Food Control 134 (2022) 108770

Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont

Honey discrimination based on the bee feeding by Laser Induced Breakdown Spectroscopy

Dimitrios Stefas [a,b], Nikolaos Gyftokostas [a,b], Panagiotis Kourelias [a,b], Eleni Nanou [a,b], Chrysoula Tananaki [c], Dimitrios Kanelis [c], Vasileios Liolios [c], Vasileios Kokkinos [d], Christos Bouras [d], Stelios Couris [a,b,*]

[a] Department of Physics, University of Patras, 26504, Patras, Greece
[b] Institute of Chemical Engineering Sciences (ICE-HT), Foundation for Research and Technology-Hellas (FORTH), Patras, 26504, Patras, Greece
[c] Laboratory of Apiculture-Sericulture, School of Agriculture and Forestry, Aristotle University of Thessaloniki, 54124, Thessaloniki, Greece
[d] Department of Computer Engineering & Informatics, University of Patras, 26504, Patras, Greece

ARTICLE INFO

Keywords:
LIBS
Laser-induced breakdown spectroscopy
Machine learning
Honey
Bee feeding
Classification
Linear discriminant analysis
Random forests

ABSTRACT

In the present work, Laser Induced Breakdown Spectroscopy (LIBS) is used for the first time to investigate the effects of artificial feeding of bees on the honey. According to LIBS technique the emission spectral characteristics of the plasma created on the surface of honey samples are analyzed. Correlation plots indicating the importance of spectral lines of elements as e.g., Calcium (Ca), Magnesium (Mg), Sodium (Na) and Potassium (K) are constructed. In addition, machine learning algorithms based on Linear Discriminant Analysis (LDA) and Random Forest Classifiers (RFC) are employed to classify the honey samples in terms of the bee food used. The constructed machine learning models were validated by both cross-validation and external validation, while the obtained accuracies exceeded 90% of correct classification, indicating the potential of LIBS technique for honey discrimination. The obtained results by LIBS were also validated by HPLC-RID, which is the standard technique used for the analysis of the main honey sugars.

1. Introduction

According to the European legislation, honey is the natural sweet substance produced by *Apis mellifera* bees from the nectar of plants or from secretions of living parts of plants or excretions of plant-sucking insects on the living parts of plants, which the bees collect, transform by combining with specific substances of their own, deposit, dehydrate, store and leave in honeycombs to ripen (Directive 2001/110/EC). This long definition determines the honey product and distinguishes it from any other sweet products that can be made artificially. The composition of honey is variable and is dependent primarily on its floral source and, also, on certain external factors, such as the beekeeping manipulations, packaging and storage conditions. Honey is mainly composed of complex carbohydrates, such as fructose, glucose, maltose and sucrose, as well as water. Other constituents are enzymes, organic-, amino- and phenolic acids, vitamins, volatile compounds, flavonoids and minerals (Da Silva et al., 2016). The main minerals present in honey originate from soil and are transported to plants by the roots. The minerals make their way into the nectar and honeydews and are incorporated into the honey produced by bees (Wang et al., 2006; Stankovska et al., 2008; Liolios et al., 2016). Thus, they are related to the type of bees' feeding (e.g., botanical origin) as well as to environmental and geographical factors. The major minerals are Ca, Mg, Na, K, Cl, P, S and minor/trace minerals include Zn, Al, Mn and Cu (Da Silva et al., 2016; Harvey, 2016; Moreno-Rojas, Cámara-Martos, & Amaro López, 2016 a; Moreno-Rojas, Cámara-Martos, & Amaro López, 2016 b). Since the inorganic content of honey affects directly the physicochemical properties of honey (e.g., conductivity) is widely used for its botanical origin discrimination (Rodopoulou et al., 2017, 2021; Rodriquez et al., 2019; Serrano et al., 2006). Because it is a high nutritional value product, it can be adulterated during its production or processed by adding sugars or low-price honey with the aim of economic benefits. Honey adulteration can be mainly achieved either with the addition of sweeteners directly to honey during the process of packaging or by feeding syrups to honeybees during the main nectar flow period.

As the nectar and honeydew flows vary strongly during the

* Corresponding author. Department of Physics, University of Patras, 26504, Patras, Greece.
E-mail address: couris@upatras.gr (S. Couris).

https://doi.org/10.1016/j.foodcont.2021.108770
Received 18 July 2021; Received in revised form 8 December 2021; Accepted 12 December 2021
Available online 15 December 2021
0956-7135/

beekeeping year, sometimes bees need additional feeding to maintain breeding activities and cover food requirements. The traditional supplement for honeybees feeding is a sucrose solution, in which sugar and water are mixed in a 1:1 ratio, but the last years many sweeteners and sugar syrups are commercially available for feeding bees. These substitutes are made from sugar cane, sugar beet, starch produced from corn, wheat, rice, or syrups of natural origin such as maple. The carbohydrate feeding of bees is divided into two types depending on the physical stage of the supplied product, solid and liquid. The first one is mainly used through winter when the colony is running short of stored honey and the second in all other cases that require complementary food for the nutritional needs of bees. Sometimes for the development of the brood, especially in situations of absence of pollen, protein supplements are required to be provided to the colonies. For the detection of adulteration in honey, chromatographic techniques have been developed such us thin layer chromatography (Puscas et al., 2013), gas chromatography−mass spectrometry (Ruiz-Matute et al., 2007), high-performance liquid chromatography (Morales et al., 2008; Zhou et al., 2014; Wang et al., 2015). In some cases, characteristic compounds as adulteration markers were found (Megherbi et al., 2009; Xue et al., 2013). Alternatively, methods based on spectroscopy, notably nuclear magnetic resonance (NMR), have been proposed (Bertelli et al., 2010; Spiteri et al., 2015). One of the most powerful techniques for detection of honey adulteration is the isotope ratio mass spectrometry (IRMS), which is based on the abundance of stable isotopes of carbon (13C/12C). With IRMS, honey adulteration using C4 sugars can be detected at levels ≥7% but the adulteration from C3 plants cannot be found (Cabañero et al., 2006; Elflein & Raezke, 2008; Guler et al., 2014). Screening techniques such as vibrational spectroscopic methods (near- and mid-infrared) in combination with multivariate data analysis provide rapid and low-cost analyses for the detection of honey adulteration (Das et al., 2017).

The aforementioned techniques are quite accurate, however they use advanced and quite complicated instrumentation, while they require well trained and specialized personnel and they are time-consuming. Very recently, a laser-based technique, Laser-Induced Breakdown Spectroscopy (LIBS) has been proposed and employed for the analysis of honey (Nespeca et al., 2020; Stefas et al., 2020; Zhao et al., 2020; Lastra-Mejías et al., 2020; Peng et al., 2020). LIBS technique utilizes a powerful enough laser beam to induce breakdown on a sample by creating a plasma. The plasma consists of excited atoms and molecules, electrons and ions and emits radiation, containing information on the sample's elemental content (Miziolek et al., 2008; De Giacomo & Hermann, 2017). LIBS can provide information about a sample's elemental composition in real time, independently of the physical state of the sample, as the plasma can be formed in any state of matter and without time-consuming pretreatment. So, in the works of Nespeca et al. (2020), Lastra-Mejías et al. (2020) and Peng et al. (2020), LIBS has been employed for the detection and evaluation of adulteration of honey after it has been mixed with sweetener syrups (as e.g., rice syrup, sugar cane syrup and high fructose corn syrup). More recently, LIBS has been applied for the elemental analysis of honey samples and for the discrimination of the honey samples in terms of their floral origin (Stefas et al., 2020).

