The role of chemometrics in single and sequential extraction assays. A review. Part II. Cluster Analysis, Multiple Linear Regression, Mixture Resolution, Experimental Design and other techniques

(Article begins on next page)

Analytica Chimica Acta, Volume 688, issue 2, 04 Mar 2011,
doi: 10.1016/j.aca.2010.12.028.

The Role of Chemometrics in Single and Sequential Extraction Assays. A Review. Part II. Cluster Analysis, Multiple Linear Regression, Mixture Resolution, Experimental Design and Other Techniques

Agnese Giacomino, Ornella Abollino*, Mery Malandrino, , Edoardo Mentasti

Department of Analytical Chemistry, University of Torino, Via Giuria 5, 10125 Torino, Italy

Abstract

Single and sequential extraction procedures are used for studying element mobility and availability in solid matrices, like soils, sediments, sludge, airborne particulate matter and so on. In the first part of this review we reported an overview on these procedures and described the applications of chemometric uni- and bivariate techniques and of multivariate pattern recognition techniques based on variable reduction to the experimental results obtained. The second part of the review deals with the use of chemometrics not only for the visualization and interpretation of data, but also for the investigation of the effects of experimental conditions on the response, the optimization of their values and the calculation of element fractionation. We will describe the principles of the multivariate chemometric techniques considered, the aims for which they were applied and the key findings obtained. The following topics will be critically addressed: pattern recognition by cluster analysis (CA), linear discriminant analysis (LDA) and other less common techniques; modelling by multiple linear regression (MLR); investigation of spatial distribution of variables by geostatistics; calculation of fractionation patterns by a mixture resolution method (Chemometric Identification of Substrates and Element Distributions, CISED); optimization and characterization of extraction procedures by experimental design; other multivariate techniques less commonly applied.

*Corresponding author at: Department of Analytical Chemistry, University of Torino, Via Giuria 5, 10125, Torino, Italy. Tel.: +39 011 6707844; fax: +39 011 6707615.
E-mail address: ornella.abollino@unito.it (O. Abollino)

Keywords: single extraction; sequential extraction; chemometrics; multivariate statistics; soil; sediment

1. Introduction

Chemometric techniques are very useful for processing large data sets [1-3], such as those obtained by applying single or sequential extraction procedures, commonly used to study element mobility and availability in solid matrices like soil, sediment, sludge, incinerator fly ash and atmospheric particulate matter [4-7]. A large amount of data is necessary to gain insight into the properties of such matrices, because they have a complex composition and their behaviour depends on the concentrations of several elements, further divided into fractions in the case of sequential extractions, and on their main chemical and physical properties, e.g. pH and organic matter content. Consequently, multivariate chemometric techniques, which allow many parameters to be taken into account simultaneously, are very helpful for visualization and interpretation of single and sequential extraction results [8]. The outcomes of such techniques can be used to classify samples and variables according to their characteristics, to understand the causes of a phenomenon, to develop a model for predicting future events. Other applications of chemometrics to extraction assays are the investigation of the effects of experimental conditions on the response, the optimization of their values and the calculation of element fractionation.

The first part of this review [9] dealt with the following topics: i) an overview on the most common extraction procedures, with their advantages and disadvantages; ii) an overview on univariate and bivariate chemometric methods, which remain indispensable for data processing and interpretation, even when sophisticated multivariate techniques are subsequent applied; iii) the principles and applications of pattern recognition techniques based on variable reduction, namely principal component analysis (PCA), factor analysis (FA) and N-way methods.

In this second part of the review we will cover a number of aspects associated to chemometrics and single and sequential extractions: i) further pattern recognition procedures; ii) modelling; iii)

geostatistics; iv) calculation of element fractionation; v) experimental design. Soil and sediment matrices will be considered, unless otherwise stated. After a short description of the principles of the techniques, we will focus on the aims for which they were applied and the key findings they provided.

The main features of each cited paper are summarized in Tables 1 and 2. Table 1 reports the operational definitions of the fractions obtained with sequential extraction procedures, the reagents used in the extraction steps and a few comments. Table 2 shows: i) the nature of the investigated matrix and its location; ii) the elements determined; iii) the extraction media; iv) the analytical method used for the determination of the extracted elements; v) the chemometric technique(s) applied; vi) the software package used. Regarding point vi), in sections 2-4 we will mention the software packages used in the cited papers. It should be recalled that also other packages can be adopted for the calculations required by each chemometric technique. Besides commercial products, free software packages are available: in many cases, free products provide the ability to perform calculations properly but have less graphical facilities than commercial ones. The choice of the most suitable package for data processing should be made taking into account the aim of the study (e.g. whether a graphical representation of results is necessary or not) and the chemometric techniques that will be applied: in particular, it is convenient to choose a product provided with several techniques of the same type (e.g. different pattern recognition techniques) and several algorithms for the same technique, so as to have the possibility to compare the results obtained with different procedures.

2. Visualization and interpretation of experimental results

2.1. Cluster analysis

2.1.1. Principles

The aim of cluster analysis (CA) is grouping samples or variables [1-3]. CA techniques can be divided into two groups: i) hierarchical, in which groups are obtained stepwise with an

agglomerative (starting from single objects and joining them in successively larger groups) or divisive (starting from one cluster comprising all objects and dividing it into successively smaller and more homogeneous clusters) approach and ii) non-hierarchical methods, in which objects are not successively joined but clusters are determined directly.

We will focus our attention on hierarchical CA (HCA), which is most frequently used. It is an unsupervised pattern recognition technique, in which the clusters are formed during the calculation and their number is not decided *a priori.* The first step in clustering objects is to evaluate their similarity (or dissimilarity): distance or correlation coefficient can be used as a measure of (dis)similarity. One way of measuring distance between two objects i and j in HCA is the Euclidean distance:

$$d_{ij} = \sqrt{\sum_{v=1}^{n} (x_{iv} - x_{jv})^2} \qquad (1)$$

where n is the number of variables. Using vector notation, equation (1) becomes:

$$d_{ij}^{\ 2} = (x_i - x_j)^T (x_i - x_j) \qquad (2)$$

where x_i and x_j are the column vectors of the two objects and T stands for “transpose”.

The smallest is the Euclidean distance, the highest is the similarity between the objects. The Euclidean distance can be graphically interpreted as the length of a vector starting from i and ending in j. It represent a particular case of the general Minkowski distance:

$$d_{ij} = \left[\sum_{v=1}^{n} (x_{iv} - x_{jv})^p \right]^{1/p} \qquad (3)$$

with p = 2. When p = 1 the Manhattan or city block distance is obtained, which is scarcely used in HCA. Some studies adopt the squared version of the Euclidean distance, which provides a faster data processing in comparison to the conventional Euclidean distance, especially in the presence of large datasets. Another approach to the calculation of distance is the Malanobis distance:

$$d_{ij}^{\ 2} = (x_i - x_j)^T C^{-1} (x_i - x_j) \qquad (4)$$

where C is the variance-covariance matrix.

The comparison of equations (2) and (4) shows that the Euclidean and Malanobis distances differ for the presence of matrix C, i.e. Malanobis distance takes correlation into account.

Other measures of distance can be found in Massart's et al. [1] and Otto's [3] textbooks; both texts also report some interesting graphs showing the differences between the distances measured with different methods. The most common measure of distance used in HCA is the Euclidean distance, even if the Malanobis distance is considered to be more reliable by Massart et al. [1] and references therein.

Instead of distance, correlation coefficients between the two row-vectors x_i and x_j can be used as a measurement of similarity. As explained by Massart et al. [1], two objects may have the same Euclidean distance from a third one, but different correlation coefficients, if the cosine of the angle between the corresponding vectors (which is equal to correlation coefficient) is different: therefore different degrees of similarity between the two objects are obtained with the two approaches. The choice between the two indexes of similarities is not straightforward; in several cases, distance and correlation coefficient are used to classify samples and variables respectively [10], but this rule is not always followed. Most of the papers cited in section 2.1.2 started from Euclidean distances [11-14] or squared Euclidean distances [15-18] for clustering samples and/or variables. Some studies adopted correlations as similarity measurements for the classification of variables [10,19,20] or, less commonly, samples [21].

The second step in HCA is clustering objects, which is more commonly performed with the agglomerative approach than with the divisive one. There are several clustering algorithms. One set of popular methods is based on the similarity matrix, which contains the similiarities between all couples of objects; initially, the two most similar objects (let us call them a, b) are identified and substituted with a new combined object (c). The similarity between object c and each of the remaining objects is calculated with different procedures, the most typical ones being single linkage, complete linkage, also called furthest neighbour, and average linkage: for instance, with

average linkage the average of the similarities between objects a and b and each remaining object is calculated. The process is repeated until all objects are linked. Massart et al. [1] and Otto [3] show a numerical example of calculation; in addition, the former text compares the results obtained with the three linkage modes mentioned above. Another popular clustering approach is Ward's method, in which the criterion for cluster forming is that a minimum increase of the sum of the squared distances of each member to the centroid (the point with mean values of the variables) of the cluster should occur. The results of HCA are usually represented as dendrograms, which show the degree of similarity among objects. Otto [3] shows a comparison of the dendrograms obtained with average linkage and Ward's methods.

The majority of the papers cited in section 2.1.2 used Ward's method of agglomeration [11,13-18,22]. Other studies applied complete linkage [12,19,21] or average linkage [10,23] clustering. In general, using different measures of similarity and/or clustering algorithms yields to different results; for this reason, we agree with Massart et al.'s recommendation to combine a clustering method with PCA or other data representations and, if possible, compare different methods of performing HCA [1].

Non-hierarchical methods are not unsupervised techniques, since they start with a decision on the number of clusters to be made; then each object is assigned to a cluster, following various procedures: for instance, in Forgy's method the distance between each object and the centroid of the clusters is calculated, and each object is assigned in the cluster with the nearest centroid. The assignment of objects to the clusters is performed repeatedly, until no change is observed in two consecutive steps [1].

Fuzzy clustering analysis is a version of CA in which there is not a sharp boundary between clusters and is useful for the treatment of data sets affected by uncertainty or vagueness, in which an overlapping of clusters can take place.

As Table 2 shows, most of the cited papers used SPSS (Statistical Package for Social Sciences) as software package for calculation. In general, SPSS is one of the most popular products for many

types of statistical calculations; it was issued in 1968, and many revisions have been made since then. Other studies adopted XLStat, which is somewhat less complete than SPSS, but has the advantage of being an Add-in package of Microsoft Excel. One paper used Statistica, another well known product for statistical calculations.

2.1.2. Applications

CA has been widely used in conjunction with extraction assays, even if less commonly than PCA. In many studies, both techniques were reported and, according to our experience, they yield similar results. For this reason, many of the papers cited in this section were already mentioned in Part I of the review [9].

CA can be used to identify groups of similar samples and anomalous specimen; alternatively, the variables can be treated as objects and their similarity or dissimilarity can be investigated. The results of this techniques have been used for several purposes: i) to distinguish among sources of elements [10,19]; ii) to identify zones with different pollution levels within a contaminated site [15-17,24]; iii) to study the chemical behaviour of main properties and element fractions [12,14,20,23]; iv) to study soil-plant relationships [11,18]; v) to distinguish soils from different areas [22] or the variations of chemical composition along a sediment core [13]; vi) to evaluate the characteristics of extractants [21]; vii) to better interpret the results of PCA [25] or compare the performances of different algorithms [26]. Examples of these applications are discussed hereafter.

Micó et al. [10] determined the total and EDTA-extractable contents of elements in agricultural soils. They computed the correlations between element concentrations and main soil properties and between element pairs, and studied the classification of variables with HCA. Two main clusters were present in the dendrogram. One comprised total and extractable concentrations of Cd, Cu, Pb and Zn, and was associated to an anthropogenic influence, since some samples had high levels of these elements. The other cluster contained total and extractable concentrations of Co, Cr, Fe, Mn and Ni, whose variability was presumed to depend on the parent rock, and represented a lithogenic

component. The results of HCA and correlation analysis were in good agreement, since the lithogenic elements had higher correlations with soil properties than anthropogenic ones (Cd, Cu, Pb and Zn), which were mainly influenced by external sources.

Pérez et al. [15] used HCA to identify groups of samples with similar characteristics within a contaminated site; they found three main groups of samples clearly differentiated by element mobility; the behaviour of variables was investigated by PCA (see Part I) [9].

Zhang et al. [19] applied correlation analysis and HCA to the results of a large database on 1310 soils, which mainly reports total concentrations, but also the available content of some elements. The variables were clustered in four groups in the dendrogram, depending on their characteristics (chemical properties or source). The authors pointed out that correlation analysis is effective in revealing the relationship between pairs of variables, but the results can be complex in the presence of a large number of variables, whereas CA enables the summarization of the multiple relationships among all the variables in a single cluster tree.

Katsaounos et al. [12] treated the results of P fractionation in soils with HCA. The geological parameters (sand, silt, organic matter, Fe, $CaCO_3$, Al, Ca) formed a separate cluster from the fractions of P. The authors did not discuss further the HCA results and interpreted the data mainly on the basis of other techniques (see Part I and section 2.3.1) [9].

