly from limestones. The high P concentration might be possibly related to anthropogenic activities related to the large fertilizers use in agriculture. Part of the Vitulano municipality lies in fact in an area with intensive vineyard occurrence. The background/baseline P distribution (Fig. 9B) is characterised by more than 95% of the survey area with P values <1,500 mg/kg. This range of concentration is usually characteristic of siliciclastic, limestone and dolostone deposits/geology. Greater concentrations ranging from 2,200 to 4,600 mg/kg were found on the slopes of the Mt. Somma-Vesuvius, and Lioni and Laviano districts, possibly linked to geogenic and anthropogenic sources, respectively.

Sodium and K concentrations ranged from 20 to 17,592 mg/kg with a mean concentration of 1,265 mg/kg, and from 804 to 63,850 mg/kg with a mean concentration of 6,852 mg/kg, respectively (Table 1). The highest values of Na (up to 10,900mg/kg) were found in soil samples on the slopes of the Somma-Vesuvius and Roccamonfina volcanoes, in the Phlegrean Fields and Ischia Island whereas those for K (up to 41,522 mg/kg) were found mostly on soils surrounding the Mt. Somma-Vesuvius (Figures 10A and 10B). These values could reliably be attributed to the occurrence of volcanic rocks and related soils in the surveyed areas. These formations are dominated by potassic and ultrapotassic lavas and pyroclastic materials. The interpolated maps reflect hence the concentrations of Na and K elements in such rocks and pyroclastics formations linked to Quaternary volcanic activities (Peccerillo, 2005; Albanese et al., 2013; Buccianti et al., 2015). In general, the soils of the Vesuvian area, formed from a more recent volcanic activity (Joron et al., 1987; De Vivo et al., 2003;Lima et al., 2003b) are much richer in Na and K than the soils on the much older Roccamonfina volcano. This is due to the fact that Na and K are relatively mobile and easily leached elements in the surficial environment.

The background maps of the Na and K (Figs. 11A and 11B) show low concentration ranging from 20.1 to 3,200 mg/kg and from 804 to 9,100 mg/kg, respectively. These relatively low ranges of concentration of Na and K were found to correspond to the same lithologies. At regional level, the highest concentration of Na and K was found in soil on the slopes of the Roccamonfina and Mt. Somma-Vesuvius, ranging from 3,200 to 9,700 mg/kg and from 9,100 to 46,900 mg/kg, respectively. This allows delimiting two ranges of background concentration of Na and K related to the local geology (Table 4).

[Figure 10 and 11 about here]

3.3. Enrichment factor or phosphorus degree of contamination

The variability coefficient (CV) of the reference elements was measured as: 26.2% (Al), 43.8% (Sc), 26.7% (Ti) and 34.2% (Zr - Table 3). Aluminium displayed the lowest value of CV, followed closely by Ti, Zr and Sc. Al and Ti are mostly related to processes forming and presence of oxides (Al_2O_3 and TiO_2) in clastic materials and are not easily affected by weathering processes. These results prove that Al remain the most stable element and was therefore chosen as reference variable to determine the P enrichment factor for this study too.

The enrichment factor scores were calculated for P using both the Al elemental concentration in continental crust (Fig. 12A), and the local Al background concentration in the survey areas (Fig. 12B).

[Figure 12 about here]

The reference with the continental crust value (Fig. 12A) showed a lower EF score (EF <2) in the northern and southern parts of the region, with medium EF scores (ranging from 2 to 4) comprising more than 50% of the study area. Higher EF scores (ranging from 20 to 40), corresponding to anomalous enrichments, were found in soil on the slopes of the Mt. Somma-Vesuvius and in the Lioni and Laviano districts. Vesuvian areas (Fig. 9) displayed higher concentration of P which can though be related to the underlying background concentration, where the enrichment can be explained by geogenic (volcanic) source. Lioni and Laviano districts lie in areas characterised by limestone-derived soils; a high P enrichment factor could therefore be linked to anthropogenic activities. The highest range of P enrichment factor scores (> 40) is registered in the provincial territory of Benevento (particularly toward the Vitulano municipality). Soils in this area are derived from sandstone, flysch deposits and limestone, where intensive agriculture activities, such as vineyards and olive plantations, are practiced. The extremely high P enrichment may be related to the use of phosphate fertilizers in agriculture practises in this area.

The calculation of P enrichment factors with local background concentrations (Fig. 12B) presented different distribution and intensity of EF scores in the study area. The Vesuvian areas presenting previously significant enrichment (EF ranging from 5 to 20, Fig 12A) were now presenting a minimum enrichment instead (EF < 2, Fig. 12B). The Vesuvian area are characterised by a high P background concentration related to the underlying parental volcanic rocks which represent materials for the subsequently formed soils throughout this area. Furthermore, Al was found at lower concentration values compared to its value in the continental crust (Table 4). Indeed, EF calculation using the P and Al local background concentrations provides a smoother EF in relation to the Vesuvian area compared to their actual local variabilities.

[Table 4 about here]

A similar result was observed in the area surrounding Lioni, Laviano and Vitulano municipalities. EF calculated in the classic way (Fig. 12A) showed a very high (EF ranging 20 to 40) and extremely high enrichment (EF > 40) as opposed to the moderate (EF ranging from 2 to 5) and significant enrichment (EF ranging from 5 to 20) shown by EF calculated using the local background (Fig. 12B). Using the local background concentration of Al and P (Fig. 12B), the widespread enrichment created by calculation made using the continental crust values as a reference fades away, replaced by a much narrower enrichment that take into account the local variability of P and Al, and highlights the anthropogenic P inputs observed around the Lioni and Laviano districts, and even more significantly in the Vitulano municipality where it may be related to the use of phosphate fertilizers in agriculture activities. A continental crust concentration based EF 'hides' the anthropogenic input behind a more general enrichment due to the fact that it does not consider the local geological variability (Reimann and de Caritat, 2005; Albanese et al., 2013; Zuzolo et al., 2017). However, when taking into account the local concentration, the anthropogenic input can be more clearly distinguished from the geogenic enrichment factor. In other words, by using as a reference the local background, it is easier to isolate anthropogenic enrichment factors from the geogenic ones.

4. Conclusions

Evidence from this study showed that compositional data transformations such as ilr transformation can help to solve the outlier artefacts and moves the composition sample space to the Euclidean Real space R^{-1}, which is failed by the classic statistical data transformation (Filzmoser and Hron, 2008).

The multivariate and integrated approach applied in this study on a multi-elemental geochemical dataset allowed to highlight the correlation between variables and helped identifying the main sources of elements in the surveyed region. Biplot based on transformed data were able to highlight the main geological features as well as the potential input of anthropogenic activities in the Campania region. In particular, four association complexes were identified:

1) Fe, Mn, Co and Ni: associated to the coprecipitation of Mn and Fe oxides-hydroxides in flysch and arenaceous material;
2) Ca and Mg: associated to limestones and dolostone outcrops occurrence of the Mt. Picentini, Mt Lattari and Mt. Cervati limestones;
3) Al, Th, As, V, Ti, and Mo: associated to the pyroclastic coverings;
4) Cu and P: potentially associated to the agriculture activities through use of phosphate fertilizers.

The robust biplots allowed to display elements in a wider space and provided grounds for an enhanced data interpretation: the length of the rays, which is linked to the variability of ilr data clr back transformed as opposed to the variables themselves, permitted to emphasize their potential sources; the groups of elements highlighted by proximal rays, were used as evidence of either local geological features and/or anthropogenic activities.

In addition, the interpolated maps by use of C-A fractal plot helped to distinguish element distributions related to their main sources , whereas using S-A multifractal plot allowed to display the background concentration Al, Na, K and P elements as reference values for Campania soils.

The maps of P Enrichment Factor scores using as reference the continental crust and the local background values showed:

1) A significant P enrichment (from 5 to 20) in soils on the slopes of the Mt. Somma-Vesuvius related to the underlying parental volcanic rocks (geology), where the geogenic component represent clearly the natural background. This enrichment disappeared when the local background reference of Al and P are used. In fact, due to the geogenic source of the P and Al in these areas, the EF tended to decrease when using the local background, confirming that those areas were not contaminated by human activities.
2) The highest range of P EF factor scores (> 40) were registered in the Benevento provincial territory (particularly in Vitulano district), Lioni and Laviano districts, with values decreasing two times when using as reference the local background values. Taking into account that the underling geology of these areas could not influence these high P EFs, it is very likely that these values may have an anthropogenic source. The above provincial districts fall in a territory where intense agriculture activities are present, allowing to infer that such soils may be affected by use of phosphate fertilizers.

