MULTIELEMENT ANALYTICAL SPECTROMETRY AS A DATA SOURCE FOR CORRELATIVE CLASSIFICATION OF SAMPLES

P. Serapinas and Ž. Ežerinskis

Joint Isotope Analysis Laboratory, Institute of Theoretical Physics and Astronomy of Vilnius University, A. Goštauto 12, LT-01108 Vilnius, Lithuania
E-mail: serapinas@pfi.lt

Fast increase of correlations between the concentrations of elements is observed when diversity of the sample sources decreases. Data of high resolution sector field ICP MS (inductively coupled plasma mass spectrometry) measurements of concentrations of 25 elements in wines were tested in the present study. As much as about ten or more pairs of correlating elements (correlation coefficient $r > 0.9$) were found for different wines of similar provenance. Correlation pattern is batch and sample specific. It represents the similar sources of elements, similar element quantity governing processes. If the number of the correlating pairs of elements is large, the mean within this list correlation coefficient can be calculated for individual samples, relative to the values of concentrations characteristic of different batches, and it can be used for identification of samples. Data on good potential for applicability of such individual sample-specific correlative characteristics in testing the provenance of samples are presented as examples.

Keywords: multielement mass spectrometry, correlation, pattern recognition

PACS: 02.50.Sk, 07.05.Kf, 07.05.Rm, 82.80.Ms

1. Introduction

Development of multielement analysis techniques, such as neutron activation analysis, inductively coupled plasma optical emission spectrometry, and especially inductively coupled plasma mass spectrometry (ICP MS) enabled the measurement of quantities of a large number of elements in the same test sample. The interest in comprehensive information on the composition of numerous samples including somewhere from about 20 to 60 elements is common in recent geochemistry and environmental research [1–5], clinical and forensic toxicology [6, 7], and food authenticity studies (e. g., [8–12] and citations therein). Various techniques of analysis of such data (Anova, factor, principal component, discriminant, cluster, dispersion and correlation analyses, neural networks [5, 8, 10, 13]) are used to discover the similarities in behaviour of different elements and the common sources of their prevalence. The data often are used for more or less detailed analysis of the governing processes. Many reasons and processes of similar changes of the quantities of elements in plants and plant products are identified and being discussed, including those related to soil and anthropogenic impact, such as weathering of minerals, atmospheric aerosols rich in

industrial fumes and exhaust gases, bioaccumulation, equipment used in production of final materials and in treatment of incoming raw ones. Efforts to link the concentrations of elements, the correlations between them and the patho-physiological situations (e. g., [14–15]) are increasing.

Implications on importance of correlations between concentrations of chemical elements, or other substances, are often appearing in literature (including the citations above) because of the informative and clear physical and chemical contents included in this concept. Nevertheless, the quantitative studies and direct application of the quantitative characteristics of correlations in analytical measurements is neither conventional nor elaborated experimental tool. The objective of this publication is to show that the modern multielement analysis enables cumulation of data for the quantitative pattern of correlations between concentrations of elements, aid quantification procedures and application of correlations not only to study the propagation of chemical substances but also for characterization and classification of samples.

2. Method of measurement and results

Results of the measurement of concentrations of chemical elements in the bottled wines produced in different countries were used in the present study. About thirty elements in 65 wine samples, from Bulgaria, Chile, France, Hungary, Italy, Spain, and USA (California)[1] mainly, were measured. The samples were diluted by a factor 1:4 for measurement of microelements and 1:30 for measurement of elements present at greater concentrations. Merck ICP Multi Element Standard Solution VI as a reference material and Tm as an internal standard were used for calibration. The reference material was selected to include the set of elements of the main interest while at the same time it was the determining factor of the element list. Nitric acid (Suprapur®, 65.3%, Merck) was added to the diluent to get a 2% acid solution for stabilization of the subsamples.

High resolution double focusing sector field ICP mass spectrometer Element2 (Thermo Finnigan MAT, Bremen, Germany), concentric 1 ml/min flow rate nebulizer, and Scott double pass spray chamber were used for the measurements. The working conditions of the instrument were: rf power 1100 W, argon plasma gas flow rate 14 l/min, sample gas 1.35 l/min. Repeatability of the measurement of the analytical signals usually did not exceed 2–3% resulting in expanded standard ($k = 2$) uncertainty of the measured concentrations of elements within ±10% mainly.

