NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

SCHOOL OF SCIENCES

DEPARTMENT OF CHEMISTRY

DOCTORAL DISSERTATION

Development of High Resolution Mass Spectrometric Methods for the investigation of food authenticity

KALOGIOURI NATASA

MSc CHEMIST

ATHENS

JULY 2017

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Ανάπτυξη Μεθόδων Φασματομετρίας Μάζας Υψηλής Διακριτικής Ικανότητας για τη μελέτη της αυθεντικότητας τροφίμων

ΚΑΛΟΓΙΟΥΡΗ ΝΑΤΑΣΑ

Α.Μ.: 001305

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ:

Νικόλαος Θωμαϊδης, Αναπληρωτής Καθηγητής ΕΚΠΑ

ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ:

Κωνσταντίνος Ευσταθίου, Καθηγητής

Μιχαήλ Κουππάρης, Καθηγητής

Νικόλαος Θωμαΐδης, Αναπληρωτής Καθηγητής

ΕΠΤΑΜΕΛΗΣ ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ

Κωνσταντίνος Ευσταθίου, Καθηγητής

Αντώνιος Καλοκαιρινός, Καθηγητής

Μιχαήλ Κουππάρης, Καθηγητής

Βικτώρια Σαμανίδου, Καθηγήτρια του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης, Τμήμα Χημείας

Νικόλαος Θωμαΐδης, Αναπληρωτής Καθηγητής

Ευάγγελος Μπακέας, Αναπληρωτής Καθηγητής

Αναστάσιος Οικονόμου, Αναπληρωτής Καθηγητής

ΗΜΕΡΟΜΗΝΙΑ ΕΞΕΤΑΣΗΣ: 18/7/2017

DOCTORAL DISSERTATION

Development of High Resolution Mass Spectrometric Methods for the investigation of food authenticity

KALOGIOURI NATASA

Registration Number: 001305

SUPERVISING PROFESSOR:

Dr. Nikolaos Thomaidis**,** Associate Professor

THREE-MEMBER CONSULTATIVE COMITTEE:

Dr. Konstantinos Efstathiou, Professor

Dr. Michail Koupparis, Professor

Dr. Nikolaos Thomaidis, Associate Professor

SEVEN-MEMBER EXAMINATION COMITTEE:

Dr. Konstantinos Efstathiou, Professor

Dr. Antony Calokerinos, Professor

Dr. Michail Koupparis, Professor

Dr. Victoria Samanidou, Professor in the Aristotle University of Thessaloniki, Department of Chemistry

Dr. Nikolaos Thomaidis, Associate Professor

Dr. Evangelos Bakeas, Associate Professor

Dr. Anastasios Economou, Associate Professor

DEFENDING DATE: 18/7/2017

ABSTRACT

Food authenticity has become increasingly important in recent years due to food fraud incidents and the handling of low quality products with misleading labels. The guarantee of the authenticity of olive oil arises great public concern because of its sensory, nutritional and economic importance. The main objective of this thesis is to develop integrated LC-HRMS workflows, including target, suspect and non-target screening strategies, coupled with advanced chemometric tools, for olive oil fingerprinting.

First, the biological activities of some minor constituents in olive oil, phenolic compounds, are reported and their occurrence and wide-scope properties in olive matrices, as well as effects of various factors on olive oil phenolic profile, are discussed. The present state of the art for their determination in foodomics science is presented, focusing on target and non-target screening HR-MS workflows coupled to chemometrics. The experimental section of the thesis consists of three sections: (1) Olive oil authenticity studies by target and non-target LC-QTOF-MS combined with advanced chemometrics, for identifying markers that classify olive oil to defective and EVOOs (Chapter 3), (2) Investigating the organic and conventional production type of EVOOs with target and suspect screening by LC-QTOF-MS, a novel semi-quantification method using chemical similarity and advanced chemometris; in order to identify a marker with a concentration threshold, by ACO/RF, that can be used to discriminate organic and conventional EVOOs (Chapter 4), and (3) Classification of Greek olive oil varieties with non-target UHPLC-QTOF-MS and advanced chemometrics; for the investigation of the fingerprints of six greek olive oil varieties and the identification of markers, with post-defined concentration thresholds, that guarantee the varietal and geographical origin (Chapter 5).

We believe that these studies have made great progress in the food authenticity field via the introduction of novel integrated HRMS screening workflows which are followed by advanced data processing, comprehensive data mining and predictive modelling tools.

SUBJECT AREA: Analytical Chemistry

KEYWORDS: Authenticity, olive oil, HR-MS, non-target screening, chemometrics

ΠΕΡΙΛΗΨΗ

Η αυθεντικότητα των τροφίμων αποτελεί ένα ιδαίτερα σημαντικό γεγονός τα τελευταία χρόνια, λόγω των πολλών περιστατικών νοθείας τροφίμων και διακίνησης προϊόντων κατώτερης ποιότητας με παραπλανητικές ετικέτες. Η εξασφάλιση της αυθεντικότητας του ελαιολάδου δημιουργεί μεγάλη ανησυχία λόγω της οργανοληπτικής, θρεπτικής και οικονομικής του σημασίας. Ο κύριος στόχος αυτής της διατριβής είναι η ανάπτυξη αναλυτικών μεθόδων υγροχρωματογραφίας φασματομετρίας μάζας υψηλής διακριτικής ικανότητας (LC-HRMS) που θα συμπεριλαμβάνουν στρατηγικές στοχευμένης, «ύποπτης» και «μη στοχευμένης σάρωσης», σε συνδυασμό με προηγμένα χημειομετρικά εργαλεία, για την εύρεση του αποτυπώματος του ελαιολάδου.

Αρχικά, αναφέρονται οι βιολογικές δράσεις των ελάσσονων συστατικών του ελαιολάδου, των φαινολικών ενώσεων, και οι επιδράσεις διαφόρων παραγόντων στο φαινολικό προφίλ. Παρουσιάζονται όλες οι σύγχρονες και αναλυτικές μέθοδοι που συνδυάζουν την επιστήμη των τροφίμων με τεχνικές μεταβολομικής, με έμφαση στις μεθόδους HR-MS. Το πειραματικό μέρος της διατριβής αποτελείται από τρία τμήματα: (1) Μελέτες αυθεντικότητας ελαιολάδου με LC-QTOF-MS σε συνδυασμό με στοχευμένη και μη στοχευμένη σάρωση και χημειομετρία για την αναγνώριση δεικτών που κατατάσσουν το ελαιόλαδο ως ελαττωματικό ή έξτρα παρθένο. (Κεφάλαιο 3), (2) Διερεύνηση του βιολογικού και συμβατικού τύπου παραγωγής των έξτρα παρθένων ελαιολάδων με LC-QTOF-MS σε συνδυασμό με στοχευμένη και «ύποπτη» σάρωση, προτείνοντας μια νέα μέθοδο ημι-ποσοτικοποίησης (Κεφάλαιο 4), και (3) Ταξινόμηση έξι ελληνικών ποικιλιών ελαιολάδου με LC-QTOF-MS μη στοχευμένη στοχευμένη σάρωση και χημειομετρία, για την αναγνώριση δεικτών με προκαθορισμένα όρια συγκέντρωσης, που εγγυώνται την ποικιλιακή και γεωγραφική προέλευση (Κεφάλαιο 5).

Πιστεύουμε ότι οι παραπάνω μελέτες έχουν σημειώσει μεγάλη πρόοδο στον τομέα της γνησιότητας των τροφίμων με την ανάπτυξη αναλυτικών μεθόδων HR-MS και μέσω της εισαγωγής μιας νέας ολοκληρωμένης ροής εργασιών που περιλαμβάνει στρατηγικές στοχευμένης, «ύποπτης» και μη στοχευμένης σάρωσης HRMS, σε συνδυασμό με, ολοκληρωμένα εργαλεία εξόρυξης, επεξεργασίας δεδομένων και προγνωστικά μοντέλα.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Αναλυτική Χημεία

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: Αυθεντικότητα, ελαιόλαδο, HR-MS, μη στοχευμένη σάρωση, χημειομετρία

To my grandfather and grandmother

Στον παππού και στη γιαγιά

“Sometimes you have to play a long time to be able to play like yourself”

- **Miles Davis** (1926-1991)

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my thesis director Dr. Nikolaos Thomaidis, for having accepted me to realize this doctoral thesis in his research group. His good advice, continuous guidance and encouragements were key motivations throughout my PhD. I am grateful to him for his trust, understanding and support both in academic and personal life.

Special thanks are attributed to the three-member committee and seven-member examination committee for their cooperation throughout this thesis and their valuable comments.

I am sincerely thankful to professor Ioannis Stratis for his continuous assistance and encouragements to keep following my goal, since I was an undergraduate student.

I would like to thank all my colleagues for their friendhip and enjoyable and memorable time we had together in the Laboratory of Analytical Chemistry, and especially my colleague and close friend, Reza Aalizadeh.

None of what I have achieved would have become true without the encouragement of my family and friends. I am deeply grateful to you for your support and encouragements. To my parents, grandfather and grandmother and my aunt who raised me up and taught me to be hardworking, honest and patient.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

The experimental part of this thesis was performed at the Laboratory of Analytical Chemistry, Department of Chemistry, University of Athens, Greece, supervised by Dr. Nikolaos Thomaidis.

The olive oil samples of the first experimental part (Chapter 3) were provided by the International Olive Oil Council and the Greek agricultural organization ELGO-DIMITRA I.O.S.V. on Lesvos. The olive oil samples of the second experimental part (Chapter 4) were collected by Michalis Pentogennis (Chemist). Finally, the olive oil samples used in the third experimental part (Chapter 5) were acquired by local producers.

This thesis is accompanied by Electronic Supplementary Material (ESM); two Microsoft office document files (ESM A and ESM B) and three excel files (ESM I, ESM II, ESM III).

CHAPTER 1

OLIVE OIL AUTHENTICITY STUDIES

1.1 Introduction

Food authenticity involves the confirmation of the stated speciation as true [1]. The authenticity is bound to the truthfulness and, therefore, food is considered authentic (or genuine) when it is not affected by any fraud [2]. The concept of authentic, applied to food, certifies that the product is in accordance with standards and force rules and with the inscriptions of the presentation label [3]. Determining the authenticity of foods can prevent false description, substitution of cheaper ingredients, and adulteration, as well as incorrect origin labeling [1]. The term authentication (authenticity test) used in food control describes the confirmation of all requirements regarding the legal product description or the detection of false or fraudulent statements [4-6].

Labelling legislation is there to ensure that food is properly described. It seeks to protect the consumer from being sold an inferior product with a false description, in addition to protecting honest traders from unfair competition. Enforcement of this legislation ensures that correctly described products remain available to the consumer and that consumer confidence is maintained, which in turn ensures a market place for these foods. Thus, the availability of analytical methods which can ensure the authenticity of foods plays a fundamental role in the operation of modern society [7].

Analysis of foods is continuously requesting the development of more robust, efficient, sensitive, and cost-effective analytical methodologies to guarantee the safety, quality, and traceability of foods in compliance with legislation and consumers' demands. The old methods used at the beginning of the 20th century based on the so-called "wet chemistry" have evolved into the current powerful instrumental techniques used in food laboratories. Besides, currently, there is also a huge interest in the health-related properties of foods as a result of an increasing public concern on how to improve health, through the so called functional foods, functional ingredients, and nutraceuticals. This improvement has

led to significant enhancements in analytical accuracy, precision, detection limits, and sample throughput, thereby expanding the practical range of food applications. Thus, there is no doubt on the importance and current need of analytical techniques developments that are able to face all these demands.

Recent developments and applications of modern instrumental analytical techniques applied advanced “omics” technologies in the fields of food science and nutrition. The aims and strategies of the metabolomics applied on foods have resulted in the proposal of a new eclectic “omics” discipline, formally baptized as “foodomics”, which is an innovative approach in the field of food authentication. Foodomics was defined for the first time in 2009 as “a new discipline that studies the food and nutrition domains through the application of advanced omics technologies to improve consumer’s well-being, health and confidence [8]. The concept of foodomics, more than a simple term covering the panel of the “omic” sciences, also includes the nutritional aspects. Foodomics allows the simultaneous characterization of large numbers of compounds in food matrices, offering to food and nutrition scientists the opportunity to acquire a far more detailed and comprehensive molecular picture of food composition [1].

Extra Virgin Olive Oil (EVOO) and Virgin Olive Oil (VOO), the emblematic food of the Mediterranean diet, is now one of the foods which are being more studied, since the characterization of its composition and quality has a great interest. Recognized for its various nutritional virtues and the beneficial health effects of several of its compounds, VOO is considered, in many countries, as a basic ingredient for a well-balanced nutrition that promotes vitality, well-being, and protects from many diseases. For a very long time, the need of characterizing the VOO composition and checking its genuineness has motivated the development of several analytical methods. These methods have allowed, besides getting a more holistic overview about VOO composition, the identification and the characterization of several of the bioactive compounds found in this interesting matrix.

Chemometrics is frequently coupled to analytical methods to improve the quality and reliability of the conclusions from experimental data (profile or fingerprints). The application of data fusion methods involve the generation of models from the combination (fusion) of outputs of different analytical methods. The main goal of

the data fusion is to enhance the synergy between the fused data set by using complementary inputs on order, to obtain better performance in the food authentication process. In fact, chemometrics can be implemented all along the different steps of EVOOs/VOOs production: raw material input control, monitoring during process, and quality control of the final product.

In the case of olive oil, the objective is to develop High Resolution Mass Spectrometric (HRMS) methods which in combination with chemometrics will reveal markers responsible for the organoleptic profile, the method production, the variety and the geographical origin. In addition, food authentication will critically depend on the establishment of databases that contain comprehensive and standardized information about the olive oil profile.

1.2 Olive oil authenticity

The authenticity of products labelled as olive oil constitutes an important issue from both commercial and health aspects. Olive oil has gained its popularity because of its quality, its health benefits derived from its consumption, and its strict purity control. But it is due to its reputation as a healthy and delectable oil and its high price, that makes it a preferred target for fraudsters. Adulteration may take place not only by accidental contamination during the stages of olive oil processing, but even more often by deliberate mislabeling of less expensive olive oil categories or admixtures, containing less expensive olive oils for the purpose of financial gain.

The olive tree (*Olea Europaea L.*) has diverged naturally to many cultivars and is cultivated mainly in the Mediterranean region; Spain, Italy, Greece, Tunisia, Turkey, Morocco and Algeria [9]. The cultivar defines the quality of the drupe and the olive oil [10]. Currently, 75% of the global production of olive oil takes place in the Mediterranean basin, mainly in Spain, Italy and Greece; an average of 22% takes place in the International Olive Oil member countries (Albania, Algeria, Argentina, Croatia, Egypt, Iran, Iraq, Israel, Jordan, Lebanon, Libya, Montenegro, Morocco, Syria, Tunisia and Turkey) and 2% in the rest of the world [11].

The number of Greek cultivars is greater than 40 and more than 90% of the territory is cultivated with 20 cultivars: Agouromanakolia, Adramytiani, Amigdalolia, Asprolia, Valanolia, Vasilikada, Gaidurelia, Dafnelia, Thiaki, Kalamon, Kalokerida, Karolia, Karidolia, Kothreiki, Kolimpada, Konservolia, Koroneiki, Koutsourelia, Lianolia Kerkiras, Mastoeidis (referred also as Athinolia or Tsounati), Mavrelia, Megaritiki, Mittolia, Strogilolia, Throumbolia, and Tragolia. [12]. Olive oil produced in Greece has excellent quality and this is because of the local climatic and soil conditions. According to the International Olive Oil Council (IOC) [13], 70% of the Greek production is categorized as EVOO, while almost the 35% is exported. Thus, it is imperative for Greece to characterize and authenticate EVOOs based on cultivar and geographical origin, in an effort to establish a brand name in the international market.

The IOC is the intergovernmental organization responsible for administering the International Agreement on Olive Oil and Table Olives, which has been negotiated at United Nations Commodity offices. Oil is the only commodity in the oil sector that has its own international accord. The International Agreement lays down the policy that members should take on the standardization of market for olive oil, olive-pomace oil and table olives. This involves adopting international rules to determine the quality of the products on sale and to monitor international trading. By signing the agreement, members undertake to make whatever arrangements are necessary, in the manner required by their legislation, to ensure the application of a number of principles and measures. These refer to the designations (grade names) and definitions of olive oils and olive-pomace oils, the use of the designation olive oil, geographical indications, and the designations and definitions of table olives. The application of these principles and provisions is compulsory in international trade and recommended in domestic trade. The members undertake to prohibit the use in their territories, for the purposes of international trade, of any designations that run counter to these principles. The IOC takes whatever measures it considers necessary to curb unfair international competition. These principles and provisions are embodied in two standards: the trade standard applying to olive oils and olive-pomace oils and the trade standard applying to table olives. A list follows including the

designations and definitions of olive oils and olive-pomace oils according to the IOC [14]:

Olive oil is the oil obtained solely from the fruit of the olive tree (*Olea europaea L.*), to the exclusion of oils obtained using solvents or re-esterification processes and of any mixture with oils of other kinds. It is marketed in accordance with the following designations and definitions:

- VOO is the oil obtained from the fruit of the olive tree solely by mechanical or other physical means under conditions, particularly thermal conditions, which do not lead to alterations in the oil, and which has not undergone any treatment other than washing, decantation, centrifugation and fi ltration. Virgin olive oil fits for consumption as is includes the following:

- EVOO has a free acidity, expressed as oleic acid, of not more than 0.8 g per 100 g and whose organoleptic characteristics correspond to those fixed for this category in the trade standard.

- VOO has a free acidity, expressed as oleic acid, of not more than 2 g per 100 g and whose organoleptic characteristics correspond to those fixed for this category in the trade standard.

- Ordinary virgin olive oil has a free acidity, expressed as oleic acid, of not more than 3.3 g per 100 g, and its organoleptic characteristics correspond to those fixed for this category in the trade standard.

- Lampante olive oil is the the oil that does not fit for consumption. It has free acidity, expressed as oleic acid, of more than 3.3 g per 100 g. It is intended for refining or for technical purposes.

- Refined olive oil is the olive oil obtained by refining methods which do not lead to alterations in the initial glyceridic structure.

- Ordinary olive oil is the oil consisting of a blend of both refined olive oil and VOO.

- Olive pomace oil is the general term for the oil obtained by treating olive pomace with solvents. Olive pomace is the residual paste left over from the production of VOO. It still contains a variable proportion of oil (5–10 %) depending on the system used to produce the VOO.

The IOC lays down minimum purity and quality criteria for each grade of olive oil and olive-pomace oil that is traded. Due to the great number of different olive varieties cultivated in different geographical regions all over the world, EVOOs and VOOs are permitted to market under a Protected Designation of Origin (PDO), Protected Geographical Indication (PGI), or Traditional Speciality Guaranteed (TSG) label, on the basis of its area and method of production [(EU) No 1151/2012] [15]. According to the European Union (EU) definition, PDO products are most closely linked to the territory. PDO products must be produced, processed and prepared in a specific region, using traditional production methods. The raw materials must also be from the defined area whose name the product bears. The quality or characteristics of the product must be due, essentially or exclusively, to its place of origin; climate, the nature of the soil and local know-how. Food products with a PGI status must have a geographical link in at least one of the stages of production, processing or preparation. The European Commission has already registered in the "Register of protected designations of origin and protected geographical indications" 16 PDO and 11 PGI olive oils, produced in Greece. The list of these oils and their geographical origin are presented in **Figure 1.1**.

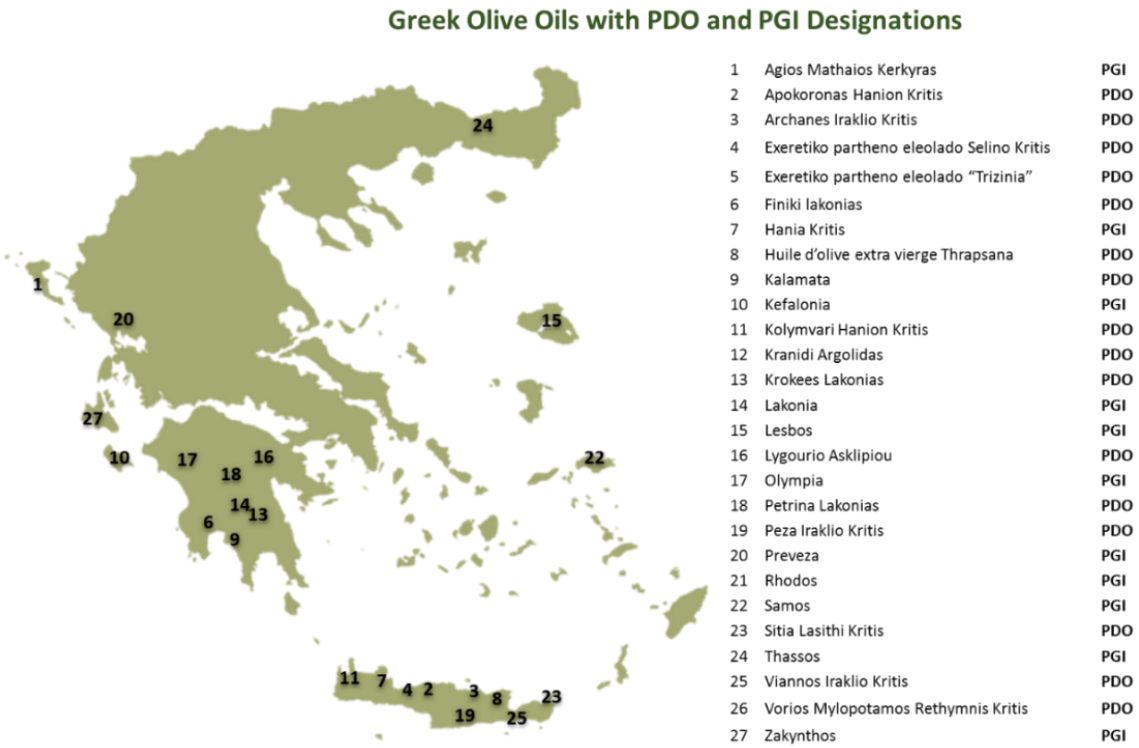


Figure 1.1: Olive Oils with PDO and PGI designations produced in Greece

The EU has established new labeling rules that make origin labeling compulsory for virgin and extra virgin labelled olive oil [Comission Regulation (EC) No 182/2009] [16]. Thus, oil produced from olives from just one EU Member State or third country has to be labeled with the name of the country of origin. VOO produced from olives from more than EU Member State has to be labeled as a “blend of Community olive oils”, while oil produced using olives from outside the EU would be labeled as a “blend of non-Community olive oils” or “blend of Community and non-Community olive oils”, or a reference to the EU and/or third countries of origin.

