JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Honey Characterization and Adulteration Detection by Pattern Recognition Applied on HPAEC-PAD Profiles. 1. Honey Floral Species Characterization

CHRISTOPHE B. Y. CORDELLA,*,†,§ JULIO S. L. T. MILITÃO,# MARIE-CLAUDE CLÉMENT,† AND DANIEL CABROL-BASS§

Agence Française de Sécurité Sanitaire des Aliments, Laboratoire d'Etudes et de Recherches sur les Petits Ruminants et les Abeilles (Unité Abeille du LPPRA-Sophia Antipolis), 105 route des Chappes, F-06902 Sophia-Antipolis Cedex, France; CNPq and Universidade Federal de Rondonia (UNIR), Km 8 Br 364, 78.900 Porto Velho, Rondonia, Brazil; and Laboratoire Arômes-Synthèses-Interactions, Université de Nice Sophia-Antipolis, 28 avenue Valrose, F-06108 Nice Cedex 2, France

An improved COFRAC (COmité FRançais d'ACréditation) method for the analysis and evaluation of the quality of honeys by high-performance anion-exchange chromatography of sugar profiles is proposed. With this method, both minor and major sugars are simultaneously analyzed and the technique is integrated in a new chemometric approach, which uses the entire chromatographic sugars profile of each analyzed sample to characterize honey floral species. Sixty-eight authentic honey samples (6 varieties) were analyzed by high-performance anion-exchange chromatography−pulsed amperometric detection. A new algorithm was developed to create automatically the corresponding normalized data matrix, ready-to-use in various chemometric procedures. This algorithm transforms the analytical profiles to produce the corresponding calibrated table of the surfaces or intensities according to retention times of peaks. The possibility of taking into account unknown peaks (those for which no standards are available) allows the maximum chemical information provided by the chromatograms to be retained. The parallel application of principal component analysis (PCA)/linear discriminant analysis (LDA) and artificial neural networks (ANN) shows a high capability in the classification of the analyzed samples (LDA, 93%; ANN, 100%) and a very good discrimination of honey groups. This work is the starting point of the elaboration of a new system designed for the automatic pattern recognition of food samples (first application on honey samples) from chromatographic analyses for food characterization and adulteration detection.

KEYWORDS: Honey; adulteration; food characterization; sugar profiles; pattern recognition; chemometrics; PCA; LDA; artificial neural networks

INTRODUCTION

Honey is a natural product that is produced throughout the world. It presents an ecological image of natural wholesomeness and reassurance in the minds of consumers, even though sporadic infant botulism cases have appeared from time to time in some European countries such as Italy, Spain, Norway, and Germany over a period of 23 years (*1*). Sugars and water represent the main chemical constituents of honey (>95%), whereas proteins, flavors and aromas, pigments, vitamins, free amino acids, and numerous volatile compounds constitute the minor components. This small percentage of the overall composition is mainly responsible for honey's organoleptic and nutritional properties. Many studies have been reported on the chemical constituents of honeys such as sugars (*2*−*8*), flavor and aroma compounds (*9*, *10*), terpenoids, norisoprenoids (*11*, *12*), flavanoids (*13*−*15*), 5-hydroxymethyl-2-furaldehyde (HMF) (*16*), phenolic compounds (*17*), aliphatic compounds, organic acids (*18*−*21*), and amino acids (*22*−*29*).

Specific approaches have been developed for characterizing food products, particularly for honeys, and for detecting their adulteration (addition of cane or beet sugars and/or sugars obtained from starch hydrolysis) (*30*, *31*). The literature shows that numerous authors have tried to characterize honey samples by peak integration from measurements of parameters such as sugars, organic acids, volatile compounds, and flavonoids (*32*, *36*), but few studies have used global fingerprints, such as chromatograms, in a pattern recognition procedure to characterize honey samples. To the best of our knowledge, no previous study about the use of the full sugar profiles [by high-

* Author to whom correspondence should be addressed: c.cordella@afssa.fr.
† Agence Française de Sécurité Sanitaire des Aliments.
§ Université de Nice Sophia-Antipolis.
# CNPq and Universidade de Federal de Rondonia.

10.1021/jf021100m CCC: $25.00
Published on Web 04/23/2003

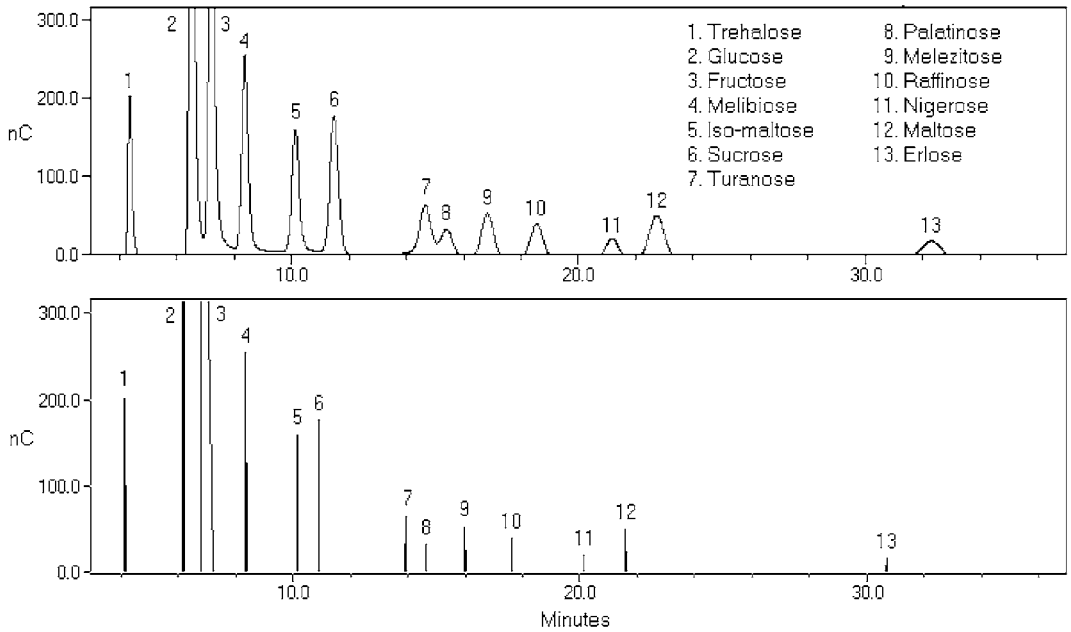

Figure 1. (Top) Calibration standard chromatogram obtained by diluting 200 μL of standard solution in 10 mL of UHQ water. (Bottom) Histogram plot of the transformed calibration standard chromatogram after the calibration process. The same treatment is applied to sample files.

performance anion-exchange chromatography−pulsed amperometric detection (HPAEC-PAD)] of honeys has been reported for floral origin characterization and adulteration detection.

The aim of this work is the development of an automated tool based on sugar analysis, using pattern recognition techniques such as artificial neural network (ANN) and classical chemometric tools [principal component analysis (PCA) or linear discriminant analysis (LDA)] to evaluate the quality of the honeys and to detect their possible adulteration by exogenous sugars. This paper presents the first steps in this study: the establishment of the chromatographic technique, the manipulation of the chromatographic data, and the comparison of performance of the chemometric tools with respect to characterization of the honey by floral species.

