Parallel factor analysis for the study of systematic error in inductively coupled plasma atomic emission spectrometry and mass spectrometry†

Antonio Moreda-Piñeiro,[a] Ana Marcos,[b] Andrew Fisher[b] and Steve J. Hill*[b]

[a]Department of Analytical Chemistry, Nutrition and Bromatology, Faculty of Chemistry, University of Santiago de Compostela, Avenida das Ciencias, s/n., 15706 -Santiago de Compostela, Spain
[b]Department of Environmental Sciences, Plymouth Environmental Research Centre, University of Plymouth, Drake Circus, Plymouth, Devon, UK PL4 8AA

This paper describes the application of a trilinear parallel factor analysis (PARAFAC) to study systematic error during the multi-element determination of a range of analytes in acid digests of solid samples (tea leaves) by ICP-AES and ICP-MS. The three variables studied were the "number of digestions", in order to assess the systematic error associated with the sample pre-treatment, and the "number of replicates" and "calibration", to provide information on the systematic error associated with the analytical determination itself. The elements under study were Co, Cr, Cu, Ni, Pb, Rb and Ti by ICP-MS, and Ba, Ca, Fe, Mg, Mn, Sr and Zn by both ICP-MS and ICP-AES. For some elements flame atomic absorption spectrometry was used for comparative purposes. A Chinese tea certified reference material containing many of the metals above was used in the study. The results obtained were compared to results from ANOVA. It was found that the systematic error, expressed as the sum of squares after PARAFAC, was quite different from the results obtained using ANOVA due to the very different way in which the models are built. The PARAFAC approach is shown to be straightforward to implement and robust.

Introduction

There are many chemical systems that provide multi-way data. A fluorescence emission spectrum measured at several excitation wavelengths for several samples, or the spectrum obtained at different wavelengths following the chromatographic separation of several samples are examples of three-way analysis found in the literature.[1] In such cases, the application of a two-way chemometrics tool such as principal component analysis (PCA) is limited, and often realistic conclusions cannot be drawn. However, parallel factor analysis (PARAFAC) is a multi-way decomposition method that can be used to solve systems characterised by multi-way data. This chemometrics tool is considered to be a generalisation of PCA to higher order arrays, although there are important differences from PCA such as the lack of rotation freedom in PARAFAC and the simultaneous estimation of components or factors.[1]

Despite PARAFAC being a relatively new chemometric technique, a number of applications have already been reported. To date, most correspond to the decomposition process in excitation–emission fluorescence spectroscopy[2–10] in order to determine hydrocarbons,[4,5] polycyclic aromatic hydrocarbons,[6,8,10] metabolites of acetylsalicylic acid[7] or carbamate pesticides.[10] The classification of fruit (apple)[2] and the monitoring of quality parameters during sugar production[9] have also been reported. The application of PARAFAC to analytical techniques such as high-performance liquid chromatography (HPLC), coupled with different detectors such as fluorescence spectrometry[6,11] and ultraviolet–visible spectrometry[12,13] have also received attention, and Smilde et al.[14] have used PARAFAC for the calibration of new stationary phases in HPLC. Other applications such as the decomposition of three-way kinetic-spectrophotometry spectra[15] or the estimation of rate constants using on-line short wavelength-near infrared data (SW-NIR) or ultraviolet–visible spectra[16,17] have been developed.

The use of PARAFAC for analysis of variance is rare; however, there is an advantage in using PARAFAC for ANOVA because of the way interacting terms are modelled. In a standard ANOVA an interaction between three variables (A, B and C) would be estimated as abc_{ijk}, while in a trilinear model, the effect would be estimated as $a_i b_j c_k$ or as a sum of such expressions if more PARAFAC components are estimated, i.e., the interaction between variables is not only estimated as a whole, but is modelled as a multiplicative effect of the different factors. If the multiplicative model is appropriate, the applied restriction ($a_i b_j c_k$ instead of merely E_{ijk}) will give a more straightforward model, and thus the interpretation of results from PARAFAC is easier than those from ANOVA.[18]

In any analysis there are at least two variables that must be taken into account to obtain precise and accurate data, the "number of replicates" and the "calibration". If a solid material is being analysed, a sample decomposition pre-treatment (a wet acid digestion, a fusion or a dry ashing) is normally required, and a third variable (the number of times that the sample pre-treatment is carried out on the same sample) must also be considered. Therefore, in order to minimise the error in the multi-element analysis of sample acid digests, three variables can be considered: the "number of replicates", the "calibration" and the "number of acid digestions". In this study, the concentration of metals in an environmental matrix (tea) was determined whilst varying the number of acid digestions prepared, the number of replicates of each acid digest and the use of two different calibration graphs. The data obtained for each metal can then be arranged in a three-way array. The application of chemometric techniques

DOI: 10.1039/b008847l

Table 1 Glossary of terms from Bro (ref. 20)

Component or factor	New variables obtained after the decomposition method
Modes or ways	The mode is the basic entity building an array: a matrix has two *modes* or *ways* (the row mode and the column mode), a three-way or three-mode array has three modes or ways
Order	The *order* of an array is the number of modes, hence a matrix is a second-order array, and a three-way array a third-order array
Dimension	The term *dimension* will be used to denote the number of levels in a mode
Rank	*Rank* is the minimum number of PARAFAC components necessary to describe an array. For a two-way array this definition reduces to the number of principal components necessary to fit the matrix

such as PCA can be used to assess the systematic error in an analytical process.[19] However, the three-way characteristics of the data set makes the interpretation difficult using a 2-way chemometrics tool, especially when the loadings from the PCA are being interpreted, and the third dimension of the problem (*i.e.*, the "calibration") masks the effects of the other two variables, *i.e.* "number of replicates" and "number of digestions".

The aim of this work was to study the systematic errors in the determination of a range of elements (Ba, Ca, Co, Cr, Cs, Cu, Fe, Mg, Mn, Ni, Pb, Rb, Sr, Ti and Zn) in tea leaves by ICP-AES and ICP-MS. In order to aid the clarity of the text, the terminology used by Bro[20] has been adapted as summarised in Table 1.

Experimental

An ICP-AES Liberty 200 instrument (Varian, Cheshire, UK) and an ICP-MS PlasmaQuad PQ2+ instrument (Fisons Instruments, Winsford, Cheshire, UK) were used in the study.

