Using sensor and spectral analysis to classify botanical origin and determine adulteration of raw honey

Zhilin Gan a, b, Yang Yang c, Jing Li d, Xin Wen a, b, Minghui Zhu a, b, Yundong Jiang a, b, Yuanying Ni a, b, *

a National Engineering Research Center for Fruits and Vegetables Processing, Beijing 100083, China
b College of Food Science and Nutritional Engineering, China Agricultural University, Beijing 100083, China
c Cloud Live Technology Group Co., Ltd., Beijing 100085, China
d Institute of Electrical Engineering of the Chinese Academy of Sciences, Beijing 100190, China

ABSTRACT

The feasibility of sensors (Electronic Nose, EN; Electronic Tongue, ET) and spectra (Near Infrared spectrum, NIR; Mid Infrared spectrum, MIR) to evaluate raw honey samples (Vitex, Jujube and Acacia) was explored. Partial least squares discriminant analysis (PLSDA) model, support vector machine (SVM) algorithms model and Interval partial least squares (iPLS) model were used to classify the botanical origin. The results indicate that spectra and sensors could classify the botanical origin of honey rapidly and accurately, since total accuracy for calibration and prediction sets was all almost 100% in EN and ET analysis by SVM model and in NIR and MIR analysis by iPLS model. Then principal components analysis (PCA) and PLSDA model were used to determine the adulterants. Total accuracy for calibration and prediction sets was all above 96% in NIR, MIR and ET by PLSDA model. The results indicate that ET is more suitable for detecting honey adulteration.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Honey is a natural food product processed by honey bees by blending the sweetened sap collected from flowers with metabolic gastric enzymes (Anjos et al., 2015). Natural honey is a very nutritious food product, which contains water (17% in general), saccharides (main constituents in honey, 75% in general, mainly glucose and fructose), amino acids, minerals (Mg, Ca, K, Na, S, P, Fe, Mn, Co, Ni, etc.), vitamins, enzymes (invertase, catalase, amylase, etc.), phenols, organic acids, pigments, volatile oils and also over 100 varieties of aromatic substances (De la Fuente et al., 2011; Mateo and Boschreig, 1997; Ouchemoukh et al., 2010; Arvanitoyannis et al., 2005; Baroni et al., 2006).

Honey needs to be rapidly evaluated and priced after collection from a bee-house, based on botanical origin classification and adulterant determination. Honey has customarily been distinguished as unifloral and multifloral depending on the botanical origin (Lenhardt et al., 2014). In China, over 20 kinds of unifloral honey are commonly produced such as jujube, vitex, acacia, rape, and linden. In general, unifloral honey types have higher market value due to their limited production and availability. Traditionally, botanical origin was determined by sensory analysis, pollen morphology, characteristic aroma analysis and characteristic interior components analysis. These methods have shown that botanical origin classification of honey is affected by color, aroma, content of pollen, method of processing and storage, content of trace substances and so on. However, these methods are expensive and time-consuming. Although the results of sensory analysis are accurate, it is objective and needs a well trained certified taste panel with a time investment. (Arvanitoyannis et al., 2005; Aparna and Rajalakshmi, 1999; Lawal et al., 2009; Bogdanov et al., 2004; Bianchi et al., 2005; Iglesias et al., 2004; Martos et al., 2000; Gonzalez-Miret et al., 2005; Conti et al., 2007).

For unethical economic gain, honey adulteration is an obvious problem in the market. Water, sucrose, inverted sugar, hydroxyethyl cellulose, dextrin and starch are adulterants which have been regularly identified by regular physicochemical analysis (Serrano et al., 2004). High performance liquid chromatography (HPLC), isotope mass spectrometry and capillary electrophoresis...
have also been used to evaluate these. However, these methods are complicated, time and labor consuming (Morales et al., 2008; Tu et al., 2011; Luo et al., 2012). C-4 botanical glycosides like corn syrup were added to honey subsequently to avoid detection. Even though corn syrup can be detected by stable carbon isotope ratio analysis (SCIRA), it is too expensive and time consuming (Cotte et al., 2007). Moreover, SCIRA is not suitable for detection of C-3 botanical glycoside such as rice syrup.