MATERIALS AND METHODS

Samples. Samples of lavender, robinia (false acacia), fir, rosemary, chestnut, thyme and TTF (multifloral) honeys were obtained from French beekeepers during the years 2000 and 2001. The samples were collected from different French geographic regions according to their floral species. The botanical origin of each sample was certified by quantitative pollen analysis according to the procedure of Louveaux et al. (*37*) and confirmed by sensory analysis. Syrup samples were obtained from French industrial suppliers (Dorsman S.AR.L.; Ickovich S.A.).

Preparation. The samples were prepared following a modified COFRAC (French Accreditation Committee) procedure (program 118). One hundred milligrams of honey was weighed, diluted, and adjusted to 100 mL in a volumetric flask with ultrahigh-quality (UHQ) water (18.2 mΩ). After homogenization, 1 mL of the honey solution was diluted with UHQ water to 10 mL in a volumetric flask.

To proceed to the automated calibration of the sample chromatograms (see Mathematical Pretreatment of the Chromatograms), calibration standards were also prepared and analyzed before each sample series of honeys. Three calibration standards were prepared at three concentration levels for 13 honey sugars (see Anion-Exchange Chromatography). As well, these calibration standards provide reference chromatograms of the main sugars of honeys.

Anion-Exchange Chromatography. Samples were analyzed by HPAEC on a Dionex 500 system (Dionex Corp.) supplied with a pulsed amperometric detector (PAD). The system was equipped with a CarboPac column (packed silica appropriate for mono-, di-, tri-, and oligosaccharide analysis). Sodium hydroxide solution (NaOH, 250 mM in water) was used as the eluant. Analyses were performed under isocratic mode (% water/% NaOH = 48:52). Flow rate was set to 0.6 mL/min. To minimize carbonate formation in the system, which leads to a dramatic reduction of the retention times, a small amount of $Ba(CH_3COO)_2$ (4 mM) was added to the alkaline eluant. Cataldi et al. (*38*) have recently shown that this practice inhibits the progressive occupancy of the active sites of the column. Classical PAD was adopted (with a gold working electrode) as the detection mode. Current was measured and integrated with respect to time to give a net faradic charge (q) for the detection cycle. By this method, the response is measured in coulombs (*39*). All experiments were conducted at room temperature. The calibration was performed with three standard solutions obtained from the dilution of 13 standard sugars in powder form (Aldrich S.A.). These standard solutions were made in order to represent the natural sugar proportionality of honeys in trehalose [α-D-Glu$_p$-(1→1)-β-D-Glu$_p$], glucose (β-D-Glu$_p$), fructose (β-D-Fru$_p$), melibiose [α-D-Gal$_p$-(1→6)-D-(1→1)-D-Glu$_p$], isomaltose [α-D-Glu$_p$-(1→6)-D-Glu$_p$], sucrose [α-D-Glu$_p$-(1→4)-D-Fru$_f$), turanose [α-D-Glu$_p$-(1→3)-D-Fru$_f$], palatinose [α-D-Glu$_p$-(1→6)-D-Fru$_f$], melezitose [α-D-Glu$_p$-(1→2)-β-D-Fru$_f$-(3→1)-α-D-Glu$_p$], raffinose (Fru-Glu-Gal), nigerose [α-D-Glu$_p$-(1→3)-D-Glu$_p$], maltose [α-D-Glu$_p$-(1→4)-D-Glu$_p$], and erlose [α-D-Glu$_p$-(1→4)-β-D-Glu$_f$-(1→2)-β-D-Fru$_f$]. The evaluation of the sugar content of honey samples was obtained from calibration curves of each sugar contained in the standard solutions (not shown here). With these chromatographic conditions, the last sugar (erlose) is detected after ~32 min (cf. **Figure 1**), and the analysis is ended at 35 min.

Mathematical Pretreatment of Chromatograms. Because injection causes some perturbation at the beginning of each chromatogram, the time period from 0 to 4 min was removed from all chromatograms. Despite these precautions, the results obtained showed shifts in the retention times (<4%) resulting in some classification difficulties. To overcome this problem, we developed an algorithm allowing the calibration of all chromatograms on the basis of the utilization of the calibration standards.

Indeed, two peaks (at least) are commonly present in an HPAEC-PAD profile of honey: fructose and sucrose (cf. **Figure 1**).

Data Matrix. The initial chromatographic files are exported under ASCII format from the Dionex software and considered as [2c, 2400r] vectors (time/s × intensity/nC). As the time column is the same for all of the samples, this column was not used for the data matrix construction. A sample data matrix is defined as Cx = {M_{sugars} × $N_{samples}$), where N represents the number of honey samples arranged by row and M represents the number of sugars selected in the

chromatograms arranged by column (cf. Automated Calibration of the Chromatograms). In this work, $N = 68$ samples consisting of {robinia, 7; lavender, 11; rosemary, 10; TTF-multifloral, 10; chestnut, 16; and fir, 14} and $M = 13$ sugars (cf. **Figure 1**). However, M can be extended as described in ref *40* even if the sugar peak in the chromatogram is not identified and labeled during standard calibration.

Automated Calibration of the Chromatograms. This procedure was implemented to remove potential shifts in the retention times observed during the analyses due to a possible carbonation of the sodium hydroxide or the aging of the chromatographic system. The program can automatically identify at least 13 honey sugars and redraw the chromatogram by positioning the sugars in their relative locations. The procedure transforms each signal proportionally to the original peak intensity or area (cf. **Figure 1**). In a group of authentic honey chromatogram files, the software performs the calibration and creates a data matrix where the intensity (or the area) of each sugar's peak is recorded. In the case of the absence of any of the 13 sugars, a value of 0 is assigned. This data matrix was used to perform a comparison between multivariate treatments and an artificial neural network (ANN) approach. Full details about this calibration procedure are given in ref *40*.

Multivariate Procedures (PCA and LDA). All treatments conducted in this study were performed with Statistica v 6.0 (from StatSoft, Paris, France).

The use of multivariate methods such as PCA and LDA allows the identification of the most important directions of variability in a multivariate data matrix, and the results are presented as two-dimensional plots.

PCA transforms the original variables into new axes, called principal components (PCs), which are orthogonal, in such a way that the data presented in those axes are uncorrelated with each other; PCA expresses as much as possible the total variation of the data in only a few principal components and in decreasing order in terms of the amount of the variation. Score plots represent the projections of the objects (samples) in the planes defined by the PCs, whereas loading plots represent the projections of the original variables in the same planes. Score and loading plots can be represented separately or in the same drawing. Objects that are projected close to each other in the score plots have similar characteristics, and, for instance, samples to the right in the score plot have high values for variables placed to the right in the loading plot. The same holds for samples placed in other locations of the graph. The variables that are projected close to each other in the loading plots are positively correlated, whereas variables lying opposite to each other tend to have a negative correlation. The more a variable is away from the axis origin, the better its contribution can be considered as a major contribution in the statistic model generated by the PC analysis.

