Analytical Methods

Step by step approach to multi-element data analysis in testing the provenance of wines

P. Serapinas [a], P.R. Venskutonis [b,*], V. Aninkevičius [c], Ž. Ežerinskis [a],
A. Galdikas [c], V. Juzikienė [c]

[a] *Institute of Theoretical Physics and Astronomy, A. Gostauto 12, Vilnius 01108, Lithuania*
[b] *Department of Food Technology, Kaunas University of Technology, Radvilenu pl. 19, Kaunas 50254, Lithuania*
[c] *Semiconductor Physics Institute, A. Gostauto 11, Vilnius 01108, Lithuania*

Abstract

The possibility to classify the provenance of a wide variety of randomly selected wines according to multi-element analysis data was tested. Large number of parameters is used for solution of such complex problem and the role of the noise increases. Stepwise approach is tested dividing the wine origin classification into some steps to simplify the problem. Outcomes of the approach are studied on the basis of the chemical analytical data obtained for 23 elements in 103 wines from seven countries. Anova was used to select the most informative elements at each step. Three or four elements often were found sufficient to discriminate between countries at 0.9 probability level. Principal component analysis was applied for concise data presentation. The possibility of application of the multivariate normal distribution to the principal components was tested and confirmed, and thoroughly used for the classification power estimates. Problems of indication of batches and adequate representation of those by samples are emphasized.
© 2007 Elsevier Ltd. All rights reserved.

Keywords: Wine fingerprinting; ICP–MS; Multi-element analysis; Principal component classification; Metrology

1. Introduction

The first successful attempts to use the multi-element analysis for classification of wines according to their provenance or country of origin can be dated as early as 1979 (Kwan, Kowalski, & Skogerboe, 1979). However, only comparatively abundant elements were measured at that time; Rb, Mn, K, Li, Mg, Ca, Na, Fe and Sr were the most cited for their capacity to discriminate wines according to their geographical origin. The advent of the inductively coupled plasma (ICP) instruments substantially increased the possibilities to measure trace and ultra-trace elements. Several reports demonstrated that trace element patterns can be used to fingerprint wines, and that these patterns, at least partially, reflect the provenance of wine or even

enable to secure its authenticity (Almeida & Vasconcelos, 2003; Capron, Smeyers-Verbeke, & Massart, 2007; Taylor, Longerich, & Greenough, 2003).

On the other hand, it became obvious that direct correlation between element concentrations in wine and soil does not exist indicating that soil–plant interactions are highly complex. Moreover, the content of detected elements is influenced by the local environmental factors and the peculiarities of handling and processing treatments applied both in viticulture and wine production. Nevertheless, promising capabilities of discrimination and classification of wines using the profile of elements in the samples and applying multi-dimensional chemometric techniques are supported by the studies in many countries, including Canada (Greenough, Mallory-Greenough, & Fryer, 2005; Taylor et al., 2003), Czech Republic (Kment et al., 2005; Šperkova & Suchánek, 2005), Germany (Gómez, Feldman, Jakubowski, & Andersson, 2004), Italy (Marengo & Aceto, 2003), New

---

Zealand (Angus, O'Keeffe, Stuart, & Miskelly, 2006), Portugal (Almeida & Vasconcelos, 2003), Romania (Done, Stoica, Cotea, Baiulescu, & Capota, 2006), Spain (Pérez-Trujillo, Barbaste, & Medina, 2003), South Africa (Coetzee et al., 2005). The role of wine making processes to the elemental pattern in the product, as well as other sources of information, including the isotope ratios of heavy elements, are also being studied (Gómez, Brandt, Jakubowski, & Andersson, 2004; Almeida & Vasconcelos, 2004; Capron et al., 2007; Suhaj & Koreňovská, 2005). Multi-element techniques, such as ICP–MS, are recognized as providing a powerful means for a rapid profiling of the trace element concentrations in a wide range of foods. The statistical methods usually include multivariate techniques as factorial analysis, discriminant, principal component analysis, artificial neural networks, classification and regression trees (Capron et al., 2007; Moreda-Pineiro, Fisher, & Hill, 2003; Pérez-Trujillo et al., 2003).

Summarizing available publications on the subject, it can be concluded that the elemental pattern of wines depends at least on soil, viticulture, environment and wine making technologies. Consequently, in case of the involvement of a higher number of geographical regions and wine making technologies the possibilities of finding unambiguous indicators for product authenticity become more and more complicated. The number and the complexity of the mathematical models needed also increase if more than two classes are to be regarded. From the other side, different mathematical approaches usually provide similar results if the essence of the problem is accounted for. The main goal of the present study was to measure the concentrations of elements in a wide variety of randomly selected wines and to aid in the elaboration of simple and transparent procedures of classification of products when the countries of their geographical origin can be both quite close and rather remote, and the number of the countries is comparatively large. Special attention is attributed to the selection of the best discriminating set of elements. Development of the probabilistic approach, based on the standard deviation estimates, to the evaluation of uncertainty of the classification was a subsequent task. Calculation of the re-substitution and prediction rates, often used in evaluations of the quality of classifications, needs very large number of samples to get the reliable estimates and possibly tends to overoptimistic classification rates.

2. Experimental

2.1. Wine samples

The bottled wine samples were purchased form the local market. In total, 103 different wines from seven countries were selected for the measurements in two experiment series. The countries of wine origin and the number of samples in the 1st series of experiments include Bulgaria (5), Chile (5), France (6), Hungary (7), Italy (6), Spain (9), USA (California) (5); in the second series of experiments the wines from Chile (20), France (20) and Spain (20) were selected.