In the present work, LIBS is used for the first time, to best of our knowledge, for the detection of the effects of external feeding of honey bees with different syrups on the honey, by analyzing the emission of the laser-produced plasma created on honey samples. The LIBS results are compared with results for the main honey sugars obtained from a High-Performance Liquid Chromatography (HPLC-RID) method. In that view, honey samples stemming from different beehives (where feeding has been performed with inverted syrup, sugar water and candy paste) were analyzed by LIBS and the effects of bee-feeding on the honey LIBS emission spectra are examined and evaluated. Furthermore, predictive models were developed employing Linear Discriminant Analysis (LDA) and Random Forests (RF) machine learning algorithms, using the acquired LIBS emission spectra, aiming to classify the honey samples based on the type of bee feeding that has been employed. In that context, the former algorithm, allowed for the visualization of the class formation, while the latter algorithm has allowed to assess and select the most important spectral features to achieve the most efficient classification of the honey samples. The results showed that the effect of bee-feeding on the honey can be successfully correlated with LIBS spectral features of the inorganic ingredients of honey, such as sodium (Na), calcium (Ca), magnesium (Mg) and potassium (K), while the spectral lines of these elements were identified as the most important for classification purposes leading to accuracies, as high as 99%.

2. Experimental

2.1. Honey production

The honey samples were produced during the period July–August 2019. For the needs of the present investigation, eighteen healthy beehives were divided into six equivalent groups in population and brood according to the feeding protocols given in Table 1. These six groups were treated according to the type of bee food used as classes: commercial inverted syrup (fructose, glucose, maltose), sugar (sucrose) and candy paste (sucrose), whereas the "Control" class corresponds to honey samples from bees that only collected nectar.

Pieces of combs with ripe honey were collected immediately after the sealing of the honey cells, as well as one month after the first collection. However, honey from bee-colonies that fed with candy paste, was not possible to collect at the last collection because the second collection took place after the blooming season and the colonies consumed the reserves of honey. The information about the samples collected from the second collection are also listed in Table 1. The honey samples were obtained by pressing the combs and filtration to remove possible wax fragments. In total, from the two sampling, 63 samples were collected and stored at −18 °C, until their analysis.

2.2. LIBS setup

A small quantity (i.e., few grams) of honey was placed in a Petri-type glass recipient and was left to form a flat surface. The recipient was thereafter placed on a x-y-z translational stage, allowing its x-y movement and the adjustment of its z-position around the focus of a laser beam perpendicularly incident on the free sample surface.

The plasma was created on the free surface of the sample by focusing the laser beam from a 5 ns Q-switched Nd: YAG laser (Quanta-Ray INDI, Spectra Physics) with a 15 cm focal length quartz lens. The laser was operating at its fundamental wavelength, at 1064 nm, with an energy per pulse adjusted at about 70 mJ. The plasma-emitted radiation was collected by means of a 5 cm focal length quartz lens and it was introduced into a quartz optical fiber bundle coupled to the entrance slit (10 μm width) of a 75 mm focal length spectrometer (AvaSpec-ULS4096CL-EVO (CMOS)) for spectral analysis. The spectrometer was equipped with a 300 lines/mm diffraction grating, and a 4096 pixels CMOS detector. From these pixels, the 2754 of them were used here, corresponding to the 200–1000 nm spectral region. A time delay (t_d) of 1.28 μs, and a gate width (t_w) of 1.05 ms were used for the temporal gating of the detector. The LIBS measurements were performed according to the following procedure: LIBS spectra of ten successive laser shots obtained at one location on the sample's surface were averaged, providing one LIBS measurement. Then, 50 such LIBS measurements were collected at different positions on the sample's surface and were employed for the subsequent statistical analysis.

2.3. Sugars analysis

The main honey sugars, fructose, glucose, sucrose and maltose were determined using an HPLC-RID method (Bogdanov et al., 1997). Briefly,

Table 1
Feeding protocols of the bees and honey samples.

Group	Bee-Feeding	Quantity	Number of training samples (1st collection)	Number of validation samples (1st collection)	Samples from the 2nd collection (after 1 month)	Class
A	inverted syrup (commercial)	2L/day	3	2	6	inverted syrup
B	inverted syrup (commercial)	0,5L/day	3	3	6	inverted syrup
C	sugar syrup (in the ratio of 1:1)	2L/day	3	3	6	syrup
D	sugar syrup (in the ratio of 1:1)	0,5L/day	3	2	6	syrup
E	candy paste (commercial)	constant presence	3	2	-	candy paste
F	No feeding	bees collect only nectar	4	2	6	control

honey mixed with methanol:water (25:75, v/v) and the solution filtered through a disposable syringe filter 0.45 μm, before the injection. The sugars were separated on a Zorbax Carbohydrate Analysis (4.6 mm ID x 150 mm × 5 μm) using a mixture of acetonitrile:water (75:25, v/v) as mobile phase at flow rate 1.8 ml/min. For the quantification, a five-point calibration curve was created and evaluated for each sugar. The resulting sugar concentrations were analyzed by multivariate analysis of variance (MANOVA) to determine statistical significances between the samples and by Linear Discriminant Analysis in order to visualize class separation via dimensionality reduction.

2.4. Data analysis

For the analysis and classification of the LIBS spectroscopic data collected from the different honey samples, two machine learning algorithms were selected, namely the Linear Discriminant Analysis (LDA) and the Random Forest Classifier (RFC). Both algorithms were used for classification purposes through the construction of predictive models. Specifically, LDA was used for the supervised visualization of the multidimensional spectroscopic (LIBS) data in a lower dimensions space, while RFC was employed for the estimation of the most important spectroscopic features regarding the discrimination of the samples as well. In more details, LDA maximizes the ratio of the between-class variance over the within-class variance through a classifier with a linear decision boundary, generated by fitting class conditional densities to the data and using Bayes' rule. The LDA model fits a Gaussian probability density to each class, with the assumption that all classes share the same variance-covariance matrix (Duda, Hart & Stork, 2000). On the other hand, RFC, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest gives a prediction for a class and the class with the most votes becomes model's final prediction. The success key of the RFC algorithm lays in the fact that many relatively uncorrelated trees, operating as a committee, will outperform any of the individual constituent trees (Breiman, 2001). The data analysis was performed using the Python programming language along with the Scikit-Learn machine learning library (Pedregosa et al., 2011).

The selection of LDA and RFC algorithms for the processing of the raw LIBS spectroscopic data was based on the following arguments. At first, both algorithms have been previously used (Stefas et al., 2020) successfully for honey analysis via LIBS, achieving high accuracies and, thus, can be considered as suitable for handling such spectroscopic data. Furthermore, both algorithms provide, intrinsically, unique methods for visualizing the classification procedure. So, LDA offers dimensionality reduction of the original data, resulting in a scatter plot depicting the samples' classification into their respective classes, while RFC offers the ability to check which variables (i.e., spectral features) are the most important for the classification procedure (Louppe et al., 2013). As a result, the combination of LDA and RFC can provide a robust methodological approach to inspect the classification of high-dimensional data efficiently and easily, as well as to construct the optimum predictive model based on the most important features.