LDA is used to determine which variables discriminate between two or more naturally occurring groups. This mathematical procedure maximizes the variance between groups and minimizes the variance within each group in such a way that outsiders can be detected more easily than by PCA. Another major purpose to which LDA is applied is the issue of predictive classification of cases.

LDA automatically computes classification functions that can be used to determine to which group each case most likely belongs. There are as many classification functions as there are groups (*41*). Each function allows the computation of classification scores for each case with respect to each group. LDA can be usefully completed by a canonical analysis to obtain the canonical scores plots that provide a visual organization of the sample scores and facilitate the interpretation of the results.

Artificial Neural Networks. Neural networks are nonlinear data processing systems capable of predicting new observations (on specific variables) from other observations (on the same or other variables). A neural network makes use of a dataset (training set) to adjust itself on the salient features of the set and develop a predictive model. Basically, each neuron receives signals from a large number of other neurons and processes them by weighted summation and nonlinear transformation to yield a signal output sent to other neurons as input. Therefore, they can be used where some information is known and some unknown information (*42*, *43*) has to be inferred. Neurons can be interconnected following various architectures, but the most commonly used is made of several layers. In this work, the input layer of the networks receives the rows of data matrix previously described, and the output layer contains the data related to the samples' groups.

The first step is to design a specific network architecture (which includes a specific number of "layers", each consisting of a certain number of "neurons"). Software is available that applies artificial intelligence techniques to aid in finding "the best" network architecture (*44*) in such a way that the size of the network matches the nature of the investigated phenomenon. In this work, different types of feedforward networks were applied on the dataset: multilayer perceptron (MLP), radial basis function network (RBF), and linear neural network (LNN). The complete theory of the ANNs will not be developed here and can be found in the literature (*45*, *46*) as well as its many applications (*47*−*52*).

RESULTS AND DISCUSSION

Performance of the System. Before the chromatograms were exported, the performance of the chromatographic apparatus was investigated. The aim was to evaluate both the detection limit (micrograms per liter) and the quantification limit (micrograms per liter) for the full set of identified sugars. This evaluation was carried out following the classical chromatographic methodology, consisting of the calculation of the signal to noise ratio (R = S/N). This ratio is determined for each peak (sugar) of the standard calibration chromatogram (13 sugars)< and the detection limit (DL) was calculated from the formula

$$\mathrm{DL}\ (\mu g/\mathrm{L}) = 3N[\mathrm{sugar}]/S \qquad (1)$$

where N = noise of the acquisition = background (nC), S = signal (nC), [sugar] = concentration for a given sugar [μg/L].

The quantification limit (QL) was calculated from the DL by the following formula:

$$\mathrm{QL}\ (\mu g/\mathrm{L}) = 5 \times \mathrm{DL} \qquad (2)$$

Good performances (data are not presented here but are available on request to the authors) were obtained by our system Dionex 500 for analyzing honey sugars. In particular, these results show that glucose has the best S/N ratio ($R = 2002.2$) and erlose the worst ($R = 73.9$). The results allow the identification of those sugars that are most easily detectable by the system and with the best accuracy. The QL is the best for glucose (26.3 μg/L) and the worst for erlose (189.5 μg/L). The values for other sugars are distributed between these two limits.

Multivariate Methods. The first step of the analysis was the application of PCA and LDA to data containing all samples (robinia, lavender, rosemary, chestnut, fir, and TTF-multifloral honeys). PCA failed to obtain two or three PCs with variance large enough to successfully separate the sample groups. Furthermore, the discriminant analysis (forward method) gave a classification rate of <90%. This result, presented in **Figure 2**, is due to similarities between rosemary and TTF-multifloral honeys (the squared Mahalanobis distance between these two groups was 3.288, whereas it was >12 between each other). Terrab et al. (*36*) obtained similar results on Moroccan honeys by GC-MS analysis of sugars, and they observed a poor discrimination between honeydew honeys and nectar honeys. The application of PCA gave worse results (cumulative variance $\approx$ 50%), but the approach employed by these authors concerning the extraction of information from chromatograms was not comparable with ours.

Table 1 shows the results obtained by stepwise discriminant analysis applied on the complete data matrix (six varieties). It can be seen that the poorest classification was obtained for TTF-multifloral honeys (60%), which needs to be treated separately

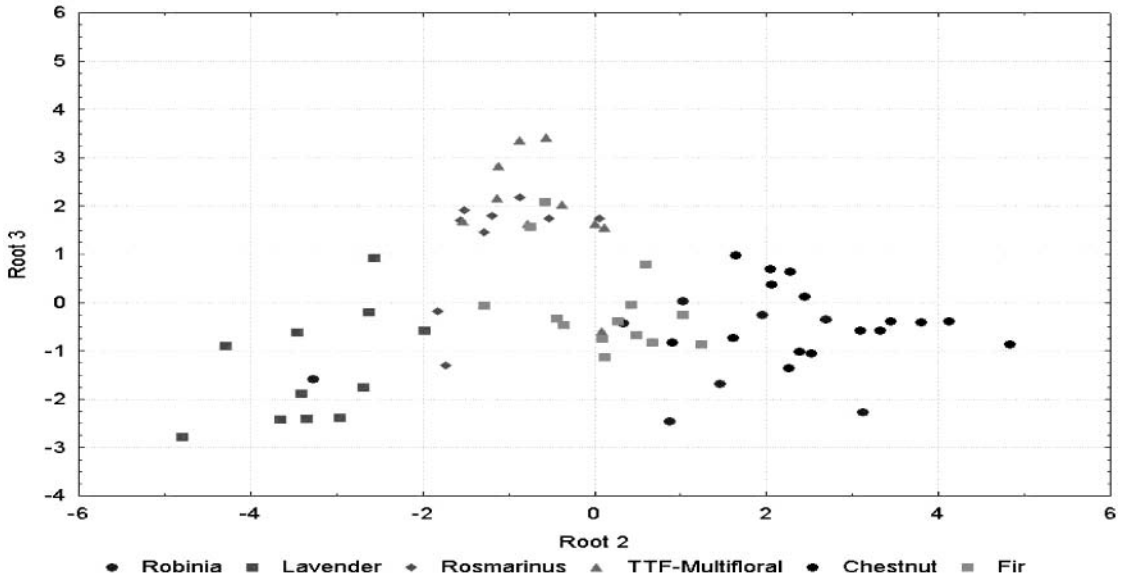

Figure 2. Plot of discriminant scores after linear discriminant analysis of 68 French honeys.

