CLASSIFICATION
OF HONEYDEW AND BLOSSOM HONEYS
BY DISCRIMINANT ANALYSIS

Technical-scientific information

Contents

S Bogdanov and M Gfeller

CLASSIFICATION
OF HONEYDEW AND BLOSSOM HONEYS
BY DISCRIMINANT ANALYSIS

Abstract

The objective of the present study is to use discriminant analysis for the classification of blossom and honeydew honeys by parameters which can be determined easily in routine honey control. Following parameters were determined in 113 blossom and 34 honeydew honeys of Swiss origin: the sugars glucose, fructose, turanose, saccharose, nigerose, maltose, isomaltose, erlose and melezitose; pH; free acidity; electrical conductivity and protein content. The honeys were classified by discriminant analysis using single and multiple variables. The best single variable for honey classification was melezitose with 96 % correct classification, while the other variables had a weaker discriminatory power. Two canonical variable functions proved especially powerful for discrimination between blossom and honeydew honeys. The first used following variables: sum of fructose and glucose content, ratio of fructose and glucose content, pH and electrical conductivity. This function classified correctly 98.2 % of the blossom and 81.2 % of of the honeydew honeys. The other function using in addition melezitose classified correctly 88.2 % of the honeydew and all blossom honeys. The results were successfully validated using 180 different Italian unifloral honeys, 84 Swiss honeys of honeydew origin and 242 blossom honeys of various botanical origin.

1 Introduction

In Switzerland and other countries honeys are very often labelled only either as honeydew (forest) or blossom (polyfloral) honey. Electrical conductivity is the honey parameter most widely used for distinguishing between these two honey types (Talpay, 1985; Bogdanov et al., 1987; Bogdanov et al., 1995; Persano Oddo and Piro, 2004). According to the EU honey standard the electrical conductivity of blossom honeys should be below 0.8 mS/cm, while that of honeydew honeys should exceed that value (European Commission, 2002). Generally, the electrical conductivity of mixed blossom-honeydew honeys lies between 0.5 and 0.8 mS/cm, while the values of pure blossom honeys are below 0.5 mS/cm (Talpay, 1985; Bogdanov et al., 1999). However, there are many exceptions to this rule. Chestnut, arbutus, erica, eucalyptus and linden honeys, just to mention several exceptions, are regarded as blossom honeys, but have often electrical conductivity values above 0.8 mS/cm (Persano Oddo and Piro, 2004; European Commission, 2002).

Other parameters such as pH value (Talpay, 1985; Bogdanov et al., 1995; Persano Oddo and Piro, 2004) or the sum of the glucose and fructose content (G+F), have also been used for the differentiation between blossom and honeydew honey (Bogdanov et al., 1987). According to the present EU honey standard, the G+F minimum value for blossom honeys should be 60 g/100g , while for honeydew honeys it is 45 g/100 g (European Commission, 2002). Other measurands, not commonly determined in routine honey control such as melezitose (Bogdanov et al., 1987; Persano Oddo and Piro, 2004), various oligosaccharides or amino acids (Bogdanov et al., 2004) have also been used.

Due to the broad variation of the honey chemical parameters, a safe distinction between honeydew and blossom honeys can only be carried out in specialized laboratories by chemical, microscopic and sensory analysis of honey. The botanical origin of honey is determined by combining the results of all three methods. On the other hand, food control laboratories generally determine chemical parameters only and depend on honey specialists, when they have to check the labelled botanical origin of honey.

Discriminant analysis is a component of all common statistical software packages. It is increasingly used for classification of groups within a group total, if the classification with separate parameters is not successful. It has been used in many cases for the classification of different unifloral honeys (Bogdanov et al., 2004), but not for the groups blossom and honeydew honey. A linear discriminant function, called "Kirkwood number" has been suggested for the differentiation between these two honey groups (Kirkwood et al., 1960):

$$X = -8.3x_1 - 12.3x_2 + 1.4x_3$$

where x_1=pH, x_2= honey ash content in % and x_3= % reducing sugars.

