Short Communication

The use of visible and near infrared spectroscopy to classify the floral origin of honey samples produced in Uruguay

E. Corbella[a] and D. Cozzolino[a,b]

[a]*Instituto Nacional de Investigación Agropecuaria, Estación Experimental INIA La Estanzuela, INIA Uruguay, Ruta 50 – km 12, Colonia, Uruguay. E-mail: corbella@inia.org.uy*

[b]*Current address: The Australian Wine Research Institute, Waite Road, Glen Osmond, PO Box 197, Adelaide, South Australia 5064. E-mail: Daniel.Cozzolino@awri.com.au*

This study reports the use of visible (vis) and near infrared (NIR) spectroscopy as a tool to classify honey samples from Uruguay, according to their floral origin. Classification models were developed using principal component analysis, discriminant partial least squares (DPLS) regression and linear discriminant analysis (LDA). Honey samples ($n = 50$) from two floral origins, namely Eucalyptus spp. and pasture, were split randomly into even calibration ($n = 25$) and validation sets ($n = 25$). Both LDA and DPLS models correctly classified, on average, more than 75% of the honey samples belonging to pasture and more than 85% of the honey samples belonging to Eucalyptus spp. These results showed that vis-NIR might be a suitable and alternative method that can easily be implemented by both the industry and retailers to classify samples according their floral origin. Vis-NIR analysis requires little sample preparation and is rapid. However, the relatively limited number of samples involved in the present work led us to be cautious in terms of extrapolating the results of this work to other floral types.

Keywords: near infrared spectroscopy, visible, classification, honey, DPLS, partial least squares discrimination, floral origin, LDA

Introduction

Determination of food authenticity is one of the most important issues in food quality control and safety.[1–3] Regulatory authorities, food processors, retailers and consumer groups are actively interested in ensuring that foods on the market are what they purport to be.[1–4] Honey is one of the most complex foodstuffs produced by nature and certainly the only sweetening agent that can be used by humans without processing.[5] In recent years, characterisation of honey by means of both chemical and sensory characteristics has received increased attention.[5–10] Quality control methods, in conjunction with multivariate statistical analysis, have been found to be able to classify honey from different geographical regions, adulteration and to describe chemical characteristics.[2,11] Traditionally, determination of the floral origin of honey is achieved by palynological analysis.[12–14] This method is based on the identification of pollen by microscopic inspection. However, the identification of the floral or botanical origins of honey is a difficult task and there is no method available today that gives unequivocal results due to variability in the amounts of pollen collected by the bees, different plant species contributing different proportions of the pollen, the amount of pollen varying from season to season, problems of falsification, pollen counting and identification and interpretation of the results.[12,14] Several chemical and physical techniques have been used to classify and identify the authenticity and botanical origin of honey produced (for example, high performance liquid chromatography and atomic absorption spectrophotometry).[5–18]

Most of the well-known applications of near infrared (NIR) spectroscopy have been involved with the development of calibrations for the quantitative prediction of chemical composition (for example, protein, fat and moisture) in both agricultural products and foods.[19] The advantages of NIR technology are not only to assess chemical composition through the analysis of the molecular bonds in the NIR region

© NIR Publications 2005, ISSN 0967-0335

(O–H, N–H, C–H), but also to build a spectrum, characteristic of the sample, which behaves as a "*fingerprint*".[1–3] The analysis of both the gross composition (moisture, pH, sugars, colour and hydroxy methyl furfuraldehyde) and the adulteration of honey using NIR spectroscopy has been examined and reported elseshwere.[20–27] The use of NIR spectroscopy combined with chemometric methods has also been reported recently to identify the botanical origin of honey produced in the European Community.[14]

The aim of this work was to investigate the potential of visible (vis) and NIR spectroscopy as a rapid and low cost technique to classify honey samples from Uruguay (South America) according to their floral origin.

