Use of elemental profiles to verify geographical origin and botanical variety of Spanish honeys with a protected denomination of origin

Michele Ghidotti, Yiannis Fiamegos [1], Catalina Dumitrascu [2], María Beatriz de la Calle [*]

European Commission, Joint Research Centre (Geel), Geel, Belgium

A B S T R A C T

Honey with Protected Denomination of Origin (PDO) could be an attractive target for fraudsters. Elemental profiles by Energy Dispersive-X Ray Fluorescence were processed by multivariate methods to classify 183 PDO honeys produced in three regions of Spain (Liébana, Granada, Tenerife). Additional honey samples (18) produced in a fourth region without PDO (El Bierzo) separated well from the PDO clusters. The manganese content was a discriminant marker of Liébana PDO and El Bierzo, that could also be differentiated from each other. Within each region, distinct clusters revealed differences between dark vs light varieties, multi- vs uni-floral honey and producers of the same PDO. The developed models were validated with 131 samples produced outside the PDO regions and El Bierzo. The proposed classification approach could be implemented as a fast screening tool to support pollen analysis in honey authentication. The reduced number of observations in some light honey models affected their performance.

1. Introduction

According to Council Regulation (EC) No 510/2006 (Council Regulation, 2006), a foodstuff or agricultural product holding a Protected Designation of Origin (PDO) must have been produced, processed and prepared in the place referred to in the designation and must have a set of characteristics derived from that geographical environment and its natural and human factors. A PDO label is a guarantee of quality and consumers are willing to pay more for products holding that label. Since economic profit is at the root of food fraud, products with a PDO are target for fraudsters. Twenty-seven honeys are registered as PDO by the European Commission (DOOR, 2020); the country with more PDO honeys is Portugal (9) followed by Spain (5), Italy (3), France (2), Slovenia (2) and Bulgaria, Croatia, Greece, Lithuania, Luxembourg, and Poland with one PDO honey each.

Traditionally, pollen analysis (melissopalynology) have been used to identify the botanical and geographical origin of honey (Bentabol Manzanares, Hernández García, Rodríguez Galdón, Rodríguez Rodríguez, & Díaz Romero, 2011; La Serna Ramos, & Gómez Ferreras, 2011) and a harmonised method have been published (Von der Ohe, Persano Oddo, Piana, Morlot, & Martin, 2004). However, this approach requires

experienced operators, a largely populated library of pollen for comparisons and a good knowledge of the flora in a particular region for geographical origin authentication (Ulberth, 2016). A discussion of the problems associated to the identification of botanical variety by melissopalynology, is included in the review made by Anklam (Anklam, 1998). An alternative to pollen analysis is the detection of the genetic material of plants used by bees to produce honey. Polymerase chain reaction (PCR) has been applied to characterise the plant profile in Corsican and Galician honeys (Laube et al., 2010).

Alternatives to melissopalynology have been developed over the last years and presented in thorough reviews (Anklam, 1998; Ulberth, 2016; Trifković, Andrić, Ristivojević, Guzelmeric, & Yesilada, 2017), with an increase in number of research papers published on honey authentication from 1327 to 7303 in the period 2006–2016 (Trifković et al., 2017). The review by Anklam (Anklam, 1998) focuses primarily in the substances used as markers to determine de botanical and geographical origin of honey, namely: aminoacids and proteins, aroma compounds, carbohydrates (sugars), enzymes (enzyme activity), fermentation products, flavonoids and other phenolic compounds, minerals and trace elements, stable isotopes, aliphatic organic acids and some other specific compounds. Trifković et al. (Trifković et al., 2017) focus in their review

on the analytical techniques used to trace honey authenticity: infra-red (IR), nuclear magnetic resonance (NMR), fluorescence spectroscopy (including the analysis of organic substances with fluorescence properties, and of inorganic compounds by total reflection X-ray fluorescence), thin layer chromatography (TLC), liquid chromatography (LC), anion exchange chromatography (HPAEC), gas chromatography (GC), stable isotope ratio mass spectrometry (IRMS), and electroanalytical methods.

Among all the listed techniques, the use of IR (Li, Huang, Xia, Xiong, & Mina, 2020; Su & Sun, 2019; Chien, Shih, Yang, & Hsiao, V.K.S., 2019), NMR (He, Liu, Liu, Zheng, Shen & Feng, 2020; Consonni, & Cagliani, 2018), IRMS (Geană, Ciucure, Costinel, & Ionete, 2020; Xu, Liu, Wu, & Cao, 2020; Zhou, Taylor, Salouros, & Prasad, 2018) and chromatography-based techniques (Arroyo-Manzanares et al, 2019; Makowicz, Kafarski, & Jasicka-Misiak, 2018; Jandrić, Frew, Fernandez-Cedi, & Cannavan, 2017) has increased a lot the last years. IRMS has been mostly used to detect the addition of exogenous sugars to honey (Geană et al., 2020), but also to classify honey based on their botanical origin (Xu et al., 2020) and on their geographical origin (Zhou et al., 2018). Although a sensitive and efficient approach, it is an expensive technique that requires developed analytical skills and it is not fit for screening purposes (Zábrodská & Vorlová, 2015).

IR, NMR and some chromatographic methods are used for untargeted analysis, that have the advantage of measuring many compounds, known and unknown, and hence making available all the information that they encrypt. IR is an economically competitive technique, easy to implement and with a high sample throughput, and has been widely applied in honey analysis in recent years. The availability of hand-held devices makes it ideal for on-site analysis. Nevertheless, as in all untargeted techniques, standardisation is strongly needed if results obtained by different laboratories are to be compared (Esslinger, Riedl, & Fauhl-Hassek, 2014). Some manufacturers of NMR instruments are investing a lot of effort to standardise the obtained results, but NMR remains an expensive technique that requires highly specialised analysts to run the analysis and to interpret the results.

Macro and trace elements have been used to discriminate food with Protected Denomination of Origin (Gonzalvez, Armenta & de la Guardia, 2009), including honey with Granada PDO (de Alda-Garcilope, Gallego-Picó, Garcinuño-Martínez & Fernández-Hernando, 2012) and from Tenerife (Hernández, Fraga, Jiménez, Jiménez & Arias, 2005). The approach has gained popularity in the last years because most control laboratories have the proper instrumentation and analytical expertise in place. Pawel (Pawel, 2009) published a review on the determination of the metal content of honey using atomic absorption (AAS) and atomic emission (AES) spectrometry. Also inductively coupled plasma-mass spectrometry (ICP-MS) (Magdas et al., 2021; Squadrone et al., 2020; Karabagias, Louppis, Badeka, Papastephanou & Kontominas, 2019; Batista et al., 2012) has been used to estimate the geographical origin of honey. As discussed in the literature (Kropf, Stibilj, Jaćimović, Bertoncelj, Golob & Korošec, 2017), analyses of honey have associated problems. Honey tends to crystallise, what can introduce heterogeneity problems when a small amount of sample is used for analysis; losses due to combustion can also take place during sample digestion. These problems can be eliminated using techniques such as neutron activation analysis (k_0-NAA) and ED-XRF that use larger amount of samples than ICP-MS, ICP-AES and AAS, and do not require sample digestion. k_0-NAA has been used to differentiate monovarietal Slovenian honeys of different botanical varieties (Kropf et al., 2017), but it requires the use of very expensive instrumentation and due to the needed infrastructure (nuclear reactor), it is not an alternative for most control laboratories.

Total reflection X-ray fluorescence has been used to classify honey according to botanical variety (Kropf et al., 2010) and recently, energy dispersive X-ray fluorescence (ED-XRF) has been successfully used to classify honeys of nine botanical varieties from seven countries (Fiamegos, Dumitrascu, Ghidotti, & de la Calle Guntiñas, 2020). The biggest advantage of ED-XRF compared to AAS, AES and ICP-MS based methods, is that no sample treatment is needed, what increases the throughput of

samples and the simplicity of the analysis. Heterogeneity is neither a problem in ED-XRF analysis due to large sample intake (normally >1 g). Also, the determination of elements such as Cl and Br is straightforward, while precautions are needed when the method used involve sample digestion (Costa, Picoloto, Hartwig, Mello, Flores & Mesko, 2015). As in IR analysis, hand-held devices are available, what facilitates the use of ED-XRF on-site (Marguí, Zawisza, Sitko, 2014). However, ED-XRF is characterised by higher limits of quantification than ICP- and AAS-based techniques. To overcome this problem, pre-concentration steps need to be introduced. Next to traditional approaches such as lyophilisation and precipitation, recent trends in pre-concentration procedures are liquid phase microextraction, use of activated thin layers and of carbon nanotubes. A revision of existing approaches to analyse elements at trace and ultratrace levels in liquid samples, has been published (Marguí et al., 2014). However, pre-concentration steps increase the complexity of the analyses and nullifies the big advantage of ED-XRF in comparison with other multi-elemental analysis techniques. The limit of quantification can also be reduced increasing the time that the sample is irradiated, but a compromise needs to be reached between reduction of limit of quantification and sample throughput.

With the aim of testing whether ED-XRF without any pre-concentration step, is a suitable method to discriminate honeys of different PDOs within a country, and of different botanical origins within the same PDO, 183 honey samples of Spanish PDOs (Liébana, Granada and Tenerife) were analysed. Most honeys were mono-varietal (chestnut, broom, pennyroyal, tajinaste, fennel, rosemary, tedera, relinchón, avocado, malpica, agave, barilla and heather), but also honeydew and multifloral honeys were analysed. The mass fractions of P, Cl, K, Ca, Mn, Fe, Cu, Ni, Zn, Rb and Sr were used as input variables for Soft Independent Modelling of Class Analogy (SIMCA) and Least Partial Square Discriminant Analysis (PLS-DA) to classify samples. Also 18 honey samples from a well-defined Spanish zone without a PDO, El Bierzo, were included in the study to evaluate the specificity of the models used for classification.