A flame atomic absorption spectrometer SpectrAA 50 (Varian), equipped with a N_2O–acetylene flame, was used for

Table 2 Operating parameters for ICP-MS

Forward power/W	1350
Gas flows/l min^{-1}	
Nebulizer	0.85
Auxiliary	1.0
Coolant	15.0
Nebulizer type	Ebdon, high solids
Data acquisition for quantitative analysis	Peak jump
Internal standard	In, 100 µg l^{-1}
Isotope masses monitored	
^{24}Mg, ^{48}Ti, ^{52}Cr, ^{54}Fe, ^{55}Mn, ^{59}Co, ^{60}Ni, ^{65}Cu, ^{66}Zn, ^{85}Rb, ^{88}Sr, ^{138}Ba, and ^{206}Pb	

Table 3 Operating conditions for ICP-AES

Forward power/W		1000
Gas flows/l min^{-1}	Coolant	15.0
	Auxiliary	1.5
Nebulizer type		Cross flow
Nebulizer pressure/kPa		150
Pump speed/rev min^{-1}		20
PMT voltage/V		650
Stabilisation time/s		10
Detection wavelengths/nm	Ba	455.403
	Ca	393.366
	Cu	324.754
	Fe	259.940
	Mg	279.533
	Mn	257.610
	Sr	407.771
	Zn	213.856

Ca determinations. A similar instrument (SpectrAA 300/400, Varian), equipped with an air–acetylene flame, was used for Cu, Fe, Mg, Mn and Zn determinations. Hollow cathode lamps (Varian) operating at the recommended current were used.

A hotplate (SH3, Stuart Scientific, UK) was used for total digestion of the tea reference material.

The chemometrics package UNSCRAMBLER, 1998 (CAMO ASA, Trondheim, Norway) was used to obtain ANOVA results in the experimental design work.

Matlab software (Mathworks, Inc., Version 5.1) and the N-way toolbox for Matlab were employed to perform the PARAFAC analysis.[21]

Reagents

The chemicals used were of ultrapure grade, and ultrapure water of resistance 18 MΩ cm^{-1} was obtained from a Milli-Q purification system (Millipore Co., Bedford, MA, USA). AnalaR nitric acid 70.0% was obtained from Merck (Poole, Dorset, UK). Stock standard solutions (1.000 or 10.000 g l^{-1}) were also supplied by Merck. Cobalt stock standard solution was purchased from Aldrich (Gillingham, Dorset, UK). Chinese tea reference material GBW 08505 was obtained from the Bureau of Analysed Samples (Middlesbrough, UK).

Acid digestion procedure

The Chinese tea reference material was digested according to a method published elsewhere.[22] Sample (0.2500 g) was weighed into a clean beaker and nitric acid (10 ml) added. The beaker was covered with a watch-glass and the sample boiled gently on a hot plate for 3 h. After cooling to room temperature, the acid digests were transferred quantitatively to 50 ml volumetric flasks. Indium (as an internal standard for ICP-MS measurements) was added to each digest to give a concentration of 100 µg l^{-1} after dilution to 50 ml. The acid digests were kept in polyethylene vials at room temperature prior to analysis.

ICP-AES/MS and FAAS measurements

The elements Ba, Co, Cr, Cs, Cu, Fe, Mg, Mn, Ni, Pb, Rb, Sr, Ti and Zn were measured by ICP-MS without dilution, using the operating parameters and the isotope masses shown in Table 2. In addition Ba, Ca, Fe, Mg, Mn, Sr and Zn were also measured by ICP-AES without dilution, using the operating conditions and emission wavelength lines given in Table 3. The optimum conditions used for ICP-MS/AES have been described elsewhere.[22] Finally, for comparative purposes, a number of elements (Ca, Cu, Fe, Mg, Mn, Zn) were also measured by FAAS, using an acetylene–air flame for all elements, except Ca, for which a N_2O–acetylene flame was used. In order to avoid the necessity of sample dilution for the Ca and Mg FAAS measurements, the less sensitive resonance lines (239.9 and 202.6 nm, for Ca and Mg, respectively) were employed. The operating parameters and details of the atomic absorption resonance lines used for the FAAS measurements are given in Table 4.

Results and discussion

Table 5 lists the elements with certified (Ba, Ca, Cu, Fe, Mg, Mn, Ni, Pb, Rb, Sr and Zn) or informative (Co, Cr and Ti) concentrations (together with the uncertainty) in the GBW 08505 tea reference material used in the study.

Data

Data sets for each element and technique, similar to those shown in Table 6 for the determination of ^{54}Fe by ICP-MS, were obtained to investigate the influence of the three variables ("number of digestions", "number of replicates" and "calibra-

Table 4 Operating parameter for FAAS

	Wavelength/ nm	Slit width/ nm	Lamp current/ mA	Air flow rate/l min^{-1}	C_2H_2 flow rate/ l min^{-1}
Ca	239.9	0.2	10	10.00[a]	6.35
Cu	324.8	0.7	10	13.50	2.00
Fe	248.3	0.2	6	13.50	2.00
Mg	202.6	0.7	10	13.50	2.00
Mn	279.5	0.2	6	13.50	2.00
Zn	213.9	0.2	10	13.50	2.00

[a]N$_2$O flow rate.

Table 5 Certified or informative concentrations in a Chinese tea certified reference material (GBW 08505) for the elements investigated and the techniques employed during this study

Element	Certified concentration/µg g^{-1}	Technique
Ba	15.7 ± 1.9	ICP-AES, ICP-MS
Ca[a]	0.284 ± 0.021	ICP-AES, FAAS
Co[b]	0.2	ICP-MS
Cr[b]	0.8	ICP-MS
Cu	16.2 ± 1.9	ICP-MS, FAAS
Fe	373 ± 23	ICP-AES, ICP-MS, FAAS
Mg[a]	0.224 ± 0.019	ICP-AES, ICP-MS, FAAS
Mn	766 ± 28	ICP-AES, ICP-MS, FAAS
Ni	7.61 ± 0.48	ICP-MS
Pb	1.06 ± 0.10	ICP-MS
Rb	36.9 ± 1.3	ICP-MS
Sr	10.8 ± 1.8	ICP-AES, ICP-MS
Ti[b]	36	ICP-MS
Zn	38.7 ± 3.9	ICP-AES, ICP-MS, FAAS

[a]Concentration expressed in %(m/m). [b]Informative concentration.

tion") on the accuracy of the methods. The variable called "number of digestions" refers to the number of acid digestions performed in parallel (one to four). The variable called "number of replicates" is the number of instrumental determinations carried out with each acid digestion (two to five). Finally, the variable "calibration" refers to two aqueous calibration graphs, prepared independently and covering an adequate range of concentrations. The procedure for calibration when using the three techniques (ICP-MS, ICP-AES and FAAS) was based on a weighted least squares approach (weighted regression). These data sets have three modes, representing the three different variables under study as in Figure 1(a).

