Journal of Food Composition and Analysis 144 (2025) 107778

Contents lists available at ScienceDirect

Journal of Food Composition and Analysis

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

FTIR spectroscopy combined with machine learning for the classification of Mediterranean honey based on origin

Dafni Dimakopoulou-Papazoglou [a,*,1], Salud Serrano [b,2], Inmaculada Rodríguez [b,3], Nikolaos Ploskas [c,4], Konstantinos Koutsoumanis [a,5,†], Eugenios Katsanidis [a,6]

[a] Department of Food Science and Technology, School of Agriculture, Faculty of Agriculture, Forestry and Natural Environment, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece
[b] Department of Food Science and Technology, University of Córdoba, Córdoba 14071, Spain
[c] Department of Electrical and Computer Engineering, Faculty of Engineering, University of Western Macedonia, Greece

ARTICLE INFO

Keywords:
FTIR spectroscopy
Machine learning
Multivariate analysis
Melissopalynological analysis
Physicochemical characteristics

ABSTRACT

This study focused on the differentiation of Mediterranean honeys based on their geographical and botanical origin using FTIR-ATR spectroscopy combined with chemometrics. A total of 156 commercial honey samples, classified as thyme, pine, or polyfloral, were gathered from five Mediterranean countries, namely Greece, Malta, Spain, Tunisia, and Turkey. Melissopalynological and physicochemical analyses were performed to characterize the honey samples. The geographical and botanical origins were identified using Principal Component Analysis (PCA) in conjunction with Random Forest (RF) and Data-Driven Soft Independent Modeling of Class Analogies (DD-SIMCA). The analysis utilized the spectral range of 1800 – 750 cm^{-1}, preprocessed with the first derivative. Both one-class (DD-SIMCA) and multiclass (RF) classification techniques demonstrated high accuracy, exceeding 90 % in most cases. Specifically, the best results in terms of differentiation of geographical origin using all samples were achieved by DD-SIMCA, yielding over 95 % accuracy for all countries, with the exception of Tunisia with an accuracy of 87 %. These findings highlight the robust predictive potential of FTIR-ATR spectroscopy combined with chemometric methods for determining both the geographical and botanical origins of honey. This methodology provides a fast, non-destructive tool for verifying the origin of Mediterranean honey, contributing to improved food traceability and supporting the honey industry.

1. Introduction

Honey, as defined by Council Directive 2001/110/EC, is a natural sweetener produced by *Apis mellifera* bees. It is regarded as a functional food due to its rich nutritional profile and diverse biological activities, including antioxidant, antimicrobial, and antiulcer properties (Afrin et al., 2020). The unique organoleptic and nutritional characteristics of honey have driven a growing demand for its consumption, resulting in increased production and imports within the European Union (EU imported 341,598 tons of honey in 2022, an increase of 110,000 tons compared to 2012 (FAO, 2022). However, fraudulent practices, such as mislabeling honey's floral origin or adulterating it with external sugars, have emerged as critical issues of concern. Monofloral honey due to its distinctive organoleptic properties, typically commands a higher price than polyfloral honeys, making it a target for consumer deception. To address these issues, the European Union has recognized the need to

* Correspondence to: Aristotle University of Thessaloniki, Department of Food Science and Technology, School of Agriculture, Faculty of Agriculture, Forestry and Natural Environment, Thessaloniki 54124, Greece.
E-mail address: ddimakop@agro.auth.gr (D. Dimakopoulou-Papazoglou).
[1] 0000–0001-8983–0577
[2] 0000–0002-5637–9430
[3] 0000–0003-2417–3910
[4] 0000–0001-5876–9945
[5] 0000–0001-7918–0275
[6] 0000–0003-4728–2650
[†] Deceased

https://doi.org/10.1016/j.jfca.2025.107778
Received 14 January 2025; Received in revised form 27 April 2025; Accepted 13 May 2025
Available online 14 May 2025

implement stringent measures to combat mislabeling and adulteration, aiming to enhance food traceability and protect consumers.

The identification of honey origin has traditionally relied on melissopalynology analysis combined with physical and chemical analyses, including moisture content, sugar profile, electrical conductivity etc. (Council Directive 2001/110/EC). Although these techniques are effective, they are often time-intensive, complex, and require skilled personnel. As a result, alternative approaches have been explored to more efficiently determine the botanical and geographical origin of honey. Recently, several studies have pointed out the effectiveness of spectroscopic techniques that focus on acquiring unique fingerprints rather than quantifying specific compounds for honey identification. Such studies utilized UV–vis spectroscopy (Ansari et al., 2018; de Souza et al., 2021; Dimakopoulou-Papazoglou et al., 2024), NIR spectroscopy (Guelpa et al., 2017; Huang et al., 2020), NMR (Bertelli et al., 2010; Rachineni et al., 2022), Raman spectroscopy (Oroian et al., 2018; Wu et al., 2022), fluorescence spectroscopy (Q. Chen et al., 2014; Yan et al., 2022), and mass spectrometry (Dinca et al., 2015). Fourier Transform Infrared Spectroscopy (FTIR) has also been utilized to identify the origin and detect adulteration of honey (Cengiz and Durak, 2019; Gallardo--Velázquez et al., 2009; Se et al., 2018). In addition, several studies have employed FTIR-ATR spectroscopy to identify the botanical origin of honey (Ciulu et al., 2021; David et al., 2022; Devi et al., 2018; Gok et al., 2015; Kasprzyk et al., 2018; Orfanakis et al., 2021; Pauliuc et al., 2021; Svečnjak et al., 2017), whereas only a few have focused on determining its geographical origin (Formosa et al., 2020; Grabato et al., 2022; Guyon et al., 2021). Specifically, FTIR spectra in tandem with multivariate statistical analysis has been successfully applied to classify the botanical origin of honey from various countries, including Croatia (Svečnjak et al., 2015), Greece (Orfanakis et al., 2021; Tsagkaris et al., 2023), Italy (Ciulu et al., 2021), India (Devi et al., 2018), Poland (Kasprzyk et al., 2018), Romania (David et al., 2022; Guyon et al., 2021), Turkey (Gok et al., 2015), etc., but there are no studies that include samples from various countries at the same time; thus, the identification of the geographical origin of the honey samples are presented in this work.

The classification of honey based on its botanical and geographical origin has been effectively achieved through the integration of multivariate analysis and machine learning techniques with various analytical and innovative methodologies (Maione et al., 2019). According to the literature, spectral data has been successfully analyzed using principal component analysis (PCA) in combination with discriminant analysis algorithms, including linear discriminant analysis (LDA), partial least squares discriminant analysis (PLS-DA), and orthogonal partial least squares – discriminant analysis (OPLS-DA) (Ansari et al., 2018; L. Chen et al., 2012; David et al., 2022; Devi et al., 2018; Gan et al., 2016; Orfanakis et al., 2021). A widely-used machine learning algorithm for classification is Random Forest (RF), which employs an ensemble of decision trees trained on randomly selected subsets of features and data samples. This ensemble approach aggregates the outputs of all the individual trees, thus allowing the model to capture a wide range of patterns in the data (Breiman, 2001). RF is a highly promising classification algorithm; however, there is only a limited number of studies exploring its application for honey identification (Batista et al., 2012; Ciulu et al., 2021; Dimakopoulou-Papazoglou et al., 2023, 2024; Martinez-Castillo et al., 2020). Furthermore, Data-Driven Soft Independent Modeling of Class Analogies (DD-SIMCA) is a one-class classification method that can be effectively applied to honey discrimination. This supervised classification technique classifies new samples as belonging to the target class based on their degree of similarity to predefined decision criteria. By exclusively leveraging information from the target class to optimize the model, DD-SIMCA is particularly well-suited for determining whether or not a sample belongs to a specific class. Despite the good performance of DD-SIMCA, only a limited number of studies have reported its application in the literature (Ansari et al., 2018; de Souza et al., 2021; Dimakopoulou-Papazoglou et al., 2023, 2024; Roshan et al., 2013; Suhandy & Yulia, 2021).

To the best of our knowledge, no prior research has focused on differentiating the origin of honey in the Mediterranean region, considering both its geographical and botanical aspects. Therefore, the present study aimed to utilize state-of-the-art machine learning techniques, which allow for the simultaneous differentiation of both geographical and botanical origin of Mediterranean honey, in order to differentiate honey samples from five different Mediterranean countries according to their origin using melissopalynology, physicochemical analyses, and FTIR spectroscopy combined with multivariate statistical analysis. Furthermore, it was the first attempt to evaluate the efficacy of both one-class (DD-SIMCA) and multiclass (RF) classification techniques for determining honey origin.

2. Materials and methods

2.1. Honey samples

A total of 156 commercial honey samples were gathered either directly from beekeepers or from the market across different geographical regions of Mediterranean countries. In particular, honey samples were collected from Greece (46 samples: 26 thyme, 10 pine, and 10 polyfloral), Tunisia (42 samples: 31 thyme and 11 polyfloral), Turkey (31 samples: 3 thyme, 10 pine, and 18 polyfloral), Spain (27 samples: 17 thyme and 10 polyfloral), and Malta (10 samples: all polyfloral). All samples were harvested during the 2021–2022 season. The samples were placed in clean plastic containers and kept refrigerated until analysis.

2.2. Melissopalynological analysis

The melissopalynological analysis was conducted following the methods described by Louveaux et al. (1978) and Von Der Ohe et al., 2004. Specifically, 10 g of each honey sample was centrifuged for 10 min at 2500 rpm, the supernatant liquid was decanted, and the sediment was dried at 40ºC before being mounted with Entellan Rapid (Merck KGaA, Madrid, Spain). At least 1200 pollen grains were counted and identified from two slides of each honey sample without any chemical treatment. The pollen slides were examined at 400x and 1000x magnification to identify the pollen types using a light microscope (Nikon Labophot-2 microscope; Nikon, Tokyo, Japan).

2.3. Physicochemical analysis

Most physicochemical analyses (except for water activity) were performed according to the Harmonized Methods of the International Honey Commission (IHC Methods, 2009).

Water content (humidity) and total sugars were determined following Chataway (1932) and Wedmore (1955). An Abbe-type refractometer (Zuzi, mod. 325, Navarra, Spain) was used, obtaining the corresponding percentage of water from the Chataway table.

Hydroxymethilfurfural (HMF) determination was made according to White method (White, Jr, 1979). A Pharmacia Biotech Ultrospec-3000 (Uppsala, Sweden) spectrophotometer was used. Results are expressed in HMF milligrams per kg of honey.

Electrical conductivity was measured at 20 °C in a 20 % (w/v) solution of honey (dry matter basis) in deionized water using a Crison model 524 conductimeter (Crison Instruments, Barcelona, Spain), according to Vorwohl (1964). The results were expressed in microSiemens per cm (μS/cm).

Measurement of water activity was made by means of a Novasina IC.500 AW-LAB apparatus (Lachen, Switzerland).

To obtain the sugar profile, fructose, glucose, sucrose, and maltose contents were determined by HPLC (High Pressure Liquid Chromatography) with RI-detection (Varian ProStar, Prostar 350 refraction index detector) using an analytical column containing amine modified silica

gel (Agilent, Microsorb 100–3 NH2, SS 150 ×4.6 mm) (IHC Methods, 2009). The mobile phase (isocratic) was acetonitrile:water (80:20, v/v), while the flow was set at 1.3 ml/min.

