

Analytical, Nutritional and Clinical Methods

Classification of monofloral honeys based on their quality control data

J. Devillers ^{a,*}, M. Morlot ^b, M.H. Pham-Delègue ^c, J.C. Doré ^d

^a Centre for the Treatment of the Scientific Information (CTIS), 3 Chemin de la Gravière, 69140 Rillieux La Pape, France

^b Bernard Michaud S.A., BP 27, Domaine Saint-Georges, Chemin de Berdoulou, 64290 Gan, France

^c LNCI, INRA, BP 23, 91440 Bures-sur-Yvette, France

^d Département Régulations, Développement et Diversité Moléculaire, USM 0502, UMR 8041 CNRS, Muséum National d'Histoire Naturelle, 63 rue de Buffon, 75005 Paris, France

Received 20 June 2003; received in revised form 28 September 2003; accepted 28 September 2003

Abstract

Four hundred and sixty-nine samples of fir, cinder heather, chestnut, lavender, acacia, rape, and sunflower honey were characterized by their moisture, conductivity, diastase activity, pH, free acidity, color, hydroxymethylfurfural and percentage of fructose, glucose, saccharose, erlose, raffinose, and melezitose. A principal component analysis performed on the corresponding matrix yielded the formation of four clusters. A stepwise discriminant analysis allowed us to obtain 100% of good predictions with only conductivity, pH, free acidity and percentage of fructose, glucose, and raffinose as variables. The simulation performances of the model were estimated from an external testing set.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: Honey; Physicochemical characteristics; Principal component analysis; Discriminant analysis; Predictive model

1. Introduction

Nowadays, the food products have to satisfy numerous quality and certification criteria before commercialization, especially in industrial countries, where there is a need to have food products of high quality with well-defined characteristics. Honey is no exception and in Europe, its composition as well as manufacture is regulated by the Council Directive 74/409/EEC of 22 July 1974. In June 2000, the Council reached political agreement on a new honey Directive to harmonize the common European market. Consequently, honey producers have now to indicate the botanical (floral or vegetable) and geographical (regional or territorial) origin of the honey. In the frame of this new Directive, the development of harmonized analytical methods to easily assess labeling compliance is also encouraged.

These requirements are not easy to satisfy. Indeed, honey is a natural substance made when the nectar and sweet deposits from plants are gathered, modified and stored in the honeycomb by honey bees. Consequently, its composition is defined by numerous factors and subject to variation.

Traditionally, the determination of the floral origin of honey is made from palynological analysis. The method is based on the identification of pollen by microscopic inspection. The shortcomings of melissopalynology have been stressed by Anklam (1998). They deal with the variability in the amounts of pollen collected by the bees, problems of falsification, and so on.

To overcome this problem, and also to save time and money, attempts have been made to predict the botanical origin of honeys from some of their physicochemical properties by means of multivariate analyses (Anklam, 1998; Anklam & Radovic, 2001; Goodall, Dennis, Parker, & Sharman, 1995; Krauze & Zalewski, 1991; Mateo & Bosch-Reig, 1997; Popek, 2002; Terrab, Diez, & Heredia, 2002). Undoubtedly, the obtained results clearly show that it is not a wishful thinking. However,

* Corresponding author. Tel.: +33-(0)4-78-08-49-84; fax: +33-(0)4-78-08-56-37.

E-mail address: j.devillers@ctis.fr (J. Devillers).

most of the studies are based on a fairly limited number of honey samples. Moreover, when models are designed from discriminant analysis (DA) or related techniques, the predictive power of these models is never estimated on an external testing set while it is the *sine qua non* in simulation modeling (Devillers & Karcher, 1991).

Consequently, the aim of this study was first to validate this kind of methodological approach from a large data base of monofloral honeys for which physicochemical characteristics were available. Second, attempts were made to determine the minimum number of physicochemical properties required to obtain the best classification of honeys according to their botanical origin.