2. Experimental

2.1. Samples

The present study was carried out in collaboration with the Regulatory Councils of the Liébana, Granada and Tenerife PDOs, respectively, and with the Asociación Berciana de Apicultores (ABERAPI). All honey samples were provided by the mentioned organisations, which provided also information about botanical origin of each sample according to pollen analysis. The study included the following varieties: 33 chestnut (20 Granada, 8 Tenerife, 5 El Bierzo), 24 broom (Tenerife), 16 pennyroyal (Tenerife), 13 tajinaste (Tenerife), 11 fennel (Tenerife), 10 rosemary (Granada), 8 tedera (Tenerife), 6 relinchón (Tenerife), 4 avocado (Tenerife), 4 malpica (Tenerife), 2 agave (Tenerife), 1 barrilla (Tenerife), 1 heather (Tenerife), 1 blackberry (El Bierzo), 49 honeydew (30 Liébana, 8 Tenerife, 11 El Bierzo) and 18 multiflora (17 Granada, 1 El Bierzo). The main set of samples consisted of 201 honey samples in total, Supplementary S1.

The capacity of the different models to flag honey samples that do not have any of the PDOs covered in this study or that were not produced in El Bierzo (specificity), was made using the results previously obtained in our laboratory with 131 honey samples from other geographical locations, referred to hereafter as NO-PDO. Summarising, the NO-PDO set of 131 samples included eleven botanical varieties (26 robinia, 20 orange, 18 lavender, 14 chestnut, 11 eucalyptus, 8 rosemary, 8 thyme, 8 lime, 7 manuka, 6 heather and 5 sunflower), and twenty countries of origin (1 Austria, 2 Bulgaria, 1 China, 3 Denmark, 8 France, 1 Germany, 1 Greece, 10 Hungary, 19 Italy, 3 Mexico, 1 Moldova Republic, 1 The Netherlands, 10 New Zealand, 1 Norway, 2 Poland, 6 Portugal, 11 Romania, 39 Spain, 6 Switzerland and 2 Uruguay, plus 2 EU blends and 1 of not determined origin). The models were constructed on the basis of

botanical origin, and since each model included honeys from different countries, the variability to be expected due to different geographical origins was covered. Models were constructed for groups with 5 or more observations. Unfortunately, the information about geographical origin stopped at the level of country and information about specific regions within a country was not available. Since no information about the exact geographical origin of the 39 Spanish honeys (11 orange, 8 rosemary, 6 lavender, 4 chestnut, 4 eucalyptus, 4 thyme, 1 heather and 1 sunflower) was available, it could be that some or all of the 8 rosemary and 4 chestnut honeys come from some of the PDO regions included in this study or from El Bierzo. Only 66 out of the 131 samples had been characterised by melissopalinology; the information provided on the labels about botanical variety for the remaining samples was regarded as being correct, as well as the information about geographical origin.

2.2. Reagents and standards

Deionised water obtained with a Milli-Q Plus system (>18.3 MΩ) (Millipore, Billerica, MA, USA), was used in blank measurements.

To evaluate the trueness of the method used, a multi-elemental stock solution containing Ag, Al, B, Ba, Bi, Ca, Cd, Co, Cr, Cu, Fe, Ga, In, K, Li, Mg, Mn, Na, Ni, Pb, Sr, Tl and Zn at 1000 mg kg^{-1} each (Merck) and individual solutions of Cl, P and Rb 1000 mg kg^{-1} (Merck), were diluted in deionised water to a final concentration of 10 mg kg^{-1}.

Two chocolate Certified Reference Materials (CRMs), ERM-BD512 (European Commission, JRC) and SRM 2384 (US NIST), were used to evaluate the influence of density in the accuracy of the method used to analyse honey. Table 1 summarises the results obtained in the analysis of the standard solution and CRMs.

Forty-seven CRMs (23 out of which, organic matrices) were used for calibration purposes. The accuracy of the method was evaluated with 21 CRMs (10 out of which, organic) and 4 reference materials of organic matrix. Detailed information on the specific CRMs used can be found elsewhere (Fiamegos & de la Calle Guntiñas, 2018).

2.3. Instrumentation and sample preparation

The Epsilon 5 (PANalytical, Almelo, The Netherlands) ED-XRF was used to carry out the determination of P, Cl, K, Ca, Mn, Fe, Ni, Cu, Zn, and Rb in the honey samples. The instrument and the performance characteristics achieved in the analysis of solid samples is described in detailed elsewhere (Fiamegos & de la Calle Guntiñas, 2018). To measure honey, holders for liquid samples were used and the measurements were carried out under a helium atmosphere and not under vacuum as it is done in the analysis of solids. The bottom of the liquid sample holders is made of 6 μm Mylar film that gave rise to blank values for some elements, in particular for P and Ca. To correct for the blank contribution, de-ionised water was measured ten times and the mass fractions obtained were recorded.

From storage to analysis, honey samples were kept at room temperature. After a thorough mixing of the samples with a metal-free spatula, approximately 10 g of honey were transferred to the ED-XRF holders for liquids, without undergoing any further pre-treatment.

The 183 PDO and 16 El Bierzo honeys were measured with two different methods. The first method was developed in our laboratory for the analysis of solid samples and has undergone a thorough validation (Fiamegos & de la Calle Guntiñas, 2018) based on the use of CRMs to evaluate the accuracy of the method, as described in 2.2. To the best of our knowledge, there is no honey CRM commercially available; however, wheat flour (DUWF-1, SOWW-1), rice flour (NIST 1568b), brown bread (BCR-191) and corn bran (BRAN-1), used to assessed the accuracy of the method, are matrices rich in carbohydrates and could be good representatives of the honey matrix, mostly composed of sugars.

One of the drawbacks of the analysis of liquid samples by ED-XRF is that in liquids a high X-ray scatter background decreases the signal-to-noise ratio (Marguí et al., 2014); the effect can be less intense in samples with a high density that are closer to a solid than an aqueous solution. A limited number of experiments were carried out to evaluate the trueness of the method when applied to the analysis of liquid samples. The multi-elemental solution containing all the elements determined in the honey samples described in 2.2 was measured, results are summarised in Table 1. Honey is a dense matrix and for that reason, trueness was also evaluated using two chocolate CRMs, ERM-BD512 and SRM 2384. During measurement, samples were kept at 40 °C and the density of chocolate at that temperature resembles to a large extend the one of honey. Unfortunately, only a reduced number of elemental mass fractions have been certified in the two chocolate CRMs. The results obtained in the analysis of the two chocolate CRMs are shown in Table 1.

For multivariate analysis the raw data obtained were used without any correction for bias or blank subtraction, avoiding in this way mathematical artefacts such as negative mass fractions due to the inherent standard deviation of the blank measurements. This could happen in honeys with low elemental mass fractions. Identical models are obtained with and without blank correction because multivariant analysis are based in analysis of co-variances, and addition or subtraction of a constant to random variables does not change their co-variance as the mean is also increased or decreased by the same constant.

In the manuscript by Fiamegos et al. (Fiamegos & de la Calle Guntiñas, 2018), working ranges and LOQs were calculated in an empirical way making use of RMs and CRMs. The drawback of this approach is that the working ranges can be only as broad as the range covered by the mass fractions in the available CRMs and RMs. The same applies to the determination of the LOQ (minimum mass fractions for which a certain trueness and precision was achieved) and the LOD (the lowest mass fraction that can be detected with a certain trueness but that does not meet the criteria for precision anymore). It is difficult to find CRMs with certified values that would met the mentioned criteria. Mass fractions of macro-elements such as P, Cl, K and Ca in food, are normally

Table 1
Analysis of Certified Reference Materials and Reference Materials to evaluate the trueness achieved with the ED-XRF method used. All values are in mg.kg^{-1}.

Element	Working range (mg kg^{-1})	LOQ (mg kg^{-1})	Expanded Uncertainty (%) (k = 2)	Standard solution		ERM-BD512		SRM 2384	
				X	x	X	x	X	x
P	10.00–15810	10.00	6.0	10	10.5 ± 0.6			3330 ± 210	2408 ± 145
Cl	10.00–19200	10.00	2.0	10	8.66 ± 0.17				
K	10.00–47500	10.00	3.0	10	8.63 ± 0.26			8650 ± 400	9197 ± 276
Ca	10.00–39960	10.00	3.5	10	12.9 ± 0.5			840 ± 74	916 ± 32
Mn	2.55–3660	2.55	11.0	10	13.0 ± 1.4	15.7 ± 0.6	14.3 ± 1.6	20.8 ± 1.3	23.6 ± 2.6
Ni	0.16–1140	0.16	25.0	10	10.4 ± 2.6	3.01 ± 0.23	2.21 ± 0.55		
Cu	1.20–112.8	1.20	10.5	10	8.77 ± 0.92	14.3 ± 0.7	11.3 ± 1.2	23.9 ± 1.0	20.0 ± 2.1
Fe	4.60–397560	4.60	6.5	10	11.8 ± 0.8			132 ± 11	155 ± 10
Zn	5.80–6952	5.80	6.5	10	12.3 ± 0.8			37.6 ± 1.9	35.8 ± 2.3
Rb	4.20– 470.0	4.20	5.0	10	18.2 ± 0.9				

X: Reference value, x: Value obtained with the ED-XRF method used.
Working ranges and LOQs for Mn, Ni, Cu, Fe, Zn, Rb, and uncertainties for all elements, as published elsewhere (Fiamegos & de la Calle Guntiñas, 2018).

higher than the actual LOQ of the method and that explains the satisfactory results obtained for the 10 mg kg^{-1} standard solutions (<LOQ for solids). For this reason, the LOQ for P, Cl, K and Ca in honey samples, was 10 mg kg^{-1}.

