ANALYTICA CHIMICA ACTA 603 (2007) 20–29

available at www.sciencedirect.com

ScienceDirect

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

3-Way characterization of soils by Procrustes rotation, matrix-augmented principal components analysis and parallel factor analysis

J.M. Andrade[a,*], *M. Kubista*[b], *A. Carlosena*[a], *D. Prada*[a]

[a] *Department of Analytical Chemistry, University of A Coruña, Campus da Zapateira s/n, E-15071 A Coruña, Spain*
[b] *Institute of Molecular Genetics, Videnska 1083, 142 20 Prague 4, Czech Republic*

ARTICLE INFO

Article history:
Received 20 June 2007
Received in revised form
3 September 2007
Accepted 20 September 2007
Published on line 29 September 2007

Keywords:
Soil
Heavy metals
Procrustes rotation
Matrix-augmented principal component analysis
Parallel factor analysis

ABSTRACT

Three different approaches for 3-way analyses, namely, Procrustes rotation, parallel factor analysis (PARAFAC) and matrix-augmented principal component analysis, have been compared considering a four-seasons study on soil pollution. Each sampling season comprised 92 roadsoil samples and 12 analytical variables (heavy metals, loss on ignition, pH and humidity). Results show that the three chemometric techniques lead to essentially the same conclusions. Hence, Procrustes rotation, a mathematical technique scarcely applied in analytical chemistry, revealed as a useful tool for 3-way data analysis with potential advantages, including its conceptual simplicity and straightforward interpretation of the results. A novel application of the consensus vectors allowed definition of "consensus scores" so that visualization of the samples and temporal patterns can be made. Results also suggested that the trilinearity assumption imbedded in PARAFAC is essentially fulfilled when studying the temporal evolution of an environmental system where no new pollution sources appear during the course of the study.

1. Introduction

Despite not being so visible for current citizens as air and water pollution, soil contamination is an important topic in today enviromental protection and remediation. Not only industries affect the soil where they are active (see e.g. [1] for an example) but also traffic, heating, etc. The latter are examples of diffuse pollution which affect the surroundings of any city. In addition to well-known smog-related events, increasing amount of suspended particles into the air, noise, CO_2, combustion-related hydrocarbons and heavy metals are of major concern. They can deposit onto the soils directly (e.g. from car exhausts) or more indirectly with rainfall, dust, deposition of other particles, etc.

These effects are particularly important where gardens and parks are used for leisure and kids play. Heavy metals can be ingested either by soil dust inhalation and through the food chain, as a result of their uptake by plants. Some studies have shown that a potentially significant source of lead intake is childrens' play-grounds in urban communities through hand to mouth contact, which is typical for children aged 1–3 years [2]. Accordingly, heavy metals monitoring in soils is an ongoing topic in environmental studies [3] even though extraction procedures and method validation can be cumbersome, as it has recently been pointed out [4,5].

Intimately intertwined to environmental monitoring is the data treatment issue. Many public bodies develop monitoring programs with sampling seasons extended over time.

* *Corresponding author*. Fax: +34 981 167065.
E-mail address: andrade@udc.es (J.M. Andrade).
0003-2670/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.aca.2007.09.043

This raises the question on how to treat such data. Today powerful chemometric tools are available to handle these so-called *N*-way data sets, many of them extensively described and exemplified in the classical text from Smilde and Co. [6]. These methods, despite being conceptually complex, represent a solution to many environmental efforts carried out nowadays. Although an extensive review is out of the scope of this paper, some recent studies can be cited: herbicides and some of their derivatives in US water reservoirs, using MA-PCA (matrix-augmented principal component analysis) and multivariate curve resolution [7] were ascertained; physicochemical parameters in rivers with different anthropogenic inputs [8] using PARAFAC, MA-PCA and factor analysis; residues of oil spills in soils by PARAFAC (parallel factor analysis) [1]; changes on water quality with time, location and season in 4-way data sets analyzed by PARAFAC and Tucker 3 [9]; heavy metals in an industrialized German area by Tucker 3 methods generalized to 4-way data [3].

2. Experimental

2.1. Samples

Four samplings (one per each annual season) were scheduled on a medium-size city (A Coruña, Galicia, NW Spain, aprox. 500,000 inhabitants in the metropolitan area), to assess the levels of nine heavy metals (Cd, Co, Cu, Cr, Fe, Mn, Ni, Pb and Zn) on public gardens, surroundings of the main road accessing the city (aprox. 100,000 vehicles per day) and a highway (see Fig. 1). In addition, three typical physicochemical parameters (humidity, pH and loss on ignition – LOI) were determined. Each sampling was performed on 92 sampling points, where composite samples were taken (0–5 cm in depth) with plasticware. In the Spanish AP9 highway several roadside soils and perpendicular transects were considered on the way from Coruña to Santiago de Compostela (approx. 56 km): samples 1, 6–10, and 15–18 were collected at roadsides; samples 2–5 and

