Fresenius J Anal Chem (1996) 354: 327–332

ORIGINAL PAPER

M. Suchánek · H. Filipová · K. Volka · I. Delgadillo
A.N. Davies

Qualitative analysis of green coffee by infrared spectrometry

Received: 6 January 1995/Revised: 9 May 1995/Accepted: 14 May 1995

Abstract Principal component analysis and linear discriminant analysis (LDA) were applied to the mid-infrared spectra for the qualitative analysis of the variety of green coffee (Arabica and Robusta). It is shown that the KBr pellet technique in combination with the LDA method can successfully be used for the identification of sample origin.

Introduction

Coffee is commercially available as one of two varieties – Arabica or Robusta. Arabica coffee generally achieves higher prices on international markets. Therefore, it is of commercial importance to identify the bean variety or to detect a mixture of varieties. Qualitative analysis of green coffee samples should provide information on the origin and kind of the coffee.

The chemical composition of the two varieties is similar, however, there are parameters that may discriminate them, such as contents of caffeine, trigonelline, chlorogenic acid and lipids [1]. On a dry-weight basis, almost half of the green coffee bean consists of polysaccharides [2].

Polysaccharide compositions of Robusta and Arabica beans are similar; the only significant difference is the higher content (approximately 3%) of arabinogalactan in the former type, but a complete structure characterization of the polysacharides in green coffee beans has not been published yet.

M. Suchánek (✉) · H. Filipová · K. Volka
Department of Analytical Chemistry, Institute of Chemical Technology, CZ-166 28 Praha 6, Czech Republic

I. Delgadillo
Department of Chemistry, Universidade de Aveiro, P-3800 Aveiro, Portugal

A.N. Davies
ISAS, Institut für Spektrochemie und Angewandte Spektroskopie, Postfach 10 13 52, D-44013 Dortmund, Germany

In cases of mistrust about the variety or origin it is necessary that a qualitative method satisfies the following requirements:

- provide the required information rapidly;
- consider all the discriminant parameters in a measurement without the necessity to examine several parameters by wet chemistry;
- be rugged, which means that it must be sensitive to small changes in the chemical composition of different types of coffee, but insensitive to inter-batch variability within one particular type of coffee;
- be simple, exhibiting a good repeatability and reproducibility.

Infrared spectroscopy is a technique that offers all these requirements, because it is rapid, simple, and gives an indication of the overall contents of fat, protein and polysaccharides as well as giving rise to a spectral profile that reflects major and minor components.

A proper processing of the infrared results and the establishment of principal metrological characteristics constitute an important part of the analytical procedure; therefore the choice of suitable chemometric procedures is of particular importance, if information on the origin of the sample and its qualitative composition is sought.

The present work deals with the application of two pattern recognition methods, viz. principal component analysis (PCA) (unsupervised learning), and linear discriminant analysis (supervised learning). The factorial design approach with subsequent variance analysis was used to test the ruggedness of the parts of the analytical procedure proposed.

Experimental

Samples and measurements

The qualitative analysis method was tested on 4 coffee samples, viz. 2 samples of Arabica coffee from Brazil and Costa Rica and 2

samples of Robusta coffee from Uganda and Costa de Marfim. The samples were ground in an ETA 6700 electric coffee mill using two grinding degree settings. A Narva DDR-GM 9458 ball mil (vibrator) (30 W power) was employed for subsequent mixing with KBr and additional grinding. The ground samples were distributed into size fractions by means of metal sieves of 0.315, 0.250 and 0.160 mm mesh size. The weight of the samples was determined on a Mettler AE240 balance with an accuracy of 0.01 mg and a standard deviation of 0.02 mg.

The absorption spectra, using the KBr disk technique, were scanned on a Nicolet 205 FTIR spectrometer over the region of 4800–350 cm^{-1}; resolving power 2 cm^{-1}, 64 scans with autoamplification. KBr disks were compressed on a Perkin–Elmer hydraulic press, applying a pressure of 10–11 t to the entire disk area (13mm in diameter).

Chemicals

Potassium bromide of reagent grade purity (Lachema, Brno, Czech Republic).

Software and calculations

PC/IR: Nicolet Instruments, version 3.10, 1991.
Principal component analysis: UNSCRAMBLER II code by CAMO A/S, Trondheim, Norway, 1993. Variance analysis of factorial experiments: ANOVA code developed by the Institute of Analytical Chemistry, 1988. The program includes an algorithm from the IBM 360 library. Linear discriminant analysis: a program in PASCAL was set up making use of the conventional algorithm for linear discriminant analysis with canonical variables [3].

