ELSEVIER

Food Chemistry 66 (1999) 365–370

Food Chemistry

www.elsevier.com/locate/foodchem

Characterization of *arabica* and *robusta* roasted coffee varieties and mixture resolution according to their metal content

M.J. Martín, F. Pablos, A.G. González*

Department of Analytical Chemistry, Faculty of Chemistry, University of Seville 41012-Seville, Spain

Received 9 December 1998; accepted 5 February 1999

Abstract

The metal content of roasted coffee samples belonging to the *arabica* and *robusta* varieties and coffee blends has been analysed. Ba, Ca, Cu, Fe, K, Mg, Mn, Na, P, Sr, Zn have been determined by inductively coupled plasma atomic emission spectrometry. Principal component and cluster analysis have been applied to characterize the coffee varieties. P, Mn and Cu have been found to be the most discriminating variables. Partial least squares regression was applied to determine the relative content of each variety in the coffee blends. This method has been applied to determine the percentage of the *robusta* variety in some commercial roasted coffee samples. © 1999 Elsevier Science Ltd. All rights reserved.

Keywords: Coffee; Metals; Pattern recognition; Chemometrics

1. Introduction

Coffee is commercially available as one of two varieties known as *arabica* and *robusta*. The coffee beverages are made from *arabica* and *robusta* roasted beans or blends of these two. *Arabica* coffees are generally considered of better quality and consequently they command higher prices. By taking this into account and considering that in recent times there is an increasing practice of selling coffees based on their varietal and/or geographic origin (Downey, Brandet, Wilson, & Kemsley, 1997), it is important to have methods to characterize these two varieties. These methods could assess the quality of the product in cases of fraudulent or accidental mislabeling. Often *arabica* and *robusta* green coffee beans can be distinguished by their size but the roasting process eliminates this visual criterion. Though the chemical composition of the two varieties is similar, it is possible to find parameters that may differentiate them. The metal content of the coffee beans has been proved to be adequate to differentiate between the *arabica* and *robusta* varieties (Clarke & Macrae, 1985; Haswell & Walmsley, 1998; Krivan, Burth, & Feria-Morales, 1993; Martín, Pablos, & González, 1998).

In this paper a study of the metal content of roasted coffee samples has been carried out. The concentrations of Ba, Ca, Cu, Fe, K, Mg, Mn, Na, P, Sr, Zn have been determined and used as chemical descriptors to differentiate between roasted coffee samples from the *arabica* and *robusta* varieties. Pattern recognition (PR) techniques such as principal component and cluster analysis have been applied. Compositional analysis of roasted coffee mixtures has been done by applying partial least squares (PLS) regression to the metal content data. By using this method the percentage of the *robusta* variety in retailed coffee samples has been determined.

2. Experimental

2.1. Analysis of the metal content

The metal determinations were carried out on a Fison-ARL 3410 inductively coupled plasma atomic emission spectrometer used under the same operational conditions as those described in a previous paper (Martín et al., 1998). Before the analysis, the samples were dried at 103°C (International Standard ISO 11294, 1994) until constant weight to determine their moisture. Assay portions of 1 g of coffee samples were mineralized with a heated mixture (1:4 v/v) of concentrated

---

* Corresponding author. Tel.: +34-9-54-557-173; fax: +34-9-54-557-168.

e-mail: agonzale@cica.es

PII: S0308-8146(99)00092-8

sulphuric and nitric acids. Then the treated samples were suitably diluted and filtered before ICP-AES analysis.

2.2. *Coffee samples*

Eighteen green coffee samples from different geographic origins were selected: Nine of the *arabica* variety and nine of the *robusta* one. They were laboratory roasted, ground and then stored in polyethylene flasks till the analysis. Table 1 shows the origin and the identification code assigned to each sample. One each of the *arabica* and *robusta* samples was selected at random, and 12 blends of these samples were prepared. The *robusta* content varied in the range 0–60% (w/w). Commercial roasted coffee samples were obtained from the market.

