

# COMBINING MULTIVARIATE ANALYSIS AND POLLEN COUNT TO CLASSIFY HONEY SAMPLES ACCORDINGLY TO DIFFERENT BOTANICAL ORIGINS

Eduardo Corbella<sup>1</sup>, and Daniel Cozzolino<sup>2</sup>\*

## ABSTRACT

This study reports the combination of multivariate techniques and pollen count analysis to classify honey samples according to botanical sources, in samples from Uruguay. Honey samples from different botanical origins, namely *Eucalyptus* spp. (n = 10), *Lotus* spp. (n = 12), *Salix* spp. (n = 5), "mil flores" (*Myrtaceae* spp.) (n = 12) and coronilla (*Scutia buxifolia* Reissek) (n = 10) were analysed using Melissopalynology (pollen identification). Principal component analysis (PCA) and linear discriminant analysis (LDA) were used to classify the honey samples accordingly to their botanical origin based on a pollen count. Honey samples of higher percentage (> 70%) of *Eucalyptus*, *Lotus* and *Scutia* pollen were 100% correctly classified, whilst samples from *Myrtaceae* spp. and *Salix* were 80 and 66% correctly classified, respectively. The use of PCA and LDA combined with pollen identification proved useful in characterizing honey samples from different botanical origins.

**Key words:** honey, Uruguay, principal component analysis, linear discriminant analysis, pollen analysis.

## INTRODUCTION

Scientists in the food and beverage industries are interested in identifying the main changes in process that may lead to a change in quality or compromises in any ingredient or finished products. Food authenticity issues in the form of adulteration and improper description have been around for a long time. Recently, the demand for natural honey has increased, consequently, methods to assure the authenticity of honey can be economically important. Several factors contribute to the quality properties of honey, such as high osmotic pressure, lower water activity, low pH, and low protein content, among others (Anklam, 1998; Bogdanov, 1999).

The authenticity of honey has two aspects, one related to honey production, and the other to description, such as geographic and botanical origin. A number of techniques have been used to determine honey authenticity and botanical origin, including the determination of aromatic compounds and flavonoids, amino acids and sugars by high performance liquid chromatography (HPLC), detection of aroma compounds by gas chromatography-mass spectrometry (GC-MS), determination of anions and cations by ion chromatography (IC), and mineral content (Anklam, 1998; Mateo and Bosch-Reig, 1998; Anapuma *et al.*, 2003; Serrano *et al.*, 2004). Spectroscopic techniques, such as mid infrared (MIR), near infrared (NIR) and Raman spectroscopy were also used to determine chemical characteristics (e.g., sugars) and contamination in honey

---

<sup>1</sup> Instituto Nacional de Investigación Agropecuaria, Estación Experimental INIA La Estanzuela, Ruta 50-km 12, Colonia, Uruguay.  
E-mail: ecorbella@adinet.com.uy

<sup>2</sup> The Australian Wine Research Institute, Waite Road, Glen Osmond, PO Box 197, Adelaide, Australia, 5064.  
Email: Daniel.Cozzolino@awri.com.au \*Corresponding author.

Received: 4 May 2007. Accepted: 10 August 2007.

samples from different origins (Cozzolino and Corbella, 2005; Bertelli *et al.*, 2007).

In recent years, characterization of honey by means of both chemical and sensory properties has received increasing attention by several authors (Latorre *et al.*, 1999; 2000; Hermosin *et al.*, 2003; Cordella *et al.*, 2003; Terrab *et al.*, 2004). Quality control methods, in conjunction with multivariate statistical analysis, have been found to be able to classify honey from different geographic regions, detect adulteration and describe chemical characteristics (Cordella *et al.*, 2002; 2003; Marini *et al.*, 2004; Devillers *et al.*, 2004). Both, pollen identification and count have been used for authentication of honey samples accordingly to floral type, although there are difficulties in assuring a correct assignment of their origin (Serrano *et al.*, 2004). However, due to its simplicity, this technique has been used extensively to identify different types of honey samples from different botanical origins.