Materials and methods

Samples

Honey samples ($n=50$) were obtained directly from the beekeepers and collected during the 2001 harvest from different locations across Uruguay (South America). In Uruguay, specific floral types of honey are obtained by the beekeepers pursuing a particular floral species for honey production, through controlling the foraging of their honeybees by hive location (near to one species of plant). Information about season, hive location and available floral sources were utilised by asking the beekeepers to accurately identify the botanical source of the honey samples. Honey samples were randomly split into two even groups, namely Eucalyptus spp ($n=25$) (hives located near to Eucalyptus spp plantations) and pasture (hives located near to pastures or crops containing pure legumes, grasses or mixtures) ($n=25$). The floral origin was confirmed by aroma, taste and colour characteristics in the honey analysed (Table 1).

The samples were taken from stainless steel tanks, kept in the dark, at a room temperature of 20–25°C, in plastic jars before chemical and NIR analysis. All the samples were fresh (< one month old) and non-crystallised. Information about honey sample chemical characteristics is detailed in previous reports.[22]

Near infrared spectroscopy

Samples were scanned from 400–2500 nm in reflectance mode (transflectance) at 2 nm intervals in a monochromator (NIRSystems 6500, Silver Spring, MD, USA). Spectral information was manipulated using Infrasoft International (ISI) version 3.1 software (ISI, Port Matilda, PA, USA). The samples (approximately 1 g) were placed in a camlock cell fitted with an aluminium-backed plate (50 mm diameter) and presented as a 0.2 mm thick film (Part number IH - 0355-2, NIRSystems, Silver Spring, MD, USA). Between samples, the cell was washed with detergent, rinsed with milli-Q water and dried using a tissue. Reflectance data were stored in absorbance units as the log of the reciprocal reflectance (log $1/R$) (where R: reflectance), using a ceramic disk as reference. The spectrum of each sample was the average of 32 successive scans (1050 data points).

Chemometrics

Spectra were exported from the ISI (v. 3.1) software in NSAS format to the Unscrambler software (version 7.5, CAMO ASA, Norway) for chemometric analysis. Principal component analysis (PCA) was performed before discriminant partial least squares (DPLS) and linear discriminant analysis (LDA) models were developed.[28] PCA was used to derive the first 20 principal components from the spectral data. These were used to examine the grouping of samples according to their floral origin and to visualise outliers. PCA

Table 1. Mean and standard deviation (*SD*) values of chemical composition of Eucalyptus spp. and pasture honey samples and the results of the *t* test for comparison of the two means.

	Eucalyptus spp. ($n=25$)		Pasture ($n=25$)		Student *t* test
	Mean	*SD*	Mean	*SD*	
M ($g\,kg^{-1}$)	177.0	9.9	175.1	16.5	NS
pH	3.1	0.47	3.7	0.65	NS
EC ($mS\,cm^{-1}$)	0.42	0.20	0.60	0.42	*
C (mm Pfund)	48.7	26.1	50.6	36.5	*
HMF ($mg\,kg^{-1}$)	7.8	5.2	17.9	4.6	*

M: moisture
C: colour
EC: electric conductivity
HMF: hydroxy methyl furfurladehyde
NS: no significant differences
*Statistical different $P<0.05$

was performed on the raw spectra and after pre-processing using both first and second derivatives to reduce baseline variation and enhance the spectral features.[29] The second derivative was performed using Savitzky–Golay derivation and smoothing (10 point and 2nd order filtering operation).