2. Materials and methods

2.1. Honey samples

A data base of 469 monofloral honey samples were provided by Bernard Michaud S.A. laboratory (F), the biggest honey producer and conditioner in France. The botanical origin of the honey was determined by means of melissopalynological analyses (Louveaux, Maurizo, & Vorwohl, 1970). Honey samples were of various geographical origins but were similarly aged. The following seven certified monofloral honey categories were included: fir (*Abies* sp.), cinder heather (*Erica cinerea*), chestnut (*Castanea sativa*), lavender (*Lavandula* sp.), acacia (*Robinia pseudoacacia*), rape (*Brassica napus*), and sunflower (*Helianthus annuus*). The number of samples in each honey category is given in Table 1.

2.2. Analytical determination

The 469 honey samples were analyzed to determine their 13 following physicochemical characteristics: moisture (Abbé refractometer, Prolabo, Toulouse, F), conductivity (conductometer LF 537, Prochilab, Bordeaux, F), diastase activity (after Schade, spectrophotometer Anthélie, Secoman, Toulouse, F), pH and free acidity (pH meter Mettler Toledo MP 225, Mettler, Toulouse, F), color (Pfund scale, Lovibond 2000 & 1000 comparator), fructose, glucose, saccharose, erlose, raffinose, and mel-ezitose content (HPLC Dionex 500, pulsed amperometric detection), hydroxymethylfurfural (HMF) (HPLC Thermo quest Separator P200, AS 100, UV detection). All the analyses were performed according to the methods in agreement with EU legislation (Bogdanov et al., 1997). This yielded the design of a 13×469 data matrix particularly suitable for multivariate analyses.

2.3. Data analysis

Principal Component Analysis (PCA) was used to obtain a reduction of dimensionality of the 13×469

data matrix and to discover the relationships between variables (physicochemical properties) and objects (honey samples) through optimal 2-D graphical displays. Briefly, PCA replaces the original variables of a data set with a smaller number of uncorrelated variables called principal components (PCs). The method is linear in that the new variables are a linear combination of the original ones (Devillers & Karcher, 1991). Due to the nature of the physicochemical data, a classical PCA on correlation matrix after autoscaling of the variables was used (Lopez et al., 1996).

For deriving a model allowing to simulate the different types of honey, a discriminant analysis (DA) was employed. Indeed, DA is a very useful tool (1) for classifying cases into different groups with a better than chance accuracy and (2) for detecting the variables allowing the better discrimination between groups. To reach this goal, a forward stepwise procedure was employed but confirmation of the results was also obtained by using the backward stepwise option. The addition or removal of a variable or effect was based on the *p* or *F* values to enter/remove (Devillers & Karcher, 1991).

All the statistical analyses were performed with ADE-4 (<http://pbil.univ-lyon1.fr/ADE-4/ADE-4.html>; Devillers & Doré, 2002) and Statistica™ Version 6 (StatSoft, France).

3. Results and discussion

The analytical results are summarized in Table 1. Inspection of this table shows that some parameters have a high discriminatory power. This is the case, for example, for the conductivity, color, erlose, raffinose or mel-ezitose. Conversely, other physicochemical properties do not widely vary from one honey to another. Thus, for example, if we consider the minimum (Min) and maximum (Max) values of humidity measured for the 7 types of honey, no significant differences are found (Table 1).

In order to cluster the seven botanical types of honeys from their physicochemical properties, a standardized PCA was used.

PC₁PC₂ accounts for 59.53% (i.e., 37.51% + 22.02%) of the total inertia of the system. Even if it should be necessary to consider the other PCs to deeply analyze the studied data matrix (Fig. 1A), we assume that this first factorial plane allows us to stress the main trends in the data.

Fig. 1B shows that the 469 honey samples are split into four leading clusters. Because a code was used to represent each category of honey, it is easy to see on the figure that these clusters correspond to specific honey types. Thus, all the fir honey samples (code number 1) are isolated in the right part of Fig. 1B and form a strong cluster. This is also the case for the chestnut