The different levels (dimensions) of each variable are shown in Table 7 corresponding to a full factorial design (FFD) $4^1 \times 4^1 \times 2^1$ (with $4^1 \times 4^1 \times 2^1 = 4 \times 4 \times 2 = 32$ samples). In this three-way PARAFAC model, the three different ways represent the number of digestions (dimension 4), number of replicates (dimension 4) and calibration (dimension 2), so, the ijk element of the three-way array contains the mean element concentration for the ith number of digestions, the jth number

of replicates and using the kth calibration graph, as shown in Figure 1(a).

Data sets for all elements and techniques similar to those for ^{54}Fe determination by ICP-MS are available on the RSC web site (Table S1†).

PARAFAC

The use of PARAFAC splits the raw data into a number of components. Each component is formed by three vectors; one score vector (for instance, for the number of replicates) and a loading vector in each variable direction (number of digestions and calibration in this case). The decomposition is such that the raw data is the sum of the products of scores and loadings for each component, e.g.,

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + E \qquad (1)$$

where x_{ijk} is the raw data, and a, b and c, the scores and loadings for F components. E is the error of the model. For $F = 2$, eqn. (1) can be represented graphically as in Fig. 1(b).

In this study, the components higher than 1 (second, third, ··· nth component) are those that offer the useful information (components that show the differences between the objects), and the first component can be considered as an estimation of the raw data. As can be seen in Tables 8 and 9, the first component produces a value very similar to that of the raw data, and the remaining information provided by the following PARAFAC components could be associated with the deviations or error of the measurements.

Selection of the number of components. The first step in the construction of a PARAFAC model is the selection of the optimum number of components. This number can be chosen to minimise the sum squared prediction error (SSPE), mathematically expressed as

$$SS = \left[\left(\sum_{f=1}^{F} a_{if} b_{jf} c_{kf} \right) - x_{ijk} \right]^2 \qquad (2)$$

where x_{ijk} is the ijkth element of the data set, and a_{if}, b_{jf} and c_{kf} are the scores and loadings for the fth component.

An alternative approach, called the core consistency diagnostic,[20] can also be used for determining the correct number of components for multi-way models. This approach has been demonstrated to be extremely powerful for determining the optimum number of PARAFAC components.[20] Both approaches, SSPE and core consistency, were applied three times for each data set (each element and technique) for one, two, and three components. The replication (the determination of each parameter three times for each number of components) is necessary because the core consistency algorithm can offer very different results when the solution for a certain number of components is inappropriate. Some examples of SSPE and core consistency plots are given in Fig. 2, and a complete set of plots

Table 6 Fe mean concentrations (µg g^{-1})in the Chinese tea reference material GBW 08505 obtained by ICP-MS

^{54}Fe (ICP-MS) certified concentration of 373 ± 23 µg g^{-1}

|		One digestion	Two digestion	Three digestion	Four digestion
Calibration I	Two replicates	386	403	399	392
Calibration I	Three replicates	379	387	385	380
Calibration I	Four replicates	365	373	374	367
Calibration I	Five replciates	361	367	372	364
Calibration II	Two replicates	406	389	389	385
Calibration II	Three replicates	392	384	378	375
Calibration II	Four replicates	380	376	371	366
Calibration II	Five replicates	377	366	363	360
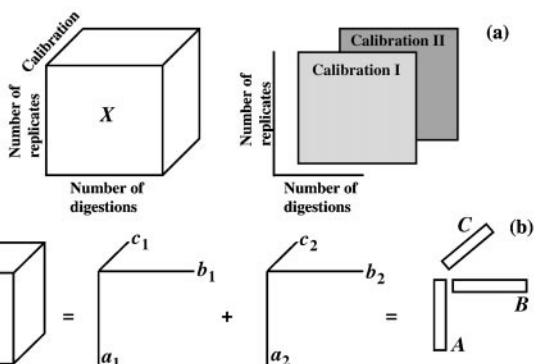
Fig. 1 Graphical representation of the three-way array of metal concentrations (a), and a two-component PARAFAC model of the data array X (b).

is available on the RSC web site under the title of Fig. S1†. The number of components must be chosen in order to obtain the lowest SSPE and also the highest value for the core consistency. PARAFAC models with higher than three-component solutions were not considered because more interpretable models are usually obtained with fewer components.[20] As a rule we can consider that if the core consistency is lower than 50.0% the number of components for that solution is not adequate. It can be seen in Fig. 2, (and in Fig. S1†), that the SSPE is decreased when the number of components in the model is high, which can lead to the choice of a larger number of components. However, although the SSPE is low for higher numbers of components, the core consistency gives the correct number of components. This can be visualised in Fig. 2 for Cu and ICP-MS, where it can be seen that the SSPE is equal to zero for a

Variables	Levels
Number of replicates (R)	2, 3, 4, 5
Number of digestions (D)	1, 2, 3, 4
Calibration (C)	I, II

number of components higher than 2 but the core consistency falls below 50.0% for the three and four-component solutions, and a two-component model must be chosen.

From the results above, two-component or three-component solutions were found to be adequate for modelling most of the different data sets. For instance, a two-component model was obtained for ^{24}Mg, ^{48}Ti, ^{52}Cr, ^{54}Fe, ^{59}Co, ^{65}Cu, ^{66}Zn, ^{85}Rb and ^{88}Sr determination by ICP-MS, and also for Ca using ICP-AES and Mg and Zn using FAAS. A three-component solution was found for ^{55}Mn, ^{60}Ni and ^{138}Ba and ^{206}Pb using ICP-MS, and Ba and Fe using ICP-AES and FAAS, respectively.

After the core consistency determinations, it was found that some of the data sets may need to be fitted using a one-component model. Examples include Fe (Fig. 2), Mg, Mn, Sr and Zn by ICP-AES, and Cu and Mn by FAAS. For these cases, the determination of the systematic error is not possible using the PARAFAC approach because the first component of the model does not give differences between the objects (data). In such cases it can be concluded that the differences between the measurements carried out using different calibrations, or by using a different number of replicates and digestions, are not significant. After a detailed study of the elements and techniques for which one-component solutions were found to be optimum, it can be seen that this data set is related to elements at a high concentration in the reference material and

Table 8 Example of PARAFAC decomposition into raw data + error for a two-component model:^{85}Rb and ICP-MS

