



# Honey exposed to laser-induced breakdown spectroscopy for chaos-based botanical classification and fraud assessment

Miguel Lastra-Mejías<sup>a</sup>, Manuel Izquierdo<sup>a</sup>, Ester González-Flores<sup>a</sup>, John C. Cancilla<sup>b</sup>,  
 Jesús G. Izquierdo<sup>c</sup>, José S. Torrecilla<sup>a,\*</sup>

<sup>a</sup> Departamento de Ingeniería Química y de Materiales, Facultad de Ciencias Químicas, Universidad Complutense de Madrid, 28040, Madrid, Spain

<sup>b</sup> Scintillon Institute, San Diego, CA, USA

<sup>c</sup> Centro de Láseres Ultrarrápidos y Departamento de Química Física I, Facultad de Ciencias Químicas, Universidad Complutense de Madrid, 28040, Madrid, Spain

## ARTICLE INFO

### Keywords:

Honey  
 LIBS  
 Chaotic parameters  
 Botanical origin  
 Adulteration  
 Classification

## ABSTRACT

Given that honey is among the top ten foods with the highest adulteration rate in the European Union, in this research, a tool has been developed to tackle this malpractice. The combination of laser-induced breakdown spectroscopy (LIBS) and chaotic parameters has been employed to classify six European honeys of different botanical origins as well as detect samples containing the usually elusive rice syrup adulteration in weight concentrations as low as 2%.

The profiles of the LIBS emission spectra can be used to faithfully classify honey in terms of botanical origin by combining information extracted directly from the spectra with simple linear modeling. In contrast, the detection of low amounts of rice syrup in honey is not as straightforward, which is why algorithms based on chaotic parameters such as shifted (lag- $k$ ) autocorrelation coefficients were employed to extract underlying information representative of adulterated samples. Since these algorithms are capable of detecting slight changes in the composition of honeys, it has been possible to identify these adulterations with a success rate greater than 90% when samples from honeys of different botanical origins are combined into the same model, and over 95% when individual honey types are analyzed.

## 1. Introduction

Honey is a natural and sweet fluid with a viscous nature and amber pigmentation, with many potential color shades as a result of a very rich and diverse composition. It is produced by honey bees (*Apis mellifera*), which use the nectar of flowers or secretions of plants as raw material [1]. This food is a natural sweetener which acts as an energetic and nutritional dietary boost when consumed. It has numerous medical and nutritional benefits due to its amount of minerals, enzymes, vitamins, and amino acids [2]. Their mineral content varies according to their botanical and geographical origin [3]. The bulk of the composition of honey is represented by carbohydrates (75–80% in weight), of which fructose and glucose are predominant. In addition, this food can contain up to 20% water (in weight), along with other compounds in lower proportions such as metals (up to 1%), proteins (up to 0.4%), amino acids, vitamins, fatty acids, enzymes, and pigments. Honey can also contain suspended particles and other substances such as pollen grains [4].

Despite following these criteria to maintain high levels of quality,

during their production they can still suffer from contaminations and/or adulterations that are detrimental to their quality and safety. So much so that honey is considered the sixth food with the highest risk of food fraud in Europe [5]. To protect this valuable food, towards the end of last century, a Codex Alimentarius was published to regulate the production and storage of honey in order to obtain parameters that guarantee its quality [6]. The deviation of the marked parameters can be produced both by malpractice in manufacturing and by the accidental inclusion of substances external to the original product, which translates into adulteration. This type of fraudulent activity has been analytically detected in honey via several methodologies. It should be noted that the techniques that attest to the authenticity of the product are constantly being developed in an attempt to respond to these increasingly common activities. The most commonly used adulterants in the beekeeping sector are rice, maize, and glucose syrups among others [7,8]. These syrups are considered sweeteners for use in foodstuffs. Some of these syrups come from C4 plants (corn syrups and glucose) and others from C3 plants (rice syrup). C3 and C4 plants differ in terms of chloroplast location as well as

\* Corresponding author.

E-mail address: [jstorre@ucm.es](mailto:jstorre@ucm.es) (J.S. Torrecilla).

<https://doi.org/10.1016/j.chemolab.2020.103939>

Received 7 June 2019; Received in revised form 8 January 2020; Accepted 14 January 2020

Available online 22 January 2020

0169-7439/© 2020 Elsevier B.V. All rights reserved.

several metabolic processes including the carbon dioxide fixation process. As honey also comes from C3 plants, rice syrup is one of the most difficult honey adulterants to detect [8]. For this reason, there are not many scientific references in this regard [7–9]. Rice syrup is a sweetener that is composed mainly of maltose and glucose.

Numerous analytical techniques have been used over the years to determine the purity of honey. The impact of the introduction of the voltammetric electronic tongue should be highlighted [10]. Also noteworthy is the application of high-performance liquid chromatography with an electrochemical detector using principal components analysis (PCA) for the classification of Chinese honeys according to their botanical origin [11]. In the area of spectroscopy, fluorescence spectroscopy [12,13] or coupled plasma mass spectrometry [14] should be mentioned, although most of them require expensive analytical equipment, complex preparation methods, and/or destruction of the analyzed samples.

A characteristic trait that can be used to identify honeys by botanical origin is their metallic composition [15,16]. This is why analytical tools capable of detecting and quantifying the metals present in honey can represent a suitable approach in their botanical characterization. In this line, Alda-Garcilope et al. successfully characterized honeys by quantifying five metals by means of flame atomic absorption spectroscopy [16]. In the same line, Wei et al. measured different metals with an approach based on laser-induced breakdown spectroscopy (LIBS) [15].

The technique based on LIBS has an inherently high sensitivity to detect metal atoms. LIBS is based on irradiating the surface of the samples using a high energy laser pulse, which results in the subsequent generation of plasma [17]. Plasma has its own characteristics that differentiates it from solids, liquids, or gases, reason why it is considered as another state of matter. It is a fluid similar to the gaseous state but electrically charged. LIBS can transform solids, liquids, or gases into plasma, allowing a rapid elemental analysis of a sample, irrespective of its original state [18]. Furthermore, there is a growing demand for the use of LIBS as an elemental analysis tool, as it can reveal complex analytical information from the plasma formed by the laser, including the specific emission spectrum of each element that composes the studied sample. The intensity of the spectral peaks or signals offer the necessary information to carry out a qualitative and quantitative analysis of the sample [19]. In addition, it enables a quick multi-element chemical analysis with relatively easy instrumentation handling, as well as requiring minimal sample preparation. This fact, together with the fast acquisition of the emission spectra, gives the technique the ability to be implemented for real-time applications. Thanks to all these characteristics, numerous studies have been carried out on different foods such as milk [20], butter [21], meat [22], or baby food [23], among others. LIBS has also been used in the field of honey with highly satisfactory results [15]. However, in many occasions, the complexity of the samples and resulting data entail a proper and sometimes elaborate mathematical treatment and interpretation of the outputted spectra. The complexity of the algorithms may vary according to the objectives to be achieved. Linear models can be applied, such as the multiple linear regression equations calculated by Gonzalez-Miret et al. to classify honey samples [24]. Chudzinska and Baralkiewicz applied PCA and clustering analysis for the authentication of honeys as a function of their metal profile [14]. Finally, one of the few articles where an estimation of the contents of some metals in honeys is carried out, covers the use of LIBS and a mathematical model based on partial least squares regression [15].