Linear discriminant analysis with canonical variables is based on a search for the linear combination of variables

$$w = v'x$$

which separates best the groups considered (whose number is k); x is the vector of experimental variables. Consider the following decomposition of the total variability of the new variable w:

$$n \cdot s_w^2 = Q_B + Q_E$$

where n is the total number of measurements, s_w^2 is the variance estimate of the new variable, and Q_B and Q_E represent the intergroup and intragroup variabilities, respectively. Now, a good separation of groups is achieved, if Q_B is as high as possible while Q_E is as low as possible. In other words, the maximum is sought for the ratio λ

$$\lambda = \frac{v'Bv}{v'Ev}$$

where B and E are the intergroup and intragroup variability matrices of the initial variables x, respectively.

The solution of the problem consists in finding the characteristic vectors v, the linear combination

$$w_1 = v_1'x$$

being referred to as the first discriminant of first canonical variable. The separability of the k groups of the set can be visualized by graphical representation of all objects on the disciminant surface of the two first canonical variables.

Analytical procedures

Two kinds of experiment were designed to examine the homogeneity of the coffee samples and the effect of the grain size on the quality of the decision-making process.

Procedure 1. The starting sample coffee was distributed uniformly over a paper sheet with a grid of squares labelled 1–100. Ten samples were taken at random by using a table of random numbers; each sample containing 5 to 6 coffee grains. The samples were ground in the coffee mill and additionally in the ball mill for 3 min. An aliquot of approximately 0.01 g was precisely weighed-in and solid KBr was added so that the total weight was 1g. The weighing was performed in triplicate for each sample. The mixture was homogenized in the ball mill for 4 min, and 3 disks about 0.25 g weight, containing 1% (m/m) coffee, were prepared from it. The disks were precisely weighed (disk weight m_{disk}), and the factors $f = 0.25/m_{disk}$ were calculated. By this factor the spectrum recorded was multiplied in order to eliminate the effect of the different thickness and weight of the disks.

Procedure 1 was only applied to Robusta, Uganda coffee. The layout is shown in Fig. 1.

Procedure 2. The starting coffee sample was uniformly spread over a sheet of paper with a grid as in Procedure 1. Coffee grains were taken from 10 squares selected at random (table of random numbers) and mixed together. The combined sample was ground in the coffee mill and additionally in the ball mill for 3 min. The powder was allowed to stand in a vacuum desiccator over P_2O_5 for 48 h and dried to constant weight. The dry sample was fractionated by using sieves of 0.315, 0.250 and 0.160 mm mesh size (Table 1). Aliquots of approximately 0.01 g were taken from each fraction weighed with an accuracy of 0.1 mg, and solid KBr prior ground in the ball mill for 3 min was added so that the total weight of the mixture was 1 g. The mixture was homogenized by mixing for 4 min, and 3 disks containing 1% (m/m) coffee were prepared from it. The experimental spectra were corrected by using factor *f* as in Procedure 1.

Procedure 2 was applied to all coffee samples. The layout is shown in Fig. 2.

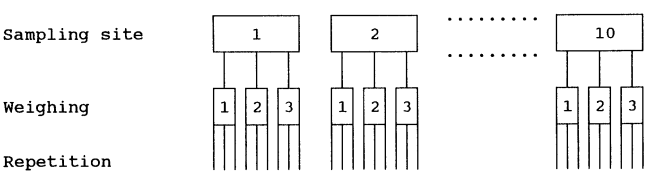


Fig. 1 Layout of procedure 1

Table 1 Sample identification and particle size of the fractions

Identification		Particle size, mm
Fraction	1	< 0.160
	2	0.160–0.250
	3	0.250–0.315
	4	> 0.315
W		undried unsieved sample
D		dried sieved sample

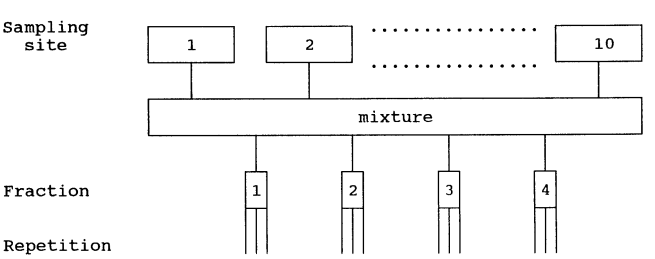


Fig. 2 Layout of Procedure 2

Results and discussion

Infrared spectra of coffee samples prepared by the above procedures were subjected to variance analysis. Absorbances at 2925, 2855, 1930, 1746, 1648, 1384, 1268, 1161, 1057 and 814 cm^{-1} (Fig. 3) were the dependent variables.