It is important to note, that for the estimation of the predictive models' accuracy and robustness, both, internal and external validation procedures were performed. For the former, a k-fold cross-validation (where k = 10) was implemented, where the training data were shuffled and split into k groups. The k-1 groups were used to train the classifier, while the remaining group was used for prediction and for assessment of the classifier's accuracy. This procedure was exhaustively repeated k-times, so that each one of the k sub-samples was used for prediction. In that way, the classifier's overall accuracy was computed within the standard deviation of each fold. For the external validation, the classifier was used to predict a second dataset with samples that have not been used to train the algorithm. In that spirit, the ability of the algorithm to generalize and predict never seen before samples can be estimated directly, ensuring the robustness of the procedure. When the classifier predicted new samples, the prediction accuracy was used to evaluate its performance and a confusion matrix was created to examine which samples have been misclassified.

3. Results and discussion

3.1. Bee-feeding effect on the LIBS spectra

The collected LIBS spectra were, at first, analyzed by identifying the most important spectral features. The most important spectral lines appearing in the LIBS spectra of the studied honeys were those of Hydrogen (H), Carbon (C), Oxygen (O), Nitrogen (N), Calcium (Ca), Magnesium (Mg), Potassium (K) and Sodium (Na). According to previous works (Nespeca et al., 2020; Stefas et al., 2020; Zhao et al., 2020; Lastra-Mejías et al., 2020; Peng et al., 2020), the most important features (in terms of feature importance and for discrimination between different samples as well) are those of the inorganic ingredients, i.e., K, Ca, Na, Mg, etc. In Fig. 1, a representative honey LIBS spectrum is presented as an example, showing the most important spectral features observed, while the insets present enlarged views of the spectral regions around the important spectral lines observed. The shown spectrum has been normalized here only for visualization purposes. The identification/assignment of these spectral features was performed based on previous works (Stefas et al., 2020) and using the NIST database (Kramida et al., 2020) as well.

Using the peak intensities of some of the stronger spectral lines of interest for honey's characterization, the presence of any correlation between the samples was searched initially, by constructing the corresponding correlation plots, as e.g., those presented in Fig. 2. The utility of such plots is based on the fact that they can help the visualization of samples' correlation with one another. For instance, in Fig. 2, the correlation plots concerning the potassium-sodium (K–Na), and calcium-magnesium (Ca–Mg) spectral lines are shown. Specifically, in Fig. 2a, the peak intensities of the K (I) 766.5 nm and the Na (I) 589.6 nm

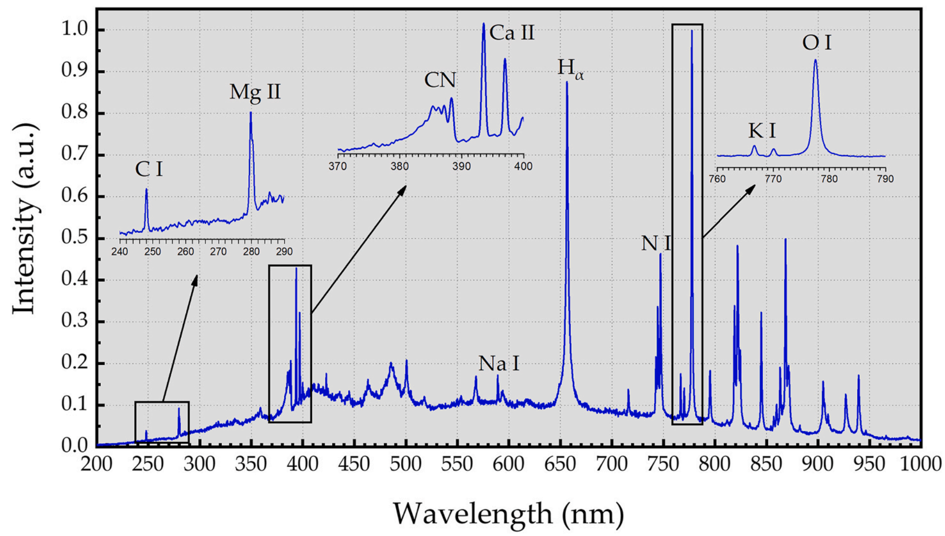

Fig. 1. A representative LIBS spectrum of honey. The spectral regions including some of the most important emission lines are shown enlarged.

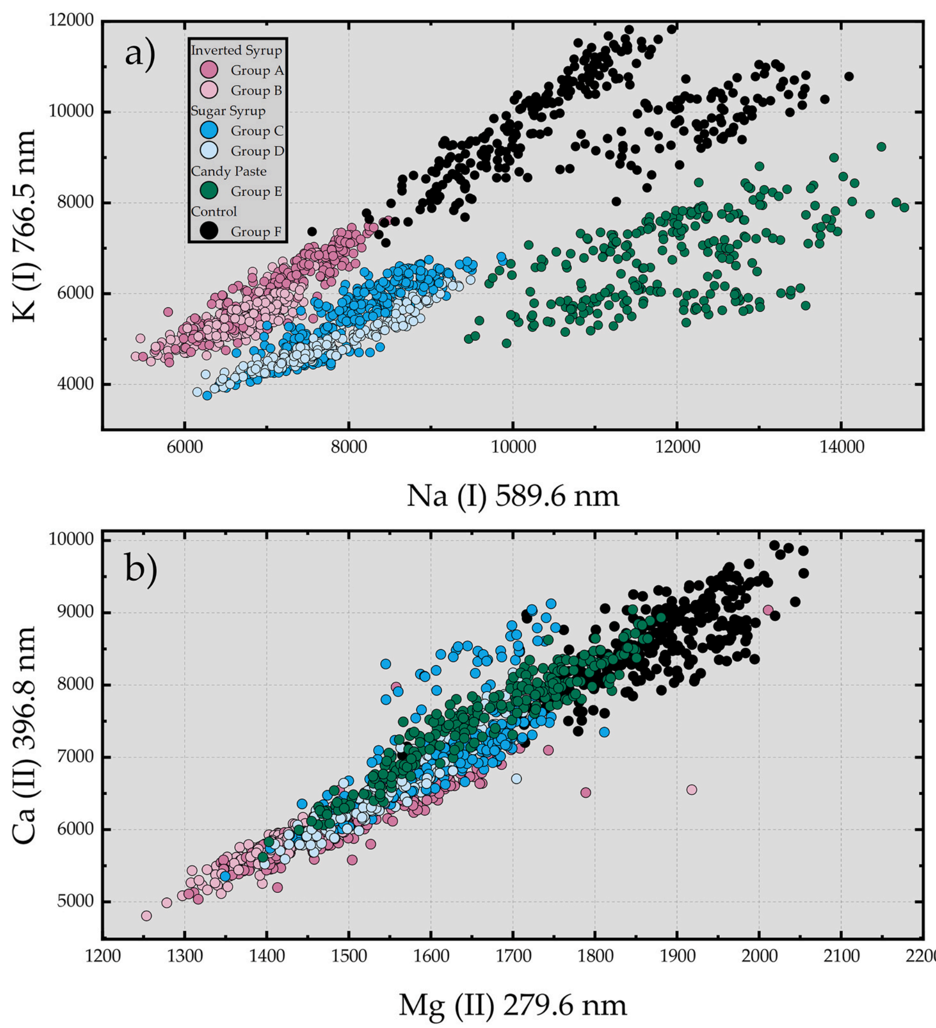

Fig. 2. Correlation plots of the intensities of a) Na (I) 589.6 nm and K (I) 766.5 nm, and b) Mg (II) 279.6 nm and Ca (II) 396.8 nm spectral lines of honey samples from different groups.