The main contributor to the standard uncertainty associated to the analysis of solid matrices is the within-pellet variation. The dispersion of results observed for the analysis of honey samples was similar to that of solid samples, hence the standard uncertainties calculated for the analysis of solids are also used in the analysis of honey.

The performance of the ED-XRF instrument was checked once a week with the reference sample FLX-S13 (Fluxana, Bedburg-Hau). The ED-XRF was recalibrated every week with the mentioned reference sample to correct the normal drift but no systematic bias was observed for any of the measured elements. Among all the organic samples included in the validation of the method, tobacco is the one in which more elements were certified and was hence used for quality control purposes. The tobacco CRM PVTL-6 (Institute of Nuclear Chemistry and Technology, Poland) was measured every week after measuring the FLX-S13 standard. No systematic bias was observed for any of the elements measured in honey along the period in which the study was carried out.

Models to classify the PDO and El Bierzo honeys, according to their geographical origin and botanical variety were built using the elemental profiles obtained with this method as independent variables.

The second method used to analyse the PDO and El Bierzo honeys, was the Auto Quantify Liquid application of the ε5 software (PANalytical). The 131 honeys samples with geographical origins other than the PDOs and El Bierzo mentioned in 2.1 (NO-PDO) had been previously analysed using the Auto Quantify Liquid application of the ε5 software (PANalytical). To be able to evaluate if these samples would be flagged as outliers by models built-up for the PDO and El Bierzo honeys, the PDO and El Bierzo honeys had also to be measured with the Auto Quantify Liquid application.

P, Cl, K, Ca, Mn, Fe, Ni, Cu, Zn, Br and Rb were quantified in all or some of the honey samples and all of them were used in multivariate analyses.

2.4. Statistical tools and multivariate analysis

Quantitative elemental profiles of honey samples were used for statistical analysis. No correction for bias or blank subtraction was made on elements concentration and the mass fractions (mg kg^{-1}) of P, Cl, K, Ca, Mn, Fe, Ni, Cu, Zn, Br and Rb were processed as such. Univariate t-tests (95% confidence interval) were performed with the software Statistica (TIBCO, Version 13.5.0.17), on sub-groups of honey samples to test significant differences in elemental mass fractions. Subsequently, the software SIMCA Version 15.0.2, Umetrics (Sartorius Stedim Biotech AS, Malmö, Sweden) was used to carry out multivariate analysis (Eriksson et al., 2013).

Principal Component Analysis (PCA), a non-supervised technique, was used to evaluate if the studied honeys form clusters separated from each other depending on their botanical and/or geographical origin. For classification purposes, two different supervised multivariate tools were used: Soft Independent Modelling of Class Analogies (PCA-Class) and Partial Least Square Discriminant Analysis (PLS-DA). The former maximises the similarities among observations within a certain population, the latter maximises the differences between two different populations. To avoid overfitting, the number of principal components was kept to three with some exceptions, such as some models in which rosemary honeys were involved, for which a drastic improvement in classification capacity was obtained using four principal components. Distances to models were normalised in units of standard deviation.

The Mahalanobis distance (DModX PS +), distance from a point to the centroid of the distribution (NIST, 2020), was used to assess the presence of extreme outliers. Samples with values larger than D$_{crit}$ (95% confidence interval) were flagged as outliers. Nevertheless, the authenticity of the samples of this study was guaranteed and outliers

could represent their natural variability. Therefore, only 2 samples from the NO-PDO set were eliminated from their respective models.

The DModX PS + was also used to detect false positives (FP) (specificity), which are samples not belonging to a certain class but not flagged as outliers by the respective class models. False negatives (FN) (sensitivity) are samples belonging to a class that are flagged as outliers by the respective class model. Sensitivity was assessed following two different approaches in function of the amount of observations available. When >15 observations were available, the populations were divided in two, one half was used to construct the model and the other half was used as test for classification purposes; in a second iteration the samples used for testing were used for modelling purposes and those initially used to construct the model became the test population. When less than 15 observations were available, different models were constructed leaving one honey sample out and using the resulting model to classify the samples left out (cross-validation).

Summarising, the approach followed to classify the honey samples consisted in two steps. In a first step samples were classified using PCA-Class; in case a sample was allocated into more than one class, it was classified as belonging to the one for which it had the highest probability of class membership. In a second step, false negatives were classified using a PLS-DA model constructed for two populations: the population with the botanical or geographical origin to which the sample belongs according to the provider of the sample (or to the label in the NO-PDO set) and the population indicated by PCA-Class as the highest probability of class membership.

Model sensitivity, specificity and accuracy were calculated according to Barbosa et al, 2016 (Barbosa et al, 2016).

3. Results and discussion

3.1. Elemental characterisation by ED-XRF

The determination of light elements such as Na, Mg and Al by ED-XRF, is characterised by high LOQ's. In this study the first element quantified in all the honey samples was P. Results summarised in Table 1 show that the method, initially validated for solid samples, can also be accurately applied to liquid samples such as an aqueous standard solution, and to dense samples such as chocolate warmed up at 40 °C, more similar to honey than the standard solution. Rb was the only element for which the mass fraction obtained in the standard solution was significantly biased. The LOQ of Rb in solid samples is 4.5 mg kg^{-1}, which is lower than the 10 mg kg^{-1} in the standard solution and therefore accurate results should have been obtained. Unfortunately, Rb was not certified in any of the two chocolate CRMs and therefore it is not possible to elucidate whether the bias is due to scatter of the radiation in the aqueous matrix or to any other matrix effect. Rb results in honey are to be considered as indicative. Phosphorous mass fraction in chocolate SRM 2384 was affected by a bias of −30%, while the result for the stock solutions was not. The determination of P and Ca are probably affected by a higher error than the other elements because results have to be corrected for the blank contribution coming from the Mylar film of the liquid holders. The concentrations of P in honey are low and in the same range as the standard solution, therefore no correction factor was applied.

Quantifiable mass fractions of Br were found in six out of the eight honeydew honey samples from Tenerife. Bromine was not included neither in the CRMs nor in the multi-elemental standard solution used to evaluate the trueness of the method for analysis of liquid samples. The LOQ of the method for Br for solid samples, 1.7 mg kg^{-1}, was also applied to honey.

The first statistical evaluation applied to the results was the comparison of the mean mass fraction and the median for each population. The median is a robust mean not affected by extreme outliers. Means and medians were in good agreement, indicating that the elemental mass fractions in the different honey populations are normally distributed.
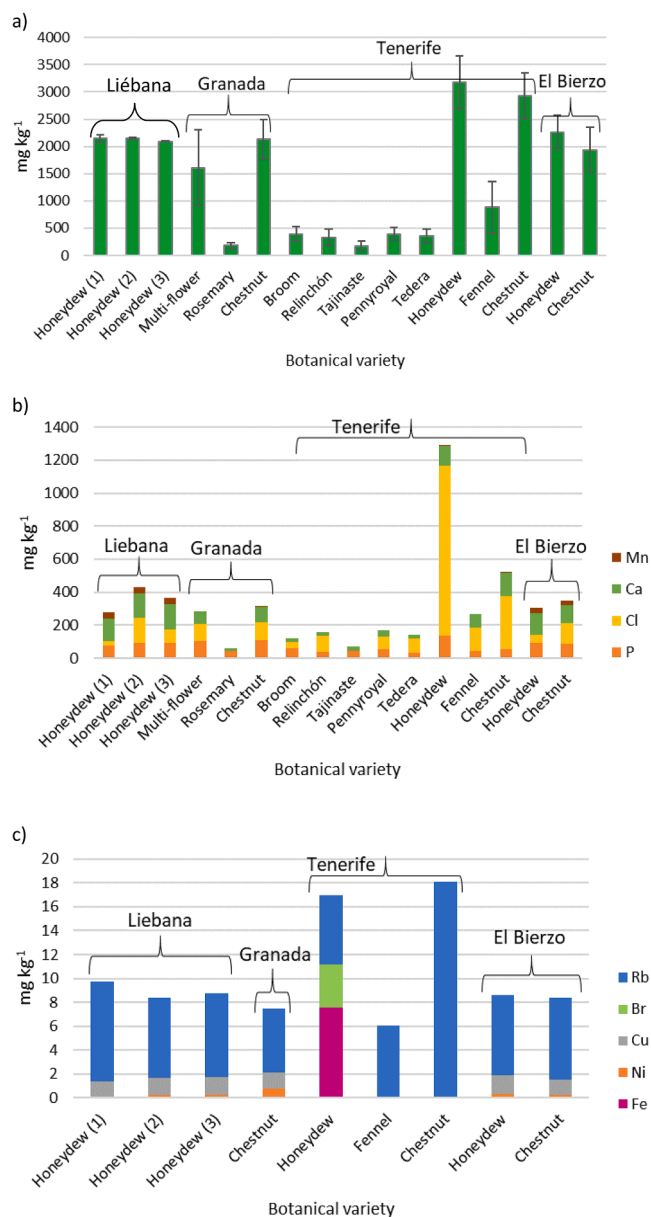
Fig. 1. Median of the elemental mass fractions found in the analysed honeys: **a)** K, **b)** Mn, Ca, Cl, P and **c)** Rb, Br, Cu, Ni, Fe.