Multivariate analysis involves the use of mathematic and statistical techniques to extract information from complex data sets. It helps to look at the sample as a whole (holistically) and not just at a single component, allowing to untangle all the complicated interactions between the constituents and understand their combined effects on the whole matrix. Nowadays, the application of supervised pattern recognition and multivariate statistical techniques, like principal component analysis (PCA), linear discriminant analysis (LDA) or discriminant analysis (DA), provides the possibility to analyse the entire food sample matrix and to make a classification possible (Ashurst and Dennis, 1996).

The aim of this work was to explore the use of multivariate techniques applied to pollen count analysis to classify honey samples collected in Uruguay according to their botanical origin.

## MATERIALS AND METHODS

### Samples, sampling and pollen analysis

Forty nine (n = 49) typical honey samples were obtained directly from the beekeepers and collected during the 2004-2005 season. Samples were harvested from different locations across Uruguay, South America. Honey samples were collected from stainless steel drums (200 kg weight) directly provided by the beekeepers.

Extraction of honey samples from combs was done by centrifugation. All samples were unheated and were analysed no later than four weeks after extraction from the hives by the beekeepers. Collection sites were selected to include different botanical origins, soil characteristics and regions. In Uruguay, species-specific floral types of honey are obtained by beekeepers, pursuing a particular floral species for honey production, through controlling the foraging of their honeybees, *Apis mellifera* (Hymenoptera: Apidae), by hive location (near to one species of plant). Information about season, hive location and available floral sources were collected by asking the beekeepers to accurately identify the floral source of the honey samples. Furthermore, the origins of honey samples according to their different floral origins were confirmed and analysed by Melissopalynology (pollen identification) using light microscopy analysis. Thus, five groups were defined, namely *Eucalyptus* spp. (n = 10), *Lotus* spp. (n = 12), *Salix* spp. (n = 5), *mil flores* (*Myrtaceae* spp.) (n = 12) and coronilla (*Scutia buxifolia*) (n = 10). Honey samples belonging to the *Myrtaceae*, *Scutia* and *Salix* species were defined as Monte (bush honey) by the beekeepers, however they were analysed separately for the purpose of this study. Additionally, one of the honey samples belonging to the *Lotus* spp. group had a high content of pollen from clover (*Trifolium* spp.). Nevertheless, it was included in this group for statistical analysis.

### Statistical and multivariate analysis

Pollen count was analysed by means of unsupervised and supervised pattern recognition techniques, namely principal component analysis (PCA) and linear discriminant analysis (LDA), respectively. PCA was used to determine which variable discriminates between honey samples of different floral origin. PCA is a mathematical procedure for resolving sets of data into orthogonal components (principal components) whose linear combinations approximate the original data to any desired degree of accuracy, in such a way the data is presented graphically in those axes (Naes *et al.*, 2002). PCA was used to derive the first principal components from the data, and used in further analysis to examine the grouping of samples, outliers and in order to visualise the relative distribution of the honey samples according to their botanical origin. PCA was performed on the pollen count data, after centering and auto-scaling of the variables (Naes *et al.*, 2002).

LDA is used to determine which variables discriminate between two or more naturally occurring groups (Otto, 1999). This mathematical procedure maximizes the variance between groups and minimizes the variance within each group, in such a way that samples belonging or not to a specific group can be detected more easily than by PCA; in this way LDA computes classification functions that can be used to determine to which group each sample most likely belongs. Each function allows the computation of classification scores for each case with respect to each group (Otto, 1999; Naes *et al.*, 2002).

Cross validation (leave one out) was used as a validation method to evaluate the performance and robustness of the classification models developed. The Unscrambler software version 7.5 (CAMO, 1996) was used to develop the PCA models, while the LDA models were developed using the JMP software (SAS Institute, 2002).

## RESULTS AND DISCUSSION

The score plot of the first two principal components (PC1 and PC2) for the classification of honey samples according to their botanical origin is shown in Figure 1. Generally, a separation was observed between honey samples according to their botanical origin, however some samples did overlap. In particular, it was observed for samples labelled as *Myrtaceae* and *Scutia* spp. These honey samples were identified as Monte (bush honey) by the bee-

keepers, lacking a dominant pollen from specific plant species (e.g., 40% or more of pollen from *Scutia* spp., 20% or more of pollen from *Myrtaceae* spp.). The first three PCs accounted for more than 96% of the variation in the honey samples analysed, where PC1 explains 71%, PC2 18% and PC3 7% of the variation related to the different sources of pollen contained in the honey samples, respectively.