2.4. ATR-FTIR spectroscopy

Prior to spectral analysis, honey samples were liquified at 40 °C for 1 hour in a water bath to ensure the dissolution of any crystals and remove the bubbles. The spectra of honeys were acquired in the mid-infrared region of 4000 – 500 cm^{-1} using a Fourier-transform infrared spectrometer (FTIR 6700 series, JASCO, Tokyo, Japan) fitted with 3-reflection ATR diamond (MIRacle ATR, Pike Technologies, Madison, Wisconsin, U.S.). A drop of the honey samples was applied to the ATR surface and spectra were recorded at room temperature with a resolution of 2 cm^{-1} and 32 scans using Spectra Manager (V.2, Jasco, Tokyo, Japan). Prior to every measurement, a background air spectrum was recorded and then subtracted from the spectra of individual samples. The ATR crystal was subjected to a cleaning process using isopropanol and drying prior to analyzing the next sample. Five spectra were collected per sample, and the resultant average spectrum was used for the interpretation of honey's chemical composition and the construction of the chemometric models.

2.5. Spectra pre-processing and chemometric analysis

Before building classification models, various pre-processing techniques were evaluated for their effectiveness in removing systematic baseline variations from the FTIR spectra. These techniques included standard normal variate (SNV), Savitzky–Golay smoothing, and derivatives (first and second derivatives calculated with a second order polynomial and 11-point window size). Additionally, mean centering was performed during the preprocessing stage. The most relevant wavelengths were identified by selecting those that yielded the highest accuracy during cross-validation iterations.

For the classification models, two approaches were used: a multiclass classification method using Random Forest (RF) algorithm, and a one-class classification technique using Data Driven - Soft Independent Modelling of Class Analogies (DD-SIMCA). A cross-validation approach was implemented across all methods, where 80 % of the honey samples were randomly designated as the training set, and the remaining 20 % were reserved for the test set to perform internal validation. A 10-fold cross-validation strategy was utilized. For model construction, RF utilized the training dataset comprising 80 % of honey samples from multiple classes, with classes defined either by honey type or country of origin for botanical or geographical classification, respectively. In contrast, Data-Driven Soft Independent Modeling of Class Analogies (DD-SIMCA) focused on building models for single-class prediction, targeting a specific geographical or botanical origin, using 80 % of the honey samples from a particular class as the training set. All models were tested at a 5 % significance level (α=0.05). In the case of RF, model performance metrics were accuracy, precision, recall, and F1-score, while for DD-SIMCA, accuracy, sensitivity, and specificity were used, as described in Dimakopoulou-Papazoglou et al. (2024).

Chemometrics analyses, regarding Principal Component Analysis (PCA) and RF were conducted in Python using the scikit-learn module, while DD-SIMCA models were implemented using a MATLAB code available at https://github.com/yzontov/dd-simca.

2.6. Data analysis

Descriptive statistical analysis to determine the average value of the analyzed data was performed. The Data Normality Test and the Kolmogorov–Smirnov Test of Normality were used to assess the normal distribution of the data. In addition, Kruskal–Wallis one-way variance analysis was conducted to test the hypotheses and Pearson Product Moment Correlation Coefficient analysis was used to evaluate the relationships between variables. The analysis of the data in the present study was conducted using IBM SPSS Statistics (version 26.0) predictive analytics software.

3. Results and discussion

3.1. Melissopalynology analysis

Pollen analysis provides a detailed picture of the botanical profile of honey samples, which is essential for classifying their botanical origin as they must meet specific thresholds according to EU legislation. As specified in (Greek Directive 127/2004) Greek Directive 127/2004, thyme honey must have a minimum of 18 % thyme pollen grains to be recognized as monofloral thyme honey. Thyme Greek honey samples showed a mean value of 48 % thyme pollen grains (*Thymus* spp.), followed by Tunisian honey samples with 27 %, Spanish with 24 %, and Turkish with 19 %. The highest concentration of thyme pollen in Greek thyme honeys aligns with findings by Karabagias et al. (2017), who studied thyme honeys originated from Greece, Spain, Marocco, and Egypt. Additionally, the pollen composition of Greek samples was widely represented by Lamiaceae, Asteraceae, Ericaceae, Oleaceae, Cistaceae (*Cistus* spp.), and Rosaceae pollen types, while Apiaceae, Fabaceae, and Myrtaceae were found in over 50 % of the samples. Pollen from *Thymus capitatus*, *Cistus* spp., *Olea europaea* L., and *Erica* spp. was detected in all honey samples. The occurrence of pollen from non-nectar-producing plants, such as *Olea europaea* and *Cistus* spp., in Greek thyme honey can be attributed to the abundant pollen presence and dispersal. *Olea europaea* and *Cistus* spp. pollen was also found in Spanish thyme honeys but less often than in Greek samples. The widespread cultivation of these plants in the Mediterranean explains their presence in thyme honey from Greece (Thrasyvoulou and Manikis, 1995; Tsigouri et al., 2004), Spain (Terrab et al., 2004), and Italy (Di Marco et al., 2017). Mediterranean plant families such as Cistaceae, Myrtaceae, and Apiaceae were identified in all countries, rendering these pollen types potential markers for Mediterranean thyme honeys. Differences in pollen were observed within Mediterranean thyme honey samples, as depicted in the microscope images in Fig. 1. In this sense, Greek samples were characterized by the presence of Ericaceae pollen, which was absent in samples from the other countries. Only Spanish thyme honeys included *Echium plantagineum* pollen, while Acacia pollen was present only in Tunisian thyme honey. Finally, pollen in Turkish samples was characterized by a combination of Asteraceae family, *Taraxacum* spp. and *Carduus* spp. with presence over 20 % and 10 %, respectively.

The majority (>90 %) of the world's pine honeydew honey is produced in Greece and Turkey. The insect *Marchalina hellenica*, which inhabits *Pinus brutia*, is only found in these countries (Duru et al., 2021; Karabagias et al., 2014). Honeydew honey is distinguished by the presence of honeydew elements, including microscopic algae and fungal spores (Özkök et al., 2018). Microscopic examination of all the samples showed the presence of fungal spores, with the most frequently observed genera being *Alternaria*, *Coleosporium,* and *Cladosporium*, among others, consistent with findings by Dimou et al. (2006) for Greek pine honeydew honeys. Furthermore, in the present study, Turkish samples showed a lower honeydew element-to-pollen grain (HDE/P) ratio compared to Greek honeys. Pine honeydew honey produced in Turkey and Greece could potentially be distinguished by their HDE/P ratio and pollen composition. The HDE/P ratio for Greek samples averaged 10.6, while for Turkish samples 7.4. Both ratios are classified as 'high density superior quality pine honey' (Louveaux et al., 1978; Sorkun, 2008). Differences in pollen composition were also noted between the two countries. Greek samples contained *Ericaceae* pollen, similar to thyme honeys, as well as *Castanea sativa*, which were absent in Turkish pine honeydew honeys. In contrast, Turkish samples were characterized by pollen from the Asteraceae family (*Taraxacum* spp. and *Carduus* spp.) with concentrations exceeding 10 %.

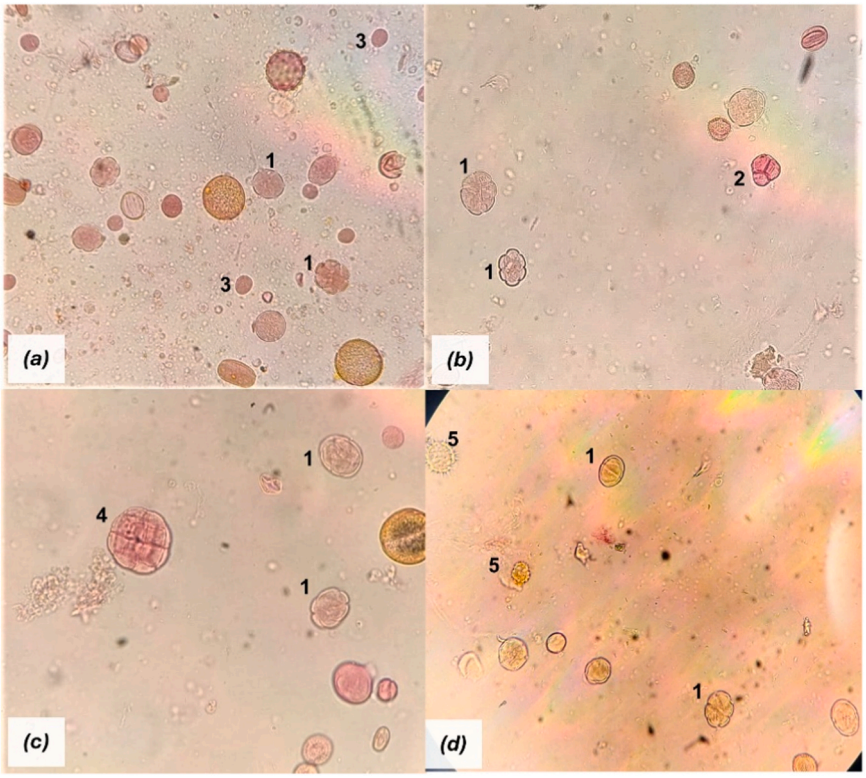

Fig. 1. Optical microscope images for *Thymus* spp. monofloral honeys from: a) Spain, b) Greece, c) Tunisia, and d) Turkey ([1]*Thymus* spp.; [2]Ericaceae family; [3]*Echium plantagineum;* [4]*Acacia* spp.; [5]Asteraceae family).

For the polyfloral samples collected from five Mediterranean countries, i.e., Spain, Greece, Malta, Tunisia, and Turkey, common and distinct pollen types were observed, suggesting potential geographical markers within the Mediterranean region. Representative microscope images of polyfloral honeys from these countries are shown in Fig. 2. As with the monofloral honeys described above, Myrtaceae and Cistaceae families were found in all the samples. Leguminosae or Fabaceae (*Plantago* spp. and *Trifolium* spp.) were also observed in most of the samples together with Asteraceae (*Helianthus* annuus, *Taraxacum* spp. and *Carduus* spp.) but showing differences in the frequency and percentage in the pollen composition. *Echium plantagineum* was only found in Spanish samples, suggesting it as a potential botanical and geographical marker for Spanish honeys. Similarly, Greek and Tunisian samples showed values exceeding 15 % for *Castanea sativa*, which could serve as a marker for honeys from these countries. Turkish samples had a high diversity, with an average of at least seventeen different pollen types. Polyfloral honeys did not display a predominant pollen type, though the Myrtaceae family commonly represented the highest percentage (>20 %) across countries, except in Turkey, where it was lower than in other countries.

The pollen analysis provides a detailed picture of the botanical composition of honey samples. A principal component analysis (PCA) was performed using all samples to understand similarities and differences between the honey samples (Fig. 3). The PCA shows a clear separation of the varietal-country groups. Thyme pollen type (*Thymus* spp.) is responsible for the separation of groups along PC1, with highest amount of thyme pollen indicated in Greek honey samples. Polyfloral honeys from Turkey seem to have the lowest amounts of *Thymus* spp. and a special composition compared to polyfloral honeys from other countries. Pine pollen type (*Pinus* spp.) influences the separation along PC2 with higher values for Greek than for Turkish pine samples.