Discriminant models were developed using DPLS regression with internal cross-validation (four groups).[28] The optimum number of factors (PLS terms) were selected on the basis of the predicted residual sum of squares (PRESS function), which should be minimised.[28] Discrimination models were performed using the dummy regression technique described elsewhere.[20,28,30] In this method, a calibration matrix is developed using dummy variables by assigning an arbitrary number if the sample belongs to a particular group or if it does not. Calibration models were developed by regressing the spectral data on the assigned reference value, namely dummy variable (1 = Eucalyptus; 2 = pasture). A honey sample was classified as Eucalyptus floral origin if its value was below 1.5 and classified as pasture if the value was above 1.5. The samples were divided at random into two sets, each comprising 25 honey samples, namely calibration ($n = 12$ pasture; $n = 13$ Eucalyptus spp.) and validation sets ($n = 13$ pasture; $n = 12$ Eucalyptus spp). Linear discriminant analysis (LDA), like DPLS, is a supervised classification technique where the number of categories and the samples that belong to each category are previously defined.[20,30] The method produces a number of orthogonal linear discriminant functions, equal to the number of categories minus one, that allow the samples to be classified into one or other category. It is based on the normal distribution and the assumption that the covariance matrices of the two groups are identical.[28] The LDA models were validated using leave-one-out cross-validation. LDA was carried out using *JMP* software (version 5.01, SAS Institute Inc., Cary, NC, USA) on the PCA sample scores on components 1 to 3 which gave the highest level of separation in the PCA models developed. Both LDA and DPLS regression models were developed using three spectral regions: vis and short NIR wavelengths 400–1100 nm, NIR region 1100–2500 nm and vis + NIR region 400–2500 nm (whole spectrum).

Results and discussion

Chemical analysis

Table 1 shows the chemical composition for honey samples analysed. Statistically significant differences were found between hydroxy methyl furfuraldehyde (HMF), colour and electric conductivity (EC) values between the two floral origins. No statistical differences were observed between moisture (M) and pH. The EC is generally accepted and used as a method for the determination of origin and botanical discrimination between honey samples. For example, blossom honeys and mixtures of blossom and honeydew honeys should have EC less than 0.8 mS cm^{-1}; however, an extreme variation in EC related with Eucalyptus honeys exists.[32,33] Therefore, the differences observed in colour, HMF and EC, in conjunction with the subjective assessment made by the beekeepers, were used as indicators of the different floral origins in the honey samples analysed.

Vis-NIR characterisation

Figure 1 shows the vis-NIR mean spectrum of honey samples identified by their floral origin. Overall, the honey samples show absorption bands in the NIR region related to1204 nm with C–H stretch second overtone, at 1468 nm with O–H stretch second overtone (water) and at 1940 nm with O–H stretch first overtone (water). The absorption band at 2102 nm was assigned to C–H deformation and combination or C–O stretch combination overtones and at 2276 nm with C–H combination, C–C stretch tones or C–O stretch–combination overtones and were both assigned to sugars in honey.[19,22,31]

PCA analysis

PCA was performed on the spectra (400 to 2500 nm) to examine qualitative differences between samples using the raw, first and second derivative, respectively. Figure 2 plots the PCA scores (PC1 vs PC2) for honey samples in the vis-NIR using the raw spectra. The PCA plot showed that different spectral attributes among samples might be associated with characteristics of the honey samples (for example, chemical composition or floral origin). It is well known that honey is rarely derived from a single botanical or floral source, since the unifloral character of honey is very hard to find in nature.[12,13] Generally, the term unifloral is related to a honey containing at least 45% of the total pollen content from one source.[12,13]

The first three PCs explain 98% of the total variation in the raw spectra. PC1 explains 90% of the total variation in the honey samples according to their floral origin. The highest eigenvectors in PC1 were observed (Figure 3) in the NIR

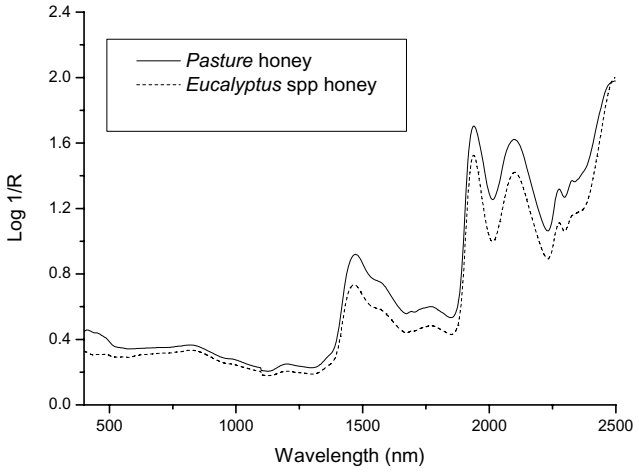

Figure 1. Visible and near infrared spectra in transflectance mode of Eucalyptus spp and pasture honey samples.