spectral lines were used for the construction of the corresponding correlation plot. As can be seen, data clusters are formed and are clearly observable, allowing for a rapid visual classification of the four types of honey samples corresponding to the four different kinds of bee-food. So, it can be seen that syrup-fed colonies' honey samples (groups A, B and C, D) exhibited in general spectral lines with lower peak intensities than the others, while the candy paste-fed colonies' honey samples (group E) and the control samples (group F) exhibited higher peak intensities. In fact, the control samples have been found to exhibit the highest intensities among the different samples' categories while they appear to form two distinct data clusters. This last observation can be attributed to the differences between the control samples originating from different beehives. Interestingly, the honey samples collected after feeding the bees with inverted syrup and syrup, presented, in general, weaker intensity spectral features compared to those of the control samples and the candy paste samples, forming two distinct clusters as well. These observations agree with other studies, employing different methods and analytical techniques, concluding that the artificial bee-feeding results to lower concentrations of such mineral elements (Özcan, Arslan & Ceylan, 2006; Moumeh et al., 2020).

Similarly, in Fig. 2b the correlation plot of the Ca (II) 396.8 nm and Mg (II) 279.6 nm spectral lines are presented. As can be seen, the control samples exhibited in general stronger peak intensities than the other samples, being clearly distinguishable from all them. However, no significant difference has been observed arising from the quantity of bee-feeding, i.e., between groups A and B, as well as C and D, respectively. So, in principle, the method is capable of distinguishing between honeys produced after bee-feeding with artificial sugars from the honey arising without bee-feeding, where bees collect only nectar.

Indeed, by analyzing the samples with HPLC, the groups corresponding to the feeding of the bees with syrups and candy paste were separated from the control group (Fig. 3). According to HPLC results, the statistical analysis showed that honey samples collected from groups with any artificial feeding had significantly different sugars concentration than the control group's samples. More specifically, comparing the control group samples, with those from colonies fed with candy paste and sugar syrup presented significantly higher concentration of sucrose, while the commercial inverted group also had significantly higher maltose content (Guler et al., 2007). For modeling the differences among different feeding conditions Linear Discriminant Analysis (LDA) was used. LDA showed two significant canonical variables. The first one accounted for 98.0% of the between-class variance and it was principally under the dependence of fructose content, while the second accounted for 1.9% and could be related to glucose and sucrose concentration.

To further examine the effect of the bee-feeding on the LIBS spectra, the spectra of the two syrups employed for the bee-feeding, i.e., the inverted syrup one, the homemade sugar water one (i.e., Syrup) and the candy paste, are shown in Fig. 4a and are compared with the spectra of some honey samples collected from beehives fed with these syrups, as well as candy paste. In Fig. 4b, the enlarged views of some spectral regions of the spectra from Fig. 4a are shown together with the spectra. As shown, the LIBS spectrum of honey differs significantly from the syrups' spectra, as the former only exhibited the spectral lines of Ca (II), Na (I) and K (I).

3.2. Classification of honeys' LIBS spectra after different bee-feeding

From a total of 33 samples, collected from various beehives (see also in Samples' section), 19 were used for the training of the machine learning algorithms, while the remaining 14 samples were employed to validate the machine learning models that have been constructed for classification purposes (see also Table 1).

In the case of the LDA algorithm, three canonical variables were used for training. The obtained accuracy, after a 10-fold cross-validation, was found to be as high as (99.6 $\pm$ 0.7)%. In Fig. 5, the scatter plot resulting from the dimensionality-reduced dataset via the LDA algorithm is shown. In this plot, each point represents a LIBS spectrum, embedded in the three-dimensional space spanned by the three canonical variables used. As can be seen, four clearly distinct data clusters were formed corresponding to the four different types of bee-feeding, reflecting the successful classification by simple visual inspection. Next, the LDA model was used for predicting the class designation of the LIBS spectra obtained from the test/validation samples. Remarkably, the predicted accuracy of the unknown data was found to attain a very high value of 98.4%. As can be seen from the confusion matrix presented in Fig. 6a, where misclassifications are highlighted by circling the corresponding numbers, no instances of the control samples were misclassified, while only a couple of instances, i.e., LIBS spectra, were misclassified. It should be noted that, for each sample, 50 LIBS independent spectra were acquired (see also in the **Materials and Methods Section**, the **2.2. LIBS Setup** and **2.3. Data Analysis sub-sections**). For example, only 4 inverted syrup spectra were misclassified as syrup spectra and vice-versa, while 3 candy paste spectra were mislabeled as syrup spectra.

Next, the RFC algorithm was utilized to develop another predictive model, in the same spirit as in the previous case for the LDA algorithm. This RFC model fitted a hundred decision tree estimators on various subsamples of the dataset and used some averaging to improve the predictive accuracy and control overfitting. Furthermore, a 10-fold cross-validation was performed, as well. The resulting accuracy was determined to be (97.6 $\pm$ 1.6)%, while the classification prediction accuracy of the unknown data attained a value of 97.0%. In the corresponding confusion matrix presented in Fig. 6b, the misclassified samples are shown. As can be seen from this confusion matrix, mislabeling was found to occur between 11 instances of syrup which were

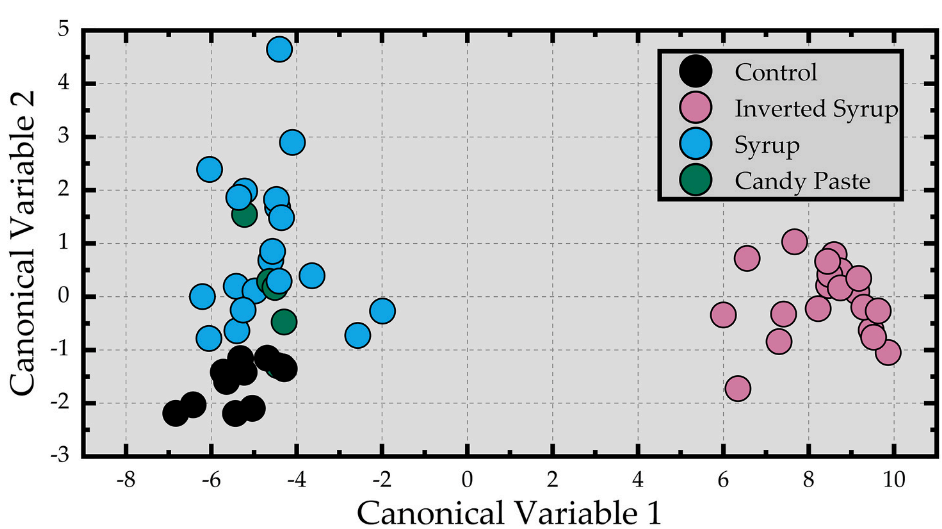

Fig. 3. Discriminant Analysis of different feeding groups.