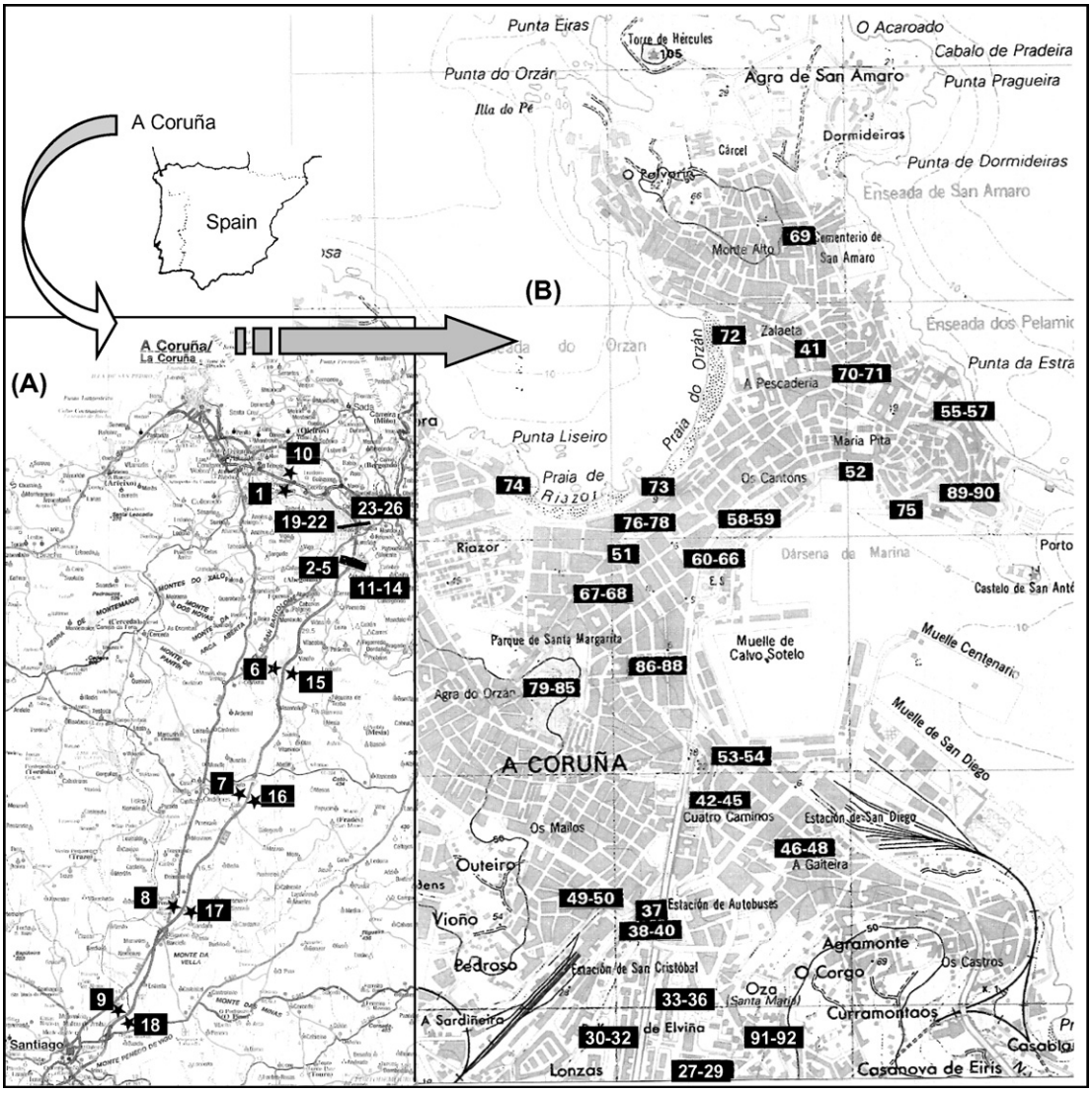

Fig. 1 – General location of the two areas under study and sampling points: (A) AP9 highway and (B) city gardens and main avenue.

11–14 are from uncultivated transects, and samples 19–26 are from transects with cultivated fields (see inset 'a' in Fig. 1). Within the city, samples 27–36 were from the roadside border of the main avenue (samples 30 and 31 are two transects on uncultivated fields) and samples 37–92 from gardens (slots, gardens, parks, etc.), see inset 'b' in Fig. 1.

2.2. Analytical procedure

Samples were air-dried, ground and sieved through a 2 mm mesh sieve. An aliquot of this fraction was used to determine humidity by heating it at 105 °C until constant weight, pH (1:2.5 in Milli-Q water, Millipore Corp.) and organic material as loss on ignition (450 °C for 6 h). The remaining part of the <2 mm fraction was heated at 60 °C for 48 h and sieved again to <0.2 mm. This fraction was used to measure the metals. 0.3000 g aliquots were extracted with HNO_3(c) (Merck, Suprapur) using teflon closed vessels and microwave heating. Metals were determined by flame- (Cu, Cr, Fe, Mn, Pb and Zn) and graphite-furnace-atomic absorption spectrometry (Cd, Co, Cr and Ni) in a 2380 Perkin-Elmer FAAS, and a 4100 Perkin-Elmer Graphite Furnace devices, respectively. Trueness was checked using the BCR-CRM 141 (calcareous loam soil) and BCR-CRM 277 (estuarine sediment) certified reference materials. More details can be found elsewhere [10,11].

2.3. 3-Way chemometric methods

2.3.1. Matrix-augmented principal component analysis

Developed by Tauler et al. [7] matrix-augmented PCA (MA-PCA) constitutes a straightforward extention of traditional PCA. The 3-way data set which is initially a data cube (samples –'rows'- × analytical variables –'columns'- × samplings –'tubes or slices' – [6]) is reordered into an extended matrix. This is called 'unfolding' of the slices, which can be done in different ways. In many cases it is preferable to unfold column-wise, i.e. maintaining the variables in columns. In our study this implies transforming a $93 \times 12 \times 4$ data set (samples × variables × samplings, $n \times p \times m$) into a $(93{\cdot}4) \times 12$ or 372×12 matrix. Then traditional PCA yields sample-related (scores) and variable-related (loadings) information. It is clear that some information is lost during the matrix augmentation, for instance, the correlation between the variables is not taken into account [12]. The nice idea put forward by Taluer and coworkers was to refold the scores matrix again after the PCA. Thus, the augmented matrix can be decomposed as $\mathbf{X}^{\text{aug}}_{(n\cdot m\times p)} = \mathbf{S}^{\text{aug}}_{(n\cdot m\times k)} \cdot \mathbf{L}'_{(k\times p)} + \mathbf{E}^{\text{aug}}_{(n\cdot m\times p)}$, where $\mathbf{X}^{\text{aug}}$ is the augmented data matrix, $\mathbf{S}^{\text{aug}}$ is the augmented scores matrix and $\mathbf{L}'$ is the $(k \times p)$ loadings matrix (the prime (′) denoting transposed) describing the composition of the k principal components (pollution sources); n = number of samples, m = number of samplings (slices), k = number of components and $\mathbf{E}$ = error matrix of residuals due to only considering k components. Clearly, the augmented scores matrix mixes information about location and time evolution. Refolding each augmented scores vector, e.g. $\mathbf{s}^{\text{aug}}_{(n\cdot m\times 1)}$ (see Refs. [7,8] for more details) to $\mathbf{R}_{(n\times m)}$, and averaging row-wise and column-wise leads to two vectors $\mathbf{r1}_{(n\times 1)}$ and $\mathbf{r2}_{(1\times m)}$. These contain the location- and time-related information, respectively. Repeating this process for each augmented scores vector, information can be obtained on how each "pollution source" influences (on average) each sampling site/sample (**r1**) and its average evolution on time (**r2**). MA-PCA has empirically been shown to give comparable results to standard PARAFAC [8].