In recent years, rapid detection techniques such as spectral technology, sensor technology and chemical kits are widely used in testing of species, habitat, freshness, nutrient quality, and drug residues in agricultural products since they are time-saving, more convenient and accurate than traditional methods. Electronic nose (EN) and electronic tongue (ET) are common methods for food analysis (Cozzolino et al., 2008; Wang et al., 2009; Ghasemi-Varnamkhasti et al., 2010; Vaslov et al., 2002; Lu et al., 2014; Pan et al., 2014; Haddi et al., 2013). A sample’s whole information so called fingerprint data, can be determined from EN and ET rather than qualification and quantification of some specific constituents. Then a discriminant model will be established by compiling fingerprint data and unknown samples could be determined by the model. In previous studies, based on volatile components collected by solid phase micro-extraction, botanical origin of honey has been detected by Mass spectrum-electronic nose (MS-EN) (Ampuero et al., 2004). The bp-ANN model founded by EN consisting of ten MOSFET and twelve MOS sensors could discriminate the botanical origin and geographical honey production area (Benedetti et al., 2004). α-Astree-ET could classify botanical origin of honey as well, in which the accuracy of the artificial neural network model was 100% (Major et al., 2011). Also PCA and ANN models were effective in classification of honey botanical origin by using ET with metallic compound electrode (Escriere et al., 2012).

NIR and MIR are extensively performed in agricultural products, food and medicine analysis (Magwaza et al., 2011; Cen and He, 2007; Balabin and Smirnov, 2011). Much research has revealed that NIR could classify the botanical origin of honey by using PCA analysis, canonical variate analysis, PLS discriminant analysis and linear discriminant analysis (Davies et al., 2002; Ruoff et al., 2007). Some reports showed that NIR was also an effective method to discriminate the adulterant of honey. Irish honey adulterated with beet syrup and high fructose syrup has been correctly discriminated and adulterant ratio could be predicted by using PLS analysis (Kelly et al., 2006a). Honey adulterated with different ratio of fructose/glucose was also determined by PLSR, K-NN and SIMCA models (Downey et al., 2003). MIR could discriminate adulterant of honey as well. Honey with glucose, fructose, sucrose and corn syrup added was detected by MIR and the discriminant accuracy could reach 90% by using LDA model (Irudayaraj et al., 2003). In another study, Irish artificial honey adulterated with fully inverted beet syrup, high-fructose corn syrup, partially invert cane syrup, dextrose syrup and beet sucrose was evaluated by MIR and well classified by using SIMCA and PLSAD models (Kelly et al., 2006b).

Compared with previous studies, this work focused on evaluating the performance of sensors (EN and ET) and spectra (NIR and MIR) in botanical origin classification and adulterant determination of raw honey. All honey samples used in this research were raw materials obtained directly from a bee-house and without any processing. Different types of discriminant analysis models were built. Sensors (ET and EN) and spectra (NIR and MIR) showed different accuracy and performance in botanical origin classification and adulterants determination.

2. Materials and methods

2.1. Sample preparation

Three types of honey samples, vitex, jujube and acacia, were provided by Beijing Baihua Apiculture Technology Development Company. Sample information is described in Table 1. Each sample, sourced from a different bee-house was filtered by 60 mesh screen in order to eliminate impurities. Samples were stored in the dark at 4 °C until analysis.

In botanical origin classification, 105 samples were divided into two sets with calibration (79) and prediction (26) as shown in Table 2, respectively. Then 35 pure honey samples (including vitex, jujube and acaica) were chosen randomly to prepare adulterant samples by adding two varieties of syrup (rice syrup and corn syrup). Both syrups were purchased from Cargill Investments (China) Co., Ltd. The physiochemical properties of syrups and honey samples were shown in Table 3. Adulterant samples were prepared by mixing pure honey solution with syrup in different concentrations (5%, 10%, 20% and 40%). The samples, 259 in total, included 105 pure honey samples and 154 adulterated samples. The division of sets was described as follow in Table 4.