As an example, the mean values of concentrations of elements in different samples from Chile (Central Valley), Spain, and Toro region of Spain and their standard deviations together with the mean and standard deviation (sd) from the whole data set are presented in Table 1. Variation of the element concentrations inside a region usually is large, mean relative standard deviation (rsd) is about 50%, obviously representing the differences in the soil of provenance, agriculture, and wine production technologies. The problem is to select, if possible, such a group of indicators that could characterize the batches of interest. It is evident that discrimination between the numerous batches according to the mean concentration values is not a simple problem. Elaboration of a correlative method for characterization of individual samples and application of such characteristics for sample classification is the main objective of the present study.

---

[1] Here and further the names of the countries are used as codes for the sample batches. More precise definition of the origin of samples is usually needed in wine pattern recognition studies. See discussion below for more details.
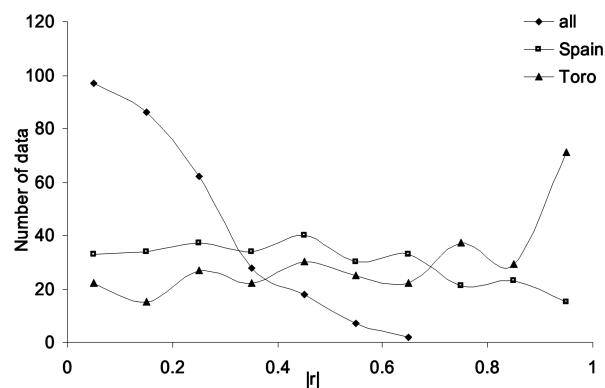

Fig. 1. Number distribution of values of the correlation coefficients (300 for each set): all data, Spain, and Toro (region of Spain). Numbers of the absolute values of correlation coefficients in the interval $|r| \pm 0.05$ are presented.

3. Contents of correlative description

Variability of the concentration data can occur because of the indicators being not representative of the batch or due to important processes that govern the quantities of the analytes. Correlation between the concentrations of elements as a source of information on similarity of the processes on which those concentrations depend is well-known in analytical science. The data even on two elements can be of interest in determining the general sources of elements and the role of different processes. Modern multielement analysis provides extremely wide possibilities for study of such correlations.

The matrix of the correlation coefficients, as a quantitative measure of correlation in the variation of concentrations, of $n = 22$ elements for batch of wines from Chile, as an example, is presented in Table 2. The diagonal values are autocorrelations, so they automatically are equal to 1. The matrix is symmetric, it includes $n(n-1)/2$ independent values. The supplementary symmetric part is included into the table only for convenience. So the number of independent indicators increases almost as n^2 and 300 values are obtained for data set of 25 elements. The largest absolute correlation coefficients (from 0.95 to 1) are marked in bold in the table indicating that significant correlations between the concentrations of some elements as well as no correlations between some other ones are found. Number distributions of the absolute values of correlation coefficients for wines from Spain and from Toro (a comparatively small region of Spain) together with those obtained from the whole measured data set are presented in Fig. 1.

In the last case, when the selection of the geographical regions is almost accidental, close to zero corre-

Table 1. Examples of concentrations (ng/g) of elements in wines from different regions.

	All data		Chile		Spain		Toro (Spain)	
	mean	sd	mean	sd	mean	sd	mean	sd
Sr	860	430	700	67	1000	650	690	65
Al	1100	825	490	120	740	515	350	90
Ca	89000	17000	70000	12000	78000	19000	63000	2600
Fe	4000	2160	2600	910	3300	1000	3800	980
Li	37	43	12	9	52	17	61	9
Tl	0.29	0.21	0.39	0.30	0.39	0.28	0.30	0.02
U	1.20	0.70	0.66	0.24	1.78	0.46	1.71	0.08
V	81	72	81	52	142	105	83	19
Cr	28	13	21	5	25	8	25	4
Co	4.4	2.0	3.1	1.1	4.2	2.8	2.8	0.4
Ni	29	11	15	2	29	11	29	4
Cu	151	122	109	46	99	104	172	126
Zn	735	520	590	350	420	285	270	79
Ga	0.33	0.17	0.27	0.07	0.26	0.23	0.12	0.03
Se	5	3	8	5	4	3	2	1
Ba	159	86	181	104	146	55	200	11

lation coefficients between concentrations of elements prevail, and absolute correlation coefficient values seldom exceed 0.5. For separate countries (Spain) distributions of correlation coefficients only slightly decrease with greater coefficient values whereas for small regions (Toro region, Spain) the number of pairs with large positive or negative correlation coefficients essentially increases. As an example, at least two groups of correlating elements can be distinguished in the matrix of correlation coefficients for the Chile Central Valley samples: Co, Mn, Zn and Be, U, Ni, Tl, Se, as well as individual pairs (V/As, Pb/Cu, etc.). Naturally, different mathematical techniques could be used to detect such regularities. In the present study we use the direct estimates of the correlation coefficients. Even in this case various techniques of data presentation and analysis are applicable. For example, multiplication of the correlation matrices can help to detect similarities and dissimilarities of the two data sets. Sums of the cross-correlation coefficients of an element with the other ones for different elements and sample batches also vary essentially, e. g., from about −4 for Al, Ca, Pb to about 3 for Mn, Co, Ni, Tl, and Zn for Chile (Table 2, bottom). The pattern for other countries is different, i. e., batch specific.