Consumers are cautious about the nutrition and health claims provided on food labeling. To increase confidence in the market and ensure high level of consumer’s protection, the European Food Safety Authority (EFSA) works for the approval of clear, accurate and corroborated nutrition and health claims as well as any other type of food labelling. Recently, a health claim on olive oil polyphenols was made stating that olive oil polyphenols contribute to the production of blood lipids from oxidative stress. The claim may be used for olive oil which contains at least 5 mg of hydroxytyrosol and its derivativies. The Regulation (EC) no 1924/2006 on “nutrition and health claims made on foods” [17], states that a health claim is any statement about a relationship between food and health. Under article 13(3) of the above regulation, a list of health claims, “other than those referred to the reduction of disease risk”, was provisioned and realised by the subsequent regulation Reg (EC) no 432/2012 [18]. The latter contains the list of these claims, among which the general function one on olive oil polyphenols. The wording for this claim is that “olive oil polyphenols contribute to the protection of blood lipids” whereas it may be used “only for olive oil, which contains at least 5 mg of hydroxytyrosol and its derivatives (e.g. oleuropein complex and tyrosol) per 20 g of olive oil”. The presentation stresses on analytical problems related to the accurate measurement of the bioactive phenols.

It is evident that the development of simple and reproducible analytical methods are urgently needed to guarantee the authenticity and traceability of PDO and PGI olive oils, as well as the country of provenance [3]. Analytical chemistry can play a significant role in food industry, by supporting antifraud authorities that

deal with authenticity and traceability issues, by helping prevent illicit practices, protecting the consumers and, finally, avoiding unfair competition.

1.3 Olive oil bioactive constituents

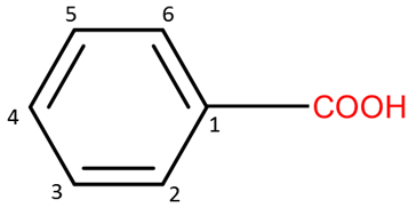
Among other fat sources, olive oil is unique due to its characteristic phenolic composition. VOO is made up of triaglycerides (more than 98%) and minor components (about 1-2%) such as squalene, a-tocopherol, phytosterols, phenolic compounds, carotenoids, aliphatic and terpenic alcohols which constitute the unsaponifiable fraction of olive oil. Phenolic compounds are bioactive compounds that are present in minor concentrations, usually in the range of milligrams per kilogram (parts per million). It is important to note that only EVOOs and some VOOs (and not refined olive oil) contain minor compounds, since most of them are removed during the refining process [19]. Phenolic compounds are important because they are natural antioxidants and contribute to the quality and organoleptic features of the oil. The concentrations of the phenolic compounds and their profile change during fruit maturity, and this is a determinant for the biological properties of the fruit and the oil produced [20-22]. Phenolic compounds have potent antioxidant activity and contribute significantly to the stability of EVOOs and VOOs against oxidation [23-26].

1.3.1 Classification of phenolic and other bioactive compounds in virgin olive oil

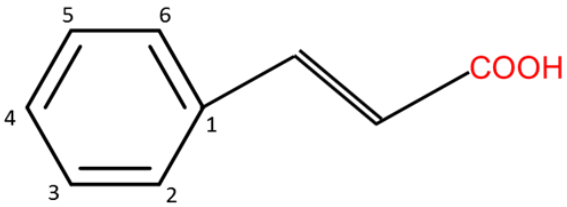
Olive oil polar phenolic fraction contains a complex mixture of compounds with different chemical structures, and it is obtained by the oil with extraction. Phenols present in EVOOs and VOOs are characterized by a chemical diversity of families such as phenolic acids, flavonoids, secoiridoids and lignans [27].

1.3.1.1 Phenolic acids

Phenolic acids were the first series of phenols determined in EVOOs and VOOs [27]. This class can be divided into two main groups: benzoic acids (with the basic chemical structures of C_6-C_1) such as gallic acid, vanillic acid, protocatechuic acid, syringic acid, hydroxybenzoic acid and cinnamic acids (with the basic chemical structures of C_6-C_3) such as caffeic acid, coumaric acid, ferulic acid, cinnamic acid and sinapic acid. The structures of the main phenolic acids identified in olive oils and olive leaves are illustrated in **Figure 1.2.**



Benzoic acid
3-Hydroxybenzoic acid: 3-OH
p-Hydroxybenzoic acid: 4-OH
3,4-Dihydroxybenzoic acid: 3,4-OH
Gentistic acid: 2,5-OH
Vanillic acid: 3-OCH_3, 4-OH
Syringic acid: 3,5-OCH_3, 4-OH
Gallic acid:3,4,5-OH



Cinnamic acid
o-Coumaric acid: 2-OH
p-Coumaric acid: 4-OH
Caffeic acid: 3,4-OH
Ferulic acid: 3-OCH_3, 4-OH
Sinapic acid: 3,5-OCH_3, 4-OH
Homovanillic acid: 5-OCH_3, 4-OH

Figure 1.2: Main phenolic acids identified in olive matrices

1.3.1.2 Phenyl alcohols

The most abundant phenolic compounds in olive oils are the phenethyl alcohols hydroxytyrosol and tyrosol (**Figure 1.3**), formed from the hydrolysis of the secoiridoid aglycones of oleuropein and lingstroside, respectively. The glycosylated forms of hydroxytyrosol and tyrosol were identified in addition only in olive druped and olive leaves [28]. Isoacteoside and campneoside are phenyl alcohols and caffeic acid sugar esters containing hydroxytyrosol or 3,4-dihydroxy-phenyl glycol moieties that have been reported in olive leaves [29].

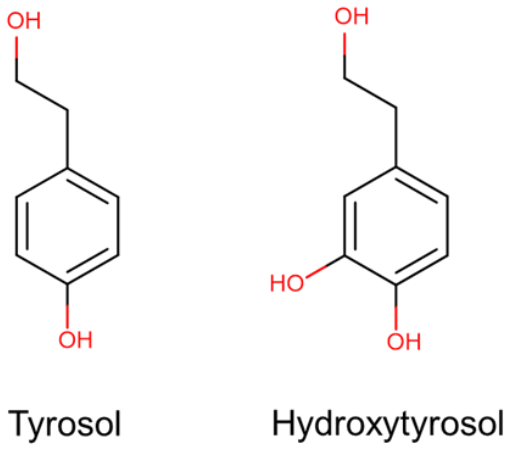


Figure 1.3: Main phenethyl alcohols identified in olive matrices

1.3.1.3 Secoiridoids

Secoiridoids are the most abundant class present in EVOOs and VOOs [22]. They are produced from the secondary metabolism of terpenes. These compounds are characterized by the presence of elenolic acid or its derivatives in its glucosidic or aglyconic form, in their molecular structure. In particular, they are formed from a phenyl ethyl alcohol (hydroxytyrosol or tyrosol), elenolic acid and, eventually, a glucoside residue. Oleuropein is an ester of hydroxytyrosol and the elenolic acid glucoside. Secoiridoids of virgin olive oil in aglyconic forms arise from glycosides in olive fruits by hydrolysis of endogenous β-glucosidases during crashing and malaxation [30]. Oleuropein, lingstroside aglyones and their derivatives are formed during ripening of olive fruits by enzymatic removal of glucose from their respective oleuropein and lingstroside glycosides [28]. In olive leaves and drupes, secoiridoids such as oleuropein, oleuropein diglucoside oleuroside, demethyl oleuropein, 10-hydroxy oleuropein, oleuropein aglycone, deacetoxy oleuropein aglycone, lingstroside, lingstroside aglycone, verbascoside, nuzhenide, oleoside and elenolic acid glucoside were also identified [28]. Verbascoside is the main hydroxycinnamic derivative in olives, and it increases during the fruit maturation. Nuzhenide is a major component of olive seeds. It has not been detected in oil, but it has been reported in paste in small quantities [31]. **Figure 1.4** illustrates the main identified secoiridoids in VOOs, olive leaves and drupes.

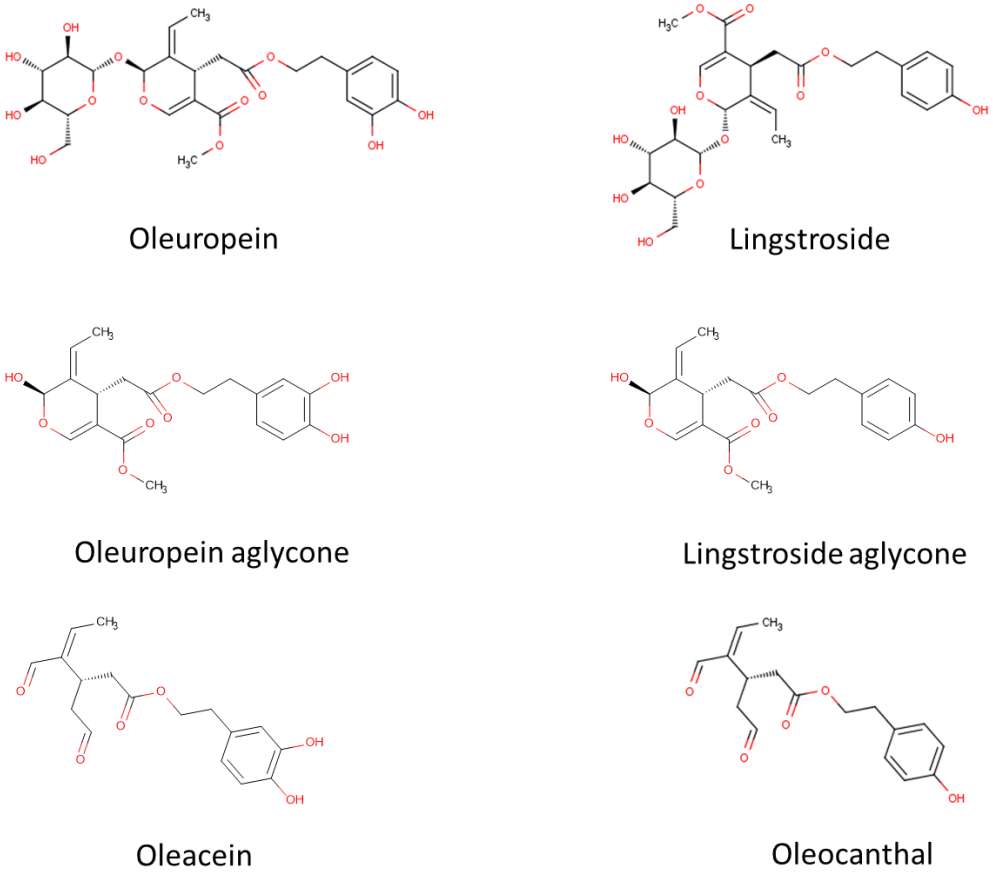


Figure 1.4: Main secoiridoids identified in olive matrices

1.3.1.4 Flavonoids

Flavonoids are largely planar molecules and their structural variation comes in part from the pattern of modification by hydroxylation, methoxylation, prenylation, or glycosylation. Flavonoids represent a small portion of olive phenols, but they contribute to the antioxidant capacity of oil's fraction [28]. Flavonoid glycosides are mainly detected in fruits and paste, whereas their aglycons, apigenin and luteolin, are found in oil. The glycosides are accumulated in fruits and paste and are hydrolyzed to their respective aglycons during production. Flavonoid aglycones are subdivided into flavones, flavonols, flavones, and flavonols depending on the presence of a carbonyl carbon at C-4, an OH group at C-3, a saturated single bond between C-2 and C-3, and a combination of no carbonyl at C-4 with an OH group at C-3, respectively (**Figure 1.5**). Flavonoids such as luteolin and apigenin are reported to be flavonoids present in virgin olive oil. Regarding olive leaves and olive drupes, a significant amount of flavonoids have

been reported. The literature reports several compounds of this class as occurring in olive leaves: luteolin glucoside, luteolin-7-o-rutinoside, apigenin-7-o-glucoside, rutin, quercetin-3-o-glucoside, chyroesiol-7-o-glucoside, taxifolin, quercetin, chryseriol, luteolin, disometin and apigenin.

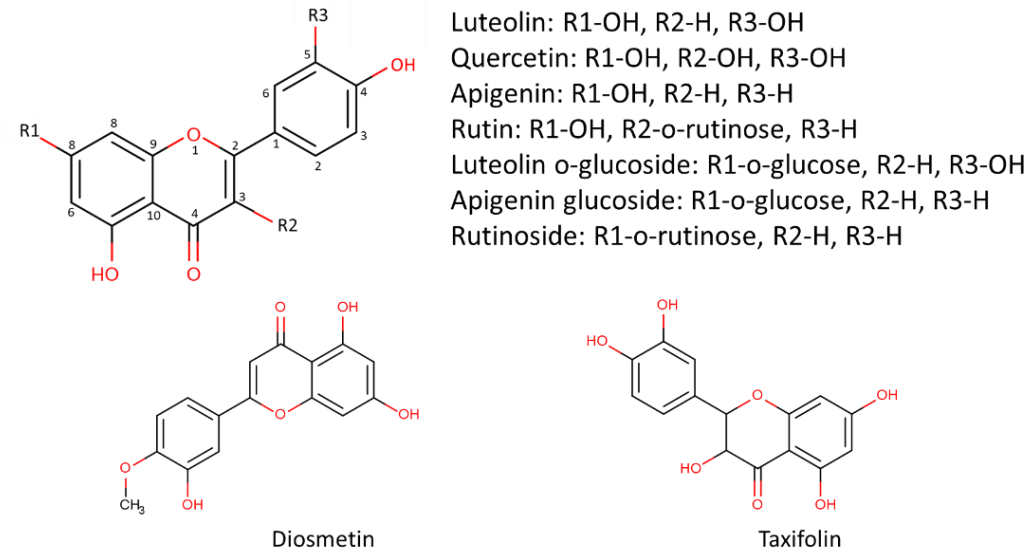


Figure 1.5: Main flavonoids identified in olive matrices

1.3.1.5 Lignans

Lignans are chemically related to lignin. They are prevalent phenolic compounds that originate from C6-C3 units. Pinoresinol, acetoxypinoresinol, hydroxypinoresinol, and syringaresinol are the main lignans identified in VOOs (**Figure 1.6**).

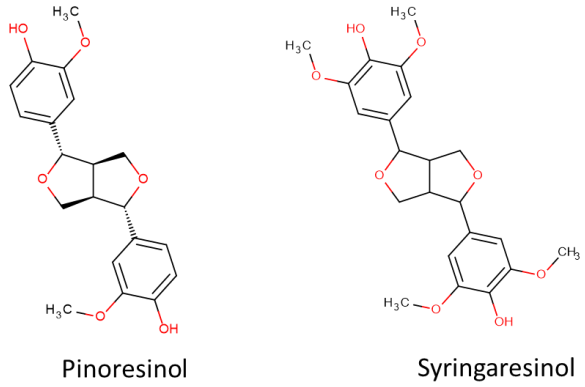


Figure 1.6: Main lignans identified in olive matrices

1.3.1.6 Non phenols

Except for the phenolic compounds, elenolic acid, a nonphenolic compound, and its derivativies and analogues constitute the iridoid part of the most important olive secondary metabolites. Their presence in olive samples indicates complex transformations that take place during olive fruit maturation and processing. Elenolic acid is present in all stages of production procedure. On the contrary, the aldehydic analogue of elenolic acid was only found in the initial stage of production procedure. Several other analogues are found only in drupes and paste; this implies degradation during malaxation or partial removal to the solid waste. The glycosylated demethylated elenolic acid, oleoside and its isomer secologanoside, were detected only in drupes and paste [32].

1.3.2 Nutritional and health benefits

Olive oil is considered a “functional” food because it has satisfactorily demonstrated that it affects beneficially certain functions in the body, beyond adequate nutrition, in a way that improves health and well-being, while it reduces the risk of certain diseases [33]. Its beneficial role in health as an internal ingredient of the Mediterranean diet is universally recognized [34].

Olive oil has been shown to demonstrate favorable health effects in relation to cardiovascular risk factors, mainly lipid profiles, blood pressure, postprandial hyperlipidemia, endothial dysfunction, oxidative stress and antithrombotic profiles. Investigations have unveiled EVOOs potential to reduce the LDL/HDL ratio. The protection against oxidative lipid damage is associated with the phenolic content of the olive oil. In addition, olive oil has hypertension-preventative properties and its use is associated with a decrease in systolic and diastolic blood pressure. Evidence from several studies have showed that the protective effects of EVOOs against chronic diseases such as atherosclerosis, cancer, obesity, diabetes and coronary diseases are related to the phenolic compounds [33].

Experimental and human cellular studies have also provided evidence on VOO and cancer, revealing a positive influence of olive oil on the initiation and

progression of the disease. The phenolic compounds have been shown to demonstrate anticancer and chemopreventive effects against colon cancer, rectum, breast, prostate, endometrium and any kind of gastrointestinal cancer [24].

The phenolic compounds are able to modulate gene expression, influencing protein expression and, subsequently, metabolite production. The phenolic fraction supports anti-oxidant, anti-inflammatory and anti-microbial activities [35]. The mechanisms by which olive oil can support these activities are varied and, most probably, interconnected. They have antibacterial, antifungal and antiviral properties [36, 37]. Due to their anti-oxidant capacity, phenolic compounds act as the first line of defense against free radicals in cellular compartments and extracellularly, as well [38]. Moreover, a decreased risk of developing rheumatoid arthritis has been linked to increased olive oil consumption [35, 38, 39]. Olive oil has also been related to protection against brain disorders and age-related disease. Existing data suggests that olive oil phenols may act as neuroprotective agents, affecting the central nervous system and may protect against Alzheimer's disease [19, 34].

The health benefits of phenolic compounds of EVOOs and VOOs are mainly due to the presence of the unique class of secoiridoids, present in all parts of olive tree [24, 40]. Specifically, hydroxytyrosol is an antioxidant with anti-inflammatory and chemopreventive effects. Tyrosol and oleocanthal also show anti-inflammatory capacity [41], while the latter has been proved that it demonstrates the same anti-inflammatory capacity with ibuprofen [42]. Oleacein and oleuropein are hypotensive and they both show anti-inflammatory properties, as well [12, 37, 43, 44].

1.3.3 Factors affecting phenolic composition

Several endogenous processes and factors interfere with the synthesis of bioactive molecules in olives. The concentration of phenolic compounds in EVOOs/VOOs is strongly affected by agronomical and technological factors, such as olive cultivar [25, 45, 46], place of cultivar [47], climate [48], degree of maturation [45, 49, 50], crop season [51], irrigation [49] and techniques of

processing and storage [47, 52]. The diversity and the interdependence between all these factors make it highly unlikely that these influences would be the same in different regions. Hence, the geographical characterization of EVOOs/VOOs addresses all these agronomic, pedoclimatic and botanical aspects which are unique to the oil of each origin [34, 53]. During maturation, the phenolic profile of olive fruits is significantly modified due to enzymic activity in a manner closely related to cultivar characteristics [54]. The phenolic content of the olive fruit changes as it grows and develops. After six months of growth, the major phenols are the glucosides of lingstroside and oleuropein. As the olives mature these compounds are deglycosylated by glucosidase enzymes to free secoiridoids. There are also variations between the ratios of individual phenols [54], unlike the glucosides, free secoiridoids can only be detected in olive oil. The phenolic content of the oil can be affected during processing, as well. Olives crushing, paste malaxation and olive oil separation affect significantly the qualitative and quantitative phenolic profile of the olive oil [9, 28]. Clashing and malaxation are the most significant steps. In fact, the main hydrophilic phenols such as oleuropein aglycone, originate through this phase by the hydrolysis of oleuropein, demethyl oleuropein and lingstroside, are catalyzed by the endogenous β-glucosidases. During malaxation, the concentration of secoiridoid aglycons and phenolic alcohols diminish in olive pastes and in the related oils, with increasing temperature and processing time [55]. Moreover, storage conditions also affect the phenolic composition and as a consequence, EVOOs after storage register a lower amount of phenols, in comparison to fress olive oils [56].

1.4 Analytical methodologies for the identification of olive oil phenolic compounds

Various analytical techniques and methods investigating the phenolic composition and quality of EVOOs and VOOs are reported in literature. The phenolic compounds are obtained mainly by Liquid Liquid Extraction (LLE) or Solid Phase Extraction (SPE) [57, 58] and, then, further analysis is carried out using different separation and spectrometric techniques. Direct phenol analysis is currently achievable using sophisticated Nuclear Magnetic Resonance (NMR)

[20, 59, 60], Infrared absorption spectroscopy (IR) [61, 62] and Capillary Electrophoresis (CE) [63, 64]. The limited volatility of many phenolic compounds has restricted the use of Gas Chromatography (GC). However, the analysis of these compounds may be accomplished with derivatization of phenolics [65, 66].

Liquid Chromatography (LC) remains the technique of choice for the qualitative and quantitative determination of several phenolic compounds in EVOOs/VOOs [58], coupled to different detectors, such as UV/Vis [41, 67, 68], Diode Array (DAD) [41, 45, 52, 69, 70] or Mass Spectrometers (MS) [57, 70-72]. Low and high molecular weight phenolic compounds can be separated by LC, due to the wide number of possible combinations between the mobile and the stationary phases. However, conventional chromatographic methods used for olive oil analysis, even if they use High Pressure Liquid Chromatography (HPLC), are time consuming and usually very tedious, and the time of analysis usually lasts longer than 30 min. Thus, it is essential to imply a target strategy in order to develop rapid and efficient procedures to perform qualitative and quantitative analysis within reduced analysis times. HPLC started improving its performance when it was coupled to MS as an identification and confirmation technique. Shortening the time of analysis, however, entails other kinds of strategies, such as using short columns and high flow rates or enlarging the choice of column packages. In this context, the use of Ultra High Pressure Liquid Chromatography (UHPLC) is gaining more interest over conventional HPLC, because HPLC column chemistry and dimensions affect resolution and sensitivity; whereas sub-2 μm particle stationary phases combined with UHPLC have helped to produce narrower peaks, rapid analysis times, and lower detection limits. Moreover, the solvent volume consumed is lower, decreasing the volumes of waste. In other words, the improvements of the analytical performance when UHPLC–MS is used are beyond the combination of conventional chromatography with MS, and can be found at different levels: resolution, as well as sensitivity of analysis, by using sub-2 μm particle size, the system operates at higher pressure, and the mobile phase flows at greater linear velocities as compared to conventional HPLC.

As for MS, it combines high sensitivity, selectivity and speed. Furthermore, it has undergone a contemporary revolution with the introduction of ambient MS. This set of techniques is based on ionization methods including Electrospray

Ionızation (ESI), which provides sensitivity, robustness, and reliability [73]. Atmospheric Pressure Chemical Ionization (APCI) is another possible API source that can be used to identify olive oil bioactive constituents [74].

The evolution of accurate High Resolution Mass Spectrometry (HRMS), coupled to GC or LC, has initiated a new trend in olive oil fingerpinting. LC-HRMS allows the identification of a wide range of compounds with different polarities. Mass analyzers commonly employed are Time-of-flight (TOF) [9, 25, 33, 75], QTOF [72, 76], Orbitrap [77], Ion Trap (IT) [71, 78], or hybrid mass analyzers; such as triple quadrupole (QqQ) [57], QTOF, Q-Orbitrap, and linear IT-Orbitrap [31, 79], permitting the generation of fragmentation patterns by MS/MS and MS^n experiments. LC-HRMS combined with systematic workflows allows reliable target analysis with reference standards, a screening for suspect analytes without reference standards, and screening for unknowns [80]. **Table 1.1** summarizes the most recent HRMS analytical methodologies developed the past 10 years for the identification of phenolic compounds in EVOOs and VOOs.