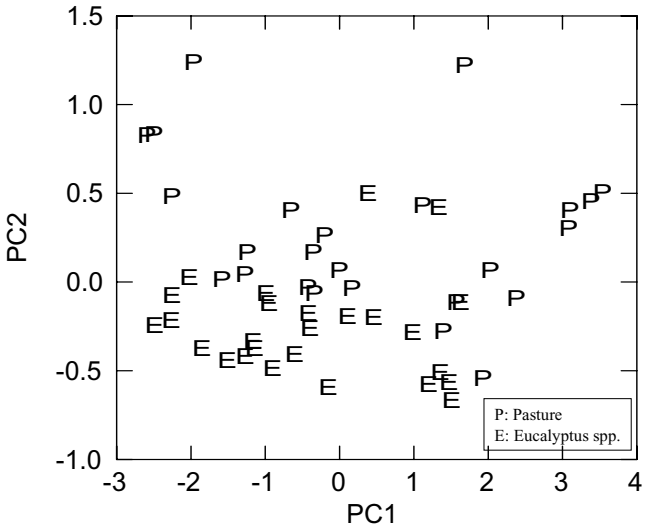

Figure 2. Principal component score plots of honey according to floral origin (vis and NIR raw spectra).

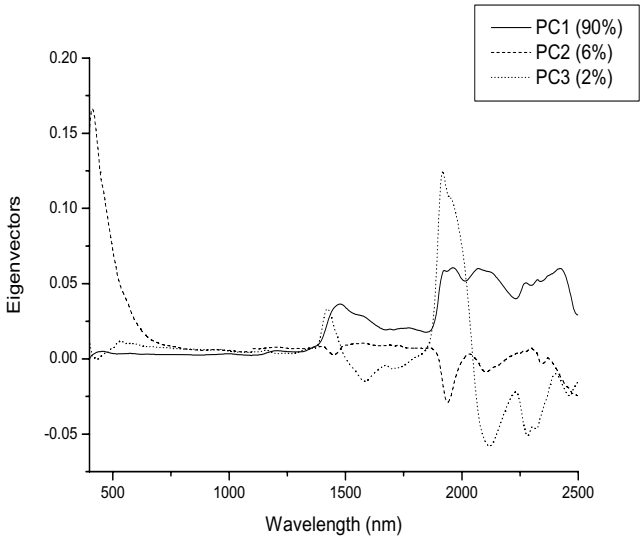

Figure 3. Eigenvectors for the first principal components for the discrimination of honey samples. according to floral origin.

region at 1488 nm related to O–H stretch second overtone (water) and around 1940 nm with O–H stretch first overtone (water). Absorption bands around 2070 nm assigned to C–H deformation and combination or C–O stretch combination overtones and at 2270 nm assigned to either C–H combination, C–C stretch tones or C–O stretch–combination overtones, are mainly associated with sugars in honey.[19,24,31] Both PC2 and PC3 explain 6 and 2% of the variation in the raw spectra, respectively. The highest eigenvectors observed in both PC2 and PC3 were mainly related to water (O–H absorption bands). Absorption bands around 1500 nm assigned to C–H stretch overtones assigned to sugars (fructose, glucose, maltose and sucrose) were also observed in PC3.[19,24] It should be noted that no eigenvectors were observed in the vis region, despite statistical significant differences being found between colour and HMF values in the set of samples analysed. Similar results were reported by other authors for the analysis of honey samples.[14,24,25]

Honey classification

The results of the classification obtained by the DPLS models are summarised in Table 2. The Eucalyptus honey

Table 2. Discriminant PLS prediction of floral origin in honey samples (validation set, $n=25$).

Data type	Wavelength range (nm)	%CC as Eucalyptus spp. ($n=12$)	%CC as pasture ($n=13$)
Raw	400–1100	85	92
	1100–2500	85	92
	400–2500	85	92
First deriv.	400–1100	85	83
	1100–2500	92	75
	400–2500	92	75
Second deriv.	**400–1100**	**100**	**92**
	1100–2500	**100**	66
	400–2500	**100**	**100**

%CC: percentage of those correctly classified
n: number of samples of each floral origin in the validation set
note: figures in bold show the best classification

Table 3. LDA classification of honey samples according to floral origin (vis + NIR).