2.3.2. Procrustes rotation

Generalized Procrustes rotation (PR) is a multivariate technique developed in the 1970s to simultaneously compare several data sets. It is based on traditional singular value decomposition (svd) to decompose a matrix into principal components, $\mathbf{X}_{(n\times p)} = \mathbf{A}_{(n\times k)} \cdot \mathbf{B}_{(k\times k)} \cdot \mathbf{L}'_{(k\times p)} + \mathbf{E}_{(n\times p)}$ (the scores matrix $\mathbf{S} = \mathbf{A}{\cdot}\mathbf{B}$). Although quite successful in biometrics, forensics and psychology, it has scarcely been applied in analytical chemistry. This is surprising because of its simple, user-oriented and straightforward fundamentals. As for the other techniques presented here, not all mathematical details are given, just some brief conceptual ideas. Interested readers are encouraged to consult the selected references. Procrustes rotation has been extensively developed and explained by Krzanowski [13] and resumed elsewhere [14,15].

The main idea of PR is to compare two or more spaces where the same variables are measured. It is advisable to compare principal components scores subspaces to avoid unstructured and random variation in the original data that may blur the general patterns, to reduce data dimensionality and because the main patterns within the datasets can be compared directly. Only the most important PCs shall be used in a PR comparison. Their number can be determined by several established tests [16], Malinowski's test [17], Wold's F-test [18] or the Wm statistic [19].

Generalized Procrustes rotation aims to compare m scores subspaces ($m = 1, \ldots, m$) by calculating a new set of factors or consensus vectors, $\mathbf{v}$, that resemble all scores subspaces. Their dimension is (1xp), since they are defined in the original p-dimensional data space. Accordingly, a first consensus vector, $\mathbf{v}_1$, has to be defined that is close to the first principal component of all M subspaces. Since a reasonable way to define "closeness" is by the cosine squared of the angle between the consensus vector and the other principal component, $\sum_{m=1}^{M} \cos^2 \alpha_{k,m}$ is maximized for each k (cardinality of the principal components). Krzanowski proved [13] that the eigenvector $\mathbf{v}_1$ corresponding to the largest eigenvalue of $W = \sum_{m=1}^{M} L'_m L_m$ fulfills those conditions. $\mathbf{L}_m$ being the $(k \times p)$ loadings matrix for data set m.

The vector $\mathbf{v}_1$ can be thought of as an average factor of all m first principal component scores. The deviation of this average factor from the first PC of a given set m is given by the angle $\alpha_{1,m} = \cos^{-1}\{(v_1{}'L_m{}'L_m v_1)^{1/2}\}$. Analogously, $\mathbf{v}_2$ is the consensus vector that corresponds to the second-largest eigenvalue of $\mathbf{W}$. $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal, which may simplify the chemical interpretation of the consensus vector. $\alpha_{2,m}$ is a measure of the difference of the second consensus vector from the second PC of set m. The process continues until k consensus vectors are obtained. Although not specifically developed for 3-way analysis, experience with PR applications in different fields [20–22] impelled us to compare its results with other more well-established 3-way techniques. In addition, like MA-PCA, PR can handle not-so-unfrequent situations when a sample was not analyzed in a given season (this is not a missing data

situation, rather all information of a sample is missing on a particular season); i.e. the slices do not have the same number of samples (rows). PR's main drawback is that, so far, it can only be applied to 3-way data.

In this paper, a novel application to visualize sample patterns after the calculation of the consensus vectors is presented. To the best of the authors' knowledge, this represents a novel application of PR. Once matrix **V** (containing the consensus vectors) is calculated, 'consensus scores' are derived for each original space ('slice' or sampling season) as $\mathbf{C} = \mathbf{X} \cdot \mathbf{V}$ and, visualized as a 'consensus scatterplot'. More graphics can be derived: (i) a comparison of the original scores plots and the consensus scores for each season; and (ii) a simultaneous comparison of all consensus scatterplots. Since conclusions drawn from those graphs may have a subjective contribution, a natural way to organize and resume such wealth of information is to take averages, in the same way as done in MA-PCA. Hence, for each consensus vector, 'season-averaged consensus scores' and 'site-averaged consensus scores' can be derived.

2.3.3. *Parallel factor analysis*

Parallel factor analysis, PARAFAC (also called trilinear decomposition [23,6]) can be introduced as a generalization of singular value decomposition to include the third way. In our empirical experience, results of PARAFAC are almost the same as those obtained with PR and it remains to be mathematically shown how results of these methods differ. Here, we will present only an empirical comparison. Below is a brief description of PARAFAC. For more details, readers are higly encouraged to refer to Smilde et al. [6].

In the same manner as a 2-way matrix (e.g. samples *x* variables) is decomposed using svd, $\mathbf{X}_{(n\times p)} = \mathbf{A}_{(n\times k)} \cdot \mathbf{B}_{(k\times k)} \cdot \mathbf{L}'_{(k\times p)} + \mathbf{E}_{(n\times p)}$ and, then, rearranged into two sets of vectors (scores and loadings matrices) $\mathbf{X}_{(n\times p)} = \mathbf{S}_{(n\times k)} \cdot \mathbf{L}'_{(k\times p)} + \mathbf{E}_{(n\times p)}$, a 3-way data array can be decomposed into three matrices $\underline{\mathbf{X}}_{(n\times p\times m)} = \mathbf{A}_{(n\times k)} \cdot \mathbf{B}_{(p\times k)} \cdot \mathbf{C}_{(m\times k)} + \mathrm{E}_{(n\times p\times m)}$. The indexes have the same meaning as above (samples, variables, sampling seasons, respectively) and underscore ($\underline{\mathbf{X}}$) is used to denote 3-way data matrices. It has been stated that, in many cases, PARAFAC is not simple to use although the results may be easy to interpret and typically lead to directly relevant conclusions [12]. PARAFAC decomposition yields three matrices. One is analogous to the traditional scores matrix and two are analogous to the current loadings matrix. Nevertheless, mathematically the differentiation is arbitrary and most authors consider the three matrices as weights. As for PR, PARAFAC requires that all 'slices' (seasons) are decomposed into the same number of components. This means that the same two matrices (e.g. **A** and **B**) will be used to model each slice (sampling season) albeit with weights given by matrix **C** (each slice/season will have its own set of weights) [6]. These weights resemble very much the consensus vectors obtained by PR (although there is a mathematical ambiguity in PARAFAC weights [6]).