3.2. Physicochemical analysis

Water content (humidity), total sugars, and water activity are indicators of proper maturation of honey and are generally not related to botanical or geographical origin. For this project, these determinations have been carried out to evaluate the quality of the honey samples. In fact, humidity is the most important determination at the time of commercialization in view of possible fermentation of honey. According to European regulations EC/2001/110, the maximum allowable humidity for honey is 20 %. Fig. 4 shows the average values of humidity, total sugar contents and water activity per country. The average humidity levels for all countries were below 18 %, preventing unwanted fermentation, except for Maltese honeys, which may be influenced by local climatic conditions.

The electrical conductivity is a parameter linked to the botanical origin, and it differs differencing between nectar honeys and honeydew honeys. However, there are no established conductivity ranges for different monofloral honeys. All thyme honey samples in this study exhibited electrical conductivities below 800 μS/cm, which align with the threshold set for nectar-origin honeys (European regulation: (EU/2014/63, 2014)). Fig. 5 presents the box plot obtained by applying the Kruskal-Wallis test for independent samples to the electrical conductivity data of thyme honeys. The median conductivity values were similar across the four countries, with Tunisia, Spain, Greece, and Turkey showing values of 348, 397, 381, and 356 μS/cm, respectively. These medians align with the average value of 400 μS/cm reported by the International Honey Commission for European thyme honey (Oddo et al., 2004). Regarding the value ranges, Spain is the one with the greatest dispersion of values (294–673 μS/cm), with Tunisia being the geographical origin that shows the narrowest range (204–527 μS/cm). The data distribution for Spain was positively skewed, whereas Greece showed a negatively skewed distribution (range: 165–575 μS/cm). In Turkey (range: 241–513 μS/cm) and Tunisia, distributions were observed to be positively and negatively skewed, respectively.

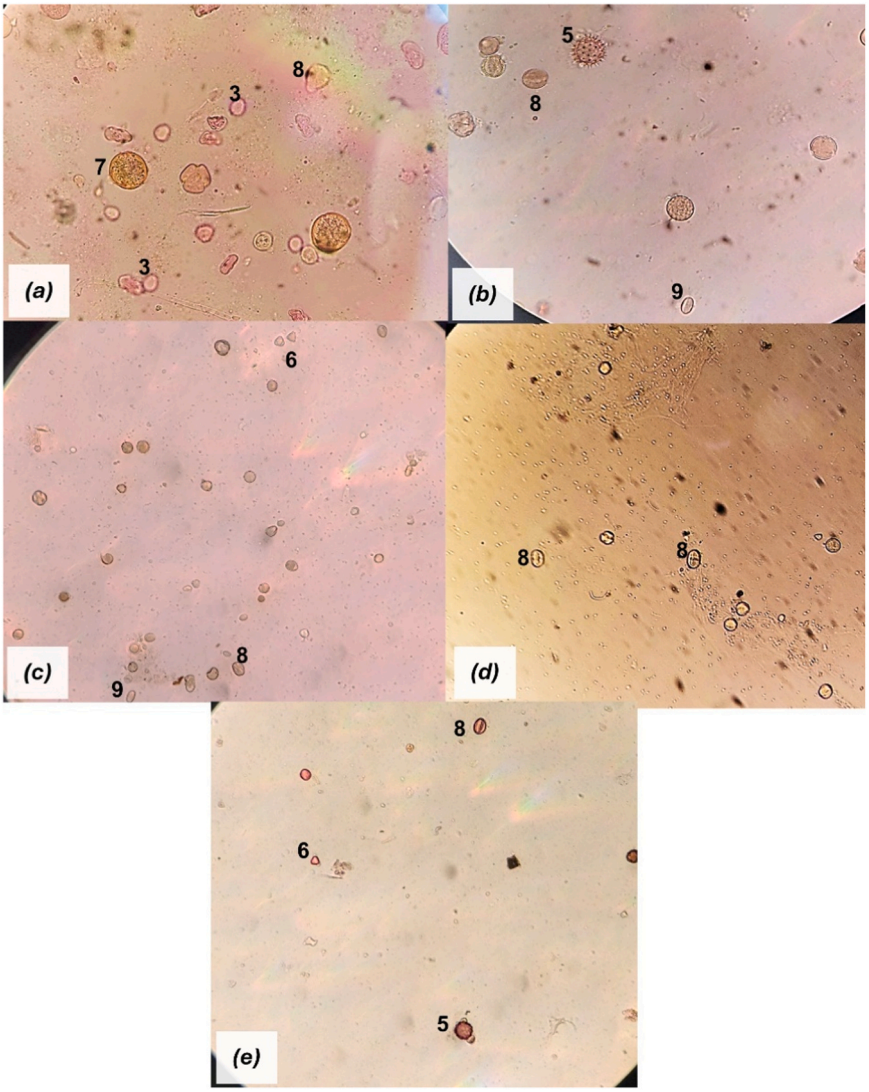

Fig. 2. Optical microscope images for polyfloral honeys from: a) Spain, b) Greece, c) Tunisia, d) Turkey, and e) Malta ([6]Myrtaceae family; [7]Cistaceae family; [8]Fabaceae family; [9]*Castanea sativa*; [3]*Echium plantagineum*; [5]Asteraceae family).

Most samples from the five countries were under the criterion of 40 mg/kg for hydroxymethylfurfural (HMF) that is established by the European normative on quality of honey. However, five samples from Tunisia (ThTN2, ThTN8, ThTN9, ThTN10, and PfTN10) exceeded this limit, likely due to the country's climatic conditions. As shown in Fig. 4, the average HMF content for the Tunisian samples was higher than the samples from the other countries, which had similar average values.

In almost all types of honeys, the main monosaccharide present in honeys is fructose followed by glucose. The concentrations of these carbohydrates, along with their ratio, have been used as indicator for honey authentication, and are particularly important for the food industry to ensure the quality of honey. The mean concentrations of sugars (fructose, glucose, sucrose, and maltose) in the analyzed honey samples are shown in Table 1. For the monofloral thyme honey, the highest fructose concentration was presented in Turkish samples, followed by Greek, Tunisian, and Spanish. For polyfloral and monofloral pine honeys, no significant differences in sugar composition were observed.

3.3. FTIR spectra

Fig. 6 shows the original ATR-FTIR spectra of representative honey samples, namely thyme, pine, and polyfloral, in the spectral region of 4000 – 500 cm^{-1}. Spectral analysis enabled the identification of characteristic absorption bands specific to honey, providing insights into its unique features which can be used to determine its origin.

The broad band from 3500 – 3000 cm^{-1}, peaking at 3293 cm^{-1}, is attributed to the O–H stretching vibration due to the water content (Anjos et al., 2015; Gok et al., 2015; Svečnjak et al., 2017). The 3000 – 2800 cm^{-1} range, and specifically at 2932 cm^{-1}, is linked to C–H stretching vibration due to the presence of carbohydrates (Gallardo-Velázquez et al., 2009), along with C–H stretching in carboxylic acids and NH_3 stretching of free amino acids, which are at low concentrations in honey (Anjos et al., 2015; Gok et al., 2015; Tewari and Irudayaraj, 2004).

The 1700 – 1600 cm^{-1} region, with a peak at 1643 cm^{-1} , is associated mainly with O–H stretching and bending vibrations of water, C═O stretching of carbohydrates and N–H bending of amide I of proteins, which are present at small amounts in honey (Ciulu et al., 2021; Gallardo-Velázquez et al., 2009; Kędzierska-Matysek et al., 2023; Tahir et al., 2017). The spectra region of 1550 – 1175 cm^{-1} is attributed to C–H, C–O, and C═C deformations, mainly from carbohydrates, phenolics, and flavonols (Gok et al., 2015; Tahir et al., 2017). In particular, the peaks at 1414 cm^{-1} and 1344 cm^{-1} are related to O–H bending of the C–OH group and O–CH and C–CH deformation vibrations in carbohydrate structures (Gallardo-Velázquez et al., 2009; Kędzierska-Matysek et al., 2023; Svečnjak et al., 2017). The peak at 1255 cm^{-1} is due to C–C

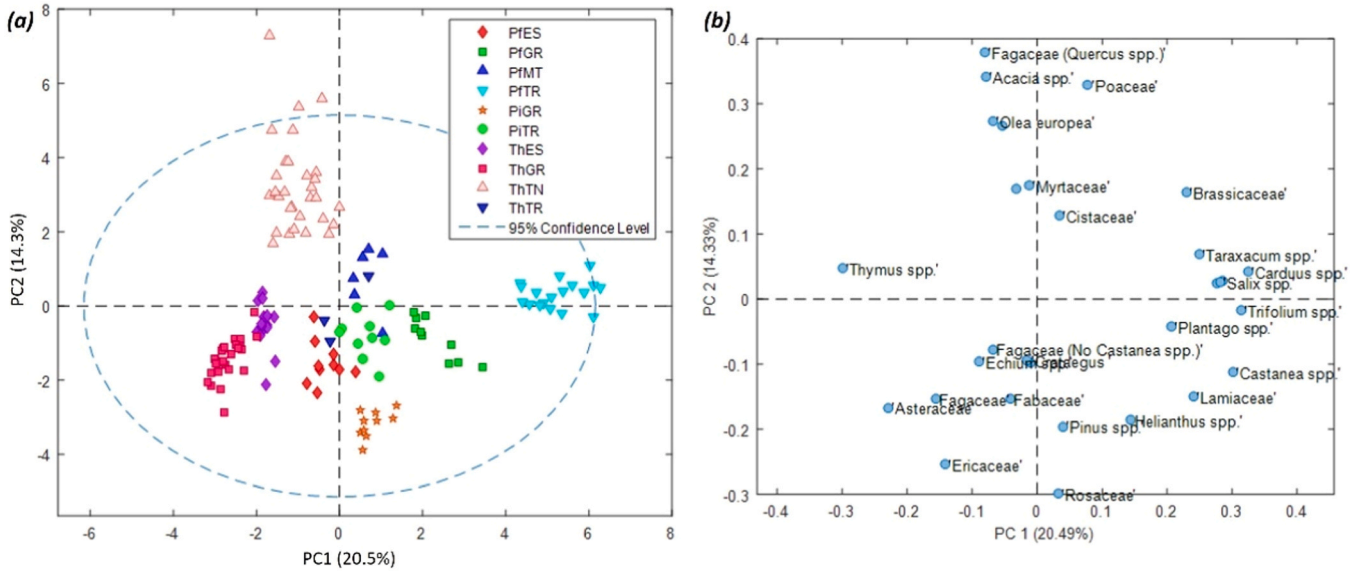

Fig. 3. Principal component analysis of pollen types in honey samples with scores plot (a) and loadings plot (b). Scores are colored according to the varietal-country combination (Pf: polyfloral, Pi: pine, Th: thyme, ES: Spain, GR: Greece, MT: Malta, TR: Turkey, TN: Tunisia).