A reliable identification requires both high resolving power and high mass spectral accuracy to increase selectivity against the matrix background and a correct molecular formula assignment to unknown compounds. It can provide powerful untargeted approaches based on advanced scanning techniques, like data-dependent or data-independent acquisitions. Moreover, for the identification and structure elucidation of unknown compounds within a reasonable time frame and with a reasonable soundness, advanced automated software solutions as well as improved prediction systems for theoretical fragmentation patterns, retention times, and ionization behavior are needed [81].

Table 1.1: Application of HR-MS methodologies in the identification of phenolic compounds in olive oils

Determination	Stationary Phase	Extraction	Mobile Phases	Time of analysis	Detection systerm	Ref.
19 phenols in varieties Arbequina, Picual, Hojiblanca	Zorbax C18 (150mm x 10 mm, 1.8 µm)	SPE with diol catridges	A: H_2O + 0.5% acetic acid, B: ACN	20 min	DAD TOF-MS	[75]
23 phenols of Tunisian Chemlali variety	Zorbax C18 (150mm x 10 mm, 1.8 µm)	LLE (MeOH:H_2O 60:40, v/v)	A: H_2O + 0.25% acetic acid, B: MeOH	27 min	DAD TOF-MS	[44]
25 phenols in 54 monovarietal EVOOs (18 Spain, 15 Italy, 10 Portugal, 4 Israel, 3 Greece, 2 California, 1 France, and 1 Turkey)	Luna PFP (150mm x 2.0 mm, 3µm)	LLE (MeOH:H_2O 80:20, v/v)	A: H_2O + 0.1% acetic acid, B: ACN; 0.1% CH_2O_2	16 min	Q-TOF-MS	[76]
20 secoiridoids in Picual, Erbequina, Frantoio and Coratina EVOOs	Halo C18 Fused-Core (100mm x 2.1 mm, 2.7 µm)	LLE (MeOH:H_2O 60:40, v/v)	A: H_2O + 0.1% formic acid, B: MeOH; 0.1% formic acid	28 min	Exactive HCD Orbitrap and Q-Exactive hybrid Orbitrap MS	[77]

Determination	Stationary Phase	Extraction	Mobile Phases	Time of analysis	Detection systerm	Ref.
20 phenols in Spanish EVOOs	Zorbax Eclipse Plus C18 (150 mm x 4.6 mm, 1.8 μm)	SPE with diol catridges	A: H_2O + 0.25% acetic acid, B: MeOH	27 min	Micro TOF-MS	[33]
28 phenos in Picholine EVOOs	Zorbax C18 (4.6 x 150 mm, 1.8 μm)	LLE (MeOH:H_2O 60:40, v/v)	A: H_2O + 0.5 % acetic acid, B: ACN	25 min	DAD IT-MS TOF-MS	[82]
18 phenos in 32 Arbequina EVOOs	Zorbax Eclipse Plus RP-C18 (150 mm x 4.6 mm 1.8 μm)	SPE with diol catridges	A: H_2O + 0.25 % acetic acid, B: MeOH	27 min	TOF-MS	[9]
19 phenols in 25 Arbequina, Koroneiki, Arbisana,Grappolo, Manzanilla, Coratina, Frantoio and MGS Mariense VOOs	Zorbax Eclipse Plus RP-C18 (150 mm x 4.6 mm, 1.8 μm)	LLE (MeOH:H_2O 60:40, v/v)	A: H_2O + 0.25 % acetic acid, B: MeOH	27 min	TOF-MS	[23]
22 phenols in 7 Tunisian VOOs	Zorbax C18 (150 mm x 4.6 mm, 1.8 μm)	LLE (MeOH:H2O	A: H_2O + 0.25 % acetic acid, B:	20 min	DAD MicroTOF-MS	[25]

Determination	Stationary Phase	Extraction	Mobile Phases	Time of analysis	Detection systerm	Ref.
		60:40, v/v)	ACN			
30 phenols in 1EVOO mix of the varieties Leccino, Frantoio, Carboncella	Zorbax Eclipse Plus RP-C18 (100 mm x 2.1 mm, 1.8 µm)	LLE (MeOH:H_2O 70:30, v/v) SPE with diol-modified catridges	A: H_2O + 0.1 % formic acid, B: ACN with 0.1 % formic acid	16 min	DAD QTOF-MS	[83]
13 phenols in 78 EVOOs from Argentina (5), Greece (1), Italy (1), Morocco (4), Peru (32), Portugal (1), Spain (26) and Syria (1)	Agilent Zorbax XDB-C18 (100 mm x 4.6 mm, 1.8 µm)	SPE diol-catridges	A: H_2O with 0.1 % formic acid, B: ACN	30 min	TOF-MS	[84]
11 oleuropein aglycone isomers in Spanish EVOOs	Agilent Zorbax Eclipse Plus (150 mm x 4.6 mm, 1.8 µm)	SPE diol-catridges	A: H_2O + 0.5 % acetic acid, B: ACN	22 min	Micro TOF-MS IT-MS^n	[78]
30 phenols in 25 EVOOS (Arbequina, Cornicabra, Hojiblanca, Picual and	Zorbax C18 (150 mm	SPE diol-	A: H_2O + 0.5 % acetic acid, B:	25 min	TOF-MS	[85]

Determination	Stationary Phase	Extraction	Mobile Phases	Time of analysis	Detection systerm	Ref.
Frantoio)	x 4.6 mm, 1.8 µm)	catridges	ACN			
23 phenols in VOOs Picual and Hojiblanca	Agilent Zorbax Eclipse Plus (150 mm x 4.6 mm, 1.8 µm)	LLE (MeOH:H_2O 60:40, v/v)	A: H_2O + 0.25 % acetic acid, B: MeOH	27 min	TOF-MS	[86]
21 phenols in Azeradj EVOOs	Zorbax C18 (150 mm x 4.6 mm, 1.8 µm)	LLE (MeOH:H_2O 60:40, v/v)	A: H_2O + 0.25 % acetic acid, B: MeOH	27 min	MicroTOF-MS	[87]
40 phenols in Tunisian EVOOs (Arbequina, Chétoui, Neb Jmal and Picholine)	Zorbax Eclipse Plus C18 (4.6 mm × 150 mm, 1.8 µm)	LLE (MeOH:H_2O 60:40, v/v)	A: H_2O with 0.25 % acetic acid, B: MeOH	27 min	TOF-MS	[88]
29 phenols in 10-months stored EVOOs	Zorbax Eclipse Plus C18 (4.6 mm × 150 mm, 1.8 µm)	LLE (MeOH:H_2O 75:25, v/v)	A: H_2O with 0.25 % acetic acid, B: MeOH	27 min	TOF-MS	[89]

EVOOs: Extra virgin olive oils, VOOs: Virgin olive oils, LLE: Liquid liquid extraction, SPE: Solid phase extraction, MeOH: methanol, ACN: Acetonitrile, Ref.: reference

1.4.1 Identification of phenolic compounds with TOF-MS

TOF resolution is directly related to the length of the flight path. Modern high resolution instruments share the characteristics of flight paths with a combined length of several meters. The introduction of a reflectron doubles the flight path and regulates the mobile energy, resulting in higher resolution. Since resolution is related to the length of flight time, TOF provides the highest resolution for relatively high m/z ion masses. Resolving power is defined at full width at full maximum (FWHM) as m/Δm, where m is the m/z and Δm is the width of the mass peak at half peak height. In TOF-MS analysis, accurate and precise mass measurements become possible. Mass measurement uncertainty in terms of mass accuracy (i.e. average mass error) and mass precision (I.e. standard deviation on the mass error) is based on calculating the relative (ppm) or absolute (mDa) difference between the measured accurate mass and the calculated exact mass of an analyte. Both mass accuracy and precision are essential for proper measurements of accurate mass [75, 80].

Hybrid tandem mass instruments, such as the Q-TOF, provide relevant structural information by obtaining product ion full spectra at accurate mass. Q-TOFMS/MS experiments confirm the existence of potential positives and are very useful in permitting the elucidation of unknown compounds. For target analysis, data-independent acquisition (IDA) is most appropriate. This approach, termed MS^E (Waters) or bbCID (Broad Band Collision Induced Dissociation) (Brukers) involves aimultaneous acquisition of accurate mass data at low and high collision energy. By applying low energy in the collision cell, no fragmentation is taking place and the information obtained is actually full scan MS spectrum. At high collision energy, fragmentation of the ions takes place and MS/MS spectra are acquired. With IDA, both molecular and fragment ion are obtained in a single acquisition without the need of pre-selection of the analytes. Data dependent acquisition is more favorable for suspect and non-target screening, due to the fact that information on specific ions can be collected.

In order to increase mass accuracy, mass calibration is performed. There are three levels of mass calibration, external, internal and lock mass calibration. External and internal calibration must include at least the mass range of interest

and can be performed with the same calibrant mixture. Lock mass calibration provides an automated way of applying the linear correction calibration to each spectrum in the analysis and it requires the presence of a continuous signal.

1.5 HRMS screening workflows

There are various workflows applied in foodomics, depending indispensably on the instrumentation and the available software. **Figure 1.7** presents the flow chart of the screening protocols: target, suspect and non-target.

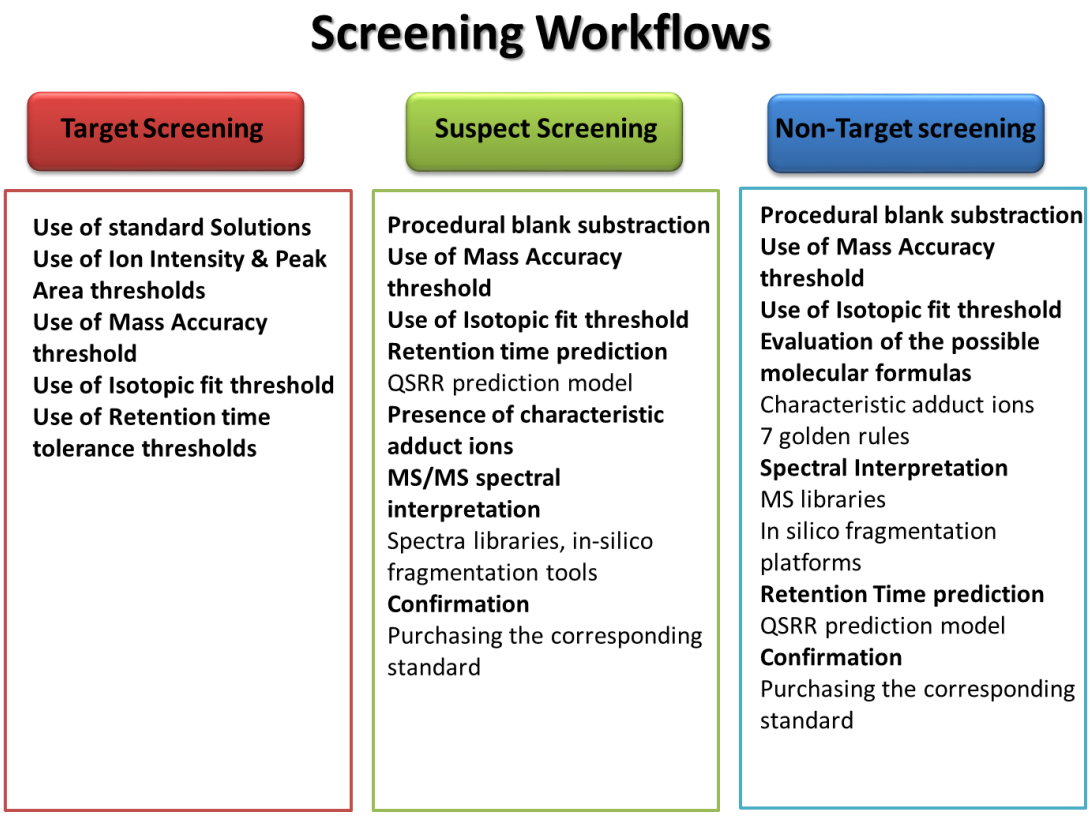


Figure 1.7: Flow chart of screening protocols: target; suspect and non-target

1.5.1 Target screening

Target analysis is based on the determination of already known analytes, and the identification is carried out with the use of standard solutions. A reference standard is required, to compare and match the experimental retention time (t_R) and the MS/MS fragmentation, and determine the concentration of the target

compound in the sample, as well. Target analysis relies on reference standards for confirmation and quantification [90]. For olive oil analysis, a target list is generated from the literature, consisting of significant classes of compounds that have been already identified in olive oil matrix.

1.5.2 Suspect screening

Suspect screening with LC-HRMS relies on accurate mass and isotope information for the precursor ion. Compounds that are expected to exist in the samples, can be screened using the exact mass of their molecular ions in negative ($[M–H]^-$) or positive ($[M+H]^+$) ESI. The exact mass of each suspect compound is extracted from the chromatogram and evaluated. Additional information is needed to reach a tentative identification, apart from the mass accuracy, isotopic fit value and MS/MS spectra interpretation are crucial. For confirmation of the unknown analyte, it is vital to use in silico fragmentation tools; such as metfrag [91], and compare MS/MS information with spectral libraries; such as MassBank [92] or FooDB [93]. Calculating the retention time for the suspects list and comparing it to the observed retention time for the experimental peaks could be an efficient filtering tool over the confirmation or rejection of the suspect compounds [94].

1.5.3 Non-target screening

Non-target screening implies the identification of compounds for which there is no previous knowledge available and is usually carried out after target and suspect screening [81]. The first step is always peak picking. In this step, comparison of the sample with control or blank samples is important to exclude irrelevant peaks. The removal of noise peaks, mass recalibration and componentization of isotopes and adducts is usually carried out automatically, as the next step. The assignment of the molecular formula to accurate mass of the peak is performed using heuristic filters, such as the seven golden rules of Kind and Fiehn [95]. Exploration of online databases, like Chemspider [96] or PubChem [97], may lead to candidate structures. Thereby, information on the parent compound (molecular formula, substructures) can help restrict the search of databases and possible

structures are likely to be proposed for the non-target compound. For ranking the candidate structures, information from MS/MS spectra has to be explored by comparing the fragmentation pattern with in-silico mass spectral fragmentation [91] or with spectral libraries [92]. In a further step, the retention time of the tentative candidates can be predicted with validated computational models based on quantitative structure retention time relationship (QSRR) [94].

1.6 Data processing

The challenge in non-target screening is the export and evaluation of the massive quantities of data generated. In order to facilitate data processing various data processing tools are available, such as XCMS [98] and MZmine [99], among others. To process HRMS data, first, binary files of all the analyzed samples are converted to mzXML files. Then, the files are processed to extract the peaks that are present in the samples. This procedure involves peak picking, grouping of peaks representing the same analyte across the samples, and a step to correct the chromatographic drift of the retention time. The next step is the regroup of the peaks for which the retention time was changed. Finally, a step of filling in the same peaks is implemented. This replaces the missing values of nondetected peaks with a low value of the intensity. Adduct peaks are removed and the internal parameters of the algorithms used for peak picking, grouping and retention time alignment are optimized. This workflow, which is illustrated in **Figure 1.8**, results in the generation of all the existing molecular feautures in the sample.

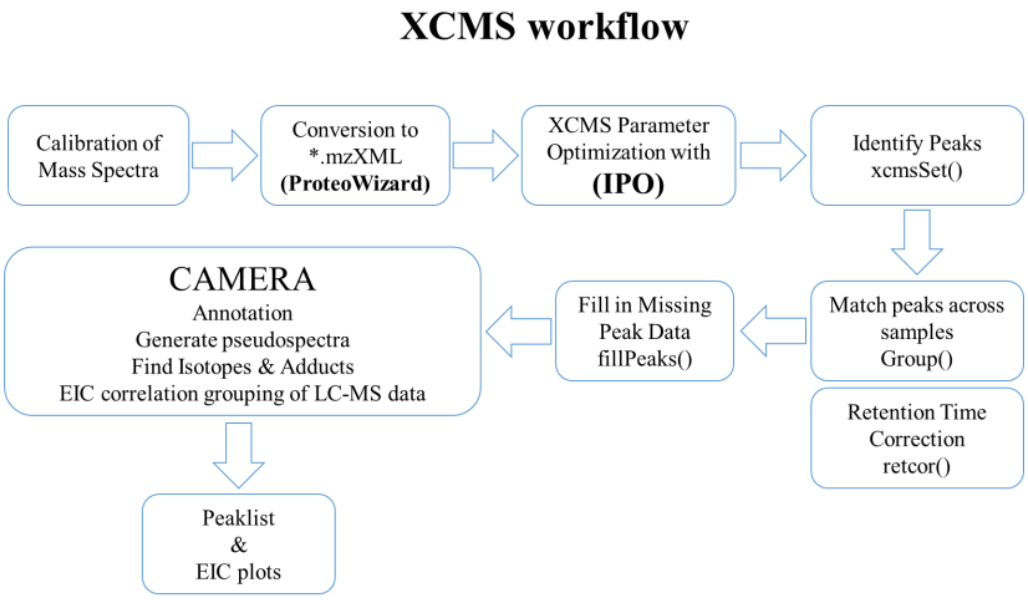


Figure 1.8: Data processing workflow with XCMS

1.6.1 Data treatment for variable reduction

The large number of variables and the associated presence of redundancy, multicollinearity, random noise, adducts or isotopic peaks and chance correlation are common problems when dealing with analytical instrumental dataset (such as LC-HRMS). The presence of irrelevant variables (in the present case the peaks list of LC-HRMS) can change the underlying data patterns and consequently, it can influence results of several multivariate methods. A normal peaks-list created by XCMS requires to be pretreated via unsupervised variable reduction strategy to exclude false positive peaks. Such a strategy refers to the procedure that aims at selecting a subset of variables (m/z), able to preserve as much information of the original peaks-list as possible, but eliminating redundancy, noise and adducts or isotopes, without taking into account their dependent classes or labels. Moreover, unsupervised reduction can facilitate the subsequent supervised selection, which can suffer from the presence of highly correlated data and chance correlation, thus giving overfitted results. WSP method (based on Wootton, Sergent, Phan-Tan-Luu's algorithm) has been developed for space-filling designs of experiments (SFD) [100] and recently has been applied for variable reduction (V-WSP) purposes for treatment of LC-HRMS peaks list [101]. V-WSP algorithm can be adapted in order to select a representative set of variables instead of all peaks list. Variables are chosen in an unsupervised

manner so as to be at a fixed minimal correlation from every variable in the defined multidimensional space. Given a data matrix with ***n*** rows (samples) and ***p*** columns (m/z), the algorithm for calculating the V-WSP method starts as follows:

Step 1: select an initial variable (seed) *j* and a correlation threshold (*thr*);

Step 2: calculate the Pearson linear correlation coefficients (*c*) between *j* and all other variables;

Step 3: eliminate variables *d* such as absolute value of *cdj* ≥ *thr*;

Step 4: variable j is selected and replaced by the variable with the highest absolute correlation value with *j* among the remaining variables;

Step 5: repeat steps 2, 3 and 4 until there are no more variables to select.

1.7 Retention time prediction

Chromatographic retention time prediction can play a fundamental role in increasing identification confidence and eliminating false positive plausible candidates during non-target and suspect screening procedure. The application of many non-target analysis approaches leads to compounds tentatively identified based solely on HRMS data and, in some cases, comparable spectral libraries. An additional and powerful tool to increase the confidence in the tentative identification of compounds, for which standards, are unavailable is reliable and accurate t_R prediction. Having considered the same LC conditions during an analysis, t_R can be interpreted by chemical structure. In order to correlate the molecular descriptors with the retention times of compounds, reliable in silico approaches based on Quantitative Structure Retention Relationships (QSRR) can be used [94]. A generic QSRR model is trained based on set of molecular descriptors, which are well explaining the elution behavior of target compounds, and experimental t_R for several compounds from different classes of emerging contaminants. This is to ensure that the model built is capable of external prediction application on a compound with completely unknown t_R.

For each compound that QSRR model is applied to predict its t_R, it is important to check whether the prediction is within the applicability domain of the model or not. The general idea of the applicability domain of the QSRR model is to define

the acceptance window for prediction results relatively to their chemical structural similarity. For instance, for a compound that has high chemical structure similarity comparing to the training set used to model t_R, the error of (or higher than) ±2 min is a relatively high value, while for a compound with lower chemical structure similarity, this threshold is accepted. Another application aspect is to show if the observed residual is due to the chemical space failure (leading to wrong prediction of t_R) or the structure tested has abnormal elution behavior. Four regions are obvious in the bubble plot, as it is illustrated in **Figure 1.9**, corresponding to different levels of acceptance for the predicted t_R. Boxes 1 (very similar to the training set used to build the model and the error (in terms of standardized residuals) less than 1.0 min) and 2 (the structure is diverse or the observed error is relatively accepted comparing the chemical structural effect and the error (in terms of standardized residuals) less than 2.0 minute) are the areas of acceptance of predicted t_R, while box 3 shows the region where the residuals are high and the predicted t_R is questioned. The last region, box 4, shows the area where the model is not applicable for a given compound if the bubble size is huge otherwise the compound is false positive (very small bubble size) and it should not be corresponded to the given t_R. Therefore, the bubble size indicates the chemical structure diversity; the larger it gets, the larger the chemical structure diversity becomes. Moreover, normalized mean distance shows if the used training set is representative for the compound tested.

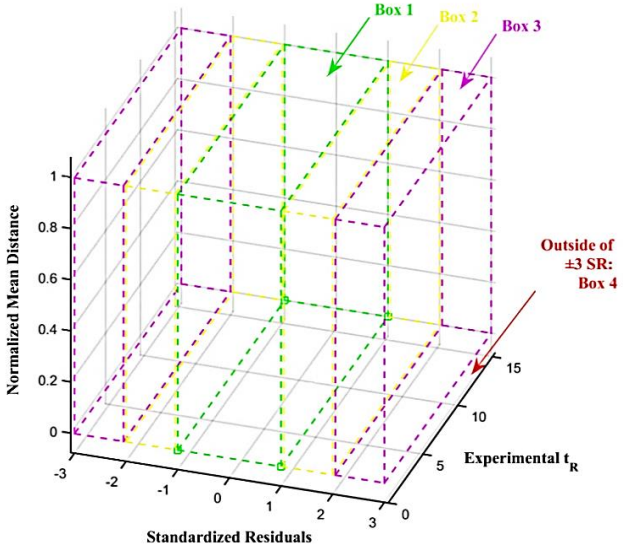


Figure 1.9: Representation of OTrAMS with its possible four boxes for predicted t_R

1.8 Data Mining

Chemometric tools are widely applied for data mining. Chemometrics could be defined as the chemical discipline that uses mathematics, statistics and formal logic to design or select optimal experimental procedures, to provide maximum revelant chemical information by analyzing chemical data, and to obtain knowledge about chemical systems [102]. The application of chemometrics is fundamental for food authentication studies since huge amounts of data must be handled.