Fig. 1 is constructed using the medians. Populations in which >50% of the observations are < LOQ are not included in Fig. 1.

As expected, the highest contents of macro- and trace-elements were found in honeydew and chestnut honeys regardless of their geographical origin, and the lowest in lighter honeys such as rosemary, broom, relinchón, tajinaste, pennyroyal and tedera. In particular, only K, Ca and P could be quantified in rosemary and tajinaste honeys, although it needs to be kept in mind that Ca and P mass fractions can only be considered as indicative for the reasons mentioned above, what is in good agreement with previous results (Fiamegos et al., 2020). Rosemary was the second poorest honey in elemental content after acacia honey. Multi-flower honeys from Granada were after honeydew and chestnut the richest honeys in macro- and trace-elements. Fennel honeys from Tenerife were situated between dark and light honeys.

Among the honeydew honeys the highest mass fractions for most elements were found in 6 out of 8 samples from Tenerife. Those samples were the only ones in the study with quantifiable amounts of Fe and Br. The chlorine mass fractions in the same samples were 10 to 30 times

higher than in the other honeydew honeys. Both Cl and Br are used for disinfection purposes, and the relatively high contents found in those samples could be due to contamination issues. The remaining two honeydew honey samples, with a Cl mass fraction comparable to the honeydew honeys from other regions, are the only two in the population with quantifiable Ni and Cu mass fractions and with the highest content of Zn. Only 3 honeydew and 2 avocado honeys, all of them from Tenerife, had quantifiable contents of Zn, Fe was quantified in a reduced number of samples, 2 multi-flower and 5 chestnut honeys from Granada and 5 honeydew, 1 tajinaste, 5 fennel, 1 chestnut and 2 avocado honeys from Tenerife. The population with a highest average content of Rb was that of the chestnut honeys from Tenerife (13.3 to 23.36 mg kg^{-1}), all the other samples had Rb mass fractions <13 mg kg^{-1}. Also the honeydew honeys from Tenerife were comparable to the other honeydew honeys in Rb mass fractions.

The highest K contents were found in the honeydew and chestnut honeys from Tenerife, Fig. 1 a. The same botanic origins in other regions had K mass fractions comparable among themselves, with ranges that overlapped within their standard deviations. In general, honeydew and chestnut honeys from Tenerife have the highest contents for all the elements, with the exception of Ca and Mn. The volcanic origin of the island of Tenerife could be the explanation for the mentioned high contents since volcanic soils are rich in minerals.

Manganese is an interesting element in the study. Honeydew honeys from Liébana and El Bierzo and chestnut honeys from El Bierzo contain approximately four times more Mn than the other samples of the same botanical origin. Also, the multiflower and blackberry honeys from El Bierzo contain more Mn than honeydew and chestnut honeys from Granada and Tenerife. Only two multiflower honeys from Granada contain quantifiable amounts of Mn but six times lower that the multi-flower honey from el Bierzo. Spanish soils are poor in manganese oxide with some exceptions, one of them is the zone where Liébana and El Bierzo are situated, where top and sub-soils are rich in MnO (Gómez-Miguel & Sotés, 2014). The differences in Mn content are so large that it could be on its own a discriminant marker of honeys from Liébana and El Bierzo.

The results obtained in this study for K, Ca, Zn and Fe in chestnut, rosemary and multi-flower honeys from Granada are in good agreement with those in the literature (de Alda-Garcilope et al., 2012) keeping in mind that differences in the elemental content can be expected depending for instance on the production year. The results obtained for Fe, Zn, K and Ca in the honeys from Tenerife were also in good agreement with values previously published (Hernández et al., 2005), although in that study no differentiation of the mass fractions based on botanical origin of the 62 analysed honeys was done. The values reported for Rb in the work by Hernández et al., vary in the range 0.00–3193 mg kg^{-1}; authors indicate that the determination of Rb with the atomic spectroscopy method used could be affected by ionisation interferences and that KCl was added as suppressor. In the present study all the values obtained for Rb were lower than 24 mg kg^{-1}, being in better agreement with those published in the literature for honeys from different geographical and botanical origins (Squadrone et al, 2020; Kropf et al., 2010) than those obtained by Hernández et al.. The mass fractions for Br, Ca, K, Rb, and Sr, obtained by ED-XRF are in good agreement with those obtained by neutron activation analysis (k$_0$-NAA) in particular for chestnut honey, keeping in mind that some of the values reported in that study are below the quantification limits of this ED-XRF method (Kropf et al., 2017). Geographical origin also needs to be considered since the honeys analysed by k$_0$-NAA are Slovenian.

3.2. Multivariate analysis of data

With the exception of the Mn case, the differences in elemental composition were not large enough to classify honey following a univariant approach. However, the t-tests (95% confidence interval) run to elucidate whether the honeys studied had significant differences in the

mass fractions of some elements based on botanical and/or geographical origin showed promising results. Therefore, multivariate analysis was undertaken to maximise the information that elemental mass fractions can provide.

The unsupervised tool PCA was used to visualize if clusters of honey are formed based on botanical variety and/or geographical origin. For classification purposes, the supervised techniques PCA-Class and binary PLS-DA were used.

3.2.1. Classification performance achieved by PCA-Class and PLS-DA

Table 2 summarises the performance characteristics (sensitivity, specificity and accuracy) of the models built-up using PCA-Class and PLS-DA when applied to the classification of Granada, Tenerife, Liébana and El Bierzo honeys. The outcomes were divided into three groups: 1) classification of honeys of different botanical origin within a PDO/region, 2) classification of honeys according to PDO's/region, and, 3) classification of the 201 honeys from Granada, Tenerife, Liébana and El Bierzo and 131 NO-PDO honeys.

The best classification performance was achieved for honeydew and chestnut honeys irrespective of which PDO they belonged, even when models were constructed with as few as eight observations, as it was the case of honeydew and chestnut honey from Tenerife. The accuracy of honeydew honey models varied from 78 to 100% and from 98 to 100% with PCA-Class and PLS-DA, respectively. The term 'accuracy' describes the proportion of correct classifications (proportion of true-positives plus true-negatives) obtained with the models. The accuracy of chestnut honey models varied from 82 to 98% and from 98 to 100% with PCA-Class and PLS-DA, respectively. The exception to this rule are the models of honeydew and chestnut honeys from El Bierzo; four honeydew honeys were classified as chestnut honey and four chestnut honeys were classified as honeydew honey, which resulted in both cases in accuracies of around 50%, both with PCA-Class and PLS-DA. This elevated number of false negatives and false positives could be caused by different criteria applied to classify chestnut honeys based on the results of pollen analysis. One honey from El Bierzo containing > 45% pollen from *Castanea sativa* was classified as chestnut honey, while chestnut honeys from Tenerife and Granada PDO must contain at least 75% pollen of *Castanea sativa*, Table 3. The accuracy of the models of honeydew and chestnut honeys from El Bierzo follow the same trends than those from Tenerife, Liébana and Granada, when applied to differentiate honeys from El Bierzo from those from other regions.

The accuracy obtained for light honey models, was lower than those of dark honeys. This is mostly due to the high rate of false negatives (low sensitivity), and it is more prominent in PCA-Class than in PLS-DA. The elemental mass fractions for most elements in light honeys, are lower than in dark honeys and so is the amount of quantifiable elements (variables with a strong contribution to the model). Tajinaste (Tenerife) is together with rosemary (Granada) the honey with the lowest elemental content and among the light ones its classification models had the poorest selectivity in PCA-Class (8 to 31%), which improved by using PLS-DA (69 to 100%). The most likely reason for the poor performance of some classification models is the low number of observations in some of the honey groups, such as rosemary, tedera and relinchón with 10, 8 and 6 observations, respectively. The sensitivity of the broom model with 24 observations is higher (20 to 83% by PCA-Class and 83 to 92% by PLS-DA) than in those previously mentioned.

Apparently, PLS-DA is a better option than PCA-Class if only few datasets are available to train the models, as it was the case for light honeys. Nevertheless, none of the two approaches is drawback free and as supervised techniques they require appropriate access to representative samples for training. However, it is virtually impossible to construct models for all possible worldwide botanical-geographical origin combinations. A honey sample that does not belong to any of the two or more populations for which models are constructed, but that is more similar to one of them than to the other/s, could be wrongly allocated to that population, resulting in a false positive. This problem

affects more PLS-DA that bases classifications on differences between populations rather than on similarities within a population. PCA-Class could overcome this problem when the models for the different populations are well defined by a number of variables with high discriminatory power, or by increasing the number of observation if the former is not possible.

3.2.2. Classification of honeys of different botanical origin within a PDO/ region

No honey, even if labelled as unifloral or mono-varietal, is obtained from a single botanical source. When analysed by melissopalinogy, a honey is considered unifloral when at least 45% of the pollen comes from a particular botanical variety, although there are exceptions to this rule (Von Der Ohe et al., 2004). Pollen from other species is inevitably present, introducing challenges in the classification by melissopalinology and by any other approach. The minimum amount of pollen of the species giving their names to unifloral honeys and honeydew honeys included in the study are summarised in Table 3, as extracted from the DOOR repository (DOOR, 2020) of the European Commission. Honeys from El Bierzo are not registered in DOOR because they do not hold a PDO; according to the information reported by ABERAPI, most honeydew honeys contain pollen from oak, holm oak, heather, chestnut and blackberry.