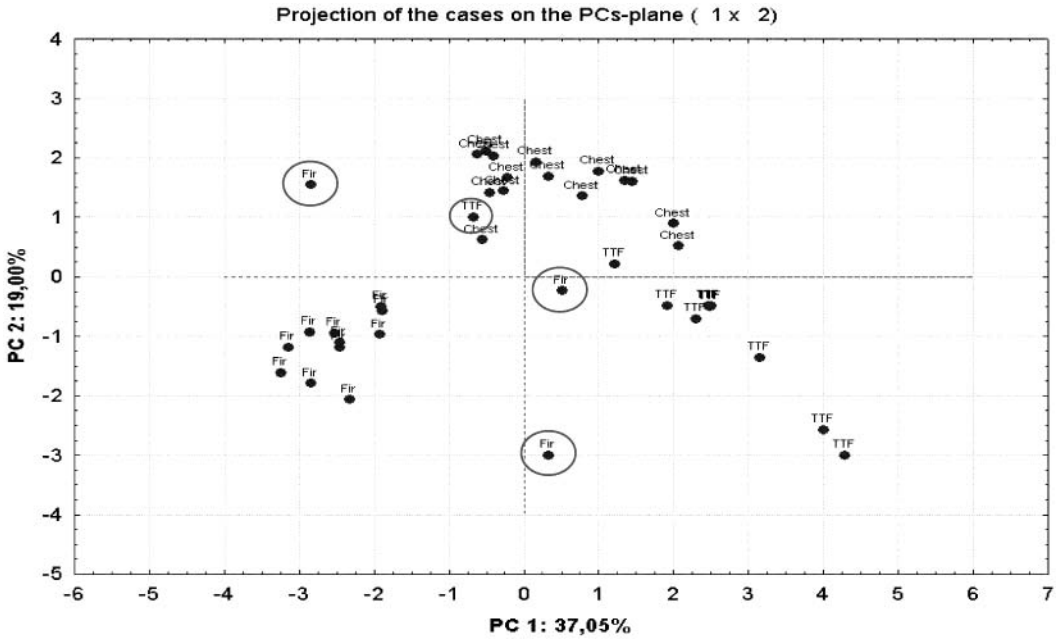

Figure 3. Plot of first and third PC scores vectors for the classification of honeys according to their floral type.

Table 1. Classification Matrix for Six Varieties of (Clear and Dark) Honeys Obtained by LDA

	% correct	robinia	lavender	rosemary	TTF	chestnut	fir
robinia	87.50	7	1	0	0	0	0
lavender	90.91	0	10	1	0	0	0
rosemary	88.89	0	1	8	0	0	0
TTF	60.00	1	0	3	6	0	0
chestnut	100.00	0	0	0	0	16	0
fir	100.00	0	0	0	0	0	14
total	89.71	8	12	12	6	16	14

from others. It can also be seen that the fir and chestnut honey groups were perfectly classified by the LDA procedure (100%). Therefore, to improve the results, we chose to perform PCA, LDA, and ANN separately on clear honeys (robinia, lavender, and rosemary) and on dark honeys (fir, chestnut, and TTF-multifloral). These two types of honeys are easily distinguished by visual inspection using the Pfund scale (*53*).

Table 2. Results of Principal Component Analysis

PC	eigenvalue	% total variance	cumulative variance (%)
1	4.446329	37.0527	37.0527
2	2.280527	19.0043	56.0571
3	1.690544	14.0878	70.1450
4	1.142643	9.5220	79.6670
5	0.702186	5.8515	85.5186

Applied separately on clear honeys, PCA gave (plot not shown here) a less effective result. The four first PCs (with eigenvalues > 1) explain only 73% of the total variance. This result needs no further comment.

Table 2 shows the eigenvalues obtained by PCA applied to dark honeys. The first four PCs explain ∼80% of the total variance, and the three honey groups of this type are relatively well distinguished by their sugar content. In other words, 80% of the total variance of the 13 variables considered could be condensed into four new variables (PCs). The score plot (**Figure**

Table 3. Classification Matrix for Clear and Dark Honeys Obtained by LDA

	% correct	robinia	lavender	rosemary
robinia	87.50	7	1	0
lavender	90.90	0	10	1
rosemary	100.00	0	0	9
total	92.86	7	11	10

Table 4. Classification Matrix for Clear and Dark Honeys Obtained by LDA

	% correct	TTF	chestnut	fir
TTF	100.00	10	0	0
chestnut	100.00	0	16	0
fir	100.00	0	0	14
total	100.00	10	16	14

Table 5. Classification Functions for Clear and Dark Honeys Obtained by LDA

variable	robinia	lavender	rosemary
fructose	–0.619	–7.496	–8.692
maltose	–5.824	–3.720	–8.097
palatinose	26.725	20.207	26.412
raffinose	–477.088	–430.819	–495.562
sucrose	1.138	2.315	2.541
erlose	–2.089	–6.552	–11.584
constant	–59.377	–50.702	–64.658

variable	TTF	chestnut	fir
trehalose	–3.357	–1.081	15.444
fructose	1.295	6.888	–5.884
palatinose	–2.087	4.110	–0.309
sucrose	–6.016	7.346	4.992
raffinose	0.124	–3.797	4.079
melezitose	–0.971	–1.978	6.019
maltose	–1.205	–9.716	–1.280
glucose	1.415	0.169	–1.874
melibiose	0.641	1.683	–1.509
nigerose	–0.936	1.849	–0.448
constant	–5.274	–9.583	–23.278

3) shows the location of the objects in the multivariate space of the first two PCs. It can be seen that the scores are arranged essentially in three groups: they include fir, chestnut, and TTF-multifloral honeys, respectively.

The circled samples in **Figure 3** correspond to outlier samples. The results of the mellissopalynology analysis reveal that these samples show some chestnut, sunflower, and/or clover pollens in their main pollen spectrum in significant amounts, indicating a contamination by the corresponding nectars. This explains the location of these samples in the PC scores vectors plot.

Tables 3 and **4** show the total classification rate obtained by LDA applied on clear and dark honeys, respectively. Ninety-three percent of the clear honey samples were correctly classified, and 100% correct classification was obtained for rosemary honey samples. Moreover, dark honey samples were perfectly classified as shown in **Table 4**. These results are partially similar to those obtained by Mateo et al. (*54*), who analyzed the sugars content of different Spanish unifloral honey types (rosemary, orange blossom, lavender, sunflower, eucalyptus, heather, and honeydew) by gas chromatography. In their work, they showed that honeydew, sunflower, heather, and eucalyptus honeys were classified (using LDA) with 100, 92.9, 83.3, and 75% correct classifications, respectively. However, for the remaining honey types, particularly rosemary and lavender honeys, which are in common with our study, the percentages of successful classifications remained weak (53.8−69.2%) compared to that given in our study. This can be explained by the number and type of sugars selected by Mateo et al.: 10 sugars of which maltulose and kojibiose do not seem to be discriminating of selected varieties. [In our study, we selected 13 sugars including trehalose, characteristic of honeydew honeys (dark honeys) and raffinose + palatinose, which were found to be characteristics of nectar honeys (clear honeys).] Moreover, Mateo et al. considered in their study both honeydew and nectar honeys in the same dataset, but we demonstrated here that results are significantly improved by considering honeydew and nectar honeys separately.