The majority of the wines were red and dry, except for some white wines (one sample from California, France, Hungary, Italy and Spain). The wines were produced in 1996–2003, the concentration of alcohol varied from 11% to 19%, the amount of sugar was from 5 to 50 g/dm^3.

To ensure acceptable stability and reproducibility of sample introduction, wine samples were diluted by a factor 1:4 for the measurement of microelements and 1:30 for the measurement of elements present in larger concentrations mainly, as well as to test for the matrix effects. Merck multi-element reference material was used for calibration to obtain the absolute concentrations of elements. As the wine samples were rather different (e.g. due to the variations in alcohol and sugar contents), the dependence of the analytical signals on the matrix was evident. Internal standardization was also used as a mean for reduction of the matrix and signal time variation effects. Be, In, Rh, Sc and Tm were tested for this purpose. As similar effects on all the internal standards were observed, Tm was selected as the most convenient and was the only internal standard for systematic measurements.

2.2. Materials and reagents

The wines were subsampled to clean PE bottles. Ultra pure water as prepared by NANOpure Barnstead/Thermolyne Co. (Dubuque, USA) purification system was used for all dilutions. ICP Multi-Element Standard Solution VI CertiPUR (Merck, Darmstadt, Germany; $k = 2$ uncertainty no more than 5%) was used as a calibration standard solution. Up to 2% of nitric acid (Suprapur®, 65.3%, Merck) was added to the diluent for the stabilization of the subsamples. Simultaneously Tm was included as an internal standard at the concentration 10 ng g^{-1}.

All plastic labware (PE bottles, centrifuge tubes, pipette tips) used for the sampling and sample treatment were new and were cleaned by soaking them for 24 h in the each bath containing the following substances: first 10% HNO$_3$, second 5% HNO$_3$ and last ultra pure water. After rinsing with ultra pure water the labware were dried in a class 10 clean bench.

2.3. ICP–MS measurement techniques and procedure

The measurements were carried out by a double focusing sector field mass spectrometer Element 2 (Thermo Finnigan MAT, Bremen, Germany). The main characteristics of the mass spectrometer and sample introduction system were as follows:

Nebulizer, sample uptake rate	Concentric, 1 ml/min (peristaltic pump was used)
Spray chamber	Scott double pass
Argon gas flow rates, L min^{-1}	

Plasma	14
Auxiliary	0.75
Sample	1.35
RF power	1100 W
Sample and skimmer cones	Nickel
Acquisition mode	E-scan
Internal standard	Tm
Wash time	3 min
Take up time	1 min
Selected elements	
Low resolution (300)	B, Be, Li, Pb, Rb, Sr, Tl, U
Medium resolution (4000)	Al, As, Ca, Co, Cr, Cu, Fe, Ga, Mg, Mn, Na, Ni, Rb, U, V, Zn
High resolution (10,000)	As, Ba, K, Mn, Na, Rb, Sr

As mentioned above, to avoid clogging of the sample introduction system and to facilitate reduction of the matrix effects wine samples were diluted before measurements. To control the stability of the analytical signals the measurements of the standard solutions were repeated before and after the measurements of the samples. The measurements of the elements were repeated at least twice at different conditions: in samples of different dilutions, at different resolutions of the mass spectrometer and on different isotopes. Repeatability of the measurements of the analytical signals usually was not worse than 2–3%.

3. Results of the measurements

The absolute concentrations of 23 elements, namely Li, Be, B, Na, Mg, Al, K, Ca, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Ga, As, Rb, Sr, Ba, Tl, U in 103 wines from seven countries were measured. The results of measurements in different resolutions, for different isotopes, in the samples of different dilution coincide within precision of 5%, however in most cases 2%. Blank signals usually were lower than 3% of the signal value, often lower than 1%, but in some cases (Be, Ga, Tl) sometimes reached up to about 20%. Possible inputs from different uncertainty sources in preparation of the test samples and reference materials, sample introduction, measurement and data analysis (as recommended by Eurachem & CITAC, 2001) including dilution, purity of diluent, weighing, clean-up, signal instability in time, matrix effects, uncertainty of the reference material and blank were accounted for in the uncertainty estimates. Total uncertainty essentially depend on element and concentration range. Mainly it does not exceeded 7% ($k = 2$) while for some elements, e.g. Tl (mainly due to irreproducibility of the signals and blank) Cu, Zn (because of the matrix effects) the measurement uncertainty was up to 10%, or even 15% (Tl). In particular, because of the large concentrations of K in the samples adequate rinsing of the apparatus needed a long time, and the procedure used in the first experimental series was improved in the second one. The ranges of variation of the measured concentra-

Table 1
Intervals of variation of the concentrations of elements in the measured wines and 3σ limits of detection (ng g^{-1})

Element	min	max	max/min	Limit of detection
Al	215	5600	26	0.02
As	0.9	29	32	0.05
B	3000	15,000	5	0.01
Ba	42	440	10	0.001
Be	0.16	6.2	39	0.01
Ca	54,000	134,000	2.5	0.03
Co	1.3	9.8	7.50	0.002
Cr	8.9	77	8.7	0.01
Cu	10	520	52	0.01
Fe	730	8800	12	0.002
Ga	0.08	0.97	12	0.01
K	630,000	2,100,000	3.3	0.05
Li	2.5	290	1160	0.01
Mg	29,000	150,000	5.1	0.01
Mn	310	3600	12	0.01
Na	940	71,000	76	0.01
Ni	11	55	5	0.01
Rb	185	6600	36	0.001
Sr	330	2200	6.7	0.001
Tl	0.04	1.10	28	0.001
U	0.20	10	50	0.005
V	1.5	465	310	0.003
Zn	125	2200	18	0.01

tions in absolute and relative units and the corresponding limits of detection are presented in Table 1.