Classification methods are major multivariate techniques aimed to find mathematical models that are able to recognize the membership of each sample to its proper class, on the basis of a set of measurements. Once a classification model has been obtained, the membership of unknown samples to one of the defined classes can be estimated. In other words, classification models find mathematical relationships between a set of descriptive variables derived from chemical measurements and a qualitative variable (such as the membership to a defined category, geographical origin, labels etc.). The method used to build the mathematical relationship can be categorized into two groups: supervised classification (like PLS-DA or LDA)) and unsupervised classification methods (PCA and clustering technique).

1.8.1 Principal Component Analysis (PCA)

PCA is a well-known chemometric technique for exploratory data analysis; it basically projects data in a reduced hyperspace, defined by orthogonal principal components [103]. These are linear combinations of the original variables, with the first principal component having the largest variance, the second principal component having the second largest variance, and so on. Thus, it is possible to select a number of significant components, so that data dimension is reduced by preserving the systematic variation in the data retained in the first selected components, while noise is excluded, being represented in the last components. Therefore, PCA enhances and facilitates data exploration and interpretation of multivariate datasets. PCA is probably the most popular multivariate statistical technique and it is widely used for handling LC-HRMS data. Through LC-HRMS,

a peak list of thousand masses can be extracted, with their intensities varying across several samples analyzed. PCA analyzes the peak list generated by LC-HRMS and its goal is to extract the important information from the peak list (loading plot) and to express this information as a set of new orthogonal variables, called Principal Components (PCs). PCA also represents the pattern of similarity of the observations and the exact masses by displaying them as points in maps (score plot) [103].

1.8.1.1 Confidence Intervals and Performance of PCA

The performance of the PCA model is evaluated using computer-based resampling techniques, such as the bootstrap and cross-validation techniques, where data is separated into a learning and a testing set. A popular cross-validation technique is the jackknife (*aka* 'leave one out' procedure). In the jackknife, each observation is dropped from the set in turn and the remaining observations constitute the learning set. The learning set is, then, used to estimate the left-out observation which constitutes the testing set. Using this procedure, each observation is estimated according to a random effect model. The predicted observations are, then, stored in a matrix denoted as $\tilde{\boldsymbol{X}}$. The overall quality of the PCA random effect model, using *M* components, is evaluated as the similarity between $\tilde{\boldsymbol{X}}$ and **X** [*M*] [103].

1.8.1.2 Covariance error ellipse

Confidence ellipses for a PCA obtained from peaks-list can be derived from covariance matrix and square-root of the corresponding eigenvalues, or based on Hotelling's T^2 method. To use these two methods, the scores, given by PCA, should be approximately normally distributed with enough samples coming from the same population. Having a prior knowledge of the subgroups of the samples is a plus to define confidence intervals more accurately, as the mean and covariance would be much close to the real scores within each subgroup [103]. As shown in **Figure 1.10**, prior knowledge of subgroups could define accurate confidence intervals to the PCA, and could represent the samples that are

outliers (Samples 41, 5 and 16); where the eclipse confidence interval based on unknown subgroups is quite misleading and cannot detect sample 41 as outlier.

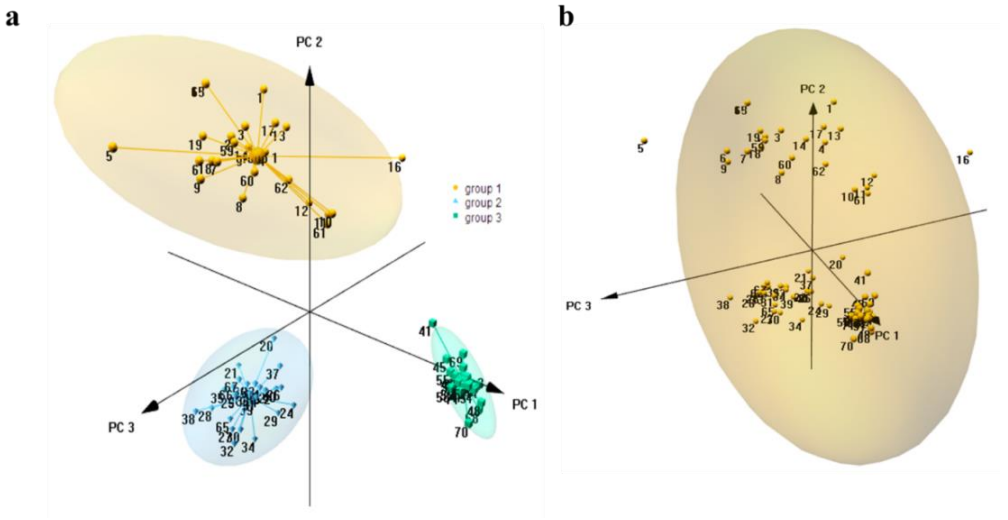


Figure 1.10: Confidence intervals (95%) derived for; (a) known and (b) unknown sub-groups.

1.8.2 Linear Discriminant Analysis (LDA)

Among traditional classifiers, Discriminant Analysis (DA) is the most known method and can be considered as the first multivariate classification technique [104]. The method is a probabilistic parametric classification technique: it maximizes the variance between categories and minimizes the variance within categories, by means of data projection from a high dimensional space to a low dimensional space. In this way, a number of orthogonal linear discriminant functions equal to the number of categories minus one is obtained. **Figure 1.11** shows the general concept of DA. Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) are used in turn, depending on the linear/non-linear class separability and on the reliability of the class covariance matrices. In fact, for LDA only the pooled covariance matrix is calculated, while for QDA the covariance matrix is calculated for each class separately [105]. QDA is a probabilistic parametric classification technique and separates the class regions by quadratic boundaries. It makes the assumption that each class has a multivariate normal distribution, while the dispersion is different in the classes. In order to estimate the class covariance matrix, the number of class objects must

be greater than the number of variables, while LDA can be applied only if the total number of the samples is greater than the number of variables.

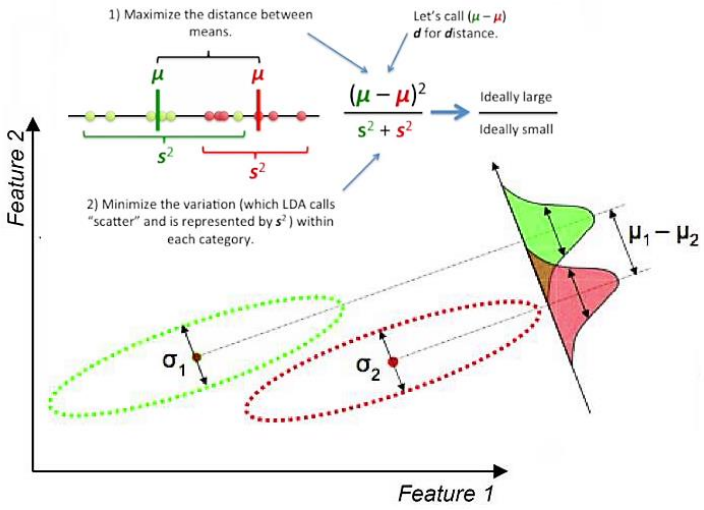


Figure 1.11: General concept of Discriminant Analysis

1.8.2.1 Partial Least Square Discriminant Analysis (PLS-DA)

Partial Least Squares Discriminant Analysis (PLS-DA) is a linear classification method that combines the properties of partial least squares regression with the discrimination power of a classification technique [106, 107]. PLS-DA is based on the PLS regression algorithm (PLS1 when dealing with one dependent Y variable and PLS2 in the presence of several dependent Y variables), which searches for latent variables with a maximum covariance with the Y-variables. The main advantage of PLS-DA is that the relevant sources of data variability are modelled by the so-called Latent Variables (LVs), which are linear combinations of the original variables (LC-HRMS peaks-list), and, consequently, it allows graphical visualization and understanding of the different data patterns and relations by LV scores and loadings. Loadings are the coefficients of variables in the linear combinations which determine the LVs and therefore, they can be interpreted as the influence of each m/z on each LV, while scores represent the coordinates of samples in the LV projection hyperspace [108]. The optimal number of LVs is usually selected by means of cross validation procedures, by choosing the latent variables which minimize the cross validation error in classification. A valid PLS-DA model will return the predicted class as a vector of size G for each sample, with values in-between 0 and 1: a g-th value closer to zero indicates that the sample does not belong to the g-th class, while a value closer to one the

opposite. Since predicted vectors will not have the form (0,0,…,1,…0), but real values in the range between 0 and 1, a classification rule must be applied. The sample can be assigned to the class with the maximum value in the Y vector or, alternatively, a threshold between zero and one can be determined for each class on the basis of the Bayes theorem (the class threshold is selected at the point where the number of false positives and false negatives is minimized).

1.8.3 Classification trees (CT)

Tree-based approaches [109] consist of algorithms based on rule induction that is a way of partitioning the data space into different class subspaces. Basically, the data set is recursively split into smaller subsets where each subset contains samples belonging to as few categories as possible. In each split (node), the partitioning is performed in such a way so that it reduces entropy (maximize purity) of the new subsets, and the final classification model consists of a collection of nodes (tree) that define the classification rule. The best split can be found using univariate strategies; in the univariate approach the algorithm searches at each binary partitioning the single m/z that gives the purest subsets. The partitioning can be formulated as a binary rule: all the samples that satisfy the rule are grouped in one subset, otherwise into another.

Random Forest (RF) starts with creation of several bootstrapped samples (subsamples) from the original matrix of samples and peaks. This step permits the estimation of the error of the left-out samples. Bagging (bootstrap aggregation) is one of the famous methods based on RF, in which several binary decision trees can be fit to a bootstrapped data (subsampled data) [109, 110]. In RF, we randomly get ***n*** times of replacement of samples into a bootstrap sample to use them as tree seeds. A large portion of these samples is used to form the training set, and the rest of the samples are used as test set; so called out of bag samples (OOB). These samples are used to calculate OOB error which is referred as prediction error of the model [109, 111]. Finally a model with large number of trees and the lowest OOB error can be selected from the different prediction trees with the following variable importance measurement [110]:

$$VI(X^f) = \frac{1}{ntree}\sum_t\left(\widetilde{OOB_t^f}\,error - OOB_t error\right) \quad (1)$$

Where: for each tree (*t*) of a forest, OOB_t is associated with the data, which are not included in the bootstrap samples, to construct *t*. OOB_t*error* is the mean square error (MSE) of a single tree on OOB_t. $\widetilde{OOB_t^f}\,error$ is the error of perturbed sample created by randomly permuting the values of X^f (variable) in OOB_t. RF internally selects features that explain the lower OOB error. To measure the importance of the *j*-th m/z after training, the values of the *j*-th m/z are permuted among the training data, and the out of bag error is again computed on this perturbed peaks list. The importance score for the *j*-th m/z is calculated by averaging the difference in out of bag error before and after the permutation over all trees. The score is normalized by the standard deviation of these differences. Features which produce large values for this score are ranked as more important than features producing small values.

1.8.4 Kohonen Self-Organizing Maps (SOMs)

Kohonen Maps are self-organizing systems applied to the unsupervised/ supervised problems [112, 113]. In Kohonen maps, similar samples are linked to the topological close neurons in the network. Basically, the neurons have as many weights as the number of m/zs in the peak list. The objective is to identify the location of those m/zs that have more similar weights in the Artificial Neural Network (ANN); the weights of the net are updated on the basis of the m/zs, i.e. the network is modified each time an m/zs is introduced for a certain number of times (epochs). An example of the structure of a Kohonen map with dimension 5x5, built for a peaks-list described by *p* m/zs is shown in **Figure 1.12**.

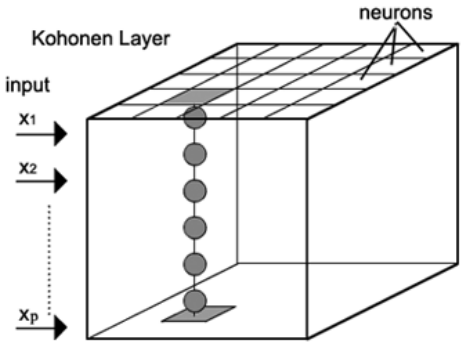


Figure 1.12: Illustration of components of Kohonen Maps

1.8.4.1 Counter Propagation Artificial Neural Networks (CP-ANNs)

CP-ANNs are very similar to the Kohonen Maps and are essentially based on the Kohonen approach, but combine characteristics from both supervised and unsupervised learning [114, 115]. CP-ANNs are able to build a mathematical model able to recognize the membership of each sample to its proper class. To do so, an output layer is added to the Kohonen ANN. CP-ANNs can be considered as extension of Kohonen maps. A CP-ANN consists of two layers, a Kohonen layer and an output layer (also called Grosberg layer) (**Figure 1.13**). When dealing with supervised classification, the class vector is unfolded into a matrix **C**, with *I* rows and *G* columns (the unfolded class information), where *I* is the number of samples and *G* the total number of classes; each entry c_{ig} of **C** represents the membership of the *i*-th sample to the *g*-th class expressed with a binary code (0 or 1). Then, the weights of the r-th neuron in the output layer ($\mathbf{y}_r$) are updated in a supervised manner on the basis of the winning neuron selected in the Kohonen layer. Considering the class of each sample *i*, the update is calculated as follows:

$$\Delta y_r = \eta \cdot \left(1 - \frac{d_r}{d_{max}+1}\right) \cdot (c_i - y_r^{old}) \quad (2)$$

Where d_r is the topological distance between the considered neuron *r* and the winning neuron selected in the Kohonen layer; c_i is the *i*-th row of the unfolded class matrix **C**, that is, a G-dimensional binary vector representing the class membership of the *i*-th sample; η is the learning rate and d_{max} is the size of the considered neighborhood that decreases during the training phase.

At the end of the network training, each neuron of the Kohonen layer can be assigned to a class, on the basis of the output weights. All the samples placed in that neuron are automatically assigned to the corresponding class. As a consequence, CP-ANNs are also able to recognize samples belonging to none of the class spaces. This happens when samples are placed in neurons whose output weights are similar (the neuron cannot be assigned to a specific class).

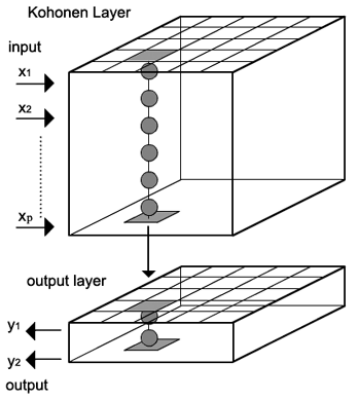


Figure 1.13: Illustration of structure of CP-ANNs

1.8.5 Receiver Operating Characteristics (ROC) curves

ROC curves are graphical tools for the analysis of classification results [116-118]. A ROC curve is a graphical plot of sensitivity and 1-specificity, for a binary classification system as its discrimination threshold is changed. A single value of sensitivity and 1-specificity can be calculated from a contingency table and consequently each contingency table represents a single point in the ROC space [119]. For each threshold value, a classification rule is calculated and the respective contingency table is obtained. The best possible classification method would yield a point in the upper left corner of the ROC space, representing maximum sensitivity and specificity, while a random classification gives points along the diagonal line from the left bottom to the top right corners. Summarizing, ROC curves are calculated for each class, separately, by changing the threshold of assignations. The area under the ROC curve (AUC) can be used as an estimator of the class separability. An example of ROC curve is shown in **Figure 1.14**, where the AUC is lower in class 2 comparing to class 1 and thus, the classification model built has lower separability and accuracy over assignment of class 2 on samples.

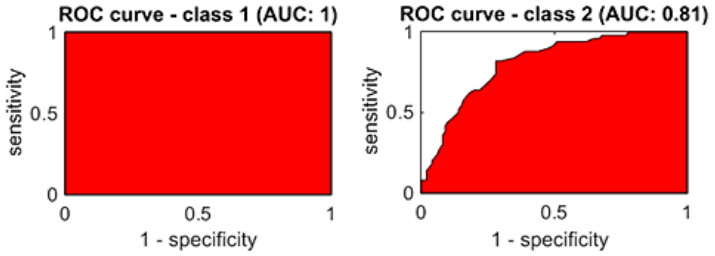


Figure 1.14: Example of ROC curves for set of samples with two classes

1.8.6 Features Prioritization

Knowing the variance explained in each projected dimension, hundred over thousands of m/zs can be prioritized. This can be achieved with the use of a method, such as Variable importance in PLS-DA projections (VIP). The idea behind this measurement is to accumulate the importance of each m/z (***j***) being reflected by loading weights (***w***) from each component [120]. The VIP measurement is defined as:

$$VIP = \sqrt{p \sum_{a=1}^{A} \left[SS_a \left(w/\|w\|^2\right)\right] / \sum_{a=1}^{A} (SS_a)} \quad (3)$$

Where: ***SSa*** is the sum of squares explained by the *a*th component. Hence, the VIP is a measure of the contribution of each m/z according to the variance explained by each PLS-DA component where $(w\boldsymbol{aj}/\|w\boldsymbol{a}\|)^2$ represents the importance of the ***j***th m/z.

1.8.7 Features Selection

ACO is a swarm intelligence algorithm that is based on the behavior of the ants searching for the food resources by their nest, using pheromone deposition without any visual information [121, 122]. It was inspired by colonies of ants, where they manage to find the shortest path connecting their nest and the source of food. As ants travel a route from a starting point to a food source, they deposit pheromone. Subsequent ants will generally choose paths with more pheromone and after many trials they will converge on an optimal path. In a typical ACO based features selection case, the algorithm begins with the generation of certain number of the ants placed randomly on the graph representing the possible combinations of every m/z. Thus, each node (in a graph) relates to a m/z and each edge shows the traversal of an ant from one m/z to another. The number of artificial pheromone [0,1] for an edge is associated with the popularity of the particular traversal by pervious ants. Therefore, ants could make probabilistic decisions to stay at which node and select which edge based on the artificial pheromone and related traversal degree. This will continue until the minimum degree for the objective function (here is the misclassification rate) has reached otherwise the information in each edge will be updated and a new set of ants will

be created and all process will be iterated once more [121, 122]. Another characteristic of the ACO algorithm is the pheromone evaporation rate ρ, a process that leads to a decrease in the pheromone trail intensity over time: it helps in avoiding rapid convergence of the algorithm towards a sub-optimal region. At each iteration, the subsets are evaluated, the pheromone is updated, a new set of ants is created and the procedure is reiterated until a stop criterion is met. **Figure 1.15** shows the general procedure of m/zs selection using ACO.

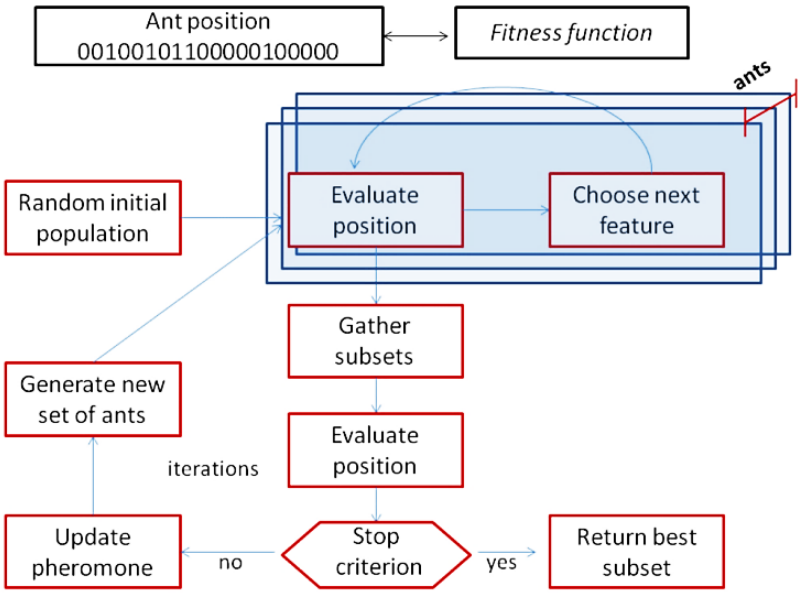


Figure 1.15: General procedure of Ant Colony Optimization

CHAPTER 2

SCOPES AND OBJECTIVES

2.1 The analytical problem

The foodomics approaches can be generally grouped as “profiling” (target) or “fingerprinting” (untargeted) strategies [1]. While profiling involves the analysis of a group of related food constituents, which are in most cases identified and quantified, “fingerprinting” is based on the determination of as many compounds as possible. Target screening provides direct functional information, whereas non-target screening usually results in a great number of features that need to be processed. The amount of data provided by non-target fingerprinting strategies is of great complexity and correct treatment of those data is of the utmost importance. It is obvious that the analytical problem has more than one perspectives that need to be evaluated.

First of all, the development of the analytical method that will be applied for food authentication purposes requires consideration of what is the question that has to be answered. In the case of olive oil, the objective is the identification of markers that are characteristic for genuine extra virgin olive oils (EVOOs). These markers should be representative for the organoleptic profile of EVOOs, the production type, and the variety, among others. Thus, the analytical problem is what kind of chemistry is required for the detection of certain classes of compounds (target screening), and for the fingerprinting of the entire metabolome of EVOOs.

In olive oil authenticity studies, while dealing with analysis of multi-class compounds having different polarities (very polar compounds; such as phenolic acids, and less polar ones; such as secoiridoids and flavonoids), it is fundamental to derive the optimum experimental conditions. In Reversed Phase Liquid Chromatography (RPLC), the more polar molecules elute first and hence, elution order can provide valuable structural information. The complexicity of the phenolic compounds, especially secoiridoids, makes the identification and characterization of their derivatives an analytical task.

Moreover, the extraction of the phenolic compounds (the use of the appropriate extracting media) constitutes a very crucial topic. The addition of the appropriate internal standard is a subject of study, as well. In general, there are no isotopic labelled natural compounds to be used as in internal standards. For this reason, most often, the internal standard used is a natural compound that does not exist in the matrix.

After the optimization of the analytical methodology and the identification of target and suspect compounds, the semi-quantification of the suspect remains a challenge, since in most cases, there are no commercial standards available. Numerous studies have been reported in literature, using the oleuropein calibration curve for the semi-quantification of the secoiridoids [9, 33, 71, 72]. However, there is a gap in literature dealing with the selection of the most appropriate standard for semi-quantification purposes. To such an end, chemical similarity analysis, which has been subject for nearly a decade, can be applied to rank the appropriate standards for semi-quantification purposes.

Non-target screening is a more challenging task. In this case, the analyst takes greater advantage of the development of instrumentation and sophisticated software in order to find the proper workflow to identify compounds that have not been previously reported. High Resolution Mass Spectrometric (HRMS) screening workflows coupled to chemometrics reveal the fingerprint of food products and hence, can be widely applied in food authenticity studies, as complimentary methods. Chemometric tools are fundamental for food authentication studies since huge amounts of data must be handled. Data mining, data fusion and feature selection are essential to propose markers and guarantee the authenticity of food products.

2.2 Research Objectives and Scope

The experimental part of this thesis is consisted of 3 individual studies.