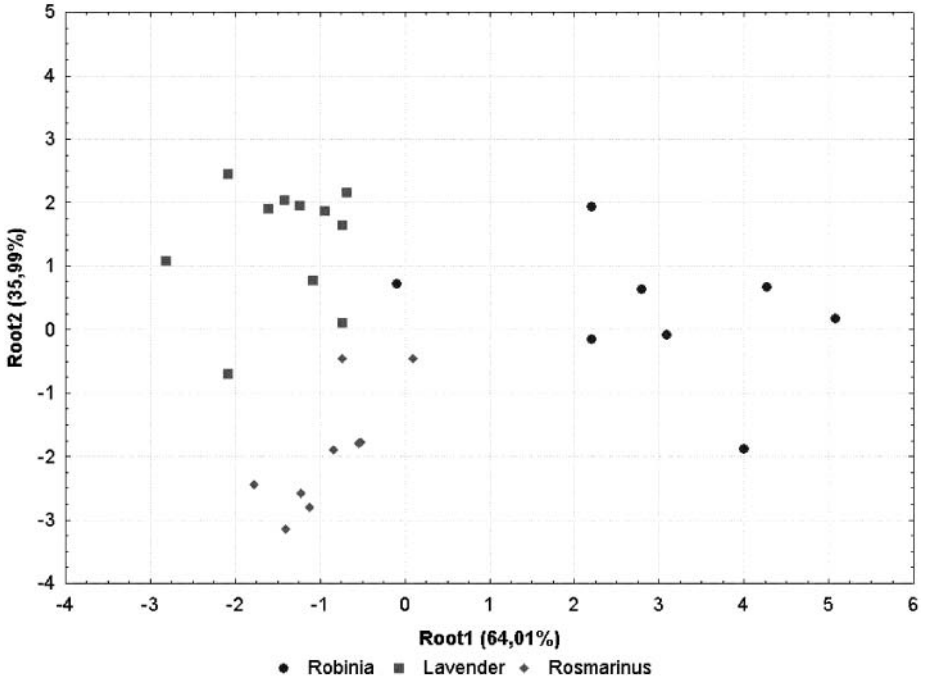

Figure 4. Separation of clear honeys using linear discriminant analysis: plot of discriminant scores.

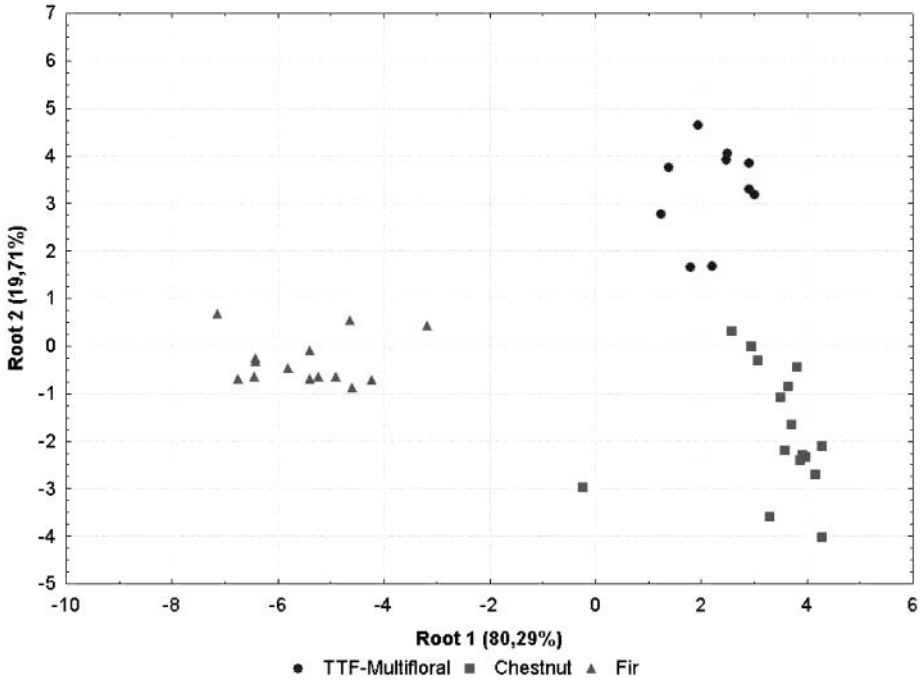

Figure 5. Separation of dark honeys using linear discriminant analysis: plot of discriminant scores.

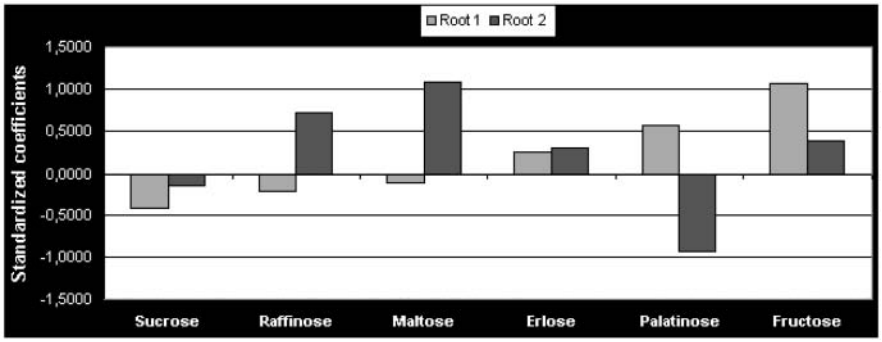

Figure 6. Contribution of variables to roots 1 and 2 in **Figure 4** (clear honeys).

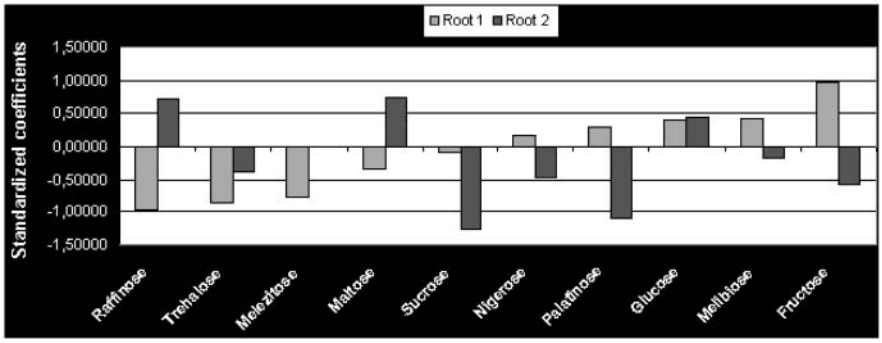

Figure 7. Contribution of variables to roots 1 and 2 in **Figure 5** (dark honeys).

Three discriminant functions were extracted for both clear and dark honeys, one function per group of honey.

Table 5 presents the variables used to build the discriminant functions. These variables were selected in terms of their Wilk's Lambda values, which indicate the contribution of each variable to the discrimination. In this case, a variable having a Wilk's Lambda value of <0.06324 (clear honeys) or <0.01006 (dark honeys) has not been selected to build the discrimination functions. It can be seen that raffinose and palatinose are the most powerful parameters in the discrimination of clear honeys, followed by maltose, erlose, sucrose, and fructose for robinia honey; fructose, erlose, maltose, and sucrose for lavender honey; erlose, fructose, maltose, and sucrose for rosemary honey; and sucrose, maltose, and trehalose for TTF-multifloral, chestnut, and fir honeys, respectively. These results are in agreement with what is commonly expected for clear and dark honeys, concerning the analysis of sugars relative importance. It should be noted that these results do not mean that the sugars having the best discriminant power have also the highest content. **Table 5** shows the contribution of other sugars in the constructed discriminant functions for the dark honeys.