4. Strategy of data analysis and pattern recognition

4.1. PCA classification of samples

Discrimination between regions of origin is based on variations of the measured concentrations of elements. As the number of elements was large the pattern recognition techniques should be applied to classify the data. The standard principal component analysis (PCA) technique was used in the present work. It is a convenient tool for the reduction of data dimension to some principal components and visualization of the similarities. It provides a new set of variables obtained as the best linear combination of the original parameters which accounts for more of the variance than any other combination. At the same time this chemometric tool enables to find out in what respects the samples are different and which variables contribute most to the difference. The Unscrambler software was used. The logarithms of the concentrations were preferred (compare Capron et al., 2007; Taylor et al., 2003) for the corresponding mathematical treatment to reduce the relative role of large deviations.

The variations between batches larger than those within batches are expected for the successful classification. The real situation is different for every element and countries of interest. Application of as wide a set of elements as possible seems most straightforward. Data from the first experiment on measurements from five to nine wines from seven countries were tested for the preliminary screening.

Poor classification possibilities were found. One can expect that rejection of the nonspecific information and inherent noise can be helpful. Anova software as included into the Excel data analysis package was used for the purpose.

In Fig. 1, the classification pattern as obtained according to concentrations of Sr, Rb and Zn is presented. For those elements F was the largest for the batch including all countries, namely $F > 8\ F_{crit}$. Three groups of the samples can be clearly distinguished: wines from Bulgaria, Hungary and Spain, samples from France and Italy, and those from Chile and California. More detailed discrimination according concentrations of those elements was not possible, and further classification was undertaken as a next independent step.

New Anova analysis of the data for all elements and countries in each group separately was carried out once more and the elements with the largest F factor for every group (in general different from the elements detected in the first test) were determined. Discrimination of wines from Chile and California was based on the concentrations of Cr, Mg and Sr ($F/F_{crit} > 2.2$, Fig. 2). Only one principal component (PC1) is sufficient to separate between the two batches. Classification of samples from Bulgaria, Hungary and Spain could be undertaken as three two-country steps (BGR–HUN, BGR–ESP, HUN–ESP) but direct classification based on a larger number of elements was also possible. The example of such classification based on data for Al, Ca, Cr, Fe, K, Li, Mn, Na, Rb, Sr ($F/F_{crit} > 1.5$) is
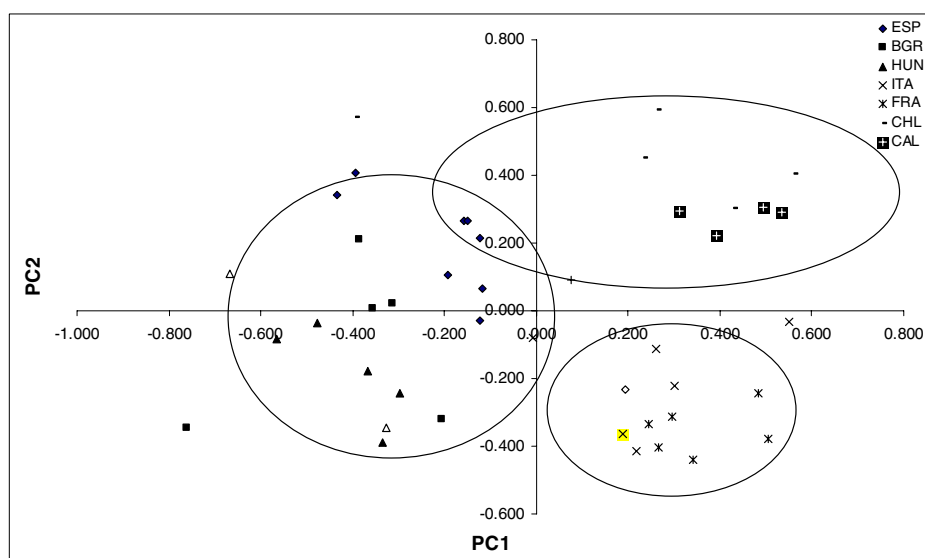

Fig. 1. Classification of wines from seven countries based on concentrations of Rb, Sr and Zn. Explained variance: 53% (PC1) and 34% (PC2). Absolute loadings of Rb and Zn to the principal components were from 0.5 to 0.8, those of Sr 0.04 and 0.26.
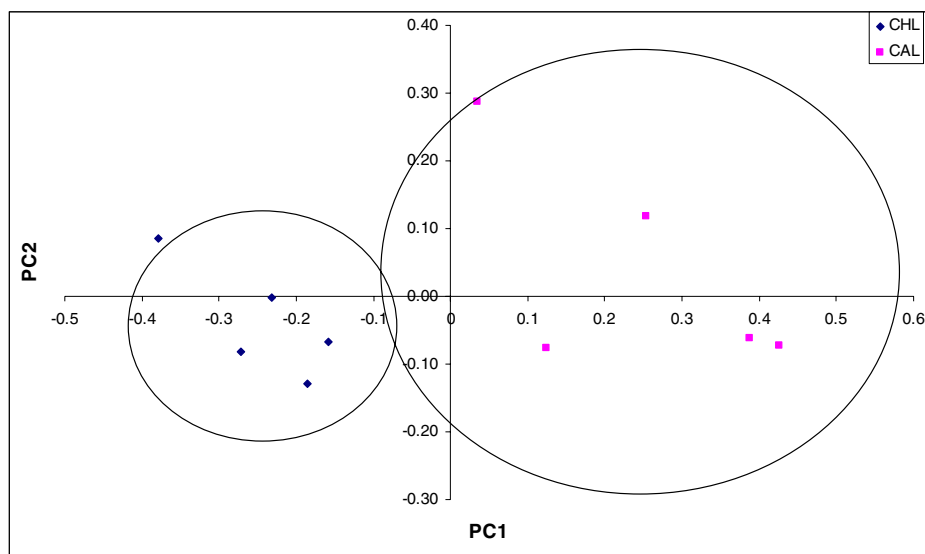