2.2. Physical and chemical analysis of honey and syrup

The conductivity was measured by a Orion 5-Star Benchtop Meter (Thermo Fisher Scientific, Waltham, USA). The pH values were tested by a PB-10 pH meter (Sartorius Corp., Bohemia, NY, USA). Proline was evaluated by spectrophotometry. Saccharides, fructose, glucose, cane sugar and maltose were determined by HPLC equipped with a YMC-Pack Polyamine II (250 mm × 4.6 mm, 5 μm) column. All of the physiochemical properties above were conducted by methods modified from previous studies (Bogdanov et al., 2004; Bogdanov, 2002).

2.3. Electronic nose

The EN system used for this study was a Heracles (Alpha Mos, Toulouse, France) which was equipped with a highly selective and sensitive specialty gas chromatograph. Volatile constituents of honey samples were collected by way of static headspace. Separation columns used in this work were DB5 and DB1701, respectively. Retention time and peak area were detected by hydrogen flame ionization detector (FIDs). Three grams of honey sample was weighed into a septa-sealed screw-cap bottle and equilibrated for 10 min at 65 °C (500 rpm). The parameters of EN were as follows: (1) Injector: temperature 180 °C, inject volume 3000 μL. (2) Temperature program: initial 40 °C, final 200 °C, heating rate 2.0 °C/s. (3) Trap temperature: 250 °C, purge time 15 s. (4) Temperature of FID 220 °C. (5) Acquisition time: 84 s (Yüzay and Selke, 2007). All determinations reported were conducted in triplicate.

2.4. Electronic tongue

The sensor array of the α-Astree ET (Alpha Mos, Toulouse, France) was comprised of seven potentiometric chemical sensors

Table 1

<table>
<thead>
<tr>
<th>Variety</th>
<th>Place of origin (different province of China)</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vitex</td>
<td>Liao Ning, Hei Bei, Bei jing, Shan Xi</td>
<td>41</td>
</tr>
<tr>
<td>Jujube</td>
<td>Liao Ning, Hei Bei, Shan Dong, Shan Xi</td>
<td>27</td>
</tr>
<tr>
<td>Acacia</td>
<td>Shan Xi</td>
<td>37</td>
</tr>
</tbody>
</table>
Table 2
Calibration and prediction sets of pure and adulterated honeys in botanical origin classification.

<table>
<thead>
<tr>
<th>Type of samples</th>
<th>Calibration set</th>
<th>Prediction set</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vitex</td>
<td>31</td>
<td>10</td>
<td>41</td>
</tr>
<tr>
<td>Jujube</td>
<td>20</td>
<td>7</td>
<td>27</td>
</tr>
<tr>
<td>Acacia</td>
<td>28</td>
<td>9</td>
<td>41</td>
</tr>
</tbody>
</table>

Table 3
Physicochemical properties of syrups and honey.

<table>
<thead>
<tr>
<th></th>
<th>Conductivity (μS/cm)</th>
<th>pH</th>
<th>Proline (mg/kg)</th>
<th>Fructose (g/100 mL)</th>
<th>Glucose (g/100 mL)</th>
<th>Cane sugar (g/100 mL)</th>
<th>Maltose (g/100 mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corn syrup</td>
<td>20.41</td>
<td>4.45</td>
<td>46</td>
<td>32.66</td>
<td>37.32</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>Rice syrup</td>
<td>15.53</td>
<td>4.54</td>
<td>41</td>
<td>42.60</td>
<td>33.32</td>
<td>0.21</td>
<td>nd</td>
</tr>
<tr>
<td>Honey</td>
<td>265</td>
<td>4.43</td>
<td>295</td>
<td>39.55</td>
<td>28.91</td>
<td>1.49</td>
<td>1.44</td>
</tr>
</tbody>
</table>

Table 4
Calibration and prediction sets of pure and adulterated honeys.