Tokalioğlu and Kartal [11] applied HCA to element concentrations in vegetables and soils using a single extraction (see also Part I) [9]. They found one main cluster comprising of elements (Cd, Co, Cr, Fe and Pb) in soil and another cluster with the elements in vegetables and some other ones (Cu, Mn, Ni and Zn) in soil. HCA did not provide information on the soil-plant relationships, since no clear separation was found as a function of either the matrix or the elements, suggesting that the extractant was not representative of the uptake of elements by plants. In a previous paper Tokalioğlu et al. [8] studied the availability of elements in vineyard soils to grapes comparing the performance of three single extractants. They treated the data with HCA separately for each reagent and found similar clusters of variables in the three dendrograms. The results of HCA were used to

support the conclusions of correlation analysis and PCA (Part I) [9].

Our research group applied HCA to the element fractionation results obtained for an agricultural soil, a contaminated soil and a marine sediment core [13,16,17,22]. In all cases, the sample clustering observed in the dendrogram was in general agreement with the groups evidenced in the PCA score plot (Part I) [9]. In addition, the outcome of HCA allowed us to better assess the level of similarity among the investigated samples. Fig. 1 shows the dendrogram obtained for element percentages extracted into the third fraction of Tessier's scheme from an industrially polluted soil [16]; a clear separation between surface (A1-A14) and vertical profile (A17b-A31) samples is apparent, with the exception of sample A19. The sub-clusters of vertical profile samples did not show a clear trend as a function of depth, probably because the profile was collected from the side of an artificial relief consisting of a mixture of industrial wastes and soil: for this reason, it was not characterized by the sequence of horizons expected for natural soils, and element distribution depended on the disposal history of the site. The top layer (A17b) was not differentiated from the layers underneath, suggesting that weathering, which should have presumably affected sample A17b more extensively than A19-A31, apparently did not influence the availability of elements bound to Fe and Mn oxides. The PCA score and loading plot for the same samples have been shown in Part I [9].

Fig. 1. Dendrogram obtained by HCA (Ward's method) for pH and element percentages extracted from contaminated soil samples (coded A1-A31) into the third fraction of Tessier's procedure, after column standardization [16].

Praveena et al. [23] applied HCA separately to data on mangrove sediment samples collected at low and high tide (see also Part I) [9]. In both cases, two clusters of variables were observed, but with different degrees of similarity. In particular, the salinity and electrical conductivity were more closely clustered for samples at high tide than at low tide, indicating a stronger contribution of seawater during high tide.

Relić et al. [20] applied HCA to sequential extraction data for river sediments, but they made few

comments to the dendrogram and mainly discussed the results considering PCA and correlation analysis (Part I) [9]. They observed some analogies between the associations of variables in the dendrogram and in PCs.

Meers et al. [21] compared 13 methods of extractions for Cd from soils with different composition and contamination levels. The dendrogram obtained by HCA, using the extractants as objects, showed that the extractions were divided into five classes, which mainly differed for the aggressiveness of the extracting reagents (see also section 2.2.2).

Alvarenga et al. [25] used HCA only to decide how to cluster data on contaminated soil in a plot of PC1 vs. PC2 (see Part I) [9], but we think that this technique, also used alone, gave interesting results. The samples unamended with organic residues were differentiated from the amended ones, which were clustered according to the effect of the added material on their chemical properties (e.g. pH, mobile element content).

Richter et al. [14] studied element fractionation in airborne particulate matter (PM_{10}). The sequential extraction results were processed with HCA. The investigated elements formed three clusters: i) As, Cd, Cu, Mn, Pb and Zn, which were extensively extracted into the first two fractions, i.e. showed high mobility; ii) Cr, Ti and V, which were mainly present in the third fraction (bound to carbonates and oxides); iii) Al, Ba, Ca, Co, Mg, Mo, Ni, which were primarily extracted into the fourth fraction (bound to silicates and organic matter) and consequently highly immobile. The authors concluded that the presence of high percentages of toxic elements, such as As, Cd and Pb, in the most mobile fractions was a cause of concern for the environment and human health.

Palumbo-Roe et al. [24] studied the bioaccessibilty of As in mine waste-contaminated soils. They applied non-hierarchical CA and found that the samples were clustered in four groups with different levels of total and bioaccessible As and of other soil properties (see also sections 2.2.2 and 3.2). An interesting aspect of this paper is that the results of CA were reported on a map of the site in order to show the spatial distribution of the clusters.

Sârbu et al. [26] applied a different statistical treatment, namely fuzzy clustering analysis, and

compared the performance of two algorithms, fuzzy c-mean algorithm and the Gufstasson-Kessel (GK) algorithm. They processed the sequential extraction results obtained by BCR on two profiles of a contaminated soil, and checked whether the sample solutions could be divided according to the extraction step. PCA, classical CA and projection pursuit (see section 2.3.4) gave unsatisfactory results, whereas fuzzy divisive hierarchical clustering provided an improved classification of the samples into each fraction. In particular, the GK algorithm provided better results than the c-mean algorithm in classifying the samples according to their characteristics.

2.2. Multiple linear regression

2.2.1. Principles

In multiple linear regression (MLR), a relationship between several independent variables x_1-x_n (predictors) and a dependent variable y (response) is expressed with a mathematical function [1-3]. In many cases, the relationship is linear, or can be linearized, and has the form

$$y = b_0 + b_1x_1 + b_2x_2 +... + b_nx_n \quad (5)$$

where b_0...b_n are the regression parameters. They can be estimated with least squares regression or (in case of non-normal error distribution) with robust regression methods.

Equations like (5) represent a model of the system under study, which can be used to investigate which variables influence its response and at what extent, and/or to predict the value of one variable when the others are known. The model is usually validated by applying it to independent experimental data (i.e. data not used to develop the model) and comparing the predicted value of the variable of interest with the measured one; in the absence of such data, cross-validation procedures can be adopted.

As observed for HCA, SPSS was the software package most extensively used for MLR calculations in the cited papers (Table 2). One study adopted Statgraphic Plus.

2.2.2. Applications

MLR considers experimental results from a different point of view with respect to pattern

recognition techniques, since it is focused on relationships between variables and not on similarities or differences among objects. The aim of MLR studies is often to predict one property from the known values of other properties, which are possibly more easily measured. As the examples reported below show, several applications of this technique in conjunction with extraction assays regard the relationships between extractable element content in soil and its uptake by plants [27-29]. MLR permits to take into account the effects of soil properties, in addition to the element concentration released with a suitable extractant (see Part I, section 2.1), on the amount of such element assimilated by plants. The models so obtained can be helpful for two main purposes: i) to predict the risks of transfer of pollutants from soil into the food chain, and consequently to decide on the proper land use; ii) to estimate the amount of micronutrients available for crops, and detect possible deficiencies. Despite the agreement between predicted and observed behaviour found by numerous researchers in their studies, there is not yet a generally accepted model for the estimation of element uptake by plants. Another less common application of MLR to biotic matrices is the study of the effect of soil properties and element content on microorganisms [30]. MLR also finds application to identify the level of contamination of a soil, in particular with the aim of predicting the amount of mobile elements from total concentrations and other properties which influence its release [31,21]. In addition, MLR has been used to investigate the parameters affecting some properties of the investigated matrix or of a specific element; for instance, much attention was devoted to the effect of soil properties on As adsorption and bioaccessibility [24,32-34] (see part I, section 2.2).

Krishnamurti et al. [27] compared seven extractants to test their suitability for predicting the plant-available Cd concentration. They grew durum wheat in greenhouse experiments and calculated the correlation between the concentration in the stem and leaves and that extracted from the soil with the different reagents. The amount extracted with NH_4Cl gave the highest correlation coefficients, hence the authors chose the results obtained with this reagent as a starting point to develop a multivariate model to predict Cd uptake in plants, taking into account some soil variables, such as

free Fe, pH, organic carbon and clay content. They observed a good agreement between actual and predicted Cd concentrations in plant leaves.

Tian et al. [28] studied the relationships between element content in the roots and unpolished grains of field-grown rice and the contents in rhizosphere soil measured in leachates obtained by single extraction, in soil solution and by diffusive gradients in thin films (DGT). Simple bivariate linear regression analyses between concentrations in plants and those leached from soil with each treatment showed that DGT gave better correlation coefficients than the other techniques. The authors used PCA only as a means of reducing multidimensional soil properties to lower-dimension parameters. They extracted two factors and included them in models having the concentrations of Cu or Pb in roots as responses and the element concentrations in soil and PCs as predictors. Correlation coefficients for models involving concentrations in soil solution and in single extracts improved in comparison to bivariate equations, because MLR took into account the effect of soil properties on element uptake by roots; on the contrary, there was only a slight improvement using DGT data. The authors concluded that DGT takes all the important soil factors into account, and that it gives a better prediction of element availability to rice than the other techniques.

Alvarez et al. [29] used MLR to compare the suitability of extraction with EDTA and of two fractionation procedures for predicting the phytoavailability of Mn and Zn for barley in agricultural alkaline soils. The phytoavailable portion of the elements was assumed to be represented by their concentrations in the stem and leaves of barley, which was the dependent variable in the models. The predictability improved when some soil properties (clays for Mn and $CaCO_3$ for Zn), the total element content, or both, were included as variables. The authors concluded that the sequential extraction results were not significantly better than EDTA extractions plus soil properties for the estimation of phytoavailability.

Illmer et al. [30] investigated the importance of Al availability for micro-organisms in forest soils (see also Part I) [9]. First of all, they studied the relationship between organic matter and other soil properties by MLR and obtained a model including moisture content, pH, concentration of

extractable Ca, stable aggregates and electrical conductivity, whereas other parameters, including concentration of KCl-extractable Al, had no significant effect on organic matter. However, significant correlations between Al and biotic parameters (e.g. microbial biomass and ATP concentration) were found for acidic soils. MLR was again applied, and each biotic parameter was put against all abiotic parameters investigated. The authors found that the content of organic matter, the pH, the electrical conductivity and the aggregate stability occurred in many of the models and had high positive or negative influence on biotic parameters, but the concentration of Al was the only abiotic factor present in every model, indicating its importance for abundance and activities of microorganisms in acidic soils.

Meers. et al. [31] sampled the soil solution from clean and contaminated soils and calculated the bivariate correlation coefficients between the main soil properties, total Zn content, Zn in the soil solution, their respective log-transformed counterparts and the relative solubility of Zn, calculated as the ratio of its concentrations in the soil solution and the total content in soil. They found that soil pH and soil texture had a strong influence on the level of Zn in the soil solution. Subsequently the authors reported a series of multivariate equations, expressing soil solution levels of Zn as a function of main soil properties and total Zn concentration in the soil. They compared the predicted and observed levels of zinc in the soil solution with the aid of scatter plots and individuated the most satisfactory equations. They also treated their data with multivariate regression formula derived from the literature. The authors pointed out that the current Flemish soil legislation links the relevant threshold values to clay and organic matter content, but not to soil pH, whereas the results of their research suggest that this parameter should be taken into account in site specific environmental risk assessment. Finally, two model speciation programmes, namely Visual Minteq and WHAM VI, were used to assess free ion activity based on soil solution composition: the results of the two models were in disagreement, suggesting the need for further investigation in the parametric background of the models.

In a later study, Meers et al. [21] compared Cd extractability from soils with 13 extracting reagents.

They constructed equations expressing extractable Cd contents as a function of main soil properties and total Cd concentrations. The variables required to describe Cd extractability with aggressive and mild reagents were different, confirming the clustering identified with HCA (see section 2.1.2). The equations were validated using literature datasets: reasonably good fits between predicted and observed concentrations were found. The authors warned that multiple regression models must be used with caution, and no single universal function will suffice to express Cd extractability for all soils. They suggested that these models can play a valuable role for a preliminary evaluation of soils potentially contaminated with Cd based on total element content and main soil properties, without the need of performing single extractions, thus reducing analysis costs and times.

Juhasz et al. [32] determined As bioaccessibility in soils contaminated by different sources of this element. They developed regression models aimed to predict bioaccessibility as a function of soil properties. The predictors that best described this characteristic in railway corridor, dip site and mine site soils were total As, total Fe and dithionite-citrate extractable Fe. A poor fit was found for soils containing geogenic As sources (gossans). The authors treated the experimental results with a model reported in a literature study and tested their own model with data reported in such study: they found that their own model had better prediction capacity than the literature model. They concluded that As bioaccessibility is controlled by a number of variables (soil properties, speciation, mineralogy, duration of contaminant–soil contact) which may be difficult to accurately model if only a few soil parameters are considered.

Palumbo-Roe et al. [24] developed two models for the prediction of As bioaccessibility in the two soil clusters previously identified with CA (see section 2.1.2). They found different explanatory variables for the two groups of soils, suggesting the binding of arsenic to different phases. Such phases were identified from the results of the sequential extraction data treatment (see section 3.2) as Fe oxyhydroxides adsorbing or coprecipitating As and sulphide, probably in the form of arsenopyrite. The presence of different phases can be the cause of the difference in bioaccessibility of As observed in the two groups of soils.