Fig. 2. Separation between Chile and California wines (Cr, Mg, Sr). Explained variance: 83% (PC1) and 16% (PC2). Absolute loadings of Cr and Sr to the principal components were from 0.5 to 0.8, those of Mg 0.12 and 0.29.

presented in Fig. 3. Bulgarian and Hungarian wines could be easy separated according to data on Cr, K, Rb and Sr. Data on the concentrations of Ca, K and Sr ($F/F_{crit} > 1.3$) were found the best for the separation of wines from Italy and France (Fig. 4).

In such a manner, the best discriminating elements can be used for every particular step. Separation between the samples from any two countries is the final result. The possibility to determine common F factors for the whole group is essential as the starting point. Nevertheless, the two-country F factors were found very useful not only for the separation of the samples from the corresponding two countries, but for the strategy of separation within small groups as well. For example, in classification of samples

from Bulgaria, Hungary and Spain F/F_{crit} for the elements accounted for varied from 42 (Rb) to 1.9 (Mn). In spite of small discrimination capability for the whole three country batch for Cr ($F/F_{crit} = 1.5$) this element was included as one of the best discriminating between BGR and HUN to improve separation between those two countries. For the same reason in this case V was omitted: in this case $F/F_{crit} = 7.3$ however it essentially reduces discrimination possibility between BGR and HUN (the two-country $F/F_{crit} = 0.1$). Similar approach can be useful even in complex cases, such as shown in Fig. 1, but the common F for the whole group provides good starting approximation.

Two principal components account for 71% of the variation of the data in Fig. 3 and about 87% of the variation
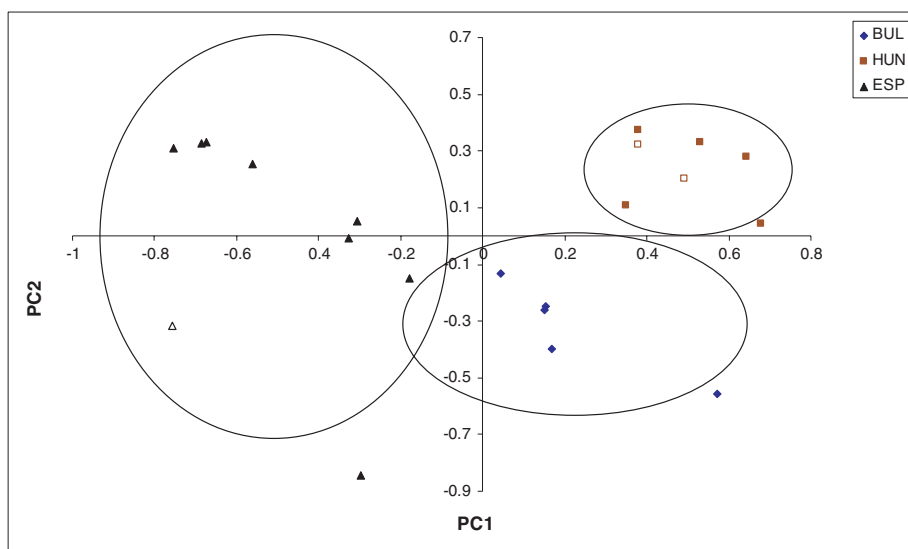

Fig. 3. Discrimination between wines from Spain, Bulgaria and Hungary according to the data on Al, Ca, Cr, Fe, K, Li, Mn, Na, Rb, Sr ($F/F_{crit} > 1.5$). Explained variance: 48% (PC1) and 23% (PC2). Absolute loadings of K to the principal components were 0.4 and 0.7, Al 0.5 and 0.4, other elements – less than 0.4.
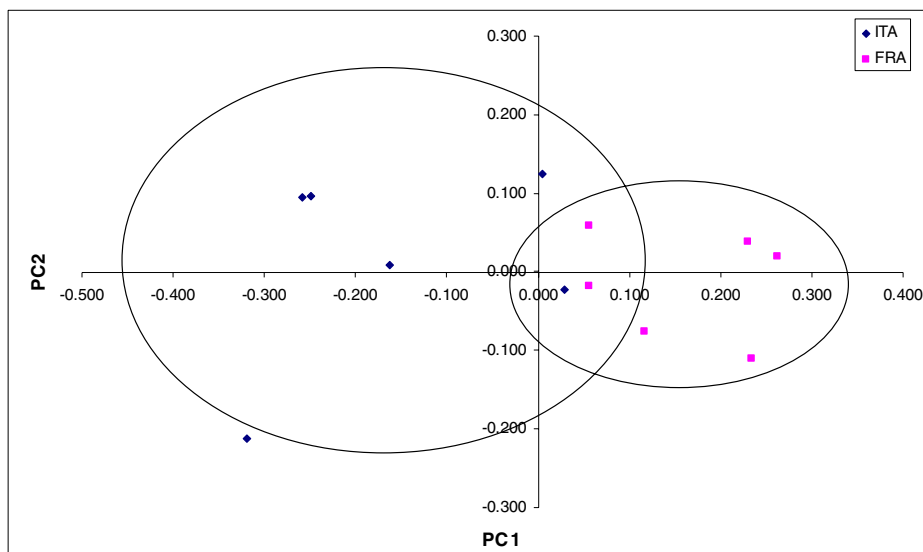