Analysis of variance (ANOVA)

Procedure 1. Ten experimental matrices were set up for the 10 wavenumbers selected. The ANOVA computer program was used for a two-factorial variance analysis with factors A and B, where factor A involved the effect of the sampling site and subsequent sample handling (10 levels for 10 sampling sites) and factor B involved the effect of taking the sample aliquot for weighing and of the weighing itself (3 levels for 3 weighed batches). The number of repetitions was 3 (for the preparation of 3 disks from each weighed batch). ANOVA gave the values of s_r^2, F_A and F_B, where s_r^2 is the residual variance of measurement and F_A and F_B are the Fisher distribution values calculated from the experimental data.

The residual variance was in the order of 10^{-4}, which documents that the disks within one weighed batch did not differ from each other significantly, i.e. the disk preparation was sufficiently well repeatable. The F_A value was invariably higher than the critical value of $F_{crit} = 2.04$ for a 5% significance level and for the degrees of freedom 9 and 60. These data demonstrate the significance of the effect of the sampling site, which should be eliminated to ensure accuracy and reproducibility of the experiment. The low F_A value for the wavenumber of 1930 cm^{-1} bears out the experimentally established fact that no band is present within that spectral range. The F_B values for the wavenumbers of 1384, 1161, 1057 and 814 cm^{-1} are lower than the critical value of $F_{crit} = 3.15$ at the 5% significance level for the degrees of freedom of 2 and 60. The fact that this was not true for all the wavenumbers examined suggests that the effect of grain size during sampling should be investigated.

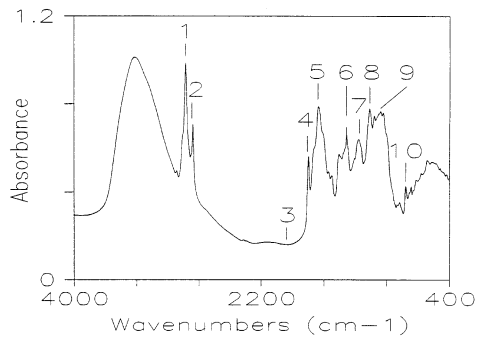


Fig. 3 Infrared spectrum of coffee sample (Robusta, Uganda) [cm^{-1}]: *1* = 2925; *2* = 2855; *3* = 1930; *4* = 1746; *5* = 1648; *6* = 1384; *7* = 1268; *8* = 1161; *9* = 1057; *10* = 814

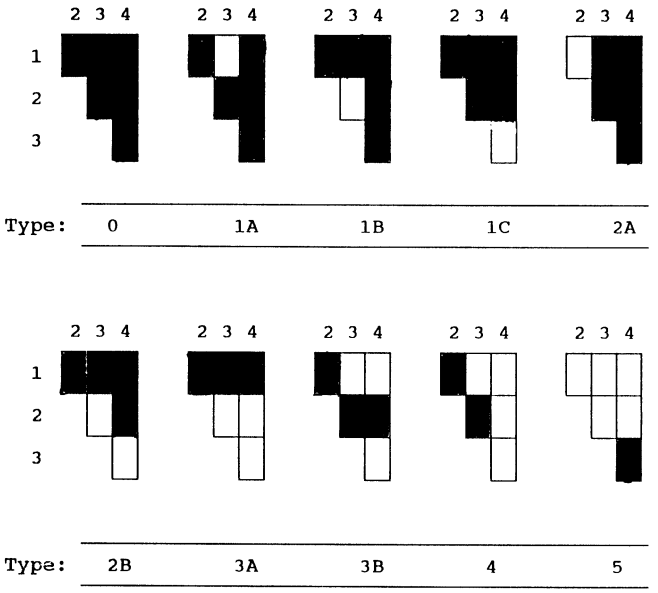
Procedure 2. Ten experimental matrices were set up for each kind of coffee. The ANOVA program was applied to single-factorial variance analysis for factor B, viz. fraction. The B factor levels are given in Table 2. The