In the first study performed, a novel RP-UHPLC-QTOF-MS/MS analytical method was developed employing target, suspect and non-target screening strategies coupled to advanced chemometric tools for the investigation of the organoleptic

profile of EVOOs. The proposed method was successfully applied in 19 real EVOOs for the identification of markers responsible for its sensory profile. The proposed target analytical method includes the determination of 14 phenolic compounds and demonstrated low limits of detection (LODs) over the range of 0.015 (apigenin) - 0.039 (vanillin) mg kg^{-1} and adequate recoveries (96-107%). A suspect list of 60 relevant compounds was compiled and suspect screening was then applied to all the samples. Semi-quantification of the suspect compounds was performed using the calibration curves of the target compounds having similar structures. Then, a non-target screening workflow was applied with the aim to identify additional compounds, responsible to differentiate EVOOs from defective oils. Robust classification-based models were built using supervised discrimination techniques, Partial Least Square Discriminant Analysis (PLS-DA) and Counter Propagation Artificial Neural Networks (CP-ANNs), for the classification of olive oils into EVOOs or defectives. Variance in Projection (VIP) was calculated to select the most significant features that affect the discrimination. Overall, 51 compounds were identified and suggested as markers. 14, 26 and 11 compounds were identified using target, suspect and non-target screening, respectively. Retrospective analysis was also performed to identify 19 free fatty acids.

In the second study, a correlation between the agronomical practices (organic or conventional) and their effects on the composition of olive oils was established. Thus, a UHPLC-QTOF-MS/MS was developed and validated, using target and suspect screening workflows, coupled with advanced chemometrics for the identification of phenolic compounds and the discrimination between organic and conventional extra virgin olive oils. The method was optimized by one factor design (OFD) and response surface methodology (RSM) to derive the optimal conditions of extraction (methanol:water (80:20, v/v), pure methanol or acetonitrile), and to select the most appropriate internal standard (caffeic acid or syringaldehyde). The results revealed that extraction with methanol:water (80:20, v/v) is the optimum and syringaldehyde 1.30 mg L^{-1} is the appropriate internal standard. The proposed method demonstrated low limits of detection in the range of 0.002 (luteolin) – 0.028 (tyrosol) mg kg^{-1}. Then, it was successfully applied in 52 olive oils of Kolovi variety. Totally, 13 target and 24 suspect phenolic

compounds were identified. Target compounds were quantified with commercially available standards. A novel semi-quantitation strategy, based on chemical similarity, was introduced for the semi-quantification of the identified suspects. Finally, Ant Colony Optimization coupled to Random Forest (ACO-RF) model selected luteolin as the only marker responsible for the discrimination, during a two-year study.

Finally, the third study examines the metabolomic profile of 51 Greek monovarietal EVOOs from the varieties: Manaki, Ladoelia, Koroneiki, Amfissis, Chalkidikis and Kolovi with non-target UHPLC-QTOF-MS/MS. Data processing was carried out with the R language and XCMS package. V-WSP algorithm was used as an unsupervised variable reduction method and decreased the number of the feautures from 287 to 250. The preliminary examination of the distribution of EVOOs toward their cultivars was achieved by Principal Component Analysis (PCA). ACO-RF was developed to prioritize over 250 features and to create a decision based classification tree. Four important markers: apigenin, vanillic acid, luteolin 7-methyl ether and oleocanthal were selected by ACO, playing a significant role in the authentication of Greek EVOOs' cultivars. The proposed method is highly applicable and can be implemented easily in routine analysis for authentication purposes as it sets concentration thresholds (based on the quantification results) over the markers identified.

CHAPTER 3

OLIVE OIL AUTHENTICITY STUDIES BY TARGET AND NON-TARGET LC-QTOF-MS COMBINED WITH ADVANCED CHEMOMETRIC TECHNIQUES

3.1 Introduction

Food authenticity has become increasingly important in recent years due to the drive for more accurate and truthful labeling. A product is characterized as authentic as long as it is firstly described accurately by the label and secondly, complies with the current legislation in force in the country where it is marked or sold [7, 123]. Authenticity constitutes a multifaceted issue that covers many aspects including characterization, adulteration, mislabeling and misleading origin [17]. Thus, there is a growing necessity to develop advanced analytical methods in combination with appropriate data processing tools that could successfully guarantee the authenticity of various food matrices.

In that respect, there is emerging concern for the guarantee of olive oil's authenticity, due to its economic importance, as well as due to its nutritional, sensory and therapeutic properties, which have been extensively elaborated by science [23, 72, 82]. The main authenticity issues that are associated with olive oil quality are adulteration, misdescription of geographical origin, the production type (conventional or organic) and the taste. The latter is the result of certain constituents present in olive oil which affect its sensory profile. According to the Olive Oil Council and its Trade Standards (COI/T.15NC no 3-25, 2003) [124], there are three positive attributes in extra virgin and virgin olive oils: fruity, bitter or pungent and sixteen negative (fusty/muddy, musty-humid-earthy, winey-vinegary, acid-sour, rancid, frostbitten olives, heated or burnt, hay-wood, rough, greasy, vegetable water, brine, metallic, esparto, grubby, cucumber). The presence of positive sensory characteristics is necessary for the classification of the olive oils as "extra virgin" (EVOOs), whereas those with negative attributes have objectionable taste and are characterized "defective". The official method for olive oil's sensory evaluation is implemented so far in a panel test developed by the International Olive Council [125]. However, as Tena et al. [126] have

recently reviewed, the official method is questioned by numerous olive oil sectors and it fails in cases that testers are not able to analyze defects at very low intensities. Moreover, it has several drawbacks since it is time-consuming, it lacks of stable and standardized reference oils with different intensities of bitterness and pungency and it also requires a group of eight to twelve testers for statistically confirmed results [76]. Therefore, as it has been recently suggested by Garcia-Gonzales and Aparicio [127] as well as Tena et al. [126], an objective measurement of virgin olive oil sensory quality should follow another strategy based on Analytical Chemistry.

Many analytical procedures have been employed to identify and quantify the volatile components that characterize olive oil flavor/aroma during the past thirty years. Between them, gas chromatography is the main technique applied for this purpose, as it has been reviewed by Escuderos [128]. Apart from the volatile fraction, however, another group of compounds widely known as “bioactive constituents”, mainly consisting of phenolic compounds, has been reported as important for the flavor of olive oil [129]. Consequently, their detection and identification in olive oils constitutes a challenging field that should be further exploited.

In this field, high resolution mass spectrometry (HRMS) has proved its excellent analytical performance allowing the analysis of a wide range of compounds in food, providing screening and tentative identification for both non-target and target compounds [123, 130, 131]. In olive oil analysis, several studies have been published regarding the qualitative and quantitative analysis of bioactive constituents with liquid chromatography-high resolution mass spectrometry (LC-HRMS), focusing in most cases on geographical origin and varietal discrimination [9, 23, 25, 33, 44, 75, 77, 82, 83]. Still, there is minor information for the sensory discrimination between extra virgin and defective olive oils. It has been suggested, that certain phenolic compounds, and more specifically certain secoiridoid derivatives, such as oleuropein and lingstroside derivatives, are responsible for bitter taste [132]. Nevertheless, the relationship between the individual hydrophilic phenols and olive oil’s sensory characteristics has not been clearly defined [127] and there is still controversy about which individual phenols

are the main contributors to taste attributes [30, 127]. Recent physicochemical and high pressure liquid chromatography (HPLC) developed methodologies for olive oil's bitterness evaluation have obtained inconsistent results with respect to the influence of different phenols. Even though Andrewes et al. [133] have suggested that decarboxymethyl lingstroside aglycone is a pungent compound, correlation between quantitative and sensory data has not been found. Moreover, Dierkes et al. [76] developed a target HPLC-HRMS profiling method to identify several relevant bitter and pungent components, but finally no correlation between the total phenolic content and the ratio bitterness/pungency could be found. These gaps in scientific literature concerning olive oil's organoleptic characteristics and their correlation with certain compounds (markers) could be fulfilled with the use of HRMS non-targeted analytical approaches.

Non-targeted methods combined with suitable chemometric tools improve the breadth of traditional targeted analysis and accelerate new prospects for novel applications [130, 131, 134]. The hyphenation of time-of-flight mass spectrometry (TOF-MS) to LC has proved its excellent analytical performance and offers a good combination of selectivity and sensitivity at high resolution and sub-second scan speeds [1]. Since, using LC-HRMS could result in the generation and detection of a large number of features (m/z), it would be a challenging task to identify them in order to investigate the authenticity of olive oil. In addition, the coupling of LC-HRMS with chemometric tools could decrease remarkably the number of detected features and introduce the most meaningful m/z that could discriminate between extra virgin olive oils and defectives [135].

Therefore, the primary purpose of the present work is the development and application of an integrated LC-HRMS workflow, including target, suspect and non-target screening approaches, coupled with supervised pattern recognition techniques, for olive oil fingerprinting. For that purpose, we developed a target quantitative method for the determination of 15 compounds and a suspect screening method with a list of 60 compounds coupled to a semi-quantitative method for the identified compounds. The identification workflow included strict rule-based filtering steps, deep MS/MS spectra interpretation and retention time prediction. Then, a non-target screening workflow was applied to establish

extensive and reliable pattern recognition models for olive oil fingerprinting by classifying olive oil samples into EVOOs and defectives. Variance in Projection (VIP) was calculated to select the most significant features that affect the discrimination.

3.2 Experimental section

3.2.1 Chemicals and standards

All standards and reagents were of high-purity grade (>95%). Methanol (MeOH) of LC-MS grade and sodium hydroxide (>99%) were purchased from Merck (Darmstadt, Germany). Ammonium acetate (≥99.0%) for HPLC and formic acid (LC-MS Ultra) were purchased from Fluka (Buchs, Switzerland). Isopropanol was purchased from Fisher Scientific (Geel, Belgium). Distilled water was provided by a Milli-Q purification apparatus (Millipore Direct-Q UV, Bedford, MA, USA). For the analytical method validation the following reagents were used: syringic acid 95% was purchased from Extrasynthèse (Genay, France),gallic acid 98 %, ferulic acid 98%, epicatechin 97%, p-coumaric (4-hydroxycinnamic acid) 98%, homovanillic acid 97%, quercetin 98% as well as oleuropein 98% and pinoresinol 95% were obtained from Sigma-Aldrich (Steinheim, Germany), hydroxytyrosol 98% was purchased from Santa Cruz Biotechnologies, caffeic acid 99% (internal standard), vanillin 99%, ethyl vanillin 98%, apigenin (4,5,7 trihydroxyflavone) 97% and tyrosol (2-(4-hydroxyphenyl)ethanol) 98% were acquired from Alfa Aesar (Karlsruhe, Germany). In order to confirm the identity of suspect and non-target compounds, luteolin 98% from Santa Cruz Biotechnologies was also acquired. For the determination of free fatty acids, hexanoic acid 99.5%, octanoic 99.5%, dodecanoic 99.5%, myristic 99.5%, pentadecanoic 99.5%, palmitic 99%, palmitoleic 98.5%, heptadecanoic 98%, heptedecenoic 99%, stearic 98.5%, oleic 99%, α-linoleic 99%, α**-**linolenic 99%, arachidic 99%, cis-eicosenoic acid 99%, heneicosanoic 99%, docosanoic 99%, tricosanoic 99% and lignoceric acid 99% were purchased from Sigma-Aldrich. Stock standard solutions of individual compounds (1000 mg L^{-1}) were solubilized in methanol and stored at -20 °C in dark brown glass. All intermediate standard solutions containing the analytes were prepared by dilution of the stock solutions in methanol.

3.2.2 Olive oil samples

Three standard defective olive oil samples with a known score (rancid, fusty and muddy) were acquired from the International Olive Council (IOC) and 3 defective and 16 extra virgin olive oils of the Kolovi and Adramytiani varieties, both monovarietal and mixtures, were provided along with the sensory evaluation by ELGO-DIMITRA I.O.S.V. in Lesvos. These samples were produced from olives during harvesting period in 2014-2015 and cultivated in different regions in Lesvos Island. To provide the sensory evaluation, the oils were subjected to an extended panel based on the EU Regulation No 1348/2013 [136] and IOC instructions [125]. The results are expressed as a median of the rates reported by eight analysts. The highest mean coefficient of variation was in all cases less than 20%. **Figure 3.1** describes the sensory profile of the extra virgin and defective olive oils acquired, represented as spider plots. More information about the samples concerning the exact organoleptic scores, the geographical origin, the variety and production type as well as the time of harvest can be found in **Table 3.1**.

Table 3.1: Organoleptic characterization, origin, time of harvest and production type of the samples

Sample	Taste	Region	Variety	Production	Harvest	Fruity	Bitter	Pungent	Fusty	Musty	Rancid
AX1	EVOO	Megaloxori	Kolovi	Conventional	January	4.8	3.8	3.7			
BD1	EVOO	Agiasos	70%Kolovi 30% Adramytiani	Conventional	November	5.0	3.3	3.4			
BL1	EVOO	Akrasi	Kolovi	Conventional	November	5.0	4.0	3.5			
BL2	EVOO	Akrasi	Kolovi	Conventional	November	3.0	2.2	2.4			
BR1	Defective	Agia Paraskevi	20%Kolovi 80% Adramytiani	Conventional	December				3.2		
EX1	EVOO	Gera	Kolovi	Conventional	December-January	4.6	2.3	2.3			
Fusty	Defective	Unknown	Unknown	Unknown	Unknown				7.8		

Sample	Taste	Region	Variety	Production	Harvest	Fruity	Bitter	Pungent	Fusty	Musty	Rancid
KAI	Defective	Gera	70%Kolovi 30% Adramytiani	Conventional	January-February				3.9		
LP1	EVOO	Moria	70%Kolovi 30% Adramytiani	Conventional	November	4.4	3.7	3.8			
Musty	Defective	Unknown	Unknown	Unknown	Unknown					4.65	
NB1	Defective	Mantamados	20%Kolovi 80% Adramytiani	Conventional	January-February				3.8		
PA1	EVOO	Megaloxori-Palaioxori	Kolovi	Organic	December	5.1	3.1	3.7			
PK1	Defective	Megaloxori	Kolovi	Conventional	January				4.7		
Rancid	Defective	unknown	unknown	Unknown	Unknown						9.5
RB1	Defective	Parakoila	Adramytiani	Conventional	December				3.4		

Sample	Taste	Region	Variety	Production	Harvest	Fruity	Bitter	Pungent	Fusty	Musty	Rancid
RB2	EVOO	Megaloxori-Palaioxori	Kolovi	Organic	December	4.3	3.2	3.3			
RB4	EVOO	Palaioxori	Kolovi	Conventional	December-January	5.2	3.9	3.8			
TP1	EVOO	Gera	70%Kolovi 30%Adramytiani	Conventional	December-January	3.6	1.7	2.2			
YH1	EVOO	Komi	70%Kolovi 30%Adramytiani	Conventional	December	4.8	3.5	3.9			

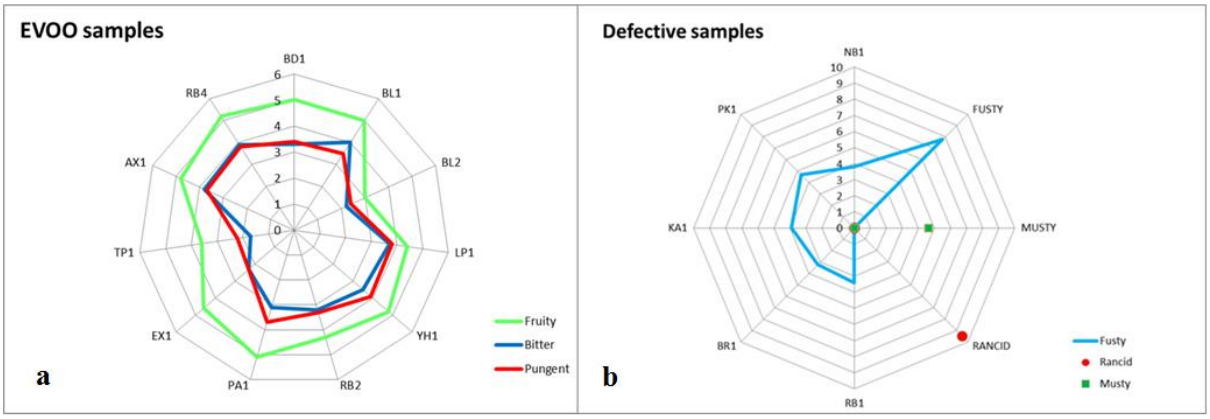


Figure 3.1: Spider plots of the organoleptic profile of olive oil samples with sensory attributes for (a) EVOO samples (fruity, bitter, pungent); (b) Defective olive oils (fusty, rancid, musty).

All samples were protected from light and humidity and stored in dark colored glasses at the ideal temperature of 14-15°C [56]. Moreover, to better preserve the quality of olive oils and increase the resistance against autoxidation, nitrogen as an inert gas was added in the bottles [137].

3.2.3 Sample Extraction

A liquid-liquid micro-extraction (LLME) method was employed for isolating the phenolic compounds from the olive oil samples as it was developed and validated by Becerra-Herrera et al. [71]. 0.5 g of each sample was weighted and 0.5 mL of methanol: water (80:20,v/v) was added in 2 mL Eppendorf tubes. Afterwards, the mixture was vortexed for 2 minutes, and centrifuged for 5 minutes at 13,400 rpm. Furthermore, the upper phase that consisted of methanol was collected and filtered through the membrane syringe filters of regenerated cellulose (CHROMAFIL® RC) (15 mm diameter, 0.22 μm pore size, provided by Macherey-Nagel (Düren, Germany). 200 μL of the methanolic phase was diluted with ultrapure water up to 0.5 mL. Finally, 5 μL of this solution was injected in the chromatographic system. Procedural blanks were also prepared and processed in the chromatographic system to detect any potential contamination.

3.2.4 Quality Control

Quality Control (QC) samples were used in order to verify that the analytical system has been stabilized prior to analyze the main batch of samples and assess its performance. A typical QC sample was prepared, as it has been suggested by Want et al. [138], by mixing all aliquots of the samples. In the beginning of the analysis, the QC was injected five times for conditioning and afterwards, it was injected at regular intervals (*i.e*. every ten sample injections) throughout the analytical run to provide a set of data from which repeatability can be assessed. The calculated Relative Standard Deviation (RSD) values for the t_R and the peak areas as well as *Δ*m errors (n=7) are presented in the **ESM I, Table S1** proving the good performance of the analytical system.

3.2.5 Instrumental analysis

RP chromatographic analysis was performed using a UHPLC system with an HPG-3400 pump (Dionex UltiMate 3000 RSLC, Thermo Fisher Scientific, Germany) interfaced to a Q-TOF mass spectrometer (Maxis Impact, Bruker Daltonics, Bremen, Germany) in negative electrospray ionization mode. Separation was carried out using an Acclaim RSLC C18 column (2.1 × 100 mm, 2.2 μm) purchased from Thermo Fisher Scientific (Driesch, Germany) with a pre-column of ACQUITY UPLC BEH C18 (1.7 μm, VanGuard Pre-Column, Waters (Ireland)). The separation was operated at column temperature of 30°C. The solvents used consisted of: (A) 90% H_2O, 10% MeOH and 5 mM CH_3COONH_4, (B) 100% MeOH and 5 mM CH_3COONH_4. The adopted elution gradient started with 1% of organic phase B (flow rate 0.2 mL min^{-1}) during one minute, gradually increasing to 39 % for the next 2 minutes, and then increasing to 99.9 % (flow rate 0.4 mL min^{-1}) in the following 11 minutes. These almost pure organic conditions were kept constant for 2 minutes (flow rate 0.48 mL min^{-1}) and then initial conditions (1% B - 99% A) were restored within 0.1 minute (flow rate decreased to 0.2 mL min^{-1}) to re-equilibrate the column for the next injection.

The Q-TOF MS system was equipped with an electrospray ionization interface (ESI), operating in negative mode with the following settings: capillary voltage of 3500 V; end plate offset of 500 V; nebulizer pressure of 2 bar (N_2); drying gas of 8 L min^{-1} (N_2); and drying temperature of 200 °C. A Q-TOF external calibration was daily performed with a sodium formate cluster solution, and a segment (0.1−0.25 min) in every chromatogram was used for internal calibration, using a calibrant injection at the beginning of each run. The sodium formate calibration mixture consisted of 10 mM sodium formate in a mixture of water/isopropanol (1:1). Full scan mass spectra were recorded over the range of 50-1000 m/z, with a scan rate of 2 Hz. MS/MS experiments were conducted using AutoMS data dependent acquisition mode based on the fragmentation of the five most abundant precursor ions per scan. For certain masses of interest, if the intensity of the m/z was low, a second analysis including the list of the selected precursor ions was carried out in AutoMS (data dependent acquisition) mode. The instrument provided a typical resolving power (FWHM) between 36,000-40,000 at m/z: 226.1593, 430.9137 and 702.8636.

3.2.6 Screening Strategies

A target list was created including 15 significant phenolic compounds that have been identified in olive oil and are reported in the literature [9, 23, 25, 33, 44, 72, 75-77, 82, 83]. The list consisted of different classes of phenolic compounds, such as phenolic acids, secoiridoids, flavonoids and lignans, with commercially available standards. The initial target list can be found in **Table 3.2**.

Table 3.2: Target list

Compound	**Molecular Formula**	**$[M-H]^-$**	**t_R (min)**
Vanillin	$C_8H_8O_3$	152.0473	4.71
Apigenin	$C_{15}H_{10}O_5$	269.0455	8.24
Caffeic acid	$C_9H_8O_4$	179.0349	1.53
Epicatechin	$C_{15}H_{14}O_6$	289.0716	4.35

Compound	**Molecular Formula**	**$[M-H]^-$**	**t_R (min)**
Ethyl vanillin	$C_9H_{10}O_3$	165.0557	5.45
Ferulic acid	$C_{10}H_{10}O_4$	193.0506	1.39
Gallic acid	$C_7H_6O_5$	169.0142	1.24
Homovanillic acid	$C_9H_{10}O_4$	181.0491	1.48
Hydroxytyrosol	$C_8H_{10}O_3$	153.0551	3.52
Oleuropein	$C_{25}H_{32}O_{13}$	539.1722	5.96
p-Coumaric acid	$C_9H_8O_3$	163.0403	1.34
Pinoresinol	$C_{20}H_{22}O_6$	357.1334	6.48
Quercetin	$C_{15}H_{10}O_7$	301.0306	7.46
Syringic acid	$C_9H_{10}O_5$	197.0417	1.43
Tyrosol	$C_8H_{10}O_2$	137.0610	4.08

A suspect list was also generated from literature including all the phenolic compounds that have already been identified in olive oil and in different organs of *Olea Europaea* (stems, leaves, drupes) in order to scan their presence in the olive oil samples of the current study. The initial suspect list consisted of 60 bioactive constituents and is presented in **Table 3.3** with the molecular formulas and the simplified molecular-input line-entry system (SMILES*)* of the suspect compounds, as well as the references of the studies in which they have been previously reported.