It is worth comparing how many parameters (weights or scores and loadings) are required to fit a model by each technique. Let us assume that *k* factors are used to describe the dataset. If we perform independent PCAs on each dataset (Procrustes rotation), we fit m*(k*n+k*p) parameters; in our case (the indexes have the same meanings as above) 4*(2*93+2*12)=840, with $k = 2$. In MA-PCA, the number of fitted parameters is k*(m*n)+k*p, or 2*(4*93)+2*12=768. In PARAFAC, they are k*n+k*p+k*m, or 2*(93+12+4)=218. Hence, the more complex the model, the less degrees of freedom and the more constrained is the solution (e.g. trilinearity requirements in PARAFAC). Hopefully, the more constrained solution is, the simpler to interpret. Thus, it is interesting to assess whether more constrained PARAFAC yields as good interpretations as PR and MA-PCA. This in turn indicates the validity of the underlying trilinearity assumption of PARAFAC when applied

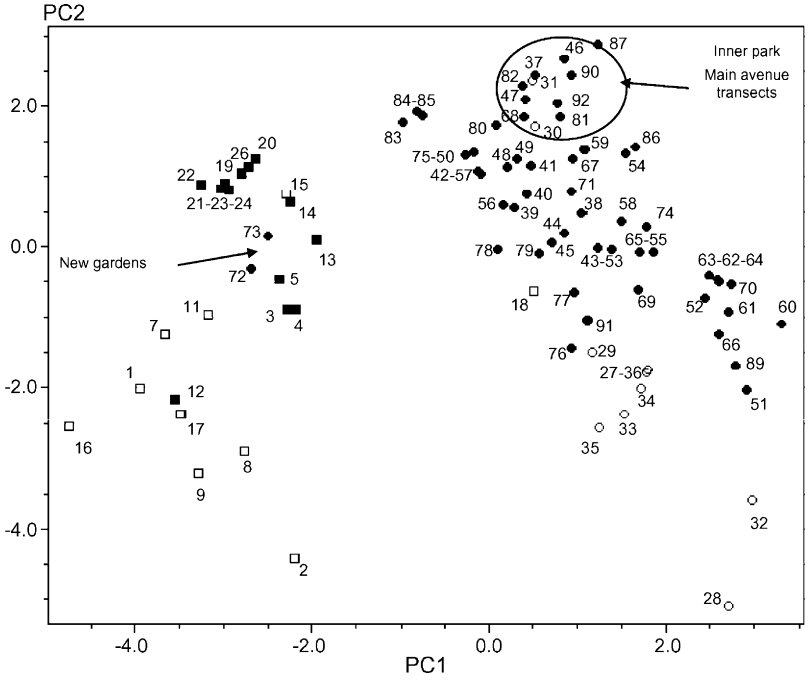

Fig. 2 – PC1-PC2 scores subspace for the autumn data. Highway (□), highway transects (■), main avenue (○), city gardens (●).

to the environmental datasets collected over time (in our case, without any new known pollution events/sources).

3. Results and discussion

Preliminary studies revealed that three samples behaved systematically different from the bulk. They were considered outliers and discarded. Two of them were located at the roadside border of the highway and, despite that part suffered only ca. 10,000 vehicles per day (vpd), they showed very high concentrations of Cu, Cr, Pb and Zn. Close study of the area revealed that the surface of the soil was very compact, without repair or gardening works and, so, the metals accumulated onto the surface. Something similar occurred for the other sample, although this corresponded to a city slot without gardening works.

Separate principal components analysis (PCA) for each sampling season (results not detailed) showed that two components (PCs) accounted for approx. 80% of all variance and that two main patterns were present. PC1 differentiated between two major groups of samples: city gardens and highway. PC2 subdivided each major group into two subgroups: highway transects (located several meters off the road) and highway roadsoils, on the one hand, and between the city gardens and the main avenue of the city, on the other hand. Results for the first sampling season (autumn) are resumed in Fig. 2 as an example (see [15] for details). All calculations are based on column-wise autoscaled data because of the different units of the analytical variables. In MA-PCA data were scaled after catenating (augmenting) the data matrices. To simplify readability, the terms 'scores' and 'loadings' will be used instead of sample-related weights and variable-related weights, respectively.

3.1. *Results with MA-PCA*

Results from the column-wise augmented MA-PCA studies are shown in Figs. 3–5. Fig. 3 presents the 'temporally averaged geographical scores' (i.e., the average behaviour of the samples in the four sampling seasons) for the first extracted factor, which explains 36% of the initial variance, the second factor explains 20.1%. This is much less than the approx. 80% explained by the first two PCs in the individual analyses, due to the more constrained solution. Still, the most relevant information is satisfactorily extracted. MA-PC1 clearly differentiates the highway, including transects, from the city gardens and main avenue. Loadings associated to this factor (Table 1) hold a clear opposition between Fe and Co (related to the natural origin of the soils) *versus* Zn, Cd, Cu and Pb, which clearly are linked to roadtraffic (Pb from gasolines, Zn and Cd from lubricants, tyres, coachwork and galvanized parts of the vehicle, and Cu from the coachwork and brake lining) [24]. It is worth noting that MA-PCA reveal this difference with just one factor but the invidual seasonal PCAs required two factors.

Moreover, city gardens' MA-PC1 scores depend on the metals accumulated over time. Negative MA-PC1 scores signify new/recently opened gardens (see Fig. 3), slots with frequent gardening, reposition of flowers, addition of fertilizers, etc., and the inner part of a main city park (highway samples have also negative MA-PC1 weights). The 'geographically averaged temporal scores' (Fig. 4) reveal that fall and winter (seasons 1 and 2) are very similar while spring (season 3) and, in particular, summer (season 4) differ.