However, given that the metal content of honeys is below 1% in most cases, the analytical tool and mathematical algorithm combination must offer high sensitivity to minor changes in these profiles to distinguish not only the different honeys, but also locate possible adulterations. This is the main reason why chaotic parameters, given their responsiveness to small changes, can be an appropriate tool for such application [25,26]. As far as we know, there is no reference regarding the use of chaotic algorithms in the honey sector. However, the sound results obtained with these mathematical procedures lead us to think that a model based on chaotic parameters would be useful to distinguish different varieties of

**Table 1**  
Information regarding the honeys and rice syrup employed in this research.

Botanical Origin of Honey	Geographical Origin	Ecological	Commercial Brand
Lemon	Spain	Yes	Bona Mel
Chestnut	Spain	Yes	Hijas del Sol
Rosemary	Spain	No	La Obrera
Multifloral	Spain	No	Granja San Francisco
Eucalyptus	France	No	Luna de Miel
Orange	France	No	Luna de Miel
Rice Syrup (not honey)	France	No	Danival

honey as well as detect adulterations by analyzing databases generated by LIBS.

Given the similarity between the chemical compositions of honey and rice syrup, and specifically the mineral composition of both products (*vide infra*), it is hypothesized that the methodology based on LIBS could be a valid tool for the determination of the presence and quantification of rice syrup in adulterated honey mixtures [9]. In this line, the main objective of this research has focused on the characterization of pure honeys with distinct botanical origins as well as the detection of small amounts of rice syrup in adulterated samples by means of LIBS. This analytical tool has been used to qualitatively achieve these aims by detecting the presence of four metals (Mg, Ca, Na, and K) that can characterize honeys and their botanical origins. Alternatively, in order to quantitatively analyze adulterated honeys from different botanical origins, the quantification of rice syrup has been carried out using feature selection and chaotic algorithms on the data generated via LIBS.

## 2. Materials and methods

This section presents and describes the samples, the analytical equipment, and the mathematical tools used in this work.

### 2.1. Honey samples

The honey samples analyzed during this research are four Spanish honeys, two of which derive from ecological farming. In addition, two French honeys have been analyzed as well. The six honeys possess unique botanical origins, and they have all been analyzed using LIBS. Information regarding their botanical origin and brand can be found in Table 1, as well as data about the rice syrup used to create artificially adulterated honey samples.

The honeys and syrup used in this work were acquired in Spanish stores. All the honeys were purchased and analyzed prior to their best-before date. Until the analyses were carried out, all the products were stored in a dark and dry environment, at a temperature of approximately 22 °C.

In order to study the potential detection of adulterated honeys via LIBS, three of the honeys (lemon, rosemary, and multifloral) were used to prepare samples mixed with rice syrup in concentrations ranging from 2 to 8% in weight (2, 4, 6, and 8%). The mixtures were made in a thermostatic bath set at 52 °C to reach adequate homogeneity. Once achieved, the adulterated samples were allowed to rest for 24 h before being measured with LIBS. Considering that the LIBS emission profile picks up the elements inside the honey samples, the temperature increase or the addition of rice syrup into the pure honey, even if it changes the color of the honey itself, will not alter the elemental composition of the sample.

### 2.2. LIBS equipment

The LIBS system used in this work is composed of a nanosecond laser, a delay generator, and a high-resolution spectrometer that sends the information to a computer where all the produced data is managed. The assembled system used in the experimentation can be seen in Fig. 1.

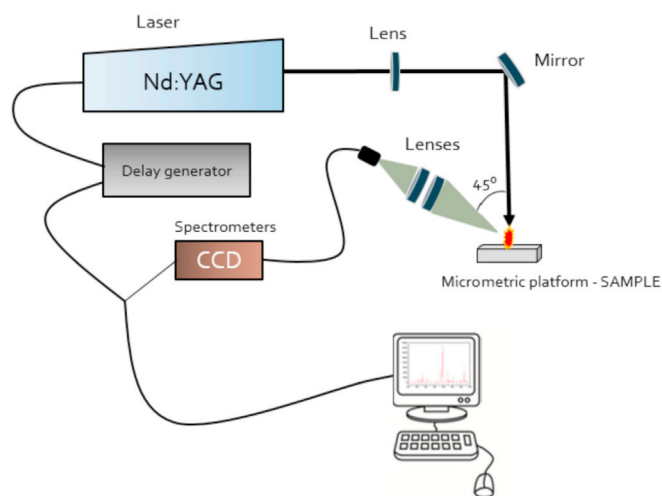


Fig. 1. LIBS system used for honey analyses.

The laser used is a pulsed Nd:YAG laser (Continuum Surelite II). Its working frequencies range from 1 to 10 Hz, with laser pulses lasting 6 ns, with a beam diameter of 4.5 mm, and a maximum energy in each pulse of 660 mJ. Its fundamental emission line is at 1064 nm. In this study, we worked at a frequency of 1 Hz and an energy of 270 mJ, measured by a power meter model AN/2E (Ophir).

The laser beam irradiating the samples, which are located on a micrometric platform with XYZ displacement (PT3, Thorlabs), was focused by a fused silica plano-convex lens of 250 mm focal distance, generating a point of 300  $\mu\text{m}$  in diameter. The light generated by the originated plasma was collected by means of two 75 mm and 35 mm focal length fused silica collimating lenses (LA4725 and LA4052, Thorlabs) at a 45° angle (see Fig. 1). This light was transported using a bifurcated UV-VIS optical fiber with a 400  $\mu\text{m}$  core (Bwtek) to two CCD spectrometers in tandem (BRC115P-V, Bwtek), covering a spectrum between 200 and 980 nm, with an approximate resolution of 0.20 nm ( $\lambda_r$ ). The entire emission spectrum consists of 3814 vectors (emission intensity-wavelength pairs). The signal received by the spectrometers is sent to a computer, where it is analyzed by the software included in the equipment (Bwspec). The integration time was 1 ms, with a delay of 4 ns between the laser pulse and the signal collection. This is done by means of a delay generator (Model DG-535, Stanford Research Systems, Inc.).

### 2.3. Analytical procedure

Regarding the analysis of the honey samples, 5 ml of each were collected in a cylindrical container of 2.1 cm in diameter and 1 cm in height. The vessel containing the sample is placed on a micrometer platform (with XYZ displacement, Fig. 1), where 13 uniformly distributed points on the surface of the sample are analyzed. In order to obtain representative and robust data from each honey sample, the measurements were taken in triplicate, resulting in 39 (13 x 3) spectra per sample (the ones from the same position were then averaged accordingly, leading to 13 final spectra per sample). In summary, 78 spectra were obtained from unadulterated honeys (6 honeys x 13 spectra per honey). Each one of these spectra is defined by the LIBS emission intensity versus wavelength, and this information has been used for the different subsequent calculations (*vide infra*). Finally, the LIBS spectra obtained were analyzed by comparison with the information contained in the NIST (National Institute of Standards and Technology) LIBS database version 5.6, which has been used to identify the emission lines of the different atoms present in the sample.

In relation to the adulterated samples prepared to design an adulteration detector, the same procedure was followed leading to 195 LIBS spectra (3 honeys x 5 adulterant concentrations (including 0%) x 13

spectra per sample). The final database was composed of a set of two-dimensional vectors (wavelength-LIBS emission intensity pairs) which have been used for the feature selection process as well as the calculation of chaotic parameters (see below).

### 2.4. Mathematical models

Linear models based on the k-nearest neighbor (kNN) method that use (i) selected features and (ii) extracted chaotic parameters from the LIBS spectra, have been developed and used to classify honey varieties and detect adulterations with rice syrup. Information from the LIBS spectra of pure honeys and their binary mixtures with rice syrup is employed for this purpose. In this section, the main characteristics of the algorithms used are presented.