Yang et al. [33] tested the bioaccessibility of As(V) added to 36 well-characterized soils with a wide range of soil properties. They used MLR to identify the properties controlling As(V) adsorption (defined as the fraction of As(V) adsorbed to the soil over 48 h contact period), bioaccessibility and sequestration (the relative change in bioaccessibility over 6 months). Fe oxide content and pH were the most important soil properties governing As(V) adsorption and sequestration respectively, whereas both these properties and, to a lesser extent, total inorganic carbon (TIC) significantly influenced the steady-state bioaccessibility of As measured after six months. They validated the models on five independent samples and applied them to predict the *in vivo* bioavailability of As, previously measured in swine dosing trials, in nine contaminated soils. They found a good agreement between the experimental and predicted values and concluded that the models can be used to estimate As(V) bioaccessibility and bioavailability on the basis of soil properties.

In a subsequent paper [34] the authors made a similar study on soils spiked with As(III). Several soil properties significantly influenced the adsorption and sequestration of As(III), indicating the complexity of the behaviour of As(III) in comparison to As(V), which was influenced by few variables. The bioaccessibility of As(III)-spiked soils was initially higher than that of As(V)-spiked soils, but it decreased with ageing to levels closer to those of the As(V)-spiked soils. Such a decrease was accompanied by the heterogeneous oxidation of As(III) to As(V). One of the models previously developed [33] was used to estimate the bioaccessibility of As in the As(III)-spiked soils and the *in vivo* bioavailability of As to juvenile swine with reasonable errors, but it failed in predicting the bioavailability measured for monkeys in a different set of samples. The authors concluded that the model can provide an initial estimate of As bioavailability in soil, which can be used to screen and prioritize As-contaminated sites and justify expensive *in vivo* studies. In addition, such estimate is supposed to be valid over the long-term, since it is based on major soil properties which should be stable over a relatively long period of time.

2.3. Other techniques

2.3.1. Linear discriminant analysis

Linear discriminant analysis (LDA) is a supervised pattern recognition technique, in which objects are grouped in classes established by the researcher *a priori* [1-3]. It is based on the calculation of new variables, called discriminant functions, obtained from linear combinations of the original variables. Such linear combinations are derived so as to best indicate the differences between the classes in contrast with the variance within the classes. LDA has two main applications: i) the evaluation of the discriminating power of variables with regard to the separation of the groups; ii) the assignment of each object to one of the groups. SPSS was used for LDA calculations in one of the papers cited below (Table 2), whereas the software used in the other paper was not reported. LDA has been scarcely applied to extraction results, but we think it deserved more attention, because it is relatively easy to use and can aid in distinguishing among samples with different levels of contamination or different element availability.

Katsaounos et al. [12] applied LDA to classify river sediments (see also Part I and section 2.1.2) [9] into three classes (clean: < 0.6 mg g^{-1} P; marginally polluted: 0.6 mg g^{-1} P; and polluted: > 0.6 mg g^{-1} P) from the results of P fractionation and speciation. The variables identified as redundant for PCA were not considered in LDA, in order to reduce uncertainty.

Dalurzo et al. [35] used LDA to differentiate Alfisols, Ultisols and Oxisols soils (see also Part I). The variables with the highest discriminating power were exchangeable Al and pH, but also different fractions of P aided in the discrimination among the different kinds of soils.

2.3.2. Artificial neural network

Artificial neural network (ANN) techniques process data by simulating the behaviour of a biological neuron. A network consists of several units, the neurons, interconnected according to suitable designs; neurons are arranged in layers, defined as “input layers”, “output layers” and intermediate “hidden layers”. A neuron receives the input signals (i.e. data) from other neurons,

aggregates them with the use of weights, applies a transfer function to transform them and passes them as the output signal [1,3]. The main applications of ANN in chemometrics are unsupervised pattern recognition (input vectors, representing the objects, are grouped according to their features), supervised patter recognition (objects are assigned to predefined classes depending on the value of the output neurons), modelling (the weights are the parameters of the model). To our knowledge, few applications of neural network techniques to extraction results exist. Their application is actually less simple than PCA and HCA and, at least in the two examples reported hereafter, the visualization of the experimental results is less straightforward. The two cited studies used Statistica for data processing but (as pointed out in the Introduction) other software packages can be adopted.

Boszke and Astel [36] used self-organizing map (SOM), a neural network-based technique, to classify floodplain soil samples, potentially affected by urban pollution sources, according to the level of contamination with Hg. The samples were divided into two clusters, corresponding to "heavily polluted" and "moderately polluted" samples. The authors concluded that SOM may be helpful in estimation of the contamination levels in samples and for the identification of "site-specific" polluted samples. They underlined the finding that different patterns of contributions of mobile and immobile Hg species were present in moderately polluted and heavily polluted samples, the former having higher percentages of mobile element than the latter. They estimated that the risk of remobilization of Hg from the investigated samples was relatively low, but the possibility of leaching of organomercury and acid-soluble species in case of high water levels or heavy floods could not be excluded.

Buszewski and Kowalkoski [37] used columns packed with artificially polluted soil samples and leached them with solutions of nitric acid (pH 2.5 – 5) simulating acid rain. They determined element concentrations and other parameters in subsequent aliquots of the eluate. They discussed the trends of element release during column leaching and the correlations between element concentrations and the main soil properties. ANN was used to model element transport. The input parameters were the initial element concentrations in the contamination solution, redox potential

and pH in the artificial acid rain, soil properties and total and inorganic carbon content in the eluate; the output parameters were the element concentrations in the eluate. The agreement between calculated and experimental elution curves was good, with some exceptions for Cd due to the small number of available samples. The authors concluded that ANN could be the future tool for modelling transport of inorganic substances in soil profiles; however, they pointed out that their model should be further refined.

In our opinion, the potential of ANN in conjunction with extraction assays has not yet been fully exploited and this technique can provide useful information on the process under study.

2.3.3. Geostatistics

Geostatistic techniques are used to process spatially or temporally correlated data [2,38]. It was developed in the field of mining, but it has found wide application in the assessment of soil pollution and, in general, for the investigation of soil properties. A geostatistic analysis typically consists of the following steps: i) the development of a model describing the spatial relationships among sampling points based on a semivariogram, i.e. a plot of semivariance as a function of distance between couples of samples; ii) the interpolation of the data, based on the semivariogram, using the so-called kriging procedure; iii) the mapping of the results, usually with contour maps, representing the estimates of the values in the investigated areas and the errors of the estimates. Geostatistics is extensively used for processing total concentrations, but its application to extractable element contents is less common. Software packages for geostatistics are usually different from those used for the other techniques considered in this review. Table 2 shows that the product most frequently adopted in the cited papers is GS^+ alone or coupled to other software packages (like ArcGIS or GIS ArcView) for mapping. Other studies employed Surfer or free packages, such as R, Variowin, Vesper, GSLIB, GSTAT, WLSFIT, KRIGE and COKRI, a Matlab program. In one case computer programs developed by another researcher (S.R. Vieria, from Instituto Agronômico de Campinas, Campinas, Brazil) was used.

Many authors investigated the spatial variability of extractable K, P and/or N and main properties in agricultural or forest soils. The aims of such studies were: i) the implementation of site-specific agricultural management practices in order to lower costs, increase crop yields and/or improve sustainable agricultural land use [39-44]; ii) the identification of the characteristics (e.g. parent material, slope gradient, land use, presence of vegetation, fertilizer input) most extensively affecting soil composition [45-49]; iii) the investigation of other phenomena, such as: the short term effects of wildfire [50]; the difference in spatial patterns before and after poultry litter application [51]; the relationship between spatial dependence in oak leaves and in soil, which can have important consequences for herbivores and nutrient cycling in the ecosystem [52]; the soil properties at a forest-prairie boundary and the effect of woody plant expansion on it [53]; the effect of soil composition on rice growth and nutrient content [42]; the sorption of P from waters in riparian wetlands [54]; the effect of pH variability on element fractionation [55].

Another application of geostatistics is the evaluation of the spatial variability of extractable concentrations of some elements, such as Cd, Cu, Pb and Zn, from the point of view of their role as micronutrients (with the exception of Cd and Pb) or as contaminants. Several topics were addressed: i) the sources of the elements [41] and the properties influencing their concentrations [56-61]; ii) the seasonal variations [62]; iii) the identification of areas with different contamination levels and consequently different risks for plant growth or human health [47,57,63-65]; iv) the effect of different data processing procedures [66,67]; v) the relationship between Cd content in durum wheat grain and in soil [68]; vi) the effect of gypsum addition and water management (flooded vs. non-flooded) on the yield of marsh vegetation [69]; vii) the identification of deficiency of micronutrients [56].

As Sigua and Hudnall [61] pointed out, the knowledge of the spatial pattern of soil characteristics is useful to select the proper land use, to apply the necessary amount of amendments or fertilizers and in general to assist the activities of environmental scientists, water resource planners, regulators, decision makers, engineers, soil scientists, and resource managers.

Geostatistics has been primarily applied to soils, but it can also be used for the investigation of sediments. Van Griethuysen et al. [69] focused their attention on the immobilizing effect of acid volatile sulphides (AVS) on elements, and estimated availability in lake sediments as the difference between simultaneously extracted metals (SEM, defined as the sum of the molar concentrations of all metals extracted during the acid volatilization step and with solubility products lower than those of MnS and FeS) and AVS. The authors examined the spatial variations of total element concentrations and main properties in comparison to spatial variations in redox conditions and available concentrations. No clear spatial dependence was detected for main properties, total and SEM concentrations, whereas AVS concentrations showed a strong spatial dependence, due to differences in lake depth, total sulphur pools and redox potential: this behaviour caused differences in availability within the lake. The authors remarked that alternative sediment quality criteria like SEM - AVS and (normalized) total concentrations may yield different priority settings for contaminated sites, due to differences in spatial variability of the underlying variables. The results were also treated by PCA, but no meaningful findings were obtained for SEM.

Castrignanò et al. [70] used factorial kriging analysis (FKA), a combination of geostatistics and PCA, to explore the characteristics of agricultural soils. This technique allows to study the behaviour and relationships among variables at different spatial scales. The strong correlation among N, exchangeable Na and cation exchange capacity (CEC) at plot-size level suggested that the short-range variation mainly derived from the spreading of fertilizers; at long spatial scale, clay was correlated with Na and CEC, supporting the hypothesis that the long-range variation was related to soil structure. The authors concluded that this kind of soil characterization is useful for farmers, since it enables them to follow crop management practices fitted with the real soil situation. Castrignanò et al. [71] applied FKA also to estimate the probability of occurrence of *Tuber melanosporum*, an ectomycorrhizal fungus which produces black truffles, in relation to some soil properties. In particular they suggested that a soft and well-aerated soil environment might be an essential condition for the growth and production of the fungus.

We think that the application of geostatistics to single and sequential extraction results should increase, because it can give useful information for risk assessment and land management. This technique requires a substantial expertise in statistics and soil science, especially in the choice of the most suitable model for the semivariogram.

2.3.4. Miscellaneous techniques

Snape et al. [72] used a commercially available software mainly designed for the treatment of environmental/ecological data (PRIMER), and applied a series of techniques aimed at investigating similarities and dissimilarities among data, namely non-metric multidimensional scaling (nMDS), analysis of similarities (ANOSIM), similarity percentages (SIMPER), relative dispersion and correlations between similarity matrices (RELATE). These techniques, which are not so common in chemometrics, were used to examine differences among marine sediments collected in different locations and to compare 0.5 and 4 h HCl-extraction data. The authors concluded that the 4 h extraction provided better definition of the low to moderately contaminated locations by picking up small differences in anthropogenic element concentrations.

In a later paper of the same research group, Stark et al. [73] described a survey on the small-scale spatial distribution of extractable element concentrations and of the abundance of soft-sediment communities in marine sediments collected along three transects (A, B and C) in the vicinity of an abandoned waste dump and a fourth transect (D) at increased distance from the dump. The values of the investigated parameters in transects A, B and C showed a high variability; nevertheless, conventional and non parametric ANOVA, nMDS and ANOSIM revealed the presence of a gradient of contaminant concentrations and of some infaunal *taxa* along the transects. Sediments in transect D were distinctly different from the ones in the other three transects for sediment communities, grain size and composition. The authors remarked that the spatial heterogeneity observed in their study indicates the need for careful sampling design and adequate spatial replications. They found significant correlations between the presence of contaminants and the

distribution and composition of soft-sediment communities. A procedure (BIOENV) available in the commercial software was used to identify the combinations of environmental variables giving highest correlations with biotic data with: the combinations of Cd, Cu, Sn, Pb concentrations and grain size were the variables that best matched the community patterns in transects A, B and C; a larger number of variables were involved when all four transects were examined together. The authors stated that the results of their study, and of other studies in the same area, form a baseline for the assessment of recovery of communities after remediation of the dump.

González et al. [74] obtained soil solution by displacement with water from columns packed with calcareous soils. They examined the correlations between the composition of the soil solution (elements content and other parameters) and of the soil samples (elements batch-extracted with ammonium acetate and main soil properties). For instance, the significant negative correlation between dissolved and extractable Ca was interpreted as the tendency of this element to be strongly retained by the exchange complex, particularly in calcareous soils in semi-arid and arid regions. Furthermore, the authors applied redundancy analysis (RDA, also called reduced rank regression, RRR, or PCA on instrumental variables) with the aid of CANOCO, a software package mainly used for ecological applications. RDA is an extension of multiple regression, used in the case of more than one response variables, therefore it is aimed at examining the relationships between a set of explanatory (independent) variables, the soil properties in this case, and a response data set, i.e. the variables measured in the soil solution in the present study; this goal is reached by calculating linear combinations of the independent variables which explain most of the total variation contained in the dependent variables [1]. In RDA the matrix of dependent variables **Y** is approximated by the regression on a limited set of combinations of the matrix of independent variables **X**. Since each Y-variable is fitted by a linear combination of X-components, and such components are linear combinations of independent variables, then Y-variables are expressed by a linear combination of the X-variables. In the study by González et al., the results of RDA showed that the characteristics of the soil solution were primarily determined by the mineralogical nature of the soil and by the

variables related with the organic fraction of the soil.