As a general observation, the findings from this study confirm the validity of using local background concentrations to better identify the ‘real’ degree of contamination as opposed to generalised continental crust values, which could emphasise ‘spurious’ enrichment due mainly to natural local concentrations of elements.

From an applied point of view, the integrated approach applied here provided a more robust qualitative and quantitative evaluation, highlighting new and vital information on the distribution and patterns of key elements (Na, K, Al and P) in soils of the Campania region. The findings from this investigation strongly point towards highly desirable follow up studies in clearly identified and discrete areas displaying high P EFs: this would allow a more detailed and thorough assessment of P footprints, which could then be used for a comprehensive human health risk evaluation due to direct and indirect exposure. This is particularly important, when considering that, if P was linked to fertilizers source, these could include impurities such as Cu, As, Zn and Pb, well-known potentially toxic elements (PTEs).

Acknowledgments

This work was supported by grant from the Ministero dell'Università e della Ricerca Scientifica - Industrial Research Project “BioPoliS” PON03PE_00107_01, funded in the frame of Operative National Programme Research and Competitiveness 2007-2013, D. D. Prot. N.713/Ric. 29/10/2010 and (EnerbioChem) PON01_01966, funded in the frame of the Operative National Programme Research and Competitiveness 2007–2013 D. D. Prot. n. 01/Ric. 18.1.2010. The authors would like to thank the anonymous reviewers for their helpful comments and suggestions.

References

Abrahim, G. M. S., Parker, P. J. 2008. Assessment of heavy metal enrichment factors and the degree of contamination in marine sediments from Tamaki Estuary, Auckland, New Zealand, Environmental Monitoring and Assessment, Vol.136, No.1-3, 227–238.

Aitchison, J. 1986. The Statistical Analysis of Compositional Data. Chapman & Hall, London (416 pp).

Albanese, S., Iavazzo, P., Adamo P, Lima A, De Vivo B., 2013. Assessment of the environmental conditions of the Sarno River Basin (south Italy): a stream sediment approach. Environ Geochem Health 35:283–297.

Albanese, S., De Vivo, S., Lima, A., Cicchella, D., 2007. Geochemical background and baseline values of toxic elements in stream sediments of Campania region (Italy). J. Geochem. Explor. 93, 21–34.

Ander, E. L., Johnson, C.C., Cave, M.R., Palumbo-Roe, B., Nathanail, C.P., Lark R.M., 2013. Methodology for the determination of normal background concentrations of contaminants in English soil. Science of the Total Environment; 454-455, 604–618. http://dx.doi.org/10.1016/j.scitotenv.2013.03.005.

APAT-ISS. 2006. Protocollo Operativo per la determinazione dei valori di fondo di metalli/metalloidi nei suoli dei siti d'interesse nazionale. Revisione 0. Agenzia per la Protezione dell'Ambiente e per i Servizi Tecnici and Istituto Superiore di Sanita; [in Italian].

Bonardi, G., D'Argenio, D., Perrone, V., 1998. Carta geologica dell'Appennino meridionale. Memorie della Societa Geologica Italiana 41.

Buat-Ménard P, Chesselet R., 1979. Variable influence of the atmospheric flux on the trace metal chemistry of oceanic suspended matter. Earth and Planetary Science Letters; 42, 399–411.

Buccianti, A., Lima, A., Albanese, S., Cannatelli, C., Esposito, R., De Vivo, B., 2015. Exploring topsoil geochemistry from the CoDA (Compositional Data Analysis) perspective: the multi-element data archive of the Campania Region (Southern Italy). J. Geochem. Explor. 159, 302–316.

Buccianti, A., Magli, R., 2011. Metric concepts and implications in describing compositional changes for world river’s water chemistry. Comput. Geosci. 37 (5), 670–676.

Buccianti, A., Mateu-Figueras, G., Pawlowsky-Glahn, V. (ed.), 2006. Compositional data analysis in the Geosciences: From theory to practice.

Buccianti, A., Pawlowsky_Glahn, V., Egoscue, J.J., 2014. Variation diagrams to statistically model the behavior of geochemical variables: Theory and applications. Journal of Hydrogeology, 519, 988-998.

Buccianti A., Lima A., Albanese S. and De Vivo B., 2018. Measuring the change under Compositional Data Analysis (CoDA): insight on geochemical system dynamics. J. Geochem Explor. 189, 100-108. https://doi.org/10.1016/j.gexplo.2017.05.006.

Buccianti, A., Lima, A., Albanese, S., Cannatelli, C., Esposito, R., De Vivo, B., 2015. Exploring topsoil geochemistry from the CoDA (Compositional Data Analysis) perspective: the multi-element data archive of the Campania Region (Southern Italy). J. Geochem. Explor. 159, 302–316.

Cave M.R, Johnson C.C, Ander E.L, Palumbo-Roe B., 2012. Methodology for the determination of normal background contaminant concentrations in English soils. British Geological Survey commissioned report, CR/12/003; 2012 [41 pp. http://nora.nerc.ac.uk/19959/].

Cheng, Q., 1999. Spatial and scaling modelling for geochemical anomaly separation. J. Geochem. Explor. 65, 175–194.

Cheng, Q., 2007. Mapping singularities with stream sediment geochemical data for prediction of undiscovered mineral deposits in Gejiu, Yunnan Province, China. Ore Geol. Rev. 32, 314–324.

Cheng, Q., Agterberg, F.P., Ballantyne, S.B., 1994. The separation of geochemical anomalies from background by fractal methods. J. Geochem. Explor. 51, 109–130.

Cheng, Q., Bonham-Carter, G.F., Raines, G.L., 2001. GeoDAS: A new GIS system for spatial analysis of geochemical data sets for mineral exploration and environmental assessment. The 20th Intern. Geochem. Explor. Symposium (IGES). Santiago de Chile, 6/5-10/5, 2001, pp. 42–43.

Cheng, Q., Xia, Q., Li, W., Zhang, S., Chen, Z., Zuo, R., Wang, W., 2010. Density/area power-law models for separating multi-scale anomalies of ore and toxic elements in stream sediments in Gejiu mineral district, Yunnan Province, China. Biogeosciences 7, 3019–3025.

Cheng, Q., Xu, Y., Grunsky, E., 2000. Integrated spatial and spectrum method for geochemical anomaly separation. Nat. Resour. Res. 9, 43–51.

Chester, R., Stoner, JH., 1973. Pb in particulates from the lower atmosphere of the eastern Atlantic. Nature; 245:27– 8.

Cicchella, D., De Vivo, B., Lima, A., 2005. Background and baseline concentration values of elements harmful to human health in the volcanic soils of the metropolitan and provincial area of Napoli (Italy). Geochem: Explor. Environ. Anal. 5, 29–40.

De Vivo, B., Ayuso, R.A., Belkin, H.E., Fedele, L., Lima, A., Rolandi, G., Somma, R., Webster, J.D., 2003. Chemistry, fluid/melt inclusions and isotopic data of lavas, tephra and nodules from 25 ka to 1944 AD of the Mt. Somma-Vesuvius volcanic activity. Mt. Somma-Vesuvius Geochemical Archive. Dipartimento di Geofisica e Vulcanologia, Università di Napoli Federico II, Open File Report 1-2003, 143 pp.

De Vivo, B., Lima, A., Albanese, S., Cicchella, D., Rezza, C., Civitillo, D., Minolfi, G., Zuzolo, D., 2016. Atlante geochimico-ambientale dei suoli della Campania (Environmental Geochemical Atlas of Campania Soils). Aracne Editrice, Roma, pp. 364. ISBN 978-88- 548-9744-1. [in Italian].

De Vivo, B., Petrosino, P., Lima, A., Rolandi, G., Belkin, H.E., 2010. Research progress in volcanology in Neapolitan area, Southern Italy: a review and alternative views. Mineral. Petrol. 99,1–28. Doi: 10.1007/s00710.009.0098.6.

De Vivo, B., Rolandi, G., Gans, P.B., Calvert, A., Bohrson, W.A., Spera, F.J., Belkin, H.E., 2001. New constraints on the pyroclastic eruptive history of the Campanian volcanic Plain (Italy). Mineralogy and Petrology 73, 47–65.

Duce R.A, Hoffmann G.L, Zoller W.H., 1975. Atmospheric trace metals at remote northern and southern hemisphere sites: pollution or natural? Science, 187:59– 61.