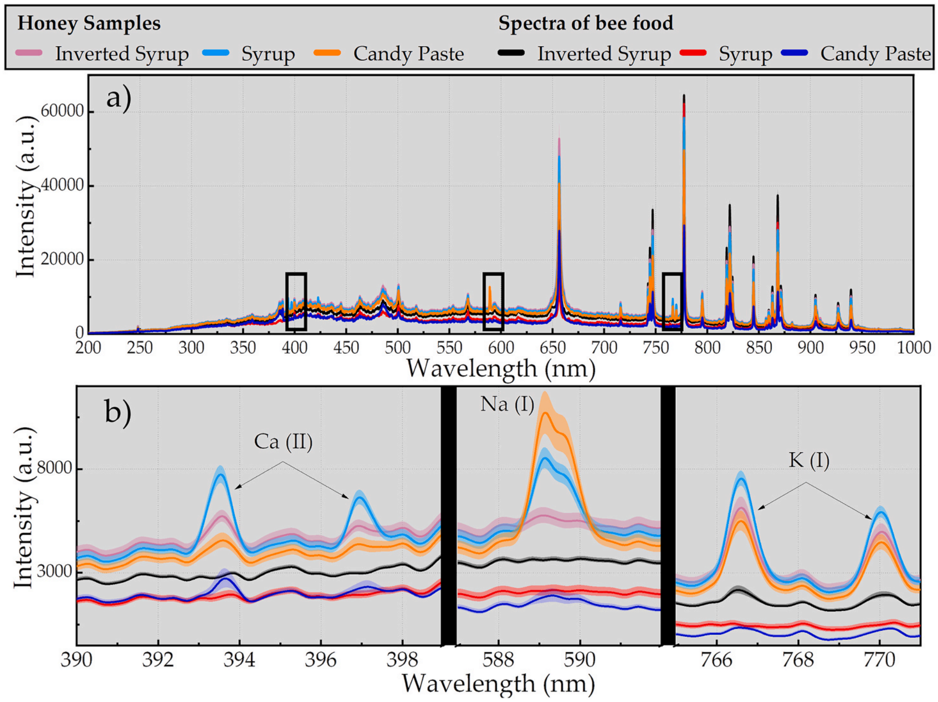

Fig. 4. a) LIBS spectra of honey samples from bees fed with inverted syrup, sugar water syrup and candy paste. For comparing the honey spectra with the spectra of the bee food, the spectra of inverted syrup, sugar water syrup and candy paste are shown, as well. b) Enlarged views of the aforementioned spectra within the spectral regions 390 nm–399 nm, 587 nm–592 nm and 765 nm–771 nm, where the spectral lines of Ca (II), Na (I) and K (I) are shown, respectively.

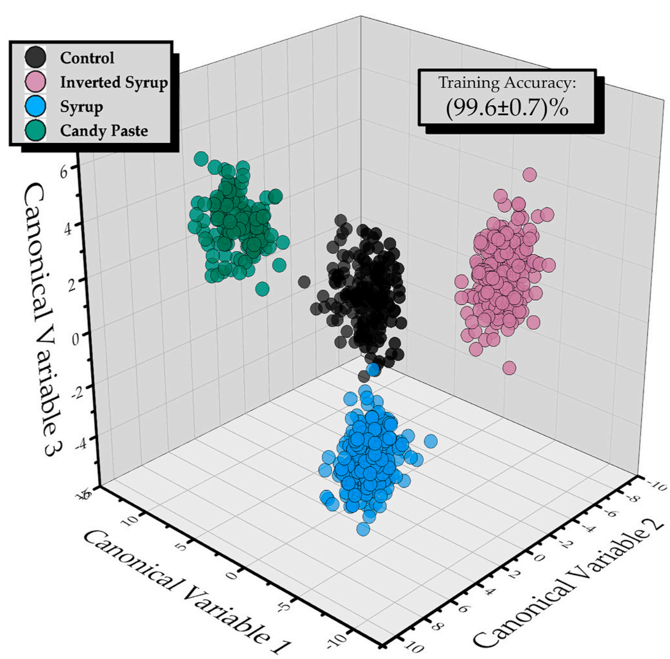

Fig. 5. Canonical Variable scatter plot indicating the supervised dimensionality reduction of the LIBS dataset achieved via the LDA algorithm.

wrongly predicted as inverted syrup ones.

A noteworthy advantage resulting from the implementation of the RFC algorithm in this study, is due to its easy-to-interpret and quick-to-calculate feature importance capability (Louppe et al., 2013). As an example, in Fig. 7 the most important features recognized by the RFC algorithm, plotted versus the corresponding wavelength are shown. As can be seen, the most important features are those corresponding to the spectral lines of the inorganic ingredients of honey, i.e., sodium Na (I) (at 589.0 and 589.6 nm), potassium K (I) (at 766.5 and 769.9 nm), calcium Ca (II) (at 393.3, 396.8 and 422.7 nm) and magnesium Mg (II) (at 279.6, 280.3 and 285.2 nm) (listed in descending order of importance), as well as the carbon C (I) line at 247.9 nm. The importance of these spectral lines for honey classification issues has been reported by other studies as well (Stefas et al., 2020; Zhao et al., 2020; Lastra-Mejías et al., 2020; Peng et al., 2020). Then, based on these observations, a new RFC model was constructed, by taking into consideration only the so determined most important features. More specifically, the dataset was reduced by selecting a threshold value of variable importance, and the RFC algorithm was retrained using less features. For example, for a variable importance threshold value of 0, all features of Fig. 7 (i.e., all the 2754 wavelengths) are taken into account, while for a variable importance value higher than e.g., 0.025 only one feature, i.e., that corresponding to the Na (I) spectral line, is considered. In this way, not only the influence of a feature or a number of features on the algorithm's accuracy (both training and testing procedures) can be evaluated, i.e., by adjusting the threshold value, but the most important spectral lines for classification purposes can be identified as well, as it has been discussed in more details elsewhere (Gyftokostas et al., 2021).

The overall performance of the RFC algorithm using different threshold values of the variable importances is presented in Fig. 8. The best results were obtained for a threshold value of 0.021, where the training accuracy is high, and the testing accuracy decreases within the standard deviation of the training accuracies stemming from the cross-validation. The obtained accuracy for this threshold value was determined to be (98.8 ± 1.2)% for the training and 98.9% for testing, corresponding to only six features. In this case, considering the threshold value which results to the highest accuracy, it appears from Fig. 7 that the emissions of Na (I), Ca (II) and K (I) are the main spectral features which contribute the most in the discrimination accuracy, all of them corresponding to the inorganic ingredients of honey.