Table 1
Physicochemical characteristics of the 469 honey samples

Botanical origin Parameter	Mean	SD ^a	Min	Max
<i>Fir (n=57)</i>				
Moisture (%)	17.60	0.581	16.20	18.80
HMF (mg/kg)	3.490	2.143	0.100	7.500
Conductivity (µs/cm)	1069	122.2	870.0	1403
Diastase activity (ID)	24.15	3.777	18.49	33.16
pH	5.151	0.286	4.750	5.790
Free acidity (meq/kg)	24.24	3.535	14.62	30.49
Color (mm Pfund)	75.00	6.682	65.00	85.00
Fructose (%)	33.37	1.422	29.56	36.46
Glucose (%)	25.63	1.592	22.25	28.69
Saccharose (%)	1.352	0.609	0.450	2.590
Erlose (%)	0.816	0.200	0.320	1.110
Raffinose (%)	1.565	0.468	0.920	2.860
Melezitose (%)	2.217	0.481	0.960	3.120
<i>Cinder heather (n=43)</i>				
Moisture (%)	18.21	0.510	17.00	19.50
HMF (mg/kg)	5.162	2.646	1.230	10.81
Conductivity (µs/cm)	604.2	66.16	469.0	769.0
Diastase activity (ID)	14.73	3.002	8.560	21.35
pH	4.064	0.154	3.780	4.360
Free acidity (meq/kg)	18.96	1.728	15.23	22.56
Color (mm Pfund)	62.56	10.49	40.00	85.00
Fructose (%)	40.17	1.213	37.90	42.59
Glucose (%)	35.75	1.396	33.25	38.96
Saccharose (%)	0.122	0.121	0	0.400
Erlose (%)	0	0	0	0
Raffinose (%)	0	0	0	0
Melezitose (%)	0	0	0	0
<i>Chestnut (n=62)</i>				
Moisture (%)	18.79	0.857	17.00	20.50
HMF (mg/kg)	2.653	2.119	0.100	8.430
Conductivity (µs/cm)	1308	363.1	785.0	1883
Diastase activity (ID)	23.29	3.747	15.93	32.00
pH	5.283	0.461	4.360	6.480
Free acidity (meq/kg)	12.20	2.517	8.210	17.98
Color (mm Pfund)	81.13	7.489	65.00	100.0
Fructose (%)	37.39	1.374	34.98	40.84
Glucose (%)	31.60	1.885	29.18	39.99
Saccharose (%)	0.250	0.280	0	1.150
Erlose (%)	0.047	0.083	0	0.330
Raffinose (%)	0.218	0.247	0	0.840
Melezitose (%)	0.423	0.307	0	1.150
<i>Lavender (n=57)</i>				
Moisture (%)	16.70	0.485	15.60	17.60
HMF (mg/kg)	3.205	1.490	0.990	6.770
Conductivity (µs/cm)	221.2	52.58	22.00	310.0
Diastase activity (ID)	14.51	1.964	10.51	19.72
pH	3.702	0.083	3.460	3.860
Free acidity (meq/kg)	14.86	1.447	10.87	17.45
Color (mm Pfund)	33.60	6.392	20.00	45.00
Fructose (%)	35.51	1.085	32.65	37.54
Glucose (%)	31.37	1.829	27.98	34.95
Saccharose (%)	2.689	0.852	0.190	4.370
Erlose (%)	0.922	0.564	0.210	3.100
Raffinose (%)	0	0	0	0
Melezitose (%)	0	0	0	0
<i>Acacia (n=34)</i>				
Moisture (%)	18.48	0.690	17.20	20.30
HMF (mg/kg)	2.462	1.338	0.590	5.900
Conductivity (µs/cm)	195.4	40.46	120.0	289.0

Table 1 (continued)