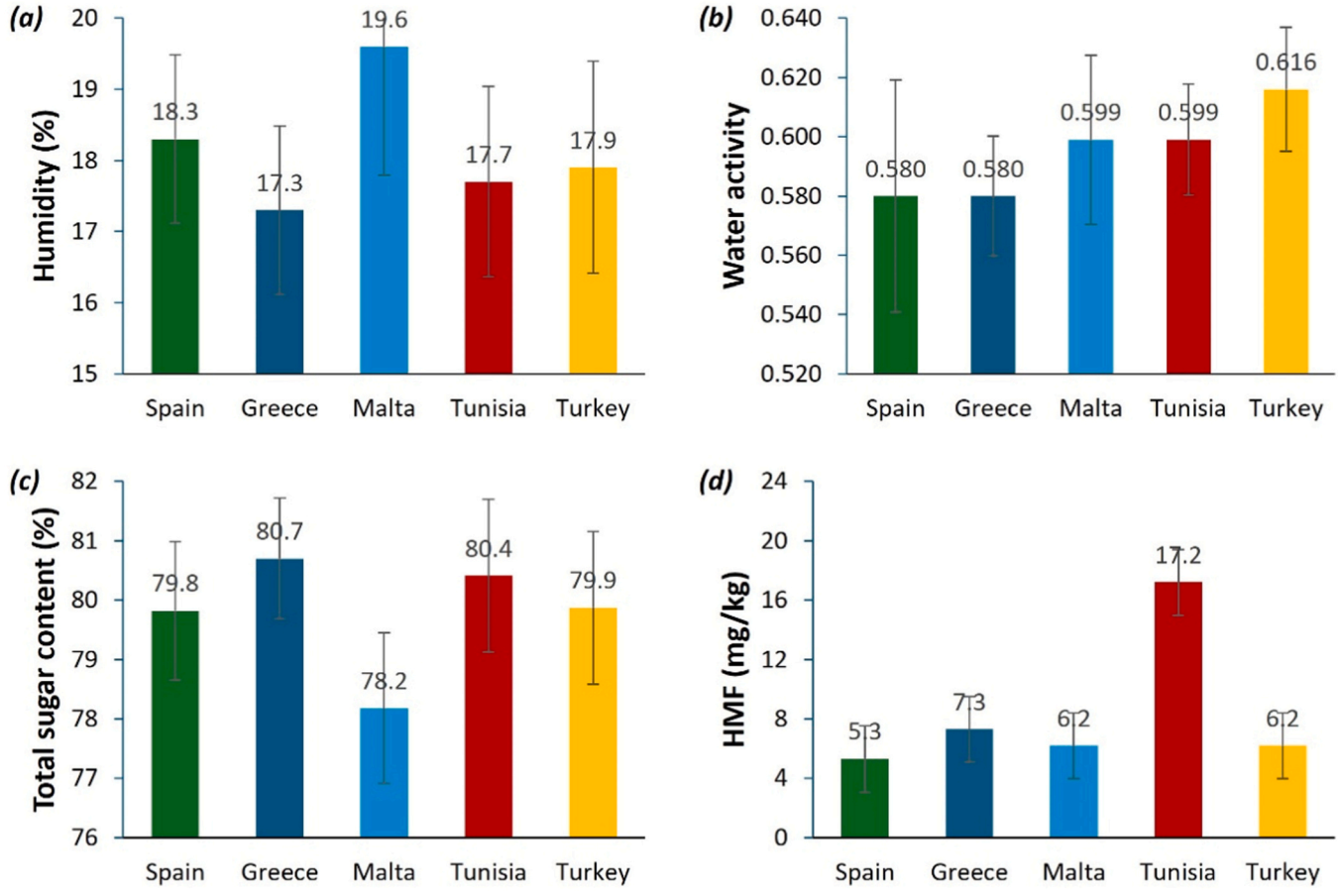

Fig. 4. Average values of humidity (*a*), total sugar content (*b*), water activity (*c*), and HMF (*d*) for honey samples per country.

stretching in the carbohydrate structure (glucose and fructose) and the deformation of CH_2 (Svečnjak et al., 2017).

The spectral region of 1175 – 950 cm^{-1} represents the C–O and C–C skeletal stretching vibrations of carbohydrates (Svečnjak et al., 2017). Specifically, the band at 1146 cm^{-1} is attributed to stretching C–H vibration or stretching C–O vibration in carbohydrate structures, while at 1100 cm^{-1} relates to stretching of the C–O band of the C–O–C linkage. According to the literature, the peak at 1148 cm^{-1} is characteristic of sucrose, the peaks at 1087 cm^{-1} and 1043 cm^{-1} indicate the presence of glucose and fructose, while the peaks at 983 cm^{-1} and 965 cm^{-1} are

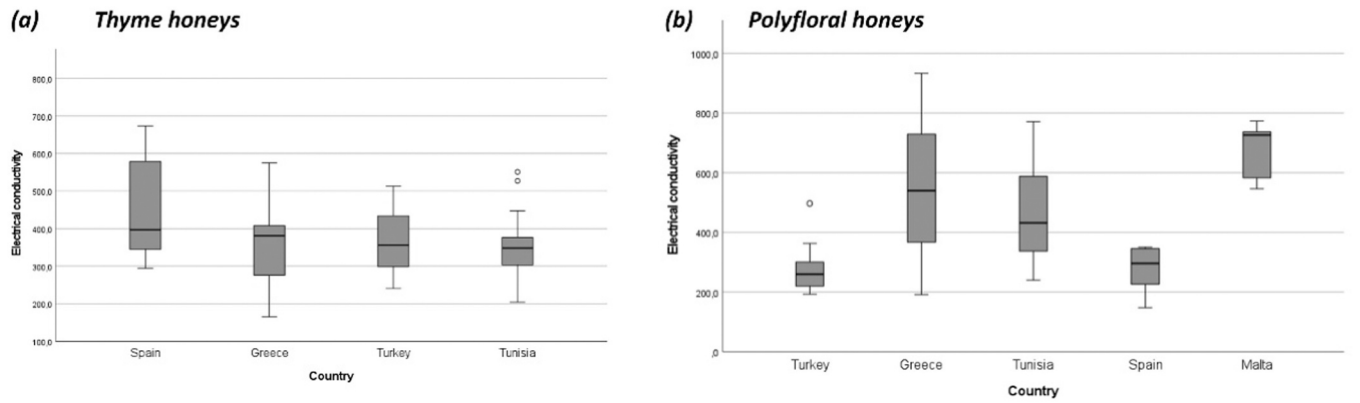

Fig. 5. Distribution of electrical conductivity (μS/cm) of thyme (*a*) and polyfloral (*b*) honey samples by country.

Table 1
Sugar composition of the honey types according to the varietal-country combination (Th: thyme, Pi: pine, Pf: polyfloral, ES: Spain, GR: Greece, MT: Malta, TR: Turkey, TN: Tunisia).

Samples	Fructose (%)	Glucose (%)	Sucrose (%)	Maltose (%)	F/G	F+G
ThGR	38.3 ± 3.9	28.2 ± 2.6	0.5 ± 0.7	1.6 ± 1.7	1.3 ± 0.3	65.0 ± 8.1
ThES	36.1 ± 2.1	25.7 ± 2.5	0.1 ± 0.2	3.7 ± 3.0	1.4 ± 0.1	61.8 ± 3.9
ThTN	36.3 ± 3.0	29.2 ± 2.3	2.8 ± 2.0	5.1 ± 1.9	1.2 ± 0.1	65.5 ± 5.0
ThTR	41.1 ± 0.8	26.0 ± 0.3	0.0 ± 0.0	5.0 ± 0.2	1.6 ± 0.0	67.0 ± 1.0
PiGR	34.6 ± 5.0	26.3 ± 4.6	0.1 ± 0.3	0.0 ± 0.0	1.3 ± 0.1	60.8 ± 9.4
PiTR	32.4 ± 1.4	24.3 ± 1.0	0.0 ± 0.0	3.0 ± 4.1	1.3 ± 0.0	56.7 ± 2.2
PfGR	35.7 ± 5.2	28.9 ± 1.9	0.1 ± 0.2	1.1 ± 0.2	1.3 ± 0.0	64.6 ± 5.9
PfES	35.7 ± 4.8	29.1 ± 2.7	0.2 ± 0.1	1.3 ± 2.8	1.2 ± 0.2	64.7 ± 4.2
PfTN	35.6 ± 3.1	29.3 ± 3.2	2.6 ± 2.1	4.7 ± 2.0	1.1 ± 0.1	64.9 ± 5.4
PfTR	38.7 ± 1.3	30.7 ± 1.7	0.0 ± 0.0	0.3 ± 1.2	1.3 ± 0.1	69.5 ± 2.3
PfMT	35.9 ± 1.8	32.4 ± 2.2	2.4 ± 1.4	3.5 ± 0.9	1.1 ± 0.1	68.4 ± 2.5

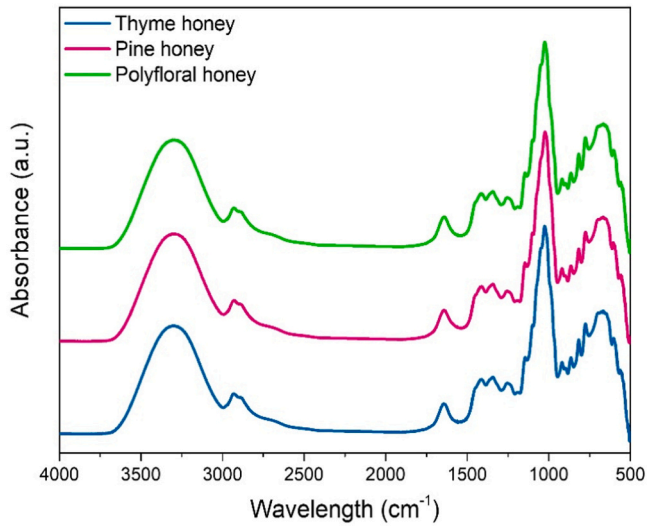

Fig. 6. Typical FTIR spectra of thyme, pine, and polyfloral honey samples.

associated with fructose (Anjos et al., 2015; Pauliuc et al., 2021; Vîjan et al., 2023). Moreover, the peak at 990 cm^{-1} reflects the stretching of C–O bonds associated with glycosidic linkages (Wang et al., 2010).

The 950 – 750 cm^{-1} region, corresponds to the anomeric region of carbohydrates, which is distinctive for each sugar (Gok et al., 2015). Within the 950 – 850 cm^{-1}, the absorption bands are attributed to the skeletal C–C vibrations of the carbon backbone in monosaccharides. The band at 918 cm^{-1} is due to the C–H bending in carbohydrates, while at 865 cm^{-1} are assigned to skeletal C–C stretching vibrations of fructose (Svečnjak et al., 2017). Additionally, bands at 817 and 775 cm^{-1} are linked to C–C–H deformation in carbohydrates, specifically of fructose.

The spectral region of 1800 – 750 cm^{-1} provides the most detailed and valuable information regarding the chemical composition of honey samples. In this range, small variations in the intensity of bands associated with different sugar fractions, as well as phenolic content were observed. Therefore, this specific spectral range, which serves as a chemical fingerprint of honey, was selected for determining its geographical and botanical origin using multivariate statistical analysis. This approach has been successfully demonstrated in previous studies by Orfanakis et al. (2021) and Svečnjak et al. (2017).