Table 1
Analysed roasted coffee samples

Variety	Origin	Code
Arabica	Brazil	1A
Robusta	Ivory Coast	2R
Arabica	Nicaragua	3A
Robusta	Vietnam	4R
Arabica	Honduras	5A
Robusta	Camerun	6R
Arabica	Salvador	7A
Robusta	Uganda	8R
Arabica	Colombia	9A
Robusta	Ivory Coast	10R
Arabica	Guatemala	11A
Robusta	Uganda	12R
Arabica	Brazil	13A
Arabica	Nicaragua	14A
Robusta	Camerun	15R
Arabica	Colombia	16A
Robusta	Uganda	17R
Robusta	Ivory Coast	18R

2.3. *Data analysis*

Eleven metals, that have been determined in the coffee samples, were considered as chemical descriptors and will be referred to as Ba, Ca, Cu, Fe, K, Mg, Mn, Na, P, Sr and Zn. A data matrix whose rows are the samples and whose columns are the variables was built (Table 2). Principal component analysis (PCA) (Chatfield & Collins, 1981) and cluster analysis (Massart & Kauffmann, 1983) were applied to discriminate between the coffee varieties. PLS (Martens & Naes, 1989) was used for the compositional analysis of the blends. The PR calculations were made by using the statistical package CSS:STATISTICA from Stafsoft^{TM} (Tulsa, OK, USA). Data processing for PLS was carried out using own developed software.

3. Results and discussion

In a recent paper (Martín et al., 1998) the most discriminating metals to differentiate between green coffee varieties were found to be P, Mn and Cu from a chemometric approach. The same one has been applied to the present data at a first stage. Once the most discriminating variables were selected, a PLS based method was used to resolve the *arabica*/*robusta* mixtures.

Table 2
Metal content (% w/w dry base) of roasted coffee samples[a]

Sample	Zn	P	Mn	Fe	Mg	Cu	Ca	Sr	Ba	Na	K
1A	2.899E-3	0.139	2.68E-3	5.98E-3	0.173	1.78E-3	0.089	4.13E-4	2.92E-4	2.30E-3	1.466
2R	1.393E-3	0.158	1.22E-3	5.61E-3	0.161	1.68E-3	0.094	4.14E-4	9.1E-5	2.14E-3	1.448
3A	1.933E-3	0.134	1.41E-3	4.62E-3	0.174	1.34E-3	0.097	5.59E-4	5.31E-4	3.34E-3	1.493
4R	2.997E-3	0.173	1.98E-3	7.34E-3	0.165	1.31E-3	0.087	4.40E-4	5.01E-4	3.38E-3	1.541
5A	5.44E-4	0.140	2.93E-3	4.66E-3	0.173	1.12E-3	0.090	5.75E-4	9.17E-4	3.01E-3	1.422
6R	3.182E-3	0.179	1.33E-3	5.06E-3	0.146	1.46E-3	0.108	6.93E-4	2.71E-4	1.99E-3	1.438
7A	3.690E-3	0.129	1.90E-3	5.27E-3	0.178	1.18E-3	0.109	5.98E-4	3.79E-4	1.17E-3	1.379
8R	4.94E-4	0.175	1.33E-3	5.41E-3	0.162	1.52E-3	0.113	6.66E-4	4.94E-4	1.69E-3	1.417
9A	1.790E-3	0.137	3.61E-3	5.20E-3	0.182	1.29E-3	0.109	1.193E-3	1.008E-3	2.33E-3	1.358
10R	7.10E-4	0.158	1.23E-3	5.17E-3	0.171	1.71E-3	0.104	5.48E-4	2.03E-4	6.6E-4	1.455
11A	5.08E-4	0.124	2.84E-3	4.03E-3	0.173	1.19E-3	0.088	5.89E-4	5.89E-4	1.06E-3	1.273
12R	1.297E-3	0.181	1.21E-3	6.37E-3	0.167	1.56E-3	0.117	7.69E-4	5.27E-4	1.47E-3	1.499
13A	5.06E-4	0.138	2.96E-3	5.09E-3	0.193	1.26E-3	0.101	5.06E-4	5.27E-4	1.70E-3	1.407
14A	8.97E-4	0.132	1.32E-3	4.08E-3	0.187	1.22E-3	0.103	5.15E-4	4.74E-4	1.60E-3	1.378
15R	7.29E-4	0.185	1.54E-3	6.09E-3	0.166	1.60E-3	0.135	8.00E-4	3.85E-4	1.21E-3	1.389
16A	5.67E-4	0.144	4.46E-3	4.94E-3	0.194	1.35E–3	0.103	7.70E-4	6.59E-4	1.77E-3	1.364
17R	5.79E-4	0.196	1.33E-3	6.30E-3	0.178	1.70E-3	0.123	8.53E-4	4.57E-4	1.10E-3	1.408
18R	1.514E-3	0.161	1.31E-3	6.25E-3	0.182	1.72E-3	0.122	5.68E-4	1.59E-4	1.59E-3	1.401