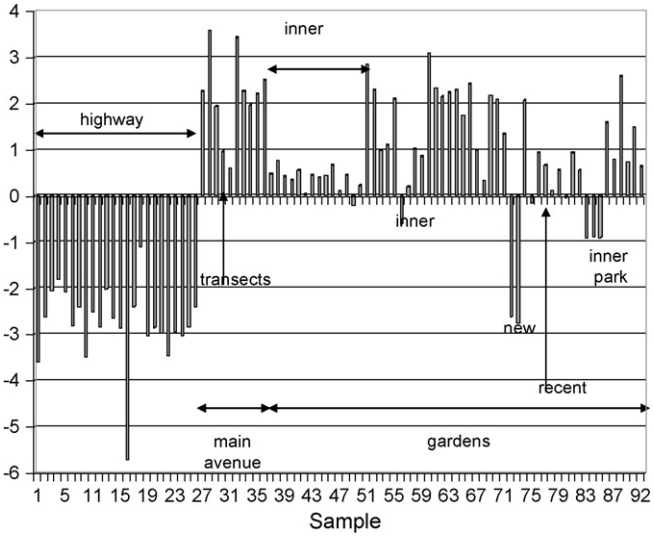

Fig. 3 – Matrix-augmented PCA. Temporary averaged geographical scores for the first factor.

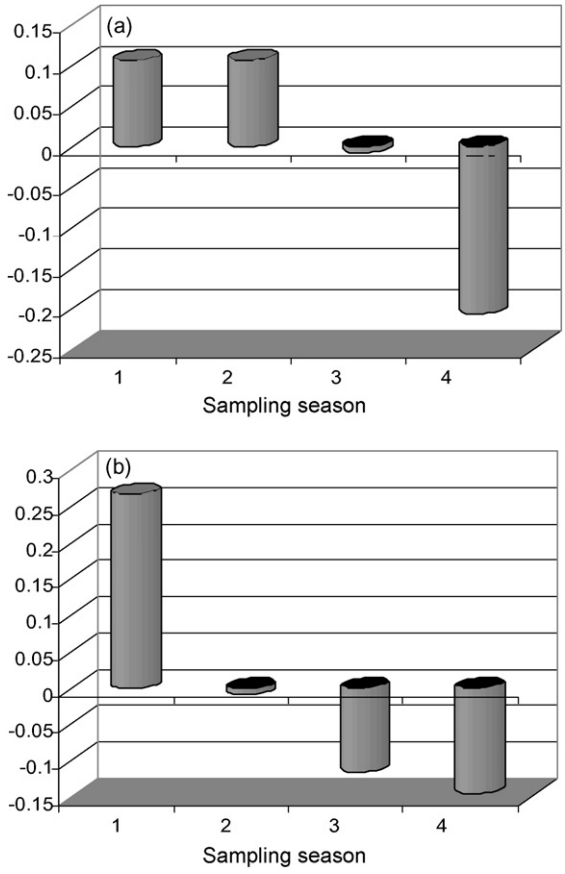

Fig. 4 – Matrix-augmented PCA, geographically averaged temporal scores for (a) PC1 and (b) PC2.

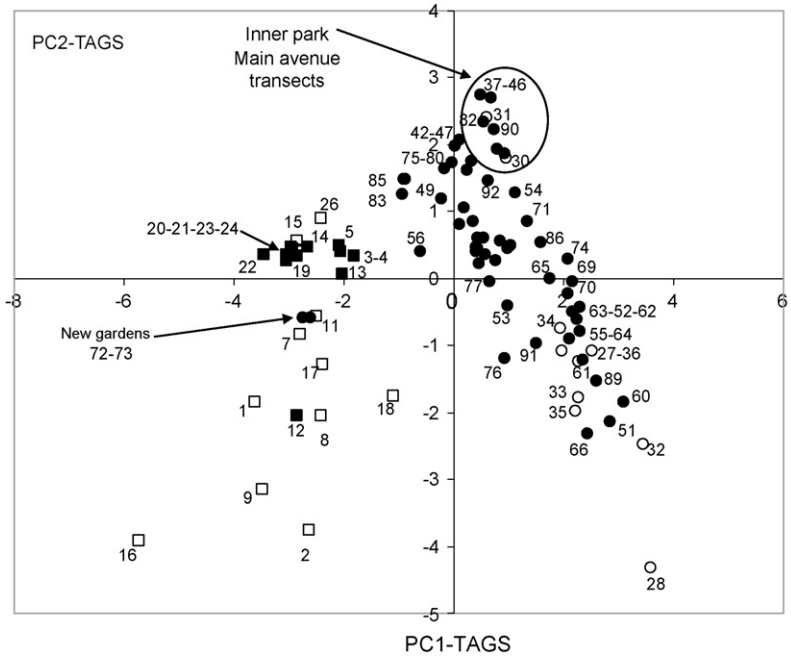

Fig. 5 – Matrix-augmented PCA, PC1 vs. PC2 temporary averaged geographical scores (TAGS). Highway (□), highway transects (■), main avenue (○), city gardens (●).

MA-PCA does not provide more information on this issue and we resourced to individual PCAs to assess each particular behaviour. Thus, it was found that autum is the season with the most obvious differences between the four types of samples (highway, highway transects, gardens, and main avenue). The rainy Galician winter reduced the differences between the subgroups to the point that they became mixed, although the two main groups remained separated. In spring, the two main blocks of samples are well differentiated and, additionally, the two subgroups of highway soils became clearly different, which can be explained by the agricultural activities and climatologic conditions. In summer, the two main blocks of samples can also be differentiated, as well as the highway and its transects, but not the city gardens and main avenue. Indeed, the latter has a different behaviour from the spring season since some samples reduced their metallic contents (particularly Cu and Pb) while several city gardens elevated them. This can be due to a reduction in traffic density along the main avenue during summer.

The MA-PC2 is loaded by Ni, Cr and Pb (with negative loadings, see Table 1). The Ni and Cr status in soils is highly dependent on their contents on the parent rocks. However, their concentrations in surface soils (as the ones considered in this work) also reflect soil-forming processes and pollution (here, mainly roadtraffic) [25]. Hence, to interpret the data the other metals acompanying them into the factors had to be studied (in a 'source apportionment' way). This way the MA-PC2 should be related to roadtraffic because of its association to Pb. Fig. 5 shows that positive or close-to-zero scores are obtained for highway transects, new gardens, main avenue transects and samples from the inner park. Highest negative loadings correspond to samples with (on average) highest con-