The largest absolute correlation coefficient values are most indicative in the correlation analysis. As an example, the correlation coefficients between some elements found for wines produced in different countries are presented in Table 3. For comparison, the correlations between the concentrations of elements in German wines from different regions were measured [9] and the coefficients larger than about 0.7 were found

significant. High correlation (0.99) was observed [11] between wine and juice, wine and soil compositions.

Most characteristic correlations, as well as the whole correlation pattern or some integral characteristics like the sums mentioned above can be used as a characteristic of the batch. Nevertheless, the correlative data so far have seemed to be inapplicable for the *characterization of a single, individual sample*. To our knowledge, such approaches are missing. We find that the multielement spectrometry opens new possibilities for the corresponding description as well.

If variation of concentrations of a large number of analytes is determined by the same source, concentrations of those elements correlate. In spite of some inevitable variations of the correlation coefficients because of the limited accuracy of determination, those correlations are indications of some quality characteristic of the batch in general and of the individual samples as well. It is natural to expect that this quality can be characterized by some mean parameter, e. g., mean correlation coefficient, applicable as a characteristic of the whole batch and the particular samples. Rather large number of pairs of the correlating elements must be included for the representative and specific description. If this number is really large, the statistical methods of analysis can be applicable to quantify the character of deviations of the concentrations in an individual sample in comparison to the mean values characteristic of the whole batch.

As an example, the analysis of the data on Chile Central Valley wine samples is presented below. As can be seen from Table 2, the high level of correlations was found for the following pairs of elements: K/Na, K/Rb,

Table 2. Matrix of correlation of element concentrations in Chile (Central Valley) wines.

	Rb	Sr	Na	Mg	Al	Ca	Mn	K	Li	Be	Tl
Rb	*1.00*	−0.17	0.85	0.73	0.22	0.07	0.55	0.90	−0.53	−0.16	0.13
Sr	−0.17	*1.00*	−0.66	0.01	**−0.95**	−0.09	0.26	−0.56	−0.62	0.70	0.54
Na	0.85	−0.66	*1.00*	0.59	0.68	0.14	0.26	**0.99**	−0.09	−0.52	−0.22
Mg	0.73	0.01	0.59	*1.00*	0.24	0.57	0.07	0.72	−0.57	−0.31	−0.09
Al	0.22	**−0.95**	0.68	0.24	*1.00*	0.35	−0.40	0.62	0.53	−0.78	−0.60
Ca	0.07	−0.09	0.14	0.57	0.35	*1.00*	−0.74	0.21	0.18	−0.05	0.07
Mn	0.55	0.26	0.26	0.07	−0.40	−0.74	*1.00*	0.26	−0.69	0.11	0.16
K	0.90	−0.56	**0.99**	0.72	0.62	0.21	0.26	*1.00*	−0.21	−0.51	−0.20
Li	−0.53	−0.62	−0.09	−0.57	0.53	0.18	−0.69	−0.21	*1.00*	−0.11	−0.13
Be	−0.16	0.70	−0.52	−0.31	−0.78	−0.05	0.11	−0.51	−0.11	*1.00*	0.94
Tl	0.13	0.54	−0.22	−0.09	−0.60	0.07	0.16	−0.20	−0.13	0.94	*1.00*
Pb	−0.92	−0.21	−0.59	−0.69	0.16	0.04	−0.70	−0.67	0.77	−0.11	−0.33
U	−0.23	0.37	−0.41	−0.52	−0.52	−0.08	−0.02	−0.46	0.26	0.91	0.89
V	−0.21	−0.09	−0.11	−0.27	−0.01	−0.71	0.40	−0.14	−0.19	−0.54	−0.71
Co	0.55	0.43	0.18	0.13	−0.54	−0.64	**0.98**	0.21	−0.77	0.28	0.33
Ni	−0.30	0.56	−0.57	−0.61	−0.74	−0.37	0.22	−0.61	0.04	0.92	0.82
Cu	−0.88	−0.14	−0.62	−0.85	0.02	−0.08	−0.58	−0.72	0.82	0.19	−0.01
Zn	0.56	0.02	0.39	−0.04	−0.22	−0.78	**0.96**	0.35	−0.48	0.04	0.12
Ga	−0.54	0.54	−0.73	−0.69	−0.68	−0.20	−0.08	−0.77	0.23	0.89	0.73
As	−0.29	0.26	−0.35	−0.34	−0.37	−0.78	0.51	−0.36	−0.36	−0.22	−0.45
Se	0.17	0.70	−0.26	0.11	−0.68	0.20	0.15	−0.19	−0.34	0.91	**0.96**
Ba	0.20	0.01	0.12	−0.30	−0.23	**−0.95**	0.86	0.07	−0.35	−0.13	−0.18
$\sum -1$	0.70	0.91	−0.94	−2.11	−3.90	−3.63	2.54	−1.08	−2.60	2.48	2.77