[a] Average of triplicate determinations.

3.1. PCA-based display methods and cluster analysis

The two principal components (PCs) issued from PCA on the data matrix explain up to 61% of the total variance, being 36.5% explained by PC1 and 24.5% by PC2. Fig. 1 shows the scores plot for the studied samples. At a glance, *arabica* samples are confined within −1.8 to −0.2 PC1 scores (the negative side of PC1) and *robusta* ones, in a narrower interval from 0.6 to 1.4 PC1 scores (the positive side of PC1). This clearly indicates that the two classes are fully separated from PCA. Factor loadings extracted from PCA indicate that P, Cu and Mn are the variables with more contribution to PC1, and hence, with high discriminating power. For

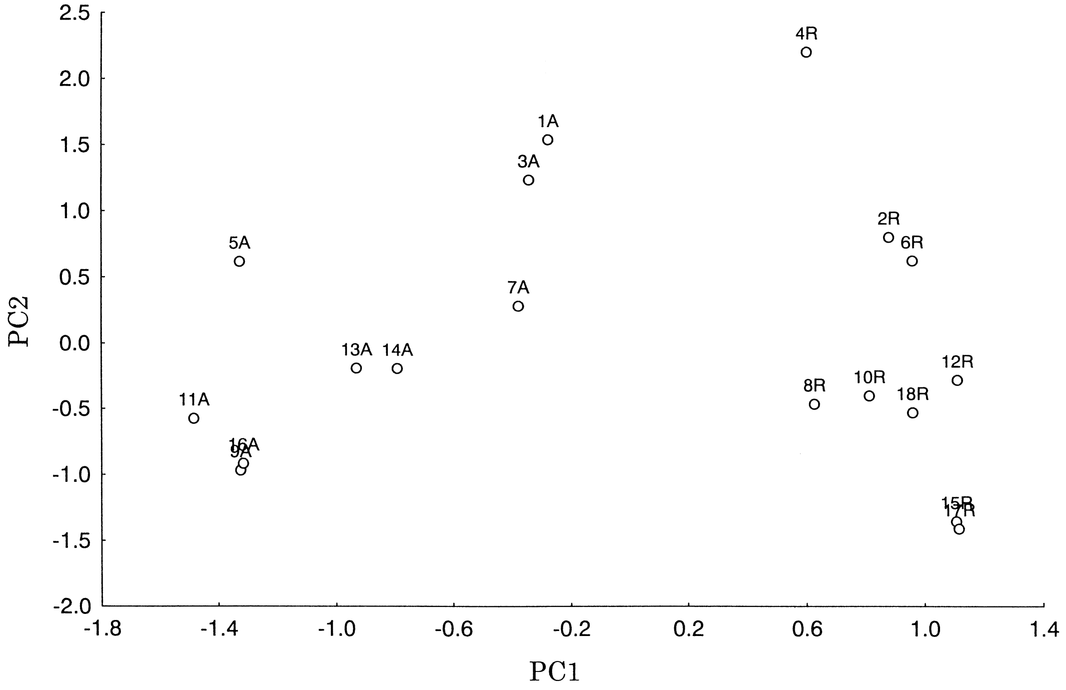

Fig. 1. PCA scores plot for the first PCs.