	Eucalyptus spp (predicted)	Pasture (predicted)
Eucalyptus spp (LDA)	12 (92%)	1 (8%)
Pasture (LDA)	3 (25%)	9 (75%)
Total	15	10

samples were classified correctly (>85%) in most of the cases, while pasture honey samples were not always correctly classified (between 70 to 100%). Wavelength regions between 400 to 1100 nm and between 400 to 2500 nm were the most suitable for correctly classifying the honey samples according to their floral origin. Table 3 shows the classification of honey samples using the LDA technique. The Eucalyptus honeys were classified correctly (>90%) while the pasture honey samples only achieved a correct classification of 75% using the LDA technique. The most useful LDA discrimination model was obtained using the vis and NIR region. In all the discriminant models evaluated, the inclusion of spectral information from the vis region was essential to obtain the most accurate honey classification. Reports by other authors suggested that pollen chromatic characteristics tend to determine and influence the honey colour properties.[34] The former suggested that the inclusion of the vis region might be the reason for the characteristics of the floral type yielding better discrimination results. Overall, the ability of the models to discriminate between or identify the floral origin of honey is based on the information from the responses of the whole matrix to the NIR radiation and not with a specific chemical or physical characteristic (for example sugars, amino acids, HMF). The NIR spectrum contains information about the entire composition (chemical and physical characteristics) of the honey matrix under analysis, yielding structural information that constitutes the fingerprint of the sample.[1–3] In general, supervised classification (for example, DPLS and LDA techniques) is used to test similar known authentic samples. This study has shown that the classification technique applied was able to extract useful information from the NIR spectra to differentiate between the floral origins of honey. Other authors reported that NIR spectroscopy is a very practical method for classifying honey samples because the data analysis system is so complex that it would be almost impossible to cheat.[14] Considering the problems associated with the conventional methods used to discriminate between different floral origins in honey, this study demonstrated the potential of NIR spectroscopy as a rapid tool for use in the honey industry. However, the limited number of samples and floral origins used in the present work lead us to be cautious in terms of extrapolating the results obtained to other conditions.

Conclusions

The results obtained in this study showed the potential of vis-NIR spectroscopy to classify honey samples according to their floral origin. In this study, the vis and NIR regions (400–2500 nm) were the most suitable for obtaining the correct honey classification. The application of discriminant techniques (DPLS and LDA) have shown excellent potential for discriminating between the floral origins of honey based on NIR spectra. The work reported here is only a feasibility study and further studies using considerably more samples (floral origin) are required before its value may be validated and adopted in routine analysis. The effect of different floral and geographical origins need to be investigated in order to provide a robust model to discriminate between floral origins using NIR spectroscopy. Analysis of pollen, to authenticate the sample origin, should also be incorporated into future studies.

Acknowledgments

The authors acknowledge the technical assistance of Mr G. Ramallo and M. Maidana at the Apiculture Project (INIA La Estanzuela) for the honey chemical analysis and the beekeepers that provided the honey samples. The work was supported by INIA–Uruguay.