Next, the constructed predictive models were used to classify 30 more samples obtained from the second honey collection which was performed one month after sealing the honey cells (see also Table 1). For this purpose, the LDA and RFC models that were previously constructed

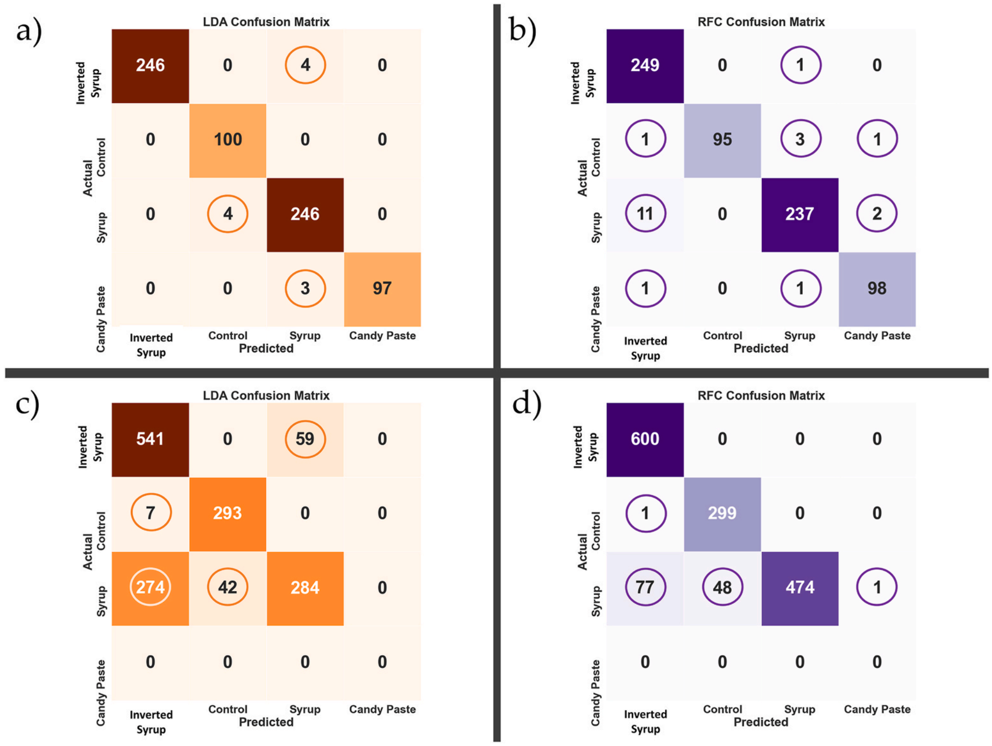

Fig. 6. Confusion matrixes for the prediction of the test/validation data for the: a) LDA algorithm, b) RFC algorithm, and the samples from the second collection for the: c) LDA, and d) RFC algorithms, respectively.

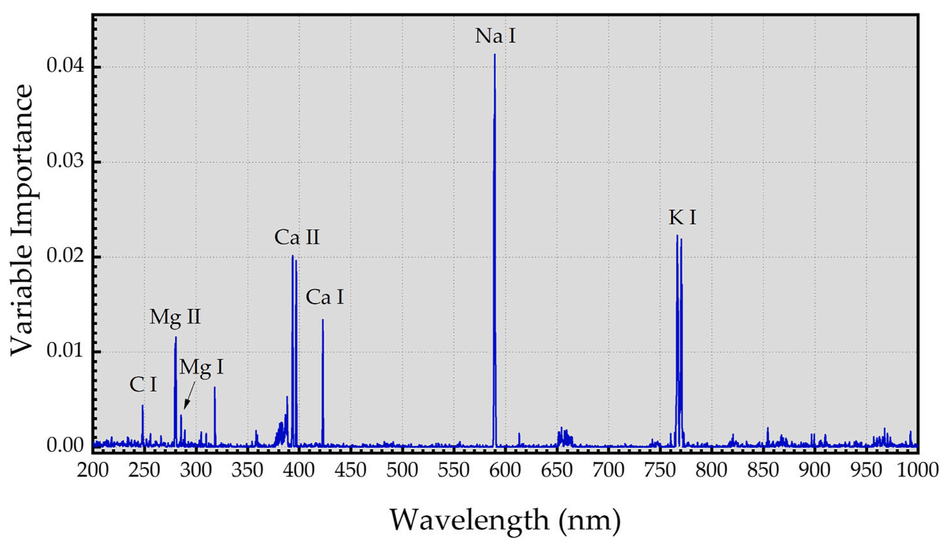

Fig. 7. Variable importances of the RFC model plotted versus the wavelength.

were used to predict the class designation of the samples. The LDA algorithm attained an accuracy of 74.5% while the RFC model reached an accuracy of 91.5%. From the analysis of these results, it becomes evident that for both models most misclassifications occurred for the inverted syrup and syrup classes. Specifically, in the corresponding confusion matrix concerning the LDA model, shown in Figs. 6c and 274 LIBS spectra of the syrup class were falsely predicted as belonging to the inverted syrup class, and 59 inverted syrup spectra were predicted as belonging to the syrup class. Similar trend was observed using the RFC algorithm (see Fig. 6d), although in this case the number of misclassifications was substantially reduced, i.e., 77 LIBS spectra belonging to the syrup class were falsely predicted as inverted syrup, while 48 spectra were misclassified as belonging to the control class.

These findings are very interesting, as they suggest that the atomic emissions of the inorganic elements (e.g., Ca, Na and K) in the LIBS spectra of the honeys from the first collection were very similar to the spectra obtained from honeys collected one month later. The very high prediction accuracies indicate that the different classes of the honeys that were collected in different time periods can be successfully identified by LIBS. This is not observed in the case of recording changes of sugars, as the enzymes in the sealed cells convert the carbohydrates, thus differentiating their concentrations over time (Al-Mahasneh et al., 2021; Guler et al., 2007). For example, if the checking of the effect of feeding in honey was based only on sucrose concentration it would be very difficult

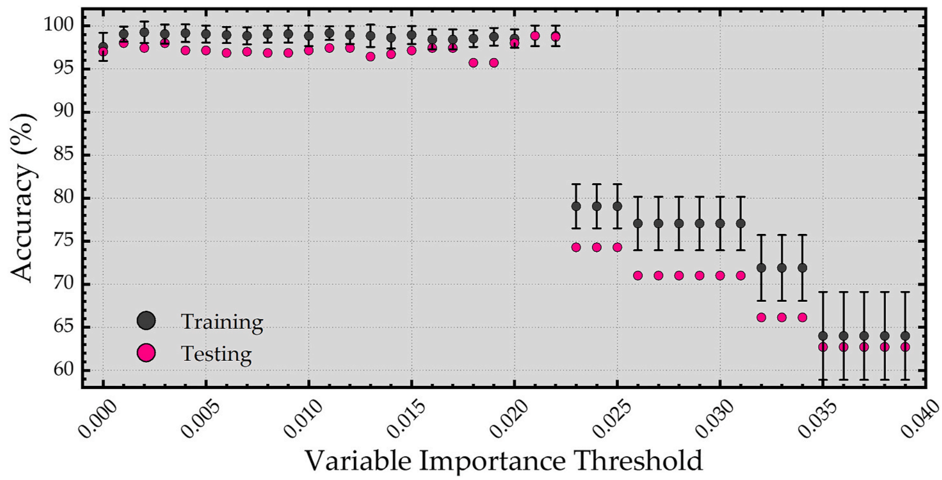

Fig. 8. Random Forest Classifier accuracies for both training and testing, for various of Variable Importance threshold values.

to be detected after some months, because the added by bees invertase convert this carbohydrate to fructose and glucose.