3.4. Classification methods – chemometrics

Subtle variations in the chemical composition of honey samples, as revealed by FTIR spectra, can be effectively analyzed and leveraged to determine their botanical and geographical origins using multivariate statistical analysis. To assess the effect of different preprocessing methods (first- and second-derivative transformation with and without SNV) on various spectral ranges (4000 – 750 cm^{-1} and 1800 – 750 cm^{-1}), we evaluated the total variance explained by the first five principal components (PC) (Table 2). The results show that using the first derivative on the 1800–750 cm^{-1} spectral range gives the best performance, explaining nearly 96 % of the total variance. Based on these findings, the first-derivative transformation was used in all the computational results that follow.

After removing systematic bias using preprocessing, PCA was initially applied to explore and visualize similarities and differences among the honey samples. PCA serves as a dimensionality reduction technique that calculates PCs, which account for the variability within the original variables. For classification purposes, the optimal number of PCs were then used to develop classification models, specifically using Random Forest (RF) and DD-SIMCA.

The score and loading plots of the first two PCs from the PCA analysis for distinguishing the geographical origin of thyme, pine, and polyfloral honey samples are depicted in Fig. 7. For thyme honey samples, clear differentiation was achieved (Fig. 7.1.a), with five PCs explaining 95.7 % of the variance. A strong differentiation was noted for pine honey samples also, since their originating from only two geographic

Table 2
Principal Component Analysis for thyme honeys using various preprocessing techniques: R^2 of first three components and total variance explained by first five components.

Spectral range	Preprocessing technique	PC1	PC2	PC3	Cumulative (PC1-PC5)
4000–750 cm^{-1}	1st derivative	68.20 %	10.58 %	8.66 %	92.77 %
4000–750 cm^{-1}	SNV, 1st derivative	69.42 %	10.61 %	7.57 %	92.49 %
4000–750 cm^{-1}	2nd derivative	26.52 %	12.49 %	3.19 %	47.26 %
4000–750 cm^{-1}	SNV, 2nd derivative	26.65 %	12.46 %	3.12 %	47.28 %
1800–750 cm^{-1}	1st derivative	71.49 %	11.03 %	8.45 %	95.90 %
1800–750 cm^{-1}	SNV, 1st derivative	74.38 %	10.95 %	6.11 %	95.61 %
1800–750 cm^{-1}	2nd derivative	32.53 %	18.33 %	4.34 %	61.36 %
1800–750 cm^{-1}	SNV, 2nd derivative	33.66 %	17.22 %	4.16 %	61.42 %

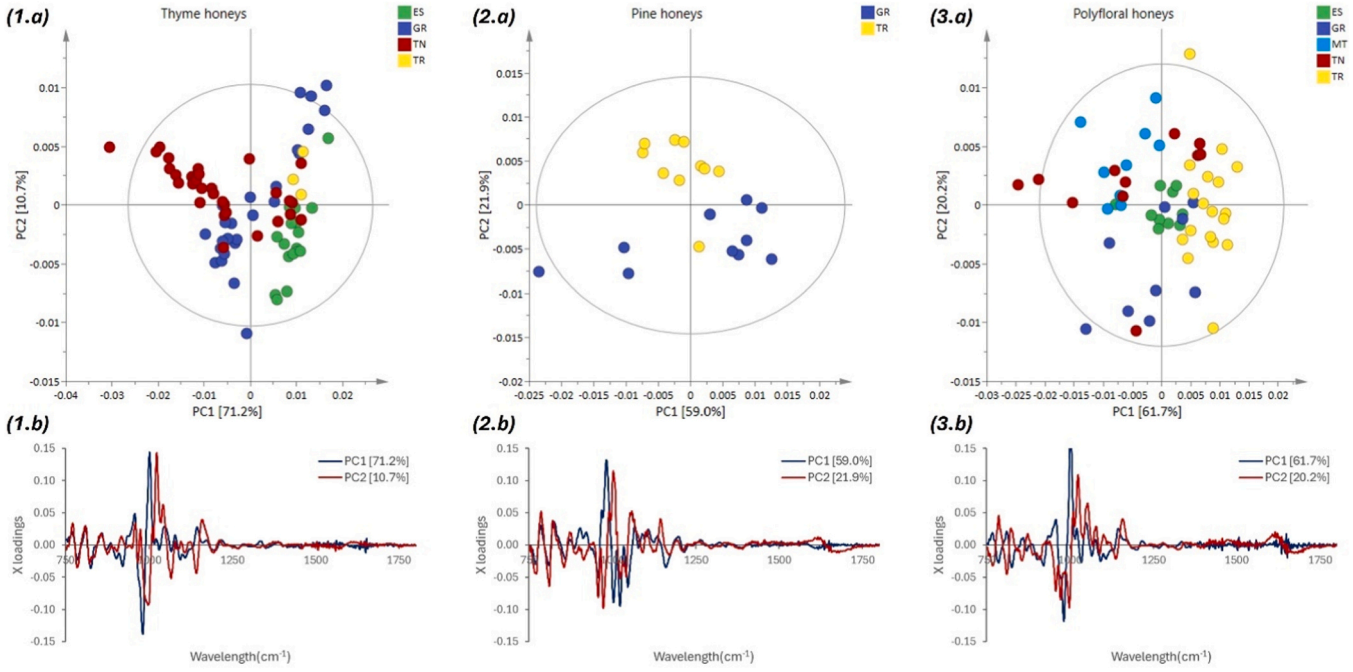

Fig. 7. Score plots (*a*) and loading plots (*b*) of the first two components for the discrimination of geographical origin of thyme (*1*), pine (*2*), and polyfloral (*3*) honey samples from PCA of FTIR spectral data pre-processed with 1st derivative in the range of 1800–750 cm^{-1} (ES: Spanish, GR; Greek, MT: Maltese, TN: Tunisian and TR: Turkish).

locations, Greece, and Turkey (Fig. 7.2.a), making it easier to obtain good results. Furthermore, a clear geographical distinction was achieved among polyfloral honey samples from the five Mediterranean countries, with the first five PCs explaining 94.7 % of the variance (Fig. 7.3.a). As shown in the loading plots (Figs. 7.1.b, 7.2.b, and 7.3.b), the spectral region that most significantly affects the analysis is 1200 – 750 cm^{-1}, related to skeletal stretching vibrations and the anomeric region of carbohydrates, while the 1600 – 1700 cm^{-1} region also has a secondary but notable impact.

In terms of differentiating the botanical origins of Greek honeys, thyme honey is clearly distinguishable from pine honey, as shown in Fig. 8. Polyfloral honeys, although also differentiated, exhibit some similarities with thyme honey. As expected, and confirmed by pollen analysis, polyfloral honeys contain a diverse array of pollen types, displaying some common pollen types as those found in thyme honey. (Fig. 9)

3.4.1. *Determination of geographical origin*

To identify the geographical origin of honey samples, both one-class classification (DD-SIMCA) and multiclass classification (RF) methods were utilized. When the Random Forest (RF) algorithm was applied to classify all honey samples based on their geographical origin, the model achieved an accuracy of approximately 68 %, with precision, recall, and F1-score values of 62 %, 62 %, and 61 %, respectively. These metrics were derived from the confusion matrix, comparing predicted versus actual values on the validation set. Accuracy represents the overall proportion of correct predictions made by the model, while precision indicates the proportion of true positives among all predicted positives, making it particularly valuable when minimizing false positives is critical. Recall, also known as sensitivity, assesses the model's effectiveness in identifying all actual positive cases, calculated as the ratio of correctly predicted positive observations to the total number of actual positives. The F1-score combines precision and recall into a single metric, serving as their harmonic mean to provide a balanced measure of the model's performance (Dimakopoulou-Papazoglou et al., 2024).

As the metrics of the general model were relatively low, separate RF models were developed for individual honey types, leading to improved results (Table 3). Specifically, the model developed to predict the geographical origin of thyme honeys (from Greece, Spain, and Tunisia) achieved an accuracy of 83.6 %. The corresponding model for pine honeys originating from Greece and Turkey demonstrated even higher performance, attaining an accuracy of 92.5 %. The high values of model's performance criteria for pine honeys (Table 3) is attributed on one hand to the fewer geographical origins and on the other hand to distinct honeydew elements (e.g., fungal spores, hyphae, algae, and wax particles) unique to each region (Dimou et al., 2006; Duru et al., 2021; Karabagias et al., 2014), which were reflected in the FTIR spectra. Regarding the model developed from thyme honeys, higher accuracy

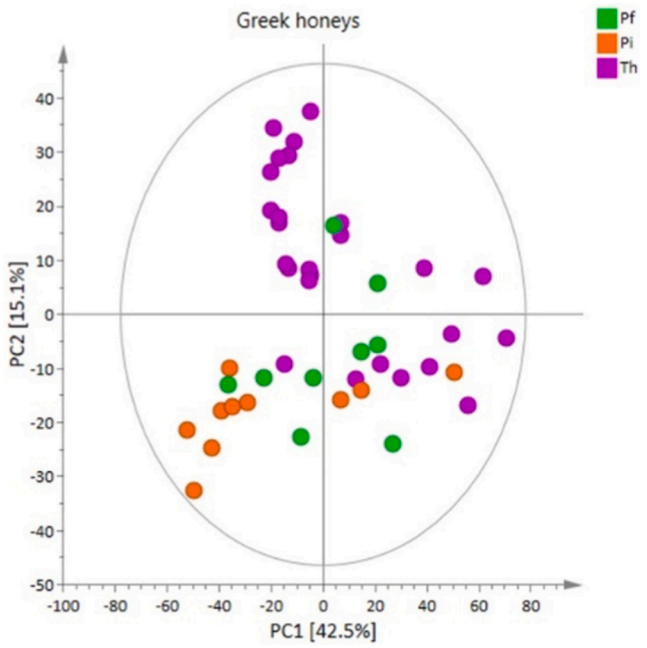

Fig. 8. PCA analysis of FTIR spectral data pre-processed with 1st derivative in the range of 1800–750 cm^{-1} for the differentiation of botanical origin of Greek honeys (Pf: polyfloral, Pi: Pine, Th: Thyme honeys).

and improved metrics compared to the general model were recorded due to the better representation of characteristic compounds in the FTIR spectra, including carbohydrate, phenolic compounds, etc., shaped by local flora. Specifically, thyme honeys contain a minimum of 18 % of thyme pollen grains, while the full pollen grains spectrum was directly affected by the geographical origin of where the hives were located, as described previously in the melissopalynology analysis. In contrast, the RF model for the discrimination of polyfloral honeys from Greece, Malta, Spain, Tunisia, and Turkey achieved only 68.5 % accuracy, with performance comparable to the general model. This limited improvement is due to the heterogeneous composition of polyfloral honeys, which contain a mix of compounds from various floral sources, that do not facilitate their better differentiation.

Generally, using DD-SIMCA models, a specific class can be predicted, such as determining whether a honey sample originates from a particular country or a specific combination of country and floral type. DD-SIMCA models were constructed to classify honeys from Greece, Malta, Spain, Tunisia, and Turkey using all honey samples, achieving an accuracy of 97.2 %, 95.4 %, 96.8 %, 86.8 %, and 96.7 %, respectively, as shown in Table 4. The performance of the DD-SIMCA models were assessed based on accuracy, sensitivity, and specificity, with specificity measuring the proportion of correctly identified negative observations relative to the total number of actual negatives.