Figures 4 and **5** show the results obtained by LDA (plot of the discriminant scores). **Figures 6** and **7** show the contribution of each variable to the two roots depicted in **Figures 4** and **5**.

Table 6. Performances of the Four Best Neural Networks Retained from 500 Tested

no. of networks/ 500 tested	network structure	total CR[a] (%)	error		
			train	select	test
76	13:13-**11**-13-6:1	89.07	0.091	0.844	1.063
78	13:13-**13**-13-6:1	90.07	0.096	0.840	2.245
91	13:13-**13**-**10**-6:1	89.88	0.100	0.784	1.719
107	13:13-**13**-13-6:1	90.07	0.0007	0.720	2.414

[a] Total classification rate.

The standardized coefficients of the root 1 variables are ranked according to their individual contribution. One can see, for example, that for clear honeys the fructose, maltose, and palatinose variables have the highest contribution in roots 1 and 2, whereas for dark honeys the variables presenting the highest contribution are sucrose, palatinose, fructose, raffinose, trehalose, and melezitose.

Artificial Neural Networks. The results presented here constitute, to our knowledge, the first application of ANNs to the recognition of the floral species of honeys. To compare the performance of a neural network approach against the multivariate procedures, we tested 500 different networks and the 100 best networks were retained. The criterion used to select networks for retention was that of balancing performance against diversity. Three types of networks were investigated in a classification process: linear, radial basis function (RBF), and three- and four-layer perceptron (MLP3 and MLP4). The highest confidence was used as the classification thresholds for all networks tested. The internal structure complexity (number [min/max] of hidden units) of the selected network types is the following: RBF [1/17]; MLP3 [1/13]; and MLP4 [1/13].

As for the multivariate approach, we applied ANNs on a full set of honey samples to see whether it was possible to find a network capable of classifying all samples with a good level of classification. The algorithms used for training the MLP networks are back-propagation (train set) and conjugate gradient descent (select set and test set) because they are described as being well suited for this type of network. For the RBF and linear networks, the training algorithms used were *K* means (KM; for train set), *K* nearest neighbor (KNN; for select set), and pseudo-invert (PI; linear least squares optimization, for test set) and PI (for train, select, and test sets), respectively. Over the 100 retained networks, only four (MLP networks) showed a total classification rate (mean of train set + select set + test set results) near or equal to 90% of good classification. No other presented a better classification rate. **Table 6** shows the results obtained by these four networks with the associated error for the train, select, and test sets.

Table 7. Description of the Best MLP Neural Network Used for Classifying Clear Honey Samples

no. of networks/ 100 tested	network structure	total CR (%)	error		
			train	select	test
98	6:6-**9**-3:1	100.00	0.000007	0.0745	0.000016

classification	robinia	lavender	rosemary
total	8	11	9
correct	8	11	9
wrong	0	0	0
unknown	0	0	0
correct (%)	100	100	100
wrong (%)	0	0	0
unknown (%)	0	0	0

The best RBF and linear networks present a total classification rate of 69.6 and 78.4%, respectively. This result tends to show that MLP networks are better suited for classifying HPAEC profiles of honey samples from the data matrix generated by our previously described algorithm.

The results were substantially improved when clear honeys were separated from dark honeys as was done during the multivariate analysis. Of 25 retained networks (100 tested), 10 MLP networks presented a perfect classification rate (100%) but differed from their train, select, and test errors. **Tables 7** and **8** present the main characteristics of the best network for clear and dark honeys, respectively.

As shown in **Figure 8**, the more compact (with the minimum number of hidden layers and hidden units) the network

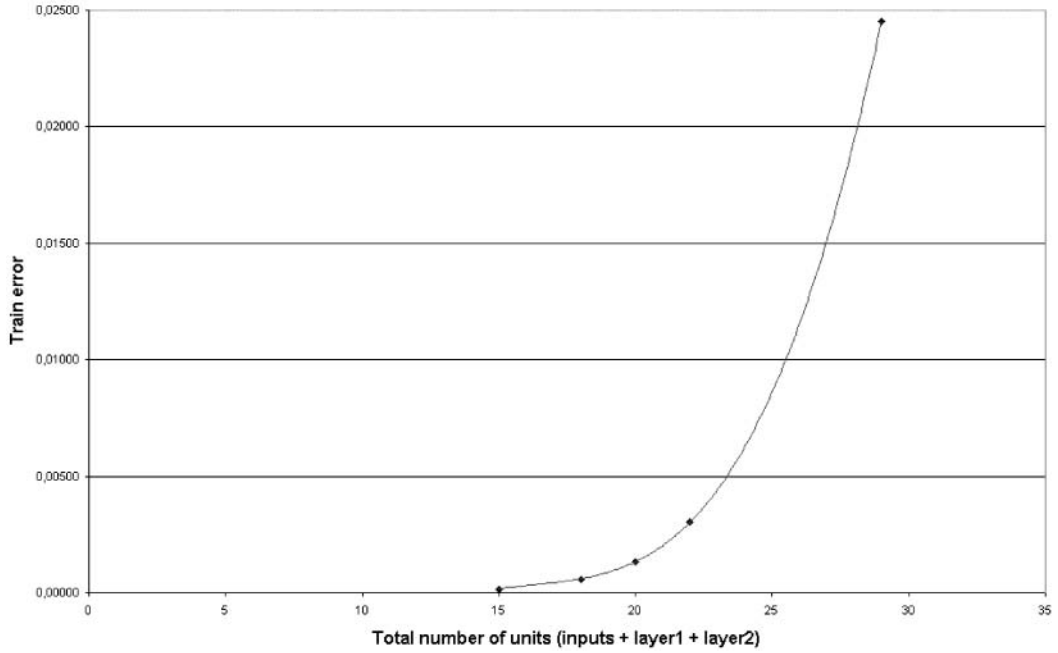

Figure 8. Representation of the train error for the 10 best MLP networks in terms of internal complexity (15–29 units dispatched over one or two hidden layers).

Table 8. Description of the Best MLP Neural Network Used for Classifying Dark Honey Samples

no. of networks/ 100 tested	network structure	total CR (%)	error		
			train	select	test
474	9:9-**11**-**12**-3:1	100.00	8.29E–5	1.06E–5	0.1248

classification	TTF-multifloral	chestnut	fir
total	10	16	14
correct	10	16	14
wrong	0	0	0
unknown	0	0	0
correct (%)	100	100	100
wrong (%)	0	0	0
unknown (%)	0	0	0

architecture, the higher is the performance of the selected network (with a lowest select error). In this case, the best network was an MLP3 net, whereas for the full set of samples an MLP4 net was the best. This can be explained by a reduction of the overlap between honey sample groups and, therefore, the structure of the network tends to diminish to lead to a smaller network having a good discriminating power.

The structure of the best neural network found for dark honeys is {9:9-**11**-**12**-3:1}, that is an MLP4 network. The classification and the prediction performances are perfect, and the three floral species (TTF, chestnut, and fir) are efficiently separated and predicted. It should be noted that no "over-learning" was encountered for the best networks that were finally selected in this work.