Fig. 4. Attempt of discrimination between wines from Italy and France (Ca, K, Sr; $F/F_{crit} > 1.3$). Explained variance: 79% (PC1) and 18% (PC2). Absolute loadings of K and Sr to the principal components were from 0.4 to 0.9, those of Ca about 0.2.

in Fig 1. In such cases it is possible that resolution of the batches can be improved if more PC will be included. In the cases considered above the additional components do not improve the resolution essentially.

The loadings of elements to the principal components presented in the figures are included in the figure captions. When only three elements are enough for the characterization of the batch it seems attractive to compare the PCA classification results with those of the direct application of the three-dimensional concentration graph. Nevertheless, the tests confirmed better classification capabilities of the PCA description even in such simple cases.

4.2. Representativity of samples and quality of the classification

As presented in figures above, the PC values for samples from the same batch, similar to the concentrations of elements, vary in some intervals. Statistical characteristics of the distributions and principles of the statistical analysis were used for evaluation of the quality of classification. The means and standard deviations of the PC values were calculated for all batches. Normal distribution was accepted as a starting approach. Area in which the data of a particular batch are expected to be found at the selected level of confidence will be limited by a constant-probability contour of the bivariate distribution (Weisstein, 2004)

$$p(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}$$
$$\times \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_x^2} - \frac{2\rho xy}{\sigma_x\sigma_y} + \frac{y^2}{\sigma_y^2}\right)\right). \quad (1)$$

Here x and y stands for the first and second principal components, σ is the standard deviation, $\rho \equiv cor(x,y)$ is the correlation coefficient between x and y, and must be zero because of the orthogonality of the principal components. Accuracy of determination of the standard deviations σ_x and σ_y is also restricted because of the limited number of measurements, and the corresponding probability distribution ought to be accounted for. We did not found in literature convenient for application formulas joining the two distributions and computer analysis of the real cases was undertaken. As a result equation of the ellipse of a constant-probability contour of the prescribed distribution in the form

$$\frac{x^2}{(t\sigma_x)^2} + \frac{y^2}{(t\sigma_y)^2} = 1 \quad (2)$$

was found as a practically acceptable, while mathematically crude, approximation. Here t is the Student's coefficient corresponding to the selected confidence level (0.9) and number of measurements. Of course, straightforward methods of hypotheses testing should be applied to get the mathematically exact results, but the graphical presentation was found convenient and useful. Eq. (2) was used to calculate the ellipses according to the corresponding standard deviations of

the values of the principal components and numbers of measurements in the figures. In such an approximation if the ellipses do not cross we accept that at the 0.9 probability level the corresponding batches can be regarded as separate. If the ellipses cross each other and the areas partially overlap significance of the common part of the probability distribution for both batches must be evaluated. It is possible by calculation of the integrals $\int p(x, y)\mathrm{d}x\,\mathrm{d}y$ over that area with the probability P (normalized to 1) distributions separate for both overlapping distributions.

As an example, inputs of the overlapping area of CHL–California and BGR–HUN–ESP (Figs. 2 and 3) distributions into corresponding total probabilities are about 3 and 2% correspondingly, and the superposition of the distributions is not important at the accepted 0.9 confidence level. Thus the two batches at that confidence level can be regarded as separate. The separation can be improved if the number of samples is increased. The superposition of the distributions in Fig. 5 (Italy–France) exceeds 10% of the total probability, and separation of those batches at that confidence level must be regarded as not possible.

The criterion for the best selection of the elements to be accounted for, including assessing a possible redundancy in variable selection, is the minimum of the standard deviations (axes of the ellipses, in general – uncertainties) of the PCA values in comparison to the difference between the mean values (centrums of the ellipses) for the batches under discrimination. The limited resolution possibilities were the reason why the 0.9 probability level was used in this publication. Transition to the 0.95 probabilities most frequently applied can easy be made by expanding the axes of the ellipses by about 20%.

As presented above, from 5 to 9 wines from one country were measured in the first series of experiments. One of the goals was to test if small data sets can be of any practical interest for the preliminary description of the batches. As follows from the material above the answer is positive. Increasing the number of samples for the homogeneous batch should result in better accuracy of the mean value of the parameters and their standard deviation and, as a consequence, smaller Student coefficient and the ellipse. In reality, there are a lot of reasons for variation of the samples, especially if those are to represent such a large region as a country and increasing of the number of randomly selected samples can add to the variety and variations of the describing parameters. As an example, if samples from some different regions of the same country will be included broadening of the distribution of the describing parameters can be expected.