When assessing the geographical origin of thyme honeys separately, DD-SIMCA models achieved accuracies of 95.2 %, 95.4 %, and 88.9 % for Greek, Spanish, and Tunisian thyme honeys, respectively. The accuracy of these models showed a slight improvement; however, sensitivity and specificity declined when the validation set included all honey samples instead of being limited to thyme honey samples. This result was anticipated, as the inclusion of a larger number of samples for model validation inherently increased variability (Table 4). The acceptance plot of validation for Greek thyme honeys (Fig. 9.1.b) shows that three out of 26 Greek thyme samples were misclassified. However, no thyme honey samples from other origins were incorrectly classified as Greek. Similarly, for Spanish thyme honeys, two out of 17 Spanish samples were misclassified, while none of the other thyme honeys were incorrectly classified as Spanish (Fig. 9.2.b). In contrast, as illustrated in the acceptance plot of validation for the Tunisian thyme honey DD-SIMCA model (Fig. 9.3.b) three Tunisian thyme honeys were misclassified and one Greek thyme sample were misclassified as Tunisian one. Consequently, the sensitivity values were 89.6 %, 82.9 %, and 90.7 %, and the specificity values were 100 %, 100 %, and 93.7 % for thyme honeys from Greece, Spain, and Tunisia, respectively (Table 4).

Regarding the models constructed to classify pine honey, the accuracy was 89.2 % for Greek and 85.0 % for Turkish honeys, when using only pine honeys, while the accuracy values improved to 99.1 % and 98.8 % respectively, when the validation set included all honey samples. According to specificity metrics, no samples from other classes were misclassified as pine honey in either model. However, one original Turkish pine sample and two Greek pine samples were misclassified, reducing the sensitivity of the models (Table 4). Increasing the number of pine honey samples for validation could improve both accuracy and sensitivity, while adding non-target class samples enhanced overall model accuracy (Table 4).

For polyfloral honeys, DD-SIMCA models resulted in a validation accuracy of 97.4 %, 99.0 %, 90.0 %, 94.5 %, and 92.3 % for honeys from Greece, Spain, Malta, Tunisia, and Turkey, respectively, using only polyfloral honey samples. Based on sensitivity and specificity values, it was more likely for an original polyfloral sample to be misclassified compared to another honey type being incorrectly classified as a specific origin or type.

The DD-SIMCA models for identifying the botanical and geographical origin of honeys based on FTIR spectra demonstrated superior accuracy compared to models developed using UV–vis spectra (Dimakopoulou-Papazoglou et al., 2024), as FTIR spectra provide more comprehensive information about the composition of the samples. Additionally, good performance in classifying honey by geographical origin was also observed by Formosa et al. (2020), who analyzed Maltese and foreign honey samples (from Greece, Italy, Sicily, France, Estonia, and others) using PLS-DA, and by Guyon et al. (2021), who differentiated French and Romanian honey samples using SIMCA.

3.4.2. *Determination of botanical origin*

Predictive models were developed to determine the botanical origin of honey using Greek thyme, pine, and polyfloral honey samples. RF was applied as a multiclass classification technique, achieving a validation accuracy of 73.3 %, with precision, recall, and F1-score values of 64.8 %, 64.0 %, and 61.9 %, respectively. The relatively low accuracy of the RF model was attributed to the misclassification of some polyfloral samples as thyme honeys. This was expected, as polyfloral honeys are derived from a variety of blossoms and trees, including *Thymus* species (Lamiaceae). Hence, by removing the polyfloral samples from the classification model, the accuracy of the binary classification model improved to 95.7 %, with an increase in precision, recall, and F1-score. Ciulu et al. (2021) also reported good performance of RF to identify the botanical origin of honeys using FTIR spectra from Italy (Sardinia), and specifically from strawberry-tree, asphodel, thistle, and eucalyptus. Additionally, the effectiveness of RF for botanical classification using UV–vis spectroscopy has also been reported by Dimakopoulou-Papazoglou et al. (2024).

Subsequently, DD-SIMCA models were developed to predict each class separately. For thyme honeys, the validation accuracy was 84.6 % when tested against the remaining Greek honey samples. However, some polyfloral samples were misclassified as thyme honeys, a pattern also observed in the PCA plot (Fig. 8), indicating similarities in the FTIR spectra of polyfloral and thyme honeys. In contrast, when building a model to predict polyfloral honeys, the validation accuracy increased to 91.6 %. The model for pine honeys achieved the highest accuracy of 95.7 %, likely due to the distinct composition of pine honey compared to other types. Similarly, the good performance of SIMCA was demonstrated by Guyon et al. (2021), who successfully differentiated honey samples from acacia, colza, linden, and sunflower origins with high accuracy in samples from France and Romania. These findings align with previous studies that utilized alternative models for classifying

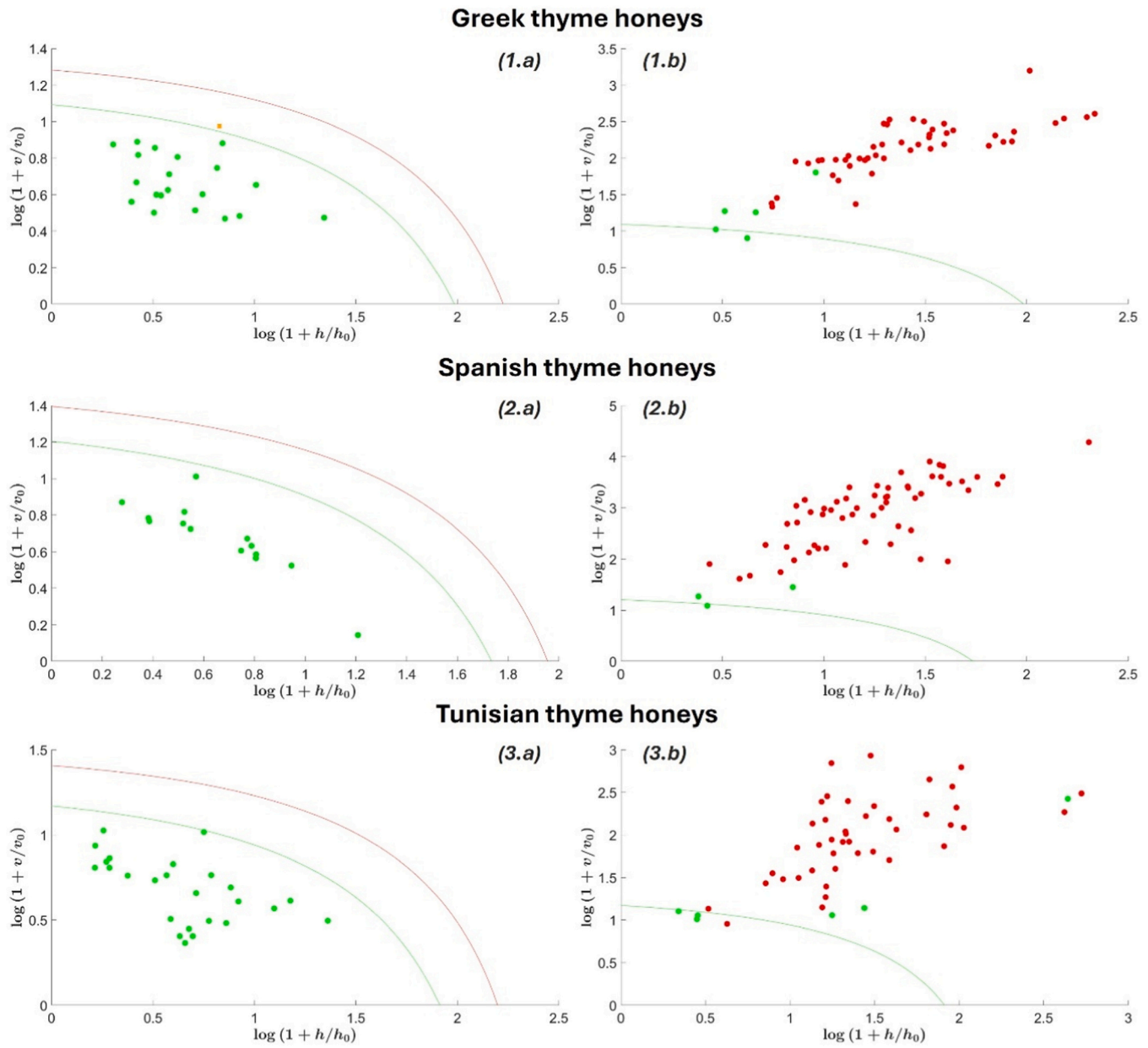

Fig. 9. Acceptance plot of the training (1.a, 2.a, and 3.a) and validation (1.b, 2.b, and 3.b) sets from DD-SIMCA models using the FTIR spectra pre-processed with 1st derivatives for Greek, Spanish, and Tunisian thyme honeys (Acceptance plot for training set: training samples are represented by green circles, with extreme samples shown as orange circles between the green and red lines. Acceptance plot for validation set: target class honey samples are shown as green circles, while non-target class samples appear as red circles.).

Table 3
Results obtained from Random Forest for identifying the geographical and botanical origin of honeys using FTIR spectra.

Samples used for analysis	Accuracy	Precision	Recall	F1-score
Discrimination of geographical origin				
All honey samples	68.00 %	61.70 %	62.02 %	60.83 %
Only thyme honeys	83.57 %	86.53 %	84.33 %	84.74 %
Only pine honeys	92.50 %	95.00 %	92.50 %	92.00 %
Only polyfloral honeys	68.46 %	63.83 %	61.40 %	60.92 %
Discrimination of botanical origin				
All honey samples	74.19 %	77.65 %	68.00 %	70.47 %
Only Greek honeys	73.33 %	64.82 %	64.00 %	61.90 %
Only Greek honeys (thyme and pines)	95.71 %	96.67 %	95.50 %	94.96 %

honey based on botanical origin. Orfanakis et al. (2021) utilized OPLS-DA to distinguish polyfloral and honeydew honeys from Greece (Creta) and achieve the high classification accuracy of 91.2 % and 82.6 %, respectively, using the spectral range 4000 – 2400 cm^{-1} and 1900 – 400 cm^{-1}. Similarly, Tsagkaris et al. (2023) demonstrated effective discrimination among different monofloral honeys (blossom, cotton, thyme, honeydew, and citrus) using PLS-DA and OPLS-DA.

4. Conclusions

This study demonstrated that FTIR spectroscopy combined with chemometrics can effectively classify Mediterranean honeys based on their botanical and geographical origins. The honey samples analyzed originated from various Mediterranean countries, including Greece, Spain, Tunisia, Malta, and Turkey, and came from different botanical sources: thyme, pine, and polyfloral. The spectral region of 1800 –

Table 4
Results obtained from DD-SIMCA for identifying the geographical and botanical origin of honeys using FTIR spectra.