#### 2.4.1. Feature selection via Relief-F

In this research, the Relief-F algorithm (RF) has been used to select the most relevant information from the LIBS spectra to design models to classify honeys in terms of botanical origin as well as state of adulteration. RF is a filter-based feature selection method that ranks features according to their ability to distinguish samples that are statistically near to each other [27,28]. When analyzing an instance, RF seeks for the nearest neighbors, one per class (same and different), and adjusts the feature weighting vector to allow ranking variables in correlation with their ability to differentiate neighbor samples from other classes. The function used to obtain the relief-F score is shown in equation (1).

$$RF(g_i) = \frac{\sum_{t=1}^n [d(g_{t,i} - g_{NM(x_t),i}) - d(g_{t,i} - g_{NH(x_t),i})]}{2} \quad (1)$$

In equation (1),  $g_{t,i}$  represents the value of the sample analyzed ( $x_t$ ) of a given feature ( $g_i$ ),  $g_{NM(x_t),i}$  and  $g_{NH(x_t),i}$  are the values of the  $i$ th feature belonging to the nearest neighbors of different and same classes, respectively [27,28]. This method was carried out with a tailored code via MATLAB 2018a.

#### 2.4.2. Chaotic parameters

In order to classify honeys of different botanical origins and detect samples adulterated with low amounts of rice syrup, chaotic parameters have been calculated based on spectral autocorrelations. These parameters consist of a set of lag-k autocorrelation coefficients calculated from the LIBS spectra (200–980 nm) [29]. The lag-k autocorrelation coefficient ( $S_{\Delta\lambda}$ ), or spectral autocorrelation, is also known as the serial correlation function [25].

It has been known since the 1940s that there is a mathematical relationship between the autocorrelation coefficient and the presence of chaos [30,31]. It was not until the end of the last century that Sugihara and May found that a strong decrease in the autocorrelation function could also be related to the existence of chaos [32]. The statements were further validated by Kettemann et al. who specified that these coefficients are a compelling tool for characterizing spectral statistics, thus describing the chaotic nature of systems [31,33]. For all these reasons, it can be agreed that the serial correlation (or autocorrelation) function can lead to a useful set of parameters in the evaluation of the chaotic nature of systems.

Spectral autocorrelation is a way to quantify to what extent  $X_{\lambda}$  versus  $X_{\lambda-k}$  presents a linear trend close to a correlation coefficient of 1 (in this research,  $X$  represents the LIBS emission from the sample at a given wavelength ( $\lambda$ ), while  $k$  symbolizes the wavelength shift). This parameter measures the intensity with which each data point depends on the wavelength delay ( $\Delta\lambda$  or  $Kw$ ). It represents the relationship between the autocovariance vs. variance of the database that describes the system. For example, for a given spectrum,  $Kw$  is between 1 ( $w = 0$ ) and 0 (much larger  $w$ ) [30]. In other words,  $Kw$  is the wavelength where the calculation started. For instance, the 226th parameter stands for the autocorrelation parameter that has been calculated starting at 226 nm

(considering the analyzed wavelength range goes from 200 to 980 nm). Kw is quantitatively described by equation (2) [34],

$$K_W = \frac{\sum_{n=1}^{N-\Delta\lambda} (X_n - \bar{X}) \cdot (X_{n-\Delta\lambda} - \bar{X})}{\sqrt{\sum_{n=1}^{N-\Delta\lambda} (X_n - \bar{X}) \cdot \sum_{n=1}^{N-\Delta\lambda} (X_{n-\Delta\lambda} - \bar{X})}} \quad (2)$$

In equation (2), N represents the number of emission points for a given spectrum. Considering that the value of  $\lambda$  oscillates between  $\lambda_0 = 200$  nm and  $\lambda_e = 980$  nm, when steps of  $\lambda \sim 4.36$  nm ( $\Delta\lambda$ ) are employed (as is the case), 179 autocorrelation parameters ( $(\lambda_e - \lambda_0) \cdot \Delta\lambda^{-1}$ ) (Kws) are obtained as a result of having 3814 emission-wavelength pairs. This range has been selected and used for all relevant information containing regions of the LIBS spectra. The resulting chaotic parameters are then used to design, optimize, and validate linear models (*vide infra*).

In summary, given that models will be designed for the classification of LIBS spectra according to the type of honey and determine their adulteration status, the input and output variables are the chaotic parameters (CPAs) extracted from the LIBS spectra and the class labels (botanical origins for the honey classifier and adulteration state (yes/no) for the rice syrup detector), respectively. Furthermore, the CPAs are easily calculated (equation (2)), even using a simple scientific calculator. In this work, the CPAs were calculated through Microsoft Excel 2013. All other statistical analyses were performed using MATLAB 2018a and Statgraphics Centurion XVIII software.

Once the total databases are fully calculated (from RF and chaotic parameters), for both applications, they were randomly divided into two groups of data to be used in the learning and validation process. These sets contain about 80% and 20% of the global database, respectively (*vide supra*). It is important to note that the data (independent and dependent variables) contained in the validation must be interpolated within the learning set. The learning and validation datasets have been generated by means of LIBS on honey samples containing known concentrations of rice syrup (ranging from 0 to 8% in weight). The concentrations of adulterants in the validation set must be within the range of the learning data set [35]. The statistical performance of the models will be quantified using the results obtained in the validation process. In total, as mentioned above, there are 234 LIBS spectra obtained from different honey samples (78 for the botanical origin classifier and 195 for the rice syrup detector).

In order to further validate the algorithms, a cross-validation has been carried out. The optimization of the parameters of each model is carried out using all but one sample, and then the performance is evaluated using the last point. This validation process is repeated as many times as samples employed, until each data point has been used to test the models [25].

#### 2.4.3. Linear modelling of gathered datasets

In this work, mathematical models based on the kNN method have been applied. It is a classic classification method which is highly employed for data mining and machine learning due to its simple implementation and its effective classification performance in a large number of fields and applications, without requiring prior learning. They are also applicable to n-dimensional samples [36].

This method was implemented to classify both pure and adulterated honey samples into their correct group according to their botanical origin and adulterant concentration. These models have been applied to the independent variables selected by RF as well as the calculated chaotic parameters. By means of this unsupervised algorithm, each unknown sample is located in the class containing the k samples closest to it. The parameter used for this decision is the Euclidean distance, and the predominant or favorable class will be selected based on lower distances. Specifically, in this work, the value of k = 10 was used [37]. The kNN models were calculated via MATLAB 2018a.

### 3. Results and discussion

This section is subdivided into two main blocks: initially, the ability of

LIBS to distinguish adulterated and unadulterated honeys of different botanical origins will be discussed. Later, mathematical tools based on RF and chaotic algorithms will be designed showing that in combination with LIBS, it will allow for the botanical classification of honeys and the detection and quantification of samples adulterated with rice syrup at different concentrations.

#### 3.1. Analysis of honey samples and LIBS results

Considering the objective of classifying samples of pure and adulterated honeys by means of LIBS, the obtained emission spectra are analyzed and discussed next.

##### 3.1.1. Analysis of pure honey samples

LIBS emission spectra allow the quantification of elements in a sample. It is possible to identify a given element by simply analyzing its characteristic wavelength. Since this is one of the objectives of this research, Fig. 2 shows as an example the LIBS emission spectrum of a sample of pure eucalyptus honey. All 234 LIBS profiles used provide characteristic information on the elemental composition of the studied honeys from both adulterated and unadulterated samples.