	Pb	U	V	Co	Ni	Cu	Zn	Ga	As	Se	Ba
Rb	−0.92	−0.23	−0.21	0.55	−0.30	−0.88	0.56	−0.54	−0.29	0.17	0.20
Sr	−0.21	0.37	−0.09	0.43	0.56	−0.14	0.02	0.54	0.26	0.70	0.01
Na	−0.59	−0.41	−0.11	0.18	−0.57	−0.62	0.39	−0.73	−0.35	−0.26	0.12
Mg	−0.69	−0.52	−0.27	0.13	−0.61	−0.85	−0.04	−0.69	−0.34	0.11	−0.30
Al	0.16	−0.52	−0.01	−0.54	−0.74	0.02	−0.22	−0.68	−0.37	−0.68	−0.23
Ca	0.04	−0.08	−0.71	−0.64	−0.37	−0.08	−0.78	−0.20	−0.78	0.20	**−0.95**
Mn	−0.70	−0.02	0.40	**0.98**	0.22	−0.58	**0.96**	−0.08	0.51	0.15	0.86
K	−0.67	−0.46	−0.14	0.21	−0.61	−0.72	0.35	−0.77	−0.36	−0.19	0.07
Li	0.77	0.26	−0.19	−0.77	0.04	0.82	−0.48	0.23	−0.36	−0.34	−0.35
Be	−0.11	0.91	−0.54	0.28	0.92	0.19	0.04	0.89	−0.22	0.91	−0.13
Tl	−0.33	0.89	−0.71	0.33	0.82	−0.01	0.12	0.73	−0.45	**0.96**	−0.18
Pb	*1.00*	0.09	0.19	−0.76	0.06	0.93	−0.63	0.32	0.13	−0.42	−0.27
U	0.09	*1.00*	−0.57	0.09	0.93	0.43	0.02	0.92	−0.35	0.74	−0.13
V	0.19	−0.57	*1.00*	0.25	−0.28	0.01	0.38	−0.32	0.93	−0.70	0.74
Co	−0.76	0.09	0.25	*1.00*	0.32	−0.62	0.90	0.03	0.42	0.35	0.75
Ni	0.06	0.93	−0.28	0.32	*1.00*	0.39	0.21	**0.95**	0.01	0.70	0.14
Cu	0.93	0.43	0.01	−0.62	0.39	*1.00*	−0.47	0.60	0.02	−0.17	−0.20
Zn	−0.63	0.02	0.38	0.90	0.21	−0.47	*1.00*	−0.09	0.42	0.04	0.89
Ga	0.32	0.92	−0.32	0.03	**0.95**	0.60	−0.09	*1.00*	−0.05	0.63	−0.08
As	0.13	−0.35	0.93	0.42	0.01	0.02	0.42	−0.05	*1.00*	−0.41	0.76
Se	−0.42	0.74	−0.70	0.35	0.70	−0.17	0.04	0.63	−0.41	*1.00*	−0.26
Ba	−0.27	−0.13	0.74	0.75	0.14	−0.20	0.89	−0.08	0.76	−0.26	*1.00*
$\sum -1$	−3.61	2.35	−1.93	2.87	2.80	−1.94	2.61	1.62	−0.87	2.23	1.44

Mn/Co, Mn/Zn, Ni/U, Ni/Ga, Cu/Pb, V/As. The mean value of the correlation coefficient between the corresponding concentrations was 0.95, with the standard deviation 0.03. Deviations of concentrations of elements of those pairs for each particular sample from the mean values can be used to calculate the corresponding mean correlation coefficient characteristic of the individual samples, i. e.,

$$r = \frac{\sum_i \left(c_{ai} - \overline{c_{ai}}\right) \cdot \left(c_{bi} - \overline{c_{bi}}\right)}{\sqrt{\sum_i \left(c_{ai} - \overline{c_{ai}}\right)^2 \cdot \sum_i \left(c_{bi} - \overline{c_{bi}}\right)^2}} . \quad (1)$$

Table 3. Comparison of correlation coefficients for some pairs of elements
for different batches.