Rhoton et al. [75] analyzed suspended sediments from a semiarid watershed and built a multivariate mixing model to identify the relative contribution of six subwatersheds (SWs) to the suspended sediment load leaving the watershed. They used the following procedure: they inserted the average values of sediment parameters into sediment properties vectors d_j, one for each subwatershed; they treated the data with the aim of finding the vector (x) containing the proportion of sediment from each subwatershed that minimizes the function $(Cx-d)^T (Cx-d) = 0$, where C is the matrix made up of the sediment property vectors of the contributing SWs, d is the sediment property vector of the outlet watershed and T stands for "transpose". Since the vectors contain mean values, a Monte Carlo routine was added to the mixing model to take into account the distribution of the data and obtain an estimate of the standard deviation of the individual SW contributions. The authors found that the greatest amount of the sediments originated in three SWs, which were characterized by the lowest soil aggregation index, and consequently by the highest erodibility. The goodness of fit of the mixing model was evaluated by comparing the experimental values of sediment properties in the outlet subwatershed with the values predicted from the relative contributions of the six SWs: a very good fit was observed for most properties. The authors also determined the composition of the soils collected from the SWs (these data were not used in the multivariate model) and discussed the relationships between soil and sediment properties. They concluded that the ability to identify primary sediment sources in watersheds contributes to a more efficient design of management practices aimed at resolving excessive runoff and erosion problems.

In projection pursuit techniques high-dimension data are projected to low dimensions [8]. Among the possible projections, the one in which the data have the highest deviation from the normal distribution is considered to have the highest "interestingness" index and to reveal more information than the other projections. Kowalik and Einax [8] showed that projection pursuit was more efficient than PCA in separating data on mobile and scarcely reactive fractions extracted from floodplain sediment depth profiles.

3. Use of chemometrics for the calculation of element fractionation

3.1. Principles

In 1997 Cave and Wragg [76] proposed the use of a single, non-specific extractant, namely nitric acid, at different concentrations, soil/solution ratios and extraction times. Their approach started from the assumption that the soil is made up of a number of physico-chemical components, each having its own chemical composition (e.g., carbonate component, iron oxide component). By leaching the sample, under certain conditions, a proportion of these components is released into solution. The concentration of an element in each extracting solution can be described as a linear sum of the amounts leached from each physico-chemical source present. The authors applied a chemometric mixture resolution procedure to the data (i.e. element concentrations in the leachate solutions) to determine the number and composition of physico-chemical components in the soil. The chemometric procedure was refined in further studies and was later called "Chemometric Identification of Substrates and Element Distributions" (CISED). We will briefly describe it by referring to the procedure reported in a paper written in 2004 [77], in which the Matlab programming language was used. CISED is based on the equation:

$$\mathbf{A} = \mathbf{B}\,\mathbf{C} \qquad (6)$$

where **A** is a k×n leachate concentrations matrix, in which each row contains the amounts of the n elements present in one of the k extracts, **B** is a k×c matrix, in which each row contains the proportions of the c components present in one of the k extracts, **C** is the c×n matrix, in which each row contains the concentrations of the n elements present in one of the c components. Matrix **A** is obtained experimentally, and the objective of the treatment is to find matrices **B** and **C**. The following steps are carried out:

- the data in matrix **A** are scaled and treated with PCA, followed by Varimax rotation, in order to identify the number of components in the samples, i.e. the value of c;
- the scores are scaled and used as a first approximation of the proportion of each component leached in each extract;

- a multiple linear regression (see section 2.2) is carried out with the sum of each row in the scaled leachate concentration matrix (i.e. the total solids present in each extract) as the dependent variable and the scaled score matrix as the independent variable. The columns of the scaled score matrix are then multiplied by their corresponding MLR coefficients. This matrix, with its rows scaled to 1, is used as the first approximation of **B**;

- a first approximation of **C** is calculated from

$$\mathbf{C} = \mathbf{A}\,\mathbf{B}^T\,[\mathbf{B}^T\,\mathbf{B}]^{-1} \qquad (7)$$

where $\mathbf{B}^T$ is the transpose of **B**;

- negative values are corrected to 0 and a second approximation of **B** is obtained from

$$\mathbf{B} = [\mathbf{C}^T\,\mathbf{C}]^{-1}\,\mathbf{C}^T\,\mathbf{A} \qquad (8)$$

the negative values in **B** are corrected to 0 and the rows are scaled to unity;

- successive values of **B** and **C** are iteratively computed until no significant correction is performed.

This mixture resolution procedure yields two kinds of results: the concentrations of the elements present in each component and the percentage of each component extracted into the various leachates. Such results are interpreted taking into account the properties of the elements and of the investigated matrix, and the geochemical nature of each component is hypothesized.

3.2. Applications

The validity of the procedure was tested with a synthetic data set: the correct number of components was identified, and a good agreement between the expected and calculated element concentrations in such components was obtained [77]. The mixture resolution procedure was applied to: i) a Standard Reference Material by the National Institute of Standard and Technology, NIST SRM 2710 (a highly contaminated soil from pasture land along Silver Bow Creek in the Butte, Montana, area), for which the correlation of the components with existing data on Tessier's fractionation were evaluated [76-78]; ii) NIST SRM 2711 (a moderately contaminated agricultural soil collected in the till layer of a wheat field) and river sediment samples prior to and after addition of humic acids, to

study the effect of the latter on element distribution [79]; iii) mine-impacted soils and soils naturally enriched in As, to study the relationships between bioaccessible As and the soil physico-chemical phases [24,80,81] (see also sections 2.1.2 and 2.2.2); iv) river sediments, to study element distribution and sediment formation processes [82]. The extracting reagent was nitric acid at different concentrations in all studies, with one exception, in which aqua regia or aqua regia and hydrogen peroxide were employed [80].

Two examples of the information obtained with CISED will be given here. Nine components were identified in NIST SRM 2711 in one study [77], in which a mineralogical analysis was also carried out. One component was extracted into water and into 0.01 – 0.05 M HNO_3; it was made up mainly of Ca (42%), Mn (20%), Zn (17%), K and Mg (> 6%) and was identified as the exchangeable fraction. The possibility that it could be connected to calcite, due to the high concentration of Ca, was ruled out because of its solubility in water and the absence of calcite observed in the mineralogical analysis. Two other components, characterized by high percentages of Fe, were extracted in the ranges 0.1 – 1 M and 0.05 – 5 M HNO_3, respectively; they were hypothesized to derive from the partial dissolution of Fe oxide/hydroxide materials and to reflect the presence of amorphous and crystalline forms, which are dissolved at different rates by each extracting reagent. In the case of river sediments treated with humic acids [79], ten components were identified. One of these components was extracted in the range 0.1 -1 M HNO_3 and was made up of Al (32%), Fe (23%), Si (16%), Na (14%) and smaller amounts of Mg, Zn, Cu; it was not influenced by the presence of humic acids, and it was interpreted as a clay component, such as chlorite. Three Fe-based components were identified. One of them was extracted into 0.05 – 0.5 M HNO_3 and was attributed to amorphous oxides; another one was soluble in 0.5 – 5 M HNO_3, and it was assigned as the crystalline oxide phase. Their extractability increased with increasing humic acid amount, probably due to the complexation of Fe. The third component was extracted into 5 M HNO_3 and it was present also in the original humic acid, which was subjected to the same extraction procedure; it was identified as the Fe humic complex derived from the original humic acid material. The small

amount of this component found in the sediment as such was interpreted as Fe-humic compound naturally present in the sample.

In our opinion, the CISED procedure is powerful, but it requires expertise both in chemometrics and in geochemistry, for the assignment of the components to the correct phases, so that its application could require the cooperation between researchers with different competencies. Supplementary data, like knowledge on soil mineralogy or the results of other sequential extraction techniques, are helpful for the correct interpretation of the results.

4. Experimental design

4.1. Principles

The term "experimental design" comprises a series of procedures used to plan experiments, with several aims: i) to understand which variables ("factors") influence the performance of the investigated system ("response"); ii) to model the effects of such factors, also taking into account their interactions; iii) to optimize the experimental conditions, i.e. find the values ("levels") of the factors which provide the optimum performance of the system; iv) to reach these goals with a limited number of experiments, in comparison to univariate approaches, in order to save time and reduce costs [1-3]. These purposes are achieved by performing a series of experiments in which several factors are varied at the same time according to predefined schemes. We can distinguish between sequential and simultaneous approaches. The former are used for optimization purposes and not for modelling, and are based on carrying out a few experiments at a time: according to the results obtained the next experiments are planned. Simultaneous strategies are suitable for both modelling and optimization, and require the execution of a series of experiments following a pre-arranged plan. A main class of simultaneous designs is the full two-level factorial design, in which two levels of each examined factor are established and experiments are carried out at all possible combinations of such levels; for k factors, 2^k experiments are required. When several factors must be examined, which would call for a very high number of experiments (e.g. for six factors: $2^6 = 64$),

fractional factorial designs are used, in which only a reduced number (e.g. 2^{k-1} or 2^{k-2}) of such experiments, chosen with suitable criteria, is carried out. Full and fractional two-level factorial designs are adopted to identify the factors with significant effect on the response, to estimate the extent of such effect and their interactions and to obtain first-order models including binary interactions. Another class of simultaneous designs includes multi-level designs, in which more than two levels of each factor are investigated and a second-order (quadratic) model is usually obtained, consisting of a constant term, first and second order terms and interactions between factors (see for instance equation 6 in section 4.2). Higher order models are seldom developed. The coefficients are calculated by multiple regression and can be used to plot the response as a function of the values of the factors, yielding response surfaces. An example of multi-level design is the central composite design, which arises from the combination of a two-level factorial design, a star design plus a centre point. The software packages Statistica and Statgraphics were used in the papers cited in section 4.2; as remarked in the Introduction, other commercial or free products can be used to set up an experimental design.

4.2. Applications

Experimental design has been applied to single and sequential extraction assays much less extensively than pattern recognition or modelling techniques. Four studies will be reported hereafter. Santamaria-Fernandez et al. [78] used experimental design to optimize the conditions for the non-specific extraction method described in section 3. In a first step, they adopted a fractional factorial design and considered several factors potentially influencing the extraction efficiency: centrifugation rate, centrifugation time, ratio between sample weight and extractant volume, number of extraction steps, acid type and acid concentration. They established maximum and minimum values for all parameters with the exception of acid type (aqua regia or nitric acid) and performed sixteen experiments, which yielded 224 solutions. The application of ANOVA allowed them to identify the significant and non-significant factors. After fixing some conditions (nitric acid,

number of extractions) they made a second set of experiments in which they studied the effects of centrifugation time, centrifugation rate and ratio between extractant volume and sample weight with a central composite design. They performed experiments in 16 different conditions and draw response surfaces reporting the values of two parameters against the amount of each extracted element. They identified the optimum conditions for every element and finally adopted compromise conditions which enabled to obtain the maximum recovery for most of the elements.

In other studies experimental design was used for the characterization of extraction procedures, by studying the effects of the variation of experimental conditions on element leachability. Such information enables to evaluate the robustness of an extraction procedure and the comparability of data obtained in different laboratories with different conditions (e.g. temperature, stirring rate). Cave and Wragg [76] studied the effect of operating conditions on the dissolution of soil components (see section 3) into nitric acid. They applied a central composite experimental design with five levels for each variable (time, T; acid concentration, A; sample to extractant ratio, S) and carried out the experiments in duplicate. They treated the data by multiple linear regression and modelled the total extracted solids in each component of the soil (identified with the procedure described in section 3) for each sample as a function of the three main effects (A, T, S), their interactions (AT, AS and TS) and the quadratic effects (A^2, T^2, S^2). The form of the model was

$$E_c = k + x_1A + x_2T + x_3S + x_4AT + x_5AS + x_6TS + x_7A^2 + x_8T^2 + x_9S^2 \quad (9)$$

where E_c is the total extracted solids for component c, k is a constant term, and $x_1 \cdot\text{-}\cdot x_9$ the linear regression coefficients.

They also calculated a regression coefficient R^2 for each component, which expresses a measure of how well the model fits the data ($R^2 = 1$ is a perfect fit). Initially they considered all the effects, their interactions and the quadratic effects. Then they removed the insignificant effects one by one. They also checked the significance of the effects by ANOVA and reported the corresponding table. Finally, they constructed surface plots of the two most significant factors against the total extracted solids for each component, and discussed the effects of such factors on element extractability. For

instance, the dissolution of one component in 0.3 – 0.7 M nitric acid, with highest concentrations extracted at very short reaction times, was interpreted as reflecting a component quickly dissolved at a reasonable acid concentration, whereas the decrease of the extraction yield with time suggested a readsorption process. This component, which was mainly made up of Pb, Cu, Zn and Mn, was attributed to very fine particulate or clayey material.