Table 1 – Comparison of loadings (analytical variable-way) for each methodology

	MA-PCA		Procrustes rotation		PARAFAC
	Factor 1	Factor 2	Consensus vector 1	Consensus vector 2	Factor 1
Cd	0.35	−0.31	−0.46	−0.08	−0.35
Co	−0.40	−0.29	0.17	−0.47	0.39
Cu	0.29	−0.35	−0.43	−0.14	−0.29
Cr	−0.28	−0.42	0.00	−0.52	0.27
Fe	−0.42	−0.20	0.24	−0.39	0.41
Mn	−0.26	−0.12	0.16	−0.21	0.26
Ni	−0.24	−0.42	−0.03	−0.48	0.23
Pb	0.29	−0.41	−0.45	−0.19	−0.29
Zn	0.35	−0.30	−0.46	−0.06	−0.35
Humidity	−0.04	0.06	0.09	0.01	0.08
LOI	−0.07	0.12	0.12	0.05	0.09
pH	0.22	−0.03	−0.21	0.07	−0.23

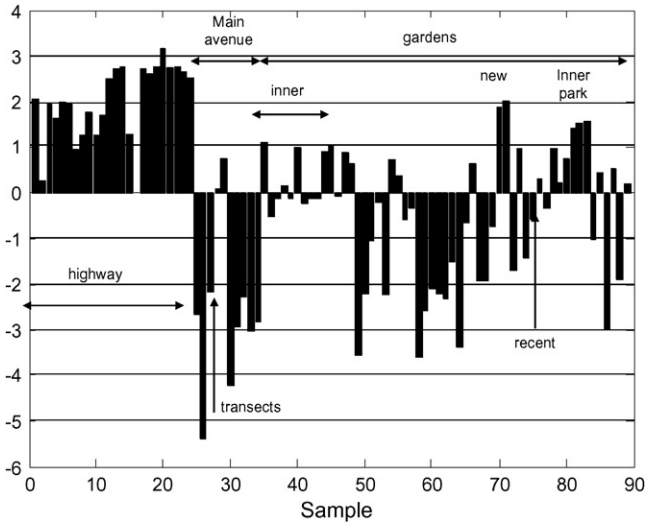

Fig. 6 – Site-averaged consensus scores for Procrustes rotation consensus vector 1.

tents of Ni and Cr (30–70 mg kg^{-1} for both), and medium-high Pb (300–600 mg kg^{-1}) contents. MA-PC2 evolved from autum to summer, like MA-PC1 (see Fig. 4). Fig. 5 agrees very well with PR and PARAFAC (see below). Only limitation is that city gardens and main avenue are not clearly differentiated on average.

3.2. *Results from Procrustes rotation*

Since two principal components were sufficient to describe each season, two components were considered also in PR. Table 1 shows that the first consensus vector is related to road traffic (Pb, Zn, Cd and Cu) whereas the second is linked to the parent soil (Co and Fe, and Cr and Ni).

Fig. 6 presents the site-averaged consensus scores for the first consensus vector. They are very similar to the MA-PC1 ones (see Fig. 3) and discriminate quite clearly between the highway (including transects) and the city gardens and the main avenue. Thus, the samples become differentiated according to their levels on roadtraffic-related metals. In general, levels are higher within the city and the main avenue. Exceptions are the two main avenue transects, new gardens and inner samples from parks (as visualized for MA-PCA above). The averaged temporal consensus vector (Fig. 7a) shows a pattern which resembles very well that from MA-PCA (Fig. 4a). This pattern is quite stable from autum to spring although there are some changes in summer. They may be explained by reduced traffic density on the main avenue (no transportation to schools, University, industries, etc., which proportionally reduced depositions on roadsoils) and by the continuous increase of traffic-related metals in some gardens at the city center that still support a high traffic density (in addition to the typical dry summer weather). The PR angles between the first consensus vector and the first PCs from each independent seasonal study were only 6.79°, 7.15°, 3.04° and 7.45° (autumn, winter, spring and summer, respectively), meaning that the seasonal behaviours indeed are very similar.

The second consensus vector reflects the soils' natural background (see Table 1) as is strongly related to Co, Fe, Cr and Ni. The fact that Cr and Ni are not associated here to typical traffic-related metals suggests that although roadtraffic can be an important anthropogenic source, the main differences between the samples are due to their parent material (characterized mainly by Fe and Co). The PR angles formed by the second consensus vector and the second PCs from the independent seasonal studies were 24.21°, 17.12°, 16.92° and 12.89° (autumn, winter, spring and summer, respectively). Despite they are still small, these are clearly larger than the PC1 angles, reflecting higher seasonal variations. This reflected also in the 'season-averaged consensus scores' (Fig. 7b).

It seems that the two PR consensus vectors divided the information condensed on the first MA-PCA factor, into a 'natural' (leaded by Co and Fe) and an 'anthropogenic' (traffic-related, leaded by Pb, Zn and Cd) factors.

Each seasonal data set (slice in the 3-way data cube) can be projected onto the consensus vectors to get "consensus scores". These can be used to assess whether the consensus vectors retrieved the main sample patterns, for instance by plotting traditional PC1-PC2 scores and PR consensus scores altogether. Fig. 8 shows that the main sample patterns were recovered by the PR vectors (as expected thanks to the low angles they formed with the original factors). It should be stressed that the samples are not expected to overlap because

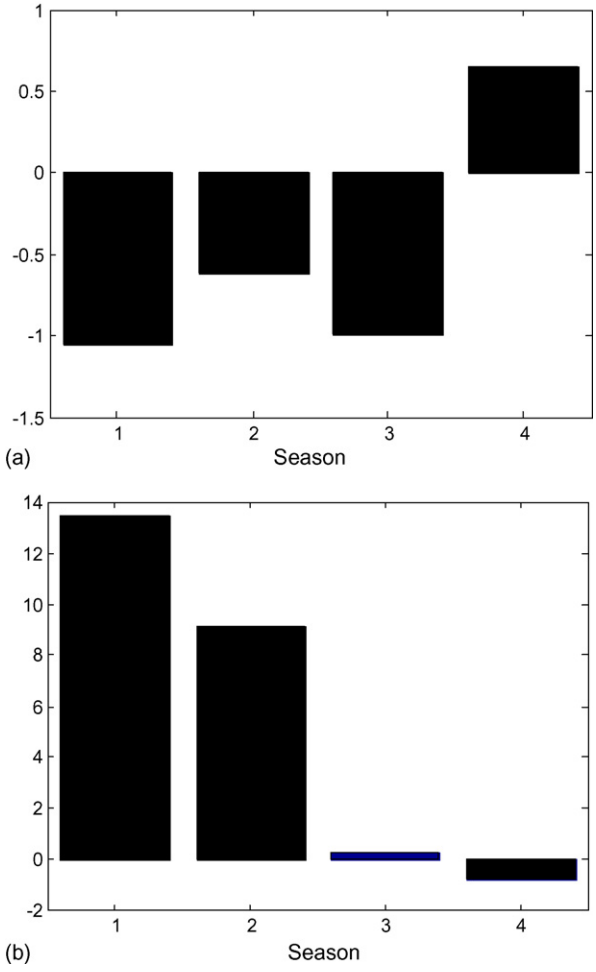

Fig. 7 – Procrustes rotation season-averaged consensus scores for consensus vector (a) 1 and (b) 2.