This function is rarely applied nowadays. One reason is that the determination of reducing sugars and total ash content has been replaced by chromatographic and by electrical conductivity measurements.

In this investigation we use discriminant analysis for the classification of blossom and honeydew honeys by parameters which can be determined easily in routine honey control.

2 Materials and methods

Honey samples

Data of unifloral honeys of Swiss origin, used in the frame of a first project for the determination of unifloral honeys in Switzerland was used (Bogdanov, 1989; Bogdanov, 1997).
The honeys had the following botanical origins: 42 chestnut, 34 rape, 12 dandelion, 6 rhododendron, 7 acacia honeys, 5 mixed blossom and 34 honeydew.

Methods
The determination of sugars, pH, free acidity and electrical conductivity was carried out according to the Swiss Food Manual (Bogdanov et al., 1995). Additionally, the protein content was determined according to Bogdanov (1981). The following sugars were determined: glucose, fructose, turanose, saccharose, nigerose, maltose, isomaltose, erlose and melezitose. The sugar concentration was expressed in g/100 g, the electrical conductivity was expressed in milli Siemens per cm.

Discriminant analysis

Discriminant analysis was carried out by the STATISTICA statistical software package (version 5). The variables, described under "Methods" were used in addition, the derived variables sum of glucose and fructose content (G+F) and the ratio between fructose and glucose (G/F) content were used. The data were submitted to a step by step discrimination analysis. For each variable set the program calculates the percentage of correctly classified cases, the canonical variable and the F value for each variable. The F value is a parameter for the importance of the variable for group classification: the higher the F value, the greater the importance of the corresponding variable for the discrimination. In the next steps we eliminated the parameters, with the smallest F value to reach the best classification with the least number of variables. The intersection point between both honey groups was calculated as the point lying at a distance equalling the same number of standard deviations from the means of each group.

3 Results and Discussion

Discriminant analysis with one variable

The results of the discriminant analysis with only one variable are summarised in table 1.

The best discrimination between both honey groups is achieved with melezitose: honeys containing less than 0.6 g/100 g of melezitose are classified as blossom honeys, those with more than 0.6 g/100 g as honeydew honeys. 95% of the blossom honeys and 97% of the honeydew honeys were classified correctly using this discriminant value. With G+F a good discrimination of both honey groups is also obtained (88-90% of correct classifications). The correct classification with electrical conductivity was between 66 and 74 % while the pH value had a relatively poor discriminant capacity (53 to 60 %).

Discriminant analysis with more variables

The first discriminant analysis of our data was carried out with all variables measured. With step by step elimination of the functions, contributing little to the correct classification, two functions with the least number of variables were selected, classifying well both honey types. The results of the discriminant analysis with the variables, yielding the best classification results are summarised in table 2. The classification of honey using these two sets of variables was nearly the same as that using all variables.

Two canonical variable functions for discrimination of the honey groups were calculated:
· Function A uses glucose, fructose, conductivity and pH, variables, determined in routine measurements
· Function B, containing in addition melezitose.
  The discriminant function A is calculated according to the following equation (table 2):

$$X = 0.28 \times (F+G) + 3.03 \times (F/G) + 1.35 \times pH - 2.65 \times C - 26{,}81$$

The value of the canonical variable for honey is obtained by substituting the values in the function. The values of the canonical variables for both groups are classified as follows:

honeydew honeys < -0.629 < blossom honeys

where -0.629 is the intersection point between the two groups.

