DETERMINATION OF BOTANICAL ORIGIN OF HONEY BY MID INFRARED SPECTROSCOPY (Mid-FTIR), COLORIMETRY AND CHEMOMETRIC ANALYSIS

Svetla Lekova, Diana Tsankova

ABSTRACT

The aim of the investigation reported is to determine the potential of determining the botanical origin of honey by a combination of middle infrared spectroscopy, colorimetry and chemometric analysis. Forty-three samples from three types of honey (acacia, linden, and honeydew) are subjected to Fourier transform mid-infrared (Mid-FTIR) spectroscopy. Transmittance spectra with wave numbers from 600 cm\(^{-1}\) to 4000 cm\(^{-1}\) are recorded. The latter are combined with some colorimetric parameters (CIELab) for training data of a linear discriminant analysis (LDA) based calibration model. Principal components analysis (PCA) is used to reduce the number of inputs (wave numbers and colour parameters) and get a proper visualization of the experimental results. The high accuracy of the proposed honey classifier is confirmed by a leave-one-out-cross validation test carried out in MATLAB environment.

Keywords: honey discrimination, Mid Infrared spectroscopy (Mid-FTIR), honey colorimetry, chemometric models, clustering analysis.

INTRODUCTION

According to the European Union regulations and standards [1, 2], “honey stipulates a pure product that does not allow for the addition of any other substance”. Its characteristic properties depend on the floral source. The determination of the botanical origin is part of the quality analysis of honey [3] and affects its commercial value [4, 5]. According to the international legislation [1, 2] the product-name “honey” can be supplemented by information referring to the botanical origin (authentication) provided that the honey comes wholly or mainly from the indicated source. So in order to prevent fraud in the labelling, means to distinguish between the different types of honey have to be developed. The content of different phenolic compounds is well recognized to reflect the type of honey and its quality, because phenolic acids and flavonoids are inherent chemical markers of the floral origin [6, 5]. At the current stage of knowledge, a reliable authentication of honey’s floral origin can be achieved by a global interpretation of sensory, pollen and physicochemical analyses carried out by an expert [7 - 9]. Most of these methods are generally too time-consuming, complex, and labour intensive for quality control application or require very specialized personnel to interpret the results.

The advantages of the technique of spectroscopy (UV, visible, near and middle infrared, fluorescent) with respect to other methods refer to its non-invasive approach as well as the relatively easy and quick data acquisition. A significant proportion of the recent work undertaken has concentrated on the mid-IR part of the electromagnetic spectrum because the fundamental vibration is observed there (as opposed to overtones or harmonics in near-IR) which makes particularly rich in information [10]. This technique is based on the follow-
ing. The sample is interrogated with an infrared beam (usually in the mid-infrared from 4000 cm\(^{-1}\) – 400 cm\(^{-1}\) or near-infrared 14 000 cm\(^{-1}\) - 4000 cm\(^{-1}\)) and the radiation is absorbed by the functional groups contained. As a result they will vibrate in one of the ways recognized (stretching or bending vibrations) [11] providing a direct correlation to (bio)chemical species [10]. The resultant infrared absorbance spectrum can be described as a chemical ‘fingerprint’, as it is characteristic of the particular sample under analysis and hence, every chemical or biochemical substance will have its own unique infrared profile [12, 13].

Fourier transform infrared (FTIR) spectroscopy is a rapid and reproducible technique of a holistic nature [10]. But, as the measurement refers to the whole IR spectrum, validated and robust chemometrics must be used in order to translate data into information [13, 14]. The approach of combining Mid-FTIR spectroscopy with multivariate calibration demonstrates the significant advantages of the chemometric models compared with conventional analytical methods in terms of decrease of the high level subjectivity in results interpretation. This approach has been used by many authors [15, 16] to determine the falsification of honey through adding other substances. A few [3, 17] applied it to determine honey’s botanical origin.

The Discriminant Analysis (DA) is probably the best known method [18] among the traditional classifiers. It can be considered the first multivariate classification technique. Some authors [19-22] have implemented linear discriminant analysis (LDA) to classify the floral origin of honey, mainly on the basis of its chemical and physical properties, as well as its mineral composition. Usually LDA and other statistical classification methods are used in a combination with the principal components analysis (PCA) as a correlation reduction method.

The purpose of this study is to investigate the potential for determining the botanical origin of honey by using a combination of middle infrared spectroscopy, colormetry and chemometric analysis. The data obtained by Mid-FTIR spectroscopy in an absorbance mode are supplemented with CIELab colorimetric parameters and then undergo subsequent statistical processing including: PCA for reducing the number of classifier’s inputs and LDA for cluster distinguishing (honey type recognition). The number of principal components (PCs) included in the calibration model is optimized by means of a leave-one-out-cross validation test carried out in MATLAB environment.