The plasma produced by the ablation of the sample gives information about the atomic, ionic, and molecular species resulting from the composition of the honey sample and the excitation of the atmosphere surrounding the sample (air in this case). The lines due to organic compounds are those corresponding to C, H, and O. Thanks to the ratios between them, i.e. C/H, C/O, and C/N, organic compounds can be identified such as aromatic nitro compounds or molecules like anthracene (examples not found in honey) [38]. The emission bands of molecules containing C<sub>2</sub> (Swan system) are related to the presence of aromatic rings in organic compounds [39], while the CN bands (violet system) are due to the interaction of carbon with atmospheric nitrogen [40]. The main metals present in the samples of the analyzed honeys are the following alkali and alkaline earth metals: Ca, Na, K, and Mg. The concentration of these metals in several honeys analyzed by Gonzalez-Miret et al. (2005) are shown in Table 2 [24].

These atoms (Ca, Na, K, and Mg) will serve as references to compare the honeys according to their botanical origin. The main emission peaks of Ca, Na, K, and Mg are found between the wavelengths 393–396 nm, 589–590 nm, 767–770 nm, and 279–280 nm, respectively (Fig. 2). It must be noted that Ca also shows a peak at around 320 nm, but only the one between 393 and 396 nm is considered as it is much more statistically relevant. Other elements such as Si or Al can also be found in the honey samples (Fig. 2) but have not been included in further analysis due to their much lower concentration (parts per billion range) when compared with the other mentioned metals [41]. To check whether there are real differences regarding these metals (Ca, Na, K, and Mg) in the analyzed honeys, a comparative study of the LIBS emission spectra was carried out (Fig. 3). It is worth mentioning that regarding the metallic composition of the rice syrup employed as adulterant, it is comparable to the honeys in terms of K and Ca, while revealing higher amounts of Na and Mg (Table 2).

Fig. 3 uncovers differences in emission intensity, not only in the wavelengths corresponding to the mentioned metals present in honeys, but also in the entire LIBS profile. As expected, these differences are related with the different composition of every honey type.

On the other hand, regarding the honey color itself, Solayman et al. stated that the percentage of metals found in honey is related to its color, as their concentration is correlated with darker honey colors [42]. Knowing this, and considering the metallic composition in the honey varieties studied here and shown in Table 2 [24], it should be expected that the honeys with greater metal contents would be chestnut, eucalyptus, and multifloral, as they apparently seem darker to the naked eye than rosemary, orange, and lemon honeys (in Table 2, these last two honeys are combined into citrus honey). This color trend seems to be in agreement with the honey images shown in Fig. 4.

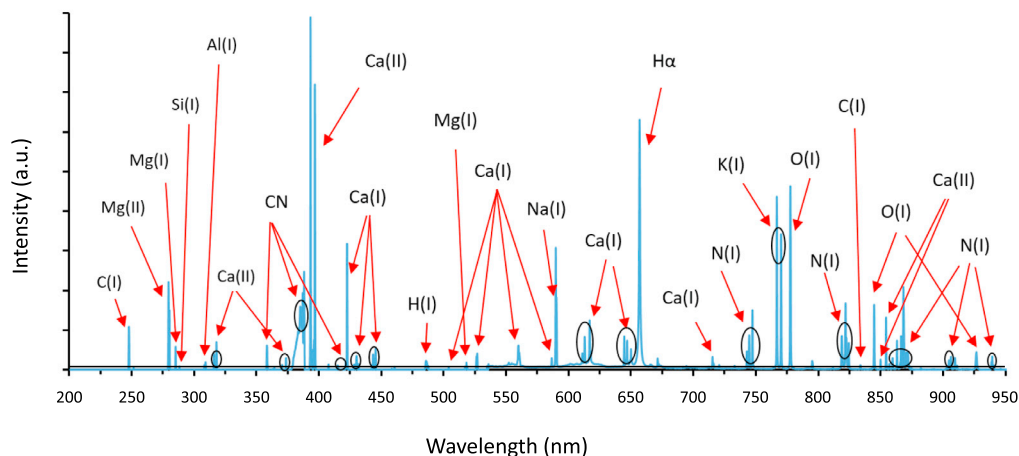


Fig. 2. LIBS emission spectrum of pure eucalyptus honey. Peaks are assigned to their corresponding element or molecule. (I) indicates neutral atoms while (II) are ionized ones.

Table 2

Amount of metals (mg/kg) found in the analyzed honeys [24] as well as rice syrup. Highest amounts of metals in honeys marked in bold.

	Citrus	Chestnut	Eucalyptus	Rosemary	Rice Syrup <sup>a</sup>
Na	83.9	93.9	144.4	80.8	440
K	237.0	1090.1	476.7	274.8	840
Ca	50.7	102.5	90.2	47.5	50
Mg	10.9	76.0	29.1	14.3	300
Total	382.5	1362.5	740.4	417.4	1630

<sup>a</sup> According to its commercial label.

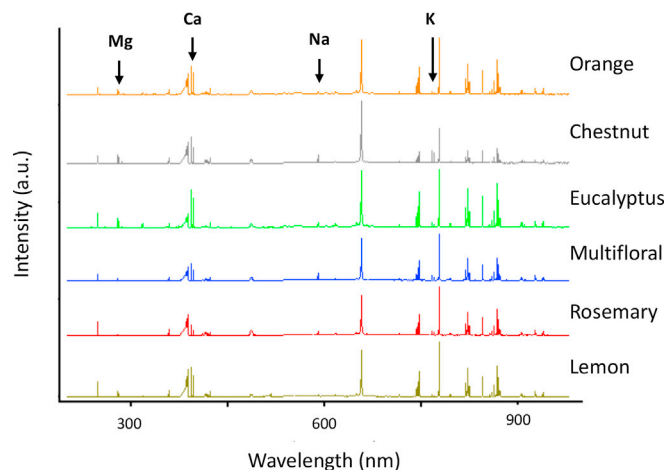


Fig. 3. LIBS spectra of the analyzed pure honeys from different botanical origins. The positions of the considered metals (Ca, Na, K, and Mg).



Fig. 4. Honey origin from left to right: rosemary, lemon, orange, multifloral, eucalyptus, and chestnut.

And finally, in order to check if the color sequence shown in Fig. 4 also corresponds with the content of the four described metals, the LIBS emission spectra of the analyzed honeys have been compared. The LIBS emission bands of Ca, Na, K, and Mg of all six pure honeys are represented in Fig. 5.