Table 3.3: Suspect list of 96 bioactive compounds present in olive oils, drupes and leaves, extrapolated from the literature

Compound	Molecular Formula	$[M-H]^-$ m/z calc.	SMILES	Ref.
Hydroxytyrosol glucoside	$C_{14}H_{20}O_8$	315.1085	OC[C@H]1O[C@@H](OCCC2=CC(O)=C(O)C=C2)[C@H](O)[C@@H](O)[C@@H]1O	[79]
Tyrosol glucoside	$C_{14}H_{20}O_7$	299.1136	[H][C@]1(CO)O[C@@]([H])(OCCC2=CC=C(O)C=C2)[C@]([H])(O)[C@@]([H])(O)[C@]1([H])O	[28]
Elenolic acid glucoside	$C_{17}H_{24}O_{11}$	403.1245	COC(=O)C1=CO[C@@H](O[C@@H]2O[C@H](CO)[C@@H](O)[C@H](O)[C@H]2O)\C(=C\C)[C@@H]1CC(O)=O	[139]
Gallocatechin	$C_{15}H_{14}O_7$	305.0666	O[C@@H]1CC2=C(O)C=C(O)C=C2O[C@H]1C1=CC(O)=C(O)C(O)=C1	[79]
Luteolin 3.7–o-diglucoside	$C_{27}H_{30}O_{16}$	609.1461	OC[C@@H]1O[C@@H](OC2=CC(O)=C3C(=O)C=C(OC3=C2)C2=CC(O[C@@H]3O[C@@H](CO)[C@H](O)[C@@H](O)[C@@H]3O)=C(O)C=C2)[C@@H](O)[C@H](O)[C@H]1O	[28]
Demethyl oleuropein	$C_{24}H_{30}O_{13}$	525.1613	C\C=C1\[C@H](O[C@@H]2O[C@H](CO)[C@@H](O)[C@H](O)[C@H]2O)OC=C([C@H]1CC(=O)OCCC1=CC=C(O)C(O)=C1)C(O)=O	[79]
Rutin	$C_{27}H_{30}O_{16}$	609.1461	CC1OC(OCC2OC(OC3=C(OC4=CC(O)=CC(O)=C4C3=O)C3=CC=C(O)C(O)=C3)C(O)C(O)C2O)C(O)C(O)C1O	[79]
Quercetin-3-o-glucoside	$C_{21}H_{20}O_{12}$	355.0882	OC[C@@H](O)[C@H]1O[C@@H](OC2=C(OC3=CC(O)=CC(O	[139]

Compound	**Molecular Formula**	**$[M-H]^-$ m/z calc.**	**SMILES**	**Ref.**
			)=C3C2=O)C2=CC(O)=C(O)C=C2)[C@H](O)[C@H]1O	
Luteolin-7-o-glucoside	$C_{21}H_{20}O_{11}$	339.0932	OC[C@@H](O)[C@H]1O[C@@H](OC2=C(OC3=CC(O)=CC(O)=C3C2=O)C2=CC(O)=C(O)C=C2)[C@H](O)[C@H]1O	[72]
Verbascoside	$C_{29}H_{36}O_{15}$	623.1981	CC1OC(OC2C(O)C(OCCC3=CC=C(O)C(O)=C3)OC(CO)C2OC(=O)C=CC2=CC=C(O)C(O)=C2)C(O)C(O)C1O	[28]
2-Methoxyoleuropein	$C_{26}H_{34}O_{14}$	569.1875	[H][C@]1(CC(=O)OC[C@@H](OC)C2=CC(O)=C(O)C=C2)\C(=C/C)[C@H](O[C@@H]2O[C@H](CO)[C@@H](O)[C@H](O)[C@H]2O)OC=C1C(=O)OC	[79]
Ligstroside	$C_{25}H_{32}O_{12}$	523.1821	CC1OC(OC2C(O)C(OCCC3=CC=C(O)C(O)=C3)OC(CO)C2OC(=O)C=CC2=CC=C(O)C(O)=C2)C(O)C(O)C1O	[28]
Diosmetin	$C_{16}H_{12}O_6$	299.0561	COC1=CC=C(C=C1O)C1=CC(=O)C2=C(O)C=C(O)C=C2O1	[79]
Kaempferol-o-rutinoside	$C_{27}H_{30}O_{15}$	539.1512	C[C@H]1[C@@H]([C@H]([C@H]([C@@H](O1)OC[C@@H]2[C@H]([C@@H]([C@H]([C@@H](O2)Oc3c(=O)c4c(cc(cc4oc3c5ccc(cc5)O)O)O)O)O)O)O)O)O	[83]
Homovanillyl alcohol	$C_9H_{12}O_3$	167.0713	COC1=CC(CCO)=CC=C1O	[27]
Hydroxyphenylacetic acid	$C_8H_8O_3$	151.0400	OC(C(O)=O)C1=CC=CC=C1	[2]
Protocatechuic acid	$C_7H_6O_4$	153.0193	OC(=O)C1=CC(O)=C(O)C=C1	[27]
Gentisic acid	$C_7H_6O_4$	153.0193	OC(=O)C1=CC(O)=CC=C1O	[27]
p- Hydroxybenzoic acid	$C_7H_6O_3$	137.0244	OC(=O)C1=CC=C(O)C=C1	[27]

Compound	Molecular Formula	$[M-H]^-$ m/z calc.	SMILES	Ref.
Benzoic acid	$C_7H_6O_2$	121.0295	OC(=O)C1=CC=CC=C1	[27]
Nuzhenide	$C_{31}H_{42}O_{17}$	685.2349	COC(=O)C1=CO[C@@H](O[C@@H]2O[C@H](CO)[C@@H](O)[C@H](O)[C@H]2O)\C(=C/C)C1CC(=O)OC[C@@H]1O[C@H](OCCC2=CC=C(O)C=C2)[C@@H](O)[C@H](O)[C@H]1O	[79]
1-Phenyl-6,7-dihydroxy-isochroman	$C_{15}H_{14}O_3$	241.0870	OC1=C(O)C=C2C(OCCC2=C1)C1=CC=CC=C1	[27]
1-(3´-methoxy-4´hydroxy)Phenyl-6,7-dihydroxy-isochroman	$C_{16}H_{16}O_5$	287.0924	COC1=C(O)C=CC(=C1)C1=C(O)C(O)=CC2=C1COCC2	[27]
Methoxy quinone	$C_7H_6O_3$	137.0244	COC1=CC(=O)C=CC1=O	[139]
Azelaic acid	$C_9H_{16}O_4$	187.0975	OC(=O)CCCCCCCC(O)=O	[83]
Absisic acid	$C_{15}H_{20}O_4$	263.1288	C\C(\C=C\[C@@]1(O)C(C)=CC(=O)CC1(C)C)=C\C(O)=O	[83]
Licodione	$C_{15}H_{12}O_5$	271.0611	OC1=CC=C(C=C1)C(=O)CC(=O)C1=CC=C(O)C=C1O	[83]
Chrysoeriol/ 3methyl-o-luteolin	$C_{16}H_{12}O_6$	299.0561	COC1=CC(=CC=C1O)C1=CC(=O)C2=C(O)C=C(O)C=C2O1	[139]
Xanthonic acid	$C_{14}H_8O_4$	239.0349	OC(=O)C1=C2C(OC3=CC=CC=C3C2=O)=CC=C1	[83]
Secologanoside	$C_{16}H_{22}O_{11}$	389.1089	OC[C@H]1O[C@@H](O[C@@H]2OC=C([C@@H](CC(O)=O)[C@H]2C=C)C(O)=O)[C@H](O)[C@@H](O)[C@@H]1O	[28]
Taxifolin	$C_{15}H_{12}O_7$	303.0510	OC1C(OC2=CC(O)=CC(O)=C2C1=O)C1=CC=C(O)C(O)=C1	[28]

Compound	**Molecular Formula**	**$[M-H]^-$ m/z calc.**	**SMILES**	**Ref.**
Olivil	$C_{20}H_{24}O_7$	375.1449	COC1=C(O)C=CC(C[C@@]2(O)CO[C@@H]([C@H]2CO)C2=CC(OC)=C(O)C=C2)=C1	[79]
Fraxamoside	$C_{25}H_{30}O_{13}$	537.1613	[H][C@@]12CC(=O)OC[C@H](OC[C@H]3O[C@@H](O[C@H](OC=C1C(=O)OC)\C2=C\C)[C@H](O)[C@@H](O)[C@@H]3O)C1=CC(O)=C(O)C=C1	[79]
Hydroxylated form of elenolic acid	$C_{11}H_{14}O_7$	257.0667	COC(=O)C1=CO[C@@H](CO)[C@@H](C=O)[C@@H]1CC(O)=O	[83]
Elenolic acid methyl ester	$C_{12}H_{16}O_6$	255.0874	COC(=O)C[C@H]1[C@H](C=O)[C@H](C)OC=C1C(=O)OC	[139]
Apigenin-7-glucoside	$C_{21}H_{20}O_{10}$	431.0983	OC[C@H]1O[C@@H](OC2=CC(O)=C3C(=O)C=C(OC3=C2)C2=CC=C(O)C=C2)[C@H](O)[C@@H](O)[C@@H]1O	[72]
Diosmin	$C_{28}H_{32}O_{15}$	607.1668	COC1=CC=C(C=C1O)C1=CC(=O)C2=C(O)C=C(O[C@@H]3O[C@H](CO[C@@H]4O[C@@H](C)[C@H](O)[C@@H](O)[C@H]4O)[C@@H](O)[C@H](O)[C@H]3O)C=C2O1	[28]
Hydroxytyrosol acetate	$C_{10}H_{12}O_4$	195.0660	CC(=O)OCCC1=CC(O)=C(O)C=C1	[33]
Decarboxymethyl oleuropein aglycone	$C_{17}H_{20}O_6$	319.1185	C\C=C(\C=O)[C@@H](CC=O)CC(=O)OCCC1=CC(O)=C(O)C=C1	[31]
Decarboxymethyl lingstroside aglycone	$C_{17}H_{20}O_5$	303.1237	C\C=C(\C=O)[C@@H](CC=O)CC(=O)OCCC1=CC=C(O)C=C1	[31]
10-Hydroxy oleuropein	$C_{19}H_{22}O_9$	393.1193	[H]C(O)\C=C1/[C@H](O)OC=C([C@H]1CC(=O)OCCC1=CC(O)	[33]

Compound	**Molecular Formula**	**$[M-H]^-$ m/z calc.**	**SMILES**	**Ref.**
aglycone			=C(O)C=C1)C(=O)OC	
Oleuropein aglycone	$C_{19}H_{22}O_8$	377.1241	COC(=O)C1=CO[C@@H](O)\C(=C\C)[C@@H]1CC(=O)OCCC1=CC(O)=C(O)C=C1	[33]
Lingstroside aglycone	$C_{19}H_{22}O_7$	361.1291	COC(=O)C1=CO[C@@H](O)\C(=C\C)[C@@H]1CC(=O)OCCC1=CC=C(O)C=C1	[33]
Syringaresinol	$C_{22}H_{26}O_8$	417.1554	COC1=CC(=CC(OC)=C1O)C1OCC2C1COC2C1=CC(OC)=C(O)C(OC)=C1	[33]
Oleoside	$C_{16}H_{22}O_{11}$	389.1089	[H][C@]1(CO)O[C@@]([H])(O[C@]2([H])OC=C(C(O)=O)[C@@]([H])(CC(O)=O)\C2=C/C)[C@]([H])(O)[C@@]([H])(O)[C@]1([H])O	[31]
1-Hydroxypinoresinol	$C_{20}H_{22}O_7$	373.1292	COC1=C(O)C=CC(=C1)C1OCC2(O)C1COC2C1=CC(OC)=C(O)C=C1	[139]
8-Hydroxypinoresinol	$C_{20}H_{22}O_7$	373.1292	[H][C@]12CO[C@H](C3=CC(OC)=C(O)C=C3)[C@@]1(O)CO[C@@H]2C1=CC(OC)=C(O)C=C1	[139]
Oleanolic acid	$C_{30}H_{48}O_3$	455.3535	CC1(C)CC[C@@]2(CC[C@]3(C)C(=CC[C@@H]4[C@@]5(C)CC[C@H](O)C(C)(C)[C@@H]5CC[C@@]34C)[C@@H]2C1)C(O)=O	[79]

Compound	Molecular Formula	$[M-H]^-$ m/z calc.	SMILES	Ref.
Maslinic acid	$C_{30}H_{48}O_4$	471.3484	CC1(C)CC[C@@]2(CC[C@]3(C)C(=CC[C@@H]4[C@@]5(C)C[C@@H](O)[C@H](O)C(C)(C)[C@@H]5CC[C@@]34C)[C@@H]2C1)C(O)=O	[79]
1-Acetoxypinoresinol	$C_{22}H_{24}O_8$	415.1398	COC1=C(O)C=CC(=C1)C1OCC2(OC(C)=O)C1COC2C1=CC(OC)=C(O)C=C1	[139]
1-Acetoxypinoresinol Isomer	$C_{22}H_{24}O_8$	415.1398	[H][C@]12CO[C@H](C3=CC(OC)=C(O)C=C3)[C@]1(CO[C@@H]2C1=CC=C(O)C(OC)=C1)OC©=O	[139]
8-Acetoxypinoresinol	$C_{22}H_{24}O_8$	415.1398	COC1=C(O)C=CC(=C1)C1OCC2(OC(C)=O)C1COC2C1=CC(OC)=C(O)C=C1	[139]
Methyl oleuropein aglycone	$C_{20}H_{24}O_8$	391.1412	CC\C=C1\[C@H](O)OC=C([C@H]1CC(=O)OCCC1=CC(O)=C(O)C=C1)C(=O)OC	[139]
Elenolic acid	$C_{11}H_{14}O_6$	241.0714	COC(=O)C1=CO[C@@H](C)[C@@H](C=O)[C@@H]1CC(O)=O	[83]
10-Hydroxy decarboxymethyl oleuropein aglycone	$C_{17}H_{20}O_7$	335.1150	[H]C(O)\C=C1\C(O)OC=C[C@H]1CC(=O)OCCC1=CC(O)=C(O	[83]
Luteolin	$C_{15}H_{10}O_6$	285.0404	OC1=CC(O)=C2C(=O)C=C(OC2=C1)C1=CC(O)=C(O)C=C1	[83]
10-Hydroxy-10Methyl oleuropein aglycone	$C_{20}H_{24}O_9$	407.1347	COC(=O)C1=CO[C@@H](O)\C(=C/C(C)O)[C@@H]1CC(=O)OCCC1=CC(O)=C(O)C=C1	[139]

Compound	**Molecular Formula**	**$[M-H]^-$ m/z calc.**	**SMILES**	**Ref.**
Aesculin	$C_{15}H_{16}O_9$	339.0721	OC[C@H]1O[C@@H](OC2=C(O)C=C3OC(=O)C=CC3=C2)[C@H](O)[C@@H](O)[C@@H]1O	[79]
Calceolarioside	$C_{23}H_{26}O_{11}$	477.1402	OC[C@H]1O[C@@H](OCCC2=CC=C(O)C(O)=C2)[C@H](O)[C@@H](O)[C@@H]1OC(=O)\C=C\C1=CC(O)=C(O)C=C1	[79]
iso-Acteoside	$C_{29}H_{36}O_{15}$	623.1981	C[C@@H]1O[C@@H](O[C@H]2[C@H](O)[C@@H](COC(=O)\C=C\C3=CC(O)=C(O)C=C3)O[C@@H](OCCC3=CC(O)=C(O)C=C3)[C@@H]2O)[C@H](O)[C@H](O)[C@H]1O	[31]
Calc: calculated, SMILES: Simplified Molecular -Input Line- Entry System, Ref: reference				

Target, suspect and non-target screening workflows were followed as they have been suggested by Krauss et al. [80] and Gago-Ferrero et al. [81]. Target screening was performed using Bruker software packages (Bruker Daltonics, Bremen, Germany) Target Analysis 1.3 and Data Analysis 4.1, as well as other tools available in these packages (Bruker Compass Isotope Pattern and SmartFormula Manually). Extracted Ion Chromatograms (EICs) were obtained using the function Find Compounds-Chromatogram (Target Analysis software), which creates the base peak chromatograms for the masses that accomplish thresholds of intensity and accuracy according to the following parameters that were set; mass accuracy window: 5 ppm, satisfactory isotopic fit was denoted only when mSigma (mSigma-Value is a measure for the goodness of fit between measured and theoretical isotopic pattern) was below or equal to 50, signal to noise threshold was set to 3, minimum area threshold was set to 800, while minimum intensity threshold was set to 200. Relative tolerance of retention time window was set lower than ±0.2 min. All the target compounds that were included in the database were identified based on mass accuracy, isotopic pattern, retention time and MS/MS fragments.

For the identification of the suspect compounds, the masses of the deprotonated ions were calculated based on the molecular formula and EICs were created in Target Analysis 1.3 with the following parameters; mass accuracy threshold up to 5 ppm, isotopic fit below or equal to 50 mSigma, ion intensity more than 800, peak area threshold 2000 and peak score (ratio area/intensity) more than 4 (preferable peak score should be between 4-38). If one or more peaks were detected using EICs, the isotopic pattern and the MS/MS fragments were examined in Data Analysis 4.1, to confirm that the peak represents the suspected compound. The comparison and interpretation of the MS/MS fragments were carried out using literature data and in silico fragmentation tools, mainly Metfrag [91] and spectral libraries such as MassBank [92]. Moreover, the possible retention time of each suspect compound was predicted and compared with the experimental retention time by an in-house developed model based on Quantitative Structure Retention Relationship (QSRR) [94], since reference

standard solutions were not commercially available for most of the suspected compounds.

Following the suspect screening, non-target screening was performed. Non-target screening involves the detection of peaks and the identification of compounds without having *a priori* information or available standards [140, 141]. Peak picking was carried out as it is explained in detail in the section Data Processing and Chemometrics. The selected peaks were tentatively identified according to mass accuracy (< 5 ppm) and isotopic pattern of the precursor ion (< 50 mSigma), their fragmentation pattern and the retention time of the extracted ion chromatographic peak. Elemental compositions of the precursor and fragment ions were suggested and plausible molecular formulas were proposed using Smart Formula Tool in Data Analysis 4.1 Software. MS/MS spectra were examined and interpreted as discussed in suspect screening in order to determine tentative candidates. QSRR prediction model was also used as a complementary tool for the identification of the non-target compounds in cases that there were no standards available [94].

The level of confidence achieved in the identification of the detected compounds was established according to Schymanski et al. [142]. Level 1 corresponds to confirmed structures where a reference standard is available, level 2 to probable structures (level 2a: evidence by spectra matching from literature or library and level 2b: diagnostic evidence where no other structure fits the experimental MS/MS information), level 3 for tentative candidate(s), level 4 to unequivocal molecular formulas and level 5 to exact mass(es) of interest. The detected compounds were labeled based on this classification.

3.2.7 Data Processing

In order to process LC-HRMS data, first of all, binary files of all the analysed samples were converted to mzXML files using Proteowizard software [143]. Then, these files were processed using the R-language and the XCMS package [98] to extract peaks that are present in the samples. This procedure involved peak picking by CentWave algorithm [98], grouping of peaks representing the same analyte across the samples and a step to correct the chromatographic drift

of retention time. The next step was to regroup the peaks of which the retention time was changed. At final step, a step of filling in the missing peaks was implemented. This replaces the missing values of none detected peaks with a small value of intensity [144]. CAMERA package was also used to deisotoping and removing the adduct peaks to avoid co-linearity during the model construction [145]. The internal parameters of the algorithms used for peak peaking, grouping and retention time alignment were optimized using the package IPO [146]. The optimal settings are presented in detail in **Table 3.4**.

Table 3.4: Parameters used for the computational analysis

Input Parameter	**negative ESI**
CentWave parameters	
ppm	17.6
Minimum peak width	15.5
Maximum peakwidth	50
prefilter	3, 1000
scan range	20 until 1840
fitgauss	true
integrate	true
Retention Time alignment based on OBI-Warp algorithm	
Distance function	cor_opt
gapInit	0.27
gapExtend	2.36
Grouping of features based on kernel density estimator	
bw	5
mzwid	0.0305
minfrac	0.6
minsamp	2
max	50

3.2.8 Chemometrics

Overall, 304 molecular features were obtained and grouped for 19 samples of olive oils. These samples were split into training and test set randomly in order to generate the classification models and then, evaluate the accuracy of classification for the external set of samples. Multivariate classification methods such as Partial Least Square Discriminant Analysis (PLS-DA) [108, 147], and Self-Organizing Maps (SOMs) [148], which are supervised pattern recognition techniques, were used to classify the olive oils into EVOOs and defective samples and investigate, subsequently, the existing relationship between the samples. Variable Importance in Projection (VIP) [108] was applied in order to distinguish and to detect the most important compounds responsible for discrimination. VIP scores estimate the importance of each variable (in this case the m/z) in the projection used to build the PLS-DA model and could be useful criteria to select significant m/z [120, 149]. VIP is a score of the variable (m/z) that shows the contribution of a variable in final Latent Variables (LVs). In order to prioritize the peaks which caused greater variation in the discrimination between samples, VIP values were calculated for the PLS-DA model. Those m/z with VIP score greater than 0.83 were considered as the most important because they cause greater variation [148] and the non-target identification workflow was applied for their identification, as it was described in **3.2.6.**

3.2.9 Method Validation

The optimized RP-UHPLC-ESI-MS method was validated to ensure that it is suitable for identification and quantification purposes. Standard addition curves were constructed for all the analytes. All the compounds were spiked in real EVOO samples. Gallic acid, p-coumaric acid, ferulic acid, syringic acid, homovanillic acid, tyrosol, hydroxytyrosol, pinoresinol, apigenin, oleuropein, vanillin, ethyl vanillin, epicatechin and quercetin were spiked at concentrations between 0.05- 5 mg kg^{-1} (10 calibration levels with 3 replicates at each level). Calibration curves were constructed with the use of the peak area of the spiked analyte subtracted by the peak area of a neat sample and divided by the peak area of the internal standard (caffeic acid, 0.5 mg kg^{-1}). Limits of detection (LODs)

and limits of quantification (LOQs) were calculated at the lowest concentration range of the analytes (0.05-1 mg kg^{-1}), by the equations:

$$LOD = \frac{3.3 \times S_a}{b} \quad (1)$$

$$LOQ = \frac{10 \times S_a}{b} \quad (2)$$

Where: S_a is the standard error of the intercept *a*; and *b* is the slope of the calibration curve.

The accuracy of the method was estimated using recoveries, at 0.5 mg kg^{-1} concentration level, calculated as follows:

$$\%RE = \frac{\text{Response extracted sample}}{\text{Response post extracted spiked sampe}} x\ 100 \quad (3)$$

Where: $Response_{extracted\ sample}$ is the average area of the analyte in matrix, which has been through the extraction process, from 3 replicates, divided each time by the peak area of the internal standard; $Response_{poste\ xtracted\ spiked\ sample}$ is the average area of each analyte, spiked into extracted matrix after the extraction procedure.

To evaluate the matrix effect, the matrix factor was calculated at 0.5 mg kg^{-1} concentration level according to the following equation:

$$MF = \frac{\text{Response of post extracted sample}}{\text{Response of standard solution}} \quad (4)$$

Where: $Response_{post\ extracted\ sample}$ is the average area of the analyte, spiked into the extracted matrix after the extraction procedure and $Response_{standard\ solution}$ is the average area count for the same concentration of analyte in a standard solution.