	CA	CHL	ARG	ESP	ITA	FRA	BGR	HUN
Sr/Li	−0.42	−0.62	0.91	0.16	**0.97**	0.32	0.59	−0.28
Na/K	−0.68	**0.99**	0.76	0.06	−0.9	0.88	−0.87	−0.7
Na/Be	**0.99**	−0.52	0.55	−0.43	0.68	0.66	0.7	0.74
Mg/U	−0.12	−0.52	0.46	−0.62	0.05	−0.93	−0.7	**0.92**
Mg/Co	0.18	0.13	**0.99**	−0.93	0.12	−0.32	−0.85	−0.58
Mg/Ba	**0.99**	−0.3	0.85	0.21	0.03	−0.12	**0.99**	0.85
Mn/Co	0.09	**0.98**	0.84	0.33	0.69	0.52	−0.21	0.7
Mn/Zn	0.52	**0.96**	0.64	0.55	0.48	0.72	−0.36	0.37
Be/Se	−0.17	0.91	−0.58	**0.96**	0.29	0.78	−0.06	−0.36
Cr/Zn	0.73	0.02	0.62	−0.03	0.34	**0.95**	0.83	0.09

Table 4. Initial data for calculation of the positive correlation coefficients for Chile wine samples *14*, *34*, *41*, *47*, and *55*
according to Eq. (1).

Pair	Element a	Relative deviations from mean value in samples					Element b	Relative deviations from mean value in samples				
		14	*34*	*41*	*47*	*55*		*14*	*34*	*41*	*47*	*55*
K/Na	K	1.045	0.907	−0.670	−0.622	−0.660	Na	1.305	0.745	−0.758	−0.696	−0.596
K/Rb	K	1.045	0.907	−0.670	−0.622	−0.660	Rb	0.320	0.304	−0.071	−0.118	−0.435
Mn/Co	Mn	0.339	−0.093	−0.061	0.532	−0.718	Co	0.172	−0.030	0.071	0.363	−0.577
Mn/Zn	Mn	0.339	−0.093	−0.061	0.532	−0.718	Zn	0.659	−0.255	−0.138	0.522	−0.789
Ni/U	Ni	−0.018	−0.177	0.212	0.041	−0.058	U	0.020	−0.394	0.582	−0.136	−0.072
Ni/Ga	Ni	−0.018	−0.177	0.212	0.041	−0.058	Ga	−0.120	−0.319	0.383	0.029	0.026
Cu/Pb	Cu	−0.128	−0.558	0.143	−0.066	0.609	Pb	−0.429	−0.641	−0.119	−0.033	1.222
V/As	V	0.004	−0.230	−0.825	0.941	0.110	As	−0.194	−0.278	−0.444	0.977	−0.062

Here i is the index of the correlating pair. In general, the correlating elements, a and b, and the corresponding mean concentration values are different for every i and batch samples being considered.

Thus, the usual statistical analysis with the calculation of covariation and averaging over the samples can be substituted by *calculation of covariation and averaging over similar* (in sense of correlation) *pairs of elements*. An example of data for calculation of the mean correlation coefficient r corresponding to the case mentioned above is presented in Table 4. The relative deviations from the mean values were used. The mean correlation coefficients for the particular samples in the range between 0.75 and 0.93 (Table 5) were obtained resulting in the mean value 0.86 with the standard deviation 0.075. This illustrates acceptable accuracy of the equivalence of averaging over the samples and correlating pairs inside the selected group of the correlating elements even for the small numbers of samples and correlating pairs. Naturally, the deviations from the same mean are used to calculate the batch characteristic correlation coefficients, while deviations from independent means for different elements enter the covariation in a sample characteristic correlation coefficient in Eq. (1). Although some lower values of the sample characteristic correlation coefficients can be expected

for this reason, the difference decreases if the numbers of representative samples and correlating pairs increase.