Conclusion. The HPAEC-PAD method and the numerical methodology for the construction of the initial data matrix presented here provide an efficient tool for the characterization of the honey floral species from ionic chromatographic profiles. The chromatographic method used in this work presents a good resolution/time of analysis ratio and shows that the HPAEC-PAD technique can be used in an automated chemometric approach for honey characterization.

The two chemometric approaches used for comparison in this work are equivalent when both clear and dark honeys are treated together (89.7% by multivariate analysis and 90.7% by ANN). However, the application of neural networks gives a better result when the two types of honey samples are separately treated (92.9% by multivariate analysis and 100% by ANN). The quality of the results shows that the matrix generated by our algorithm preserves the information content required for good chemical data representativeness.

Given that our goal is the elaboration of an automated tool for the recognition of honey varieties and the detection of their eventual adulteration, the next step is the utilization of the same tools for detecting adulterated samples and the determination of the sensitivity of the method.

ACKNOWLEDGMENT

We thank Dr. Tom Forrest for revision of the manuscript and his availability during the development of this work.

LITERATURE CITED

(1) Opinion of the Scientific Committee on Veterinary Measures Relating to Public Health on Honey and Microbiological Hazards (Adopted June 19−20, 2002). European Commission, Health and Consumer Protection Directorate-General. Directorate C−Scientific Opinions, C2−Management of scientific committees; scientific cooperation and networks.

(2) Swallow, K. W.; Low, N. H. Analysis and Quantitation of the Carbohydrates in Honey Using High-Performance Liquid Chromatography. *J. Agric. Food Chem.* **1990**, *38*, 1828−1832.

(3) Bogdanov, S.; Vit, P.; Kilchenmann, V. Sugar Profile and Conductivity of Stingless Bee Honeys from Venezuela. *Apidologie* **1996**, *27*, 445−450.

(4) Sporn, P.; Plhak, L.; Friedrich, J. Alberta Honey Composition. *Food Res. Int.* **1992**, *25*, 93−100.

(5) Pérez-Arquillué, C.; Conchello, P.; Arino, A.; Juan, T. J.; Herrera, A. Physicochemical Attributes and pollen Spectrum of Some Unifloral Spanish Honeys. *Food Chem.* **1995**, *54*, 167−172.

(6) Lombard, A.; Buffa, M.; Patetta, A.; Manino, A.; Marletto, F. Some Aspects of the Carbohydrate Composition of Callaphidid Honeydew. *J. Apic. Res.* **1987**, *26* (4), 233−237.

(7) Astwood, K.; Lee, B.; Manley-Harris, M. Oligosaccharides in New Zealand Honeydew Honey. *J. Agric. Food Chem.* **1998**, *46*, 4958−4962.

(8) Low, N. H.; South, W. Determination of Honey Authenticity by Capillary Gas Chromatography. *J. AOAC Int.* **1995**, *78*, 1210−1218.

(9) Guyot, C.; Bouseta, A.; Scheirman, V.; Collin, S. Floral Origin Markers of Chestnut and Lime Tree Honeys. *J. Agric. Food Chem.* **1998**, *46*, 625−633.

(10) Pérez, R. A.; Sánchez-Brunete, C.; Calvo, R. M.; Tadeo, J. L. Analysis of Volatiles from Spanish Honeys by Solid-Phase Microextraction and Gas Chromatography−Mass Spectrometry. *J. Agric. Food Chem.* **2002**, *50*, 2633−2637.

(11) D'Arcy, B. R.; Rintoul, G. B.; Rowland, C. Y.; Blackman, A. J. Composition of Australian Honey Extractives. 1. Norisoprenoids, Monoterpenes, and Other Natural Volatiles from Blue Gum (*Eucalyptus leucoxylon*) and Yellox Box (*Eucalyptus melliodora*) Honeys. *J. Agric. Food Chem.* **1997**, *45*, 1834−1843.

(12) Guyot-Declerck, C.; Chevance, F.; Lermusieau, G.; Collin, S. Optimized Extraction Procedure for Quantifying Norisoprenoids in Honey and Honey Products. *J. Agric. Food Chem.* **2000**, *48*, 5850−5855.

(13) Bogdanov, S. Determination of Pinocembrin in Honey Using HPLC. *J. Apic. Res.* **1989**, *28* (1), 55−57.

(14) Martos, I.; Cossentini, M.; Ferreres, F.; Tomás-Barberán, F. A. Flavonoid Composition of Tunisian Honey and Propolis. *J. Agric. Food Chem.* **1997**, *45*, 2824−2829.

(15) Tomás-Barberán, F. A.; Martos, I.; Ferreres, F.; Radovic, B. S.; Anklam, E. HPLC Flavonoid Profiles as Markers for the Botanical Origin of European Unifloral Honeys. *J. Sci. Food Agric.* **2001**, *81*, 1−12.

(16) Lo Coco, F.; Valentini, C.; Novelli, V.; Ceccon, L. High-Performance Liquid Chromatographic Determination of 2-Furaldehyde and 5-Hydroxymethyl-2-furaldehyde in Honey. *J. Chromatogr. A* **1996**, *749*, 95−102.

(17) Andrade, P. Analysis of Honey Phenolic Acids by HPLC, its Application to Honey Botanical Characterisation. *J. Liq. Chromatogr. Relat. Technol.* **1997**, *20*, 2281−2288.

(18) Del Nozal, M. J. High Performance Liquid Chromatographic Determination of Organic Acids in Honeys From Different Botanical Origin. *J. Liq. Chromatogr. Relat. Technol.* **1998**, *21*, 3197−3214.

(19) Cherchi, A.; Spanedda, L.; Tuberoso, C.; Cabras, P. Solid-Phase Extraction and High-Performance Liquid Chromatographic Determination of Organic Acids in Honey. *J. Chromatogr. A* **1994**, *669*, 59−64.

(20) Ferreres, F.; Andrade, P.; Tomas-Barberan, F. A. Natural Occurrence of Abscisic Acid in Heather Honey and Floral Nectar. *J. Agric. Food Chem.* **1996**, *44*, 2053−2056.

(21) Wilkins, A. L.; Lu, Y.; Tan, S. T. Extractives from New Zealand Honeys. 5. Aliphatic Dicarboxylic Acids in New Zealand Rewarewa (Knightea excelsa) Honey. *J. Agric. Food Chem.* **1995**, *43*, 3021−3025.

(22) Wootton, M.; Edwards, R. A.; Faraji-Haremi, R. Effect of Accelerated Storage Conditions on The Chemical Composition and Properties of Australian Hdoneys. 2. Changes in Sugar and Free Amino Acid Contents. *J. Apic. Res.* **1976**, *15* (1), 29−34.

