## ORIGINAL RESEARCH ARTICLE

# Chemometric tools for the characterisation of honey produced in La Pampa, Argentina, from their elemental content, using inductively coupled plasma optical emission spectrometry (ICP-OES).

**José M. Camiña[1]\*, Miguel A. Cantarelli[1], Valeria A. Lozano[1], María S. Boeris[1], María E. Irimia[2], Raúl A. Gil[3] and Eduardo J. Marchevsky[3]**

[1]Departamento de Química. Facultad de Ciencias Exactas y Naturales. Universidad Nacional de La Pampa. Av. Uruguay 151 (6300) Santa Rosa. La Pampa. Argentina.
[2]Universidad Católica de San Juan. A. José Ignacio de la Roza 1616 (5400) San Juan, Argentina.
[3]Area de Química Analítica. Facultad de Química, Bioquímica y Farmacia. Universidad Nacional de San Luis. Chacabuco y Pedernera (6300) San Luis. Argentina.

\*Corresponding author. Email: : jcaminia@exactas.unlpam.edu.ar **or** jcaminia@yahoo.com.

## Summary

Thirty two samples of natural honey produced in the province of La Pampa (Argentina) were characterised on the basis of their phosphorous, aluminium, iron, calcium, magnesium and sodium contents. Analytical determinations were carried out using inductive coupled plasma optical emission spectrometry (ICP-OES). For characterisation, chemometric methods used were Principal Components Analysis (PCA) and Cluster Analysis (CA), while for classification, Linear Discriminant Analysis (LDA) was used. The results show that samples obtained within 50 km from the centre of the province (around Santa Rosa, the capital city), were different from those coming from the rest of the province, generating a classification based on geographical origin. Phosphorous content was the most significant variable in the classification of the PCA model.

## Herramientas quimiométricas para la caracterización de la miel producida en La Pampa, Argentina, a partir de su contenido elemental, usando espectrometría de emisión óptica con plasma acoplado inductivamente (ICP-OES).

### Resumen

Treinta y dos muestras de miel producida naturalmente en la provincia de La Pampa (Argentina) fueron caracterizadas en función de su contenido en fósforo, aluminio, hierro, calcio, magnesio y sodio. Las determinaciones analíticas se realizaron utilizando espectrometría de emisión óptica con plasma acoplado inductivamente (ICP-OES). Para la caracterización, los métodos quimiométricos utilizados fueron el Análisis de Componentes Principales (PCA) y el Análisis de Agrupamiento (CA), mientras que

para la clasificación se utilizó el Análisis Discriminante Lineal (LDA). Los resultados indican que las muestras obtenidas dentro de los 50 km desde el centro de la provincia (alrededores de Santa Rosa, la capital), fueron diferentes de las procedentes del resto de la provincia, generando una clasificación basada en el origen geográfico. El contenido en fósforo fue la variable más importante para la clasificación mediante el modelo de PCA.

**Keywords:** honey, classification, chemometrics, Principal Components Analysis, Cluster Analysis, Discriminant Analysis.

# Introduction

Honey is a natural food produced by bees from nectar or other secretions of flowers. Bees transform the original nectar into honey by using specific enzymes and storing it in cells within bee hives. Honey has a content of 80–85 % carbohydrates, 15–17 % water, 0.3 % proteins, 0.2 % ash, and minor quantities of amino-acids and vitamins as well as other components in small amounts. Honey composition depends on its botanical origin, a useful approach for its geographical characterization. The mineral composition of honey depends on nectar composition (Hernandez *et al.*, 2005) and minor elements coming from the soil. Usually, minor components are used to classify and discriminate food samples. Ash represents the total mineral content, potassium being the dominant element, followed by others, such as chloride, sodium, calcium, and magnesium. In a previous paper (Camiña *et al.*, 2004), we used chemometric methods such us the Partial Least Square (PLS) method to predict concentration of several elements present in honey. We also established relationships between physicochemical parameters and honey quality, using Principal Components Analysis (PCA) (Lozano *et al.*, 2006). Other studies have successfully used the content of several metals to classify honey (Herrero *et al.*, 1996; Latorre *et al.*, 1998; Barbaste *et al.*, 2002). The criterion for classification nevertheless depends on regional conditions such as soil chemical composition, but can also depend on other factors, such as genetic ones (from plants and bees), or agricultural practices, due to the use of fertilizers and pesticides. Selected variables can play an important role and must be properly chosen.