Maybe the most useful application of the approach described in this work is the classification of one honey within a PDO/region according to their botanical origin. The sources of variation of the elemental honey composition is reduced after elimination of the variations due to geographical origin. Fig. 2 a shows the PCA score plot of the Tenerife broom-tajinaste honeys, showing that these two light honeys tend to form separated clusters based on their elemental composition. Neither by PCA-Class nor by PLS-DA any broom honey was classified as tajinaste nor the other way around, Table 2.

The method could also be applied to identify small differences among honeys with the same botanical origin related to different locations within the same PDO/region and/or to differences in the way the honey is processed. An interesting case is that of the honeydew honeys from Liébana. All samples received from the Liébana PDO are honeydew honeys from three different beekeepers, referred to as 1, 2 and 3 in Fig. 2 b and in Table 2. The PCA score plot in Fig. 2 b shows three perfectly differentiated clusters, relating each one of them to a different beekeeper.

Another application within a PDO/region is the classification of honeys as multi- or unifloral, the latter being normally more expensive on the market than the former. Fig. 2 c shows the PCA score plot, of the three types of honey received from Granada: multiflower, rosemary and chestnut. As expected, the rosemary and chestnut honeys form two clusters perfectly separated; chestnut honeys are very rich in major and trace elements and rosemary has very low mass fractions for most elements. A third cluster is formed by most multiflower honeys although three of them cluster together with the rosemary honeys; this makes 8 rosemary honeys to be classified by PCA-Class as multi-flower honeys, whose cluster is characterised by a larger dispersion than that of rosemary. PLS-DA succeeded to classify correctly 100% of the rosemary honeys. The three mentioned multi-flower honeys should receive special attention during pollen analysis since their elemental composition strongly resembles that of rosemary honeys. The same applies to one chestnut honey projected in the PCA score plot together with the multi-flower honeys and that was classified as such by PLS-DA.

The set of honeys with Tenerife PDO included 4 avocado, 4 malpica, 2 agave, 1 barrilla and 1 heather. Although the small number of samples in each population did not allowed a thorough evaluation of data, the PCA-Class models of the other botanical varieties from Tenerife were challenged to test if they were specific enough to flag those honeys as outliers. Promising results were obtained since only two tendencies were observed; avocado honeys were accepted by the honeydew honey model and malpica honeys by the fennel model. More samples are needed from

Table 2
Classification performance characteristics of the models constructed.

Botanical variety/ Geographical origin	Performance within region				Performance between regions included in this study				Performance between regions in and outside this study			
	SIMCA		PLS-DA		SIMCA		PLS-DA		SIMCA		PLS-DA	
Liébana												
Honeydew (1) [10]	Sen.: 100 % Spe.: 100 % Acc.: 100 %		Sen.: 100 % Spe.: 100 % Acc.: 100 %		Sen.: 63 %	FN:10 honeydew El Bierzo, 1 not classified	Sen.: 100 % Spe.: 95 % Acc.: 98 %	FP:1 Honeydew El Bierzo	Sen.: 80 % Spe.: 100 % Acc.: 100 %	FN:4 heather, 2 not classified	Sen.: 100 % Spe.: 100 % Acc.: 100 %	
Honeydew (2) [10]	Sen.: 100 % Spe.: 100 % Acc.: 100 %		Sen.: 100 % Spe.: 100 % Acc.: 100 %		Spe.: 100 % Acc.: 78 %							
Honeydew (3) [10]	Sen.: 100 % Spe.: 100 % Acc.: 100 %		Sen.: 100 % Spe.: 100 % Acc.: 100 %									
Granada												
Chestnut [20]	Sen.: 75 % Spe.: 96 % Acc.: 87 %	FN:5 not classified FP:1 multi-flower	Sen.: 95 % Spe.: 100 % Acc.: 98 %	FN:1 multi-flower	Sen.: 75 % Spe.: 92 % Acc.: 82 %	FN:1 chestnut El Bierzo, 4 not classified FP:1 chestnut El Bierzo	Sen.: 100 % Spe.: 92 % Acc.: 97 %	FP:1 chestnut El Bierzo	Sen.: 70 % Spe.: 100 % Acc.: 91 %	FN: 4 chestnut, 2 heather	Sen.: 100 % Spe.: 100 % Acc.: 100 %	
Rosemary [10]	Sen.: 20 % Spe.: 100 % Acc.: 81 %	FN:8 multi-flower	Sen.: 100 % Spe.: 100 % Acc.: 100 %		Sen.: 20 % Spe.: 100 % Acc.: 84 %	FN:7 pennyroyal, 1 tajinaste	Sen.: 100 % Spe.: 100 % Acc.: 100 %		Sen.: 50 % Spe.: 100 % Acc.: 95 %	FN:2 robinia, 3 rosemary *	Sen.: 80 % Spe.: 100 % Acc.: 98 %	FN:2 robinia
Multi-flower [17]	Sen.: 82 % Spe.: 73 % Acc.: 77 %	FN:1 chestnut, 2 not classified FP:8 rosemary	Sen.: 100 % Spe.: 97 % Acc.: 98 %	FP:1 chestnut	Sen.: 94 % Spe.: 40 % Acc.: 50 %	FN:1 pennyroyal FP:14 broom, 5 relinchón, 2 fennel, 4 tedera, 10 tajinaste, 13 pennyroyal (Tenerife)	Sen.: 94 % Spe.: 81 % Acc.: 84 %	FN:1 pennyroyal FP:2 broom, 3 relinchón, 2 fennel, 2 tedera, 6 pennyroyal (Tenerife)	Sen.: 59 % Spe.: 83 % Acc.: 80 %	FN:3 eucalyptus,2 rosemary, 1 chestnut, 1 heather FP:4 robinia, 4 orange, 1 eucalyptus, 1 heather, 3 lavender, 1 lime, 4 rosemery, 1 sunflower, 3 thyme	Sen.: 88 % Spe.: 92 % Acc.: 92 %	FN:1 eucalyptus,1 rosemary FP: 6 lavender, 2 orange, 1 robinia,1 sunflower
Tenerife												
Honeydew [8]	Sen.: 75 % Spe.: 100 % Acc.: 98 %	FN:2 not classified FN:2 not classified	Sen.: 100 % Spe.: 99 % Acc.: 99 %	FP:1 chestnut FN:1 honeydew	Sen.: 63 % Spe.: 100 % Acc.: 94 %	FN:3 not classified	Sen.: 100 % Spe.: 100 % Acc.: 100 %		Sen.: 100 % Spe.: 100 % Acc.: 100 %		Sen.: 100 % Spe.: 100 % Acc.: 100 %	FP:1 chestnut

Table 2 (*continued*)

Botanical variety/ Geographical origin	Performance within region				Performance between regions included in this study				Performance between regions in and outside this study			
	SIMCA		*PLS-DA*		*SIMCA*		*PLS-DA*		*SIMCA*		*PLS-DA*	
Chestnut [8]	Sen.: 75 % Spe.: 100 % Acc.: 98 %		Sen.: 88 % Spe.: 100 % Acc.: 99 %		Sen.: 75 % Spe.: 100 % Acc.: 94 %	FN:2 not classified	Sen.: 100 % Spe.: 100 % Acc.: 100 %		Sen.: 75 % Spe.: 99 % Acc.: 98 %	FN:1 chestnut, 1 heather FP:1 chestnut	Sen.: 100 % Spe.: 99 % Acc.: 99 %	
Fennel [11]	Sen.: 73 % Spe.: 92 % Acc.: 89 %	FN:1 pennyroyal, 2 not classified	Sen.: 73 % Spe.: 98 % Acc.: 95 %	FN:1 pennyroyal, 2 not classified	Sen.: 82 % Spe.: 100 % Acc.: 97 %	FN:2 multi-flower (Granada)	Sen.: 82 % Spe.: 100 % Acc.: 97 %	FN: 2 multi-flower (Granada)	Sen.: 30 % Spe.: 100 % Acc.: 86 %	FN:5 lime, 2 heather, 1 eucalyptus	Sen.: 82 % Spe.: 100 % Acc.: 97 %	FN:1 lime, 1 heather
Broom [24]	Sen.: 83 % Spe.: 99 % Acc.: 95 %	FN:1 pennyroyal, 1 fennel & pennyroyal, 2 not classified FP: 1 pennyroyal	Sen.: 92 % Spe.: 100 % Acc.: 98 %	FN:1 fennel, 1 pennyroyal	Sen.: 71 % Spe.: 100 % Acc.: 82 %	FN:14 multi-flower (Granada)	Sen.: 92 % Spe.: 100 % Acc.: 97 %	FN:2 Multi-flower (Granada)	Sen.: 21 % Spe.: 99 % Acc.: 84 %	FN:12 thyme, 6 lavender, 1 rosemary FP :1 robinia	Sen.: 83 % Spe.: 99 % Acc.: 95 %	FN:3 lavender, 1 thyme FP :1 robinia
Tajinaste [13]	Sen.: 8 % Spe.: 100 % Acc.: 87 %	FN: 11 pennyroyal, 1 not classified	Sen.: 69 % Spe.: 100 % Acc.: 96 %	FN:3 pennyroyal, 1 not classified	Sen.: 23 % Spe.: 99 % Acc.: 86 %	FN:10 multi-flower (Granada) FP: 1 rosemary (Granada)	Sen.: 100 % Spe.: 100 % Acc.: 100 %		Sen.: 31 % Spe.: 82 % Acc.: 75 %	FN:8 rosemary, 1 lavender FP:8 robinia, 4 rosemary, 1 orange, 1 lavender	Sen.: 77 % Spe.: 87 % Acc.: 86 %	FN:2 rosemary, 1 lavender FP:5 robinia, 3 rosemary, 1 orange, 1 lavender
Pennyroyal [16]	Sen.: 56 % Spe.: 74 % Acc.: 71 %	FN:4 fennel, 1 tedera, 1broom, 1 not classified	Sen.: 56 % Spe.: 87 % Acc.: 82 %	FN:2 tedera, 1 relinchón, 1fennel, 3 not classified	Sen.: 19 % Spe.: 87 % Acc.: 73 %	FN:13 multi-flower (Granada) FP:7 rosemary (Granada), 1 multi-flower (Granada)	Sen.: 63 % Spe.: 98 % Acc.: 91 %	FN:6 multi-flower (Granada) FP:1 multi-flower (Granada)	Sen.: 56 % Spe.: 84 % Acc.: 79 %	FN:4 thyme, 2 orange, 1 robinia FP:6 robinia, 3 orange, 2 rosemary, 1 lavender, 1 thyme	Sen.: 67 % Spe.: 96 % Acc.: 92 %	FN:3 thyme, 1 orange, 1 robinia FP:1 orange, 1 rosemary, 1 thyme
Tedera [8]	Sen.: 50 % Spe.: 99 % Acc.: 95 %	FN:2 pennyroyal, 1 fennel, 1 relichón	Sen.: 60 % Spe.: 98 % Acc.: 95 %	FN:3 pennyroyal	Sen.: 38 % Spe.: 90 % Acc.: 85 %	FN:4 multi-flower (Granada), 1 not classified FP:7 rosemary (Granada)	Sen.: 82 % Spe.: 94 % Acc.: 92 %	FN:2 multi-flower (Granada) FP:3 multi-flower (Granada), 1 rosemary (Granada)	Sen.: 63 % Spe.: 90 % Acc.: 87 %	FN:1 orange, 1 rsemary, 1 thyme FP:4 robinia, 2 lavender, 1 orange, 1 rosemary	Sen.: 63 % Spe.: 96 % Acc.: 93 %	FN:1 orange, 1 rsemary, 1 thyme FP:2 lavender, 1 orange
Relinchón [6]	Sen.: 20 % Spe.:	FN:4 pennyroyal, 1 fennel	Sen.: 67 % Spe.:	FN:2 pennyroyal	Sen.: 17 %	FN:5 multi-flower (Granada)	Sen.: 50 %	FN: 3 multi-flower (Granada)	Sen.: 17 % Spe.:	FN:3 rosemary, 2 thyme FP:2 rosemary, 1 robinia, 1 orange	Sen.: 50 % Spe.:	FN:2 rosemary, 1 thyme