Original data—

		One digestion	Two digestions	Three digestions	Four digestions
Calibration I	Two replicates	39.75	38.88	37.70	37.84
Calibration I	Three replicates	39.67	38.80	38.22	38.22
Calibration I	Four replicates	40.28	39.16	38.63	38.46
Calibration I	Five replicates	40.69	39.32	38.58	38.29
Calibration II	Two replicates	34.60	34.60	35.49	36.12
Calibration II	Three replicates	35.66	34.98	35.74	35.86
Calibration II	Four replicates	33.07	33.94	35.10	35.43
Calibration II	Five replicates	34.22	34.96	35.71	35.77

First component (raw data)—

		One digestion	Two digestions	Three digestions	Four digestions
Calibration I	Two replicates	39.30	38.53	37.76	37.76
Calibration I	Three replicates	39.42	38.65	37.88	37.88
Calibration I	Four replicates	39.73	38.95	38.17	38.17
Calibration I	Five replicates	39.89	39.11	38.33	38.33
Calibration II	Two replicates	38.21	37.46	36.71	36.71
Calibration II	Three replicates	38.33	37.58	36.82	36.82
Calibration II	Four replicates	38.62	37.87	37.11	37.11
Calibration II	Five replicates	38.78	38.02	37.26	37.26

Second component (error)—

		One digestion	Two digestions	Three digestions	Four digestions
Calibration I	Two replicates	0.37	0.26	0.13	0.11
Calibration I	Three replicates	0.30	0.21	0.11	0.09
Calibration I	Four replicates	0.56	0.40	0.21	0.16
Calibration I	Five replciates	0.45	0.32	0.17	0.13
Calibration II	Two replicates	0.37	-2.58	-1.34	-1.05
Calibration II	Three replicates	0.30	-2.12	-1.10	-0.86
Calibration II	Four replicates	0.56	-3.94	-2.04	-1.61
Calibration II	Five replicates	0.45	-3.19	-1.65	-1.30

Table 9 Example of PARAFAC decomposition into raw data + error for a three-component model: ^{138}Ba and ICP-MS

Original data—

		One digestion	Two digestions	Three digestions	Four digestions
Calibration I	Two replicates	16.54	15.80	15.41	15.47
Calibration I	Three replicates	16.44	15.66	15.18	15.34
Calibration I	Four replicates	16.17	15.33	14.89	15.03
Calibration I	Five replciates	15.95	15.03	14.69	14.83
Calibration II	Two replicates	17.05	17.10	16.94	16.52
Calibration II	Three replicates	16.84	17.00	16.81	16.61
Calibration II	Four replicates	16.64	16.49	16.47	16.16
Calibration II	Five replicates	16.64	16.23	16.30	16.08

First component (raw data)—

		One digestion	Two digestions	Three digestions	Four digestions
Calibration I	Two replicates	16.73	15.77	15.77	15.77
Calibration I	Three replicates	16.62	15.66	15.66	15.66
Calibration I	Four replicates	16.21	15.28	15.28	15.28
Calibration I	Five replciates	15.92	15.00	15.00	15.00
Calibration II	Two replicates	17.96	16.93	16.93	16.93
Calibration II	Three replicates	17.84	16.81	16.81	16.81
Calibration II	Four replicates	17.40	16.40	16.40	16.40
Calibration II	Five replicates	17.09	16.10	16.10	16.10

Second component (error)—

		One digestion	Two digestions	Three digestions	Four digestions
Calibration I	Two replicates	−0.46	0.08	−0.20	−0.30
Calibration I	Three replicates	−0.46	0.08	−0.19	−0.30
Calibration I	Four replicates	−0.36	0.06	−0.15	−0.23
Calibration I	Five replicates	−0.24	0.04	−0.10	−0.16
Calibration II	Two replicates	−1.09	0.18	−0.46	−0.71
Calibration II	Three replicates	−1.08	0.18	−0.46	−0.70
Calibration II	Four replicates	−0.85	0.14	−0.36	−0.55
Calibration II	Five replicates	−0.57	0.10	−0.24	−0.38

Third component (error)—

		One digestion	Two digestions	Three digestions	Four digestions
Calibration I	Two replicates	0.12	0.03	−0.22	−0.12
Calibration I	Three replicates	0.18	0.04	−0.32	−0.18
Calibration I	Four replicates	0.17	0.04	−0.31	−0.17
Calibration I	Five replciates	0.15	0.03	−0.27	−0.15
Calibration II	Two replicates	−0.12	−0.02	0.21	0.12
Calibration II	Three replicates	−0.17	−0.04	0.31	0.17
Calibration II	Four replicates	−0.16	−0.03	0.30	0.16
Calibration II	Five replicates	−0.14	−0.03	0.26	0.14

measured by ICP-AES or FAAS. The precision offered by these techniques has been found to be better than that given by ICP-MS, as demonstrated in another publication using principal component analysis and experimental designs.[23] In addition, for the elements and techniques referred to above, the extension of systematic error found for the three sources of error (calibration, number of replicates and number of digestions) was very small.[23] Therefore, PARAFAC also detects that the systematic error is small in these cases, and one-component PARAFAC models are proposed.

In the same way, a three-component model was found to be optimum for some cases (Ba, Mn and Ni determination by ICP-MS, Ba by ICP-AES and Fe by FAAS), which are the determinations which offer the worst precision.

Scores and loadings vectors from PARAFAC. Some examples of the score and loading vectors of one-, two- and three-components are given in Fig. 3, for Cu (FAAS), ^{48}Ti (ICP-MS) and Ba (ICP-AES), respectively. A complete set of loading vector plots is available on the RSC web site as Fig. S2†. It can be seen that two Y-axes are given in some plots. This happens when more than one component is obtained and when the element concentration in the reference material is high. Since we have assumed that the first component estimates the raw data and the second and third components the systematic error, if the element concentration is high, the first solution must be represented using a higher Y scale than for the second and third components (systematic errors are small). Therefore in such cases, the left Y-axis will represent the first component and the right Y-axis the second and third components.