Botanical origin Parameter	Mean	SD ^a	Min	Max
Diastase activity (ID)	19.03	6.568	10.23	33.25
pH	3.897	0.127	3.620	4.120
Free acidity (meq/kg)	8.954	1.171	6.300	11.36
Color (mm Pfund)	7.647	4.126	5.000	25.00
Fructose (%)	39.81	1.107	38.20	42.90
Glucose (%)	26.88	1.149	24.20	28.95
Saccharose (%)	2.049	1.239	0.230	5.300
Erlose (%)	1.554	0.610	0.450	2.500
Raffinose (%)	0	0	0	0
Melezitose (%)	0	0	0	0
<i>Rape (n = 96)</i>				
Moisture (%)	18.46	0.655	17.00	19.80
HMF (mg/kg)	3.196	1.665	0.210	5.950
Conductivity (µs/cm)	203.1	44.35	110.0	269.0
Diastase activity (ID)	26.85	5.911	11.20	36.80
pH	4.019	0.119	3.700	4.260
Free acidity (meq/kg)	10.66	1.318	6.510	12.30
Color (mm Pfund)	25.99	4.079	20.00	35.00
Fructose (%)	37.90	1.218	34.20	39.60
Glucose (%)	40.74	1.320	35.20	42.40
Saccharose (%)	0	0	0	0
Erlose (%)	0	0	0	0
Raffinose (%)	0	0	0	0
Melezitose (%)	0	0	0	0
<i>Sunflower (n = 120)</i>				
Moisture (%)	18.19	0.566	16.60	19.40
HMF (mg/kg)	3.191	1.849	0.230	9.560
Conductivity (µs/cm)	306.2	57.47	230.0	500.0
Diastase activity (ID)	25.04	5.653	16.23	38.56
pH	3.888	0.087	3.660	4.090
Free acidity (meq/kg)	19.91	3.423	14.23	26.59
Color (mm Pfund)	47.25	6.606	30.00	55.00
Fructose (%)	38.76	1.048	36.25	41.10
Glucose (%)	37.62	1.071	35.10	40.20
Saccharose (%)	0.227	0.146	0	0.500
Erlose (%)	0	0	0	0
Raffinose (%)	0	0	0	0
Melezitose (%)	0	0	0	0
<i>Total (n = 469)</i>				
Moisture (%)	18.09	0.874	15.60	20.50
HMF (mg/kg)	3.287	2.006	0.100	10.81
Conductivity (µs/cm)	519.2	437.7	22.00	1883
Diastase activity (ID)	22.41	6.574	8.560	38.56
pH	4.247	0.616	3.460	6.480
Free acidity (meq/kg)	16.03	5.548	6.300	30.49
Color (mm Pfund)	47.62	23.23	5.000	100.0
Fructose (%)	37.56	2.339	29.56	42.90
Glucose (%)	34.30	5.366	22.25	42.40
Saccharose (%)	0.742	1.066	0	5.300
Erlose (%)	0.330	0.567	0	3.100
Raffinose (%)	0.219	0.539	0	2.860
Melezitose (%)	0.325	0.746	0	3.120

Min, minimum; Max, maximum.

^aSD, standard deviation.

honeys (code number 3), which are all displayed in the middle bottom part of Fig. 1B. Cinder heather honey (number 2), rape honey (number 6) and sunflower honey (number 7) samples are mixed together and located in the left part of Fig. 1B into a fairly compact cluster. The

top of Fig. 1B is only occupied by the lavender honey (number 4) and acacia honey (number 5) samples. It is noteworthy that two samples of acacia honey and one sample of rape honey are mixed between their corresponding cluster. The classical way to see the variables