Table 2 Description of ANOVA for the Procedure 2 and objects identification in Figs. 4–7[a]

Kind of coffee	Fraction, level B	Level B number	F_{krit} (5%)	Number in Figs. 4, 5	Number in Figs. 6, 7
Uganda	1			1	
	2			2	
	3	6	3.10	3	1
	4			4	
	D			5	
	W			6	
Brasil	1			7	
	2			8	
	3	4	4.07	9	2
	4			10	
Costa Rica	1			11	
	2			12	
	3	5	3.48	13	3
	4			14	
	D			15	
Costa de Marfin	1			16	
	2			17	
	3			18	
	4	5	3.48	19	4
	D			20	

[a] Fraction W is undried unsieved sample; fraction D is dried unsieved sample; the other fraction cf. Table 1

Table 3 Types of contrast matrix for various kinds of green coffee[a]



$\tilde{\nu}$, cm^{-1}	Robusta Uganda	Costa de Marfin	Arabica Brasil	Arabica Costa Rica
2925	2B	1C	0	2A
2855	3A	1C	0	2A
1930	4	1A	5	3B
1746	2B	1C	0	2A
1648	1B	0	0	1C
1384	2B	1C	0	1C
1268	1B	1C	0	1C
1161	1B	1C	0	1C
1057	1B	0	0	1C
814	2B	1C	0	1C

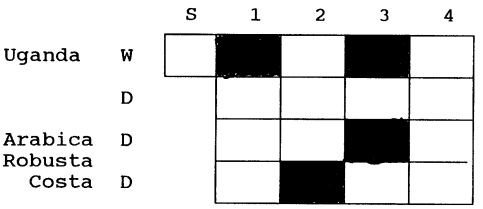
[a] Numbers 1–4 correspond to different sieved fractions, dark areas correspond to significant contrast, light areas correspond to insignificant contrast (signficance level 5%)

number of repetitions was 3, corresponding to the preparation of 3 disks from each fraction.

The computer processing gave values of s_r^2 and F_B. The residual variance was in the order of 10^{-4} for all kinds of coffee, indicating that the disks within a given fraction did not differ significantly; hence the disk preparation was sufficiently well repeatable. The F_B value, describing the effect of properties of the fractions on the absorbance was much greater than the critical value F_{crit} for all kinds of coffee and at all wavelengths except for the wavelength of 1930 cm^{-1}. This implies that the fractions differ from each other significantly, which may be due to the grain size effect on the refraction and reflection phenomena during the interaction of infrared rays with particles. The sieving process may also result in a separation into particles having different properties. The low F_B value at 1930 cm^{-1} can be explained in terms of absence of any band at that wavenumber.

Next, the similarity of the fractions was examined based on contrasts. The aim of this examination was to find which fractions were most or least similar to each other, or which fraction contributed most of the high F_B value.

Table 4 Contrast matrix for wavenumber 1930 cm^{-1}[a]



[a] Letter W corresponds to undried unsieved green coffee; letter D corresponds to dried unsieved green coffee; dark areas correspond to significant contrast, light areas correspond to insignificant contrast (significance level 5%)

When comparing two dependent variable values (e.g. absorbances) A_1 and A_2, the contrast is their difference,

$$\phi = A_1 - A_2$$

Contrast ϕ is significant if the confidence interval

$$L_{1,2} = \phi \pm t \cdot s_r \cdot (2/r)^{1/2}$$

does not include zero. In this equation, t is the Student criterion for the given significance level and degree of freedom, s_r is the residual standard deviation, and r is the number of repetitions.

For each coffee kind and wavenumber, the average absorbances of the fractions were calculated from three repetitions, and the contrasts and confidence intervals were established from them. The results are given in Tables 3 and 4. The tables demonstrate that the results of contrast analysis are similar for all wavenumbers except 1930 cm^{-1}, where all fractions are similar and contrasts insignificant. For Robusta, Uganda coffee, fractions 2 and 3 and fractions 3 and 4 (coarsest) are most similar, whereas fraction 1 is quite different and exhibits an appreciably higher absorbance. For Arabica, Brazil coffee, all contrasts are significant, hence, the fractions are different. For the samples of Arabica, Costa Rica, the coarsest fractions 3 and 4 are similar at all wavenumbers. This also applies to Robusta, Costa de Marfim, where only the contrasts between fractions 3 and 4 are insignificant.