The more negative or positive the value for the canonical variables in comparison to the intersection point, the more correct the classification as blossom or honeydew honey should be. Honeys with values around the intersection of -0.629 are probably honeys, containing both honeydew and nectar. 98.2 % of the blossom honeys and 91.2 % of the honeydew honeys were classified correctly with this function. The correct classification with the best single classification variable G+F, was significantly lower with an overall correct

classification of 89 %. Thus, by means of simple measurements, such as pH, electrical conductivity, glucose and fructose determination, and by using the discriminant canonical function formed by these variables, the classification of honey can be improved, compared to the conventional determination of honeydew honey origin by means of a separate determination of electrical conductivity, G+F or pH value. This function is similar to Kirkwood's discriminant function, using similar variables: pH, % reducing sugars and % ash content (see introduction). The variable contributing most of all to the discrimination of honey types is G+F. G+F corresponds roughly to the reducing sugar content used in the Kirkwood function, because fructose and glucose are the main reducing sugars. The electrical conductivity, used as a variable in this study directly correlates to the ash content of honey used in the Kirkwood classification function.

In figure 1 the distribution of the canonical variables of the blossom and honeydew samples is shown graphically. While the discriminant values of the blossom honeys are nearly normally distributed, the values of honeydew honeys are spread over a wide range and have a distribution which is far from normal. This can be explained by the different origin of honeydew honey. It is known that honeydew honeys can derive from honeydew produced on different trees and by various aphids. The fact, that honeydew honeys are often mixed with various amounts of nectar honey, might also contribute to the abnormal distribution of the honeydew honey canonical values.

It is surprising, that the importance of electrical conductivity for the classification with the canonical variable is relatively small. In routine honey control the purity and the quality of honeydew honey is determined mostly by conductivity measurements. However, there are blossom honeys with a relatively high conductivity (chestnut, arbutus, erica, eucalyptus, linden), which can be classified to be of honeydew origin on the basis of the electrical conductivity. Function 1 however, should allow the routine classification of the great majority of blossom honeys.

If the melezitose content is used as an additional variable (discriminant function 2 in table 2), the classification of blossom honeys is only slightly better, while that of honeydew honeys increases to 100 %.

The distribution of the canonical variables (fig. 1 B) is similar to that without melezitose (fig. 1 A). In this case, too, the distribution of the values of the honeydew honeys is far from normal. The determination of melezitose content has to be carried out with costly chromatographic methods, and therefore is not performed in routine honey analysis. On the other hand, F+G can be determined with simple enzymatic methods (Bogdanov et al., 1995).

4 Conclusion

Validation of the models with other data

We applied function 1 on data of 180 different Italian unifloral honeys, 140 of blossom and 40 of honeydew origin (Accorti et al., 1986). Our classification function A classified all of the honeydew and 92% of the blossom honeys correctly. The worst classification with 70 % correct results was found for the arbutus and euclalyptus honeys, while for the other honeys the classification was better than 90 %.

Recently, a second project dealing with the description of Swiss unifloral honeys was elaborated (Bogdanov et al., 2005). The two discriminant functions were tested on the data of 84 honeydew and 242 blossom honeys of different botanical origin: 84 mixed blossom, 24 acacia, 16 rhododendron, 54 chestnut, 27 dandelion and 37 rape honeys. 96 % of the blossom honeys were correctly classified both by model A and B. The correct classification was almost 100 % for blossom honeys, other than chestnut honey. On the other side, only 84 % of the chestnut honeys were classified as blossom honeys, the rest being classified as honeydew honey. Indeed, it is known, that chestnut honeys contain sometimes more or less honeydew. 88 % of the honeydew honeys were correctly classified by function A, while 94 % of them were correctly classified by function B. Two out of 14 linden honeys were classified as honeydew honeys. It is known that linden honeys can often contain honeydew.

With the present study the classification of blossom and honeydew honeys by two discriminant functions is proposed. One of them, using the easily measurable parameters fructose, glucose, conductivity and pH, is especially suitable for routine honey analysis.

5 References

Accorti,M.; Persano Oddo,L.; Piazza,M.G. and Sabatini,A.G. (1986) Schede di caratterizzazione delle principali qualita di miele Italiano. Apicoltura **2**: 1-35.