Boonjob et al. [83] critically compared and evaluated two automated flow-through fractionation methods, namely sequential injection microcolumn extraction (SI-MCE) and sequential injection stirred-flow chamber extraction (SI-SFCE). The main difference between the methods is the sample housing, i.e. microcolumn or stirred flow-chamber. Element extractability in a standard reference material (NIST 1633b, coal fly ash) was investigated with the BCR scheme. The authors used a two-level full-factorial design to investigate the effect of experimental parameters, namely sample weight and extraction flow rate, on element leachability for the two methods. The response data was the sum of the extractable amounts of a given element in the three BCR steps. The results were reported as Pareto charts [1], i.e. histograms in which the length of each bar is proportional to the absolute value of each estimated effect (i.e. sample weight and extractant flow rate) and interaction between them, on element extractability. The results showed that the two parameters and their interactions had no significant influence on total extractability (with a few exceptions) as a consequence of the multistage nature of dynamic fractionation. The authors concluded that SI-MCE and SI-SFCE are not operationally defined procedures: therefore the extraction conditions can be modified without statistically significant changes on percentage leachability, provided that exhaustive extraction is ensured and appropriate sample containers are used.

In a further paper from the same research group, Rosende et al. [84] adopted a two-level full factorial design to study the effects of extractant flow rate, sample weight and sample fluidization (the latter is a qualitative factor, i.e. the presence or absence of fluidization were considered) on the performance of an automated flow-through extraction system equipped with a column of large volume capacity. Such column was devised to overcome a major drawback of most dynamic

extraction procedures, i.e. the small sample amount involved, which does not ensure representativeness in case of heterogeneous matrices; the extractant flows upward through the column, and the drag force it exerts on the sample particles matches the apparent weight of the particle bed: this assembly ensures sample fluidization, i.e. a homogeneous suspension of the bulk sample. The fluidized-bed column method was tested by applying the Toxicity Characteristic Leaching Procedure (TCLP) to highly inhomogeneous municipal solid waste incineration (MSWI) bottom ashes [9,85]. The results were represented with Pareto charts and the significance of factors' influence and their second-order interactions on the analytical response for each element was evaluated by ANOVA and by examining the coefficients of first-order regression equations of the full factorial design: high values of the coefficients indicated a highly significant effect of the corresponding factor on the leachability of an element and vice versa. Sample fluidization had the strongest (positive) effect on the extraction efficiency, followed by sample amount; extractant flow rate was significant only for Cu and Pb. The interactions between parameters were statistically significant for several elements, showing that a one-at-a-time approach for investigation of factor effects would have been unsuitable. The authors concluded that sample fluidization was imperative for proper performance of the extraction system. Finally, the macrocolumn fluidized-bed apparatus provided shorter operation times and better repeatability and accuracy than the batch procedure, and better repeatability than microcolumn flow-through techniques.

5. Concluding remarks

Chemometric techniques have been applied to single and sequential extraction results mainly for pattern recognition and modelling, but such techniques can be used also for other purposes. In particular, we believe that experimental design should deserve more attention; from the analytical and methodological points of view, this technique is extremely useful to optimize the extraction conditions, identify the effects of the experimental parameters on element leachability and evaluate the robustness of a procedure. In addition, it would be desirable to explore new procedures for

computing element fractionation, in order to evaluate the possibility to tackle the drawbacks of sequential extraction, as it was done with CISED.

Regarding data visualization and interpretation, PCA, HCA and MLR are the multivariate techniques most extensively used in conjunction with single and sequential extractions, followed by FA, N-way methods and LDA. Other strategies, such as nMDS, RDA and projection pursuit have rarely been applied. Presently, geostatistics is mainly being applied to total element concentrations or to the available portion of macronutrients, but it would be worthwhile to use it more frequently to investigate the spatial distribution of available concentrations of trace elements, obtained with single or sequential extractions, both with agronomical and environmental purposes: this treatment might permit to obtain more information on the relationships between trace element and main soil properties and to identify more punctually the areas with high environmental risks, or with a deficiency or excess of micronutrients.

The application of two or more multivariate methods, as several researchers did, is helpful to confirm some findings and/or to better exploit the data in order to acquire useful information. As we pointed out in the first part of the review, it is crucial to be aware of the meaning and the steps of whatever technique is being used, and respect the underlying assumptions.

A cooperation among chemometricians, experts in the matrices under study and analytical chemists is highly desirable. The contribution of specialists in chemometrics is especially necessary in the case of procedures that require algorithms developed by the authors, or complex processing, as well as in the exploration of the applicability of further multivariate techniques to extraction assay results.

Acknowledgement

We thank the Italian Ministry of Instruction, University and Research (MIUR, Rome, PRIN project) and the Italian National Research Program for Antarctica (PNRA) for financial support.

References

[1] D.L. Massart, B.G.M. Vandenginste, L.M.C. Buydens, S. De Jono, P.J. Leqy, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics, Parts A and B, Elsevier, Amsterdam, 1997.

[2] J.W. Einax, H.W. Zwanziger, S. Geiб, Chemometrics in Environmental Analysis, VCH, Weinheim, 1997.

[3] M. Otto, Chemometrics: Statistics and Computer Application in Analytical Chemistry, Wiley-VCH, Weinheim, 1999.

[4] J.R. Bacon, C.M. Davidson, Analyst 133 (2008) 25-46.

[5] C.R.M. Rao, A. Sahuquillo, J.F. López-Sánchez, Water, Air, Soil Pollut. 189 (2008) 291-333.

[6] P. Quevauviller (Ed.), Methodologies for Soil and Sediment Fractionation Studies, Royal Society of Chemistry, Cambridge, 2002.

[7] C. Gleyzes, S. Tellier, M. Astruc, TrAC, Trends Anal. Chem. 21 (2002) 451-467.

[8] C. Kowalik, J.W. Einax, Acta Hydrochim. Hydrobiol. 34 (2006) 4 -436.

[9] O. Abollino, M. Malandrino, A. Giacomino, E. Mentasti, Part I of this review.

[10] C. Micó, L. Recatalá, M. Peris, J. Sánchez, Soil Sediment Contam. 17 (2008) 467-485.

[11] S. Tokalioğlu, S. Kartal, Intern. J. Environ. Anal. Chem. 83 (2003) 935-952.

[12] C.Z. Katsaounos, D.L. Giokas, I.D. Leonardos, M.I. Karayannis, Water Res. 41 (2007) 406-418.

[13] M. Malandrino, E. Mentasti, A. Giacomino, O. Abollino, E. Dinelli, S. Sandrini, L. Tositti, Toxicol. Environ. Chem. 92 (2010) 453–475.

[14] P. Richter, P. Grin, I. Ahumada, A. Giordano, Atmos. Environ. 41 (2007) 6729-6738.

[15] G. Pérez, M. Valiente, J. Environ. Monit. 7 (2005) 29-36.

[16] O. Abollino, A. Giacomino, M. Malandrino, E. Mentasti, M. Aceto, R. Barberis, Water Air Soil Pollut. 137 (2006) 315-338.

[17] O. Abollino, M. Aceto, M. Malandrino, E. Mentasti, C. Sarzanini, R. Barberis, Environ. Pollut. 119 (2002) 177-193.

[18] Ş. Tokalioğlu, Ş. Kartal, A.A. Gűneş, Int. J. Environ. Anal. Chem. 84 (2004) 691-705.

[19] C. Zhang, D. Fay, D. McGrath, E. Grennan, O.T. Carton, Geoderma 146 (2008) 378-390.

[20] D. Relić, D. Đorđević, A. Popović, T. Blagojević, Environ. Int. 31 (2005) 661-669.

[21] E. Meers, G. Du Laing, V. Unamuno, A. Ruttens, J. Vangronsveld, F.M.G. Tack, M.G. Verloo, Geoderma 141 (2007) 247-259.

[22] O. Abollino, M. Aceto, M. Malandrino, E. Mentasti, C. Sarzanini, F. Petrella, Chemosphere 49 (2002) 545-557.

[23] S.M. Praveena, A. Ahmed, M. Radojevic, M.H. Abdullah, A.Z. Aris, Int. J. Environ. Res. 2 (2008) 139-148.

[24] B. Palumbo-Roe, B. Klinck, J. Environ. Sci. Health. Part A 42 (2007) 1251-1261.

[25] P. Alvarenga, A.P. Gonçalves, R.M. Fernandes, A. de Varennes, G. Vallini, E. Duarte, A.C. Chuna-Queda, Chemosphere 74 (2009) 1292-1300.

[26] C. Sârbu, K. Zehl. J.W. Einax, Chemom. Intell. Lab. Syst. 86 (2007) 121-129.

[27] G.S.R. Krishnamurti, L.H. Smith, R. Naidu, Aust. J. Soil Res. 38 (2000) 823-836.

[28] Y.Tian, X. Wang, J. Luo, H. Yu, H. Zhang, Environ. Sci. Technol. 42 (2008) 7649-7654.

[29] J.M. Alvarez, L.M. Lopez-Valdivia, J. Novillo, A. Obrador, M.I. Rico, Geoderma 132 (2006) 450-463.

[30] P. Illmer, U. Obertegger, F. Schinner, Water Air Soil Pollut. 148 (2003) 3-14.

[31] E. Meers, V.R. Unamuno, G. Du Laing, J. Vangronsveld, K. Vanbroekhoven, R. Samson, L. Diels, W. Geebelen, A. Ruttens, M. Vandegehuchte, F.M.G. Tack, Geoderma 136 (2006) 107-119.

[32] A.L. Juhasz, E. Smith, J. Weber, M. Rees, A. Rofe, T. Kuchel, L. Sansom, R. Naidu, Chemosphere 69 (2007) 69-78.

[33] J.K Yang, M.O. Barnett, P.M. Jardine, N.T. Basta, S.W. Casteel, Environ. Sci. Technol. 36

(2002) 4562-4569.

[34] J.K. Yang, M.O. Barnett, J. Zhuang, S.E. Fendorf, P.M. Jardine, Environ. Sci. Technol. 39 (2005) 7102-7110.

[35] H.C. Dalurzo, S. Vazquez, C. Fernández Lopez, J. Prause, Agrochimica 6 (2007) 319-328.

[36] L. Boszke, A. Astel, Environ. Monit. Assess. 152 (2009) 133-147.

[37] B. Buszewskl, T. Kowalkowski, Environ. Eng. Sci. 23 (2006) 589-595.

[38] J.W. Einax, U. Soldt, Chemom. Intell. Lab. Syst. 46 (1999) 79-91.

[39] E. Heilmann, P. Leinweber, G. Ollesch, R. Meißner, J. Plant Nutr. Soil Sci. 168 (2005) 307-315.

[40] Z. Shi, K. Wang, J.S. Bailey, C. Jordan, A.H. Higgins, Soil Use Manage. 18 (2002) 353-362.

[41] Z.M. Wang, K.S. Song, B. Zhang, D.W. Liu, X.Y. Li, C.Y. Ren, S.M. Zhang, L. Luo, C.H. Zhang, Plant Soil Environ. 55 (2009) 110–120.

[42] N. Moritsuka, J. Yanai, M. Umeda, T. Kosaki, Soil Sci. Plant Nutr. 50 (2004) 565-573.

[43] T. Karyotis, A. Charoulis, T. Mitsimponas, E. Vavoulidou, Commun. Soil Sci. Plant Anal. 36 (2005) 403-414.

[44] M. Jalali, Geoderma 140 (2007) 42–51.

[45] A. Rodríguez, J. Durán, J.M. Fernández-Palacios, A. Gallardo, Geoderma 151 (2009) 303–310.

[46] P. Bengtson, N. Basiliko, C.E. Prescott, S.J. Grayston, Soil Biol. Biochem. 39 (2007) 2429–2435.

[47] A. Gassner, J. Fleckenstein, S. Haneklaus, E. Schnug, Commun. Soil Sci. Plant Anal. 33 (2002) 3347-3357.

[48] Z. Wang, B. Zhang, K. Song, D. Liu, C. Ren, S. Zhang, L. Hu, H. Yang, Z. Liu, Commun. Soil Sci. Plant Anal. 40 (2009) 2389-2412.

[49] A. Paz-González a; M.T. Taboada, Commun. Soil Sci. Plant Anal. 31 (2000) 2135-2146.

[50] A. Rodríguez, J. Durán, J.M. Fernández-Palacios, A. Gallardo, For. Ecol. Manage. 257 (2009) 739–746.

[51] A. Sadeghi, C.D. Graff, J. Starr, G. McCarty, E. Codling, K. Sefton, Soil Sci. 171 (2006) 850-857.

[52] A. Gallardo, F. Covelo, Plant Soil 273 (2005) 269-277.

[53] A. Bekele, W.H. Hudnall, Plant Soil 280 (2006) 7-21.

[54] G.L. Bruland, C.J. Richardson, J. Environ. Qual. 33 (2004) 785–794.

[55] G. Muñoz -Meléndez, A. Korre, S.J. Parry, Environ. Pollut. 109 (2000) 497-504.

[56] J.Shi, J. Xu, P. Huang, J. Soils Sediments 8 (2008) 415-423.