In recent years, the designation of origin has been an important factor to define quality and cost parameters (Barbaste *et al.*, 2002), but chemometric tools can also help to find fraud and adulteration of natural honeys (Hernandez *et al.*, 2005).

The province of La Pampa produces a significant proportion of the total honey production in Argentina. This production is destined for export, generally as non specific honey without floral origin and mainly for food factories. Nevertheless, honey production represents an important activity in the economy of the province. From an economic perspective, therefore, applying the geographical origin criterion as a quality parameter in order to produce certified honey with designation of origin could be a useful approach.

In this paper we discuss the characterization and classification of 32 natural honey samples from unspecified multifloral origins. Phosphorous, aluminium, iron, calcium, magnesium and sodium contents in honey were used as variables. These elements derive from the soil, due to the fact that plants use them as nutrients. The soil elemental content can therefore be used for characterizing and classifying honey, showing characteristics of the soil where honey was produced. This can include anthropogenic activities such as intensive agricultural practices (e.g. the use of fertilizers and pesticides) and can help to define geographical areas and to find differences between them. The aim was to identify differences between honey samples from the province of La Pampa.

# Materials and Methods

### Samples
Thirty two natural honey samples were obtained from their original production place, from December 2005 to March 2006. Fifteen samples came from different localities of the capital department within 50 km around the city of Santa Rosa (Capital area "C" samples) whilst the remaining seventeen samples came from other localities of the province of La Pampa (Non Capital "NC" samples).

### Apparatus
Emission measurements were made using an inductively coupled plasma optical emission spectrometer (ICP-OES), Baird model 2070 ICP, with 100 cm optical length Czerny Turner monochromator. Weight measurements were made using a Mettler H20 analytical balance with a precision of 0.01 mg.

### Reagents
Standard stock solutions of all elements were prepared from Merck reagents atomic absorption spectroscopy grade. Nitric and perchloric acids, Merck suprapure grade, were used in the mineralisation step. Both the standard solutions and the samples were diluted using de-ionized water.

### Sample mineralisation
Five grams of honey were accurately weighed into a 120 ml glass beaker. Then 25 ml of nitric acid and 10 ml of perchloric acid were added and heated in a sand bath, to mineralise organic material until almost dry. The acid clear solutions were transferred to 50 ml flasks and diluted with deionised water. The mineralisation procedure was carried out as in previous work (Camiña *et al.*, 2004) showing an adequate recovery percentage of around 97 %. Samples and standard solutions were measured in ICP-OES equipment by direct nebulisation.

### Results processing
Principal components analysis (PCA) was carried out using the Unscrumbler 6.11 software package (CAMO AS, Norway). Cluster analysis (CA) was carried out using Multivariado software (Argentina). Linear discriminant analysis (LDA) was carried out using InfoStat software (Argentina). The original data matrix

consisting of 32 rows and 6 columns was auto-scaled for column, subtracting the media of each column to every sample and dividing it for their standard deviation.

# Results

## Sample statistical data

Table 1 shows statistical data including mean, standard deviation, minimum and maximum concentration levels, for capital (C), non capital (NC) and total honey samples.

## Characterization of honey samples

*Principal Components Analysis*

Principal Components Analysis (PCA) is a multivariate tool that can process an enormous amount of data produced by computers and measurement techniques. Principal components analysis was used to search for data trends, by combining the original variables. This multivariate technique provides a partial view of the data in a space with a reduced number of dimensions while preserving most of their variability. The PCA focus is to discover the true dimension of data (reducing the number of original variables) and to allow for the identification of new latent variables (Johnson, 2000).

The PCA model was validated by the cross-validation method, meaning that a sample is left out one time and used to find the validation error. Then, all samples are left out one time and the total error is obtained. (Martens *et al.*, 1996; Beebe *et al.*, 1998).

Fig. 1 shows the percentage of variance explained as a function of the number of principal components indicating that 4 PC explain 96.97 % of the original information in the calibration step and a 93.81 % in the validation step.