Similar correlation coefficients can be calculated for other batches, e. g., other countries. In the particular case, the correlation coefficient representing the pairs of correlating elements characteristic of Chile, as presented in Table 4, was much lower (mean values −0.09, −0.20, 0.14, and −0.31 correspondingly) when calculated for wines from France, Italy, Argentina, and California (compare with Table 5) illustrating high potential of the possibility of discrimination between batches. Nevertheless, while attempting to classify all the measured wines from 5 countries mentioned above by application of the positive correlations of 6 pairs of elements, three French wines could not be classified unequivocally. Switching from 6 pairs of correlating elements to 8, as well as taking account of negative correlations enabled full classification. Such a statistical approach is possible only due to a large enough number of the correlating pairs available for the modern multielement spectrometry. Possibility of the quantification of an individual sample is decisive for such discrimination.

From the practical examples it follows that the classification according to the data on about five to ten reference samples and a similar number of correlating pairs can be possible, while, as follows from the additional

Table 5. Example of application of correlation coefficients calculated for individual samples for classification of the samples (see text for details).

Region			Chile			France	Italy
Sample	*14*	*34*	*41*	*47*	*55*	*56*	*30*
Largest positive correlation coefficient as for							
California	0.27	0.29	−0.08	0.14	0.54	0.01	0.34
Chile	**0.82**	**0.87**	**0.75**	**0.93**	**0.92**	0.59	0.19
France	−0.14	0.14	0.25	0.43	−0.12	**0.93**	−0.04
Italy	−0.19	0.53	−0.92	−0.22	−0.04	**0.77**	**0.91**
Toro (Spain)	−0.09	0.17	−0.23	0.09	−0.32	0.03	0.32
Largest negative correlation coefficient as for							
California	0.11	−0.43	−0.18	0.41	−0.20	0.34	0.12
Chile	**−0.76**	**−0.93**	**−0.70**	**−0.77**	**−0.92**	−0.14	0.21
France	0.01	0.04	0.25	−0.08	0.49	**−0.72**	0.51
Italy	0.43	0.58	0.45	0.37	0.40	0.36	**−0.88**
Toro (Spain)	0.61	0.72	0.57	0.11	−0.11	0.22	−0.47

analysis of the accuracy of the correlation coefficients as presented below, about ten reference samples and correlating pairs ought to be preferred as the lower limit.

4. Accuracy of results

First of all the reliability of description depends on the accuracy of initial data, i. e., the accuracy of the correlation coefficients. It is well known from statistics that for $n = 10$ or more measurements the correlation can be detected at the level of confidence $P = 0.95$ if the experimental value $|r_n| \cdot \sqrt{n-1}$ exceeds the critical value 1.9 (for comparison, the corresponding critical value for $P = 0.999$ is 2.6). Ten measurements is an acceptable number to regard the correlation coefficient of about 0.9 or 1 as significant. In correspondence with this condition, element pairs with correlation coefficients of about or larger than 0.9 by absolute value were selected for analysis. It is hard to evaluate the dependence of the accuracy of correlation coefficients on the accuracy of the input data analytically in details. In Fig. 2 the results of some modelling calculations concerning data sets characteristic of the multielement spectrometry are presented.

Random data, scattered within the ranges from ±2% to ±30% of the measured values, were added to the initial data and variations of the coefficients of correlations were quantified for 20, 10, and 5 measured values. The results, as presented in Fig. 2, confirm the importance of the number of the measured values (samples, correlating pairs of elements) and show good potential of the ICP mass spectrometry: if the relative standard uncertainties of the data do not exceed 10% (usual for ICP MS measurements) then for about ten measured values
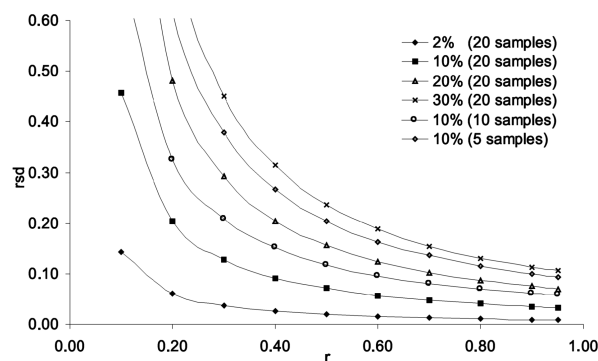

Fig. 2. Dependence of the relative standard deviation of the coefficients of correlation on the absolute value of the correlation coefficient for 20, 10, and 5 samples and different scatter of the initial data.

the rsd of the correlation coefficients within 0.1 can be expected for correlation coefficients as low as 0.6.