**EXPERIMENTAL**

**Materials and Methods**

**Honey Spectrum Acquisition and Color Measuring**

Forty-three samples of three different types of honey (acacia - 13 samples; linden - 17 samples; and honeydew - 13 samples) were purchased from supermarkets and private producers. Prior to the spectral measurement, the honey samples were placed in a water container at 50°C until the soluble substances fully dissolved. Then the samples were annealed at room temperature (25°C-26°C). The honey spectral characteristics were recorded by a Mid-FTIR spectrometer Varian 660 ranging from 400 cm\(^{-1}\) to 4000 cm\(^{-1}\) at 1.929 cm\(^{-1}\) sampling space. To exclude the spectra noisy parts, only the range between 600 cm\(^{-1}\) and 4000 cm\(^{-1}\) was taken under consideration for further data analysis. Then the absorbance spectra were derived from the transmittance spectra (Transmittance, [%]) in accordance with:

\[
\text{Absorbance} = -\log_6 (\text{Transmittance}/100).
\]

Given that various spectral features arise in the different regions of the honey spectrum, data set structures consisted of two different spectral intervals, 3600 cm\(^{-1}\)-2400 cm\(^{-1}\) and 1800 cm\(^{-1}\) - 700 cm\(^{-1}\) (fingerprint region) [17]. They were considered together in order to find the best calibration model. The spectral regions pointed above were chosen on the ground of the mid-infrared bands of some important chemical groups reported previously [23, 17, 24].

The measurements were performed at room temperature in correspondence with the methods of the European Honey Commission [25]. The colorimetric study of honey was carried out using a software package VISIONlite ColorCalc for spectrophotometer Helios Omega. The mode ‘Advanced’ was used, i.e. the calculations were performed in the range of 380 nm - 780 nm. The honey samples were placed in a cuvette 10 mm x 10 mm and the color parameters in CIELab colorimetric system were measured.

**Statistical Multivariate Analysis: PCA and LDA**

The aim of the PCA method is to reduce the multivariate data dimensionality whilst preserving as much
of the relevant information as possible. PCA transfers the data (observations of possibly correlated variables) linearly to a new coordinate system. Thus a new set of variables, the principal components, is obtained. It is a set of linear functions of the original variables. PCs are uncorrelated. The greatest variance of data projection comes to lie on the first coordinate, the second greatest variance - on the second coordinate, and so on. The full set of principal components is as large as the original set of variables.

The linear discriminant analysis (LDA) is a classic algorithm for classification, with, as its name suggests, a linear decision surface. The basic idea of LDA is to find a linear transformation, such that the ratio of the between-class scatter and the within-class one is maximized. Samples are projected to a new space with the smallest within-class distance and the largest inter-class one [26]. Although LDA usually gives a good discrimination performance, it suffers from some deficiencies, if the variables are highly correlated or the class boundaries are complex or nonlinear [27]. To avoid such deficiencies, in the former case, variables are often transformed by correlation-reducing methods such as PCA, and in the latter case, LDA could be replaced by QDA.

Let the PC-LDA based classifier of honey consists of n + 3 inputs (n is the number of wave numbers included in honey absorbance spectrum characteristics) and 3 outputs (classes) corresponding to 3 different types of honey (acacia, linden and honeydew). The three additional inputs are designed for the 3 colorimetric indicators (parameters L, a and b) of the CIELab system. The proposed combination of FTIR absorbance spectra with the three colorimetric parameters of CIELab system aims to increase the accuracy of predicting the floral origin of honey. From 2 to 15 PCs were used consistently in LDA based classifiers and by means of a leave-one-out-cross validation test carried out in MATLAB environment the number of PCs was optimized by the “accuracy of prediction / classification” criterion.

RESULTS AND DISCUSSION

FTIR Absorbance Spectra and Colorimetry of Honey

Absorbance spectra of the three brands of honey with wave numbers ranging from 4000 cm$^{-1}$ to 600 cm$^{-1}$ are shown in Fig. 1. The mid-infrared bands could be interpreted using studies previously reported [23, 17, 24]. The first peak (3600 cm$^{-1}$ - 3000 cm$^{-1}$) corresponds to water and carbohydrates O-H stretching, while the two peaks at 2950 cm$^{-1}$ - 2800 cm$^{-1}$ - to carbohydrates (methyl and methylene groups) C-H stretching [17, 24]. In the so called fingerprint region, a complex series of carbohydrate (mainly glucose and fructose) absorption, and H-O-H banding of H$_2$O is observed [17, 24].

The average values and the standard deviations of FTIR spectra of the three types of honey (acacia, linden and honeydew) are shown in Fig. 2. The differences between the classes are evident from the average values (Fig. 2a) of those classes FTIR spectra. But the relatively high values of the standard deviations (Fig. 2b) indicate big scattering of the samples of the classes and eventually their intermixing.