4. Conclusions

In the present study, LIBS is used, for the first time, not only to study the effect of bee-feeding on honey samples but also to classify them on the basis of the type of the bee food. Thus, it was proved that the effect of the type of bee feeding can be deduced by certain spectral lines of honey LIBS spectra such as magnesium, calcium, potassium and sodium, which were found to be the most important for the discrimination of the honey samples. In addition, the effect of bee-feeding for further period of time seems not to contribute significantly at honey spectra. According to these results the bee-feeding seems to reach a maximum level beyond that no significant difference can be observed. This is quite important since with specific amount of feed more beehives and more high-quality honey can be produced. Also, the LDA algorithms seems to be a less generalized model for prediction of honeys (according to the predictive accuracy of the second honey collection). It must be pointed out that LDA is quite useful for visualization of multidimensional data, but when it comes in terms of robustness and accuracy can be deceiving. For this reason, RFC algorithm were applied, since it is considering one of the most robust and generalized predictive models that is confirmed by the obtained results.

CRediT authorship contribution statement

Dimitrios Stefas: Data curation, Investigation, Software, Writing – original draft. **Nikolaos Gyftokostas:** Data curation, Investigation, Software, Writing – original draft. **Panagiotis Kourelias:** Data curation, Investigation, Software, Writing – original draft. **Eleni Nanou:** Data curation, Investigation, Software, Writing – original draft. **Chrysoula Tananaki:** Methodology, Project administration, Resources, Supervision, Writing – review & editing. **Dimitrios Kanelis:** Data curation, Investigation. **Vasileios Liolios:** Data curation, Investigation. **Vasileios Kokkinos:** Data curation, Investigation, Software, Writing – original draft. **Christos Bouras:** Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing. **Stelios Couris:** Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing, All authors have read and agreed to the published version of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Support from Greek national funds through the Public Investments Program (PIP) of General Secretariat for Research and Technology (GSRT), under the Emblematic Action “The bee routes” (project code: 2018ΣΕ01300000) is acknowledged. D.S. acknowledges support from the «Andreas Mentzelopoulos Foundation».

References

Al-Mahasneh, M., Al-U’datt, M., Rababah, T., Al-Widyan, M., Kaeed, A. A., Al-Mahasneh, A. J., & Abu-Khalaf, N. (2021). Classification and prediction of bee honey indirect adulteration using physiochemical properties coupled with K-means clustering and simulated annealing-artificial neural networks (SA-ANNs). *Journal of Food Quality*. , Article 6634598. https://doi.org/10.1155/2021/6634598, 2021.

Bertelli, D., Lolli, M., Papotti, G., Bortolotti, L., Serra, G., & Plessi, M. (2010). Detection of honey adulteration by sugar syrups using one-dimensional and two-dimensional high-resolution nuclear magnetic resonance. *Journal of Agricultural and Food Chemistry, 58*(15), 8495–8501. https://doi.org/10.1021/jf101460t

Bogdanov, S., Martin, P., & Lullmann, C. (1997). *Harmonised methods of the European honey commission* (pp. 1–59). Apidologie (Extra Issue).

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32. https://doi.org/10.1023/A:1010933404324

Cabañero, A. I., Recio, J. L., & Rupérez, M. (2006). Liquid chromatography coupled to isotope ratio mass spectrometry: A new perspective on honey adulteration detection. *Journal of Agricultural and Food Chemistry, 54*(26), 9719–9727. https://doi.org/10.1021/jf062067x

Council Directive 2001/110/EC. Council directive 2001/110/EC of 20 December 2001 relating to honey. European Union, Council of the European Union http://data.europa.eu/eli/dir/2001/110/2014-06-23.

Da Silva, P. M., Gauche, C., Gonzaga, L. V., Costa, A. C. O., & Fett, R. (2016). Honey: Chemical composition, stability and authenticity. *Food Chemistry, 196*, 309–323. https://doi.org/10.1016/j.foodchem.2015.09.051

Das, C., Chakraborty, S., Acharya, K., Bera, N. K., Chattopadhyay, D., Karmakar, A., & Chattopadhyay, S. (2017). FT-MIR supported Electrical Impedance Spectroscopy based study of sugar adulterated honeys from different floral origin. *Talanta, 171*, 327–334. https://doi.org/10.1016/j.talanta.2017.05.016

De Giacomo, A., & Hermann, J. (2017). Laser-induced plasma emission: From atomic to molecular spectra. *Journal of Physics D Applied Physics, 50*(18), Article 183002. https://doi.org/10.1088/1361-6463/aa6585

Duda, R. O., Hart, P. E., & Stork, D. G. (2000). In *Pattern classification* (2nd ed.). Wiley-Interscience.

Elflein, L., & Raezke, K. P. (2008). Improved detection of honey adulteration by measuring differences between 13C/12C stable carbon isotope ratios of protein and sugar compounds with a combination of elemental analyzer - isotope ratio mass spectrometry and liquid chromatography - isotope ratio mass spectrometry (δ13C-EA/LC-IRMS). *Apidologie, 39*(5), 574–587. https://doi.org/10.1051/apido:2008042

Guler, A., Bakan, A., Nisbet, C., & Yavuz, O. (2007). Determination of important biochemical properties of honey to discriminate pure and adulterated honey with sucrose (Saccharum officinarum L.) syrup. *Food Chemistry, 105*(3), 1119–1125. https://doi.org/10.1016/j.foodchem.2007.02.024

Guler, A., Kocaokutgen, H., Garipoglu, A. V., Onder, H., Ekinci, D., & Biyik, S. (2014). Detection of adulterated honey produced by honeybee (Apis mellifera L.) colonies fed with different levels of commercial industrial sugar (C3 and C4 plants) syrups by the carbon isotope ratio analysis. *Food Chemistry, 155*, 155–160. https://doi.org/10.1016/j.foodchem.2014.01.033

Gyftokostas, N., Stefas, D., Kokkinos, V., Bouras, C., & Couris, S. (2021). Laser-induced breakdown spectroscopy coupled with machine learning as a tool for olive oil authenticity and geographic discrimination. *Scientific Reports, 11*, 5360. https://doi.org/10.1038/s41598-021-84941-z

Harvey, L. J. (2016). Calcium: Properties and determination. In B. Caballero, P. M. Finglas, & F. Toldrá (Eds.), *Encyclopedia of food and health* (pp. 590–595). Oxford: Academic Press. https://doi.org/10.1016/B978-0-12-384947-2.00102-1.

Kramida, A., Ralchenko, Y., Reader, J., & NIST ASD Team. (2020). *NIST atomic spectra database (version 5.8)*. https://doi.org/10.18434/T4W30F. Available online:.

Lastra-Mejías, M., Izquierdo, M., González-Flores, E., Cancilla, J. C., Izquierdo, J. G., & Torrecilla, J. S. (2020). Honey exposed to laser-induced breakdown spectroscopy for chaos-based botanical classification and fraud assessment. *Chemometrics and Intelligent Laboratory Systems, 199*, Article 103939. https://doi.org/10.1016/j.chemolab.2020.103939

Liolios, V., Tananaki, C., Dimou, M., Kanelis, D., Goras, G., Karazafiris, E., et al. (2016). Ranking pollen from bee plants according to their protein contribution to honey bees. *Journal of Apicultural Research, 54*, 582–592. https://doi.org/10.1080/00218839.2016.1173353

Louppe, G., Wehenkel, L., Sutera, A., & Geurts, P. (2013). Understanding variable importances in forests of randomized trees. *Advances in Neural Information Processing Systems, 26*, 431–439.