Application of correlative characteristics of individual samples for pattern recognition includes the listing of characteristic correlating pairs of elements for each batch, calculation of corresponding mean correlation coefficients for individual samples, and analysis of the data. An example of the data obtained is presented in Table 5. Statistical criteria of hypothesis testing, dispersion analysis, or outlier tests can be used to classify the data. The values in Table 5 that were found in agreement with the reference values for the regions were marked in bold. As an example, according to the list of positively correlating pairs of elements the sample *56* could be classified as from France or Italy, but the negative correlation coefficient corresponds only to that for France.

Selection of the acceptable geographical region or some other criteria for classification of batches is a problem. As can be seen from Fig. 1, when very dif-
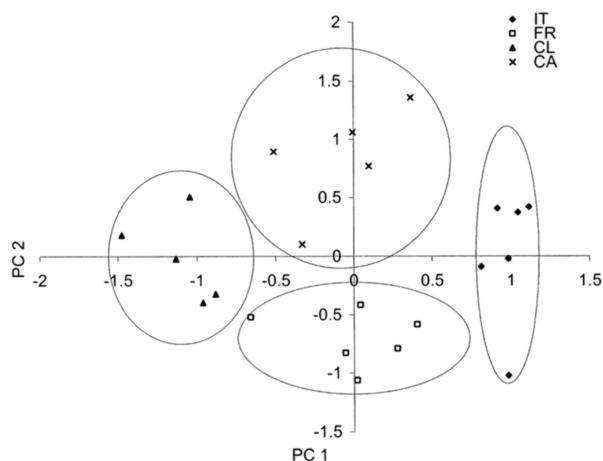
Fig. 3. Scatter plot of PCA analysis of correlation coefficients for wines from Italy, France, Chile, and California.

ferent samples are included into a batch, the correlation coefficients decrease. Sufficient number of correlating pairs of elements can be regarded as a test that the variability within the batch is not too wide to regard that batch as a potential classification unit. In correspondence with the material presented above the analysis of both the positive and the negative correlations including at least about ten independent pairs of elements in each of the two groups is needed. No doubt the samples must be representative of the batch.

Multivariate techniques, such as Principal Component Analysis (PCA) and some other, are used for classification and concise presentation of large data sets. In Fig. 3 the results of the PCA analysis of the correlation coefficients, similar to Table 5, for wine samples from California, Chile, France, and Italy are presented. The axes of the ellipses correspond to the confidence probability of 0.95 (Student's coefficients included) for each of the two principal components. As follows from the picture, the classification of the samples at such a confidence level is possible. Nevertheless, two principal components explain only about 58% of the variations of the data, three components explain 74% of the variance, and even four components explain only 82% of the variance. So the problem is that dimension of the space that can be used for figures is too small. It is natural, as the indicators corresponding to all the regions (if the correlation coefficients involve a sufficient number of pairs) are equally essential, and dimension of the space cannot be reduced without significant loss of information. Just the data that are in some special intervals of the values are most important here. The statistical data analysis seems more promising in the case. However, especially if the range of correlation coefficients being regarded is wide, caution is needed concerning possible data devi-

ation from the normal distribution. If only the information about the whole batch in general is considered and the integral data for batches such as the full set of correlation coefficients, sums over elements (bottom of Table 2), etc., are used, then application of the concise presentation affordable by the multivariate techniques is suitable and convenient.

As follows from the material presented above, the modern analytical spectrometry provides a lot of data for analysis of the correlation relations between the analytes that may be of interest for identification and studies of the physical and chemical processes and identification of the individual samples. Application of correlations between elements enables easy elimination of non-specific information and related noise and highlights the information of the physical interest. The method is proposed to extend the applicability of the quantitative correlative description used for characterization of the batches to single, individual, samples, and to enable direct application of the correlations between elements for sample classification.

Acknowledgements

References

[1] C. Reimann, G. Kashulina, P. de Caritat, and H. Niskavaara, Multielement, multi-medium regional geochemistry in the European Arctic: Element concentration, variation and correlation, Appl. Geochem. **16**(7–8), 759–780 (2001).
[2] J. Hu, B.S. Zheng, R.B. Finkelman, B.B. Wang, M.S. Wang, S.H. Li, and D.S. Wu, Concentration and distribution of sixty-one elements in coals from DPR Korea, Fuel **85**(5–6), 679–688 (2006).
[3] K. Plessow, K. Acker, H. Heinrichs, and D. Moller, Time study of trace elements and major ions during two cloud events at the Mt. Brocken, Atmos. Environ. **35**(2), 367–378 (2001).
[4] J. Chiarenzelli, L. Aspler, C. Dunn, B. Cousens, D. Ozarko, and K. Powis, Multielement and rare earth element composition of lichens, mosses, and vascular plants from the Central Barrenlands, Nunavut, Canada, Appl. Geochem. **16**(2), 245–270 (2001).
[5] T. Ciesielski, M.V. Pastukhov, P. Fodor, Z. Bertenyi, J. Namiesnik, and P. Szefer, Relationships and bioac-

cumulation of chemical elements in the Baikal seal (Phoca sibirica), Environ. Pollut. **139**(2), 372–384 (2006).