The eigenvectors for the first three PCs used to develop the PCA plot are shown in Figure 2. The highest eigenvector in PC1 is explained by the high percentage of pollen of *Eucalyptus*, whilst the highest eigenvector in PC2 is explained by the highest percentage of pollen of *Lotus* and *Trifolium*, whilst PC3 is explained by the highest percentage of *Trifolium*. Although no clear separation was observed among the honey samples, the PCA scores were related to different botanical origins. It is therefore possible that some of the samples were not exactly of the floral origin defined by the beekeeper, which may explain the PCA score grouping overlap. Even honeys classified as predominantly belonging to one floral origin could also be, to some extent, considered blends or mixtures of different species. The classification results obtained using LDA are shown in Table 1. Honey samples belonging to *Scutia*, *Eucalyptus* and *Lotus* were 100% correctly classified. On the other hand, honey samples identified as *Salix* and *Myrtaceae* spp. were 80 and 66% correctly classified, respectively.

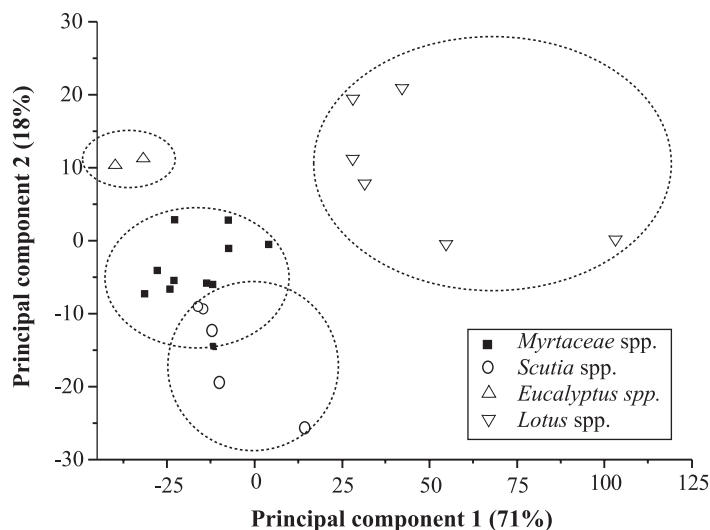


Figure 1. Principal component score plot of honey samples from 2005 based on pollen content.

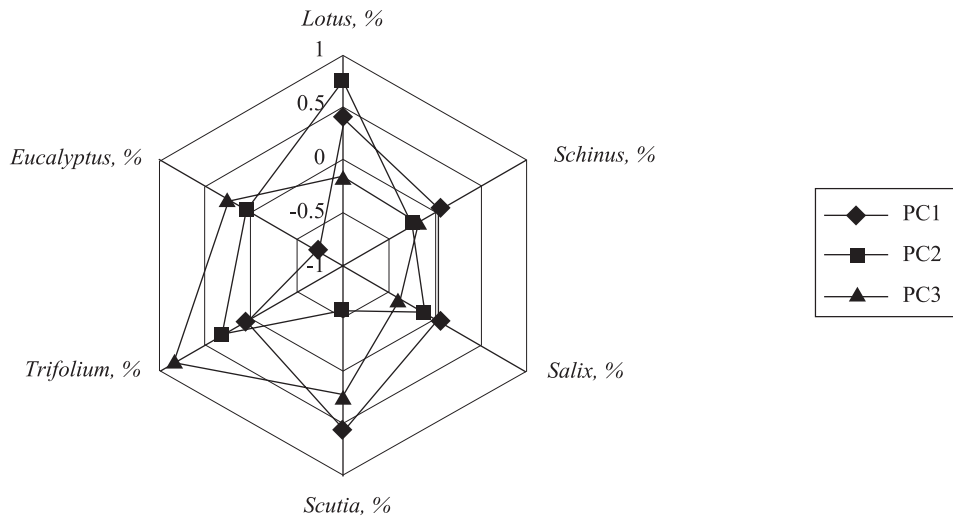


Figure 2. Eigenvectors for the first three principal components of honey samples from 2005 based on pollen content.