D. Ducci, Tranfaglia G., 2005. L'impatto dei cambiamenti climatici sulle risorse idriche sotteranee in Campania. Boll. Ordine Geologi della Campania, 1-4, 13-21; [in Italian].

Egli, M., Mirabella, A. and Sartori, G., 2008. The role of climate and vegetation in weathering and clay mineral formation in late Quaternary soils of the Swiss and Italian Alps. Geomorphology, 102, 307-324, doi: 10.1016/j.geomorph.2008.04.001.

Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Barcelo-Vidal, C., 2003. Isometric logratio transformations for compositional data analysis. Math Geol., 35(3), 279-30.

Egozcue, J.J., Pawlowsky-Glahn, V., (2005) Groups of parts and their balances in compositional data analysis. Mathematical Geology 37(7):795−828.

ESRI 2012. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.

Filzmoser, P., Hron, K., 2008. Outlier detection for compositional data using robust methods. Math. Geosci. 40 (3), 233–248.

Filzmoser, P., Hron, K., Reimann, C., 2009a. Principal component analysis for compositional data with outliers. Environmetrics, 20 (6), 621–632.

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

Gabriel, K.R., 1971. The biplot graphic display of matrices with application to principal component analysis. Biometrika 58 (3), 453.

Goldberg, E.D, (Edt). 1972. Baseline studies of pollutants in the marine environment and research recommendations, The International Decade of Ocean Exploration (IDOE) Baseline Conference, 24– 26 May 1972. New York.

Hakanson, L., 1980. An ecological risk index for aquatic pollution control: a sedimentological approach. Water Research,14, 975–1001.

Han, J., Kamber, M., 2001. Data Mining: Concepts and Techniques. Morgan-Kaufmann Academic Press, San Francisco.

Hron, K., Templ, M., Filzmoser, P., 2010. Imputation of missing values for compositional data using classical and robust methods. Comput. Stat. Data Anal., 54 (12), 3095–3107.

ISTAT, 2016. Resident Population on 1st January, 2017. https://www.istat.it/en/population-and-households.

Joron, J.L., Metrich, N., Rosi, M., Santocroce, R., Sbrana, A., 1987. Chemistry and petrography. In: Santacorce, R. (Ed.), Somma-Vesuvius. CNR Quad. Ric. Sci. 114, 105–174.

Lima, A., De Vivo, B., Cicchella, D., Cortini, M., Albanese, S., 2003a. Multifractal IDW interpolation and fractal filtering method in environmental studies: an application on regional stream sediments of Campania Region (Italy). Appl. Geochem. 18 (12), 1853-1865. Doi: 10.1016/S0883 2927(03)00083-0.

Lima, A., Danyushevsky, L.V., De Vivo, B., Fedele, L., 2003b. A model for the evolution of the Mt. Somma-Vesuvius magmatic system based on fluid and melt inclusion investigations. In: De Vivo, B., Bodnar, R.J. (Eds), Melt Inclusions in Volcanic Systems: Methods, Applications and Problems. Series: Development in Volcanology. Elsevier, Amsterdam, 272 pp.

Loska, K., Cebula, J., Pelczar, J., Wiechula, D., Kwapilinski, J., 1997. Use of enrichment and contamination factors together with geoaccumulation indexes to evaluate the content of Cd, Cu, and Ni in the Rybnik water reservoir in Poland. Water, Air, and Soil Pollution, 93, 347-65.

Mandelbrot, B.B., 1983. The Fractal Geometry of Nature. Freeman, San Francisco, pp. 468.

Maronna, R., Martin, R., Yohai, V., 2006. Robust Statistics: Theory and Methods. John Wiley pp. 436. ISBN: 978-0-470-01092-1.

Martin, J.M, Whitfield M., 1983. The significance of the river input of chemical elements to the ocean. In: Wong CS, Boyle E, Bruland KW, Burton JD, Goldberg E eds) Trace metals in sea water. Plenum, New York, pp 265-296.

Minolfi, G., Albanese, S., Lima, A., Tarvainen, T., Fortelli, A., De Vivo, B., 2018. A regional approach to the environmental risk assessment - human health risk assessment: case study in the Campania region. J. Geochem. Explor.184, Part B, 400-416. .

Mondillo, N., Boni, M., Balassone, G., and Rollinson, G. 2011. Karst bauxites in the Campania Apennines (southern Italy): A new approach, Periodico di Mineralogia, 80, p. 407-432.

Nziguheba G., Smolders E., 2008. Inputs of trace elements in agricultural soils via phosphate fertilizers in European countries. Sci Total Environ, 390, 53-57.

Olea, R.A., Raju, N.J., Egozcue, J.J., Pawlowsky-Glahn, V., Sing, S. 2018. Advancements in hydrochemistry mapping: Methods and application to groundwater arsenic and iron concentrations in Varanasi, Uttar Pradesh, India. Stochastic Environmental Research and Risk Assessment, vol. 32, no. 1, p. 241−259.

Otero, N., Tolosana-Delgado, R., Solera., A., Pawlowsky-Glahn, V., Canals, A., 2005. Relative vs. absolute statistical analysis of compositions: A comparative study of surface waters of a Mediterranean river. Water Res, 39, 1404–1414.

Pawlowsky-Glahn, V. Egozcue, J. J. 2001. Geometric approach to statistical analysis on the simplex. Stochastic Environmental Research and Risk Assessment (SERRA), 15 (5), 384-398.

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

Peccerillo, A., 2005. Plio-quaternary volcanism in Italy. In: Petrology, Geochemistry, Geodynamics. Springer-Verlag, Berlin Heidelberg, pp. 365, ISBN 978-3-540- 29,092-6.

Peirson D.H., Cawse P.A., Cambray R.S., 1974. Chemical uniformity of airborne particulate material, and a maritime effect. Nature, 251, 675– 9.

Pison, G., Rousseeuw, P.J., Filzmoser, P., Croux, C. 2003. Robust factor analysis. J. Multivar. Anal., 84, 145–172.

Rahn, K.A., 1976. The chemical composition of the atmospheric aerosol. Technical Report of the Graduate School of Oceanography, University of Rhode Island, Kingston, R.I., USA.

Reimann, C., Garrett, R.G., Filzmoser, P., 2005. Background and threshold - critical comparison of methods of determination. Sci Total Environ. 346, 1-16.

Reimann, C., De Caritat, P., 2005. Distinguishing between natural and anthropogenic sources for elements in the environment: regional geochemical surveys versus enrichment factors. Sci Total Environ., 337, 91–107.

Reimann, C., Filzmoser, P., Garrett, R., 2002. Factor analysis applied to regional geochemical data: problems and possibilities. Appl. Geochem., 17 (3), 185-206.

Reimann C., Birke M., Demetriades A., Filzmoser P., O'Connor P. (Editors) and GEMAS Team, (2014). Chemistry of Europe's agricultural soils — part A: methodology and interpretation of the GEMAS data set. Geologisches Jahrbuch (Reihe B), Schweizerbarth: Hannover, pp. 528.

Reimann, C., Filzmoser, P., Garrett, R.G., Dutter, R., 2008. Statistical data analysis explained. In: Applied Environmental Statistics with R. Chichester. Wiley, pp. 362. ISBN: 978-0-470-98581-6.

Rolandi, G., Bellucci, F., Heizler, M.T., Belkin, H.E., De Vivo, B., 2003. Tectonic controls on genesis of ignimbrites from the Campanian Volcanic Zone, Southern Italy. In: De Vivo, B., Scandone, R. (Eds.), Ignimbrites of the Campania Plain, Italy. Mineral.Petrol., 79, 3-31.

Rousseeuw, P.J., Van Driessen, K., 1999. A fast algorithm for the minimum covariance determinant estimator. Technometrics, 41, 212–223.

Saeedi, M., Li, L.Y., Salmanzadeh, M. 2012. Heavy metals and polycyclic aromatic hydro-carbons: pollution and ecological risk assessment in street dust of Tehran. J. Hazard. Mater., 227-228, 9–17.

Salminen, R., Gregorauskiene, V., 2000. Considerations regarding the definition of a geochemical baseline of elements in the surficial materials in areas differing in basic geology. Appl. Geochem., 15, 647–653.

Scheib, A.J., Birke, M., Dinelli, E., The Gemas Project Team, 2014. Geochemical evidence of aeolian deposits in European soils. Boreas 43 (1), 175–192. http://dx.doi.org/10. 1111/bor.12029.