 [6] P. Heitland and H.D. Koster, Biomonitoring of 30 trace elements in urine of children and adults by ICP-MS, Clin. Chim. Acta **365**(1–2), 310–318 (2006).

 [7] J.P. Goulle, L. Mahieu, J. Castermant, N. Neveu, L. Bonneau, G. Laine, D. Bouige, and C. Lacroix, Metal and metalloid multielementary ICP-MS validation in whole blood, plasma, urine and hair – Reference values, Forensic Sci. Int. **153**(1), 39–44 (2005).

 [8] A. Moreda-Pineiro, A. Fisher, and S.J. Hill, The classification of tea according to region using pattern recognition techniques and trace metal data, J. Food Composition and Analysis **16**(2), 195–211 (2003).

 [9] M.M.C. Gomez, I. Feldman, N. Jakubowski, and J.T. Andersson, Classification of German white wines with certified brand of origin by multielement quantitation and pattern recognition techniques, J. Agric. Food Chem. **52**(10), 2962–2974 (2004).

[10] J.-P. Perez-Trujillo, M. Barbaste, and B. Medina, Chemometric study of bottled wines with denomination of origin from the Canary Islands (Spain) based on ultra-trace elemental content determined by ICP-MS,

Anal. Lett. **36**(3), 679–697 (2003).

[11] C.M.R. Almeida and M.T.S.D. Vasconcelos, Multielement composition of wines and their precursors including provenance soil and their potentialities as fingerprints of wine origin, J. Agric. Food Chem. **51**(16), 4788–4798 (2003).

[12] J. Šperková and M. Suchánek, Multivariate classification of wines from different Bohemian regions (Czech Republic), Food Chem. **93**(4), 659–663 (2005).

[13] R. Jimenezespinosa, A.J. Sousa, and M. Chicaolmo, Identification of geochemical anomalies using principal component analysis and factorial kriging analysis, J. Geochem. Explor. **46**(3), 245–256 (1993).

[14] R. Rahil-Khazen, B.J. Bolann, and R.J. Ulvik, Correlations of trace element levels within and between different normal autopsy tissues analyzed by inductively coupled plasma atomic emission spectrometry (ICP-AES), Biometals **15**(1), 87–98 (2001).

[15] M. Dinya, E. Szekely, K. Szentmihalyi, G. Tasnadi, and A. Blazovics, Major and trace elements in whole blood of phlebotomized patients with porphyria cutanea tarda, Trace Elem. Med. Biol. **19**(2–3), 217–220 (2005).

DAUGIAELEMENTĖS ANALIZINĖS SPEKTROMETRIJOS TAIKYMAS KORELIACINIAM BANDINIŲ KLASIFIKAVIMUI

P. Serapinas, Ž. Ežerinskis

Jungtinė izotopinės analizės laboratorija, VU Teorinės fizikos ir astronomijos institutas, Vilnius, Lietuva

Santrauka

Pastebimas staigus koreliacijos koeficientų tarp elementų koncentracijų didėjimas, kai gruopuojamų bandinių kilmė yra panašesnė. 25 elementų koncentracijos buvo išmatuotos aukščiausia skiriamąja geba indukcinės plazmos masių spektrometrijos metodu. Apie dešimt ar net daugiau koreliuojančių elementų porų (kurių koreliacijos koeficientas $r > 0,9$) buvo aptikta skirtinguose vynuose iš panašių šaltinių. Koreliuojančios poros yra grupės ir mėginio ypatumas. Jos rodo panašumą tarp elemento koncentracijų ir jų šaltinių. Jeigu koreliuojančių porų skaičius tarp elementų yra ganėtinai didelis, tai pasinaudoję tuo sąrašu mes galime apskaičiuoti vidutinius koreliacijos koeficientus pavieniams bandiniams bet kurią klasifikuojamą grupę atitinkančiam koreliuojančių elementų porų sąrašui. Toks metodas gali būti naudojamas mėginiams atpažinti.