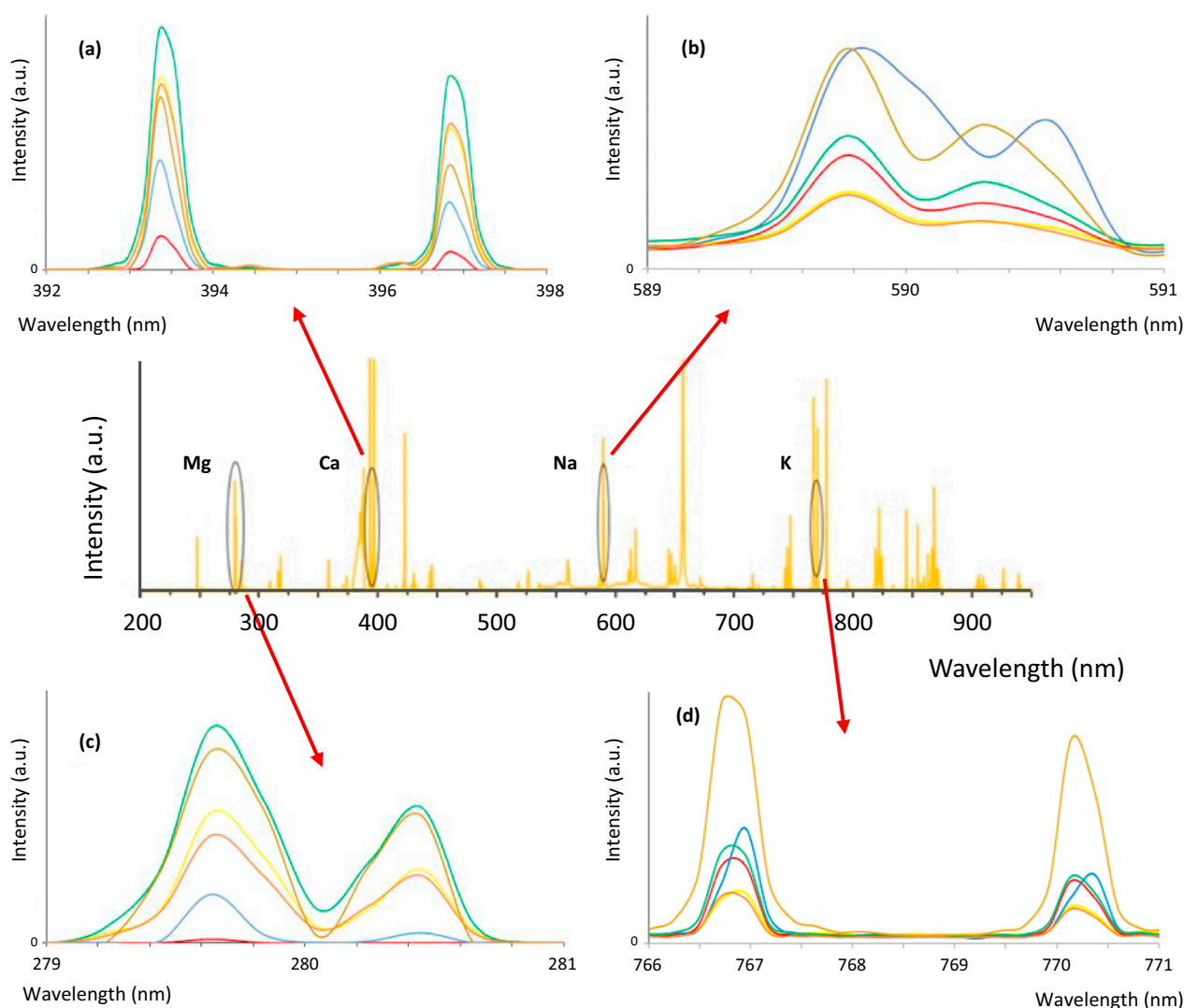
It can be noticed in Fig. 5 that, overall, the honey from multifloral, eucalyptus, and chestnut origins possess higher metal contents, which matches the mentioned color-based hypothesis (Table 2; Fig. 4) [24,42]. Specifically, eucalyptus honey has the highest Ca and Mg content, multifloral reveals the greatest Na amounts (very close to chestnut), and chestnut possesses the most K. On the other hand, as could be expected, there is a remarkable similarity between the LIBS profiles of lemon and orange honeys. This can be attributed to their similar botanical origin (both honeys come from bees that only collect nectar from citrus flowers). Finally, citrus and rosemary honeys have the lowest metal content, as well as the lightest color shades. These facts are consistent with other similar previously published studies involving honey analyses [14–16,24,43]. However, in each of these studies, the concentrations were not exactly the same due to the different geographical origins of the honeys, which directly affect their resulting metal content. With all this, considering the differences in LIBS emission intensity shown in Fig. 5, depending on the botanical origin of the honeys, it can be said that there are differences between pure honeys of different botanical origins, and they can be successfully identified via LIBS.

### 3.1.2. Analysis of adulterated honey samples

In addition to the bands corresponding to air, carbon, oxygen, hydrogen, and other non-honey-related compounds mentioned above, and considering the same LIBS emission bands for the wavelengths determined and shown in Figs. 2 and 3, the presence of the four characteristic metals (Ca, Na, K, and Mg) can also be detected in adulterated honeys (lemon, rosemary, and multifloral have been used). As an example, Fig. 6 shows the different LIBS emission profiles for pure and adulterated rosemary honey with rice syrup (2, 4, 6, and 8% in weight).

Looking into the emission peaks generated by these four metals, they do not seem to visibly differ in the LIBS spectra of the rosemary honey samples with varying amounts of rice syrup (Fig. 6). The same can be said for both lemon and multifloral honeys (data not shown). As might be expected, this fact makes sense as the concentrations of adulterants in all cases are very low and similar (always  $\leq 8\%$  in weight). Therefore, to classify pure and adulterated honey samples, the complete profiles of the LIBS emission spectra have been analyzed in order to use all the available information that can be extracted from them.

Considering that each emission spectrum is composed of 3814 two-dimensional vectors (*vide supra*), defining adequate mathematical



**Fig. 5.** Expanded LIBS emission bands of (a) Ca, (b) Na, (c) Mg, (d) and K in the honeys. Line colors correspond to the pure honeys from different botanical origins: lemon (yellow), chestnut (brown), rosemary (red), multiflora (blue), eucalyptus (green), and orange (orange). The center graph is an arbitrary example of a LIBS spectra from honey to show where the amplified regions (a–d) are located. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

classifying models is not trivial. For this reason, RF has been employed to identify the most relevant features of the LIBS database, whereas chaotic parameters have been calculated from it, as they are known to be capable of capturing slight differences among similar spectra [25]. Therefore, RF and chaos-based models have been developed for a dual purpose: to classify pure honey samples according to their botanical origin, as well as to distinguish adulterated honey samples from pure ones.

Although the metals present in honey are characteristic of their botanical origin, in order to fulfil the objectives of the work (i.e. to carry out the characterization of adulterated and unadulterated honeys as accurately as possible), it is necessary to have as much information as possible. For this reason, in the following sections of the work, the full LIBS emission profile will be used.

### 3.2. Classification of pure honeys by botanical origin

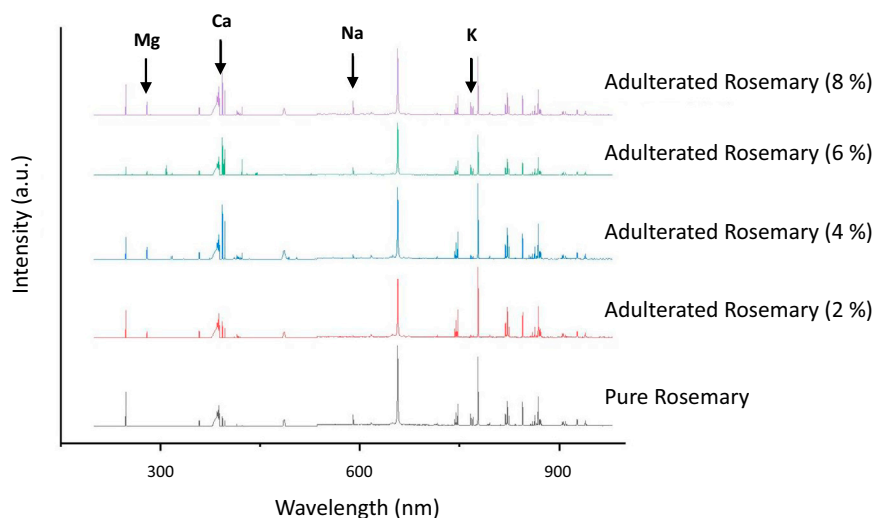
All the LIBS emission spectra generated from the samples of the six pure honeys analyzed (78 spectra) were used to determine the most relevant features using RF (equation (1)) as well as calculate the chaotic parameters (equation (2)). In other words, the data collected initially from the samples containing 3814 points corresponding to the LIBS

spectra, has been reduced by using two methods. On one hand, the original LIBS spectra were reduced to 179 points by the RF algorithm and, on the other, they were each characterized by 179 calculated autocorrelation parameters.