Sutherland, R. A., Tolosa, C. A., Tack, F. M. G., Verloo. M. G., 2000. Characterization of selected element concentration and enrichment ratios in background and anthropogenically impacted roadside areas, Archives of Environmental Contamination and Toxicology, 38, 428–438.

Tarvainen, T., Jarva, J., 2011. Using geochemical baselines in the assessment of soil contamination in Finland. In: Johnson C.C, Demetriades A, Locutura J, Ottesen R.T, eds. Mapping the Chemical Environment of Urban Areas. Chichester, UK: John Wiley & Sons Ltd; p. 223–31.

Taylor, S.R., McLennan, S.M., 1995. The geochemical evolution of the continental crust. Reviews of Geophysics, 33, 241– 65.

Thiombane, M., Zuzolo, D., Cicchella, D., Albanese, S., Lima, A., Cavaliere, M., De Vivo, B. 2018a. Soil geochemical follow-up in the Cilento World Heritage Park (Campania, Italy) through exploratory compositional data analysis and C-A fractal model. J. Geochem. Explor. 189, 85–99.

Thiombane, M., Martin-Fernandez, J.A., Albanese, S., Lima, A., Doherti, A., De Vivo, B. 2018b. Exploratory analysis of multi-element geochemical patterns in soil from the Sarno River Basin (Campania region, southern Italy) through compositional data analysis (CODA). J. Geochem. Explor. https://doi.org/10.1016/j.gexplo.2018.03.010.

Tóth, G., Guicharnaud, R-A., Tóth, B., Hermann, T., 2014. Phosphorus levels in croplands of the European Union with implications for P fertilizer use, European Journal of Agronomy, 55, 42-52. ISSN 1161-0301, Doi 10.1016/j.eja.2013.12.008.

U.S. Environmental Protection Agency (EPA), 2001. Guidance for Characterizing Background Chemicals in Soil at Superfund Sites, External Review Draft, Office of Emergency and Remedial Response, OSWER. 9285.7-41. [Replaced by Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites, EPA 540-R-01-003, September 2002.].

Wedepohl K. H., 1995. The composition of the continental crust. Geochimica et Cosmochimica Acta, 59, 1217–32.

Wu, S., Zhou, S., Li, X., 2011. Determining the anthropogenic contribution of heavy metal accumulations around a typical industrial town: Xushe, China. J. Geochem. Explor. 110, 92–97.

Zuo, R., 2011. Decomposing of mixed pattern of arsenic using fractal model in Gangdese belt, Tibet, China. Appl. Geochem. 26, S271–S273.

Zuo, R., Wang, J., Chen, G., Yang, M., 2015. Identification of weak anomalies: a multifractal perspective. J. Geochem. Explor., 148, 12–24.

Zuo, R., Wang, J., 2016. Fractal/multifractal modelling of geochemical data: a review. J. Geochem. Explor., 164, 33–41.

Zuzolo, D., Cicchella, D., Catani, V., Giaccio, L., Guagliardi, I., Esposito, L., De Vivo, B. (2017). Assessment of potentially harmful elements pollution in the Calore River basin (southern Italy). Environ Geochem Health 39:531-548. DOI 10.1007/s10653-016-9832-2.

Figure and Table captions

Figures

Figure 1. . Simplified geological map (A) and land use (B) of the Campania region (southern of Italy).

Figure 2. Edaplot (combination of histogram, density trace, one-dimensional scattergram and Boxplot in just one display) and CP plot of Na of the raw data, log-transformed data, and ilr transformed data.

Figure 3. Edaplot (combination of histogram, density trace, one-dimensional scattergram and Boxplot in just one display) and CP plot of K through the raw data, log-transformed data, and ilr transformed data.

Figure 4. Edaplot (combination of histogram, density trace, one-dimensional scattergram and Boxplot in just one display) and CP plot of P of the raw data, log-transformed data, and ilr transformed data.

Figure 5. Biplots for first and second principal components of factor analysis for raw data (classical, left plot) and ilr coordinates clr back-transformed (robust, right plot) of the survey area (Campania region, Southern Italy).

Figure 6. Flow chart of data processing for contamination degree modelling using GIS environment

Fig. 7. (A) The interpolated Z1 map. Note the red and blue colours highlight higher and lower proportion of Al and P, respectively. (B) The interpolated Z2 map. The red and blue colours highlight higher and lower proportion of Na and K, respectively.

Figure 8. Interpolated maps of Al (A) and P (B) elemental distribution in the survey area; ranges of concentration are based on the C-A fractal plot held bellow.

Figure 9. Background/baseline maps of Al (A) and P (B) elemental concentration in the survey area.

Figure 10. Interpolated maps of Na (A) and K (B) elemental distribution in the survey area; ranges of concentration are based on the C-A fractal plot held bellow.

Figure 11. Background/baseline maps of Na (A) and K (B) elemental concentration in the survey area.

Figure 12. Interpolated maps of the enrichment factor scores into the study areas; (A) map created using the continental crust reference values of Al and P (Wedepohl, 1995); (B) map is based on the local reference background concentrations of Al and P (this study) as reference elements in each pixel.

Tables

Table 1. Descriptive statistic of 3669 topsoils samples from the Campania region, Southern of Italy. RMS and Std. Deviation are the root mean square and standard deviation, respectively.

Table 2. Sequential binary partition table of the 18 investigated variables and the obtained balances (Z_1-Z_{17}). Parts coded with + and - are the elemental associations involved in the calculation of the i-th order partition, respectively.

Table 3. Variability Test of Al, Sc, Ti and Zr elements; MD is the Median, MAD corresponds to median absolute deviation, MAD= median {|xi–median (xi)|}, that is the median value (50th percentile) of the deviations of all individual xi values (concentrations) from the median value (concentration). CV= coefficient of variation.

Table 4. Geogenic background/Baseline value ranges for Al, P, Na and K elemental concentration in the continental crust (Wedepohl, 1995) and in soils of the Campania region according to the lithology.

Figure 1. Simplified geological map (A) and land use (B) of the Campania region (southern of Italy).

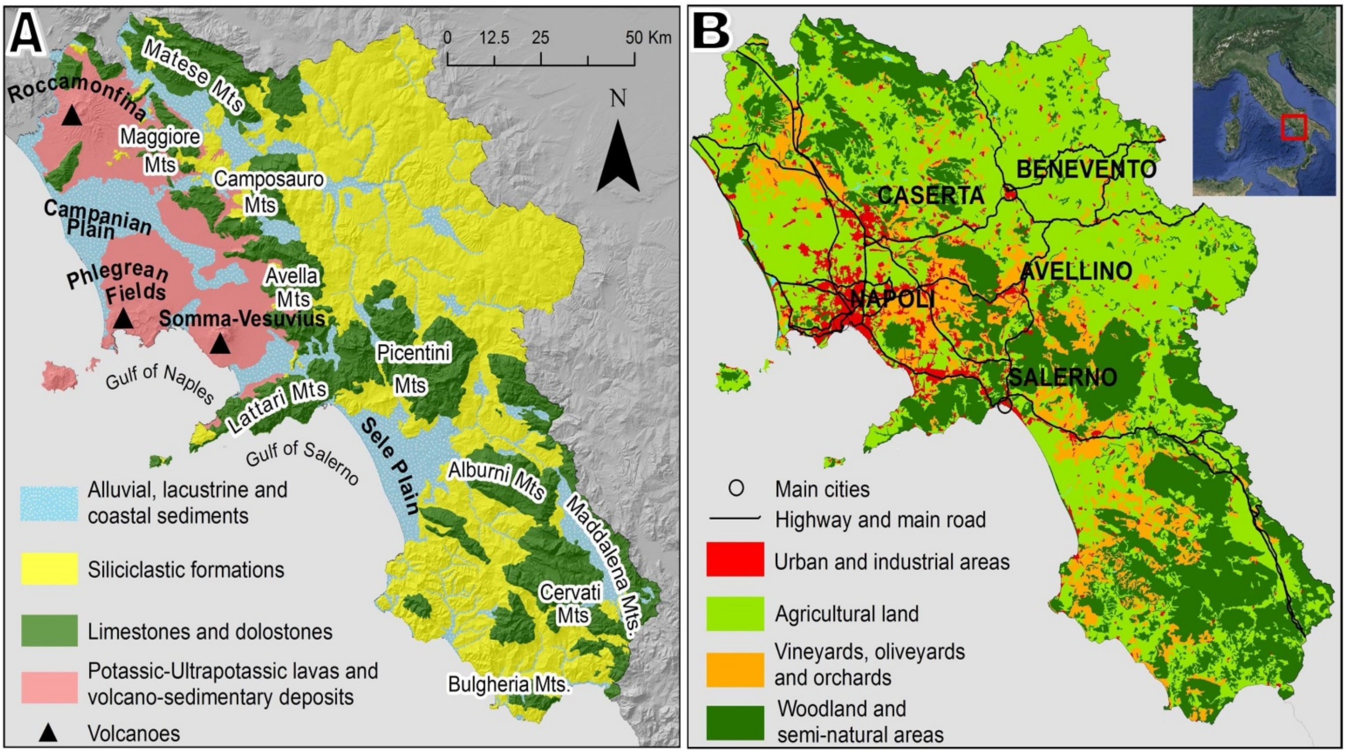

Figure 2. Edaplot (combination of histogram, density trace, one-dimensional scattergram and Boxplot in just one display) and CP plot of Na of the raw data, log-transformed data, and ilr transformed data.