We conclude that the first PARAFAC factor accounts for the mean data. This can be seen in the plots in Fig. 3 and also in Fig. S2†, where the loading vectors for the first component offer constant values independent of the number of replicates, digestions and calibration. This means that the first component does not give information about the differences between the data. These differences, *i.e.*, the systematic error, could be assessed using the loadings of the second PARAFAC component, or in the case of a three-component model, from the loading of the second and third PARAFAC factors. Thus, the multiplication of the loadings from PARAFAC factors higher than one can be used to evaluate the systematic error in
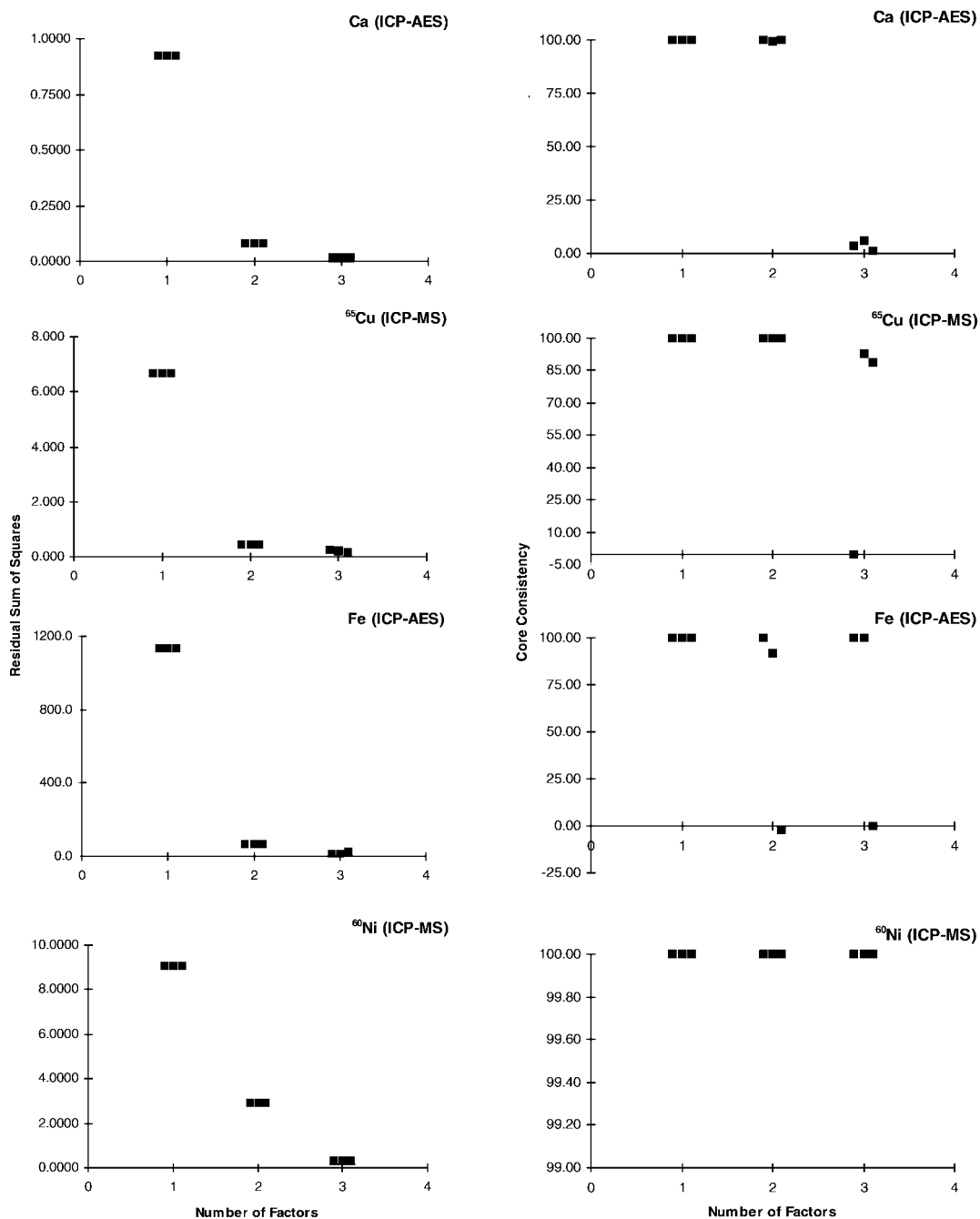
Fig. 2 Sum of square predicted error (left) and core consistency (right) plots for Ca (ICP-AES), ^{65}Cu (ICP-MS), Fe (ICP-AES) and ^{60}Ni (ICP-MS).

the system. In Fig. 3 and also in Fig. S2† it can be seen that the loading vectors for the second and third component offer very different values for all numbers of replicates and digestions, and for the calibration.

After analysis of the loading vectors, it can be concluded that the "number of replicates" is a less significant variable for ICP-AES and FAAS determinations because the loading vectors remain more or less constant for all replicates tested. However, this variable is very significant for some ICP-MS measurements, such as for ^{59}Co and ^{52}Cr, ^{60}Ni, ^{206}Pb, ^{85}Rb, and ^{48}Ti (Fig. S2† and Fig. 4). The "number of digestions" and "calibration" variables appear to be very significant for all elements and techniques. The values of the loading vectors for the "number of digestions" tend to zero when the number of acid digestions is higher, indicating that the error can be minimised by increasing the number of digestions. This can be observed in Fig. 3 for Ba (ICP-AES), and for other cases in Fig. S2†. Similar results to those shown above have been

obtained using the ANOVA output from experimental designs[23] and these indicate that PARAFAC may be applied satisfactorily instead of ANOVA to study the three-way data set.