Megherbi, M., Herbreteau, B., Faure, R., & Salvador, A. (2009). Polysaccharides as a marker for detection of corn sugar syrup addition in honey. *Journal of Agricultural and Food Chemistry, 57*(6), 2105–2111. https://doi.org/10.1021/jf803384q

Miziolek, A. W., Palleschi, V., & Schechter, I. (2008). *Laser induced breakdown spectroscopy*. Cambridge: Cambridge University Press.

Morales, V., Corzo, N., & Sanz, M. L. (2008). HPAEC-PAD oligosaccharide analysis to detect adulterations of honey with sugar syrups. *Food Chemistry, 107*(2), 922–928. https://doi.org/10.1016/j.foodchem.2007.08.050

Moreno-Rojas, R., Cámara-Martos, F., & Amaro López, M.Á. (2016a). Sodium: Properties and determination. In B. Caballero, P. M. Finglas, & F. Toldrá (Eds.), *Encyclopedia of food and health* (pp. 19–23). Oxford: Academic Press. https://doi.org/10.1016/B978-0-12-384947-2.00636-X.

Moreno-Rojas, R., Cámara-Martos, F., & Amaro López, M.Á. (2016b). Potassium: Properties and determination. In B. Caballero, P. M. Finglas, & F. Toldrá (Eds.), *Encyclopedia of food and health* (pp. 439–445). Oxford: Academic Press. https://doi.org/10.1016/B978-0-12-384947-2.00553-5.

Moumeh, B., Dolores Garrido, M., Diaz, P., Peñaranda, I., & Linares, M. B. (2020). Chemical analysis and sensory evaluation of honey produced by honeybee colonies fed with different sugar pastes. *Food Sciences and Nutrition, 8*, 5823–5831. https://doi.org/10.1002/fsn3.1843

Nespeca, M. G., Vieira, A. L., Júnior, D. S., Neto, J. A. G., & Ferreira, E. C. (2020). Detection and quantification of adulterants in honey by LIBS. *Food Chemistry, 311*, Article 125886. https://doi.org/10.1016/j.foodchem.2019.125886

Özcan, M., Arslan, D., & Ceylan, D. A. (2006). Effect of inverted saccharose on some properties of honey. *Food Chemistry, 99*(1), 24–29. https://doi.org/10.1016/j.foodchem.2005.07.009

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Peng, J., Xie, W., Jiang, J., Zhao, Z., Zhou, F., & Liu, F. (2020). Fast quantification of honey adulteration with laser-induced breakdown spectroscopy and chemometric methods. *Foods, 9*(3), 341. https://doi.org/10.3390/foods9030341

Puscas, A., Hosu, A., & Cimpoiu, C. (2013). Application of a newly developed and validated high-performance thin-layer chromatographic method to control honey adulteration. *Journal of Chromatography A, 1272*, 132–135. https://doi.org/10.1016/j.chroma.2012.11.064

Rodopoulou, M. A., Tananaki, C., Dimou, M., Liolios, V., Kanelis, D., Goras, G., & Thrasyvoulou, A. (2017). The determination of the botanical origin in honeys with over-represented pollen: Combination of melissopalynological, sensory and physicochemical analysis. *Journal of the Science of Food and Agriculture, 98*(7), 2705–2712. https://doi.org/10.1002/jsfa.8764

Rodopoulou, M. A., Tananaki, C., Kanelis, D., Liolios, V., Dimou, M., & Thrasyvoulou, A. (2021). A chemometric approach for the differentiation of 15 monofloral honeys based on physicochemical parameters. *Journal of the Science of Food and Agriculture*. https://doi.org/10.1002/jsfa.11340. accepted article (under publication).

Rodriquez, I., Camara - Martos, F., Flores, G. M., & Serrano, S. (2019). Spanish avocado (*Persea americana* Mill.) honey: Authentication based on its composition criteria, mineral content and sensory attributes. *Lebensmittel-Wissenschaft und -Technologie-Food Science and Technology, 111*, 561–572. https://doi.org/10.1016/j.lwt.2019.05.068

Ruiz-Matute, A. I., Soria, A. C., Martínez-Castro, I., & Sanz, M. L. (2007). A new methodology based on GC-MS to detect honey adulteration with commercial syrups. *Journal of Agricultural and Food Chemistry, 55*(18), 7264–7269. https://doi.org/10.1021/jf070559j

Serrano, S., Espejo, R., Villarejo, M., & Jordal, M. L. (2006). Diastase and invertase activities in Andalusian honeys. *International Journal of Food Science, 42*(1), 76–79. https://doi.org/10.1111/j.1365-2621.2006.01213.x

Spiteri, M., Jamin, E., Thomas, F., Rebours, A., Lees, M., Rogers, K. M., & Rutledge, D. N. (2015). Fast and global authenticity screening of honey using 1H-NMR profiling. *Food Chemistry, 189*, 60–66. https://doi.org/10.1016/j.foodchem.2014.11.099

Stankovska, E., Stafilov, T., & Šajn, R. (2008). Monitoring of trace elements in honey from the Republic of Macedonia by atomic absorption spectrometry. *Environmental Monitoring and Assessment, 142*, 117–126. https://doi.org/10.1007/s10661-007-9913-x

Stefas, D., Gyftokostas, N., & Couris, S. (2020). Laser induced breakdown spectroscopy for elemental analysis and discrimination of honey samples. *Spectrochimica Acta Part B Atomic Spectroscopy, 172*, Article 105969. https://doi.org/10.1016/j.sab.2020.105969

Wang, S., Guo, Q., Wang, L., Lin, L., Shi, H., Cao, H., & Cao, B. (2015). Detection of honey adulteration with starch syrup by high performance liquid chromatography. *Food Chemistry, 172*, 669–674. https://doi.org/10.1016/j.foodchem.2014.09.044

Wang, H., Inukai, Y., & Yamauchi, A. (2006). Root development and nutrient uptake. *CRC Critical Reviews In Plant Sciences, 25*(3), 279–301. https://doi.org/10.1080/07352680600709917

Xue, X., Wang, Q., Li, Y., Wu, L., Chen, L., Zhao, J., & Liu, F. (2013). 2-acetylfuran-3-glucopyranoside as a novel marker for the detection of honey adulterated with rice syrup. *Journal of Agricultural and Food Chemistry, 61*(31), 7488–7493. https://doi.org/10.1021/jf401912u

Zhao, Z., Chen, L., Liu, F., Zhou, F., Peng, J., & Sun, M. (2020). Fast classification of geographical origins of honey based on laser-induced breakdown spectroscopy and multivariate analysis. *Sensors, 20*(7), 1878. https://doi.org/10.3390/s20071878

Zhou, J., Qi, Y., Ritho, J., Duan, L., Wu, L., Diao, Q., Li, Y., & Zhao, J. (2014). Analysis of maltooligosaccharides in honey samples by ultra-performance liquid chromatography coupled with evaporative light scattering detection. *Food Research International, 56*, 260–265. https://doi.org/10.1016/j.foodres.2014.01.014