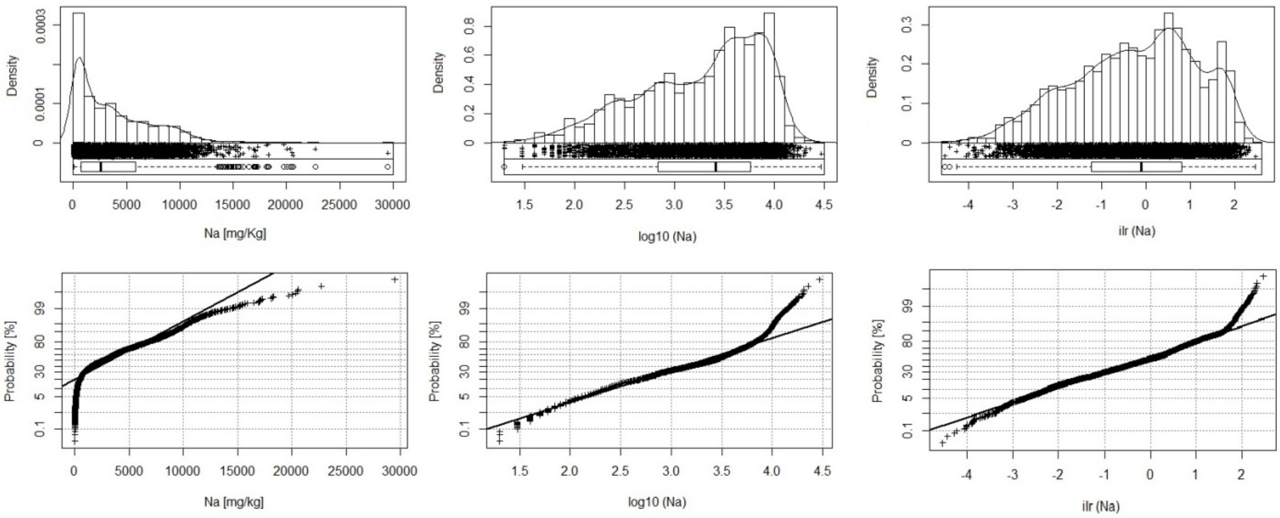

Figure 3. Edaplot (combination of histogram, density trace, one-dimensional scattergram and Boxplot in just one display) and CP plot of K through the raw data, log-transformed data, and ilr transformed data.

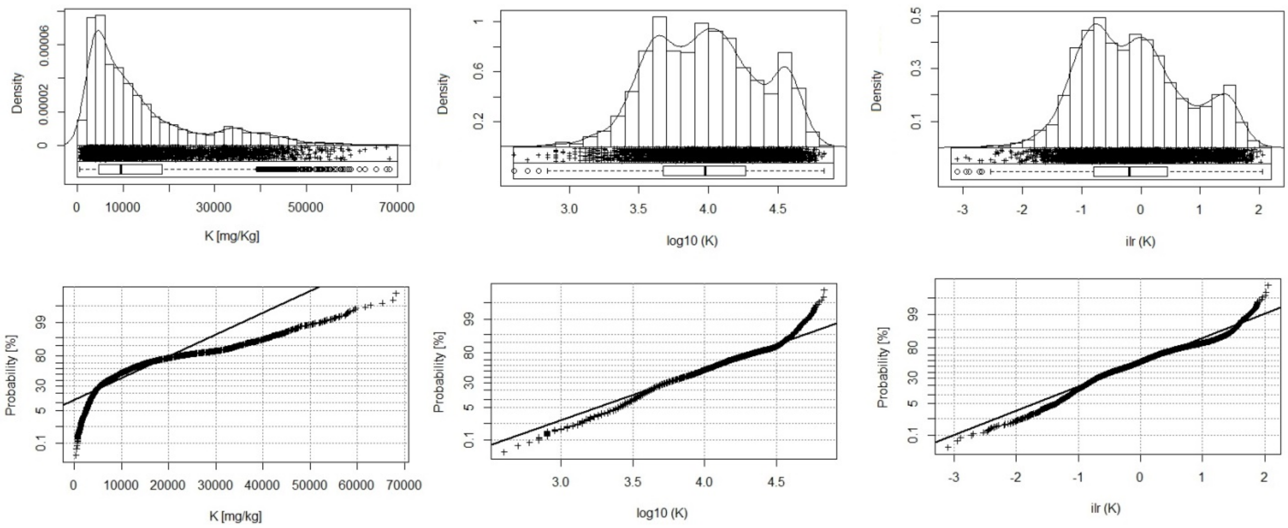

Figure 4. Edaplot (combination of histogram, density trace, one-dimensional scattergram and Boxplot in just one display) and CP plot of P of the raw data, log-transformed data, and ilr transformed data.

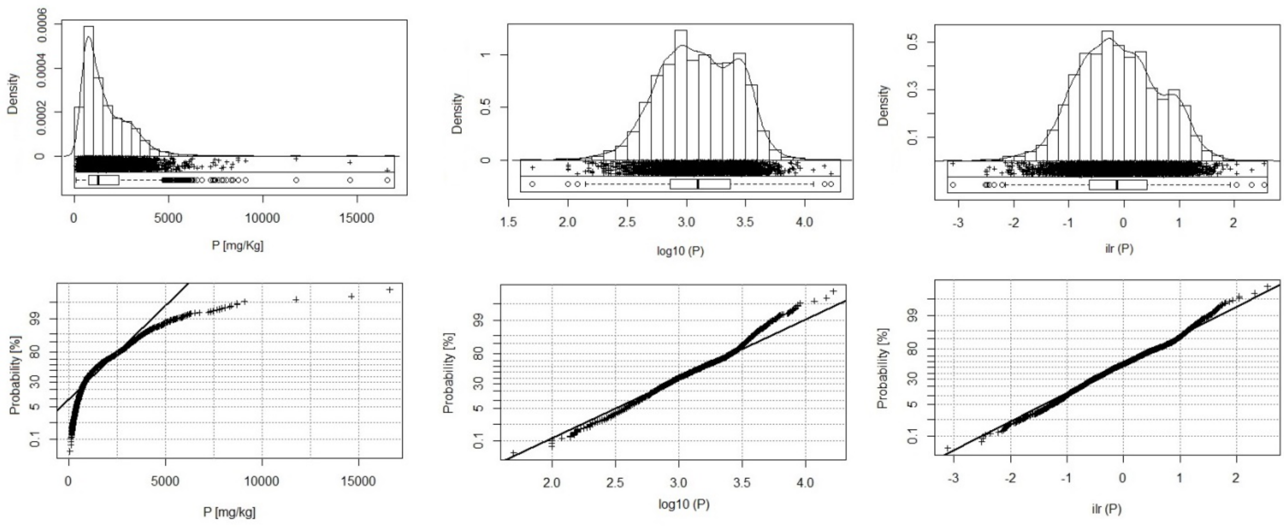

Figure 5. Biplots for first and second principal components of factor analysis for raw data (classical, left plot) and ilr coordinates clr back-transformed (robust, right plot) of the survey area (Campania region, Southern Italy).