In the second series of experiments, 20 different red wines from each of the following countries: Chile, France and Spain were measured. If all the data are included the resolution of the former pictures is slightly improved. In more details, the transformation of the distributions is analyzed in Fig. 5a, b where the data for all wines from Chile, France and Spain as measured in both experiments for Rb, Sr, Li and Zn, and the corresponding 0.9 probability
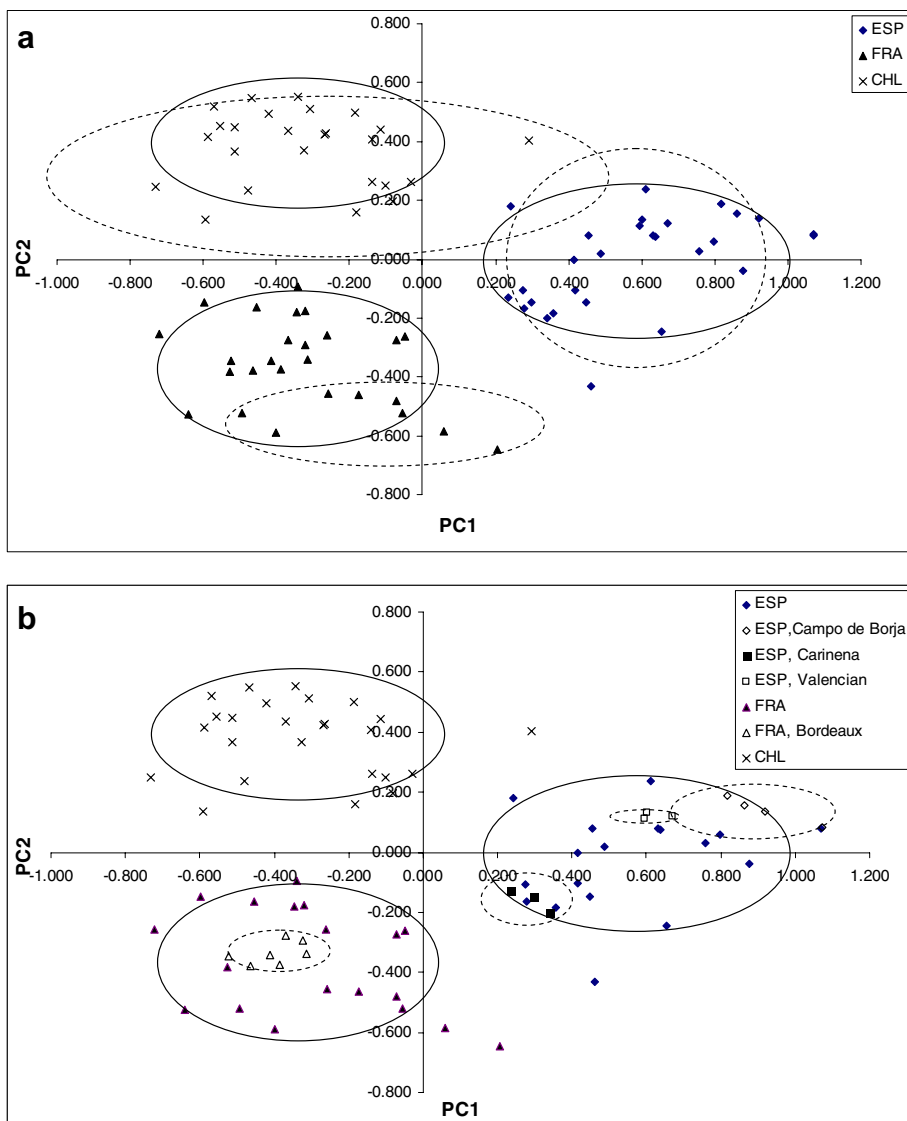
Fig. 5. Two principal component pattern for wines from Chile, Spain and France (solid line) measured in both experiments (Li, Rb, Sr and Zn, $F/F_{crit} > 11$) and some regions (dotted line). Explained variance: 59% (PC1) and 27% (PC2). Absolute loadings of the elements to the principal components were from 0.1 to 0.8.

ellipses as solid lines are included. Li (absent in Fig. 2) was included to strengthen the differences between the smaller regions within the countries (see below). In parallel the ellipses corresponding to the samples measured in the first experiment only (dashed lines) are presented. Both distributions for the wines from Spain are almost the same (0.9 probability level), distribution of the larger number of samples for Chile gets closer while that for France changes essentially. Consequently more detailed information on the samples should be regarded.

First of all the type of the wine; as mentioned above only red wines were measured in the second series of experiments. In the first series, the red wines also dominate, however, some white wines from Spain, Italy, California, France and Hungary were included. Naturally, those single points obtained on white wines do not change essentially the general view of the probabilistic pictures but often tend

to the borders of the corresponding 0.9 probability distributions (explicitly marked as transparent points in Figs. 1, 3). The differences depend on the elements being regarded and particular combinations of those elements can be important, which indicates that the two classes of wines such ought to be regarded as different products. As example of some other kind: the lowest point for Spanish wines in Fig. 3 corresponds to a wine bottled in Denmark. Besides, some marked wines were detected as quite different from their batches just according to the measured concentration values. More detailed analysis of the data showed that those were bottled in foreign countries, and therefore, they were excluded from the current analysis.

All Chile wines (measured in both experiments) were from Central Valley. In this case (Fig. 5), the variety of samples (according to Rb, Sr, Li, Zn) was sufficiently represented in the first experiment and increasing of the

number of samples narrows the distribution. Concerning the French wines addition of the new samples essentially increased the diversity of the geographical regions being considered, and the distributions obtained in the two experiments are different. Naturally, it would be of interest to separate the smaller geographical regions within the distribution representing all the country. Ellipses indicating samples from Bordeaux (France), Valencian, Aragon (Cariñena, Campo de Borja) (Spain) regions are presented in Fig. 5b as dotted lines. Certainly, to obtain the optimal classification into regions the next independent steps (new tests of the *F* values corresponding to the classifications being regarded) should be undertaken.