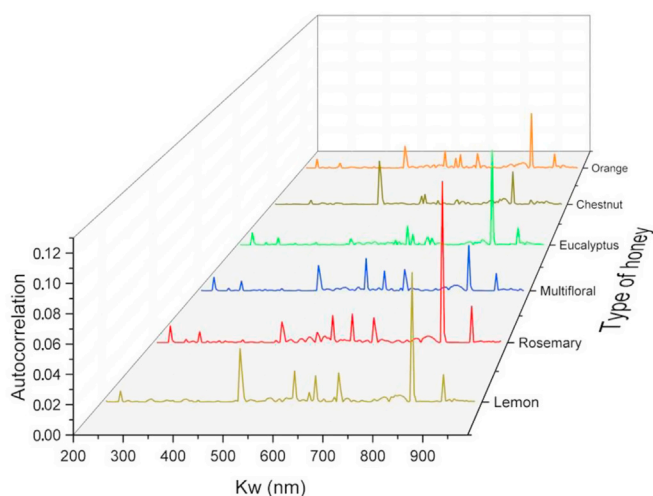
In the first approach, the RF algorithm selected the most important emission-wavelength pairs, many of which were related to the described metal peaks. In particular, the RF method selected many values corresponding to Ca (I) (more than 36%). The bulk of the remaining independent variables selected by RF to characterize the pure honeys by LIBS corresponded to the regions of K (I), Mg (II), Ca (II), and Na (I) (see Fig. 2).

On the other hand, the 179 calculated autocorrelation points ( $K_w$ ) for each honey were used to describe the characteristic profiles. Fig. 7 shows the different chaotic profiles of the six analyzed pure honeys.

In relation to the similarity that existed between the LIBS spectra of the citrus honeys (lemon and orange; Fig. 5), in this case, with this type of chaotic parameters, the profiles are completely different for a range covering from  $K_w = 200$ –980 nm (Fig. 7). As can be seen, the values obtained for around  $K_w = 850$  nm are unique for most honeys. In the case of eucalyptus and lemon honeys, although the peaks for this  $K_w$  value are similar, they have different ones for other  $K_w$  values. In the case of



**Fig. 6.** LIBS emission spectra of pure and rosemary honey adulterated with rice syrup at different concentrations. The wavelength positions of Ca, Na, K, and Mg are shown.

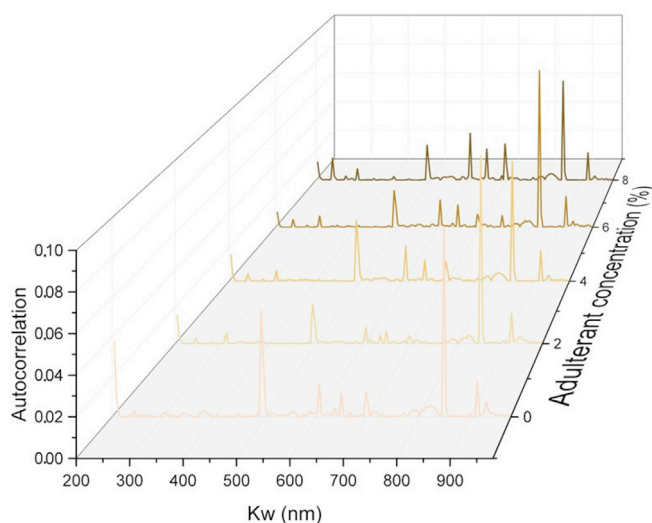


**Fig. 7.** Profiles of the calculated autocorrelation values of the six analyzed pure honeys.

multiflora and chestnut honeys, which also have similar  $Kw = 850$  nm values, they reveal differences between  $Kw = 525$  nm and  $Kw = 625$  nm. Finally, rosemary honey possesses the highest peaks of all honeys for the values of  $Kw = 850$  nm and  $Kw = 910$  nm.

Since this trend is repeated for most of the measured samples, considering these parameters as independent variables, it may be possible to design a model with a strong classifying power. In order to evaluate the potential of these chaos-based independent variables, as well as the ones located by RF, simple linear models have been used to classify the honey samples in terms of botanical origin.

The models tested were kNNs, where the input variables of the unsupervised models are the features selected by RF or the calculated chaotic parameters. Among the main parameters of this linear model, it is worth mentioning that the distance applied is the Euclidean, and the value of  $k$  is 10 (amount of analyzed neighbors). With this simple linear model, the average classification accuracy when using the most relevant LIBS emission-wavelength pairs according to RF was 91.3%, whereas when the chaotic parameters were employed, the accuracy improved to 95.8% (performances calculated for a cross-validation). Taking into account these values, the user could use either type of model in almost all



**Fig. 8.** Profiles of the calculated autocorrelation values of lemon honey adulterated with rice syrup (0–8% in weight).

cases.

The main difference of these approaches is that the chaotic parameter-based model has a general vision of the whole LIBS spectra, while the RF based model is only receiving information from specific peaks mathematically selected from whole LIBS spectra. On the other hand, it was observed that the greatest number of classification errors occurred for the multiflora honey, as well as a relevant number of misclassifications between lemon and orange (citric) honeys.

### 3.3. Detection of adulterated honeys

In this case, a binary classification has been developed to separate pure honeys from adulterated ones with rice syrup (2–8% in weight). All the LIBS emission spectra measured from the samples of the 3 adulterated honeys (lemon, rosemary, and multiflora) with four different concentrations of rice syrup as adulterant (2%, 4%, 6%, and 8% in weight; 195 spectra, including pure honey) were used for the selection of the most relevant regions of the LIBS emission spectra using RF and for the calculation of the chaotic parameters, following the same procedure as in the previous case. Fig. 8 shows, as an example, the differences in the

**Table 3**

Summary of results obtained by the kNNs that used the chaotic parameters in the classification of samples as adulterated or unadulterated according to the three rice syrup concentration thresholds, for all honey combined into one model as well as every individual honey variety analyzed separately.

Threshold (% in weight of rice syrup)	Classification accuracy of the kNNs during cross-validation (%)			
	All honey types	Lemon	Rosemary	Multifloral
2	87.2	92.4	93.0	90.3
4	90.1	95.0	94.6	92.2
6	91.2	95.8	95.1	93.0

profiles of lemon honey containing different concentrations of rice syrup.

In this case, the input variables that RF has selected are not only related to the aforementioned metals (K (I), Mg (II), Ca (I), Ca (II) and Na (I); see Fig. 2), but also, 42% of these independent variables are related with the presence of N<sub>2</sub> and O<sub>2</sub> of the surrounding air. There is no doubt that the inclusion of this information into the kNN model hinders its performance. Probably, this is the main reason why very poor statistical results were achieved (misclassification rate higher than 30%).

On the other hand, regarding the chaotic parameters, as might be expected, the differences between the ones obtained from the different LIBS spectra are less palpable than in the case of the classification of pure honeys by botanical origin (samples are much more similar in this case in terms of composition). It should be noted that we are trying to identify honeys with an adulterant concentration as low as 2%, which is far below the typical fraud.