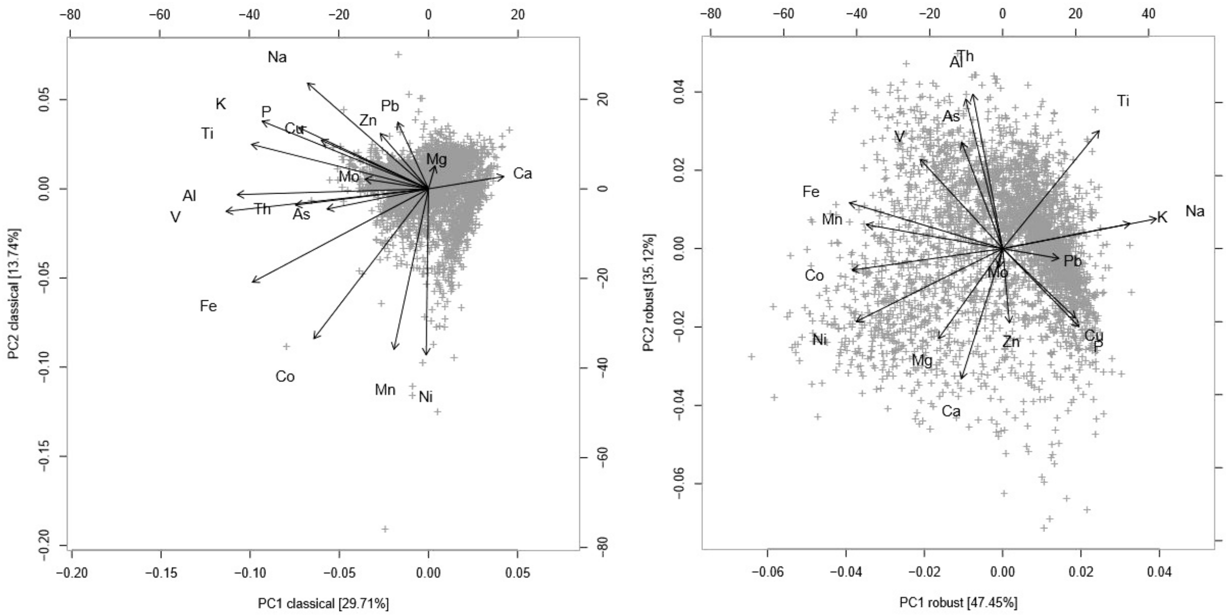

Figure 6. Flow chart of data processing for contamination degree modelling using GIS environment

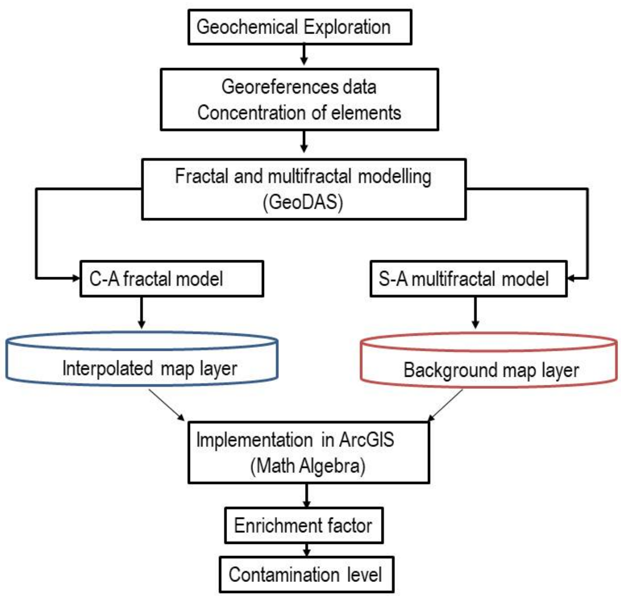

Fig. 7. (A) The interpolated Z_1 map. Note the red and blue colours highlight higher and lower proportion of Al and P, respectively. (B) The interpolated Z_2 map. The red and blue colours highlight higher and lower proportion of Na and K, respectively.

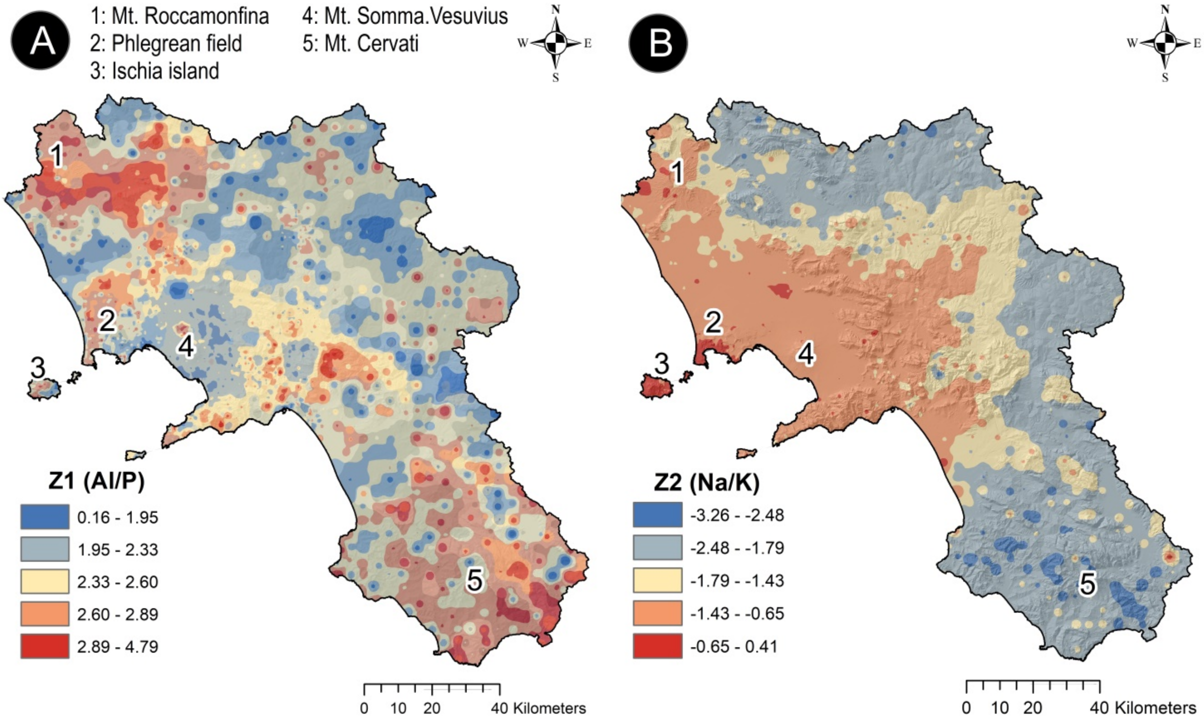

Figure 8. Interpolated maps of Al (A) and P (B) elemental distribution in the survey area; ranges of concentration are based on the C-A fractal plot held bellow.