(23) Davies, A. M. C. The Application of Amino Acid Analysis to the Determination of the Geographical Origin of Honey. *J. Food Technol.* **1976**, *11*, 515−523.
(24) Davies, A. M. C. Amino Acid Analysis of Honeys From Eleven Countries. *J. Apic. Res.* **1975**, *14* (1), 29−39.
(25) Perez, C.; Conchello, P.; Arino, A.; Yanguela, J.; Herrera, A. Dosage des Acides Aminés des Protéines de Différents Miels Espagnols. *Sci. Aliments* **1989**, *9*, 203−207.
(26) Brückner, H.; Langer, M.; Lüpke, M.; Westhauser, T.; Godel, H. Liquid Chromatographic Determination of Amino Acid Enantiomers by Derivatization with *o*-phthaldialdehyde and Chiral Thiols. Applications with Reference to Food Science. *J. Chromatogr. A* **1995**, *697*, 229−245.
(27) Pawlowska, M.; Armstrong, D. W. Evaluation of Enantiomeric Purity of Selected Amino Acids in Honey. *Chirality* **1994**, *6*, 270−276.
(28) Gilbert, J.; Shepherd, M. J.; Wallwork, M. A.; Harris, R. G. Determination of the Geographical Origin of Honeys by Multivariate Analysis of Gas Chromatographic Data on their Free Amino Acid Content. *J. Apic. Res.* **1981**, *20* (2), 125−135.
(29) Pirini, A.; Conte, L. S.; Francioso, O.; Lercker, G. Capillary Gas Chromatographic Determination of Free Amino Acids in Honeys as a Means of Discrimination between Different Botanical Sources. *J. High Resolut. Chromatogr.* **1992**, *15* (3), 165−170.
(30) Cordella, C.; Moussa, I.; Martel, A.-C.; Sbirrazzuoli, N.; Lizzani-Cuvelier, L. Recent Developments in Food Characterization and Adulteration Detection: Techniques Oriented Perspectives. *J. Agric Food Chem.* **2002**, *50*, 1751−1764.
(31) Anklam, E. A Review of the Analytical Methods to Determine the Geographical and Botanical Origin of Honey. *Food Chem.* **1998**, *63*, 549−562.
(32) Krauze, A.; Zalewski, R. Classification of Honeys by Principal Component Analysis on the Basis of Chemical and Physical Parameters. *Z. Lebensm. Unters. Forsch.* **1991**, *192*, 19−23.
(33) López, B.; Latorre, M. J.; Fernández, M. I.; García, M. A.; García, S.; Herrero, C. Chemometric Classification of Honeys According to Their Type Based on Quality Control Data. *Food Chem.* **1996**, *55*, 281−287.
(34) Peña Crecente, R.; Herrero Latorre, C. Pattern Recognition Analysis Applied to Classification of Honeys from Two Goegraphical Origins. *J. Agric. Food Chem.* **1993**, 560−564.
(35) Mateo, R.; Bosch-Reig, F. Classification of Spanish Unifloral Honeys by Discriminant Analysis of Electrical Conductivity, Color, Water Content, Sugars, and pH. *J. Agric. Food Chem.* **1998**, *46*, 393−400.
(36) Terrab, A.; Vega-Pérez, J. M.; Diez, M. J.; Heredia, F. J. Characterization of Northwest Moroccan Honeys by Gas Chromatographic−Mass Spectrometric Analysis of their Sugar Contents. *J. Sci. Food Agric.* **2001**, *82*, 179−185.
(37) Louveaux, J.; Maurizio, A.; Vorwohl, G. Methods of Melissopalynology. *Bee World* **1978**, *59* (4), 139−157.
(38) Cataldi, T. R. I.; Campa, C.; Casella, I. G.; Bufo, S. A. Determination of Maltitol, Isomaltitol, and Lactitol by High-PH Anion-Exchange Chromatography with Pulsed Amperometric Detection. *J. Agric. Food Chem.* **1999**, *47*, 157−163.
(39) LaCourse, W. R.; Johnson, D. C. Optimization of Waveforms for Pulsed Amperometric Detection of Carbohydrates Based on Pulsed Voltametry. *Anal. Chem.* **1993**, *65*, 50−55.
(40) Cordella, C. B. Y.; Militão, J. S. L.; Cabrol-Bass, D. New Software for Automated Calibration of HPLC Profiles Usable in Pattern Recognition Procedures: Application to HPAEC-PAD Profiles of Honeys. *Anal. BioAnal. Chem.* **2002**, submitted for publication.
(41) Hair, J. F. et al. *Multivariate Data Analysis*, 3rd ed.; MacMillan: New York, 1992. Jenrich, R. I. Stepwise Discriminant Analysis. In *Statistical Methods for Digital Computers*; Enslein, K., Ralston, A., Wilf, H. S., Eds.; Wiley: New York, 1960; pp 76−95.
(42) Patterson, D. *Artificial Neural Networks*; Prentice Hall: Singapore, 1996.
(43) Fausett, L. *Fundamentals of Neural Networks*; Prentice Hall: New York, 1994.
(44) *Neural Networks in Statistica Electronic Manual*, verrsion 6; StatSoft, Inc.: Tulsa, OK, 2001; Statistica (data analysis software system), www.statsoft.com.
(45) Zupan, J.; Gasteiger, J. Neural Networks: A New Method for Solving Chemical Problems or Just a Passing Phase? *Anal. Chim. Acta* **1991**, *248*, 1−30.
(46) Basheer, I. A.; Hajmeer, M. Artificial Neural Networks: Fundamentals, Computing, Design, and Application. *J. Microbiol. Methods* **2000**, *43*, 3−31.
(47) Song, X.-H.; Chen, Z.; Yu, R.-Q. Artificial Neural Networks Applied to Odor Classification for Chemical Compounds. *Comput. Chem.* **1993**, *17* (3), 303−308.
(48) Briandet, R.; Kemsley, E. K.; Wilson, R. H. Approaches to Adulteration Detection in Instant Coffees using Infrared Spectroscopy and Chemometrics. *J. Sci. Food Agric.* **1996**, *71*, 359−366.
(49) Angerosa, F.; Di Giacinto, L.; Vito, R.; Cumitini, S. Sensory Evaluation of Virgin Olive Oils by Artificial Neural Network Processing of Dynamic Headspace Gas Chromatographic Data. *J. Sci. Food Agric.* **1996**, *72*, 323−328.
(50) Sun, L.-X.; Danzer, K.; Thiel, G. Classification of Wine Samples by Means of Artificial Neural Networks and Discrimination Analytical Methods. *Fresenius' J. Anal. Chem.* **1997**, *359*, 143−149.
(51) Storbeck, F.; Daan, B. Fish Species Recognition using Computer Vision and A Neural Network. *Fish. Res.* **2001**, *51*, 11−15.
(52) Therdthai, N.; Zhou, W. Artificial Neural Network Modelling of the Electrical Conductivity Property of Recombined Milk. *J. Food Eng.* **2001**, *50*, 107−111.
(53) USDA. Pfund determination of the color of extracted honey. In Circular 24-1927, revised 1933; U.S. Department of Agriculture: Washington, DC, 1933; 20 pp.
(54) Mateo, R.; BoschReig, F. Sugar Profiles of Spanish Unifloral Honeys. *Food Chem.* **1997**, *60* (1), 33−41.

Received for review November 6, 2002. Revised manuscript received March 4, 2003. Accepted March 7, 2003. Part of this work was financially supported by a European grant.

JF021100M