Three different mathematical models have been developed using the calculated chaotic parameters to detect adulterated honey samples. The difference of the models was the threshold that was defined in terms of rice syrup amount: 2%, 4%, and 6% in weight. For instance, in the case of classification with a 2% threshold, this means that a sample is classified as adulterated in those cases where the amount of adulterant is identified as 2% or more. The same applies to the 4 and 6% thresholds. The models tested to carry out these classifications are similar to the ones used in the classification of pure honeys. Specifically, unsupervised models based on kNN were also employed, where the input variables were the chaotic parameters. Additionally, the distance applied in the classification was the Euclidean distance and the number of neighbors checked in the model (k) was 10. As in the validation of the models used for the classification of pure honey samples, cross-validation has been used to validate the models designed. The main results of the three models are shown in Table 3.

As can be noticed, the classification accuracy increases in correlation with the adulteration concentration cut-off, reaching a 91% correct classification rate for samples containing 6–8% rice syrup (Table 3). This makes sense as the chaotic parameters from samples that possess higher amounts of rice syrup must be farther from the ones calculated regarding the pure honey samples (Fig. 8). It is relevant to highlight that the adulterant concentrations found in real adulterated honeys are usually well above the concentrations used in this work, which further strengthens the results of the chaotic models.

The chaotic kNN model offers far better results than the RF kNN. A plausible cause may be that the chaos-based model has in consideration the whole LIBS spectra profile, while the RF variables solely cover the peaks selected, which include clearly confounding data. On the other hand, the errors in the classifications can also derive from the different botanical origins that have been used together to design these kNN models. To assess this issue, three different kNN models were also designed for every individual adulterated honey (lemon, rosemary, and multifloral) using the chaotic parameters as independent variables (Table 3). As can be seen by comparing the results shown in Table 3 (global model versus individual honey models), the classification accuracy has improved considerably (2–6% approximate improvement). In

particular, the multifloral honey variety is the one with the highest classification error considering all thresholds (9.7, 7.8, and 7.0% error). In contrast, they are better for rosemary honey (7.0, 5.4, and 4.9%) and the lowest for the lemon variety (7.6, 5.0, and 4.2%). The adulteration of all three varieties was reliably detected by combining LIBS and chaotic parameters.

#### 4. Conclusion

This work has revealed the ability of laser-induced breakdown spectroscopy (LIBS) in the characterization of pure ecological and non-ecological commercial honeys with different botanical origins. Also, success has been reached in the classification of honeys according to their botanical origin as well as in the detection of adulterated samples containing rice syrup (less than 8% in weight) by combining LIBS, chaotic parameters based on autocorrelation coefficients, and a simple linear unsupervised classifier. The results have been fitting, achieving classifications with less than 5% error in terms of botanical origin identification, and less than 13% for the location of adulterated samples with concentrations as low as 2% in weight of rice syrup. In the latter case, if the adulteration concentration cut-off is considered to be from 4% or 6% in weight, the classification errors decrease to around 10% and 9% (Table 3), respectively. Finally, if individual honeys are considered, the adulteration detection rate increases up to 93–96% for a 6% concentration cut-off (Table 3).

Derived from these results, LIBS can be used for the characterization of single-varietal and multifloral honey samples as well as the detection of adulterations. This aspect facilitates the integration of this analytical tool with different chemometric approaches. In this work, simple linear classifiers have been used to unmask the potential of chaotic parameters in the classification of pure honeys and detection of adulterated ones. For this reason, the integration of LIBS and chaotic parameters opens a compelling door for the detection of fraudulent activities in the honey sector as well as during the production of other goods.

#### Declaration of competing interest

No conflicts of interest.

#### CRediT authorship contribution statement

**Miguel Lastra-Mejías:** Conceptualization, Methodology, Formal analysis. **Manuel Izquierdo:** Validation, Investigation. **Ester González-Flores:** Validation. **John C. Cancell:** Software, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Supervision. **Jesús G. Izquierdo:** Methodology, Supervision. **José S. Torrecilla:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - original draft, Supervision, Writing - review & editing, Funding acquisition.

#### Acknowledgements

This work has been partially funded by the FEI program of the Complutense University of Madrid under the project reference FEI-EU-17-03, and FEI 16/123. Also, the Ultrafast Laser Center of the Complutense University of Madrid is thanked for providing access to the facilities.

#### References

- [1] J. Siddiqui, S.G. Musharraf, M.I. Choudhary, A.U. Rahman, Application of analytical methods in authentication and adulteration of honey, *Food Chem.* 217 (2017) 687–698, 217.
- [2] L. Wu, B. Du, Y.V. Heyden, L. Chen, L. Zhao, M. Wang, X. Xue, Recent advancements in detecting sugar-based adulterants in honey e A challenge, *Trends Anal. Chem.* 86 (2017) 25–38.