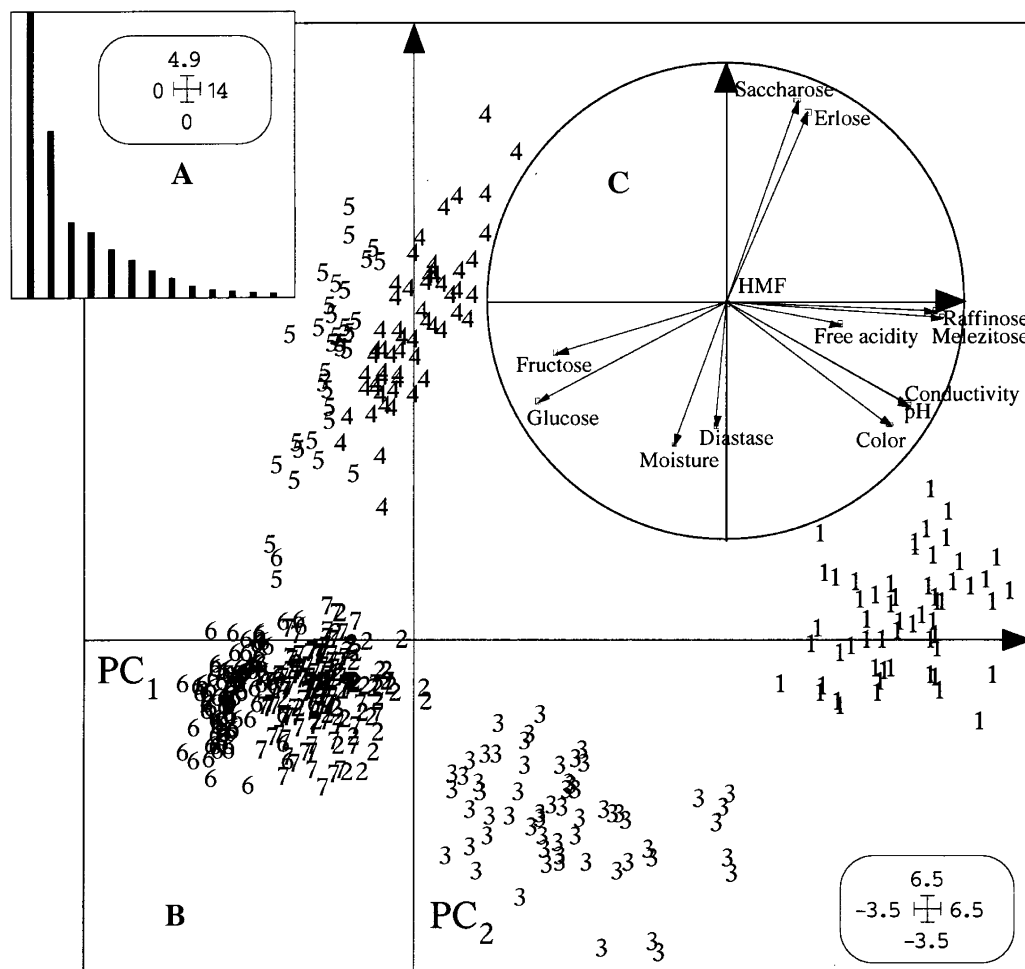


Fig. 1. Ordination resulting from the PCA of the 469 honey samples described by 13 physicochemical properties. Histogram of the eigenvalues (A), samples (B), correlation circle (C).

responsible for the formation of the four clusters in Fig. 1B is to inspect the correlation circle (Fig. 1C). Thus, for example, in Fig. 1C, HMF (hydromethylfurfural) being located at the origin of PC_1PC_2 , this variable does not influence the formation of the clusters in Fig. 1B. Conversely, the specific location of the fir honey (code number 1) samples in Fig. 1B is due to their high percentages of raffinose and melezitose in comparison with the other honey types. A more convenient strategy for interpreting the clusters in Fig. 1B consists in directly plotting the values of each variable on this map. This yields a collection of 13 graphs, one per studied variable (Fig. 2). On each graph the larger the black circle, the greater the corresponding value and hence, the higher its influence on the formation of the corresponding cluster. Thus, for example, Fig. 2 confirms that HMF does not participate in the formation of the clusters. Conversely, erlose mainly explains that lavender honey and acacia honey form a cluster in Fig. 1B.

PCA yielding a fairly good separation of the honey types from their physicochemical properties, it was le-

gitimate to try to derive a quantitative model from discriminant analysis (DA). Obviously, the challenge was to model the seven types of honey and not their clusters as they appear in Fig. 1B. The data base was split into a training set for building the DA model and an external testing set for estimating its simulation performances. Because in DA, it is required to have balanced classes, the number of rape and sunflower honey samples randomly allocated to the testing set was higher than this for the other honeys (see Table 1). The training set of 364 samples included 52 fir honey samples, 38 cinder honey samples, 56 chestnut honey samples, 52 lavender honey samples, 30 acacia honey samples, 68 rape honey samples, and 68 sunflower honey samples. The external testing set of 105 samples included 5 fir honey samples, 5 cinder honey samples, 6 chestnut honey samples, 5 lavender honey samples, 4 acacia honey samples, 28 rape honey samples, and 52 sunflower honey samples.