[57] X. Zhang, F. Lin, Y. Jiang, K. Wang, X.L. Feng, Environ. Monit. Assess. 155 (2009) 205-213.

[58] A. Paz González, M.T. Taboada Castro, S.R. Vieira, Can. J. Soil Sci. 81 (2001) 469-479.

[59] L. Boruvka, L. Mladkova, O. Drabek, J. Inorg. Biochem. 99 (2005) 1796-1806.

[60] D.W. Franzen, T. Nanna, W.A. Norvell, Agron. J. 98 (2006) 1015-1022.

[61] G.C. Sigua, W.H. Hudnall, J. Soils Sediments 8 (2008) 193-202.

[62] E.E. Golia, A. Dimirkou, S.A. Floras, Commun. Soil Sci. Plant Anal. 40 (2009) 376-390.

[63] P. Burgos, E. Madejón, A. Pérez de Mora, F. Cabrera, Int. J. Earth Obs. Geoinf. 10 (2008) 11–25.

[64] F. Covelo, A. Rodríguez, A. Gallardo, Plant Soil 311 (2008) 109-119.

[65] C.W. Lin, Environ. Geochem. Health 24 (2002) 23-33.

[66] J. Wu, W.A. Norvell, R.M. Welch, Geoderma 134 (2006) 187– 199.

[67] R. Caridad-Cancela, E. Vidal Vázquez, S.R. Vieira, C.A. Abreu, A. Paz González, Commun. Soil Sci. Plant Anal. 36 (2005) 253-274.

[68] J. Wu, W.A. Norvell, D.G. Hopkins, R.M. Welch, Soil Sci. Soc. Am. J. 66 (2002) 268–275.

[69] C. van Griethuysen, E.W. Meuboom, A.A. Koelmans, Environ. Toxicol. Chem. 22 (2003) 457-465.

[70] A. Castrignanò, L. Giugliarini, R. Risaliti, N. Martinelli, Geoderma 97 (2000) 39–60.

[71] A. Castrignanò, P. Goovaerts, L. Lulli, G. Bragato, Geoderma 98 (2000) 95-113.

[72] I. Snape, R.C. Scouller, S.C. Stark, J. Stark, M.J. Riddle, D.B. Gore, Chemosphere 57 (2004) 491-504.

[73] J.S. Stark, I. Snape, M.J. Riddle, S.C. Stark, Mar. Pollut. Bull. 50 (2005) 276-290.

[74] C. González, J.R. Quintana, L. Moreno, A. Vázquez, A.L. Lafuente, A. Romero, Geoderma 137 (2007) 352-359.

[75] F.E. Rhoton, W.E. Emmerich, D.A. DiCarlo, D.S. McChesney, M.A. Nearing, J.C. Ritchie, Soil Sci. Soc. Am. J. 72 (2008) 1102-1112.

[76] M.R. Cave, J. Wragg, Analyst 122 (1997) 1211-1221.

[77] M.R. Cave, A.E. Milodowski, E.N. Friel, Geochem. Explor. Environ. Anal. 4 (2004) 71-86.

[78] R. Santamaría-Fernández, A. Moreda-Piñeiro, S.J. Hill, J. Environ. Monit. 4 (2002) 330-336.

[79] R. Santamaría-Fernández, M.R. Cave, S.J. Hill, J. Environ. Monit. 5 (2003) 929-934.

[80] B. Palumbo-Roe, M.R. Cave, B.A. Klinck, J. Wragg, H. Taylor, K.E. O'Donnell, R.A. Shaw, Environ. Geochem. Health 27 (2005) 121-130.

[81] J. Wragg, M. Cave, P. Nathanail, J. Environ. Sci. Health Part A 42 (2007) 1303-1315.

[82] R. Santamaría-Fernández, M.R. Cave, S.J. Hill, Anal. Chim. Acta 557 (2006) 344-352.

[83] W. Boonjob, M. Rosende, M. Miró, V. Cerdà, Anal. Bioanal. Chem. 394 (2009) 337-349.

[84] M. Rosende, M. Miró, V. Cerdà, Anal. Chim. Acta 658 (2010) 41-48.

[85] US EPA, Test Methods for Evaluation of Solid Waste, vol. IA. Laboratory Manual Physical/Chemical Methods, SW 846, 40 CFR Parts 403 and 503, third ed., US Government Printing Office, Washington, DC, 1995.

Captions to figures

Fig. 1. Dendrogram obtained by HCA (Ward's method) for pH and element percentages extracted from contaminated soil samples (coded A1-A31) into the third fraction of Tessier's procedure, after column standardization [16].

Table 1
Sequential extraction procedures adopted in the papers cited in this review.

Procedure	Comment	Ref.
Tessier's sequential extraction: exchangeable (1 M $MgCl_2$, pH 7); bound to carbonates (1 M CH_3COONa/ CH_3COOH, pH 5); bound to Fe-Mn oxides (0.04 M $NH_2OH{\cdot}HCl$ in 25% v/v CH_3COOH, 96°C); bound to organic matter (HNO_3/H_2O_2, pH 2, 85°C; 3.2 M CH_3COONH_4 in 20% v/v HNO_3); residual ($HF/HClO_4$). The original reference reports also the possibility to use 1 M CH_3COONa, pH 8.2 and 0.3 M $Na_2S_2O_4$/0.175 M Na-citrate/0.025 M H-citrate for the first and third fractions respectively.	One of the first sequential extraction procedures developed. Most of the other procedures derive from it. It was the most extensively applied scheme before the introduction of the BCR protocol	[16,22,29,55]
Revised BCR sequential extraction: exchangeable, water- and acid-soluble (0.11 M CH_3COOH); reducible (0.5 M $NH_4OH{\cdot}HCl$, pH 1.5); oxidisable (H_2O_2; 1 M CH_3COONH_4, pH 2); residual (aqua regia).	Developed by SMT in order to harmonize fractionation procedures and ensure comparability. It provides a detailed description of operative conditions. For these characteristics it is the most extensively applied sequential extraction technique nowadays	[11,13,15,25,26,29,83]
5-step sequential extraction: exchangeable (1 M CH_3COONH_4, pH 7); bound to carbonates and easily reducible phases (0.6 M HCl, pH 4; 0.1 M $NH_2OH{\cdot}HCl$ in 0.01 M HCl, pH 2); bound to moderately reducible phases (0.2 M $(NH_4)_2C_2O_4$/0.2 M $H_2C_2O_4$, pH 3); bound to organic matter and sulphides (H_2O_2/HNO_3, pH 2, 85°C; 3.2 M CH_3COONH_4); bound to acid-soluble residue (6 M HCl, 85°C).	Based on Tessier's scheme. The differences are: exchangeable fraction following Kersten and Förstner's scheme [see part I]; reagents and definitions of second, third and fifth fractions	[20]
4-step sequential extraction: water-soluble (H_2O); environmentally mobile (1 M CH_3COONH_4, pH 7); bound to carbonates and oxides (1 M $NH_2OH{\cdot}HCl$ in 25% v/v CH_3COOH); organic and refractory-associated (HNO_3/ HF/ H_2O_2).	The reagents and the operational definitions of the fractions substantially differ from the popular Tessier's and BCR schemes.	[14]

Table 1 (continued)

Procedure	Comment	Ref.
5-step sequential extraction: labile inorganic and labile organic (0.5 M $NaHCO_3$); inorganic moderately labile, chemisorbed on Fe, Al and organic moderately labile, chemisorbed on humic acids (0.1 M NaOH); within small stable aggregates, physically inaccessible and within small stable aggregates, physically inaccessible (0.1 M NaOH + sonication); Ca-bound (1 M HCl); residual inorganic (HCl); residual organic (HCl/H_2O_2).	Sequential extraction scheme for P. It is distinctly different from procedures for metals, owing to the anionic nature of P compounds.	[35]
5-step sequential extraction: exchangeable (1 M $Mg(NO_3)_2$); organically bound or associated with organic matter (0.7 M NaOCl); in crystalline Mn oxide or coprecipitated (0.2 M $(NH_4)_2C_2O_4 \cdot H_2O/H_2C_2O_4$); in crystalline Fe oxide or coprecipitated ($Na_2S_2O_4$).	Sequential extraction scheme for Fe, Mn and Al. The reagents and the operational definitions of the fractions substantially differ from the popular Tessier's and BCR schemes.	
4-step sequential extraction : soil solution and labile (water and anion exchange resin); labile, inorganic, organic and microbic (0.5 M $NaHCO_3$, pH 8.2); in humic and fulvic acids and in Al and Fe phosphates (0.1 M NaOH); hardly soluble (1 M H_2SO_4).	Sequential extraction scheme for P. Characterized by the use of an anion exchange resin in addition to chemicals	[39]
4-step sequential extraction : plant-available and water-extractable (H_2O); weakly sorbed-bioavailable organic and inorganic (0.5 M $NaHCO_3$, pH 8.2); strongly bound chemisorbed-potentially bioavailable (0.1 M NaOH); apatite or Ca-bound and non-bioavailable (1 M HCl).	Sequential extraction scheme for P. The reagents are similar to the ones reported in the previously mentioned schemes, but the definitions of the fractions are different	[12]
6-step sequential extraction : organomercury ($CHCl_3$; 0.01 M $Na_2S_2O_3$); water-soluble (H_2O); acid-soluble (0.5 M HCl); associated to humic matter (0.2 M NaOH); elemental (aqua regia, 150°C); residual, HgS (aqua regia).	Sequential extraction scheme for Hg. The reagents and the operational definitions of the fractions substantially differ from the popular Tessier's and BCR schemes	[36]

Table 1 (continued)

Procedure	Comment	Ref.
3-step sequential extraction: exchangeable (1 M CH_3COONH_4); readily soluble nonexchangeable (0.01 M HCl); recalcitrant nonexchangeble (0.2 M sodium tetraphenylborate).	Sequential extraction scheme for K. It is mainly focused on the exchangeability of the element	[42]
Physiologically Based Extraction Test (PBET): simulated stomach conditions (simulated stomach fluid: 1.25 g pepsin, 0.50 g sodium malate, 0.50 g sodium citrate, 420 µL lactic acid and 500 µL acetic acid made up to 1 L with H_2O/HCl, pH 2.5, 37°C); simulated small intestine conditions (pH 7 with $NaHCO_3$, addition of bile salts and pancreatine to the simulated stomach fluid).	Aimed at studying element bioaccessibility. It simulates gastrointestinal tract environment	[80,81]
Modified PBET: stomach phase (simulated stomach fluid: 1.25 g pepsin, 0.50 g sodium malate, 0.50 g sodium citrate, 420 µL lactic acid and 500 µL acetic acid made up to 1 L with H_2O/HCl, pH 2.5, 37°C); small intestine 1 (pH 7 with $NaHCO_3$, addition of 175 mg bile salts and 50 mg pancreatine to the simulated stomach fluid); small intestine 2 (the same as small intestine 1 after an additional 2 hours incubation).	Aimed at studying element bioaccessibility. It differs from the above cited PBET method for the presence of two steps in simulated intestine conditions	[24]
Simplified PBET (SBET): 30.03 g L^{-1} glycine/HCl, pH 1.5, 37°C.	Aimed at studying element bioaccessibility. It considers only one gastric phase.	[32]

Table 2

Selected applications of chemometric techniques to single or sequential extraction results. The papers are arranged in the order in which they have been discussed in the "Application" sections.