- [3] I. Conti, P. Medrzycki, A. Iannone, F.V. Grillenzoni, F. Corvucci, D. Dagnino, G. Casazza, C. Montanari, M.G. Mariotti, Preliminary survey of the nutritional and palynological traits of honey bee-foraged pollen from Liguria (Italy), *J. Apicult. Res.* 57 (2018) 572–579.
- [4] P.M. da Silva, C. Gauche, L.V. Gonzaga, A.C. Costa, R. Fett Honey, Chemical composition, stability and authenticity, *Food Chem.* 196 (2016) 309–323.
- [5] J. Moore, J. Spink, M. Lipkus, Development and application of a database of food ingredient fraud and economically motivated adulteration from 1980 to 2010, *J. Food Sci.* 77 (2012) R118–R126.
- [6] Comisión del codex alimentarius, 1999.
- [7] S.G. Musharraf, S.A. Fatima, A.J. Siddiqui, M.I. Choudhary, Atta-ur-Rahman. H-1-NMR fingerprinting of brown rice syrup as a common adulterant in honey, *Anal. Methods* 8 (2016) 6444–6451.
- [8] Y. Liu, F.Y. Qu, L.P. Luo, W.H. Xu, M. Zhong, Detection of rice syrup from acacia honey based on lubrication properties measured by tribology technique, *Tribol. Int.* 129 (2019) 239–245.
- [9] N. Czupa, C.J.C. Phillips, B. Kovacs, Composition of acacia honeys following processing, storage and adulteration, *J. Food Sci. Technol.* 56 (2019) 1245–1255.
- [10] M. Oroian, S. Ropciuc, Romanian honey authentication using voltammetric electronic tongue. Correlation of voltammetric data with physico-chemical parameters and phenolic compounds, *Comput. Electron. Agric.* 157 (2019) 371–379.
- [11] J. Zhao, X. Dua, N. Cheng, L. Chen, X. Xue, J. Zhao, L. Wu, W. Cao, Identification of monofloral honeys using HPLC–ECD and chemometrics, *Food Chem.* 194 (2016) 167–174.
- [12] Kr Nikolova, T. Eftimov, A. Aladjadjiyan, Fluorescence spectroscopy as method for quality control of honey, *Adv. Res.* 2 (2014) 95–108.
- [13] M. Lastra-Mejías, A. Torreblanca-Zanca, R. Aroca-Santos, J.C. Cancilla, J.G. Izquierdo, J.S. Torrecilla, Characterization of an array of honeys of different types and botanical origins through fluorescence emission based on LEDs, *Talanta* 185 (2018) 196–202.
- [14] M. Chudzinska, D. Baralkiewicz, Estimation of honey authenticity by multielements characteristics using inductively coupled plasma-mass spectrometry (ICP-MS) combined with chemometrics, *Food Chem. Toxicol.* 48 (2010) 284–290.
- [15] K.W. Se, S.K. Ghoshal, R.A. Wahab, Laser-induced breakdown spectroscopy unified partial least squares regression: an easy and speedy strategy for predicting Ca, Mg and Na content in honey, *Measurement* 136 (2019) 1–10.
- [16] C. Alda-Garcilope, A. Gallego-Picó, J.C. Bravo-Yagüe, R.M. Garcinuño-Martínez, P. Fernández-Hernando, Characterization of Spanish honeys with protected designation of origin “Miel de Granada” according to their mineral content, *Food Chem.* 135 (2012) 1785–1788.
- [17] D.A. Cremers, L.J. Radziemski, *Handbook of Laser-Induced Breakdown Spectroscopy*, John Wiley & Sons, Ltd, England, 2006.
- [18] A.W. Miziolek, V. Palleschi, I. Schechter, *Laser-induced Breakdown Spectroscopy (LIBS). Fundamentals and Application*, Cambridge University Press, New York, 2006, 2006.
- [19] S. Guirado Gutiérrez Málaga, *Tecnología LIBS subacuática. investigación, desarrollo y aplicación al estudio de materiales arqueológicos sumergidos*, 2015.
- [20] B. Sezer, S. Durna, G. Bilge, A. Berkkan, A. Yetişemeyen, I.H. Boyaci, Identification of milk fraud using laser-induced breakdown spectroscopy (LIBS), *Int. Dairy J.* 81 (2018) 1–7.
- [21] H.T. Temiz, B. Sezer, A. Berkkan, U. Tamer, I.H. Boyaci, Assessment of laser induced breakdown spectroscopy as a tool for analysis of butter adulteration, *J. Food Compos. Anal.* 67 (2018) 48–54.
- [22] F.O. Leme, D.M. Silvestre, A.N. Nascimento, C.S. Nomura, Feasibility of using laser induced breakdown spectroscopy for quantitative measurement of calcium, magnesium, potassium and sodium in meat, *J. Anal. Atomic Spectrom.* 33 (2018) 1322–1329.
- [23] R. Cama-Moncuñill, Y. Dixit, X. Cama-Moncuñill, M.P. Casado-Gavaldá, Maria Markiewicz-Keszyccka, Patrick J. Cullen, Carl Sullivan, Multipoint NIR spectroscopy for simultaneous analyses of dairy products – Part B: Quantification, *NIR News* 28 (2017) 13–16.
- [24] M.L. Gonzalez-Miret, A. Terrab, D. Hernanz, M.A. Fernandez-Recamales, F.J. Heredia, Multivariate correlation between color and mineral composition of honeys and by their botanical origin, *J. Agric. Food Chem.* 53 (2005) 2574–2580.
- [25] M. Lastra-Mejías, R. Aroca-Santos, M. Izquierdo, J.C. Cancilla, M.L. Mena, J.S. Torrecilla, Chaotic parameters from fluorescence spectra to resolve fraudulent mixtures of fresh and expired protected designation of origin extra virgin olive oils, *Talanta* 195 (2019) 1–7.
- [26] J.C. Sprott, *Chaos and Time-Series Analysis*, Oxford University Press Inc., New York, 2003.
- [27] B. Wu, C.C. Chen, T.M. Kechadi, L.Y. Sun, A comparative evaluation of filter-based feature selection methods for hyper-spectral band selection, *Int. J. Rem. Sens.* 34 (2013) 7974–7990.
- [28] Z. Zhao, F. Morstatter, S. Sharma, S. Alelyani, A. Anand, H. Liu, *Advancing Feature Selection Research – ASU Feature Selection Repository*, Arizona State University, AZ (USA), 2011. Technical Report.
- [29] J.S. Torrecilla, V. Fernández-Ruiz, M. Cámara, M.C. Sánchez Mata, The ability of spectrum autocorrelation models to predict the lycopene concentration in foods through visible spectroscopic data, *Talanta* 85 (2011) 2479–2483.
- [30] L. Zechmeister, A.L. LeRosen, W.A. Schroeder, A. Polgar, L. Pauling, Spectral characteristics and configuration of some stereo isomeric carotenoids including polyycopene and pro- $\gamma$ -carotene, *J. Am. Chem. Soc.* 65 (1943) 1940–1951.
- [31] Z.H. Ding, Z.R. Lu, J.K. Liu, J. Ke, Parameters identification of chaotic systems based on artificial bee colony algorithm combined with cuckoo search strategy, *Sci. China Technol. Sci.* 61 (2018) 417–426.
- [32] G. Sugihara, R. May, Nonlinear forecasting as a way of distinguishing chaos from measurement error in a data series, *Nature* 344 (1990) 735–741.
- [33] S. Kettemann, D. Klakow, U. Smilansky, Characterization of quantum chaos by the autocorrelation function of spectral determinants, *J. Phys. Math. Gen.* 30 (1997) 3643–3662.
- [34] H. Kant, T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge, U.K., 2005.
- [35] Guidance Document on the Validation of (Quantitative) Structure Activity Relationship [(Q)SAR] Models, No. 69, OECD, Series on Testing and Assessment, Organisation of Economic Cooperation and Development, Paris, France, 2007.
- [36] M. Lee, T.M. Gatton, K.K. Lee, A monitoring and advisory system for diabetes patient management using a rule-based method and KNN, *Sensors* 10 (2010) 3934–3953.
- [37] S. Zhang, Senior Member, IEEE, Xuelong Li, Fellow, IEEE, Ming Zong, Xiaofeng Zhu, Ruili Wang, Efficient kNN classification with different numbers of nearest neighbors, *IEEE Trans. Neural Network. Learn. Syst.* 29 (2018) 5.
- [38] P. Lucena, A. Doña, L.M. Tobaría, J.J. Laserna, New challenges and insights in the detection and spectral identification of organic explosives by laser induced breakdown spectroscopy, *Spectrochim. Acta B.* 66 (2011) 12–20.
- [39] L. St-Onge, R. Sing, S. Bechard, M. Sabsabi, Carbon emissions following 1.064  $\mu\text{m}$  laser ablation of graphite and organic samples in ambient air, *Appl. Phys. A* 69 (1999) s913–s916.
- [40] S. Trautner, J. Jasik, C.G. Parigger, J.D. Pedarnig, W. Spindelhofer, J. Lackner, P. Veis, J. Heitz, Laser-induced optical breakdown spectroscopy of polymer materials based on evaluation of molecular emission bands, *Spectrochim. Acta* 174 (2017) 331–338.
- [41] S.K. Altun, H. Dinç, N. Paksoy, F.K. Temamoğulları, M. Savrunlu, Analyses of mineral content and heavy metal of honey samples from south and east region of Turkey by using ICP-MS, *Int. J. Anal. Chem.* (2017). Article ID 6391454.
- [42] M. Soleyman, M.A. Islam, S. Paul, Y. Ali, M.I. Khalil, N. Alam, S.H. Gan, Physicochemical properties, minerals, trace elements, and heavy metals in honey of different origins: a comprehensive review, *Compr. Rev. Food Sci. Food Saf.* 15 (2016) 219–233.
- [43] A.M.G. Paramas, J.A.G. Barez, R.J. Garcia-Villanova, T.R. Pala, R.A. Albajar, J.S. Sanchez, Geographical discrimination of honeys by using mineral composition and common chemical quality parameters, *J. Sci. Food Agric.* 80 (2000) 157–165.