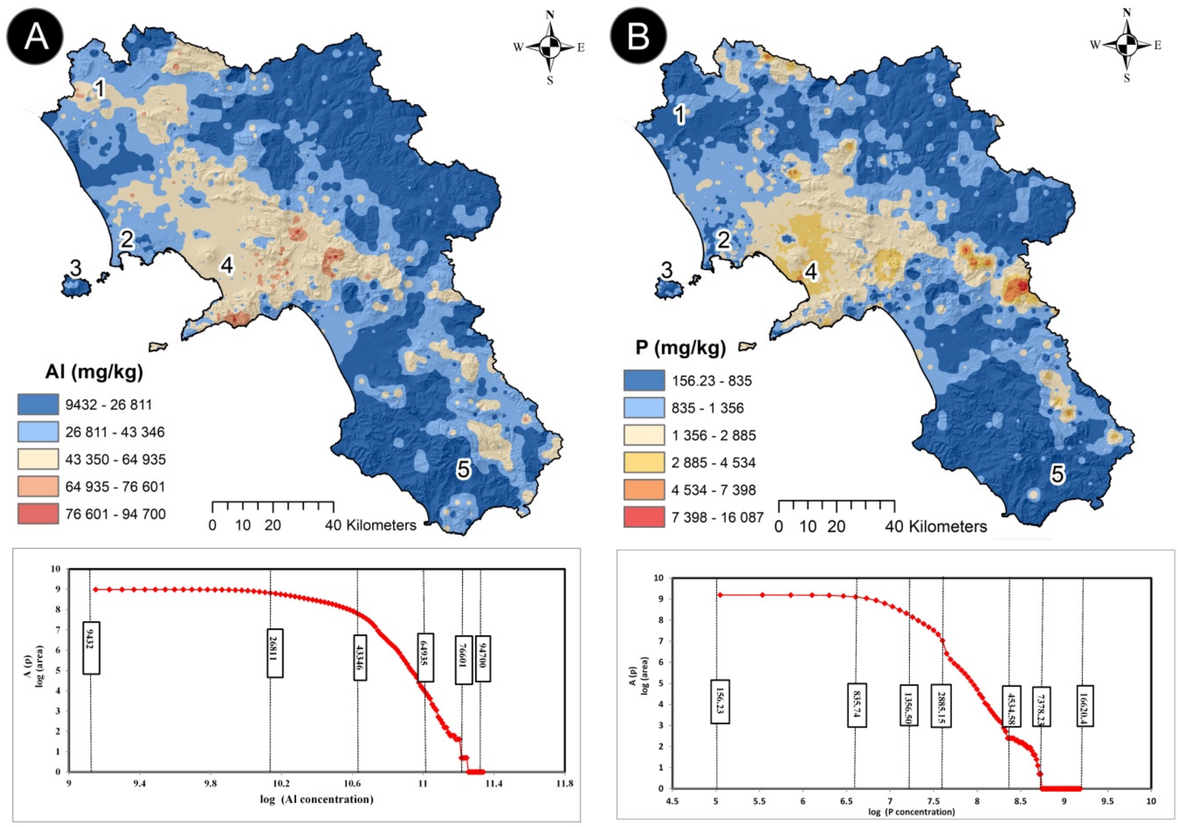

Figure 9. Background/baseline maps of Al (A) and P (B) elemental concentration in the survey area.

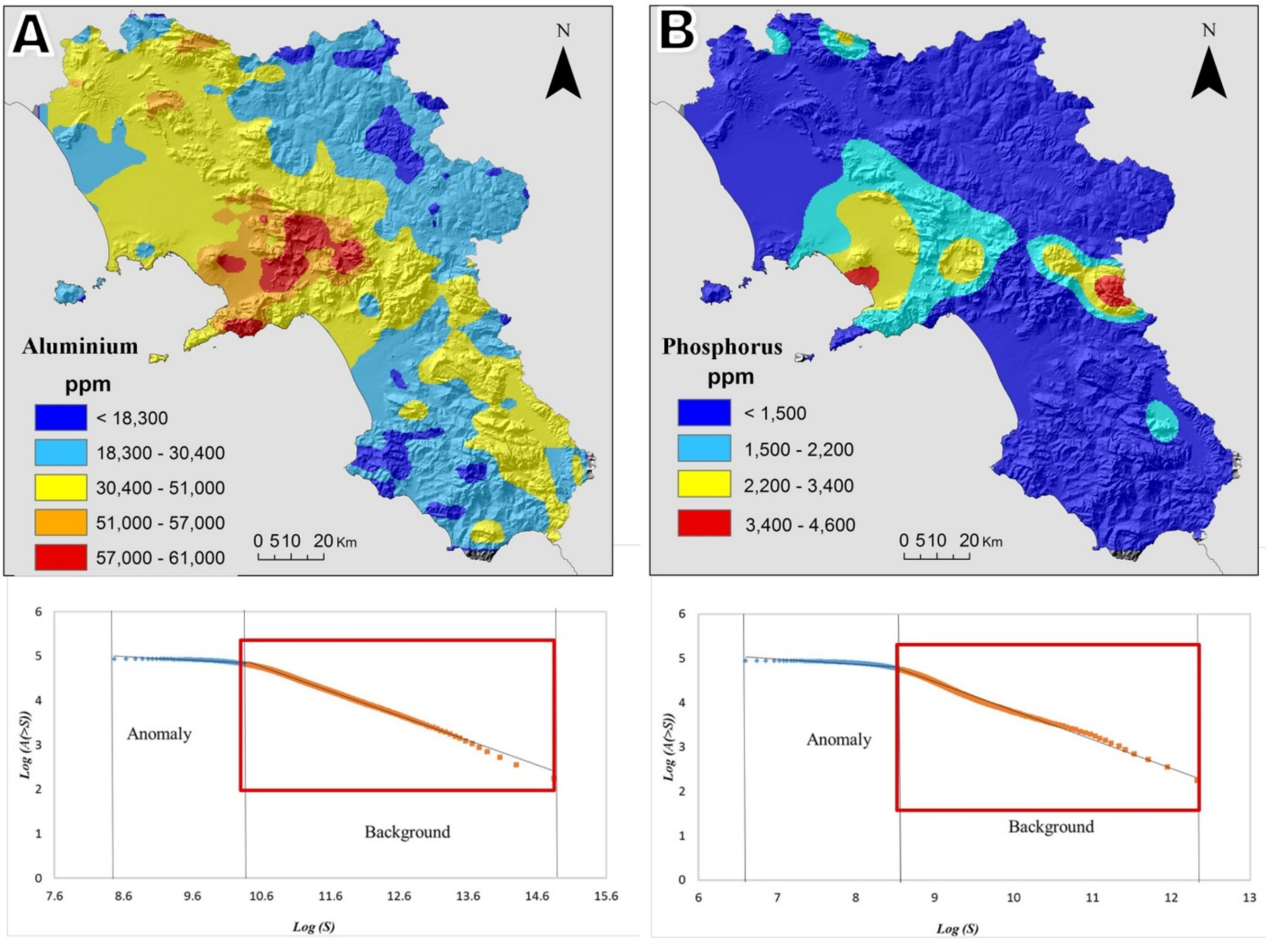

Figure 10. Interpolated maps of Na (A) and K (B) elemental distribution in the survey area; ranges of concentration are based on the C-A fractal plot held bellow.

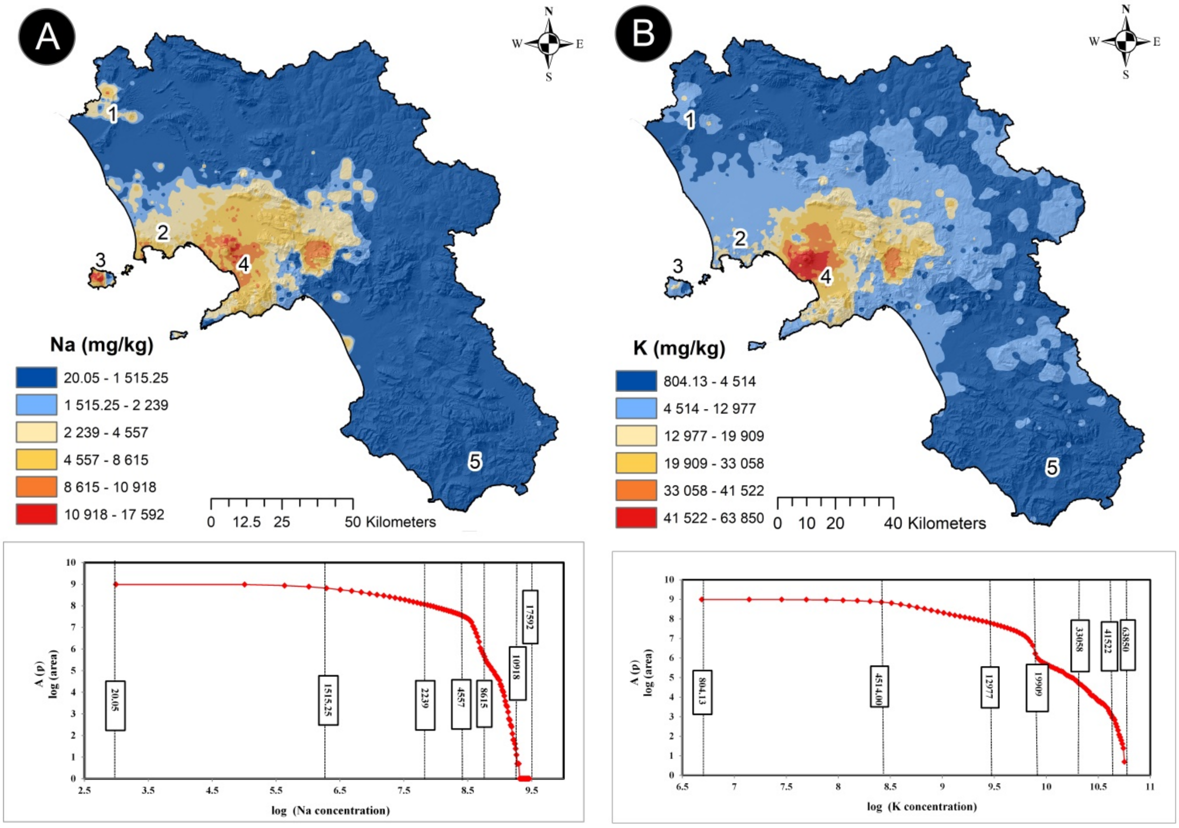

Figure 11. Background/baseline maps of Na (A) and K (B) elemental concentration in the survey area.