Fig. 2 shows the scores plot obtained by PCA. Two groups, Capital (C) and Non Capital (NC) samples are clearly separated and enclosed by two ellipses. Capital honey samples are placed at the top of the score plot, while Non Capital samples are placed
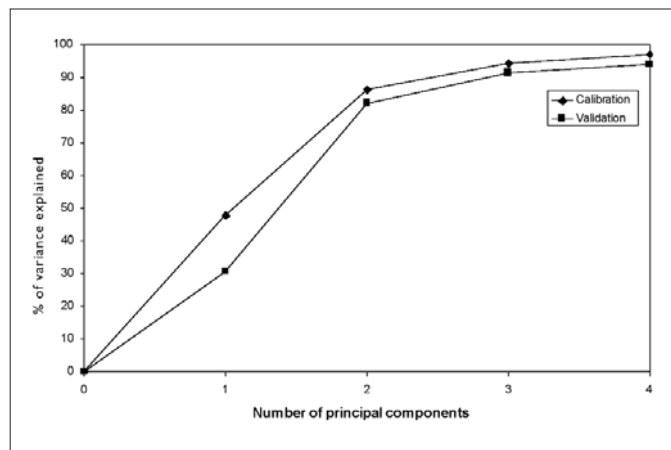


**Fig 1**. % of variance explained as function of number of principal components.
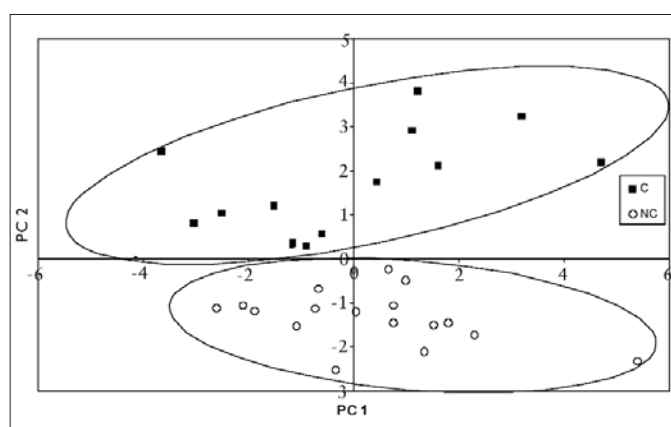


**Fig 2**. Scores plot of natural honey using PC1 vs. PC 2.

at the bottom. The differences between both Capital and Non Capital honey samples are based on their phosphorous contents, shown in the loadings plot (Fig. 3). In this plot, the first component (PC 1) mainly describes the calcium, sodium and

**Table 1**. Concentration values for phosphorous, iron, aluminium, calcium, magnesium and sodium in natural honey samples from the province of La Pampa.

| Region | N | Statistical parameter | P* | Fe* | Al* | Ca* | Mg* | Na* |
|---|---|---|---|---|---|---|---|---|
| **Capital (C)** | 15 | Mean | 40.88 | 3.66 | 1.49 | 46.97 | 19.24 | 21.07 |
| | | SD | 15.53 | 3.19 | 1.64 | 19.80 | 10.09 | 11.09 |
| | | Minimum | 18.38 | 1.13 | 0.02 | 18.60 | 6.00 | 6.10 |
| | | Maximum | 66.62 | 10.32 | 5.22 | 87.73 | 46.57 | 42.85 |
| **Non Capital (NC)** | 17 | Mean | 14.51 | 4.31 | 3.97 | 57.99 | 25.88 | 25.86 |
| | | SD | 15.79 | 2.13 | 2.84 | 17.00 | 7.43 | 8.91 |
| | | Minimum | 1.17 | 2.07 | 0.00 | 31.55 | 13.25 | 14.54 |
| | | Maximum | 24.27 | 8.53 | 13.03 | 102.06 | 43.24 | 45.59 |
| **All** | 32 | Mean | 26.87 | 4.01 | 2.81 | 52.83 | 22.77 | 23.61 |
| | | SD | 17.46 | 2.65 | 2.64 | 18.90 | 9.26 | 10.12 |
| | | Minimum | 1.17 | 1.13 | 0.00 | 18.60 | 6.00 | 6.10 |
| | | Maximum | 66.62 | 10.32 | 5.22 | 102.06 | 43.24 | 42.85 |