Matrix	Elements	Analytical technique	Extraction procedure	Chemometric treatment	Software	Ref.
Agricultural soils with vegetable crops (Lower Vinalopò region, Spain)	Cd, Co, Cr, Cu, Fe, Mn, Ni, Pb, Zn	FAAS, GF-AAS	Single extraction (0.05 M EDTA, pH 7)	Correlation analysis, HCA	SPSS 13.0	[10]
Contaminated soils from an abandoned mining area (Salsigne, France)	As, Cd, Cu, Ni, Pb, Zn	ICP-AES, ICP-MS	BCR	Correlation analysis, PCA, HCA	SPSS 10.0, XLStat 5.2	[15]
Soils from a national database project (Ireland)	K, Mg, P (plus total Al, As, Ba, Ca, Cd, Ce, Co, Cr, Cu, Fe, Ga, Ge, Hg, La, Li, Mg, Mn, Mo, Na, Nb, Ni Pb, Rb, S, Sb, Sc, Se, Sn, Sr, Ta, Th, Ti, Tl, U, V, W, Y, Zn)	ICP-AES, ICP-MS	Single extraction (acetate buffer)	Correlation analysis, HCA	SPSS 14	[19]
Surface river sediments (Louros River, Greece)	P	UV-vis spectro-photometry	4-step sequential extraction, speciation within each extract	PCA, HCA, LDA	SPSS 13.0	[12]
Soils from urban garden (Kayseri, Turkey)	Cd, Co, Cr, Cu, Fe, Mn, Ni, Pb, Zn	FAAS	BCR	Correlation analysis, PCA, HCA	SPSS 9.05	[11]
Soils from around a zinc smelter (Kayseri, Turkey) and grapes	Ca, Cd, Co, Cr, Cu, Fe, Mg, Mn, Ni, Pb, Zn	FAAS	Single extractions (0.1M HCl in 0.025 M H_2SO_4, 1 M CH_3COONH_4, aqua regia)	Correlation analysis, PCA, HCA	SPSS 10.0	[18]

Table 2 (continued)

Matrix	Elements	Analytical technique	Extraction procedure	Chemometric treatment	Software	Ref.
Marine sediments (Terra Nova Bay, Antartica)	Cd, Cr, Cu, Fe, Mn, Ni, Pb, Zn	ICP-AES, GF-AAS	BCR	PCA, HCA	XLStat	[13]
Contaminated soils (Piedmont, Italy)	Al, Cu, Cr, Fe, La, Mn, Ni, Pb, Sc, Ti, V, Y, Zn	ICP-AES	Tessier	PCA, HCA	XLStat	[16]
Agricultural soils (Piedmont. Italy)	Al, Cd, Cr, Cu, Fe, Mn, Ni, Pb, Ti, Zn	ICP-AES, GF-AAS	Single extraction (0.02 M EDTA in 0.5 M CH_3COONH_4), Tessier	PCA, HCA	XLStat	[22]
Contaminated soils (Piedmont, Italy)	Al, Cd, Cu, Cr, Fe, La, Mn, Ni, Pb, Sc, Ti, V, Y, Zn, Zr	ICP-AES, GF-AAS	Single extractions (water,0.5 M CH_3COOH, 0.02 M EDTA in 0.5 M CH_3COONH_4)	PCA, HCA	XLStat	[17]
Mangrove sediments (Mengkabong Lagoon, Sabah, Malaysia)	Al, Ca, Cu, Fe, K, Na, Mg, Pb, Zn	FAAS	Single extraction (Na, K, Ca, Mg: CH_3COONH_4; other elements: aqua regia)	PCA, HCA	Not reported	[23]
Alluvial river sediments (Danube river, Pančevo Oil Refinery, Serbia)	Cu, Fe, Mn, Ni, Pb, Zn,	FAAS	5-step sequential extraction	Correlation analysis, PCA, HCA	SPSS for Windows 10	[20]

Table 2 (continued)

Matrix	Elements	Analytical technique	Extraction procedure	Chemometric treatment	Software	Ref.
Soils with different composition and pollution levels (Flanders, Belgium)	Cd	ICP-AES, GF-AAS	Soil solution, single extractions (0.01 M $CaCl_2$, 0.1 M $Ca(NO_3)_2$, 0.1 M $NaNO_3$, 1 M NH_4NO_3, 1M CH_3CONH_4, 1 M $MgCl_2$, 0.11 M CH_3COOH, 0.1 M HCl, 0.5M HNO_3, 0.02 M EDTA+0.5 M CH_3COONH_4+0.5 M CH_3COOH pH 4.65, 0.005 M DTPA+0.01 M $CaCl_2$+0.1 M TEA pH 7.3, aqua regia)	Correlation analysis, HCA, MLR	SPSS 11.0	[21]
Contaminated soil treated with organic residues (Aljustrel mining area, Portugal)	Cu, Pb, Zn	FAAS, GF-AAS	Single extractions (0.01 M $CaCl_2$ pH 5.7, 0.5 M $CH_3COO\ NH_4$, 0.5 M CH_3COOH, 0.02 M EDTA), BCR	Correlation analysis, PCA, HCA	Statistica 6.0	[25]
Airborne particulate matter, PM_{10} (Santiago, Chile)	Al, As, Ba, Ca, Cd, Cr, Cu, Mg, Mn, Mo, Ni, Pb, Ti, V, Zn	ICP-MS	4-step sequential extraction	HCA	Not reported	[14]
Mine waste-contaminated soils (Devons Great Consols Mine, UK)	Al, As, Ba, Ca, Cr, Cu, Fe, K, Li, Mg, Mn, Na, Ni, P, Pb, S, Sr, V, Zn	ICP-AES	Modified PBET, CISED	CISED, CA, MLR	Not reported	[24]

Table 2 (continued)

Matrix	Elements	Analytical technique	Extraction procedure	Chemometric treatment	Software	Ref.
Contaminated soil profiles (Bad Liebenstein, Thuringia, Germany).	Cd, Co, Cr, Cu, Eu, Fe, Mn, Ni, Pb, Sb, Se, Y, Zn	ICP-AES, ICP-MS	BCR	Fuzzy CA	Not reported	[26]
Agricultural soils (South Australia)	Cd	GF-AAS	Single extractions (0.01 and 0.05 M $CaCl_2$, 0.1 M Na_2EDTA, 0.005 M DTPA-TEA, 1 M NH_4NO_3, 0.02 M AAAC-EDTA, 1 M NH_4Cl	Correlation analysis, MLR	Not reported	[27]
Soils cultivated with rice (Jiangsu province, China)	Cd, Cu, Pb, Zn	ICP-AES, ICP-MS	Single extractions (0.01 M $CaCl_2$, 0.11 M CH_3COOH), soil solution, DGT	PCA, MLR	SPSS 13.0	[28]
Uncontaminated soils (Central Spain)	Mn, Zn	AAS	Single extraction (0.05 M EDTA), modified Tessier, BCR,	Correlation analysis, MLR	Statgraphic Plus 5.0	[29]
Acidic forest soils (Tyrol, Austria)	Al	AAS	1 M HCl	Correlation analysis, FA, MLR	Not reported	[30]
Contaminated and uncontaminated soils and sediments (Flanders, Belgium)	Al, Ca, Cd, Cr, Cu, Fe, K, Mg, Mn, Na, Ni, Pb, Zn	ICP-AES, FES	Soil solution	Correlation analysis, MLR	SPSS 10.0, Excel 9.0, Surfer 6.04	[31]
Contaminated soils (Australia)	As	ICP-AES	SBET	MLR	SPSS 15.0	[32]

Table 2 (continued)

Matrix	Elements	Analytical technique	Extraction procedure	Chemometric treatment	Software	Ref.
Soils spiked with As(V) (USA)	As(V)	AAS	Streamlined PBET	MLR	Not reported	[33]
Soils spiked with As(III) (USA)	As(III)	AAS	Streamlined PBET	MLR	Not reported	[34]
Alfisols, Ultisols, Oxisols soils (Misiones Province, Argentina)	Al, Ca, Fe, Mg, Mn, P	Not reported	P: 11-step fractionation; Al, Fe, Mn: 5-step sequential extraction; single extractions (Al: KCl; Ca and Mg: CH_3COONH_4)	FA, LDA	Not reported	[35]
River sediments and floodplain soil (Warta River, Poland)	Hg	CV-AFS	6-step sequential extraction for Hg	Correlation analysis, ANN	Statistica 6.0	[36]
Soil profiles (Torun, Poland)	Cd, Ni, Pb	FAAS	Column leaching with modeled acid rain	Correlation analysis, ANN	Statistica 6-0	[37]
Agricultural soils (Harz mountains, northeastern Germany)	K, Mg, P	ICP-AES	4-step sequential extraction (P), double lactate extraction (K, Mg, P)	Correlation analysis, geostatistics	Surfer	[39]
Soils from grassland fields (Northern Ireland)	K, Mg, N, P	ICP-AES	Single extractions (P: $NaHCO_3$; K: 1 M CH_3COONH_4; S: $Ca(H_2PO_4)_2$)	Geostatistics	GS^+	[40]

Table 2 (continued)

Matrix	Elements	Analytical technique	Extraction procedure	Chemometric treatment	Software	Ref.
Agricultural soils (northeast China)	K, N, P	Colorimetry	Single extractions (K: 1 M CH_3COONH_4)	Correlation analysis, geostatistics	GS^+, GIS ArcView 3.2, Spatial Analyst	[41]
Soils from rice fields (Takatsuki City, Japan)	K, N, P	FES, colorimetry	3-step sequential extraction (K), water and Bray method (P), single extractions (N: 2 M KCl)	Correlation analysis, geostatistics	GS^+	[42]
Soils (Philippoi peatland, northern Greece)	Ca, Cu, Fe, K, Mg, Mn. Na, Zn	AAS	Single extraction (Fe, Cu, Zn, Mn: 0.005 M DTPA; Ca, K, Mg, Na: 1 M CH_3COONH_4)	Geostatistics	Not reported	[43]
Soil (Field Research Station, Dastjerd, western Iran)	K	Not reported	Single extraction (CH_3COONH_4), successive extractions (0.01 M $CaCl_2$)	Geostatistics	GS^+	[44]
Forest soil (La Palma Island, Canary Islands, Spain)	N, P	Colorimetry	Single extractions (N: 0.5 M K_2SO_4; P: 0.5 M $NaHCO_3$)	Geostatistics	R 2.7.2 for Linux	[45]
Forest soils (coastal British Columbia, Canada)	N, P	Colorimetry	Single extraction (N: chloroform), ion exchange resin probe	Geostatistics	GS^+	[46]

Table 2 (continued)

Matrix	Elements	Analytical technique	Extraction procedure	Chemometric treatment	Software	Ref.
Agricultural cambisol (Kassow, Germany)	P	Colorimetry	Single extractions (0.01 M $CaCl_2$, 0.5 M CH_3COONH_4/0.5 M CH_3COOH/0.02 M Na_2-EDTA; 0.43 M HNO_3; aqua regia; 0.1M Ca-lactate/0.1 M Ca-acetate/ 0.3M CH_3COOH)	Correlation analysis, geostatistics	Variowin 2.2	[47]
Agricultural soils (northeast China)	K, N, P	Colorimetry	Single extractions (K: 1M CH_3COONH_4)	Correlation analysis, geostatistics	GS^+, ArcGIS	[48]
Agricultural and forest soils (Lugo Province, Spain)	Ca, Cu, Fe, K, Mg, Mn, Zn	FAAS	Single extraction (Ca, K, Mg: CH_3COONH_4; Cu, Fe, Mn, Zn: 0.05 M EDTA)	Geostatistics	Computer programs developed by Vieira	[49]
Forest soil (La Palma Island, Canary Islands, Spain)	N, P	Colorimetry	Single extractions (N: 2 M KCl; P: 0.5 M $NaHCO_3$)	Geostatistics	R 2.4.1 for Linux	[50]
Agricultural soils (coastal plain region of Maryland)	P	-	Single extraction (Mehlich III extractant: 0.2 M CH_3COOH, 0.25 M NH_4NO_3, 0.015 M NH_4F, 0.013M HNO_3, 0.001M EDTA)	Geostatistics	Vesper, Surfer	[51]

Table 2 (continued)

Matrix	Elements	Analytical technique	Extraction procedure	Chemometric treatment	Software	Ref.
Forest soils (northwestern Spain)	N, P	Colorimetry	Single extraction (N: 2 M KCl; P: 2.5 % CH_3COOH)	Correlation analysis, geostatistics	R	[52]
Forest soils (Kitsatchie National Forest, Winn Parish, Louisiana)	Ca, Mg, K, Fe, Mn	ICP-AES	Single extraction (Mehlich III extractant: 0.2 M CH_3COOH, 0.25 M NH_4NO_3, 0.015 M NH_4F, 0.013M HNO_3, 0.001M EDTA)	Correlation analysis, geostatistics	GS^+	[53]
Forest soil (North Carolina)	Al, Fe, P	Not reported	Single extraction (Al, Fe: oxalate)	Correlation analysis, geostatistics	GS^+	[54]
Soil from a landfill (Bedfordshire,UK)	Cr, Cu, Zn	ICP-AES	Tessier	Correlation analysis, geostatistics	Variowin 2.2	[55]
Soil (northern Zhejiang Province, around Taihu Lake, China)	B, Cu, Fe, Mn, Mo, Zn	ICP-AES	Single extraction (DTPA)	Geostatistics	GS^+, ArcGIS	[56]
Surface soils from seven land uses. (Fuyang County, China)	Cu	AAS	Single extraction (DTPA, $CaCl_2$, TEA)	Correlation analysis, geostatistics	GS^+, ArcGIS	[57]
Forest soils (La Coruña, Spain)	Co, Cr, Fe, Mn, Ni, Zn	ICP-MS	Single extraction (0.05 M EDTA)	Correlation analysis, geostatistics	KRIGE, COKRI, Surfer	[58]

Table 2 (continued)

Matrix	Elements	Analytical technique	Extraction procedure	Chemometric treatment	Software	Ref.
Forest soils (Jizera Mountains, Bohemia, Czeck Republic)	Al	ICP-AES	Single extraction (0.5 M KCl, 0.05M $Na_4P_2O_7$)	Correlation analysis, geostatistics	GS^+,Variowin 2.21	[59]
Agricultural soils (North Dakota)	B, Cd, Cu, Se, Zn	AAS, ICP-AES, HG-AAS, colorimetry	Single extraction (Cd, Cu, Zn: DTPA; B, Se: H_2O)	Geostatistics	GS^+	[60]
Surface soil samples (Luisiana)	Al, Ca, Fe, Mg, Na, S,	ICP-AES	Single extraction (water)	Geostatistics	Not reported	[61]
Agricultural soils (central Greece)	Cd, N	GF-AAS	Single extractions (Cd: DTPA)	Correlation analysis, geostatistics	ArcInfo	[62]
Soil affected by the Aznancóllar mine spill (Spain)	As, Cd, Cu, Pb, Zn	Not reported	Single extraction (0.05 M EDTA)	Correlation analysis, geostatistics	Vesper 1.6	[63]
Forest soil (northwestern Spain)	N, P	Colorimetry	Single extraction (N: 1 M KCl; P: 2.5% CH_3COOH)	Correlation analysis, geostatistics	R 1.8 for Linux	[64]
Contaminated soils (Nantou Country, Taiwan)	Pb	AAS	Single extraction (1 M HCl)	Geostatistics	Surfer	[65]
587 soils (North Dakota)	Zn	AAS	Single extraction (DTPA)	Correlation analysis, geostatistics	GSLIB	[66]
Cultivated soils (Galicia, Spain)	Cu, Mn, Pb, Zn	ICP-AES	Single extractions (0.05 M EDTA and $CaCl_2$)	Geostatistics	GSTAT, PCRaster	[67]