Samples used for analysis	Accuracy	Sensitivity	Specificity
Differentiation of geographical origin – using all honey samples			
Greek honeys	97.23 %	92.67 %	100.00 %
Spanish honeys	96.84 %	84.44 %	100.00 %
Maltese honeys	95.44 %	87.00 %	96.28 %
Tunisian honeys	86.78 %	93.10 %	88.41 %
Turkish honeys	96.69 %	89.35 %	99.19 %
Differentiation of geographical origin of Thyme honeys – using only thyme samples			
Greek honeys	95.18 %	89.62 %	100.00 %
Spanish honeys	95.40 %	82.94 %	100.00 %
Tunisian honeys	88.85 %	90.65 %	93.70 %
Differentiation of geographical origin of Thyme honeys – using all honey samples			
Greek honeys	96.79 %	87.31 %	99.22 %
Spanish honeys	97.87 %	82.35 %	100.00 %
Tunisian honeys	89.85 %	91.29 %	91.53 %
Differentiation of geographical origin of Pine honeys – using only pine samples			
Greek honeys	89.17 %	87.00 %	100.00 %
Turkish honeys	85.00 %	82.00 %	100.00 %
Differentiation of geographical origin of Pine honeys – using all honey samples			
Greek honeys	99.12 %	87.00 %	100.00 %
Turkish honeys	98.78 %	82.00 %	100.00 %
Differentiation of geographical origin of Polyfloral honeys – using only polyfloral samples			
Greek honeys	97.40 %	87.00 %	100.00 %
Spanish honeys	99.00 %	95.00 %	100.00 %
Maltese honeys	90.00 %	87.00 %	92.29 %
Tunisian honeys	94.49 %	90.00 %	96.60 %
Turkish honeys	92.27 %	81.11 %	100.00 %
Differentiation of botanical origin – using only Greek honey samples			
Thyme honeys	84.58 %	89.62 %	94.74 %
Pine honeys	95.68 %	85.00 %	99.71 %
Polyfloral honeys	91.58 %	81.11 %	95.83 %

750 cm^{-1} was the most useful for the classification of the samples using Random Forest (for multiclass classification) and DD-SIMCA model (for one-class classification). These techniques successfully discriminated the geographical and botanical origins of the samples, achieving an accuracy exceeding 90 % in most cases. The findings underscore that FTIR spectroscopy, being highly sensitive to the overall chemical composition of samples, provides reliable results for determining the botanical and geographical origins of honey when combined with multivariate statistical analysis. This quick, non-destructive method, coupled with user-friendly chemometric techniques, can be a promising analytical tool in the honey industry, enhancing food traceability.

However, there are a few limitations to consider. Factors like seasonal changes or environmental conditions, which can affect the composition of honey, were not considered in this study. These variables could influence the consistency of the results. Additionally, while the study included honeys from different Mediterranean countries and floral sources, the overall number of samples was relatively small. Including more samples from a wider range of regions and plant origins would make the models more robust and reliable.

CRediT authorship contribution statement

Rodríguez Inmaculada: Writing – review & editing, Investigation. **Serrano Salud:** Writing – original draft, Methodology, Investigation, Formal analysis. **Dimakopoulou-Papazoglou Dafni:** Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **Katsanidis Eugenios:** Writing – review & editing, Supervision, Project administration, Methodology. **Koutsoumanis Konstantinos:** Writing – review & editing, Project administration, Funding acquisition. **Ploskas Nikolaos:** Writing – review & editing, Formal analysis, Data curation.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

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.

Acknowledgement

This paper is part of the PRIMA programme, supported by European Union's Horizon 2020 research and innovation programme, under grant agreement No 1932, project MEDIFIT (Call 2019 1 Agrofood IA).

The authors gratefully acknowledge the valuable contributions of Konstantinos Koutsoumanis, who sadly passed away during the final acceptance of this manuscript.

Data availability

Data will be made available on request.

References

Afrin, S., Haneefa, S.M., Fernandez-Cabezudo, M.J., Giampieri, F., al-Ramadi, B.K., Battino, M., 2020. Therapeutic and preventive properties of honey and its bioactive compounds in cancer: an evidence-based review (Cambridge Core). Nutr. Res. Rev. 33 (1), 50–76. https://doi.org/10.1017/S0954422419000192.

Anjos, O., Campos, M.G., Ruiz, P.C., Antunes, P., 2015. Application of FTIR-ATR spectroscopy to the quantification of sugar in honey. Food Chem. 169, 218–223. https://doi.org/10.1016/j.foodchem.2014.07.138.

Ansari, M.J., Al-Ghamdi, A., Khan, K.A., Adgaba, N., El-Ahmady, S.H., Gad, H.A., Roshan, A., Meo, S.A., Kolyali, S., 2018. Validation of botanical origins and geographical sources of some Saudi honeys using ultraviolet spectroscopy and chemometric analysis. Saudi J. Biol. Sci. 25. https://doi.org/10.1016/j.sjbs.2017.09.005.

Batista, B.L., da Silva, L.R.S., Rocha, B.A., Rodrigues, J.L., Berretta-Silva, A.A., Bonates, T.O., Gomes, V.S.D., Barbosa, R.M., Barbosa, F., 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 (1), 209–215. https://doi.org/10.1016/j.foodres.2012.07.015.

Bertelli, D., Lolli, M., Papotti, G., Bortolotti, L., Serra, G., Plessi, M., 2010. Detection of honey adulteration by sugar syrups using one-dimensional and two-dimensional high-resolution nuclear magnetic resonance. J. Agric. Food Chem. 58. https://doi.org/10.1021/jf101460t.

Breiman, L., 2001. Random forests. Mach. Learn. 45. https://doi.org/10.1023/A:1010933404324.

Cengiz, M.F., Durak, M.Z., 2019. Rapid detection of sucrose adulteration in honey using Fourier transform infrared spectroscopy. Spectrosc. Lett. 52. https://doi.org/10.1080/00387010.2019.1615957.

Chataway, H.D., 1932. The determination of moisture in honey. Can. J. Res. 6 (5), 532–547. https://doi.org/10.1139/cjr32-041.

Chen, L., Wang, J., Ye, Z., Zhao, J., Xue, X., Heyden, Y.V., Sun, Q., 2012. Classification of Chinese honeys according to their floral origin by near infrared spectroscopy. Food Chem. 135 (2), 338–342. https://doi.org/10.1016/j.foodchem.2012.02.156.

Chen, Q., Qi, S., Li, H., Han, X., Ouyang, Q., Zhao, J., 2014. Determination of rice syrup adulterant concentration in honey using three-dimensional fluorescence spectra and multivariate calibrations. Spectrochim. Acta Part A: Mol. Biomol. Spectrosc. 131. https://doi.org/10.1016/j.saa.2014.04.071.

Ciulu, M., Oertel, E., Serra, R., Farre, R., Spano, N., Caredda, M., Malfatti, L., Sanna, G., 2021. Classification of unifloral honeys from Sardinia (Italy) by ATR-FTIR spectroscopy and Random Forest. Molecules 26 (1). https://doi.org/10.3390/molecules26010088.

David, M., Hategan, A.R., Magdas, D.A., Berghian-Grosan, C., Simionescu, B., 2022. Botanical origin assessment of honey based on ATR-IR spectroscopy: a comparison between the efficiency of supervised statistical methods and artificial intelligence. Appl. Sci. 12 (19). https://doi.org/10.3390/app12199645.

de Souza, R.R., Sousa Fernandes, D.D., Diniz, P.H.G.D., 2021. Honey authentication in terms of its adulteration with sugar syrups using UV–Vis spectroscopy and one-class classifiers. Food Chem. 365. https://doi.org/10.1016/j.foodchem.2021.130467.

Devi, A., Jangir, J., K.A, A.-A., 2018. Chemical characterization complemented with chemometrics for the botanical origin identification of unifloral and multifloral honeys from India. Food Res. Int. 107, 216–226. https://doi.org/10.1016/j.foodres.2018.02.017.

Di Marco, G., Manfredini, A., Leonardi, D., Canuti, L., Impei, S., Gismondi, A., Canini, A., 2017. Geographical, botanical and chemical profile of monofloral Italian honeys as