References

1. G. Downey, *J. Near Infrared Spectrosc*. **4,** 47 (1996).
2. P.R. Arhurst and M.J. Dennis, *Food Authentication*. Chapman-Hall, London, UK (1996).
3. G. Downey, *Trends Anal. Chem.* **17,** 418 (1998).
4. G. Downey, *NIR news.* **11,** 8 (2000).
5. M.T. Iglesias, C. de Lorenzo, M.C. Polo, P.J. Martín-Álvarez and E. Pueyo, *J. Agric. Food Chem.* **52,** 84 (2004).
6. A. Pereyra-Gonzalez, L. Burin and M.P. Buera, *Food Res. Int.* **32,** 185 (1999).
7. I. Hermosin, R.M. Chicon and M.D. Cabezudo, *Food Chem.* **83,** 263 (2003).
8. A. Terrab, A.G. Gonzalez, M.J. Diez and F.J. Heredia, *J. Sci. Food Agric*. **83,** 637 (2004).
9. M.J. Latorre, R. Peña, S. García and C. Herrero, *The Analyst* **125,** 307 (2000).
10. J. Devillers, M. Morlot, M.H. Pham-Delegue and J.C. Dore, *Food Chem.* **86,** 305 (2004).
11. Ch. Cordella, J.S. Militao, M.C. Clement and D. Cabrol-Bass, *J. Agric. Food Chem.* **51,** 3234 (2003).
12. E. Anklam, *Food Chem.* **63,** 549 (1998).
13. E. Anklam and B. Radovic, *Am. Lab*. 62 (2001).
14. A.M.C. Davies, B. Radovic, T. Fearn and E. Anklam, *J. Near Infrared Spectrosc.* **10,** 121 (2002).
15. S. Sivakesava, and J. Irudayaraj, *J. Sci. Food Agric.* **81,** 683 (2001).

16. S. Sivakesava and J. Irudayaraj, *Int. J. Food Sci. Tech.* **37,** 351 (2002).
17. M.M. Paradkar and J. Irudayaraj, *Food Chem.* **76,** 231 (2001).
18. A.C. Soria, M. González, C. de Lorenzo, I. Martínez-Castro and J. Sanz, *Food Chem.* **85,** 121 (2004).
19. B.G. Osborne, T. Fearn and P.H. Hindle, *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*, 2nd Edn. Longman Scientific and Technical, Harlow, Essex, UK (1993).
20. D. Kelly, G. Downey and V. Fouratier, *J. Agric. Food Chem.* **52,** 33 (2003).
21. M. García-Álvarez, J.F. Huidobro, M. Hermida and J.L. Rodríguez Otero, *J. Agric. Food Chem.* **48,** 5154 (2000).
22. D. Cozzolino and E. Corbella, *J. Apic. Res.* **42,**16 (2003).
23. J. Ha, M. Koo and H. Ok, *J. Near Infrared Spectros.* **6,** A367 (1998).
24. P.Y. Qiu, H.B. Ding, Y.K. Tang and R.J. Xu, *J. Agric. Food Chem.* **47,** 2760 (1999).
25. G. Downey, V. Fouratier and D. Kelly, *J. Near Infrared Spectros.* **11,** 447 (2003).
26. L. Dvash, O. Afik, S. Shafir, A. Schaffer, Y. Yeselson, A. Dag and S. Landau, *J. Agric. Food Chem*. **50,** 5283 (2002).
27. M. García-Álvarez, S. Ceresuela, J.F. Huidobro, M. Hermida and J.L. Rodriguez Otero, *J. Agric. Food Chem.* **50,** 419 (2002).
28. T. Naes, T. Isaksson, T. Fearn and T. Davies, *User-friendly guide to multivariate calibration and classification*. NIR Publications, Chichester, UK. (2002).
29. W.R. Hruschka, in *Handbook of Near-Infrared Analysis*, Ed by D.A. Burns and E.W. Ciurczak. Marcel Dekker Inc. New York, USA, p. 365 (1992).
30. M. Otto, *Chemometrics*, Wiley-VCH, Weinheim, Germany (1999).
31. I. Murray, in *Proceedings of the International NIR/NIT Conference,* Ed by J. Hollo, K.J. Kaffka and J.L. Gonczy. Akademiai Kiado, Budapest, Hungary, p. 13 (1986).
32. S. Bogdanov, P. Martin and C. Lullman, *Apidologie* 1 (1997).
33. S. Bogdanov, *Bee World*. **90,** 61 (1999).
34. A. Terrab, M.L. Escudero, M.L. Gonzalez-Miret and F.J. Heredia, *J. Sci. Food Agric*. **84,** 380 (2004).

Received: 18 June 2004
Revised: 4 April 2005
Accepted: 7 April 2005
Web Publication: 29 July 2005