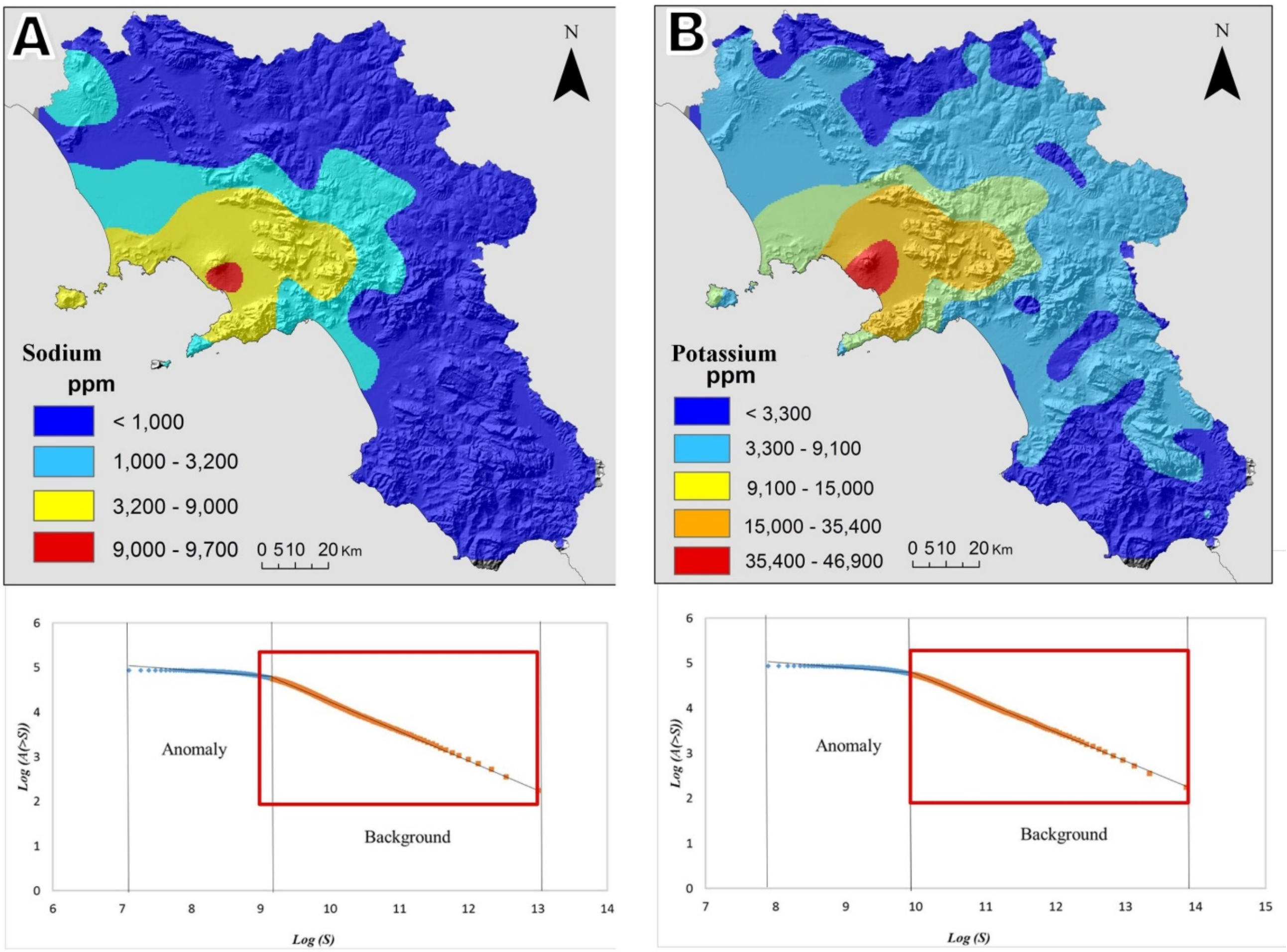

Table 1. Descriptive statistic of 3669 topsoils samples from the Campania region, Southern of Italy. RMS and Std. Deviation are the root mean square and standard deviation, respectively.

Elements	Unit	Minimum	Maximum	Mean	Median	RMS	Std. Deviation	Skewness	Kurtosis
Al	mg/kg	2100	94700	40584	41600	44014	17036	0.01	-0.75
As	mg/kg	0.6	163	12.6	12.1	14.6	7.4	6.1	87.8
Ca	mg/kg	800	295200	35495	22300	51036	36675	2.38	7.1
Co	mg/kg	0.5	79	10.7	10.3	11.7	5.1	1.6	12.3
Cu	mg/kg	2.5	2394	109.3	62.2	191	156.8	5.5	44.5
Fe	mg/kg	1600	154600	25031	2510	26282	8012	1.19	18.2
K	mg/kg	400	68200	14008	9500	18755	12472	1.39	1.2
Mg	mg/kg	700	104600	7347	5800	10483	7479	5.06	35.8
Mn	mg/kg	77	7975	863	779	970.6	443.4	5.44	55.7
Mo	mg/kg	0.06	62	1.5	1.3	2.2	1.6	18.9	637
Na	mg/kg	20	29490	3667	2600	5129	3587	1.23	1.68
Ni	mg/kg	0.5	100	16.2	14.8	19.6	10.9	2.42	9.97
P	mg/kg	50	16620	1641	1250	2063	1250	2.20	12.1
Pb	mg/kg	3.1	2052	73.8	54.2	117.7	91.7	7.7	100.2
Th	mg/kg	0.3	60	12.8	12.4	14.5	6.7	1.1	3.5
Ti	mg/kg	5	3270	1159	1240	1314	618.7	-0.23	-0.65
V	mg/kg	5	224	66.9	62	73.5	30.3	0.51	-0.29
Zn	mg/kg	11.4	3210	119	91	168.1	118.5	9.2	164.7

Table 2. Sequential binary partition table of the 18 investigated variables and the obtained balances (Z_1-Z_{17}). Parts coded with + and - are the elemental associations involved in the calculation of the i-th order partition, respectively.

Balances	Ti	Th	As	V	Al	P	Na	K	Mo	Cu	Zn	Pb	Ca	Mg	Fe	Mn	Co	Ni
Z_1					+	-												
Z_2							+	-										
Z_3					+	+	-	-										
Z_4	+	+	+	-	-													
Z_5				+	-													
Z_6	+	+	-															
Z_7	+	-																
Z_8						+	+	+	+	+	+	+	-	-				
Z_9									+	-								
Z_{10}											+	-						
Z_{11}										+	-							
Z_{12}								+	-									
Z_{13}													+	+	-	-	-	-
Z_{14}													+	-				
Z_{15}															+	+	-	-
Z_{16}															+	-		
Z_{17}																	+	-

Table 3. Variability Test of Al, Sc, Ti and Zr elements; MD is the Median, MAD corresponds to median absolute deviation, MAD= median {|xi–median (xi)|}, that is the median value (50th percentile) of the deviations of all individual xi values (concentrations) from the median value (concentration). CV= coefficient of variation.

Stat. Parameters	Al (ppm)	Sc (ppm	Ti (ppm)	Zr (ppm)
MD	41600	2.3	896	5.1
MAD	10900	1.01	240	1.7
CV (%)	26.2	43.8	26.7	34.2

Table 4. Geogenic background/Baseline value ranges for Al, P, Na and K elemental concentration in the continental crust (Wedepohl, 1995) and in soils of the Campania region according to the lithology.

Elements	Wedepohl, 1995*	Local background in this Study		
	Cont. Crust	Siliciclastic deposits	Limestone and Dolostone	Volcano-clastic deposits
Al (mg/kg)	150500	9432 - 30400	30400 - 51000	51000 - 61000
Na (mg/kg)	35600	20.4 - 1000	20.4 - 1000	1000 - 9700
K (mg/kg)	31900	804 - 9100	804 - 9100	9100 - 46900
P (mg/kg)	1500	156 - 1500	156 - 1500	1500 - 4600

* Average in continental crust. (Wedepohl, 1995)