For the calculation of ME, 1 was subtracted by the quotient (4) and multiplied by 100, so that the negative result indicates suppression and the positive result indicates enhancement of the analyte signal.

The precision of the method was demonstrated in terms of repeatability (intra-day precision) and intra-laboratory reproducibility (inter-day precision), at spiked

concentrations of 0.5 mg kg^{-1}. Repeatability was expressed as the $\%RSD_r$ values of 10 replicate analyses (n = 10) in the same day. Reproducibility experiments expressed as the $\%RSD_R$ value of 3 replicates of two consecutive days ($n \times k = 3 \times 2 = 6$).

3.3 Results and discussion

3.3.1 Target screening results

Data dependent method was used to scan the presence of target compounds in real olive oil samples and 14 target compounds were determined. Those were ferulic acid, gallic acid, homovanillic acid, p-coumaric acid and syringic acid from the group of the phenolic acids, tyrosol and hydroxytyrosol from the class of phenolic alcohols, vanillin and ethyl vanillin from the group of phenolic aldehydes, apigenin, quercetin and epicatechin from flavonoids, pinoresinol which is a lignan and the secoiridoid oleuropein. The mass accuracy of the precursor ions, as well as of the qualifiers of the detected compounds were < 5 ppm compared to standard solutions and the isotopic fit was < 50 mSigma in all cases. The most abundant fragments provided by the MS/MS (AutoMS) spectra were confirmed using Metfrag [91] as well as literature records. Target screening results are summarized in **Table 3.5**.

Table 3.5: Target screening results

Compound	**Molecular Formula**	**$[M-H]^-$ m/z std.**	**$[M-H]^-$ m/z exp.**	**t_R (min)**	**Δt_R (min)**	**Fragments *m/z***	**Elemental Formula**
Vanillin	$C_8H_8O_3$	151.0400	151.0400	4.73	+0.02	71.0140 95.0140 108.0217 136.0162	$C_3H_3O_2$ $C_5H_3O_2$ $C_6H_4O_2$ $C_7H_4O_3$
Apigenin	$C_{15}H_{10}O_5$	269.0455	269.0453	8.24	0	149.0248 151.0037	$C_8H_5O_3$ $C_7H_3O_4$
Caffeic acid	$C_9H_8O_4$	ND	-	-	-	-	-
Epicatechin	$C_{15}H_{14}O_6$	289.0716	289.0716	4.37	+0.02	137.0248	$C_7H_5O_3$

Compound	**Molecular Formula**	**$[M-H]^-$ m/z std.**	**$[M-H]^-$ m/z exp.**	**t_R (min)**	**Δt_R (min)**	**Fragments *m/z***	**Elemental Formula**
						151.0416	$C_8H_7O_3$
Ethyl vanillin	$C_9H_{10}O_3$	165.0557	165.0558	5.44	-0.01	67.0190	C_4H_3O
						92.0266	C_6H_4O
						108.0215	$C_6H_4O_2$
						137.0245	$C_7H_5O_3$
Ferulic acid	$C_{10}H_{10}O_4$	193.0506	193.0506	1.40	+0.01	134.0370	$C_8H_6O_2$
						178.0271	$C_9H_6O_4$
Gallic acid	$C_7H_6O_5$	169.0142	169.0140	1.25	+0.01	125.0246	$C_6H_5O_3$
Homovanillic acid	$C_9H_{10}O_4$	181.0506	181.0506	1.50	+0.02	59.0134	$C_2H_3O_2$
						122.0369	$C_7H_6O_2$
						137.0610	$C_8H_9O_2$
						154.0266	$C_7H_6O_4$
Hydroxytyrosol	$C_8H_{10}O_3$	153.0557	153.0557	3.53	+0.01	123.0446	$C_7H_7O_2$
Oleuropein	$C_{25}H_{32}O_{13}$	539.1770	539.1770	5.96	0	59.0138	$C_2H_3O_2$
						89.0244	$C_3H_5O_3$
						101.0242	$C_4H_5O_3$
						111.0083	$C_5H_3O_3$
						121.0295	$C_7H_5O_2$
						307.0823	$C_{15}H_{15}O_7$
						377.1242	$C_{19}H_{21}O_8$
p-Coumaric acid	$C_9H_8O_3$	163.0400	163.0402	1.34	0	119.0506	C_8H_7O
						93.0349	C_6H_5O
Pinoresinol	$C_{20}H_{22}O_6$	357.1343	357.1340	6.49	+0.01	151.0399	$C_8H_7O_3$
Quercetin	$C_{15}H_{10}O_7$	301.0353	301.0350	7.42	-0.04	121.0294	$C_7H_5O_2$
						151.0024	$C_7H_3O_4$
Syringic acid	$C_9H_{10}O_5$	197.0455	197.0454	1.44	+0.01	182.0218	$C_8H_6O_5$
Tyrosol	$C_8H_{10}O_2$	137.0608	137.0610	4.07	-0.01	81.0262	C_5H_5O
						93.0345	C_6H_5O
						112.0530	$C_6H_8O_2$

exp: experimental, std: standard, ND: not detected

After verification and confirmatory analysis representative qualifier ions of the detected compounds were cross-verified with fragments presented in previous works. The MS/MS fragments of the target compounds are reported in **Table 3.5**. The MS/MS spectrum of quercetin shows fragment at m/z: 151.0024 corresponding to $C_7H_3O_4$ [72, 79]. The qualifier ions of oleuropein m/z: 307.0823 and m/z: 377.1241 corresponding to $C_{15}H_{15}O_7$ and $C_{19}H_{21}O_8$, respectively, have been also reported by Kanakis et al. [31]. Pinoresinol shows characteristic fragmentation at m/z: 151.0399, corresponding to $C_8H_7O_3$ [71]. The EICs of the target compounds identified are presented in **Figure 3.2**.

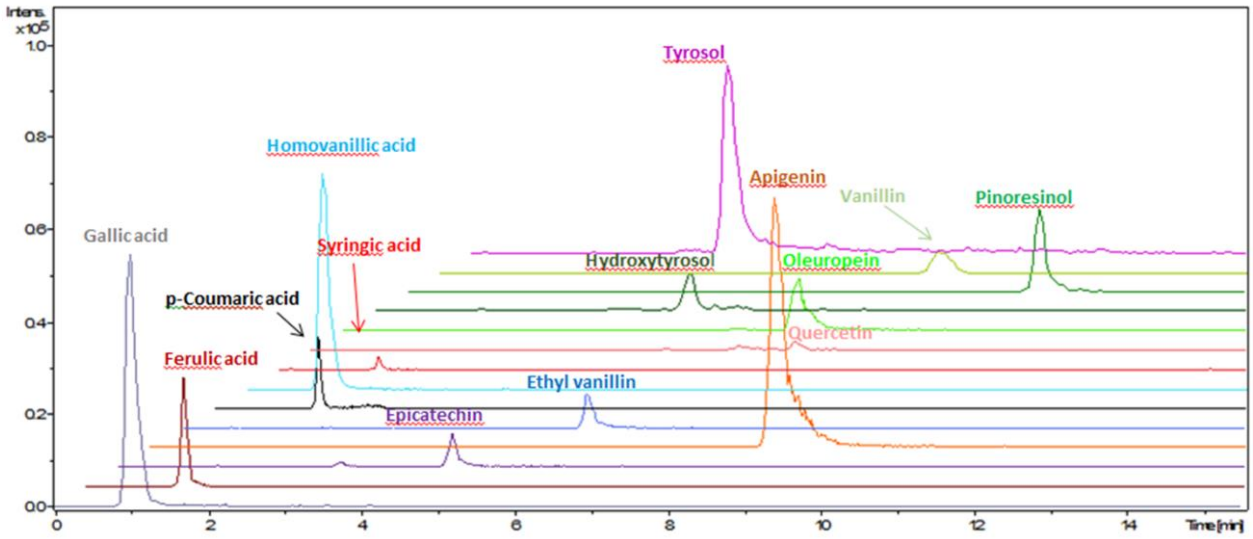


Figure 3.2: EICs of the target analytes in an EVOO sample

All the analytical parameters including precision (RSD%), limits of detection (LOD) and quantification (LOQ), linearity (calibration curves and regression coefficient (r^2)), as well as recoveries (RE%), matrix factor (MF) and matrix effect (ME%) were calculated and are presented in **Table 3.6.**

Table 3.6: Results of the validation of the target screening method

Compound	LOD (mg kg^{-1})	LOQ (mg kg^{-1})	Intra-day precision RSD_r% (n=10)	Inter-day precision RSD_R% (n=2×3)	Equation y= (a±Sa)+(b±Sb)x linear range: 1-5 mg kg^{-1}	r^2	RE %	MF	ME%
Gallic acid	0.029	0.091	4.5	5.2	y = (0.02±0.01)+(0.11±0.02)x	0.995	104	0.98	-2.23
p-Coumaric acid	0.030	0.092	1.2	4.1	y = (0.01±0.02)+(0.03±0.01)x	0.993	97.1	0.94	-6.39
Ferulic acid	0.034	0.086	2.2	2.7	y = (0.03±0.02)+(0.032±0.005)x	0.995	101	0.95	-4.89
Syringic acid	0.029	0.090	2.1	2.4	y = (0.006±0.002)+(0.07±0.01)x	0.992	101	0.94	-6.28
Homovanillic acid	0.027	0.089	2.4	6.3	y = (0.001±0.002)+(0.02±0.01)x	0.995	100	0.94	-6.41
Tyrosol	0.030	0.091	4.1	3.2	y = (-0.006±0.003)+(0.041±0.004)x	0.997	97.3	0.96	-4.11
Hydroxytyrosol	0.034	0.077	2.9	4.3	y = (-0.022±0.004)+(0.09±0.06)x	0.997	105	0.95	-4.92
Pinoresinol	0.023	0.075	2.1	2.7	y = (-0.005±0.004)+(0.03±0.01)x	0.997	107	0.92	-8.09
Apigenin	0.015	0.046	3.5	4.1	y = (0.10±0.07)+(0.52±0.37)x	0.996	102	0.95	-4.41
Oleuropein	0.022	0.068	4.9	2.9	y = (-0.019±0.002)+(0.13±0.02)x	0.998	98.1	0.94	-6.44
Vanillin	0.039	0.078	1.1	4.9	y = (-0.02±0.01)+(0.11±0.07)x	0.996	96.4	0.93	-7.47
Ethyl vanillin	0.028	0.084	3.7	6.3	y = (-0.027±0.001)+(0.161±0.004)x	0.996	99.5	0.96	-4.20
Epicatechin	0.029	0.088	4.8	5.5	y = (-0.02±0.01)+(0.096±0.003)x	0.993	96.7	0.94	-6.57
Quercetin	0.023	0.069	2.9	4.3	y= (-0.005±0.001)x+(0.03±0.02)x	0.995	98.0	0.91	-9.41

LOD: limit of detection, LOQ: limit of quantification, r^2: regression coefficient, RSD: relative standard deviation, RE: recovery, MF: matrix factor, ME: matrix effect

Calibration curves were constructed and were linear with r^2 > 0.99 in all cases. Precision limit was ≤ 4.9% RSD for intra-day experiments and ≤ 6.4% RSD for inter-day experiments, indicating the good precision of the method developed. LODs and LOQs were adequate and ranged between 0.015 (apigenin) - 0.034 (vanillin) mg kg^{-1} and 0.046 - 0.091 mg kg^{-1}, respectively. Analytes showed satisfying recovery efficiency (between 96 and 108%). Low matrix suppression was observed for all the phenolic compounds (ME%: 2.23-9.41).

In the next step, the target compounds detected were quantified in all samples; taking into consideration that quantitative analysis is crucial to offer a comprehensive overview of the phenolic composition of EVOOs. The concentrations of ferulic acid, gallic, homovanillic, p-coumaric and syringic acid, tyrosol and hydroxytyrosol, vanillin and ethyl vanillin, apigenin, luteolin, quercetin and epicatechin, pinoresinol and oleuropein were quantified from their corresponding calibration curves using caffeic acid as IS. Quantitative results of the target compounds can be found in the **ESM I, Table S2 (**results are expressed as the mean values ± SD (n=3), in mg kg^{-1})).

3.3.2 Suspect screening

In suspect screening, 26 bioactive compounds out of the 60 of the initial suspect list were tentatively identified in real olive oil samples with ion intensities above 800 and peak areas more than 2000 in all cases. The results showed high mass accuracy (< 5 ppm) and acceptable isotopic fit values (< 50 mSigma). Peak score describes the Peak Area/Peak Intensity ratio and was calculated over the range 4-19 for all the suspect compounds [81]. MS/MS spectra were examined and verified with MetFrag [91] as well as literature records. **Table 3.7** summarizes the suspect screening results, providing information about the identification criteria and the level of identification of each compound. QSSR was employed for the prediction of the possible retention time in cases that no reference standards were available. The difference between the experimental retention time and the predicted was less than 1 min for all the suspect compounds, except for syringaresinol (2 min difference).

Table 3.7: Identified compounds through suspect screening, identification criteria and level of identification

Compound	Molecular Formula	$[M-H]^-$ m/z calculated	$[M-H]^-$ m/z experimental	t_R (min)	t_R (min) (predicted)	Fragments m/z	Elemental Formula	Peak score A/I	Level Ident.
Elenolic acid	$C_{11}H_{14}O_6$	241.0714	241.0714	4.51	4.26	59.0137 95.0496 127.0400 151.0402 171.0300	$C_2H_3O_2$ C_6H_7O $C_6H_7O_3$ $C_8H_7O_3$ $C_7H_7O_5$	10	2a
Elenolic acid Isomer	$C_{11}H_{14}O_6$	241.0722	241.0714	3.92	*	95.0494	C_6H_7O	4	3
Elenolic acid Isomer II	$C_{11}H_{14}O_6$	241.0722	241.0714	1.31	*	95.0494	C_6H_7O	6	3
Elenolic acid methyl ester	$C_{12}H_{16}O_6$	255.0874	255.0874	4.58	4.53	67.0189 69.0346 185.0455	C_4H_3O C_4H_5O $C_8H_9O_5$	8	2a
Luteolin	$C_{15}H_{10}O_6$	285.0404	285.0404	7.55	Confirmed with a standard	133.0295 151.0036	$C_8H_5O_2$ $C_7H_3O_4$	12	1
Hydroxylated form of elenolic acid	$C_{11}H_{14}O_7$	257.0667	257.0667	1.36	*	59.0104 137.0603 181.0535	$C_2H_3O_2$ $C_8H_9O_2$ $C_9H_9O_4$	11	2a

Compound	**Molecular Formula**	**[M-H]$^-$ m/z calculated**	**[M-H]$^-$ m/z experimental**	**t_R (min)**	**t_R (min) (predicted)**	**Fragments m/z**	**Elemental Formula**	**Peak score A/I**	**Level Ident.**
Hydroxylated form of elenolic acid isomer	$C_{11}H_{14}O_7$	257.0667	257.0665	1.41	*	-	-	8	3
Hydroxytyrosol acetate	$C_{10}H_{12}O_4$	195.0660	195.0662	6.71	6.48	134.0373 149.0608 161.0246	$C_8H_6O_2$ $C_9H_9O_2$ $C_9H_5O_3$	15	2b
Hydroxytyrosol acetate isomer	$C_{10}H_{12}O_4$	195.0660	195.0662	5.74	6.48	59.0136 134.0373 161.0246	$C_2H_3O_2$ $C_8H_6O_2$ $C_9H_5O_3$	12	3
Decarboxymethyl oleuropein aglycone	$C_{17}H_{20}O_6$	319.1185	319.1185	5.61	6.14	69.0342 95.0502 123.0451 139.0608 165.0556 183.0660 195.0656	C_4H_5O C_6H_7O $C_7H_7O_2$ $C_8H_{11}O_2$ $C_9H_9O_3$ $C_9H_{11}O_4$ $C_{10}H_{11}O_4$	16	2a
Decarboxymethyl lingstroside aglycone	$C_{17}H_{20}O_5$	303.1237	303.1237	6.42	6.76	124.0531 137.0608 147.0453 165.0556	$C_7H_8O_2$ $C_8H_{10}O_2$ $C_9H_7O_2$ $C_9H_9O_3$	15	2a

Compound	**Molecular Formula**	**[M-H]$^-$ m/z calculated**	**[M-H]$^-$ m/z experimental**	**t_R (min)**	**t_R (min) (predicted)**	**Fragments m/z**	**Elemental Formula**	**Peak score A/I**	**Level Ident.**
						183.0662	$C_9H_{11}O_4$		
10-Hydroxy oleuropein aglycone	$C_{19}H_{22}O_9$	393.1193	393.1191	4.82	5.48	137.0244 181.0502	$C_7H_5O_3$ $C_9H_9O_4$	8	2b
Oleuropein aglycone Isomer	$C_{19}H_{22}O_8$	377.1241	377.1241	5.88	6.88	111.0087 149.0241 263.0926 275.0923	$C_5H_3O_3$ $C_8H_5O_3$ $C_{14}H_{15}O_5$ $C_{15}H_{15}O_5$	16	3
Oleuropein aglycone	$C_{19}H_{22}O_8$	377.1241	377.1247	7.29	6.88	111.0088 149.0244 195.0644 275.0919 307.0823	$C_5H_3O_3$ $C_8H_5O_3$ $C_{10}H_{11}O_4$ $C_{15}H_{15}O_5$ $C_{15}H_{15}O_7$	19	2a
Lingstroside aglycone	$C_{19}H_{22}O_7$	361.1291	361.1293	6.63	6.83	259.0975 291.0875	$C_{15}H_{15}O_4$ $C_{15}H_{15}O_6$	19	2a
Syringaresinol	$C_{22}H_{26}O_8$	417.1554	417.1556	6.18	8.10	127.0408 181.0503	$C_6H_7O_3$ $C_9H_9O_4$	9	2b

Compound	**Molecular Formula**	**[M-H]$^-$ m/z calculated**	**[M-H]$^-$ m/z experimental**	**t_R (min)**	**t_R (min) (predicted)**	**Fragments m/z**	**Elemental Formula**	**Peak score A/I**	**Level Ident.**
Oleoside	$C_{16}H_{22}O_{11}$	389.1089	389.1087	7.91	*	113.0244 139.0032 149.0240 165.0552 183.0666	$C_5H_5O_3$ $C_6H_3O_4$ $C_8H_5O_3$ $C_9H_9O_3$ $C_9H_{11}O_4$	12	2a
Oleoside Isomer	$C_{16}H_{22}O_{11}$	389.1089	389.1087	5.7	*	165.0552	$C_9H_9O_3$	11	3
1-Hydroxypinoresinol	$C_{20}H_{22}O_7$	373.1292	373.1292	6.39	6.39	121.0294 151.0401 163.0402	$C_7H_5O_2$ $C_8H_7O_3$ $C_9H_7O_3$	8	2b
1-Hydroxypinoresinol Isomer	$C_{20}H_{22}O_7$	373.1292	373.1294	6.42	*	-	-	9	3
Oleanolic acid	$C_{30}H_{48}O_3$	455.3535	455.3540	13.94	12.69**	44.9980 120.0940 152.1202 407.3316	CHO_2 C_9H_{12} $C_{10}H_{16}O$ $C_{29}H_{43}O$	8	2a
Maslinic acid	$C_{30}H_{48}O_4$	471.3484	471.3485	12.82	12.58	44.9982 405.3156 423.3423	CHO_2 $C_{29}H_{41}O$ $C_{29}H_{43}O_2$	12	2a

Compound	**Molecular Formula**	**[M-H]$^-$ m/z calculated**	**[M-H]$^-$ m/z experimental**	**t_R (min)**	**t_R (min) (predicted)**	**Fragments m/z**	**Elemental Formula**	**Peak score A/I**	**Level Ident.**
1-Acetoxypinoresinol	$C_{22}H_{24}O_8$	415.1398	415.1399	6.42	7.20	151.0402 280.0951 343.1188	$C_8H_7O_3$ $C_{14}H_{16}O_6$ $C_{19}H_{19}O_6$	10	2b
Methyl oleuropein aglycone	$C_{20}H_{24}O_8$	391.1412	391.1418	7.51	7.37	59.0140 67.0192 99.0456 111.0087 137.0608 291.0875	$C_2H_3O_2$ C_4H_3O $C_5H_7O_2$ $C_5H_3O_3$ $C_8H_9O_2$ $C_{16}H_{15}O_6$	6	2b
10-Hydroxy-10-methyl oleuropein aglycone	$C_{20}H_{24}O_9$	407.1347	407.1347	6.71	6.75	99.0453 111.0087 121.0295 135.0453 137.0243 149.0245 163.0402 179.0351 195.0665 241.0871	$C_5H_7O_2$ $C_5H_3O_3$ $C_7H_5O_2$ $C_8H_7O_2$ $C_7H_5O_3$ $C_8H_5O_3$ $C_9H_7O_3$ $C_9H_7O_4$ $C_{10}H_{11}O_4$ $C_{15}H_{13}O_3$	19	2b

Compound	**Molecular Formula**	**[M-H]$^-$ m/z calculated**	**[M-H]$^-$ m/z experimental**	**t_R (min)**	**t_R (min) (predicted)**	**Fragments m/z**	**Elemental Formula**	**Peak score A/I**	**Level Ident.**
10-Hydroxy decarboxymethyl oleuropein aglycone	$C_{17}H_{20}O_7$	335.1150	335.1151	4.28	5.52	59.0139 85.0296 121.0292 151.0401 153.0557 155.0716 199.0613	$C_2H_3O_2$ $C_4H_5O_2$ $C_7H_5O_2$ $C_8H_7O_3$ $C_8H_9O_3$ $C_8H_{11}O_3$ $C_9H_{11}O_5$	8	2a

Peak Score (A/I): ratio of peak area to intensity

Level Ident.: Level of identification (Level 1 corresponds to confirmed structures where a reference standard is available; level 2a: evidence by spectra matching from literature or library and level 2b: diagnostic evidence where no other structure fits the experimental MS/MS information; level 3 for tentative candidate)

*t_R: The retention time prediction results are not reliable and other verification methods such as MS/MS fragmentation pattern should be applied.

**t_R: The retention time prediction result is not reliable because oleanolic acid is found to be outside of applicability domain of the model.

The initial suspect list mainly consisted of all the possible secoiridoid derivativies of the oleuropein complex, due to the fact that oleuropein is the major secoiridoid found in the pulp of olives and its concentration decreases during the maturation process to form derivatives. Oleuropein aglycone, 10-hydroxy oleuropein aglycone, methyl oleuropein aglycone, 10-hydroxy-10-methyl oleuropein aglycone, 10-hydroxy decarboxymethyl oleuropein aglycone, lingstroside aglycone, decarboxymethyl oleuropein aglycone (oleacein) and decarboxymethyl lingstroside aglycone (oleocanthal) were tentatively identified at level 2 (level 2a or level 2b as it is summarized in **Table 3.7**). The EICs of the identified secoiridoids are illustrated in **Figure 3.3**. One isomer of oleuropein aglycone was identified at level 3. The identification of oleuropein aglycone, lingstroside aglycone, oleacein and oleocanthal is of high importance because they have been correlated with the positive attributes of bitter and pungent taste [76]. Moreover, oleacein and oleocanthal are both considered important due to their decisive role in health protection [24]. Studies have demonstrated that oleacein exhibits anti-inflammatory and antimicrobial activities, skin protection and reduction of disorder due to the metabolic syndrome [82], whereas oleocanthal presents breast anticancer and potent antioxidant activity [41]. The precursor ions of oleacein and oleocanthal were detected at m/z: 319.1185 and m/z: 303.1237, respectively. Both compounds present one single broad peak in the EICs of the full scan (AutoMS). The qualifier ions detected at m/z: 69.0342, 95.0502, 139.0608, 183.0660 and 195.0656 correspond to C_4H_5O, C_6H_7O, $C_8H_{11}O_2$, $C_9H_{11}O_4$ and $C_{10}H_{11}O_4$, respectively, and they have also been reported by Dierkes et al. [76]. The peak at m/z: 165.0556 corresponds to $C_9H_9O_3$ [31, 79]. The MS/MS spectrum of oleacein is presented in **Figure 3.3**.