Bogdanov S. (1981) Bestimmung von Honigprotein mit Coomassie Brillantblau G 250. Mitt. Gebiete Lebensm. Hyg. **72**: 411-417

Bogdanov S. (1989) Blütensortenhonige in der Schweiz, Schweiz. Bienenztg. 112, 681-684.

Bogdanov S. (1997) Charakterisierung von Schweizer Sortenhonigen, Agrarforschung 4, 427-430.

Bogdanov S., Bieri K., Figar M., Figueiredo V., Iff D., Känzig A., Stöckli H., Zürcher K. (1995) Kapitel 23 Bienenprodukte: 23A Honig, Schweiz. Lebensmittelbuch.

Bogdanov,S.; Bieri,K.; Kilchenman,V. and Gallmann,P. (2005) Schweizer Sortenhonige. ALP forum **23 d**: 3-55.

Bogdanov,S.; Bieri,K.; Kilchenman,V. and Gallmann,P. (2005) Schweizer Sortenhonige. ALP forum 23 d: 3-55.

Bogdanov S., Lüllmann C., Martin P., von der Ohe W., Russmann H., Vorwohl G., Persano Oddo L., Sabatini A.G., Marcazzan G.L., Piro R., Flamini C., Morlot M., Lhertier J., Borneck R., Marioleas P., Tsigouri A., Kerkvliet J., Ortiz A., Ivanov T., D'Arcy B., Mossel B., Vit P. (1999) Honey quality, methods of analysis and international regulatory standards: review of the work of the International honey comission, Mitt. Lebensm. Hyg. 90, 108-125.

Bogdanov S., Rieder K., Rüegg M. (1987) Neue Qualitätskriterien bei Honiguntersuchungen, Apidologie 18, 267-278.

Bogdanov S., Ruoff K., Persano Oddo L. (2004) Physico-chemical methods for the characterisation of unifloral honeys: a review, Apidologie 35, S4-S17.

European Commission (2002) Council Directive 2001/110/EC of 20 December 2001 relating to honey. Official Journal of the European Communities L10: 47-52.

Kirkwood K.C., Mitchell T.J., Smith D. (1960) An examination of the occurrence of honeydew in honey, Analyst 85, 412-416.

Persano Oddo L., Piro R. (2004) Main European unifloral honeys: descriptive sheets, Apidologie 35, S38-S81.

Talpay B. (1985) Spezifikationen für Trachthonige, Dtsch. Lebensmittel Rundschau 81, 148-152.

variable	blossom honeys		honeydew honeys		D
	$\overline{x}$ +/-sd	% corr	$\overline{x}$ +/-sd	% corr	
Fructose %	39.6+/-2.71	84	33.8+/-2.5	91	36.7
Fructose/Glucose	1.33+/-0.24	50	1.33+/-0.13	50	1.34
Fructose+Glucose	70.1+/-3.6	90	59.3+/-4.8	88	65.5
Melezitose %	0.2+/-0.3	95	5.1+/-3.4	97	0.6
pH	4.5+/-0.8	60	4.5+/-0.26	53	4.5
Conductivity (mS/cm)	0.74+/-0.54	66	1.04+/-0.19	74	0.96

Table 1 **Classification of blossom and honeydew honeys with discriminant analysis using single variables**
% corr - correct classification in %, D - discriminant value to discriminate best between the two honey groups,
$\overline{x}$, average, sd- standard deviation