Thus, the problem of classification of the samples is in close relation with the problem of the adequate representation of the diversity of the batches by samples. Possibly the two tasks are strongly interconnected and just the study of both could give a more global perspective.

As mentioned above, the normal distribution approach was used to evaluate the quality of discrimination between the batches. As the dependences between the input data and the values of the principal components are very complicated the character of the distribution of the principal components is not self evident. Very large data sets (hundreds of samples) would be needed to test the character of the distribution directly. Our test included the following approaches. First of all, as the 0.9 probability confidence level was selected the ellipses in the figures should include about 90% of the corresponding samples. Really part of the samples found inside the ellipses in Figs. 1–5 varies from 0.8 to 1 while the mean value was $86\% \pm 4\%$ ($k = 2$) in good agreement with the predetermined probability. In addition, the third and fourth moments of the distributions of the first three principal components were calculated for the largest batches, namely those regarded in Figs. 1 and 5. The mean ratio of the third moment of the distributions and the third power of the standard deviation was -0.14 ± 0.32 ($k = 2$) while the ratio of the fourth moment and σ^4 was $2.8 \pm 0,7$ ($k = 2$) in good agreement with the values characteristic for the normal distribution (0 and 3 as the corresponding values are expected).

In general, as mentioned above, increase of the number of samples and number of the principal components could be used to improve the quality of classification. Improvement of the accuracy of measurements, especially for the elements at small concentrations, is an additional possibility. Nevertheless, cases are expected where the classification will be not good enough or even impossible.

According to Fig. 4, Italy and France should be regarded as one batch. Identification of the smaller classes, possibly based on quite different approaches, may help further steps. Really, the states may be not specific enough classification criterion concerning wines, especially if those states are neighbors, while some specific differences depending on traditions, standardization, legal approaches etc. are possible and should not be missed in organization of the measurements. As one of the approaches, smaller geographical

regions, vineyards, etc. can be introduced as the classification batches for the next step. We had too few material for the detailed classification of the French and Italian wines by regions but, as an example, Bordeaux wines could be discriminated from Italian ones according to the data on concentrations of Ca, Li and Sr. In particular, essential change of criteria of classification can be indispensable if the new principles of identification of the classes are introduced.

Naturally, discovering of the most specific criteria is very important at any classification step. In every step discussed above the classes are getting smaller. Decreasing the differences between the regions will result in smaller variations of the concentrations of elements. At the same time, one can expect that the remaining reasons of variation will be more common for larger number of elements, i.e. the correlations between concentrations of elements could increase due to some common agrochemical or technological processes. As an example, positive correlation between concentrations of Ca, Mg, Li, Ni was observed for Spanish wines in general while for Campo de Borja this list includes Cu, Fe, Zn in addition but concentrations of all those elements tend to anticorrelate with concentrations of Al, B, Ba, Rb. Correlations between concentrations of Al, Ga, U, V were found characteristic to French wines while for Bordeaux wines positive correlations between concentrations of Ba, Ni, Rb, Sr, and some other correlating elements, were found in addition. Method of application of correlations to description of the individual samples and application for classification purposes was described by Serapinas and Ežerinskis (2006). The method was not used for classification here as only some regions were represented enough by samples in this study.

5. Conclusions

The results obtained demonstrate high potential of the multi-element analytical data description by the principal components for the wine pattern recognition, even when large variety of regions is included. Step by step approach from large to smaller batches was found a promising method for a better exploitation of the discriminating power of the most discriminating elements at the particular step. Selection of elements is essential to reduce the noise level and can be carried out by the usual statistical procedures. Anova provides sufficient information on the importance of an element. The data at the limits of the ranges of the concentrations can readily provide preliminary information on what regions can or cannot be regarded as probable regions of origin of the sample. The lowest measured concentrations of element, as compared to the whole concentration range for the same element, seem to be particularly promising: it is hardly probable to reduce the concentration of a single or few elements in wine selectively while accidental or fraud contamination is easy. Even for such abundant elements as Mg, Al, K, Rb and Sr, or minor elements Be, Cr, Ga and Tl individual potential biasing countries or groups of such countries could be found.

Knowledge and reduction of the uncertainties of the measurement can be decisive in this case. Some measured wines originally attributed to one of the countries studied, really were bottled in foreign countries, and were identified as biasing according to the measured concentration values or from the PC analysis.

Classification of food products, e.g. diary products (Karoui & De Baerdemaeker, 2007), olive oils (D'Imperio, Dugo, Alfa, Mannina, & Segre, 2007), honey (Bertelli, Plessi, Sabatini, Lolli, & Grillenzoni, 2007) on the basis of elemental, molecular analysis and other indicators is a fast developing approach. Highly advanced study, based on very extensive material for wines was published recently by Capron et al. (2007). We hope that the stepwise approach from large to smaller variety of batches, as applied in the present paper, based on simple software technique can be useful especially when the number of classes essentially increases. Besides, usually very good classification rates, as provided in the publications above, are hardly compatible with the statistical nature of the distributions of the data presented and probabilistic evaluation of the uncertainty of the data. Further studies of the predicting power of the classification results seem to be desirable.