The loadings in the space formed by the first and second components (two-component solution) or by the second and third components (three-component solution) are given in Fig. 5, and also in the file Fig. S3† on the RSC web site. The interpretation of these plots is as follows: for the two-component models [^{65}Cu (ICP-MS) in Fig. 5] all the loadings have a very similar value for the abscissa (Factor 1), and in general, it can be seen that the loadings are closer to (0,0) when the number of replicates and also the number of digestions are increased. For a three-component solution [shown in Fig. 5 for ^{60}Ni (ICP-MS) and ^{206}Pb (ICP-MS)] it can also be seen that the loadings offer a low value for the factor 2 and 3 (closer to (0,0)) for a high number of digestions (about 3 and 4 digestions). An increase in the number of replicates produces a similar effect to
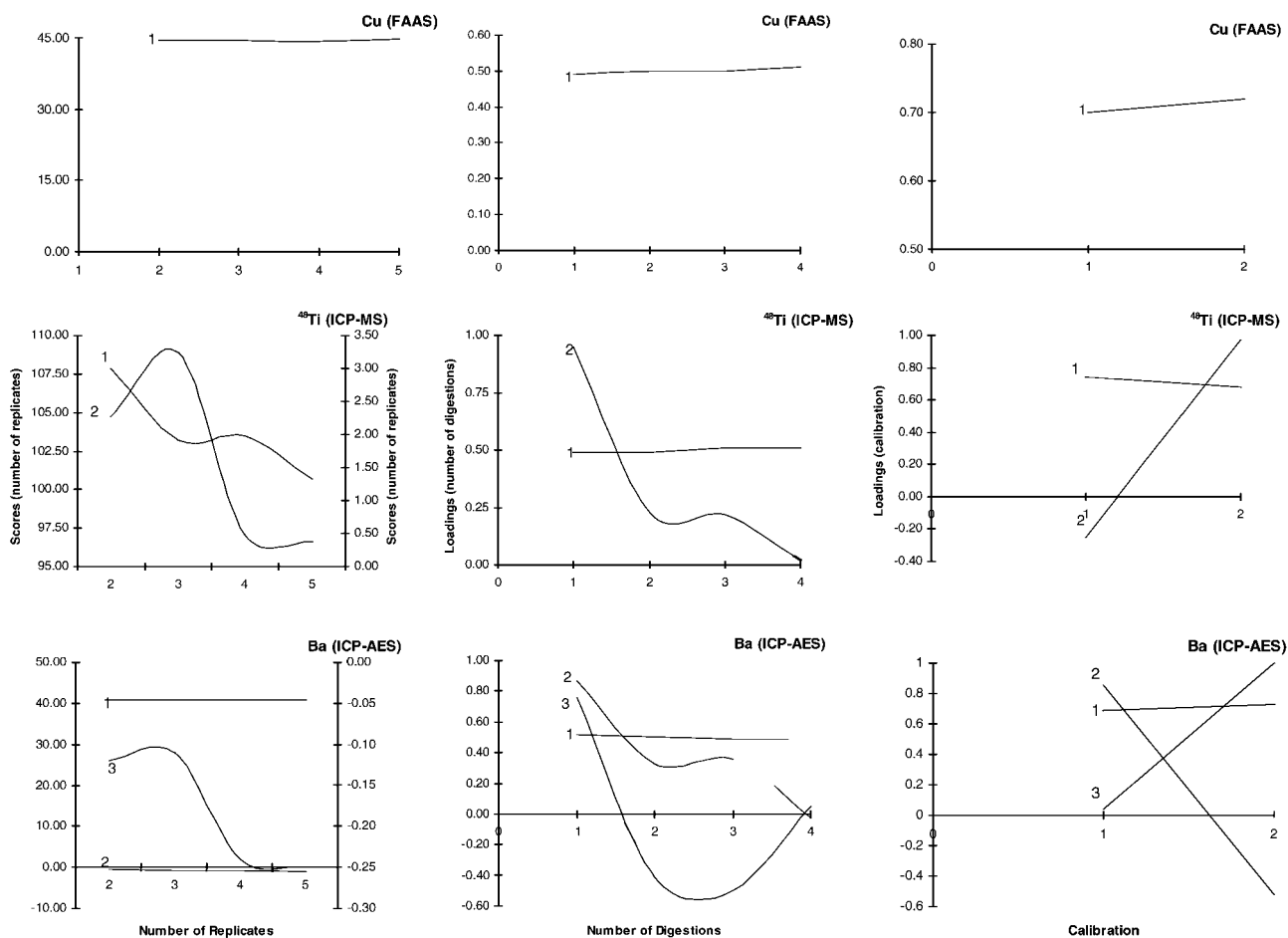
Fig. 3 Loading vectors from one-component (Cu, FAAS), two-component (^{48}Ti, ICP-MS) and three-component (Ba, ICP-AES) PARAFAC models: first PARAFAC component refers to left Y-axis; second and third PARAFAC components refer to right Y-axis.
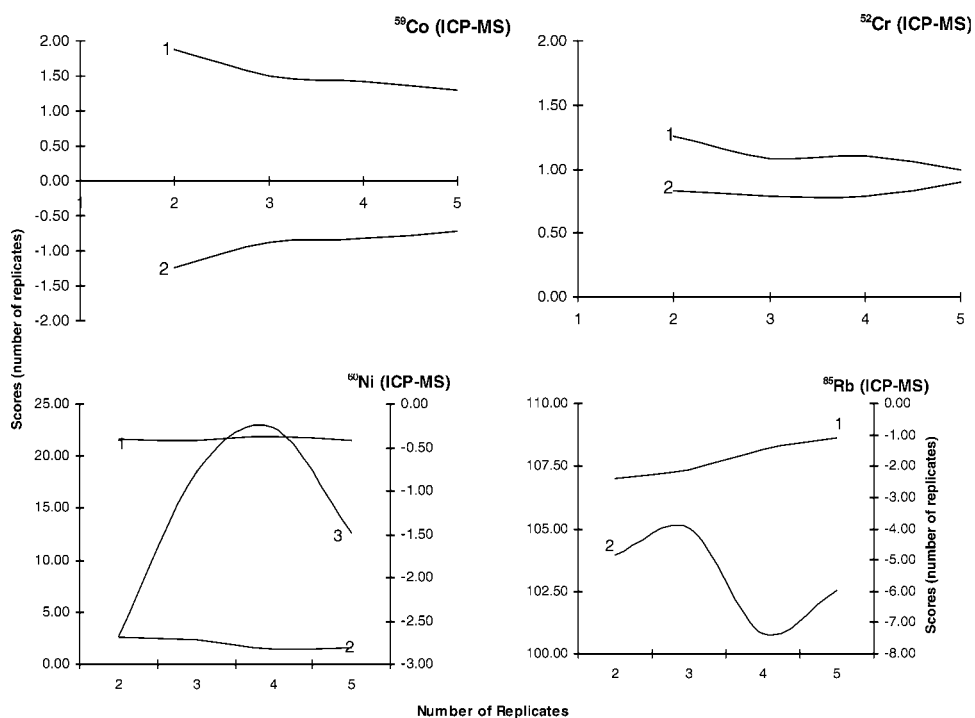