food quality guarantee and territory brand. Plant Biosyst. Int. J. Deal. All Asp. Plant Biol. 151 (3), 450–463. https://doi.org/10.1080/11263504.2016.1179696.
Dimakopoulou-Papazoglou, D., Ploskas, N., Koutsoumanis, K., Katsanidis, E., 2024. Identification of geographical and botanical origin of Mediterranean honeys using UV–vis spectroscopy and multivariate statistical analysis. J. Food Meas. Charact. 18 (5), 3923–3934. https://doi.org/10.1007/s11694-024-02465-2.
Dimakopoulou-Papazoglou, D., Ploskas, N., Serrano, S., Silva, C.S., Valdramidis, V., Koutsoumanis, K., Katsanidis, E., 2023. Application of UV–Vis spectroscopy for the detection of adulteration in Mediterranean honeys. Eur. Food Res. Technol. 249 (12), 3043–3053. https://doi.org/10.1007/s00217-023-04347-1.
Dimou, M., Katsaros, J., Klonari, K.T., Thrasyvoulou, A., 2006. Discriminating pine and fir honeydew honeys by microscopic characteristics. J. Apic. Res. 45 (2), 16–21. https://doi.org/10.1080/00218839.2006.11101319.
Dinca, O.-R., Ionete, R.E., Popescu, R., Costinel, D., Radu, G.-L., 2015. Geographical and botanical origin discrimination of Romanian honey using complex stable isotope data and chemometrics. Food Anal. Methods 8 (2), 401–412. https://doi.org/10.1007/s12161-014-9903-x.
Duru, M.E., Taş, M., Çayan, F., Küçükaydın, S., Tel-Çayan, G., 2021. Characterization of volatile compounds of Turkish pine honeys from different regions and classification with chemometric studies. Eur. Food Res. Technol. 247 (10), 2533–2544. https://doi.org/10.1007/s00217-021-03817-8.
EU Council Directive 2001/110/EC of 20 December 2001 relating to honey. Off J Eur Commun Legis 10:47–52 (2002). Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32001L0110&from=EN.
EU Council Directive 2014/63/EU of 15 May 2014 relating to honey (2014). Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014L0063.
FAO 2022 ⟨https://www.fao.org/faostat/en/#data/TCL⟩.
Formosa, J.P., Lia, F., Mifsud, D., Farrugia, C., 2020. Application of ATR-FT-MIR for tracing the geographical origin of honey produced in the Maltese islands. Foods 9 (6). https://doi.org/10.3390/foods9060710.
Gallardo-Velázquez, T., Osorio-Revilla, G., Loa, M.Z., Rivera-Espinoza, Y., 2009. Application of FTIR-HATR spectroscopy and multivariate analysis to the quantification of adulterants in Mexican honeys. Food Res. Int. 42 (3), 313–318. https://doi.org/10.1016/j.foodres.2008.11.010.
Gan, Z., Yang, Y., Li, J., Wen, X., Zhu, M., Jiang, Y., Ni, Y., 2016. Using sensor and spectral analysis to classify botanical origin and determine adulteration of raw honey. J. Food Eng. 178, 151–158. https://doi.org/10.1016/j.jfoodeng.2016.01.016.
Gok, S., Severcan, M., Goormaghtigh, E., Kandemir, I., Severcan, F., 2015. Differentiation of Anatolian honey samples from different botanical origins by ATR-FTIR spectroscopy using multivariate analysis. Food Chem. 170, 234–240. https://doi.org/10.1016/j.foodchem.2014.08.040.
Grabato, J.R., Pilario, K.E., Micor, J.R.L., Mojica, E.-R.E., 2022. Geographical and entomological differentiation of Philippine honey by multivariate analysis of FTIR spectra. J. Food Compos. Anal. 114, 104853. https://doi.org/10.1016/j.jfca.2022.104853.
Greek directive 127, /2004, Government Gazette B-239/23-2-2005; Annex II, article 67 of Greek Food Code (Classification of monofloral honeys). Available: ⟨https://www.aade.gr/sites/default/files/2020-03/67a-iss1.pdf⟩.
Guelpa, A., Marini, F., Plessis, A., Slabbert, R., Manley, M., 2017. Verification of authenticity and fraud detection in South African honey using NIR spectroscopy. Food Control 73. https://doi.org/10.1016/j.foodcont.2016.11.002.
Guyon, F., Logodin, E., Magdas, D.A., Gaillard, L., 2021. Potential of FTIR- ATR diamond in discriminating geographical and botanical origins of honeys from France and Romania. Talanta Open 3, 100022. https://doi.org/10.1016/j.talo.2020.100022.
Huang, F., Song, H., Guo, L., Guang, P., Yang, X., Li, L., Zhao, H., Yang, M., 2020. Detection of adulteration in Chinese honey using NIR and ATR-FTIR spectral data fusion. Spectrochim. Acta A Mol. Biomol. Spectrosc. 235. https://doi.org/10.1016/j.saa.2020.118297.
IHC Methods. (2009). *International Honey Commission. Harmonized Methods.* ⟨https://www.ihc-platform.net/ihcmethods2009.pdf⟩.
Karabagias, I.K., Badeka, A., Kontakos, S., Karabournioti, S., Kontominas, M.G., 2014. Characterisation and classification of Greek pine honeys according to their geographical origin based on volatiles, physicochemical parameters and chemometrics. Food Chem. 146, 548–557. https://doi.org/10.1016/j.foodchem.2013.09.105.
Karabagias, I.K., Louppis, A.P., Karabournioti, S., Kontakos, S., Papastephanou, C., Kontominas, M.G., 2017. Characterization and classification of commercial thyme honeys produced in specific Mediterranean countries according to geographical origin, using physicochemical parameter values and mineral content in combination with chemometrics. Eur. Food Res. Technol. 243 (5), 889–900. https://doi.org/10.1007/s00217-016-2803-0.
Kasprzyk, I., Depciuch, J., Grabek-Lejko, D., Parlinska-Wojtan, M., 2018. FTIR-ATR spectroscopy of pollen and honey as a tool for unifloral honey authentication. The case study of rape honey. Food Control 84, 33–40. https://doi.org/10.1016/j.foodcont.2017.07.015.
Kędzierska-Matysek, M., Teter, A., Florek, M., Matwijczuk, A., Niemczynowicz, A., Matwijczuk, A., Czernel, G., Skałecki, P., Gładyszewska, B., 2023. Use of physicochemical, FTIR and chemometric analysis for quality assessment of selected monofloral honeys. J. Apic. Res. 62 (4), 863–872. https://doi.org/10.1080/00218839.2021.1900637.
Louveaux, J., Maurizio, A., Vorwohl, G., 1978. Methods of Melissopalynology. BEE World 59 (4), 139–157. https://doi.org/10.1080/0005772X.1978.11097714.
Maione, C., Barbosa, F., Barbosa, R.M., 2019. Predicting the botanical and geographical origin of honey with multivariate data analysis and machine learning techniques: a review. Comput. Electron. Agric. 157, 436–446. https://doi.org/10.1016/j.compag.2019.01.020.
Martinez-Castillo, C., Astray, G., Mejuto, J.C., Simal-Gandara, J., 2020. Random forest, artificial neural network, and support vector machine models for honey classification. eFood 1 (1), 69–76. https://doi.org/10.2991/efood.k.191004.001.
Oddo, L.P., Piro, R., Bruneau, É ., Guyot-Declerck, C., Ivanov, T., Piskulová, J., Flamini, C., Lheritier, J., Morlot, M., Russmann, H., 2004. Main European unifloral honeys: descriptive sheets. Apidologie 35 (. 1), S38–S81.
Orfanakis, E., Markoulidakis, M., Philippidis, A., Zoumi, A., Velegrakis, M., 2021. Optical spectroscopy methods combined with multivariate statistical analysis for the classification of cretan thyme, multi-floral and honeydew honey. J. Sci. Food Agric. 101. https://doi.org/10.1002/jsfa.11182.
Oroian, M., Ropciuc, S., Paduret, S., 2018. Honey adulteration detection using Raman spectroscopy. Food Anal. Methods 11. https://doi.org/10.1007/s12161-017-1072-2.
Özkök, A., Yüksel, D., Sorkun, K., 2018. Chemometric evaluation of the geographical origin of Turkish Pine Honey. Food Health 4 (4), 274–282.
Pauliuc, D., Ciursă, P., Ropciuc, S., Dranca, F., Oroian, M., 2021. Physicochemical parameters prediction and authentication of different monofloral honeys based on FTIR spectra. J. Food Compos. Anal. 102, 104021. https://doi.org/10.1016/j.jfca.2021.104021.
Rachineni, K., Rao Kakita, V.M., Awasthi, N.P., Shirke, V.S., Hosur, R.V., Chandra Shukla, S., 2022. Identifying type of sugar adulterants in honey: Combined application of NMR spectroscopy and supervised machine learning classification. Curr. Res. Nutr. Food Sci. 5. https://doi.org/10.1016/j.crfs.2022.01.008.
Roshan, A.-R.A., Gad, H.A., El-Ahmady, S.H., Khanbash, M.S., Abou-Shoer, M.I., Al-Azizi, M.M., 2013. Authentication of monofloral Yemeni sidr honey using ultraviolet spectroscopy and chemometric analysis. J. Agric. Food Chem. 61. https://doi.org/10.1021/jf402280y.
Se, K.W., Ghoshal, S.K., Wahab, R.A., Ibrahim, R.K.R., Lani, M.N., 2018. A simple approach for rapid detection and quantification of adulterants in stingless bees (Heterotrigona itama) honey. Food Res. Int. 105, 453–460. https://doi.org/10.1016/j.foodres.2017.11.012.
Sorkun, K., 2008. Türkiye' nin Nektarlı Bitkileri, Polenleri ve Balları (Nectar. Plants, Honeys Pollens Turk.), Palme Publ. 341. ISBN 978-9944-341- 67-7.
Suhandy, D., Yulia, M., 2021. The use of UV spectroscopy and SIMCA for the authentication of indonesian honeys according to botanical, entomological and geographical origins. Molecules 26. https://doi.org/10.3390/molecules26040915.
Svečnjak, L., Bubalo, D., Baranović, G., Novosel, H., 2015. Optimization of FTIR-ATR spectroscopy for botanical authentication of unifloral honey types and melissopalynological data prediction. Eur. Food Res. Technol. 240 (6), 1101–1115. https://doi.org/10.1007/s00217-015-2414-1.
Svečnjak, L., Prđun, S., Rogina, J., Bubalo, D., Jerković, I., 2017. Characterization of Satsuma mandarin (Citrus unshiu Marc.) nectar-to-honey transformation pathway using FTIR-ATR spectroscopy. Food Chem. 232, 286–294. https://doi.org/10.1016/j.foodchem.2017.03.159.
Tahir, H.E., Xiaobo, Z., Zhihua, L., Jiyong, S., Zhai, X., Wang, S., Mariod, A.A., 2017. Rapid prediction of phenolic compounds and antioxidant activity of Sudanese honey using Raman and Fourier transform infrared (FT-IR) spectroscopy. Food Chem. 226, 202–211. https://doi.org/10.1016/j.foodchem.2017.01.024.
Terrab, A., Recamales, A.F., Hernanz, D., Heredia, F.J., 2004. Characterisation of Spanish thyme honeys by their physicochemical characteristics and mineral contents. Food Chem. 88 (4), 537–542. https://doi.org/10.1016/j.foodchem.2004.01.068.
Tewari, J., Irudayaraj, J., 2004. Quantification of saccharides in multiple floral honeys using Fourier transform infrared microattenuated total reflectance spectroscopy. J. Agric. Food Chem. 52. https://doi.org/10.1021/jf035176+.
Thrasyvoulou, A., Manikis, J., 1995. Some physicochemical and microscopic characteristics of Greek unifloral honeys. Apidologie 26 (6), 441–452. https://doi.org/10.1051/apido:19950601.
Tsagkaris, A.S., Bechynska, K., Ntakoulas, D.D., Pasias, I.N., Weller, P., Proestos, C., Hajslova, J., 2023. Investigating the impact of spectral data pre-processing to assess honey botanical origin through Fourier transform infrared spectroscopy (FTIR). J. Food Compos. Anal. 119, 105276. https://doi.org/10.1016/j.jfca.2023.105276.
Tsigouri, A., Passaloglou-Katrali, M., Sabatakou, O., 2004. Palynological characteristics of different unifloral honeys from Greece. Grana 43 (2), 122–128. https://doi.org/10.1080/00173130310017643.
Vîjan, L.E., Mazilu, I.C., Enache, C., Enache, S., Topală, C.M., 2023. Botanical origin influence on some honey physicochemical characteristics and antioxidant properties. Foods 12 (11). https://doi.org/10.3390/foods12112134.
Von Der Ohe, W., Oddo, L.P., Piana, M.L., Morlot, M., Martin, P., 2004. Harmonized methods of melissopalynology. Apidologie 35 (. 1), S18–S25. https://doi.org/10.1051/apido:2004050.
Vorwohl, G., 1964. Messung der elektrischen Leitfähigkeit des Honigs und der Verwendung der Messwerte zur Sortendiagnose und zum Nachweis von VerfÄlschungen mit Zuckerfütterungshonig. Z. Biene 7 (2), 37–47.
Wang, J., Kliks, M.M., Jun, S., Jackson, M., Li, Q.X., 2010. Rapid analysis of glucose, fructose, sucrose, and maltose in honeys from different geographic regions using Fourier transform infrared spectroscopy and multivariate analysis. J. Food Sci. 75. https://doi.org/10.1111/j.1750-3841.2009.01504.x.
Wedmore, E.B., 1955. The accurate determination of the water content of honeys: part I. Introduction and results. BEE World 36 (11), 197–206. https://doi.org/10.1080/0005772X.1955.11094904.

White, J.W., Jr, 1979. Spectrophotometric method for hydroxymethylfurfural in honey. J. Assoc. Off. Anal. Chem. 62 (3), 509–514. https://doi.org/10.1093/jaoac/62.3.509.

Wu, X., Xu, B., Ma, R., Niu, Y., Gao, S., Liu, H., Zhang, Y., 2022. Identification and quantification of adulterated honey by Raman spectroscopy combined with convolutional neural network and chemometrics. Spectrochim. Acta Part A: Mol. Biomol. Spectrosc. 274. https://doi.org/10.1016/j.saa.2022.121133.

Yan, S., Sun, M., Wang, X., Shan, J., Xue, X., 2022. A novel, rapid screening technique for sugar syrup adulteration in honey using fluorescence spectroscopy. Foods 11. https://doi.org/10.3390/foods11152316.