It is highly probable that the level at that the step by step classification stops as well as the level of confidence of classification now is mainly determined not by analytical or mathematical discrimination capabilities but by reasonable identification of the batches and sufficiently exhaustive representation of the corresponding variations of the concentrations of elements by samples. Simple increase of the number of samples is hardly a promising solution of the problem. It needs better understanding of the genesis of the differences and ways in which they manifest themselves, and corresponding derivation of the discrimination criteria. Both studies on identification of batches and discrimination between them are highly interconnected.

Acknowledgement

References

Almeida, C. M. R., & Vasconcelos, M. T. S. D. (2003). Multi-element composition of wines and their precursors including provenance soil and their potentialities as fingerprints of wines origin. *Journal of Agricultural and Food Chemistry, 51*, 4788–4798.

Almeida, C. M. R., & Vasconcelos, M. T. S. D. (2004). Does the winemaking process influence the wine $^{87}Sr/^{86}Sr$? A case study. *Food Chemistry, 85*, 7–12.

Angus, N. S., O'Keeffe, T. J., Stuart, K. R., & Miskelly, G. M. (2006). Regional classification of New Zealand red wines using inductively-coupled plasma-mass spectrometry (ICP–MS). *Australian Journal of Grape and Wine Research, 12*, 170–176.

Bertelli, D., Plessi, M., Sabatini, A. G., Lolli, M., & Grillenzoni, F. (2007). Classification of Italian honeys by mid-infrared diffuse reflectance spectroscopy (DRIFTS). *Food Chemistry, 101*, 1565–1570.

Capron, X., Smeyers-Verbeke, J., & Massart, D. L. (2007). Multivariate determination of the geographical origin of wines from four different countries. *Food Chemistry, 101*, 1585–1597.

Coetzee, P. P., Steffens, F. E., Eiselen, R. J., Augustyn, O. P., Balcaen, L., & Vanhaecke, F. (2005). Multi-element analysis of South African wines by ICP–MS and their classification according to geographical origin. *Journal of Agricultural and Food Chemistry, 53*, 5060–5066.

D'Imperio, M., Dugo, G., Alfa, M., Mannina, L., & Segre, A. L. (2007). Statistical analysis on Sicilian olive oils. *Food Chemistry, 102*, 956–965.

Done, J. C., Stoica, A. I., Cotea, V. V., Baiulescu, G. E., & Capota, P. (2006). Determination of some microelements in Dealurile Moldovei viticultural area wines. *Revista de Chimie, 57*, 239–244.

Eurachem, & CITAC (2001). *Quantification of uncertainty in chemical measurement* (2nd ed.).

Greenough, J. D., Mallory-Greenough, L. M., & Fryer, B. J. (2005). Geology and wine 9: Regional trace element fingerprinting of Canadian wines. *Geoscience Canada, 32*, 129–137.

Gómez, M. M. C., Brandt, R., Jakubowski, N., & Andersson, J. T. (2004). Changes of metal composition in German white wines through the winemaking process. A study of 63 elements by inductively coupled plasma–mass spectrometry. *Journal of Agricultural and Food Chemistry, 52*, 2953–2961.

Gómez, M. M. C., Feldman, I., Jakubowski, N., & Andersson, J. T. (2004). Classification of German white wines with certified brand of origin by multi-element quantitation and pattern recognition techniques. *Journal of Agricultural and Food Chemistry, 52*, 2962–2974.

Karoui, R., & De Baerdemaeker, J. (2007). A review of the analytical methods coupled with chemometric tools for the determination of the quality and identity of diary products. *Food Chemistry, 102*, 621–640.

Kment, P., Mihaljevič, M., Ettler, V., Šebek, O., Strnad, L., & Rohlová, L. (2005). Differentiation of Czech wines using multi-element composition – A comparison with vineyard soil. *Food Chemistry, 91*, 57.

Kwan, W.-O., Kowalski, B. R., & Skogerboe, R. K. (1979). Pattern recognition analysis of elemental data. Wines of *Vitis vinifera* cv. Pinot Noir from France and the United States. *Journal of Agricultural and Food Chemistry, 27*, 1321–1326.

Marengo, E., & Aceto, M. (2003). Statistical investigation of the differences in the distribution of metals in Nebbiolo-based wines. *Food Chemistry, 81*, 621–630.

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

Pérez-Trujillo, J.-P., Barbaste, M., & Medina, B. (2003). Chemometric study of bottled wines with denomination of origin from the Canary Islands (Spain) based on ultra-trace elemental content determined by ICP–MS. *Analytical Letters, 36*, 679–697.

Serapinas, P., & Eżerinskis, Ž. (2006). Multi-element analytical spectrometry as a data source for correlative description of samples. *Lithuanian Journal of Physics, 46*, 505–512.

Šperkova, J., & Suchánek, M. (2005). Multivariate classification of wines from different Bochemian regions (Czech Republic). *Food Chemistry, 93*, 659–663.

Suhaj, M., & Koreňovská, M. (2005). Application of elemental analysis for identification of wine origin. A review. *Akademiai Kiado, 34*, 393–401.

Taylor, V. F., Longerich, H. P., & Greenough, J. D. (2003). Multi-element analysis of Canadian wines by inductively coupled plasma mass spectrometry (ICP–MS) and multivariate statistics. *Journal of Agricultural and Food Chemistry, 51*, 856–860.

Weisstein, E. W. (2004). Bivariate normal distribution. *MathWorld – A Wolfram web resource*. http://mathworld.wolfram.com/Bivariate-NormalDistribution.html.