Table 2 (continued)

Botanical variety/ Geographical origin	Performance within region		Performance between regions included in this study		Performance between regions in and outside this study	
	SIMCA	PLS-DA	SIMCA	PLS-DA	SIMCA	PLS-DA
	Spe.: 99% Acc.: 94%	99% Acc.: 97%	Spe.: 100% Acc.: 94%	Spe.: 100% Acc.: 96%	95% Acc.: 89%	96% Acc.: 93% FP:2 rosemary, 1 orange
El Bierzo						
Honeydew [11]	Sen.: 55% FN:4 chestnut, 1 not classified FP:3 chestnut, 1 multi-flower Spe.: 43% Acc.: 50%	Sen.: 64% FN:4 chestnut Spe.: 43% Acc.: 56%	Sen.: 73% FN:3 not clasified FP:10 honeydew Liébana Spe.: 74% Acc.: 74%	Sen.: 91% FN:1 Honeydew Liébana Spe.: 100% Acc.: 98%	Sen.: 73% FN:2 heather, 1 not classified Spe.: 100% Acc.: 96%	Sen.: 91% Spe.: 100% Acc.: 99% FN:1 heather
Chestnut [5]	Sen.: 0% FN:4 honeydew, 1 not classified FP:4 chestnut, 1 blackberry Spe.: 62% Acc.: 44%	Sen.: 20% FN:4 honeydew, 1 FP:4 honeydew, 1 blackberry, 1 multi-flower Spe.: 54% Acc.: 44%	Sen.: 80% FN:1 chestnut (Granada) FP:1 chestnut (Granada) Spe.: 96% Acc.: 94%	Sen.: 80% FN:1 chestnut (Granada) Spe.: 100% Acc.: 97%	Sen.: 80% FN:1 heather FP:1 chestnut, 1 heather, 1 manuka, 1 Thyme Spe.: 93% Acc.: 92%	Sen.: 100% Spe.: 97% Acc.: 97% FP:1 chestnut, 1 heather

Sensitivity = TP/(TP + FN), Specificity = TN/(TN + FP), Accuracy= (TP + TN)/(TP + TN + FP + FN) * Rosemary in the Rest population not belonging to the batch or rosemary honeys from Granada PDO. [n] Amount of observations in a population.

the mentioned botanical varieties to carry out conclusive studies, including sensitivity and specificity figures.

3.2.3. Classification of honey according to PDO/region

It is not likely that honey from a certain PDO will be sold with the label of a different PDO since very likely there would not be any financial gain involved. However, this study was carried out to evaluate the specificity of our models and thus their ability to detect honeys coming from Spanish regions other than that covered by their own PDO. This is relevant when applied to honeys of the same botanical origin. Fig. 2 d shows the PLS-DA score plot of the honeydew honeys from Liébana and from El Bierzo, which geographically are the closest regions among all included in this work. Honeydew honeys from both regions are characterised by a high Mn content, and no classification could be done exclusively based on Mn mass fractions, as could be the case with Tenerife honeydew honeys. The first thing observed is the larger dispersion of the honeydew honeys from El Bierzo in comparison with those from Liébana, which could be the consequence of the tighter restrictions in the full production process of PDO honeydew honeys. Other possible explanation for the mentioned difference in dispersion is that El Bierzo is a county around five times larger than Liébana, and the observed dispersion reflects the geological, climatological and biotope intrinsic differences of El Bierzo. This dispersion makes some honeydews honeys from Liébana not to be flagged as outliers by the PCA-Class honeydew honey model of El Bierzo, while the PCA-Class model of honeydew honeys from Liébana is characterised by a specificity of 100%, Table 2.

Again, the models constructed for honeydew and chestnut honeys by PLS-DA have the highest sensitivity and specificity, in all cases higher than 90%. Only the model of chestnut honeys from El Bierzo had a lower sensitivity, 80%, which is the result of the low number of observations and the large heterogeneity in that population.

Interestingly, the PLS-DA models for rosemary and tajinaste, the two types of honey with the lowest elemental contents, have in both cases 100% sensitivity and specificity. The low elemental contents make them different from all the other light honeys. On the other hand, their elemental profiles were different enough to differentiate the two populations from each other.

For the remaining honeys, sensitivity and specificity varied between 82 and 100% with the exception of pennyroyal and relinchón, in the

Table 3
Information about composition regarding pollen of botanical species.