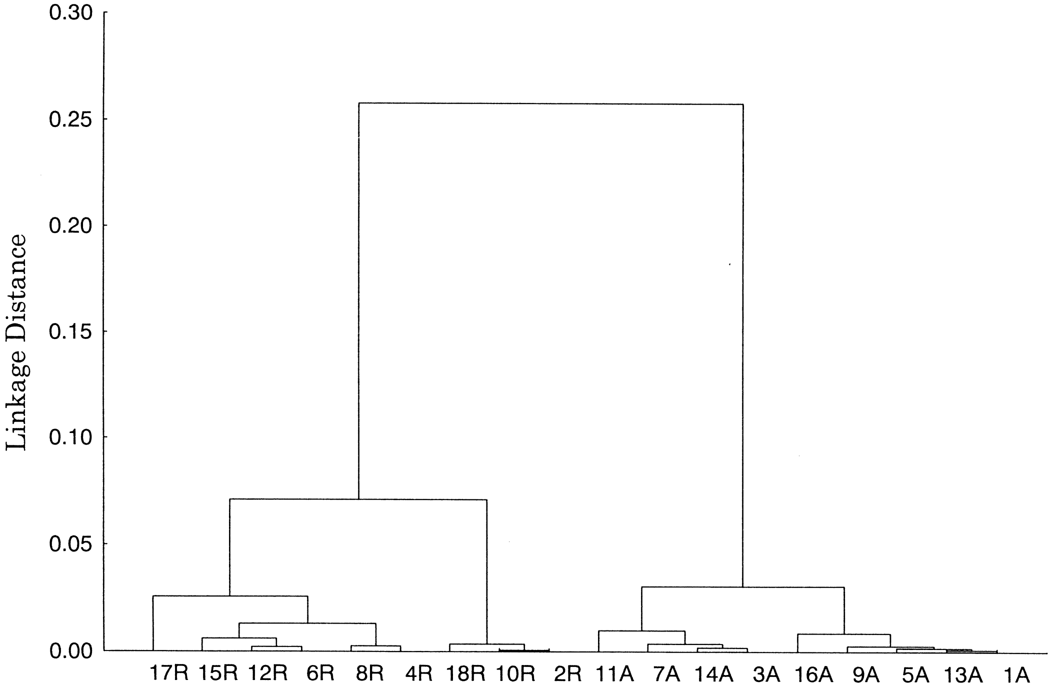

Fig. 2. Dendrogram of coffee samples by using P, Mn and Cu variables.

assessing this statement, a hierarchical cluster analysis of samples was carried out using the variables P, Cu and Mn. The Euclidean distance was used as similarity measurement and the Ward's method as amalgamation rule (Martín et al., 1998). In Fig. 2, the corresponding dendrogram is presented. Two well-separated clusters appear corresponding to the two studied classes without any conflict situation.