Because HMF did not participate in the clustering of the honey samples (Fig. 1), DA was performed from the 12 remaining variables. The forward and backward stepwise options were used for selecting the most

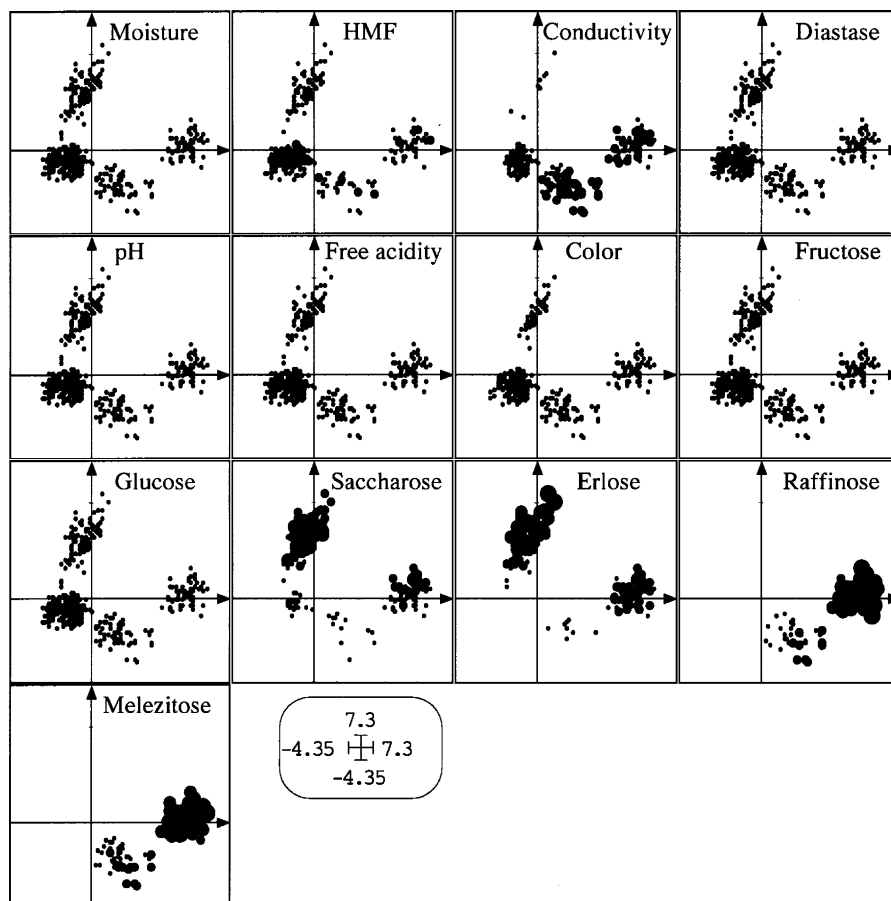


Fig. 2. Projection of the values of each variable on PC_1PC_2 (Fig. 1B). The higher the circle, the greater the value.

interesting variables allowing to obtain the better simulation results with the training set and more important, with the external testing set. A DA model only including conductivity, pH, free acidity, and percentage of fructose, glucose, and raffinose as variables yielded 100% of good classification with both the training and testing sets. The raw canonical discriminant function (DF) coefficients are given in Table 2. They can be used to compute the raw canonical scores for each case and DF. From them, it is possible to verify the modeling results or to perform new simulations. The standardized canonical DF coefficients, the eigenvalues for each DF, and the cumulative proportion of common variance extracted by each DF are displayed in Table 3. Inspec-

tion of Table 3 reveals that the first discriminant function is principally under the dependence of % glucose, pH, and conductivity. The percentages of fructose and raffinose and free acidity dominate the second one. Our results are in agreement with those found in the literature. Thus, Mateo and Bosch-Reig (1997) and Terrab, Vega-Pérez, Diez, and Heredia (2001) showed that it was possible to discriminate unifloral honeys from their sugar profile. Terrab, Diez & Heredia (2002) stressed the importance of pH for discriminating eucalyptus, citrus, lythrum, apiaceae, and honeydew honey samples. However, while the free acidity parameter was selected by their DA model, it is noteworthy that the standardized coefficient of this variable was inferior to that ob-