| Honey description	Pollen requirements (DOOR, 2020)
Tenerife	
Broom	> 30% *Spartocytisus supranubis*
Tajinaste	> 9% *Echium* spp.
Chestnut	> 75% *Castanea sativa*
Relinchón	> 4% *Hirstfeldia ineana*
Fennel	> 2% *Foeniculum vulgare*
Pennyroyal	> 1% *Bistropogon origarcifolium*
Tedera	> 3% *Aspalthium bituminosum*
Avocado	> 2% *Persea americana*
Heather	> 25% *Erica* spp.
Barrilla	> 13% *Mesembry-anthemum crystalinum*
Malpica	> 4% *Carlina xeranthemoides*
Agave	Traces *agave americana*
Honeydews	> 0.3% honeydew elements, variable pollen spectrum with a high proportion of *anemophilus* species
Granada	
Chestnut	> 75% *Castanea sativa* (absence of *Erica spp*)
Rosemary	> 15% *Rosemarinus officinalis* or > 10% accompanied by > 5% of the *Lamiaceae* family.
Multi-flower	> 5% of the *lamiaceae* family
Liébana	
Honeydew	Oak and holm oak honeydew predominate, in addition *Rubus spp*. and *Erica spp*. Pollens and other pollens from Liébana characteristic flora.
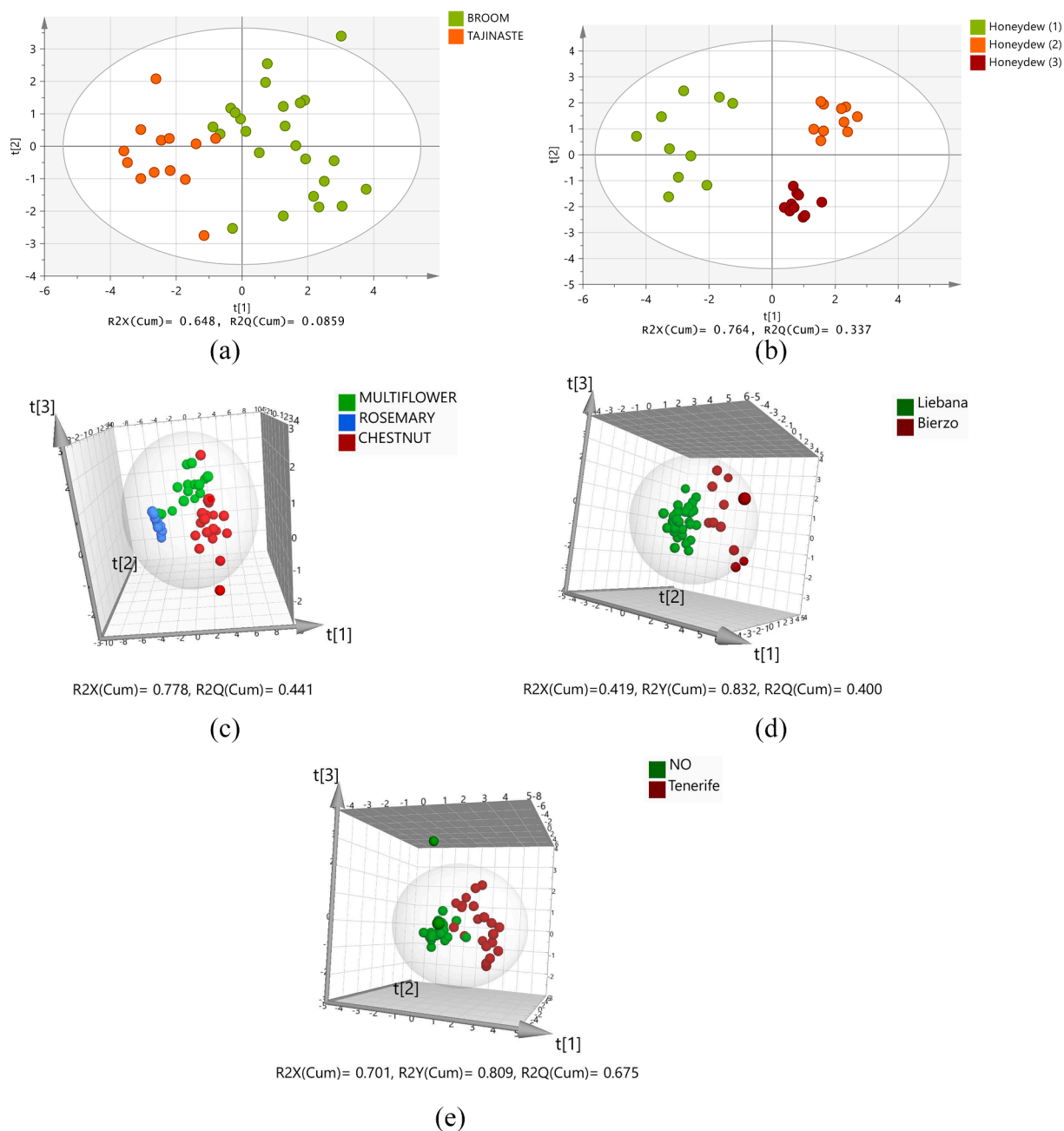
Fig. 2. a) PCA score plot (showing the two first principal components) of broom and tajinaste honeys from Tenerife, **b)** PCA score plot (showing the two first principal components) of honeydews from three different beekeepers from Liébana, **c)** PCA score plot (showing the three first principal components) of chestnut, rosemary and multiflower honeys from Granada, **d)** PLS-DA score plot (showing the three first principal components) of honeydews from Liébana and El Bierzo, **e)** PLS-DA score plot (showing the three first principal components) of broom honeys from Tenerife and robinia honeys from the NO-PDO population. Ellipse: Hotelling's T2 (95%).

latter case probably due to the small number of observations. Both the PCA-Class and PLS-DA models of pennyroyal honeys are characterised by poor sensitivity and specificity, maybe due to a large intrinsic variability of this type of honey.

Otherwise, most false negative and false positive results in this between-region study involve the multi-flower honeys from Granada: *unifloral* honeys from other regions wrongly classified as multi-flower and multi-flower wrongly classified as *unifloral* from other regions. The Granada multi-flower model clearly flagged as outlier the multi-flower honey from El Bierzo probably because multi-flower honeys are better mirrors of the local flora than the *unifloral* ones.

3.2.4. Robustness of the PDO classification models to detect non-PDO honeys

The last part of the study was the evaluation of the performance of the built-up models to differentiate the PDO and El Bierzo honeys, from honeys with different botanical and geographical origin, including honeys from outside Spain and from outside Europe, NO-PDO. The honeydew, chestnut and fennel honey models were challenged with dark NO-PDO honeys (chestnut, heather, eucalyptus, linden and manuka) to see whether the model is specific enough to flag them as not belonging to the PDO honeys. The rosemary, tajinaste, pennyroyal, tedera and relinchón models were challenged with light NO-PDO honeys (robinia, orange, lavender, rosemary, sunflower and thyme). The multiflower Granada model was challenged with light and dark NO-PDO

honey models. Sensitivity studies were also carried out.

As in previous studies, PLS-DA models were more accurate than PCA-Class models. PLS-DA models for honeydew and chestnut honey had sensitivities between 91 and 100% and specificity between 97 and 100%. Models of Tenerife light honeys were characterised by relatively poor sensitivities (high amount of false negatives), very likely due to the small number of observations: relinchón (50%, 6 observations), tedera (63%, 8 observations), tajinaste (77%, 13 observations) and broom (83%, 24 observations), the pennyroyal model being an exception to this rule (67%, 16 observations). Also the model of Granada rosemary honeys is an exception to this rule with 80% sensitivity and a relatively small number of observations 10 observations, although to achieve those results, the amount of principal components needed to be increased from three to four. An increase of the amount of principal components did not improve the results in the case of pennyroyal and tajinaste. The number of observations in the tedera and relinchón models is not high enough to apply this approach without taking the risk of overfitting. Fennel seems to be a type of honey that from the elemental point of view can be placed between the dark and the light honeys, and the respective classification model was characterised by 82% sensitivity with 11 observations.

The specificity of the light honey models varied in the range 86% for tajinaste and 98% for rosemary. An example of the clusters formed by some light honeys is shown in Fig. 2 e. The false positives among the NO-PDO honeys follow some patterns, which was expected from the characteristics of the different honeys. Robinia honeys, poorer in elemental content than rosemary honey (Fiamegos et al., 2020) and tajinaste, represent the highest number of false positives in the tajinaste model, followed by rosemary. Lavender followed by orange accounted for the highest amount of false positives in the Granada multiflower model. Lavender and orange trees grow in the region of Granada, and the multiflower honeys from that region may contain pollen from those two types of plants. The PDO Granada also has uniflora lavender and orange honeys which were not included in this study because they were not available when the study was conducted.

4. Conclusions

ED-XRF is a suitable screening technique to be used in the classification of honeys from a certain PDO or region, providing important information to support pollen analysis. The performance of the classification models built-up in this way for dark honeys is particularly good even with a reduced number of samples. The method can also be applied to light honeys but would profit from increasing the number of training samples. The method could be particularly useful for the pre-classification of honeys within a PDO or region as uniflora or multiflower, allowing to concentrate the efforts of pollen analysis in a reduced number of samples. Ranges for certain major and trace elements could be added to the list of characteristics of a certain honey in the DOOR repository, to be used for classification purposes.

CRediT authorship contribution statement

Michele Ghidotti: Formal analysis, Data curation, Validation, Investigation, Methodology. **Yiannis Fiamegos:** Formal analysis, Data curation, Validation, Methodology. **Catalina Dumitrascu:** Formal analysis, Methodology. **María Beatriz de la Calle:** Data curation, Validation, Investigation, Project administration, Resources, Supervision.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2020.128350.

References

Anklam, E. (1998). A review of the analytical methods to determine the geographical and botanical origin of honey. *Food Chemistry, 63*(4), 549–562.

Arroyo-Manzanares, N., García-Nicolás, M., Castell, A., Campillo, N., Viñas, P., López-García, I., et al. (2019). Untargeted headspace gas chromatography – Ion mobility spectrometryanalysis for detection of adulterated honey. *Talanta, 205*, Article 120123. https://doi.org/10.1016/j.talanta.2019.120123.

Barbosa, R., Silva de Paula, E., Paulelli, A. C., Moore, A. F., Oliveira Souza, J. M., Lemos Batista, B., et al. (2016). Recognition of organic rice samples based on trace elements and support vector machines. *Journal of Food Composition and Analysis, 45*, 95–100. https://doi.org/10.1016/j.jfca.2015.09.010.

Batista, B. L., da Silva, L. R. S., Rocha, B. A., Rodrigues, J. L., Berretta-Silva, A. A., Bonates, T. O., et al. (2012). Multi-element determination in Brazilian honey samples by inductively coupled plasma mass-spectrometry and estimation of geographic origin with data mining techniques. *Food Res Int., 49*, 209–215. https://doi.org/10.1016/j.foodres.2012.07.015.

Bentabol Manzanares, A., Hernández García, Z., Rodríguez Galdón, B., Rodríguez Rodríguez, E., & E., Díaz Romero, C.,. (2011). Differentiation of blossom and honeydew honeys using multivariate analysis on the physicochemical parameters and sugar composition. *Food Chemistry, 126*, 664–672. https://doi.org/10.1016/j.foodchem.2010.11.003.

Chien, H. Y., Shih, A. T., Yang, B. S., Vincent, K. S., & Hsiao, V. K. S. (2019). Fast honey classification using infrared spectrum and machine learning. *Mathematical Biosciences and Engineering, 16*(6), 6874–6891. https://doi.org/10.3934/mbe.2019344.

Consonni, R., & Cagliani, L. R. (2018). The potentiality of NMR-based metabolomics in food science and food authentication assessment. *Analytical Science Advances.* https://doi.org/10.1002/mrc.4807.

Costa, V. C., Picoloto, R. S., Hartwig, C. A., Mello, P. A., Flores, E. M. M., & Mesko, M. F. (2015). Feasibility of ultra-trace determination of bromine and iodine in honey by ICP-MS using high sample mass in microwave-induced combustion. *Anal Bioanal Chem, 407*, 7957–7964. https://doi.org/10.1007/s00216-015-8967-9.