(*Concentrations expressed in mg kg-1)
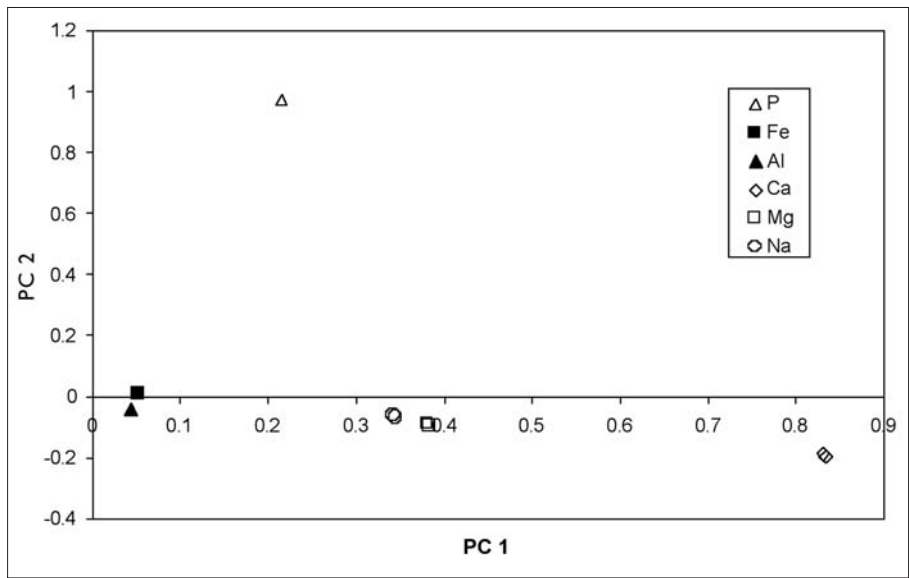
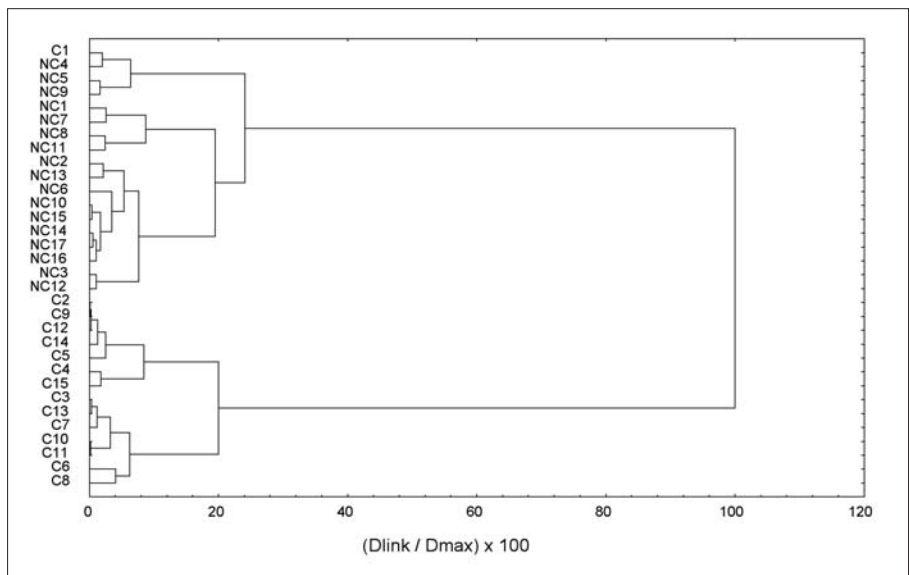**Fig 3**. Loadings plot, showing the effect of variables on principal components.