Table 2
Raw canonical discriminant function (DF) coefficients

Parameter	DF1	DF2	DF3	DF4	DF5	DF6
Intercept	1.86497	-11.8603	-23.9458	11.61571	24.09737	-9.36904
Conductivity	0.00248	0.0021	0.0035	-0.00251	0.00176	-0.00382
pH	1.80681	0.8415	1.9151	1.28633	-0.71464	3.15825
Free acidity	0.04535	-0.2026	0.0170	-0.35657	0.05531	0.10215
Fructose	0.11645	0.5884	-0.0661	-0.36868	-0.57651	-0.04337
Glucose	-0.49064	-0.3237	0.4782	0.09455	-0.01768	-0.04882
Raffinose	1.44196	-3.3296	-0.3258	2.18030	-2.84663	-1.67979

Table 3

Standardized canonical discriminant function (DF) coefficients, eigenvalues, and cumulative proportion of common variance extracted by each DF

Parameter	DF1	DF2	DF3	DF4	DF5	DF6
Conductivity	0.38723	0.321437	0.542023	-0.392096	0.274808	-0.596570
pH	0.41624	0.193857	0.441181	0.296331	-0.164631	0.727565
Free acidity	0.10875	-0.485966	0.040685	-0.855171	0.132658	0.244989
Fructose	0.14616	0.738527	-0.082936	-0.462750	-0.723601	-0.054436
Glucose	-0.71373	-0.470927	0.695567	0.137536	-0.025716	-0.071012
Raffinose	0.28290	-0.653231	-0.063923	0.427753	-0.558480	-0.329557
Eigenvalue	21.85513	7.251658	6.290780	2.782504	0.659128	0.027844
Cumulative proportion	0.56230	0.748881	0.910735	0.982325	0.999284	1.000000

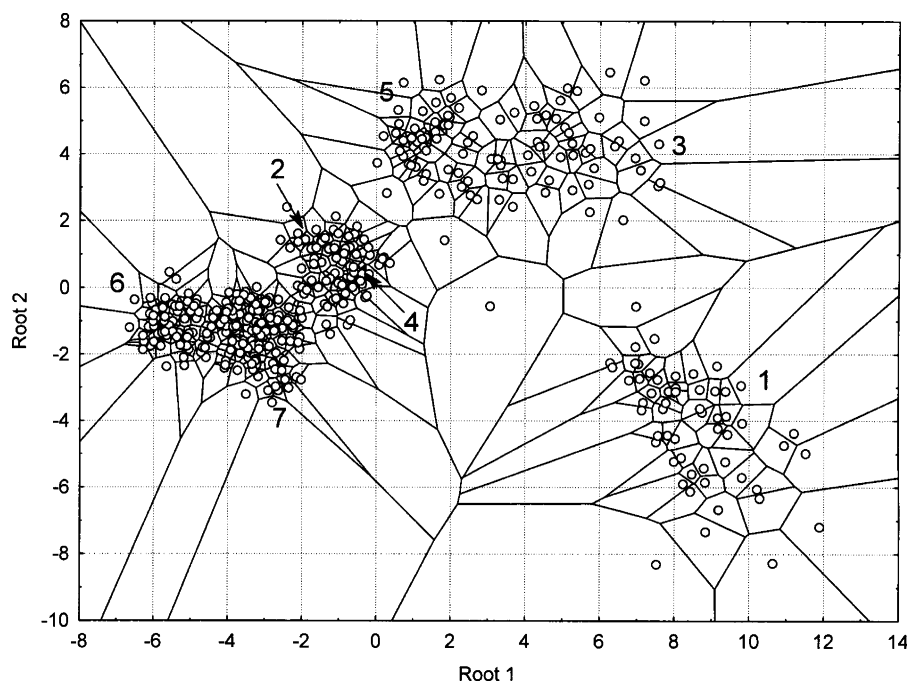


Fig. 3. Scatterplot of the canonical scores obtained with the two first discriminant functions. Nos. 1–7 represent the different honey categories.

tained which the lactonic acidity/free acidity ratio parameter.