Another very straightforward way to choose discriminating variables by direct inspection is to draw the variable–variable plots. The scatter plots of the samples using as axis variables P-Cu and P-Mn couples are depicted in Figs. 3 and 4, respectively. As it can be seen, in both cases the separation of *arabica* and *robusta* samples is fair. As a conclusion, it may be stated that variables P, Cu and Mn are very suitable for discriminating

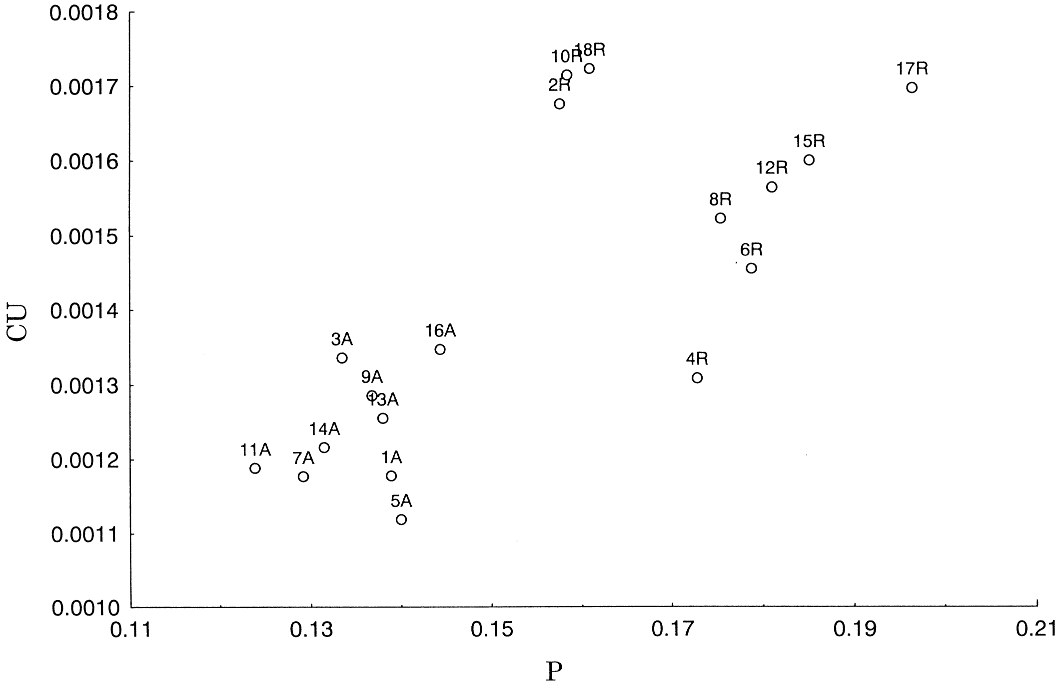

Fig. 3. Variable-variable plot, P-Cu.