**Fig 4**. Cluster analysis dendrogram of natural honey samples using the Ward method.
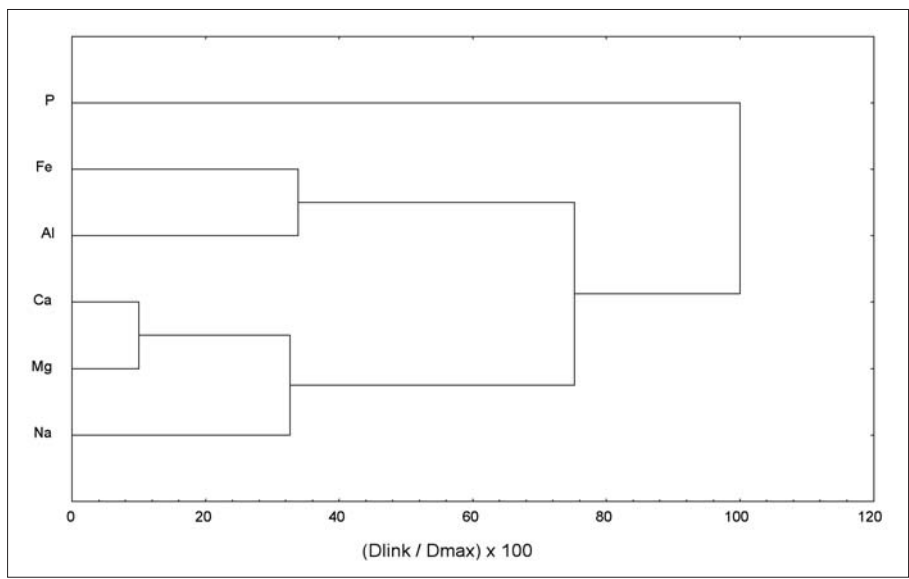


**Fig 5**. Cluster analysis dendrogram of variables using the Ward method.

magnesium behaviour; aluminium and iron are described to a lesser degree. PC1 enlarges the ellipses in the Fig. 1 scores plot. The second principal component (PC 2) mostly reflects the phosphorous content, showing the main differences between both groups in the scores graphic. This is a significant result as phosphorous content has not until now been used as a variable for chemometric honey classification.

*Cluster Analysis*
Cluster Analysis is used to classify objects, characterized by the values of a set of variables, into groups. It is therefore an alternative to Principal Components Analysis for describing the structure of a data table (Massart, *et al.*, 1997). Because of its unsupervised nature, Cluster Analysis is frequently used to screen data for clustering of samples. Cluster Analysis involves techniques that produce classification from unclassified data, allowing groups based on their similarity (Johnson, 2000).

Cluster analysis was carried out using an auto-scaled matrix of 32 samples and 6 variables. The Ward method was used as a hierarchical linkage criterion of amalgamation, and it is based on the minimum variance (Massart *et al.*, 1997; Johnson, 2000). A Pearson Coefficient was used as association criterion, based on the binary correlation coefficient.

Fig. 4 shows the tree diagram or dendrogram obtained using both the Ward method and the Pearson Coefficient. The C group may be statistically different from the NC group; only C1 appearing in the NC group. Cluster Analysis allows confirmation of the classification results obtained by PCA, using different multivariate methods.

Fig. 5 shows the dendrogram obtained for the six variables. There are no significant differences between the variables, but phosphorous behaves differently from the rest of the elements. Trivalent, divalent and monovalent ions are grouped and separated between them.

**Classification of honey samples**
*Linear Discriminant Analysis*
Linear Discriminant Analysis (LDA) is a classification method used to find a theoretical value resulting in the best possible discrimination between a priori established groups. Discrimination is based on weighing the theoretical values for each variable in such a way as to maximize between-group variance in comparison to within-group variance. (Johnson, 2000) Discriminant Analysis models involve sets of equations that are linear combinations of the independent variables, resulting in the maximum possible separation between groups (Hernandez *et al.*, 2005); these equations are known as Discriminant function (D) (Mongay Fernandez, 2005). The number of D equations depends on the number of K groups and the number of P variables, so S describes spaces of D function as follows:-

$$S = \text{minor} (K-1, P)$$

And the discriminant function can be described as:
$$D_S = a_{S0} + a_{S1} X_1 + a_{S2} X_2 + \ldots + a_{SP} X_P$$

In this paper, all variables were studied to confirm a normal fit distribution. Then, the LDA model was constructed using 75% of the honey samples, which were randomly chosen within each group; the remaining 25 % were used to validate it. Therefore, the training matrix consisted of 24 samples (12 C and 12 NC samples) and the validation matrix consisted of 3 C and 5 NC samples. Wilk's lambda of model was 0.164, a lower value than the one obtained in previous works (Hernandez *et al.*, 2005). The variables used were the same as for PCA and CA: P, Fe, Al, Ca, Mg and Na. S was = 1 and the discriminant function obtained using all variables was:-

$$D = -0.42 + 1.42 \times P - 1.1 \times Fe - 0.84 \times Al - 0.29 \times Ca - 0.98 \times Mg - 0.01 \times Na$$

Table 2 shows the results of the LDA model, showing a good fit. No sample in the training and validation matrix was misclassified, which indicates a prediction ability of 100 %.