The average values and the standard deviations of the color parameters of the three types of honey are given in Table 1.

PC-LDA Based Models for Honey Discrimination

PCA is carried out in order to visualize the data referring to the different honey samples and to identify their similarities and differences. The spectral dimensionality is reduced to a small number of principal components using PCA. The scores scatter plot of the 1st and 2nd PCs is shown in Fig. 3a. The samples form three clusters (acacia, linden and honeydew), which are overlapped. Here, the type of honey is determined solely on the ground of the inscription on the label by the manufacturer. The two PCs suitably visualize the honey’s spectra,
but the information provided by them is not enough to distinguish properly the different types of honey. Therefore PCA is applied to a combination of FTIR spectra characteristics and the three indicators (L, a, b) of the colorimetric system CIELab. In this case the first two PCs explain as high as 96.49 % of the combined data variance (79.84 % for PC-1 and 16.65 % for PC-2). The result (Fig. 3b) shows a better differentiation between the different types of honey, with a little exception in the classes ‘acacia’ and ‘linden’.

From 2 PCs to 12 PCs (obtained from the enriched data: spectral characteristics + color parameters) are chosen to develop the PC-LDA models. The leave-one-out-cross-validation test is used to check the performance of the classifiers. The results referring to the honey’s botanical origin prediction obtained on the ground of

Table 1. Honey color parameters.

<table>
<thead>
<tr>
<th>CIELab Values (Ill.D65/10 deg Observer 380-780 nm)</th>
<th>Acacia honey</th>
<th>Linden honey</th>
<th>Honeydew honey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average value ± Standard deviation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>92.45±5.46</td>
<td>85.35±6.63</td>
<td>51.27±12.80</td>
</tr>
<tr>
<td>a</td>
<td>0.29±1.78</td>
<td>3.05±4.78</td>
<td>27.94±5.83</td>
</tr>
<tr>
<td>b</td>
<td>32.01±15.00</td>
<td>57.65±16.72</td>
<td>74.54±11.63</td>
</tr>
</tbody>
</table>

Table 2. Discrimination accuracy of PC-LDA model under different number of PCs.

<table>
<thead>
<tr>
<th>Number of PCs</th>
<th>PC-LDA (%)</th>
<th>Number of PCs</th>
<th>PC-LDA (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>86.05</td>
<td>8</td>
<td>97.67</td>
</tr>
<tr>
<td>3</td>
<td>86.05</td>
<td>9</td>
<td>93.02</td>
</tr>
<tr>
<td>4</td>
<td>90.70</td>
<td>10</td>
<td>90.70</td>
</tr>
<tr>
<td>5</td>
<td>93.02</td>
<td>11</td>
<td>93.02</td>
</tr>
<tr>
<td>6</td>
<td>88.37</td>
<td>12</td>
<td>93.02</td>
</tr>
<tr>
<td>7</td>
<td>88.37</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
these classifiers are shown in Table 2. The optimal result (97.67 % accuracy) corresponds to the case presented by 8 PCs. The performance of the validation test of the PC-LDA model (with 8 PCs) is shown in Fig. 4 and Table 3. As evident from Table 3, sample 1 from the class ‘linden’ is predicted as ‘acacia’. The model predicts correctly 42 out of 43 samples.

CONCLUSIONS

A combination of Mid-FTIR, colorimetry and LDA was proposed for distinguishing the botanical origin of honey. The LDA calibration model using 8 PCs predicted correctly 42 out of 43 samples. 97.67 % prediction accuracy (100 % class ‘acacia’, 94.12 % class ‘linden’, and 100 % class ‘honeydew’) was achieved. The high accuracy of the proposed honey classifier was confirmed by a leave-one-out-cross validation test carried out in MATLAB environment. Unfortunately, the reference classes were determined only by means of labeling by the manufacturer. More unifloral types of honey have to be included to improve the performance of the calibrations and the reliability of the results. Furthermore, the reference samples have to be characterized by physico-chemical and pollens analysis.

The obtained results indicated that Mid-FTIR spectroscopy combined with colorimetry and chemometrics can be used as an alternative, quick, and low cost method for the identification of monofloral types of honey.

Acknowledgements

The paper presents research and development, supported by Scientific Fund of Internal Competition of the University of Food Technologies, Plovdiv, under the Research Project No.7/16-H and the Science&Research Programme of the University of Chemical Technology and Metallurgy, Sofia (Project 11 588/2016).

REFERENCES

11. B. Stuart, D.J. Ando, Biological applications of infrared spectroscopy, 1997, Chichester: Published on behalf of ACOL (University of Greenwich) by John Wiley.