Fig. 4 Number of replicate loading vectors from two-component (^{59}Co, ICP-MS) and three-component (^{60}Ni, ICP-MS) PARAFAC models: first PARAFAC component refers to left Y-axis; second and third PARAFAC components refer to right Y-axis.
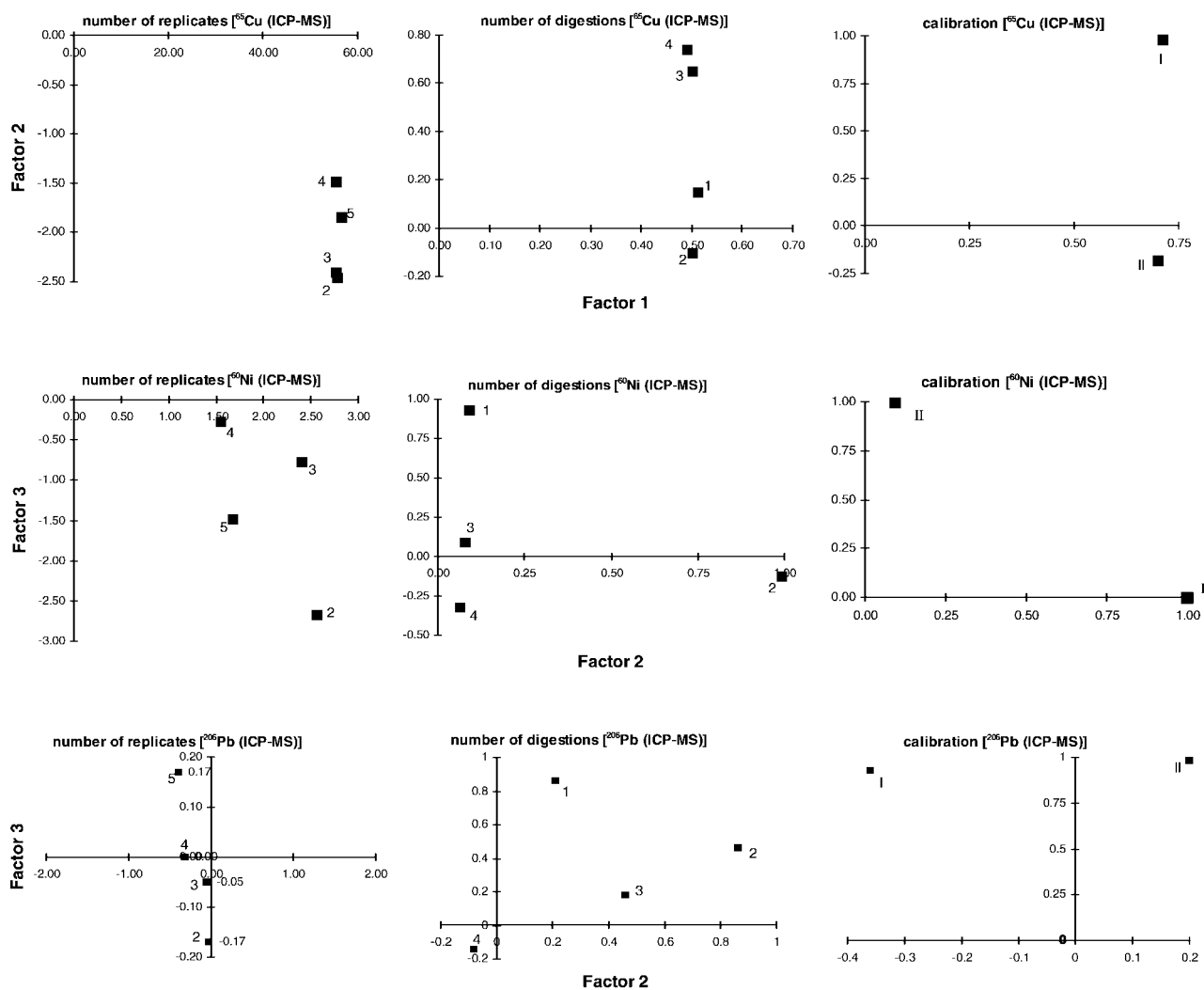
Fig. 5 Projections of loadings in the space formed by the first and second components (two-component solution) for ^{65}Cu (ICP-MS) and by the second and third components (three-components solution) for ^{60}Ni (ICP-MS) and ^{206}Pb (ICP-MS).

that given by the number of digestions; however, it is less important for ICP-AES and FAAS determinations than for ICP-MS determination (Fig. S3†).

Error estimation from PARAFAC

Two examples from PARAFAC decomposition involving both two- and three-component solutions are shown in Tables 8 and 9, respectively. The figures under the heading first component, second components, *etc.*, are the results of the multiplication of loadings according to eqn. 1. It can be seen that the first component gives a data matrix very similar to the original data (also shown in the tables). The second component (for a two-component solution), or the sum of the second and third components for a three-component model, are the differences between the original data and the modelled data given by PARAFAC decomposition. So, the sum of all these terms gives the raw data (original data) and the second component (two-component solution) or the sum of the second and third components (three-component solution) can be assumed to be the error for this measurement. Using this assumption, the sum of squares of second, or second and third, components can be calculated, and these sums for each case are given in Table 10.

Results from analysis of variance (ANOVA)

Analysis of variance (ANOVA) was performed using the full factorial design shown in Table 11, which facilitates all

interactions and the three variables used in the PARAFAC model. The proposed ANOVA model is given below,

Metal concentration,

$$[\text{M}]_{ijk} = R_i + D_j + C_k + Rd_{ij} + Rc_{ik} + DC_{jk}$$

ANOVA tables for the elements and techniques investigated are given in ref. 23. Results from ANOVA were difficult to interpret since some two order interactions were found to be significant, *e.g.*, R/C ("number of replicates"/"calibration") and D/C ("number of digestions"/"calibration"). For instance, the two-order interaction "number of digestions" and "calibration" (D/C) is significant, a scenario that occurs for all cases studied except for the determination of Mg by FAAS.[23] This means that both variables ("number of digestions" and "calibration") are significant. Therefore, to reduce the uncertainty, both the number of digestions and calibrations must be increased. This result may be expected for the variable "number of digestions", but not for the variable "calibration" (more than one aqueous calibration is necessary to reduce systematic errors). In addition, the significance of the two-order interaction D/C means that the uncertainty attributed to the variable "calibration" can be minimised by increasing the number of acid digestions. Similarly, the uncertainty attributed to the variable "number of digestions" can be reduced by using two different aqueous calibrations instead of increasing the number of digestions. Therefore, even for simple systems with

Table 10 Sum of squares obtained after PARAFAC and ANOVA applications

Element	Technique	Sum of squares PARAFAC $(DRC)^a$	Sum of squares ANOVA $(D+R+C+DR+DC+RC)$
Ba[b]	ICP-MS	7.269	16.621
Ba[b]	ICP-AES	2.765	14.478
Ca	ICP-AES	7.23×10^{-3}	2.23×10^{-3}
Ca	FAAS	1.6×10^{-4}	3.40×10^{-3}
Co	ICP-MS	3.476	0.202
Cr	ICP-MS	2.721	0.494
Cu	ICP-MS	17.307	15.610
Fe	ICP-MS	3.63×10^3	4.56×10^3
Fe[b]	FAAS	2.68×10^3	7.04×10^3
Mg	ICP-MS	8.50×10^{-5}	3.08×10^{-3}
Mg	FAAS	2.63×10^{-5}	2.26×10^{-4}
Mn[b]	ICP-MS	5.48×10^3	2.90×10^3
Ni[b]	ICP-MS	27.503	5.184
^{206}Pb[b]	ICP-MS	0.253	0.139
Rb	ICP-MS	54.909	138.320
Sr	ICP-MS	2.420	8.361
Ti	ICP-MS	16.268	95.153
Zn	ICP-MS	31.238	105.830
Zn	FAAS	213.107	45.403

$^a D$ = number of digestions; R = number of replicates; C = calibration.
[b]Three-components model.