***Table 2***. Result of recognition and prediction ability for linear discriminant analysis model.

| Matrix | Predicted group | | % correct |
|---|---|---|---|
| | C | NC | |
| **Training step** | | | |
| C | 12 | 0 | 100 |
| NC | 0 | 12 | 100 |
| % total | | | 100 |
| **Validation step** | | | |
| C | 3 | 0 | 100 |
| NC | 0 | 5 | 100 |
| % total | | | 100 |

# Discussion

Three multivariate methods are proposed to classify samples of natural honey using phosphorous, aluminium, iron, calcium, magnesium and sodium as variable elements. They can be easily analysed with several analytical methods such as the ICP-OES. By means of the PCA method, the use of phosphorous content as PCA characterization criterion produced good results, suggesting a major phosphorous presence in the soil from the Capital group compared to the Non Capital group. Calcium, sodium and magnesium enlarge Capital and Non Capital groups, showing that these elements are present in soils from both groups. The importance of the phosphorous variable in the loading plot can be interpreted in PC2, which shows a latent variable (i.e. anthropogenic activity), due to more intensive agricultural practices in the Capital (C) area in comparison with the Non Capital (NC) area through the use of phosphorous agents (fertilizers and pesticides) in soil and plants. In the same way, PC1 can be interpreted as another latent variable (i.e. soil characteristic) due to Ca influence in loading plots, because the province of La Pampa is characterized by calcareous soils.

CA as an unsupervised grouping method divided samples in two groups C and NC, similar to the PCA results. From 32 samples, only sample C1 was not correctly grouped using CA. The results show both the importance of the selected variables, as well as the association and linkage criterion, to obtain adequate grouping. The variables dendrogram plot shows differences

between P, Fe-Al, Ca-Mg and Na, which can be interpreted as a grouping criterion based on the oxidation state of these elements.

LAD used as classification method, allowed confirmation of differences between both Capital and Non Capital samples based on the geographical criterion, with a predictive ability of 100 %, which shows a good fit of the model in the training and prediction set. The same variables (P, Fe, Al, Ca, Mg and Na), were used for three multivariate methods, PCA, CA and DA, obtaining similar results in all cases.

The results show that the variables criterion used could be different depending on the geographical origin of the natural honey (Latorre et al., 1998; Hernandez et al., 2005). Phosphorous was the most important variable to define a criterion based on geographical classification in the province of La Pampa. Inductively coupled plasma optical emission spectrometry (ICP-OES) was used as a reliable analytical method. There is an important number of works using this method and other spectroscopic techniques, to determine multielemental content in honey samples (Viñas et al., 1997; Caroli et al., 1999; Caroli et al., 2000; Forte et al., 2001). Other analytical methods used in honey samples involve electro-analytical methods, such as stripping analysis (Li et al., 1995; Sanna et al., 2000; Muñoz et al., 2006) and ion chromatography with electrochemical detection (Buldini et al., 2001). In all cases, the determination of heavy metals in honey samples was carried out. In such works, a few transition elements were analyzed and methods were validated by spectrometric methods, as ICP-OES, ICP-MS and ETAAS. In this work, the ICP-OES method was used to determine phosphorous, iron, aluminium, calcium, magnesium and sodium in natural honey samples, due to the ability of this method to obtain a fast multielemental analysis by direct nebulisation, including phosphorous, which is not a reference to be analysed by electrochemical method in honey samples.

The results suggest that the proposed method is adequate to classify samples of natural honey from nearby areas within the province of La Pampa, Argentina. This method can be useful to identify fraud of origin in natural honey and can be useful to define the designation of origin in regional natural honey.

# Acknowledgements

# References

BARBASTE, M; MEDINA, B; SARAVIA, L; ORTIZ, M C; PEREZ TRUJILLO, J P (2002) Analysis of SIMCA models for denomination of origin of wines from Canary Islands, build by means of trace and ultratrace metal content. *Analytica Chimica Acta* 462: 161–174.