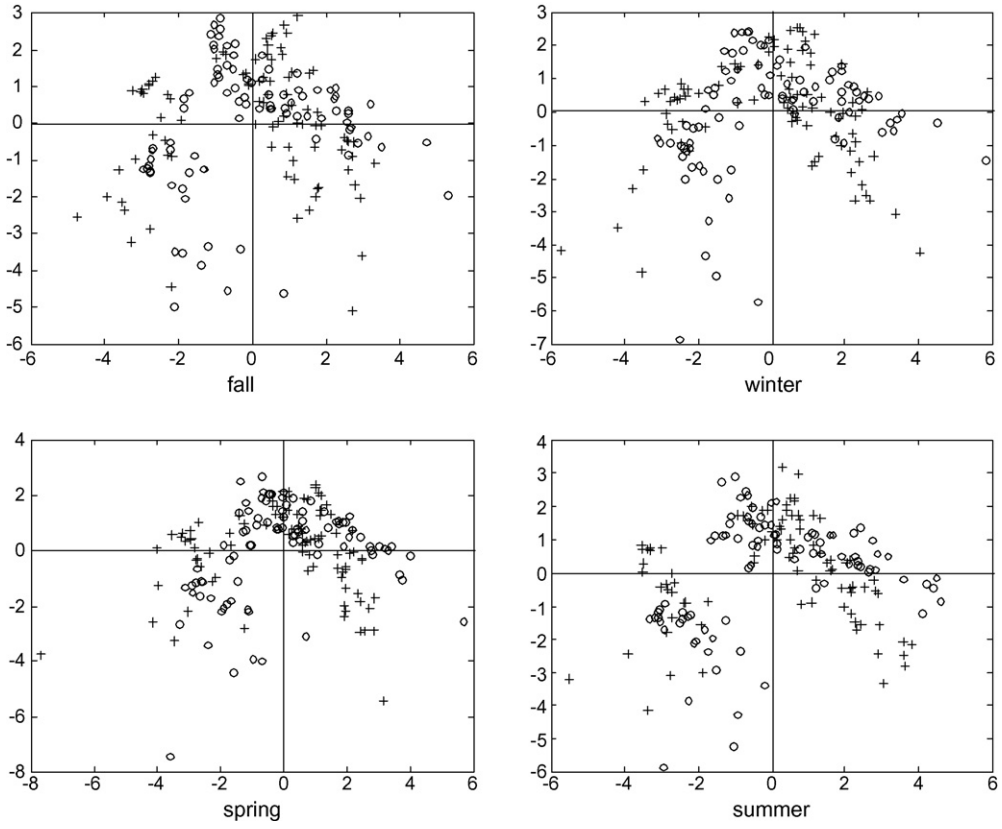

Fig. 8 – Comparison among the independent PCA scores (+) and the Procrustes consensus scores (○) for each sampling season (slice in the data cube).

the subspaces are different. Only the general appearance and sample distribution should be compared. We found that the patterns are essentially identical within each season. Among the different seasons, changes are found only along the second factors. They are not large but sufficient to mix the highway samples with the highway transects, and the avenue with some city gardens (during winter and spring). Now, the 'PR site-averaged consensus scores' can be calculated and

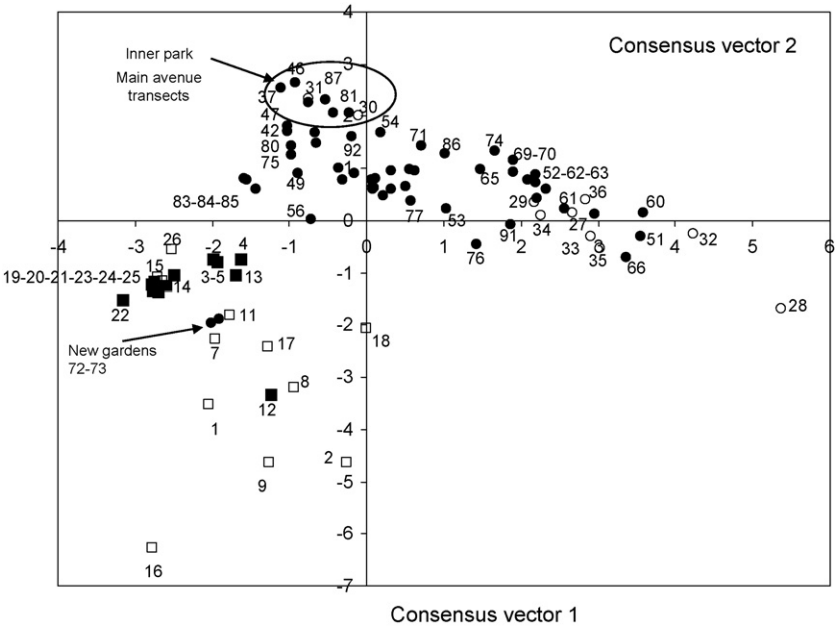

Fig. 9 – Procrustes rotation site-averaged consensus scores, consensus 1 vs. consensus 2. Highway (□), highway transects (■), main avenue (○), city gardens (●).