Table 11 $4^1 \times 4^1 \times 2^1$ Full Factorial Design matrix

Experiment	Number of replicates	Number of digestions	Calibration
1	2	1	I
2	3	1	I
3	4	1	I
4	5	1	I
5	2	2	I
6	3	2	I
7	4	2	I
8	5	2	I
9	2	3	I
10	3	3	I
11	4	3	I
12	5	3	I
13	2	4	I
14	3	4	I
15	4	4	I
16	5	4	I
17	2	1	II
18	3	1	II
19	4	1	II
20	5	1	II
21	2	2	II
22	3	2	II
23	4	2	II
24	5	2	II
25	2	3	II
26	3	3	II
27	4	3	II
28	5	3	II
29	2	4	II
30	3	4	II
31	4	4	II
32	5	4	II

three variables, the interpretation of ANOVA results is difficult, and clearly for systems involving a large number of variables the interpretation is yet more complex.

In contrast to ANOVA, the proposed PARAFAC model is as follows:

$$\text{Metal concentration, } [\text{M}]_{ijk} = RDC_{ijk}$$

Thus PARAFAC offers a more straightforward model as the two-order interactions, three-order interactions and so on, are not taken into account to describe the system. As can be seen, the three-order interaction, RDC_{ijk}, was not significant for ANOVA models, but this did form the basis for the PARAFAC model, and so the two methods point to quite different solutions.

The sum of squares obtained after ANOVA are given together with those obtained by PARAFAC in Table 10 to indicate the differences obtained using the two approaches. According to Bro and Heimdal[18] this difference may be expected because PARAFAC and ANOVA provide very different models of the data. ANOVA suggests six effects while PARAFAC suggests one. Furthermore, PARAFAC offers a straightforward generalisation of the effects related to other levels of the variables, and not just the ones explicitly used in the experiment.

The results of the sum of squares in Table 10 indicate that ICP-MS affords the worst precision when compared with ICP-AES and FAAS. This is in agreement with results reported elsewhere following the study of systematic errors using alternative chemometric techniques.[23]

Conclusions

PARAFAC has been used successfully to decompose a three-way data set formed from multi-element determinations in order to assess the systematic error of such measurements. The scores and loadings for the first PARAFAC component are an estimation of the raw data (without error) and the scores and loadings for the second and third PARAFAC components provide an estimation of the systematic error of the measurement.

PARAFAC also gives an easy to interpret and robust model of the data because two-order interactions are not used to build the model. Such two-order interactions are used in the ANOVA model and prove difficult to interpret when used on multi-variable systems. Even for the study of a very simple system (only three variables), PARAFAC may provide an attractive alternative procedure to assess the sum of squares and the uncertainty (normally obtained from a standard ANOVA). In addition, because of the multi-way nature of PARAFAC, the visualisation of scores and loadings for systems with more than two variables is easier than when using a two-way technique such as PCA.

In this study the systematic error was found to be more important in ICP-MS determinations than for ICP-AES or FAAS. In fact, a one-component PARAFAC model was found for several data sets related to ICP-AES and FAAS determinations, indicating a small systematic error and good precision for these two techniques. The variables "number of digestions" and "calibration" were found to be the most significant and they contribute more to the uncertainty of the measurements for both ICP-AES and ICP-MS techniques. Finally, the findings of this study are in agreement with those obtained using ANOVA and an experimental design approach,[23] although the results obtained are easier to interpret.

References

1 R. Bro, *Chemometrics Intell. Lab. Syst.*, 1997, **38**, 149.
2 P. Seiden, R. Bro, R. Poll and L. Munck, *J. Agric. Food Chem.*, 1996, **44**, 3202.
3 L. Noergaard, *J. Chemometrics*, 1996, **10**, 615.

4   A. R. Muroski, K. S. Booksh and M. L. Myrick, *Anal. Chem.*, 1996, **68**, 3534.
5   K. S. Booksh, A. R. Muroski and M. L. Myrick, *Anal. Chem.*, 1996, **68**, 3539.
6   J. L. Beltran, J. Guiteras and R. Ferrer, *Anal. Chem.*, 1998, **70**, 1949.
7   J. C. G. Esteves da Silva and S. A. G. Novais, *Analyst*, 1998, **123**, 2067.
8   J. L. Beltran, R. Ferrer and J. Guiteras, *Anal. Chim. Acta*, 1998, **373**, 311–319.
9   R. Bro, *Chemometrics Intell. Lab. Syst.*, 1999, **46**, 133.
10 R. D. Jiji, G. A. Cooper and K. S. Booksh, *Anal. Chim. Acta*, 1999, **397**, 61.
11 J. L. Beltran, J. Guiteras and R. Ferrer, *J. Chromatogr.*, 1998, **802**, 263.
12 P. Hindmarch, K. Kavianpour and R. G. Brereton, *Analyst*, 1997, **122**, 871.
13 G. G. Andersson, B. K. Dable and K. S. Booksh, *Chemometrics Intell. Lab. Syst.*, 1999, **49**, 195.
14 A. K. Smilde, P. H. Van der Graaf, D. A. Doornbos, T. Steerneman and A. Sleurink, *Anal. Chim. Acta*, 1990, **235**, 41.
15 J. C. G. Esteves da Silva and C. J. S. Oliveira, *Talanta*, 1999, **49**, 889.
16 S. Bijlsma, D. J. Louwerse, W. Winding and A. K. Smilde, *Anal. Chim. Acta*, 1998, **376**, 339.
17 S. Bijlsma, D. J. Louwerse and A. K. Smilde, *J. Chemometrics*, 1999, **13**, 311.
18 R. Bro and H. Heimdal, *Chemometrics Intell. Lab. Syst.*, 1996, **34**, 85.
19 R. L. Watters, *Spectrochim. Acta, Part B*, 1991, **46**, 1593.
20 R. Bro, Multi-way Analysis in the Food Industry, Ph.D. thesis, Amsterdam, 1998.
21 The N-way toolbox for Matlab, version 1.02, KVL, Denmark, 1999.http://newton.foodsci.kvl.dk/matlab/nway/toolbox/index.htm.
22 A. Marcos, A. Fisher, G. Rea and S. J. Hill, *J. Anal. At. Spectrom.*, 1998, **13**, 521.
23 A. Moreda-Piñeiro, A. Marcos, A. Fisher and S. J. Hill, *J. Anal. At. Spectrom*, 2001, DOI: 10.1039/b008847l.