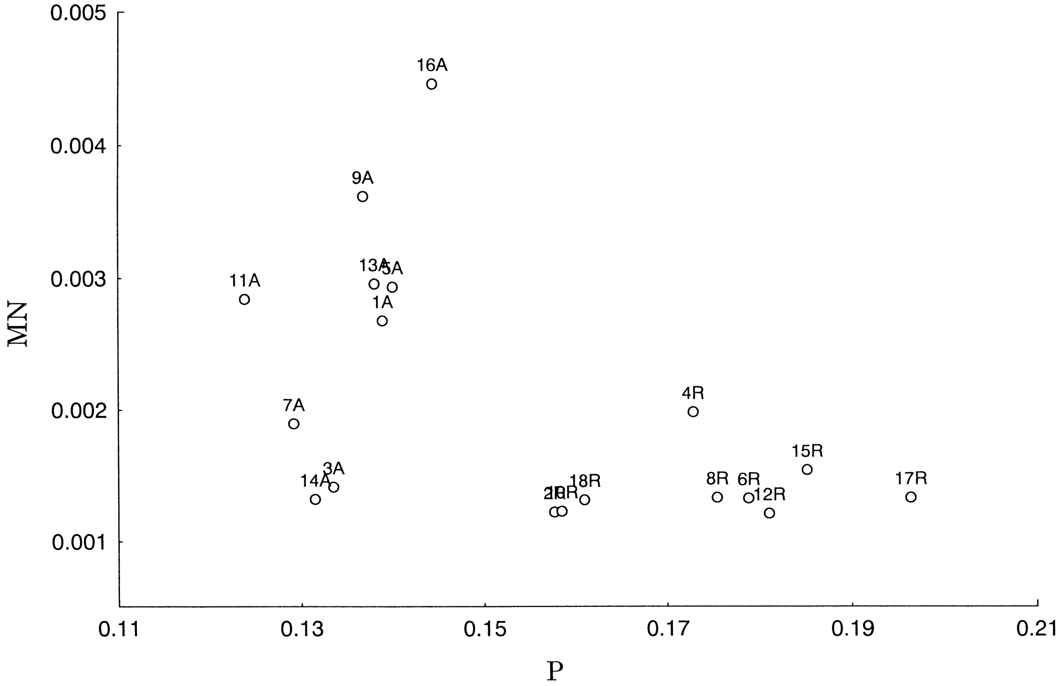

Fig. 4. Variable–variable plot, P-Mn.

between *arabica* and *robusta* roasted coffee varieties, as it was found formerly for green ones (Martín et al., 1998).

3.2. *Resolution of* arabica/robusta *mixtures*

As it was indicated in Section 2.2, *arabica/robusta* roasted coffee mixtures were prepared containing up to 60% *robusta*. A single 100% pure *robusta* sample was considered, however, for comparison purposes. These mixtures were metal analysed by ICP-AES as previously reported. The metal content in P, Cu and Mn of each blend expressed in weight percentage (dry basis) is gathered in Table 3. PLS calibration was performed for resolving the different *arabica/robusta* mixtures as Briandet, Kemsley, and Wilson (1996) pointed out. Using internal cross validation (Martens & Naes, 1989), the optimal regression model was obtained with one PLS latent variable. The cross-validation predicted values for the percentage *robusta* in mixtures were plotted versus the actual values as depicted in Fig. 5. The behaviour is fairly linear with a correlation coefficient of 0.996. In order to apply this approach to industrially processed samples, three commercial ground roasted coffee samples of 100% *robusta*, 100% *arabica* and a blend with 25% *robusta* (according to the label claim) were analyzed in this way. The results obtained were 94.0, 103.7 and 25.3%, respectively.

Table 3
Metal content (% w/w dry base) of roasted coffee mixtures[a]

Robusta (%)	P	Mn	Cu
0	0.136	2.48E-3	1.58E-3
5	0.138	2.36E-3	1.67E-3
10	0.140	2.29E-3	1.71E-3
15	0.142	2.21E-3	1.68E-3
20	0.143	2.20E-3	1.71E-3
25	0.147	2.12E-3	1.76E-3
30	0.148	2.06E-3	1.75E-3
35	0.152	2.06E-3	1.76E-3
40	0.155	1.98E-3	1.78E-3
45	0.158	1.97E-3	1.80E-3
50	0.159	1.92E-3	1.83E-3
55	0.161	1.85E-3	1.86E-3
60	0.162	1.84E-3	1.92E-3
100	0.179	1.42E-3	2.03E-3

[a] Average of triplicate determinations.

4. Conclusion

The content of P, Mn and Cu can be used for discriminate *arabica* and *robusta* roasted coffee varieties. Besides, these descriptors enable the resolution of *arabica/robusta* mixtures with a prediction error of about 7%. These findings can be considered as a feasibility study that should be extended by analyzing a large number of commercial samples in order to obtain more reliable rules for authentication purposes.