Table 1. Linear discriminant classification rates for honey samples (2004 and 2005) based on pollen count.

	<i>Scutia</i>	<i>Eucalyptus</i>	<i>Lotus</i>	<i>Myrtaceae</i>	<i>Salix</i>
<i>Scutia</i>	10 (100%)	0	0	0	0
<i>Eucalyptus</i>	0	10 (100%)	0	0	0
<i>Lotus</i>	0	0	12 (100%)	0	0
<i>Myrtaceae</i>	0	2	0	8 (66%)	2
<i>Salix</i>	1	0	0	0	4 (80%)

Percentages of samples correctly classified in parentheses.

In developing classification or discrimination models, it is known that when more properties (variables) are used for classification, more objects (samples) are needed to get a robust model. From the results obtained, it is shown that three of the five botanical origins analysed showed 100% correct classification. Honey samples being labelled as *Eucalyptus* and *Lotus* spp. showed that the pollen of these species is predominant in the honey samples analysed. However, in samples labelled either as *Myrtaceae* and *Salix* spp., the blend with other botanical sources makes a correct classification difficult. This is explained by the fact that *Myrtaceae* and *Salix* spp. are not single species, and have a diverse pollen source compared to the other three types analysed.

In general, supervised classification (e.g., discriminant analysis) is used to test similarity of known authentic samples. However, questions about misclassification arising from this study still need to be

addressed by using larger data sets or by incorporating authentic samples of monofloral honey. The limited number of samples in some of the honey floral categories studied in the present work led us to be cautious about extrapolating these results to other conditions. The combination of multivariate techniques with pollen count have made it possible to ascertain, on a rigorous, scientific basis, the apicultural importance of the different plant species, whereas previously this evaluation was based on empirical general field observations made by beekeepers.

## CONCLUSIONS

The results obtained in this study showed the potential of combining multivariate techniques (LDA and PCA) with pollen count to classify the botanical origin of honey samples produced in Uruguay. The limited number of samples in some of the honey floral categories studied in the present work, however, led us to be cautious about extrapolating

these results to other conditions. Further experiments need to be carried out in order to address questions about misclassification, by using larger data sets or by incorporating authentic samples of monofloral honey samples.

### ACKNOWLEDGMENTS

The authors acknowledge the technical assistance of Mr. G. Ramallo at the Apiculture Project (INIA La Estanzuela) for the honey chemical analysis and the beekeepers who provided the honey samples. The work was supported by INIA (Instituto Nacional de Investigación Agropecuaria), Uruguay.

## Clasificación del origen botánico de la miel mediante la combinación de análisis multivariado y recuento de polen

### RESUMEN

Este estudio reporta la combinación de técnicas de análisis multivariado y de polen para clasificar el origen botánico de muestras de miel provenientes de Uruguay. Muestras de miel de diversos orígenes botánicos, a saber *Eucalyptus* spp. (n = 10), *Lotus* spp. (n = 12), *Salix* spp. (n = 5), mil flores (*Myrtaceae* spp.) (n = 12) y coronilla (*Scutia buxifolia* Reissek) (n = 10) fueron analizadas usando Melisopalínología (identificación del polen). Análisis de componentes principales (APC) y de discriminantes lineales (ADL) fueron utilizados para clasificar las muestras de la miel de acuerdo a su origen botánico basado en el recuento de polen. Las muestras de miel que contenían más de un 70% de polen de *Eucalyptus*, *Lotus* y *Scutia buxifolia* fueron clasificadas correctamente en un 100% de los casos, mientras que las muestras de miel identificadas como de *Myrtaceae* spp. y *Salix* fueron clasificadas correctamente en un 80 y 66% de los casos. El uso de APC y de ADL combinado con la identificación del polen probó ser una herramienta útil para caracterizar muestras de miel de diversos orígenes.