The results of this analysis indicate that for a good evaluation of the kind of coffee, the samples should be ground and size-distributed because the spectral patterns depend significantly on the particle geometry.

Principal component analysis (PCA)

Principal component analysis was applied to normalized experimental data obtained by Procedure 2. The average absorbances of the fractions were arranged into an experimental data matrix, where the

selected wavenumbers were the parameters and the kinds of coffee and fractions corresponded to the objects.

The results are shown in Figs. 4 and 5. The UNSCRAMBLER II code was used to obtain two new variables, of which the first principal component contains the majority (95%) of information required. Figure 4 shows graphically how the initial parameters affect the first two principal components. The assignment of numbers to the objects is given in Table 2. Figure 4 presents an assessment of the objects with respect to the new variables. Points 5, 6, 15, and 20, which correspond to unsieved samples, are outliers. Objects 1, 7, 11, and 16, which correspond to the finest fraction 1, are also far from the cluster of points in the left-hand part of the graph. With regard to this, objects corresponding to the unsieved coffee samples were eliminated from the experimental matrix. Figure 5 then emerged as the result of the principal component analysis; the assignment of numbers to the objects is given in Table 3. Points 1, 5, 9, and 13, again corresponding to the finest fraction 1, are actually outliers. Additional distant objects, No. 6 and 10, correspond to the second finest fraction 2.

The results of the principal component analysis indicate that this calculation procedure cannot be employed to discriminate the samples of the various kinds of coffee even after sieving fractionation.

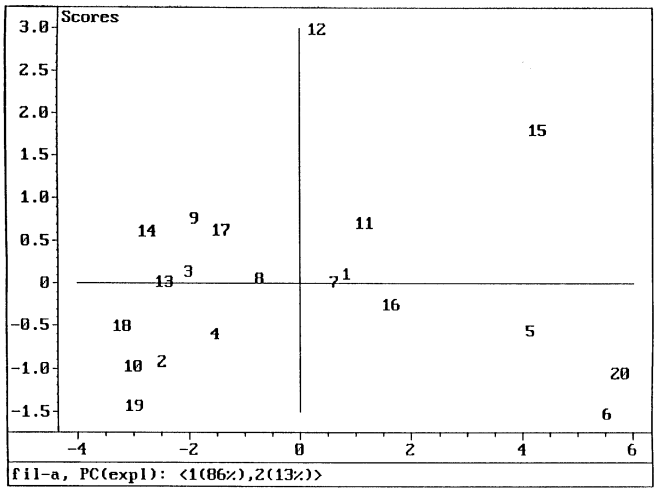


Fig. 4 Principal component analysis of all coffee samples (sample numbering cf. Table 2)

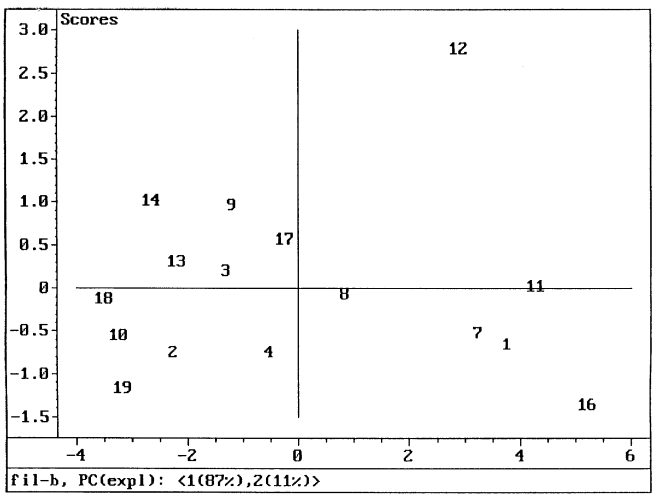


Fig. 5 Principal component analysis of sixteen samples (unsieved samples eliminated; sample numbering cf. Table 2)

Linear discriminant analysis (LDA)

The results of LDA are shown in Figs. 6 and 7. The numbers 1 through 4 correspond to the kinds of coffee according to Table 2; the different fractions of the same kind of coffee are not discriminated by numbers. Figure 7 shows the situation where only 4 fractions obtained by sieving the samples were taken into account, whereas the properties of all fractions including the unsieved fraction are shown in Fig. 6. The point clusters in this figure are spread to a higher extent and occasionally overlap, whereas in Fig. 7 the groups of objects are