BEEBE, K; PELL, R; SEASHOLTZ, M (1998) *Chemometric: a Practical Guide*. Wiley and Sons; New York, USA, 183 pp.

BULDINI, P L; CAVALLI, S; MEVOLI A; SHARMA, J L (2001) Ion chromatographic and voltammetric determination of heavy and transition metals in honey. *Food Chemistry* 73 (4): 487–495.

CAMIÑA, J M; BOERIS, M S; MARTINEZ, L D; LUCO, J M;  MARCHEVSKY, E J (2004) Simultaneous Determination of Cu, Zn and Fe in Honey by Use of Partial  Least Square Regression PLS-2 Method. *Chemia Analityczna* 49 (5): 717–727.

CAROLI, S; FORTE, G; IAMICELLI, A L; GALOPPI, B (1999) Determination of essential and potentially toxic trace elements in honey by inductively coupled plasma-based techniques. *Talanta* 50: 327–336.

CAROLI, S; FORTE, G; ALESSANDRELLI, M; CRESTI, R; SPAGNOLI, M; D'ILIO, S; PAUWELS, J; KRAMER, G (2000) A pilot study for the production of a certified reference material for trace elements in honey. *Microchemical Journal* 67: 227–234.

FORTE, G; D'ILIO, S; CAROLI, S (2001) Honey as a candidate reference material for trace elements. *Journal of AOAC International* 84 (6): 1972–1980.

HERNANDEZ, O; FRAGA, J; JIMENEZ, A; JIMENEZ, F; ARIAS, J (2005) Characterization of honey from the Canary Islands: determination of mineral content by atomic absorption spectrophotometry. *Food Chemistry* 93: 449–458.

HERRERO, C; LÓPEZ, B;  LATORRE, M; GARCÍA, S; GARCÍA, S; FERNANDEZ, M (1996) Chemometric classification of honeys based quality control data. *Food Chemistry* 55: 283–287.

JOHNSON, D E (2000) *Métodos multivariados aplicados al análisis de datos*. Internacional Thompson Editores; Mexico DF, Mexico; 327 pp.

LATORRE, M J; PEÑA, R; PITA, C; BOTANA, A; GARCÍA, S; HERRERO, C. (1998) Chemometric classification of honeys based on their type II. Metal content data. *Food Chemistry*  66: 263–268.

LI, Y; WAHDAT, F; NEEB, R (1995) Digestion-free determination of heavy metals (Pb, Cd, Cu) in honey using anodic stripping differential pulse voltammetry and potentiometric stripping analysis. *Fressenius Journal of Analytical Chemistry* 351 (7): 678–682.

LOZANO, V A; BOERIS, M S; SCOLES, G E; PATTACINI, S H; MARCHEVSKY; CAMIÑA, J M (2006) Evaluación de Parámetros Físicos y Químicos en Miel a Través del Uso de Análisis de Componentes Principales. *Afinidad* 63 (521): 40–47.

MARTENS, H; NAES, T (1996) *Multivariate Calibration*. J. Wiley and Sons Ltd; New York, USA, 73 pp.

MASSART, D L; VANDEGINSTE, B G M; BUYDENS, L M C; DE JONG, S; LEWI, P J; SMEYERS-VERBEKE, J (1997) *Handbook of Chemometrics and Qualimetrics Vol. B*. Elsevier; Amsterdam, Netherlands, 57 pp.

MONGAY FERNANDEZ, C (2005) *Quimiometría*. Publicaciones Universidad de Valencia; Valencia, España, 247 pp.

MUÑOZ, E; PALMERO, S (2006) Determination of heavy metals in honey by potentiometric stripping analysis and using a continuous flow methodology. *Food Chemistry* 94 (3): 478–483.

SANNA, G; PILO, M I; PIU, P C; TAPPARO, A; SEEBER, R (2000) Determination of heavy metals in honey by anodic stripping voltammetry at microelectrodes. *Analitica Chimica Acta* 415 (1-2): 165–173.

VIÑAS, P; LÓPEZ-GARCÍA, I; LANZÓN, M; HERNÁNDEZ-CÓRDOBA, M (1997) Direct determination of lead, cadmium, zinc, and copper in honey by electrothermal atomic absorption spectrometry using hydrogen peroxide as a matrix modifier. *Journal of Agricultural Food Chemistry* 45: 3952–3956.