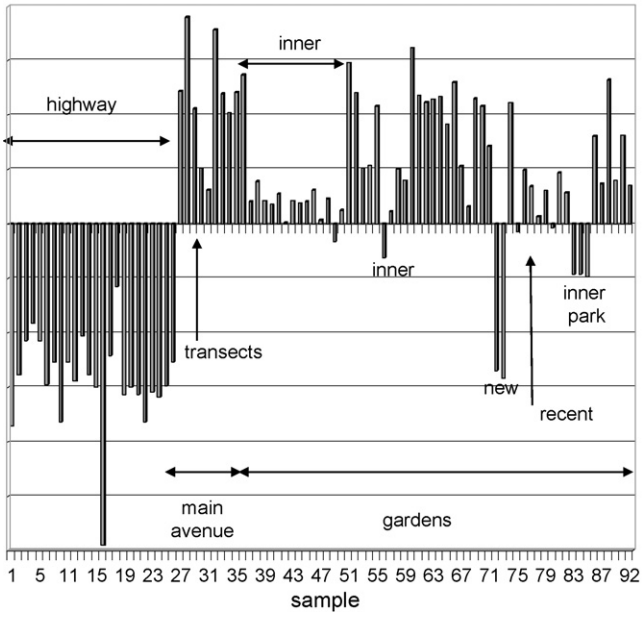

Fig. 10 – PARAFAC mode 1 (sample-related) weights.

depicted (Fig. 9). Here, it can be observed that results are essentially the same as those from MA-PCA (Fig. 5) and no further details are required (the numbers of several overlapping samples were omitted for clarity).

3.3. *Results from PARAFAC*

With PARAFAC one factor was sufficient to decompose the 3-way data matrix. The sample-related weights are shown in Fig. 10. They show the same pattern as those from the two analyses above (Figs. 3 and 6). Hence, a clear difference between the highway area, the city gardens and main avenue is found. The new gardens and inner part of the main park are discriminated from the rest of the city gardens (lowest weights on the right-hand side of the figure). The two transects from the main avenue have lower weights than the other samples from the main avenue. Remaining samples behave in the same way as in the studies above. The analytical variables-related weights revealed a clear opposition between traffic-related metals (Cd, Zn, Pb and Cu) and the soil-related metals (Co and Fe, plus some participation of Cr, Mn and Ni), see Table 1. The season-related weights decrease the importance of this factor steadily from the 1st (fall) to the 4th (summer) season, as in the analyses above (Figs. 4 and 7) and it is not displayed again. These findings are consistent with the studies above.

4. Conclusions

Results presented in this paper demonstrate that the three techniques employed here lead to essentially the same conclusions along the sample-, variable- and season-related ways. This strongly suggests that in absence of new and unexpected pollution events the trilinear assumption behind PARAFAC is satisfied in environmental studies that focus on temporal evolution.

Besides, the results showed that Procrustes rotation is a good option to address 3-way datasets. The consensus vectors yield conclusions that can be compared to those derived from matrix-augmented PCA and PARAFAC. Procrustes rotation can be used to calculate geographical- and temporal-averaged scores, much in the same way as in MA-PCA. This allows for insightful representations where the sample patterns can be viewed and compared (either between slices or with the individual PCAs). Different from the other methods is that Procrustes rotation measures the similarity (expressed as an angle) between the calculated consensus factor and the corresponding factor in each slice. These Procrustes rotation similarity angles can reveal in a snapshot in which season/s the sample patterns changed. On the contrary, PARAFAC and MA-PCA require detailed studies of each slice to ascertain the differences on the site-averaged scores patterns.

Another difference is that Procrustes rotation revealed two important pollution patterns while MA-PCA and PARAFAC condensed them in a single factor. This might be attributed to the fact that Procrustes rotation holds more degrees of freedom and, therefore, the information can be divided in different factors, which would be an advantage to interpret complex systems.

Acknowledgements

Prof. Dr. Mikael Kubista acknowledges a sabbatical grant from the Spanish Ministry of Education (SAB2005-0162).

REFERENCES

[1] V. Gaganis, N. Pasadakis, Anal. Chim. Acta 573–574 (2006) 328.
[2] M. Wilhelm, I. Lombeck, F.K. Ohnesorge, Sci. Total Environ. 141 (1994) 275.
[3] I. Stanimirova, K. Zehl, D.L. Massart, Y. Vander Heyden, J.W. Einax, Anal. Bioanal. Chem. 385 (2006) 771.
[4] P.J. Jenks, Spectros. Europe 19 (1) (2007) 30.
[5] C. Micó, L. Recatalá, M. Peris, J. Sánchez, Spectros. Europe 19 (1) (2007) 23.
[6] A. Smilde, R. Bro, P. Geladi, Multi-Way Analysis, Willey, United Kingdom, 2004.
[7] R. Tauler, D. Barceló, E.M. Thurman, Environ. Sci. Technol. 34 (16) (2000) 3307.
[8] M. Felipe-Sotelo, J.M. Andrade, A. Carlosena, R. Tauler, Anal. Chim. Acta 583 (2007) 128.
[9] K.P. Singh, A. Malik, V.K. Singh, N. Basant, S. Sinha, Anal. Chim. Acta 571 (2006) 248.
[10] A. Carlosena, P. López-Mahía, S. Muniategui, E. Fernández, D. Prada, J. Anal. At. Spectrom. 13 (1998) 1361.
[11] A. Carlosena, J.M. Andrade, D. Prada, Talanta 47 (1998) 753.
[12] R. Brereton, The Alchemist (2000) 2000/09/01, http://www.chm.bris.ac.uk/org/chemometrics/pubs/chemweb.html.
[13] W.J. Krzanowski, Principles of Multivariate Analysis, Clarendon Press, Oxford (UK), 2000.
[14] J.M. Andrade, M.P. Gómez-Carracedo, W.J. Krzanowski, M. Kubista, Chemom. Intell. Lab. Syst. 72 (2004) 123.
[15] A. Carlosena, J.M. Andrade, M. Kubista, D. Prada, Anal. Chem. 67 (1995) 2373.

[16] A. Elbergali, J. Nygren, M. Kubista, Anal. Chim. Acta 379 (1999) 143.
[17] E.R. Malinowski, Factor Analysis in Chemistry, second ed., Willey, New York, 1991.
[18] S. Wold, Technometrics 24 (1978) 397.
[19] H.T. Eastment, W.J. Krzanowski, Technometrics 24 (1982) 73.
[20] M. Kubista, Chemom. Intell. Lab. Syst. 7 (1990) 273.
[21] J.M. Andrade, M.P. Gómez-Carracedo, E. Fernández, A. Elbergali, M. Kubista, D. Prada, Analyst 128 (2003) 1193.
[22] J.M. Andrade, M. Holík, J. Halámek, Talanta 63 (2004) 865.
[23] GenEx User's Manual, MultiD, Sweden (2007) (www.multid.se).
[24] R. García, E. Millán, Sci. Total Environ. 146/147 (1994) 157.
[25] A. Kabata-Pendias, H. Pendias, Trace Elements in Soils and Plants, CRC Press, Boca Raton, Florida, 1984.