**Palabras clave:** miel, Uruguay, componentes principales, análisis de discriminantes, análisis de polen.

### LITERATURE CITED

- Anapuma, D., K.K. Bhat, and V.K. Sapna. 2003. Sensory and physico-chemical properties of commercial samples of honey. *Food Chem.* 83:183-191.
- Anklam, E. 1998. A review of the analytical methods to determine the geographical and botanical origin of honey. *Food Chem.* 63:549-562.
- Ashurst, P.R., and M.J. Dennis. 1996. *Food authentication*. 399 p. Blackie Academics and Professionals, London, UK.
- Bertelli, D., M. Plessi, A.G. Sabatini, M. Lolli, and F. Grillenzoni 2007. Classification of Italian honeys by mid-infrared diffuse reflectance spectroscopy (DRIFTS). *Food Chem.* 101:1565-1570.
- Bogdanov, S. 1999. Honey quality and international regulatory standards: review by the International Honey Commission. *Bee World* 90:61-69.
- CAMO. 1996. The Unscrambler, version 7.5. CAMO Process AS, Oslo, Norway.
- Cordella, Ch., J.S. Militao, M-C. Clement, and D. Cabrol-Bass. 2003. Honey characterization and adulteration detection by pattern recognition on HPAEC-PAD profiles. 1. Honey floral species characterization. *J. Agric. Food Chem.* 51:3234-3242.
- Cordella, Ch., I. Moussa, A-C. Martel, N. Sbirrazzuoli, and L. Lizzani-Cuvelier. 2002. Recent developments in food characterisation and adulteration detection: technique-oriented perspective. *J. Agric. Food Chem.* 50:1751-1764.
- Cozzolino, D., and Corbella, E. 2005. The use of visible and near infrared spectroscopy to classify the floral origin of honey samples produced in Uruguay. *J. Near Infrared Spectros.* 13:63-68.
- Devillers, J., M. Morlot, M.H. Pham-Delegue, and J.C. Dore. 2004. Classification of monofloral honeys based on their quality control data. *Food Chem.* 86:305-312.

- Hermosin, I., R.M. Chicon, and M.D. Cabezudo. 2003. Free amino acid composition and botanical origin of honey. *Food Chem.* 83:263-268.
- SAS Institute. 2002. JMP Software. Version 5.01, Cary, North Caroline, USA.
- Latorre, M.J., R. Peña, S. García, and C. Herrero. 2000. Authentication of Galician (N.W. Spain) honey by multivariate techniques based on metal content data. *Analyst* 125:307-310.
- Latorre, M.J., R. Peña, C. Pita, A. Botana, S. García, and C. Herrero. 1999. Chemometric classification of honey samples according to their type. II. Metal content data. *Food Chem.* 66:263-268.
- Marini, F., A.L. Magri, F. Balestrieri, F. Fabretti, and D. Marini. 2004. Supervised pattern recognition applied to the discrimination of the floral origin of six types of Italian honey samples. *Anal. Chim. Acta* 515:117-125.
- Mateo, R., and F. Bosch-Reig. 1998. Classification of Spanish unifloral honeys by discriminant analysis of electrical conductivity, color, water content, sugars and pH. *J. Agric. Food Chem.* 46:393-400.
- Naes, T., T. Isaksson, T. Fearn, and T. Davies. 2002. A user-friendly guide to multivariate calibration and classification. 344 p. NIR Publications, Chichester, UK.
- Otto, M. 1999. *Chemometrics*. 314 p. Wiley-VCH. Weinheim, Germany.
- Serrano, S., M. Villarejo, R. Espejo, and M. Jodral. 2004. Chemical and physical parameters of Andalusian honey: classification of Citrus and Eucalyptus honeys by discriminant analysis. *Food Chem.* 87:619-625.
- Terrab, A., M.L. Escudero, M.L. Gonzalez-Miret, and F.J. Heredia. 2004. Colour characteristics of honey as influenced by pollen grain content: a multivariate study. *J. Sci. Food Agric.* 84:380-386.