<table>
<thead>
<tr>
<th>Type of samples</th>
<th>Calibration set</th>
<th>Prediction set</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pure honey samples</td>
<td>79</td>
<td>26</td>
<td>105</td>
</tr>
<tr>
<td>Adulterated samples</td>
<td>115</td>
<td>39</td>
<td>154</td>
</tr>
<tr>
<td>Adulterated 5%</td>
<td>22</td>
<td>8</td>
<td>30</td>
</tr>
<tr>
<td>Adulterated 10%</td>
<td>35</td>
<td>12</td>
<td>47</td>
</tr>
<tr>
<td>Adulterated 20%</td>
<td>35</td>
<td>11</td>
<td>46</td>
</tr>
<tr>
<td>Adulterated 40%</td>
<td>23</td>
<td>8</td>
<td>31</td>
</tr>
</tbody>
</table>

3. Results

3.1. Physicochemical properties of honey and syrup

The mean values of the physicochemical parameters are shown in Table 3. Generally, the ratios and constituents of saccharides and pH were very similar in honey and syrup, while conductivity and proline were the more meaningful parameters to evaluate honey and syrup due to the big difference. The difference in conductivity is likely due to natural honey’s high mineral content. (Mateo and Boschreig, 1998).

3.2. Botanical origin classification

3.2.1. PLSDA model of EN, ET, NIR and MIR

PLSDA model of EN was established based on discriminant analysis results of seven principal components by using 20 sensors. PLSDA model of ET was found according to the results from seven sensors’ evaluation. Five principal components were then chosen to optimize the results.

PLSDA model of MIR was built by discriminant analysis results of ten principle components.

PLSDA model of NIR was established by discriminant analysis results of five principle components.

The PLSDA model results of these four methods were shown in Table 5. It was shown that the error rate was less than one in every method, both in calibration set and prediction set. In the MIR study, the accuracy of calibration and prediction reached 100% for all samples. Compared with MIR, misjudgment and outliers were still present in the NIR, EN and ET studies. The results indicated that MIR performed best using the PLSDA model.

3.2.2. SVMDA model of sensors

Type C of SVM was applied in this study and the kernel function chosen was RBF (Radial basis function). Cost and gamma, the main parameters of this function, were 100, 0.001 in EN and 100, 0.0031623 in ET. Also SVs was 28 and 20 in EN and ET, respectively. Although the results shown in Fig. 1 confirmed that one Vitex sample and one Acacia sample which came from the calibration set misjudged Acacia and Jujube in EN and ET, respectively, yet no errors occurred in prediction set. This means the SVM model can raise the accuracy of botanical origin classification when compared with PLSDA model in EN and ET analysis. Traditionally, PCA, artificial neural network (ANN), linear discriminant analysis (LDA), discriminating factor analysis (DFA) were the best common models applied to EN and ET data for identification of honey botanical origin (Huang et al., 2014; Ampuero et al., 2004; Tiwari et al., 2013;
3.2.3. iPLSDA model of spectra

iPLSDA model was run to optimize the botanical origin classification results by using NIR and MIR. The spectra regions of NIR and MIR were divided into different sections. 6310–5847 cm\(^{-1}\) for NIR, 3397–3298 cm\(^{-1}\), 2893–2592 cm\(^{-1}\) and 1381–980 cm\(^{-1}\) for MIR were the optimized region. iPLSDA models of NIR and MIR were found by discriminant analysis results of six principle components. The accuracy was 100% for both the calibration and prediction sets. Compared with a previous study (Chen et al., 2012) which used the BP-ANN model for NIR analysis, iPLSDA model in this work more accurately distinguished the botanical origin of Acacia, Vitex and Jujube. Fig. 2 is the Scatter plot for discriminating honey types by PLSDA based on NIR data ((a) Acacia (b) Jujue (c) Vitex). Only one error occurred in each type of honey. The results indicated that iPLS could optimize the spectra region which reduced computational cost. Moreover, the discrimination among the samples was improved due to the more obvious distribution of different samples. In the optimized region of NIR, O–H secondary multiple frequency, C–H secondary multiple frequency and combination band were the predominant characteristics. C–H, O–H, H\(_2\)O, C–O and C–C stretching bands as well as O–C–H, C–C–H and C–O–H were the prominent information in MIR.