To visualize the discrimination power of the model, a simple 2D scatterplot of the canonical scores obtained with the two first discriminant functions has been represented in Fig. 3 with Voronoi tessellations to better underline the clusters (Okabe, Boots, Sugihara, & Chiu, 1999). While only about 75% of the variance are extracted with these two functions (Table 3), a fairly good separation of the honeys is obtained. Indeed, only cinder heather honey samples (number 2) and lavender honey samples (number 4) are not well separated.

In this study, an attempt was made to discriminate seven types of unifloral honeys from a minimum of their physicochemical characteristics. A discriminant model, using conductivity, pH, free acidity and the percentage

of fructose, glucose, and raffinose of the honeys as variables, allows to perfectly reach this goal. This model allows us to save time and money in the determination of the floral origin of honey. However, to be used in practice, it should be necessary to extend its domain of application to other categories of unifloral honeys. This will be the goal of our future work.

Acknowledgements

We are grateful to P. Beaune, F. Britis, R. Lurdos, C. Métaux, and J.C. Vernet (Bernard Michaud SA laboratory, France) for technical assistance. This study was partly granted by the Programme Communautaire sur l'“Apiculture 2002”.

References

- Anklam, E. (1998). A review of the analytical methods to determine the geographical and botanical origin of honey. *Food Chemistry*, 63(4), 549–562.
- Anklam, E., & Radovic, B. (2001). Suitable analytical methods for determining the origin of European honey. *American Laboratory*, May, 60–64.
- Bogdanov, S., Martin, P., Lüllmann, C., Borneck, R., Flamini, C., Morlot, M., Heretier, J., Vorwohl, G., Russmann, H., Persano-Oddo, L., Sabatini, A. G., Marazzan, G. L., Marioleas, P., Tsigouri, K., Kerkvliet, J., Ortiz, A., & Ivanov, T. (1997). Harmonised methods of the European honey commission. *Apidologie*, 1–59 (extra issue).
- Devillers, J., & Doré, J. C. (2002). e-Statistics for deriving QSAR models. *SAR QSAR in Environmental Research*, 13(3–4), 409–416.
- Devillers, J., & Karcher, W. (1991). *Applied multivariate analysis in SAR and environmental studies* (p. 530). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Goodall, I., Dennis, J. J., Parker, I., & Sharman, M. (1995). Contribution of high performance liquid chromatographic analysis of carbohydrates to authenticity testing of honey. *Journal of Chromatography A*, 706, 353–359.
- Krauze, A., & Zalewski, R. (1991). Classification of honeys by principal component analysis on the basis of chemical and physical parameters. *Zeitschrift für Lebensmittel-Luntersuchung und –Forschung*, 192, 19–23.
- Lopez, B., Latorre, M. J., Fernandez, M. I., Garcia, M. A., Garcia, S., & Herrero, C. (1996). Chemometric classification of honeys according to their type based on quality control data. *Food Chemistry*, 55(3), 281–287.
- Louveaux, J., Maurizo, A., & Vorwohl, G. (1970). Commission internationale de botanique apicole de l'U.I.S.B. Les méthodes de la méliko-palynologie. *Apidologie*, 1(2), 211–227.
- Mateo, R., & Bosch-Reig, F. (1997). Sugar profiles of Spanish unifloral honeys. *Food Chemistry*, 60, 33–41.
- Okabe, A., Boots, B., Sugihara, K., & Chiu, S. N. (1999). *Spatial tessellations: Concepts and applications of Voronoi diagrams* (2nd ed., p. 696). Wiley.
- Popek, S. (2002). A procedure to identify a honey type. *Food Chemistry*, 79, 401–406.
- Terrab, A., Diez, M. J., & Heredia, F. J. (2002). Characterisation of Moroccan unifloral honeys by their physicochemical characteristics. *Food Chemistry*, 79, 373–379.
- Terrab, A., Vega-Pérez, J. M., Diez, M. J., & Heredia, F. J. (2001). Characterisation of northwest Moroccan honeys by gas chromatographic-mass spectrometric analysis of their sugar components. *Journal of the Science of Food and Agriculture*, 82, 179–185.