		Function A
	F value	A = 0.28x(F+G)+3.03x(F/G)+1.35xpH - 2.65xC-26.81
F+G	198.44	
F/G	23.45	intersection point= -0.629
Conductivity	20.18	values for blossom honey: $\overline{x}$ = 0.878 +/- 0.881 (sd)
pH	17.79	values for honeydew $\overline{x}$ = -2.884 +/- 1.327 (sd)
		correct classification : 98.2% B and 91.2% H
		Function B
	F-value	B = 2.49x(F/G)+0.20x(F+G) - 0.30xMel + 1.09xpH - 2.18xC - 19.78
Melezitose	225.67	
F+G	41.62	intersection point= -0.259
F/G	13.96	values for blossom honey: $\overline{x}$ = 0.99 +/- 0.69 (sd)
Conductivity	14.13	values for honeydew honey $\overline{x}$ = -3.29 +/- 1.67 (sd)
pH	11.85	correct classification : 100% B and 88.2 % H

Table 2 **Classification of blossom and honeydew honeys with discriminant analysis**
Variables:
Function A Fr+Gl sum of fructose and glucose content, Fr/Gl: ratio between the fructose and glucose content; C - conductivity;
Function B same as function 1 with the addition of Mel, melezitose content
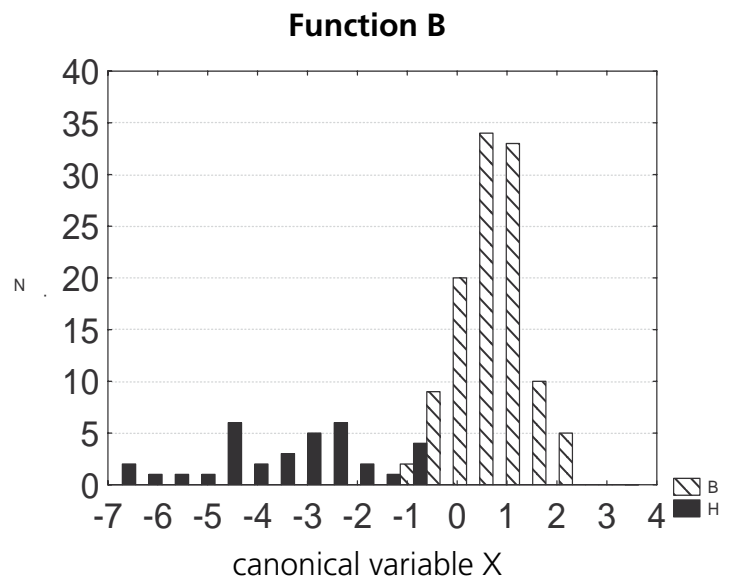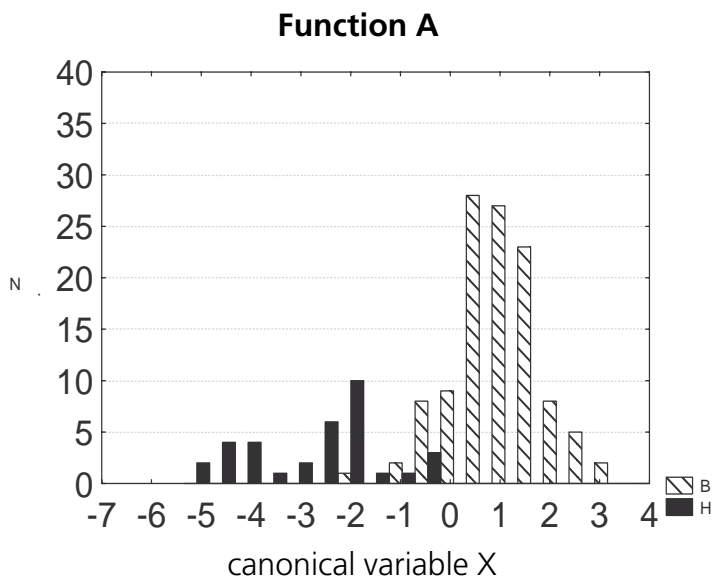B: blossom, H: honeydew honey; $\overline{x}$, average; sd - standard deviation

Figure 1 **Classification of blossom and honeydew honey with discriminant analysis with two functions** (see table 2), B: blossom honey; H: honeydew honey