3.3. Adulteration determination

3.3.1. PCA analysis

The PCA scatter plot of EN is shown in Fig. 3 (a). Adulterated samples and pure honey samples could not be well discriminated by EN due to the crossover although they were separated to some degree. Also it was inferred that obvious difference of volatile components among part of the samples existed because of a few discrete points. There are two reasons for this difference: (1) constitution of volatile substance is different in three kinds of pure honey, even in different samples of the same botanical origin. Every sample was derived from different bee-house, so it could be influenced by the natural environment. (2) The fragrance characteristic has been changed after adding the syrup. So it is difficult for EN to classify adulterant samples since the volatile components are complicated.

ET-PCA which is described in Fig. 3(b) shows that pure honey could be divided into three groups according to different botanical origin. The corresponding adulterants were discriminated as three groups as well. Additionally, pure honey and adulterant could be clearly distinguished by ET.

As shown in Fig. 3(c), pure honey and adulterant were uniformly distributed leading to NIR-PCA’s inability to classify adulteration of honey. Both division and crossover existed in Fig. 3(d). This means pure honey and adulterant could not be accurately discriminated by MIR in PCA analysis.

It was concluded that honey samples of this study could be better discriminated by ET, compared using EN, NIR and MIR in PCA.
analysis, no matter whether among adulterant samples, or between pure honey and adulterant. Saccharide intermolecular bonds were the predominant bond type present in the spectra. Based on the result of 3.1, the ratio and constituent of saccharides and pH were very similar in honey and syrup. So adulterant could not be determined by PCA analysis in NIR and MIR. More pretreatment of data and foundation of discriminant models are needed. Combined with 3.1, difference of conductivity between pure honey and syrup is mainly because more minerals exist in pure honey. ET is more sensitive to minerals. Furthermore, there is gustatory difference between pure honey and adulterants which can be exactly classified by the sensors of ET. In brief, ET is more suitable for detecting honey adulteration.

3.3.2. PLSDA models

Fewer than 105 and 154 pure honey and adulterant samples, respectively, were usable due to outliers.

The results from the PLSDA model for EN are shown in Table 6. Six principle components were collected in EN analysis and outliers were eliminated. Misjudgment still existed in calibration and prediction. Thus, EN was probably not suitable to determine the adulteration of honey based on the PLSDA model. The reason could be as follows: (1) Three different kinds of honey collected in this study were substantially different in terms of volatile components. (2) Even samples of same origin which were collected from different area were obviously different according to the EN footprints. (3) The “smell” affected by freshness of samples was different due to the collecting and processing time.

ET-PLSDA model was established by five principle components after eliminating three abnormal adulterant samples. The total discriminant accuracy of calibration and prediction were 98.43% and 100%, respectively, as shown in Table 6. That means adulterant can be accurately determined by combining the ET-PLSDA model with the results from ET-PCA, which can be explained by syrup’s lack of trace substances such as acids, phenols, enzymes, minerals and amino acids neither such physicochemical properties like pH, conductivity which all exist in honey and are easily detected by ET. However, a previous study (Zakaria et al., 2011), demonstrated that single modality assessment of EN and ET was ineffective to discriminate the adulterated honey by PCA and LDA.

Table 7 shows PLSDA results for discriminating honey adulteration using NIR and MIR by pretreatment of SNV + 2nd derivative + 15 point smooth + auto-scale. Except for one misjudgment that appeared in calibration set of MIR, the accuracy of other sets all reached 100% as shown in Table 7. SNV + 2nd derivative + 15 point smooth + auto-scale performed best and was selected to build the

Fig. 2. Scatter plot for discriminating honey types by PLSDA based on NIR data. (a) Acacia; (b) Jujube; (c) Vitex.
Fig. 3. PCA score plot of pure and adulterant honeys based on EN, ET, NIR and MIR data. (a) EN (b) ET (c) NIR (d) MIR.