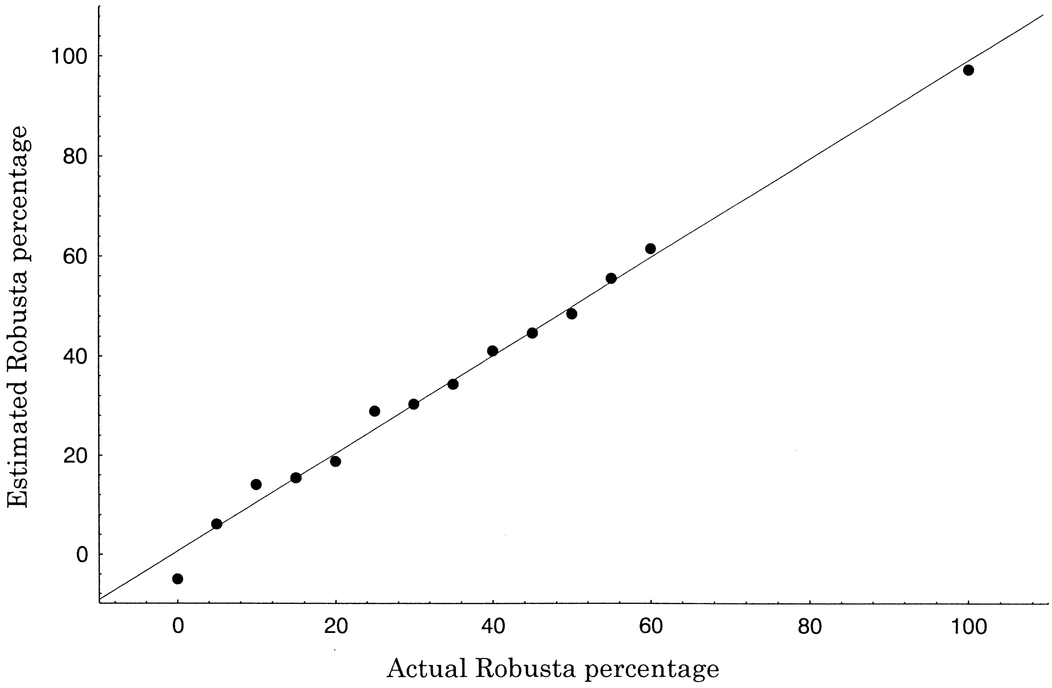

Fig. 5. PLS model-predicted versus actual % *robusta* in *arabica/robusta* mixtures.

References

Briandet, R., Kemsley, E. K., & Wilson, R. H. (1996). Discrimination of *arabica* and *robusta* in instant coffee by fourier transform infrared spectroscopy and chemometrics. *Journal of Agricultural and Food Chemistry, 44*, 170–174.

Chatfield, C., & Collins, A. J. (1981). *Introduction to multivariate analysis*. London: Chapman & Hall.

Clarke, R. J., & Macrae, R. (Eds.). (1985). *Coffee* (Vol I). London: Elsevier.

Downey, G., Briandet, R., Wilson, R. H., & Kemsley, E. K. (1997). Near- and mid- infrared spectroscopies in food authentication: coffee varietal identification. *Journal of Agricultural and Food Chemistry, 45*, 4357–4361.

Haswell, S. J., & Walmsley, A. D. (1998). Multivariate data visualization methods based on multielemental analysis of wines and coffees using total-reflection x-ray fluorescence analysis. *J. Anal. At. Spectrom, 13*, 131–134.

International Organisation of Standardization ISO 11294 (1994). Roasted ground coffee—determination of loss in mass at 103°C (routine method). Technical committee ISO/TC 34, *Agricultural food products*, Subcommittee SC IS, *Coffee*.

Krivan, V., Barth, P., & Feria-Morales, A. (1993). Multi-element analysis of green coffee and its possible use for the determination of origin. *MiKroKim. Acta, 110*, 217–236.

Martens, H., & Naes, T. (1989). *Multivariate calibration*. Chichester: Wiley.

Martín, M. J., Pablos, F., & González, A. G. (1998). Characterization of green coffee varieties according to their metal content. *Analytica Chimica Acta, 358*, 177–183.

Massart, D. L., & Kauffmann, L. (1983). *Interpretation of analytical data by use of cluster analysis*. New York: Wiley.