Council Regulation (EC) No 510/2006 of 20 March 2006 on the protection of geographical indications and designations of origin for agricultural products and foodstuffs.

de Alda-Garcilope, C., Gallego-Picó, A., Bravo-Yagüe, J. C., Garcinuño-Martínez, R. M., & Fernández-Hernando, P. (2012). Characterization of Spanish honeys with protected designation of origin "Miel de Granada" according to their mineral content. *Food Chemistry, 135*, 1785–1788. https://doi.org/10.1016/j.foodchem.2012.06.057.

DOOR https://ec.europa.eu/agriculture/quality/door/list.html?&recordStart=16&filter.dossierNumber=&filter.comboName=&filterMin.milestone__mask=&filterMin.milestone=&filterMax.milestone__mask=&filterMax.milestone=&filter.country=&filter.category=PDOPGI_CLASS_14&filter.type=PDO&filter.status=.REGISTERED. (Last accessed on 05-05-2020).

Eriksson, L., Byrn, T., Johansson, E., Trygg, J., & Vikström, C. (2013). *Multi-and Megavariate Data Analysis*. Umetrics Academy, Malmö, Sweden: Basic Principles and Applications.

Esslinger, S., Riedl, J., & Fauhl-Hassek, C. (2014). Potential and limitations of non-targeted fingerprinting forauthentication of food in official control. *Food Research International, 60*, 189–204. https://doi.org/10.1016/j.foodres.2013.10.015.

Fiamegos, Y., & de la Calle Guntiñas, M. B. (2018). Validation strategy for an ed-xrf method to determine trace elements in a wide range of organic and inorganic matrices based on fulfilment of performance criteria. *Spectrochimica Acta Part B, 150*, 59–66. https://doi.org/10.1016/j.sab.2018.10.009.

Fiamegos, Y., Dumitrascu, C., Ghidotti, M., & de la Calle Guntiñas, M. B. (2020). Use of energy-dispersive X-ray fluorescence combined with chemometric modelling to classify honey according to botanical variety and geographical origin. *Analytical and Bioanalytical Chemistry, 412*, 463–472. https://doi.org/10.1007/s00216-019-02255-6.

Geană, E. I., Ciucure, C. T., Costinel, D., & Ionete, R. E. (2020). Evaluation of honey in terms of quality and authenticity based on the general physicochemical pattern, major sugar composition and δ^{13}C signature. *Food Control, 109*, Article 106919. https://doi.org/10.1016/j.foodcont.2019.106919.

Gómez-Miguel, V. D., & Sotés, V. (2014). *El Manganeso y la Viticultura: Una revisión.* Alimentación y Medio Ambiente: Ministerio de Agricultura.

Gonzalvez, A., Armenta, S., & de la Guardia, M. (2009). Trace-element composition and stable-isotope ratio for discrimination of foods with Protected Designation of Origin. *Trends in Analytical Chemistry, 28*(11), 1295–1311. https://doi.org/10.1016/j.trac.2009.08.001.

He, C., Liu, Y., Liu, H., Zheng, X., Shen, G., & Feng, J. (2020). Compositional identification and authentication of Chinese honeys by ^{1}H NMR combined with multivariate analysis. *Food Research International, 130*, Article 108936. https://doi.org/10.1016/j.foodres.2019.108936.

Hernández, O. M., Fraga, J. M. G., Jiménez, A. I., Jiménez, F., & Arias, J. J. (2005). Characterization of honey from the Canary Islands: Determination of the mineral content by atomic absorption spectrophotometry. *Food Chemistry, 93*, 449–458. https://doi.org/10.1016/j.foodchem.2004.10.036.

Jandrić, Z., Frew, R. D., Fernandez-Cedi, L. N., & Cannavan, A. (2017). An investigative study on discrimination of honey of various floral and geographical origins using UPLC-QToF MS and multivariate data analysis. *Food Control, 72*, 189–197. https://doi.org/10.1016/j.foodcont.2015.10.010.

Karabagias, I. K., Louppis, A. P., Badeka, A., Papastephanou, C., & Kontominas, M. G. (2019). Nutritional aspects and botanical origin recognition of Mediterranean honeys based on the "mineral imprint'' with the application of supervised and non-supervised statistical techniques. *European Food Research and Technology, 245*, 1939–1949. https://doi.org/10.1007/s0021 7-019-03306 -z.

Kropf, U., Korošec, M., Bertoncelj, J., Ogrine, N., Nečemer, M., Kump, P., et al. (2010). Determination of the geographical origin of Slovenian black locust, lime and chestnut honey. *Food Chemistry, 121*, 839–846. https://doi.org/10.1016/j.foodchem.2009.12.094.

Kropf, U., Stibilj, V., Jaćimović, R., Bertoncelj, J., Golob, T., & Korošec, M. (2017). Elemental Composition of Different Slovenian Honeys Using k_0-Instrumental Neutron Activation Analysis. *Journal of AOA C International, 100*(4), 871–880. https://doi.org/10.5740/jaoacint.17-0146.

La Serna Ramos, I. E., & Gómez Ferreras, C. (2011). An example of the role of exotic flora in the geographical characterisation of honey: Schinus molle L. in the Canary Islands (Spain). *Grana, 50*, 136–149. https://doi.org/10.1080/00173134.2011.578656.

Laube, I., Hird, H., Brodmann, P., Ullmann, S., Schöne-Michling, M., Chisholm, J., et al. (2010). Development of primer and probe sets for the detection of plant species in honey. *Food Chemistry, 118*, 979–986. https://doi.org/10.1016/j.foodchem.2008.09.063.

Li, Y., Huang, Y., Xia, J., Xiong, Y., & Mina, S. (2020). Quantitative analysis of honey adulteration by spectrum analysis combined with several high-level data fusion strategies. *Vibrational Spectroscopy, 108*, Article 103060. https://doi.org/10.1016/j.vibspec.2020.103060.

Magdas, D. A., Guyon, F., Puscas, R., Vigouroux, A., Gaillard, L., Dehelean, A., et al. (2021). Applications of emerging stable isotopes and elemental markers for geographical and varietal recognition of Romanian and French honeys. *Food Chemistry, 334*, Article 127599. https://doi.org/10.1016/j.foodchem.2020.127599.

Makowicz, E., Kafarski, P., & Jasicka-Misiak, I. (2018). Chromatographic fingerprint of the volatile fraction of rare Hedera helix honey and biomarkers identification. *European Food Research and Technology, 244*(12), 2169–2179. https://doi.org/10.1007/s00217-018-3127-z.

Marguí, E., Zawisza, B., & Sitko, R. (2014). Trace and ultratrace analysis of liquid samples by X-ray fluorescence spectrometry. *Trends in Analytical Chemistry, 53*, 73–83. https://doi.org/10.1016/j.trac.2013.09.009.

NIST, https://www.itl.nist.gov/div898/software/dataplot/refman2/auxillar/matrdist.htm (Last accessed on 10 May 2020.

Pawel, P. (2009). Determination of metal content in honey by atomic absorption and emission spectrometries. *Trends in Analytical Chemistry, 28*(1), 117–128. https://doi.org/10.1016/j.trac.2008.09.015.

Squadrone, S., Brizio, P., Stella, C., Pederiva, Brusa, S. F., Mogliotti, P., Garrone, A., & Abete, M. C. (2020). Trace and rare earth elements in monofloral and multifloral honeys from Northwestern Italy; A first attempt of characterization by a multi-elemental profile. *Journal of Trace Elements in Medicine and Biology, 61*, Article 126556. https://doi.org/10.1016/j.jtemb.2020.126556.

Su, W. H., & Sun, D. W. (2019). Mid-infrared (MIR) Spectroscopy for Quality Analysis of Liquid Foods. *Food Engineering Reviews, 11*, 142–158. https://doi.org/10.1007/s12393-019-09191-2.

Trifković, J., Andrić, F. A., Ristivojević, P., Guzelmeric, E., & Yesilada, E. (2017). Analytical Methods in Tracing Honey Authenticity. *Journal of AOAC International, 100*(4), 827–839. https://doi.org/10.5740/jaoacint.17-0142.

Ulberth, F. (2016). Advances in Testing for Adulteration in Honey, Advances in Food Authenticity Testing. In Woodhead Publishing Series in Food Science, Technology and Nutrition (pp. 729–753), DOI:10.1016/B978-0-08-100220-9.00026-6.

Von der Ohe, W., Persano Oddo, L., Piana, M. L., Morlot, M., & Martin, P. (2004). Harmonized methods of melissopalynology. *Apidologie, 35*, S18–S25. https://doi.org/10.1051/apido:2004050.

Xu, J. X., Liu, X., Wu, B., & Cao, Y. Z. (2020). A comprehensive analysis of ^{13}C isotope ratios data of authentic honey types produced in China using the EA-IRMS and LC-IRMS. *J Food Sci Technol, 57*(4), 1216–1232. https://doi.org/10.1007/s13197-019-04153-2.

Zábrodská, B., & Vorlová, L. (2015). Adulteration of honey and available methods for detection: A review. *Acta Veterinaria Brno, 83*(10), 85–102. https://doi.org/10.2754/avb201483S10S85.

Zhou, X., Taylor, M. P., Salouros, H., & Prasad, S. (2018). Authenticity and geographic origin of global honeys determined using carbon isotope ratios and trace elements. *Scientific Reports, 8*, 14639. https://doi.org/10.1038/s41598-018-32764-w.