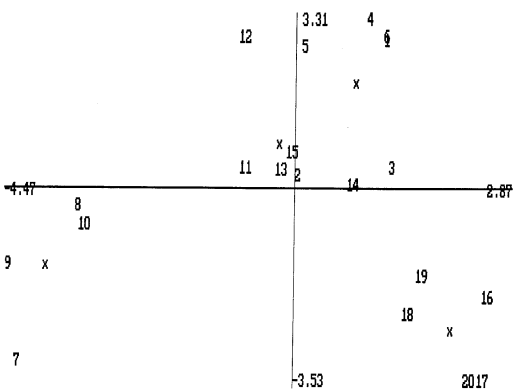


Fig. 6 Linear discriminant analysis of all (20) coffee samples (sample numbering cf. Table 2)

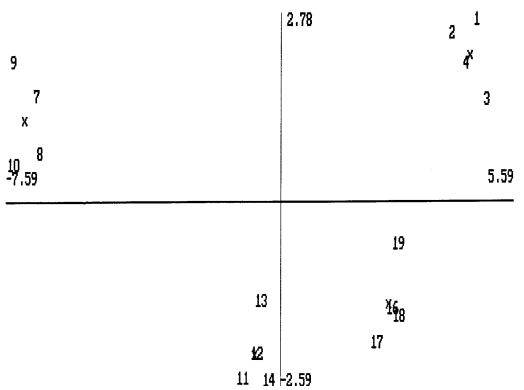


Fig. 7 Linear discriminant analysis of sieved coffee samples (sample numbering cf. Table 2)

clearly separated by the kinds of coffee. This suggests that the method can be employed to discriminate between different kinds of coffee, or to attribute an unknown sample to a given group of objects, if the given experimental procedure including sieving is strictly adhered to.

Linear discriminant analysis thus proves to be well suited for a qualitative discrimination between different kinds of coffee based on the selection of suitable peaks within the infrared spectral region of 4800–350 cm^{-1}. It is important that the samples must be fractionated by sieving because different fractions form a set of patterns of one class.

Recommended experimental procedure for the analysis of coffee

The experiments performed gave evidence that green coffee can be conveniently analyzed by infrared spectroscopy. The following experimental procedure is recommended:

Spread the sample of green coffee uniformly over a paper sheet containing a grid of squares labelled 1 to 100. Take coffee grains from 10 sampling sites selected at random (by using tables of random numbers) and mix them together. Grind the sample so obtained in a coffee mill and additionally in a ball mill for 3 min. Allow the ground coffee to stand in a vacuum dessicator over P_2O_5 for 48 h. Separate the dry sample into 4 fractions by using sieves of 0.315, 0.250 and 0.160 mm mesh size. Weigh precisely an amount of approximately 0.01 g of each fraction and add KBr prior ground in the ball mill for 3 min so that the total weight of the mixture is 1 g. Homogenize the mixture for 4 min, and prepare from it 3 disks approximately 0.25 in weight. Measure the absorption spectra of the disks, which will contain 1% (m/m) coffee, on an FTIR spectrometer over the region of 4800–400 cm^{-1} at a resolution of 2 cm^{-1}; 64 scans, autoamplification. After measurement, determine the precise weight of the disk and calculate the factor f = 0.25/disk weight (in grams). Multiply the absorbances by this factor in order to eliminate the effect of different weights and thicknesses of the disks. Calculate the average spectrum for each fraction, and the absorbances at the following wavenumbers (cm^{-1}): 2925, 2855, 1930, 1648, 1384, 1268, 1161, 1057, and 814 respectively. An experimental matrix of absorbance values with 10 parameters (columns) and 16 objects (rows) is used as training set for discriminant analysis. Optimal results of discriminant analysis are several clusters of coffee samples separated by their origin. This method is suitable for the identification of an unknown coffee sample. The unknown sample has to be treated as described. The position of unknown sample points in the cluster diagram may serve for the identification of the sample origin.

References

1. Smith AW (1985) In: Clarke RJ, Mcrae R (eds) Coffee, Vol 1. Chemistry. Elsevier, New York
2. Bradbury AGW, Halliday DJ (1990) J Agric Food Chem 38:389
3. Hebák P, Hustopecký J (1987) Multivariate statistical methods (in Czech). SNTL, Prague