Table 6
PLSDA results for discriminating honey adulteration using EN and ET data.

<table>
<thead>
<tr>
<th>Type of samples</th>
<th>EN</th>
<th>ET</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Calibration</td>
<td>Prediction</td>
</tr>
<tr>
<td></td>
<td>Recognition ratio % Misjudged/Total</td>
<td>Recognition ratio % Misjudged/Total</td>
</tr>
<tr>
<td>Pure honey</td>
<td>95.95</td>
<td>3/74</td>
</tr>
<tr>
<td>Adulterated total</td>
<td>93.52</td>
<td>7/108</td>
</tr>
<tr>
<td>Adulterated 5%</td>
<td>95.24</td>
<td>1/21</td>
</tr>
<tr>
<td>Adulterated 10%</td>
<td>93.94</td>
<td>2/33</td>
</tr>
<tr>
<td>Adulterated 20%</td>
<td>87.88</td>
<td>4/33</td>
</tr>
<tr>
<td>Adulterated 40%</td>
<td>100</td>
<td>0/21</td>
</tr>
</tbody>
</table>

Table 7
PLSDA results for discriminating honey adulteration using NIR and MIR data.

<table>
<thead>
<tr>
<th>Type of samples</th>
<th>NIR</th>
<th>MIR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Calibration</td>
<td>Prediction</td>
</tr>
<tr>
<td></td>
<td>Recognition ratio % Misjudged/Total</td>
<td>Recognition ratio % Misjudged/Total</td>
</tr>
<tr>
<td>Pure honey</td>
<td>100</td>
<td>0/79</td>
</tr>
<tr>
<td>Adulterated total</td>
<td>100</td>
<td>0/115</td>
</tr>
<tr>
<td>Adulterated 5%</td>
<td>100</td>
<td>0/22</td>
</tr>
<tr>
<td>Adulterated 10%</td>
<td>100</td>
<td>0/35</td>
</tr>
<tr>
<td>Adulterated 20%</td>
<td>100</td>
<td>0/35</td>
</tr>
<tr>
<td>Adulterated 40%</td>
<td>100</td>
<td>0/23</td>
</tr>
</tbody>
</table>
models compared to SNV + auto-scale and SNV+1st derivative+15point smooth + auto-scale. It is probably because 2nd derivative can amplify the difference between adulterated honey and pure honey. Thus, PLSDA model is suitable for NIR and MIR. In a previous study (Zhu et al., 2010), NIR combined with chemometrics methods has been used to detect adulteration of honey samples. Least square support vector machine (LS-SVM) model was established and the total accuracy was 95.1%. Our model has more potential due to its higher accuracy.

4. Conclusions

Botanical origin of honey could be rapidly determined by EN, ET, NIR and MIR. Compared with PLSDA, iPLS and SVM could improve methods has been used to detect adulteration of honey samples. Data processing is simpler when using ET as well. Additive, ET is more sensitive to substances like amino acids, minerals, phenols, monosaccharides and disaccharides which exist in honey shown with NIR and MIR. So it can be inferred that ET may lead to more extensive application for determination of honey adulterants. Further research is needed because few studies have published on this topic.

Acknowledgement

We thank the Beijing Baiahua Apiiculture Technology Development Company for providing the raw honey. Also, we are grateful to Prof. Han, Dr. Zhou, and Dr. Lv, who all work in College of Engineering, China Agricultural University, and Bruker Optics, Inc., Germany, for guiding us in the use of NIR and MIR. We acknowledge funding of Natural Science Foundation of China (NO. 31171675) and Special Fund for Agric-research in the Public Interest (No. 20150314-01) from the Ministry of Agriculture of the People’s Republic of China.

References

Bogdanov, S., 2002. Harmonised methods of the international honey commission. Swiss Bee research Centre, FAM, Liebefeld, CH-3003 Bern, Switzerland.
honey characterization and botanical classification by an electronic tongue. Talanta 85, 569–574.