Table 2 (continued)

Matrix	Elements	Analytical technique	Extraction procedure	Chemometric treatment	Software	Ref.
Grain and agricultural soil (North Dakota)	Cd, Na, S	ICP-AES, IC	Single extraction (Cd and Na: 0,005 M DTPA; S: H_2O)	Geostatistics	GS^+	[68]
Lake sediments (The Netherlands)	Al, Ca, Cd, Cr, Cu, Fe, Mn, Ni, P, Pb, S, Zn	ICP-AES, ICP-MS	Single extraction (1 M HCl, expressed as SEM, see text)	Correlation analysis, geostatistics	WLSFIT, Surfer	[69]
Agricultural soils from an experimental area (Italy)	K, Na, P	Not reported	Single extractions (K, Na: CH_3COONH_4; P: $NaHCO_3$);	Correlation analysis, geostatistics, FKA	Not reported	[70]
Soil from an experimental area (Volperino, Italy)	Fe, Mn	ICP-AES	Single extraction (Fe and Mn:0.005 M DTPA/0.01 M CaCl/ 0.1 M tetraethylammonium	Correlation analysis, geostatistics, FKA	Not reported	[71]
Marine sediments (Antarctica)	Ag, As, Cd, Cr, Cu, Fe, Mn, Ni, Pb, Sn, Zn	ICP-MS	1 M HCl	nMDS, ANOSIM, SIMPER, RELATE	PRIMER 5.2.2	[72]
Marine sediments adjacent to an abandoned waste dump, (Brown Bay, Casey Station, Antarctica)	Ag, As, Cd, Cr, Cu, Fe, Mn, Ni, Pb, Sb, Sn, Zn	ICP-MS	1 M HCl	nMDS, ANOSIM, BIOENV procedure	PRIMER	[73]
Calcaric Fluvisols (Spain)	Ca, K, Mg, Na, Si (+ anions)	FAAS, FAES	Column leaching with water	Correlation analysis, RDA	Canoco 4.5	[74]

Table 2 (continued)

Matrix	Elements	Analytical technique	Extraction procedure	Chemometric treatment	Software	Ref.
Soils and suspended sediments (Walnut Gulch Experimental Watershed, Arizona, USA)	Ca, Fe, K, Mg, Mn, Na, ^{137}Cs, ^{40}K ^{226}Ra	AAS	Single extractions (Fe and Mn: $Na_4P_2O_7$, acid ammonium oxalate, sodium citrate–bicarbonate–dithionite; Ca, K, Mg, Na: CH_3COONH_4)	Model	Not reported	[75]
Floodplain sediments (Bad Liebenstein, Thuringia)	Cd, Co, Cu, Cr, Eu, Fe, Mn, Ni, Pb, Sb, Se, Y, Zn	Not reported	Four sequential extractions	PCA, Projection pursuit	Not reported	[8]
NIST SRM 2710 (contaminated soil in the Butte, Montana area)	Al, Ba, Ca, Cd, Cu, Fe, K, Mg, Mn, Na, Ni, P, Pb, S, Si, Sr, Ti, V, Zn	ICP-AES	CISED	CISED	Matlab	[77]
NIST SRM 2710 (contaminated soil in the Butte, Montana, area)	Al, Ba, Ca, Cd, Cu, Fe, K, Mg, Mn, Na, Ni, P, Pb, Si, Ti, V, Zn	ICP-AES	CISED	CISED, experimental design	Statistica 5.1, MathCadPlus	[76]
NIST SRM 2710 (contaminated soil in the Butte, Montana, area)	Al, Ba, Ca, Cd, Cu, Fe, K, Mg, Mn, Na, P, Pb, S, Si, Sr, Ti, V, Zn	ICP-AES	CISED	CISED, experimental design	Not reported	[78]
NIST SRM 2711 (moderately contaminated agricultural soil) and river sediment (Carnon River, Cornwall, UK)	Al, Ba, Ca, Cd, Cu, Fe, K, Mg, Mn, Na, P, Pb, S, Si, Sr, Ti, V and Zn	ICP-AES	CISED	CISED	Not reported	[79]

Table 2 (continued)

Matrix	Elements	Analytical technique	Extraction procedure	Chemometric treatment	Software	Ref.
Ironstone-derived soil from the archive of British Geological Survey	Al, As, B, Ba, Ca, Cd, Co, Cr, Cu, Fe, K, Li, Mg, Mn, Mo, Na, Ni, P, Pb, S, Se, Si, Sr, V, Zn	ICP-AES	CISED, PBET	CISED	Not reported	[80]
Ironstone-derived soil (Wellingborough, Northamptonshire, UK)	Al, As, Ba, Ca, Co, Cr, Cu, Fe, K, Mg, Mn, Na, Ni, P, Pb, S, Se, Si, Sr, Ti, Zn	XRF, ICP-AES	CISED, PBET (As)	CISED, MLR	Not reported	[81]
Estuarine sediments (Arosa Estuary, Galicia, Spain)	Al, Ba, Ca, Cd, Cu, Fe, K, Mg, Mn, Na, P, Pb, S, Si, Sr, Ti, V, Zn	ICP-AES	CISED	CISED	Not reported	[82]
NIST SRM 1633b (coal fly ash)	Cr, Cu, Ni, Pb, Zn	ICP-AES	BCR (sequential injection dynamic extraction)	Experimental design	Statgraphics	[83]
Municipal solid waste incineration bottom ashes (Palma de Mallorca, Spain)	Cd, Cr, Cu, Pb, Zn	ICP-AES	TCLP (column dynamic extraction)	Experimental design	Statgraphics	[84]

List of abbreviations. AAS: atomic absorption spectroscopy (the atomizer was not indicated); CV-AFS Cold Vapour Atomic Fluorescence Spectroscopy; DGT: Diffusive Gradients in Thin Films; DPASV: Differential Pulse Anodic Stripping Voltammetry; FA: Factor Analysis; FAAS: Flame Atomic Absorption Spectroscopy; FES: Flame Emission Spectroscopy; GF-AAS: Graphite Furnace Atomic Absorption Spectroscopy; HG-AAS: Hydride Generation Atomic Absorption Spectroscopy; IC: Ion Chromatography; ICP-AES: Inductively Coupled Plasma Atomic Emission Spectroscopy; ICP-MS: Inductively Coupled Plasma Mass Spectrometry; PBET: Physiologically Based Extraction Test; XRF: X Ray Fluorescence.

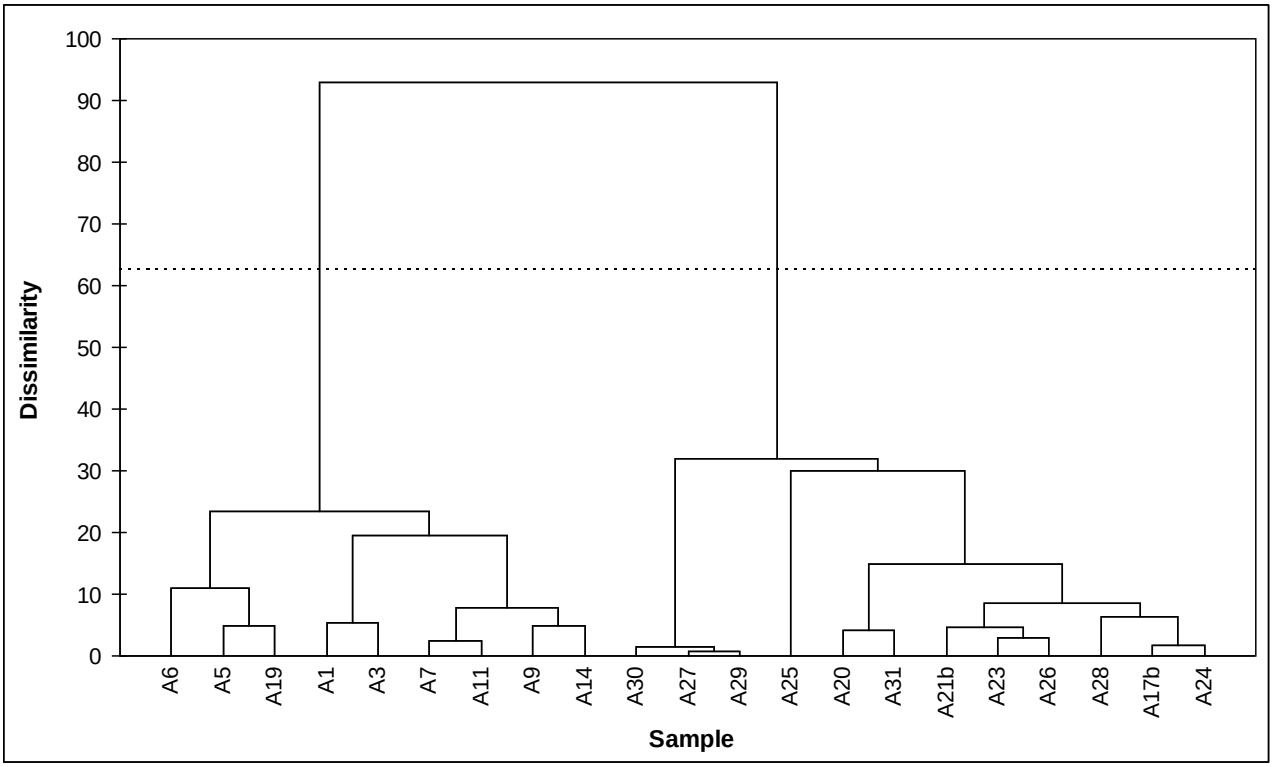
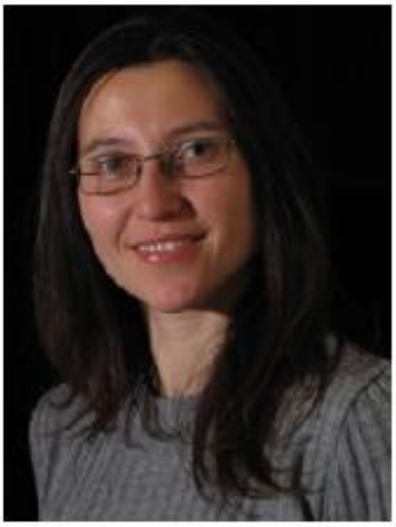

Dissimilarity
100
90
80
70
60
50
40
30
20
10
0
A6
A5
A19
A1
A3
A7
A11
A9
A14
A30
A27
A29
A25
A20
A31
A21b
A23
A26
A28
A17b
A24
Sample

Fig. 1



Agnese Giacomino graduated in Chemistry in 2003 at the University of Torino. She received her PhD in Chemical Sciences in 2007 at the University of Torino. Now she is holder of a research grant at the Department of Analytical Chemistry of the University of Torino. Her research activities are finalized to study the behaviour of metals in different natural matrices (soils, sediments, vegetables, seawater...), using chemometrics for data processing, to characterize the composition of pharmaceutical formulations, to develop new procedures of remediation of contaminated soils, to study analytical methods for the determination of trace metals.

Ornella Abollino received her PhD in Chemical Sciences in 1991 from the University of Torino. She is presently Associate Professor of Analytical Chemistry at the Faculty of Pharmacy, University of Torino. Her research activities are mainly focused on the following topics: development of voltammetric and spectroscopic procedures for the determination and speciation of trace metals; study of element fractionation in sediment and soils from remote and anthropized areas; characterization of metal content in pharmaceutical formulations; interaction between trace metals and plants; application of chemometric techniques for the processing of experimental results related to the above mentioned matrices.

Mery Malandrino received her PhD in Chemical Sciences in 2001 at the University of Torino. She is presently Researcher of Analytical Chemistry at the Faculty of Sciences of the University of Torino. She is mainly involved in the development of sensitive analytical procedures capable of characterising complex environmental matrices, such as soils, sediments and atmospheric particulate matter. She uses chemometrics to interpret experimental results. The purpose of her studies is to gain insight into the behaviour of elements in uncontaminated ecosystems and their influence on climate changes. Furthermore, she studies the development of eco-compatible decontamination procedures for polluted soils.

Edoardo Mentasti has been full professor of Analytical Chemistry at the Faculty of Sciences of the University of Torino from 1980 to 2007 and is now retired. He has been Editor-in-chief of the Journal Annali di Chimica from 1996 to 2006. His main research interests are the development of preconcentration and speciation procedures for trace metal ions coupled to determination by atomic spectroscopy, the development of voltammetric methods of analysis, the characterization of environmental compartments (seawaters, lacustrine ecosystems, sediments, soils) with the aid of chemometric techniques, the interaction between trace metals and clays, the remediation of contaminated soils.