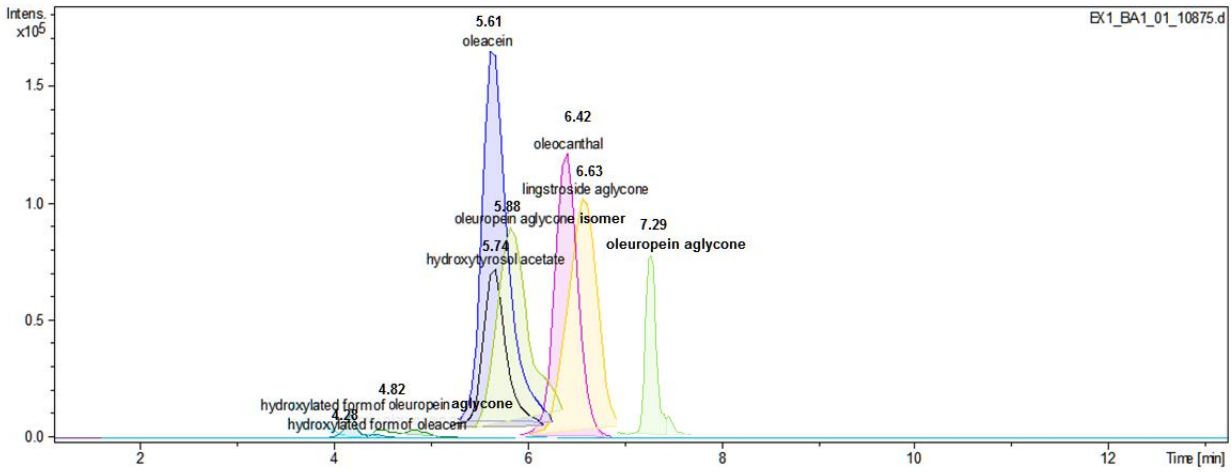


Figure 3.3: EICs of the suspect secoiridoids identified in an EVOO sample

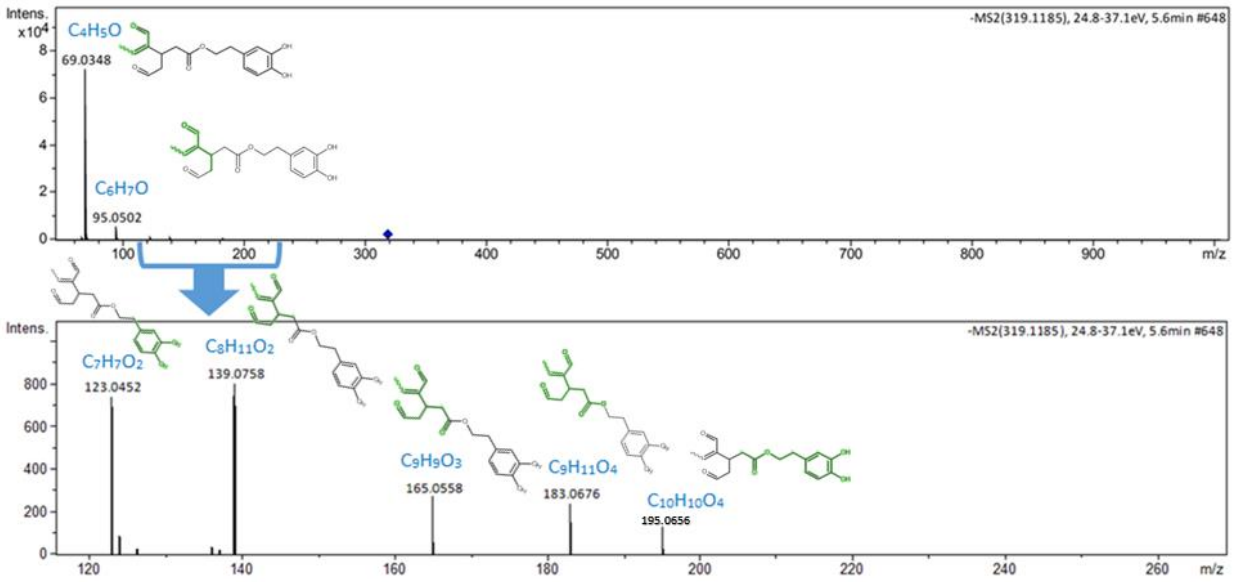


Figure 3.4: MS/MS spectrum of oleacein

As regards the MS/MS spectrum of oleocanthal, it presents a similar fragment with oleacein at m/z: 165.0556 corresponding to $C_9H_9O_3$ [76, 79]. Moreover, the peak at m/z: 183.0662 corresponds to $C_9H_{11}O_4$ and been reported in previous works [76]. Oleuropein aglycone (identification level 2a) eluted as two different peaks (with retention times t_R=5.86 min and t_R=7.21 min), suggesting the existence of an isomer (identification level 3). Three qualifier ions were the same for both compounds (111.0087 and 111.0088 [83], 149.0241 [83] and 149.0244

[76], 275.0919 and 275.0923, which have been reported by Kanakis et al. [31] and Dierkes et al. [76], corresponding to $C_5H_3O_3$, $C_8H_5O_3$ and $C_{15}H_{15}O_5$, respectively). The fragments m/z 195.0644, corresponding to $C_{10}H_{11}O_4$, has been reported by Kanakis et al. [31] and Dierkes et al. [76]. Lingstroside aglycone (identification level 2a) presents two qualifier ions 259.0975 and 291.0875 [31, 76] corresponding to $C_{15}H_{15}O_4$ and $C_{15}H_{15}O_6$, respectively. As for 10-hydroxy decarboxymethyl oleuropein aglycone (identification level 2a), the fragment at m/z 199.0614 corresponding to $C_9H_{11}O_5$ has already been reported by Kanakis et al. [31].

Elenolic acid, which is a non-phenol and has been described as marker of maturation of olives [132] was identified at level 2a. The MS/MS spectrum of elenolic acid is presented in **Figure 3.5**. Elenolic acid methyl ester and the hydroxylated form of elenolic acid were also tentatively identified at level 2a. For the hydroxylated form of elenolic acid, four fragments were recorded and 137.0603 corresponding to $C_8H_9O_2$ as well as 181.0535 matching $C_9H_9O_4$ have also been suggested by Capriotti et al. [83]. In addition, two isomers of elenolic acid and one isomer of the hydroxylated form of elenolic acid were identified at level 3.

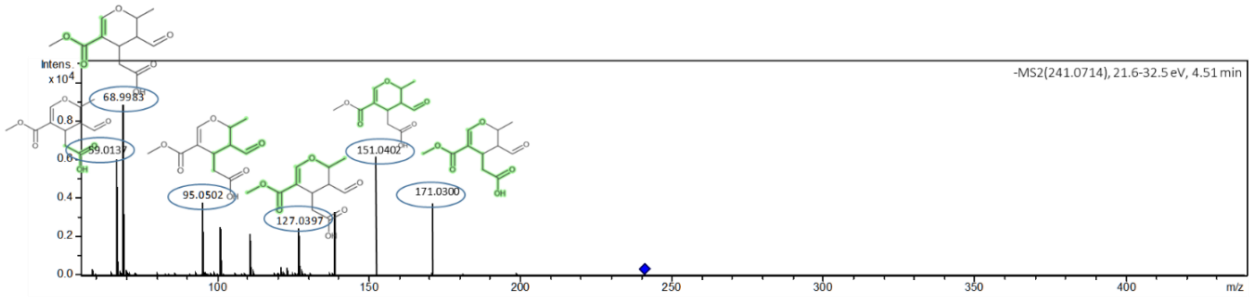


Figure 3.5: MS/MS spectrum of elenolic acid

From the class of phenethyl alcohols**,** two hydroxytyrosol derivativies were detected, hydroxytyrosol acetate (identification level 2b) and one isomer of hydroxytyrosol acetate (identification level 3). **Figure 3.6** illustrates the EIC of hydroxytyrosol acetate, showing two broad peaks; one for hydroxytyrosol acetate eluting at 6.71 min and one for its isomer eluting at 5.74 min, and the MS/MS spectra of both compounds.

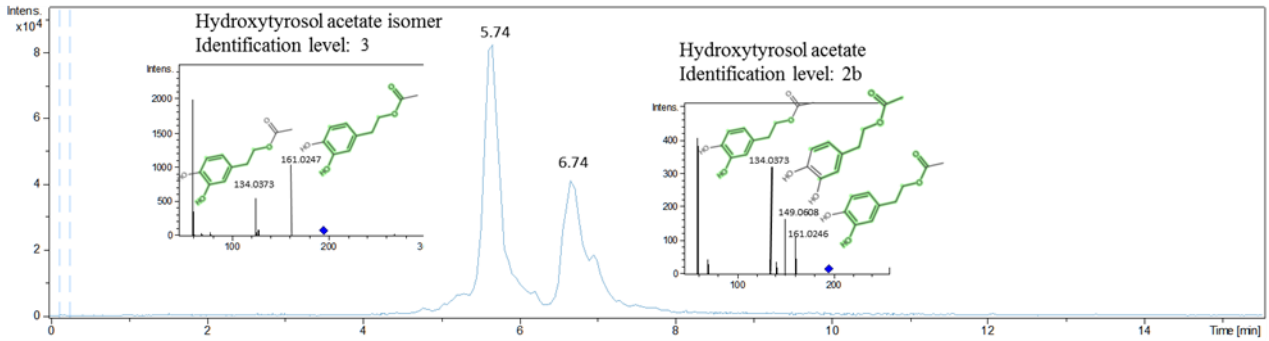


Figure 3.6: EIC and MS/MS spectra of hydroxytyrosol acetate (t_R=6.71 min) and its isomer (t_R=5.74 min)

Next, from the class of lignans, which have been suggested as varietal markers [24, 25] and present antiviral activities [24], 1-hydroxypinoresinol and 1-acetoxypinoresinol, as well as syringaresinol were identified at level 2b. One isomer of 1-hydroxypinoresinol was identified at level 3. **Figure 3.7** shows the EICs of the identified lignans.

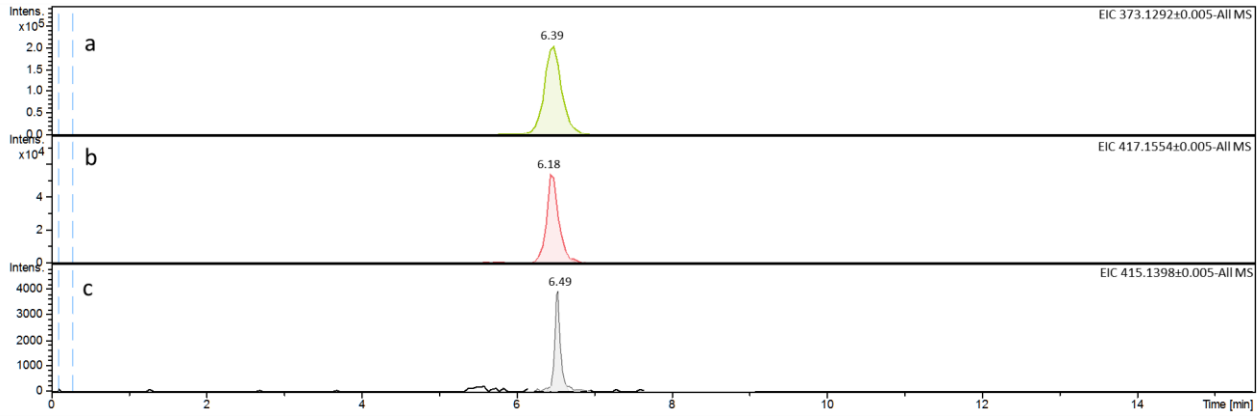


Figure 3.7: EICs of (a) hydroxypinoresinol; (b) acetoxypinoresinol (c) syringaresinol

From the class of flavonoids, the presence of luteolin was confirmed with a reference standard (identification level 1). Finally, the triterpenic acids, oleanolic and maslinic acids were identified at level 2a. Oleoside was identified at level 2a and one isomer of oleoside was identified at level 3.

After the determination, the suspect compounds belonging to the classes of lignans, flavonoids and secoiridoids were semi-quantified on the basis of target compounds having similar structures, as it has been suggested in previous works [9, 33, 71, 72]. 1-Acetoxypinoresinol, 1-Hydroxypinoresinol and syringaresinol

were semi-quantified using the pinoresinol calibration curve. Hydroxytyrosol acetate was semi-quantified using the hydroxytyrosol calibration curve and all the secoiridoids were semi-quantified using oleuropein calibration curve. Semi-quantification results are presented in **ESM I, Table S3** (concentrations are expressed in mg kg^{-1}, as the mean values with ± standard deviation (n=3)**)**.

A remarkable difference is observed in the total phenolic content between the EVOOs and the defective samples. The total phenolic content of the EVOOs ranged between 222-318 mg kg^{-1}, whereas the defectives (RB1, BR1, PK1, KA1, NB1, fusty, musty and rancid) demonstrate impressively low values between 47 and 136 mg kg^{-1}. Therefore, it can be concluded that the taste of olive oil is in direct relationship with the concentration of the phenolic compounds. The total phenolic content of each olive oil sample (sum of the quantified target compounds plus the semi-quantified concentrations of the suspect phenolic compounds) is presented in **Figure 3.8** (concentrations are expressed in mg kg^{-1}, as the mean values with ± standard deviation (n=3)).

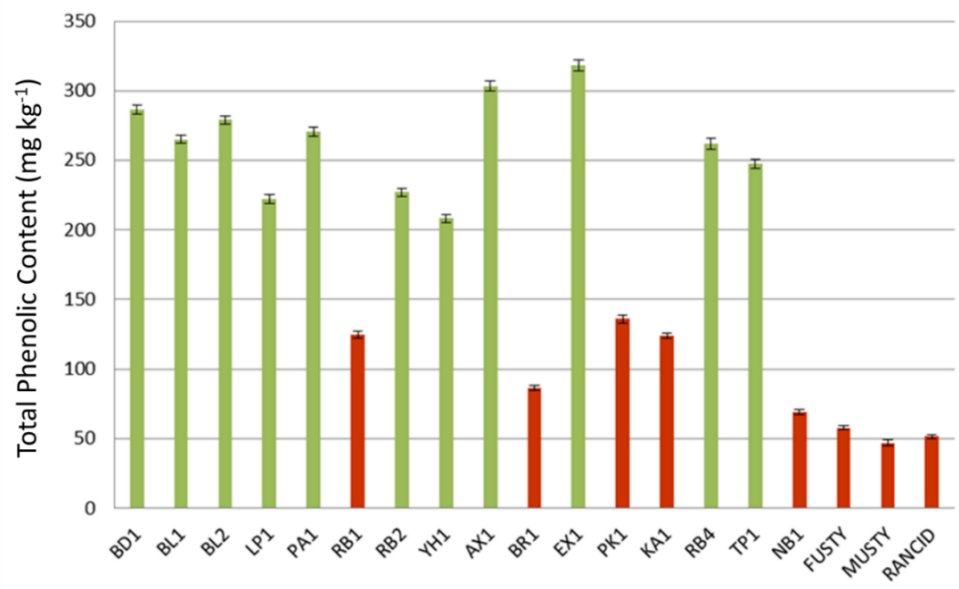


Figure 3.8: Total phenolic content (mean values ± SD (n=3)) of the analysed samples (EVOOs in green; defective samples in red)

3.3.2.1 QSRR applicability domain study

A novel tool was developed to visualize the correlation between the retention times, chemical structure similarity, and standardized residuals to fully

understand the origin of residuals between the experimental and predicted retention time. The applicability domain tool was applied to analyze the predicted t_R.

For a compound that has high chemical structure similarity with the training set used for modeling purposes, the error of (or higher than) ±2 min is a relatively high value, while for a compound with lower chemical structure similarity, this threshold is accepted. Therefore, standardized residual values, which incorporate the effect of the structure diversity of the suspect compounds, were used. These values can result in better justification of whether retention time prediction error is due to wrong structure or simply the proposed structure is outside the applicability domain and the retention time prediction cannot be used to support the identification of a suspect compound. The exported results are drawn as bubble plot (**Figure 3.9**) and reveal whether the observed residual is due to the wrong chemical structure (leading to wrong prediction of retention time) or the tested structure is out of the applicability domain of the model. Three regions are obvious in the bubble plots, corresponding to different levels of acceptance for the predicted retention time. Boxes 1 and 2 are the areas of acceptance of predicted retention time, while box 3 shows the region where the residuals are high and the predicted retention time is questioned. The bubble size indicates the chemical structure diversity; the larger it gets the larger the chemical structure diversity becomes. Sometimes, despite the large leverage values (chemical structural diversity), a prediction model may predict t_R with lower error. In this respect, additional parameters are required to study whether the training set is representative for the suspect or not. This could make a justification to reject or accept the compounds with high leverages even if the predicted t_R is located in box 1 and box2. Normalized mean distance was used to show if the used training set is representative for the tested compound and to provide a justification about the compounds with large bubble size [94].

The results derived (based on the so called OTrAMS technique) led to the rejection of two compounds (oleoside isomer and 1-hydroxypinoresinol isomer). These isomers are located in box 4 with accepted leverage values suggesting that these compounds cannot have the elution pattern as recorded. Moreover, six

compounds (elenolic acid, elenolic acid isomer, elenolic acid isomer II, hydroxylated form of elenolic acid, hydroxylated form of elenolic acid isomer and oleoside) with acceptable leverages are located in box 3, suggesting that the prediction results are not reliable and other verification methods such as MS/MS fragmentation pattern should be applied. Moreover, oleanolic acid was also found to be outside of applicability domain of the model despite a good prediction result. Therefore, the training set used to build the model is not representative for this compound and t_R information and the results of prediction should be used as a support for increasing the verification level. The rest of compounds listed as suspects showed to be within box 1 and 2, and the t_R information with the prediction results can be used for any justifications.

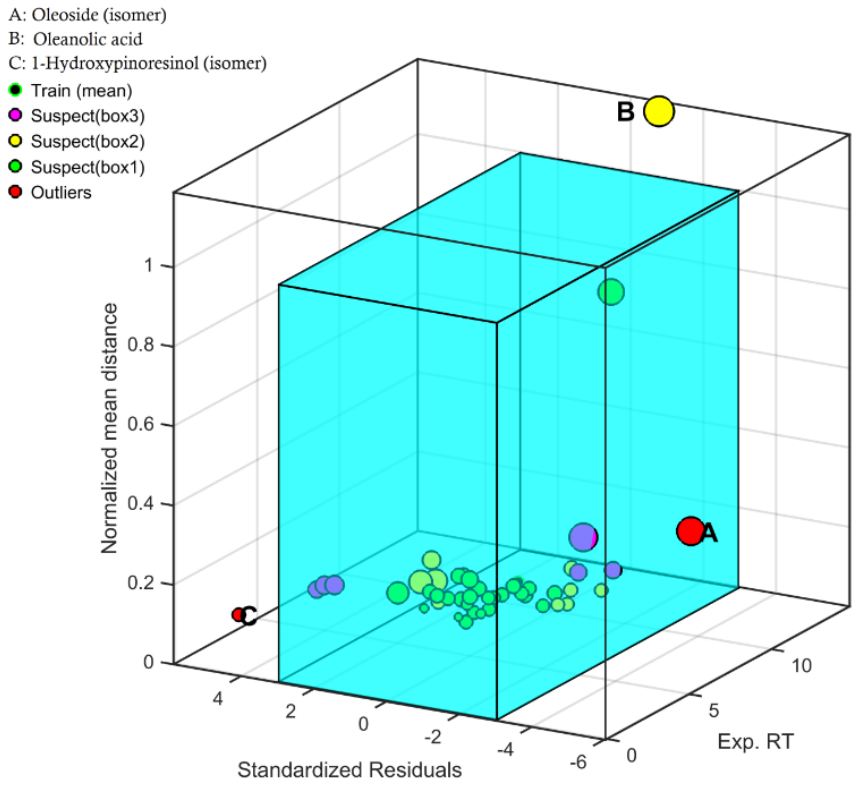


Figure 3.9: The applicability domain study for the predicted t_R of the suspect compounds

3.3.3 Non-target screening

Non-target screening workflow resulted in the generation of 304 features. Then, Variable Importance in Projection (VIP) was applied in order to distinguish the

most important compounds responsible for discrimination. In order to prioritize the peaks which cause greater variation in the discrimination between samples, VIP values were calculated for the PLS-DA model [149] and those with VIP score greater than 0.83 were considered as the most important [120, 147, 149]. 151 Out of the 304 features, 151 were calculated with a VIP score above 0.83 and were followed with subsequent analysis. 15 out of the 151 important compounds already existed in the target and suspect list (tyrosol, hydroxytyrosol, apigenin, oleuropein, ethyl vanillin, elenolic acid, decarboxymethyl oleuropein aglycone (oleacein), decarboxymethyl lingstroside aglycone (oleocanthal), hydroxytyrosol acetate, 10-hydroxy oleuropein aglycone, oleuropein aglycone, lingstroside aglycone, 10-hydroxy-10-methyl oleuropein aglycone, methyl oleuropein aglycone and 10-hydroxy decarboxymethyl oleuropein aglycone) and were, subsequently, excluded from the non-target list.

In an effort to identify the remaining 136 masses, an inclusion list consisting of their precursor ions m/z was created and the QTOF operated in Auto MS/MS mode, in order to obtain the MS/MS spectra of the unknown analytes. From the 136 non-targets, 7 compounds were successfully identified and confirmed with reference standards (identification level 1). These compounds were hexanoic acid, octanoic acid, palmitic acid, α-linolenic, α-linoleic, oleic, as well as arachidic acid. What is more interesting is that their EICs revealed great variations in the intensities of the free fatty acids among the samples. The most significant variations were observed in the EICs of hexanoic and octanoic acid as they are presented in **Figure 3.10(a)** and **Figure 3.10(b)**, respectively. Their presence in defective samples proves that they can be established as markers. The intensities of hexanoic and octanoic acid were high for rancid and RB1 (which is a defective olive oil characterized as musty) and minimum for the EVOOs. On the contrary, the peaks of the rest of the free fatty acids, palmitic acid, linolenic, linoleic, oleic and arachidic acid were most intense in the EVOOs.

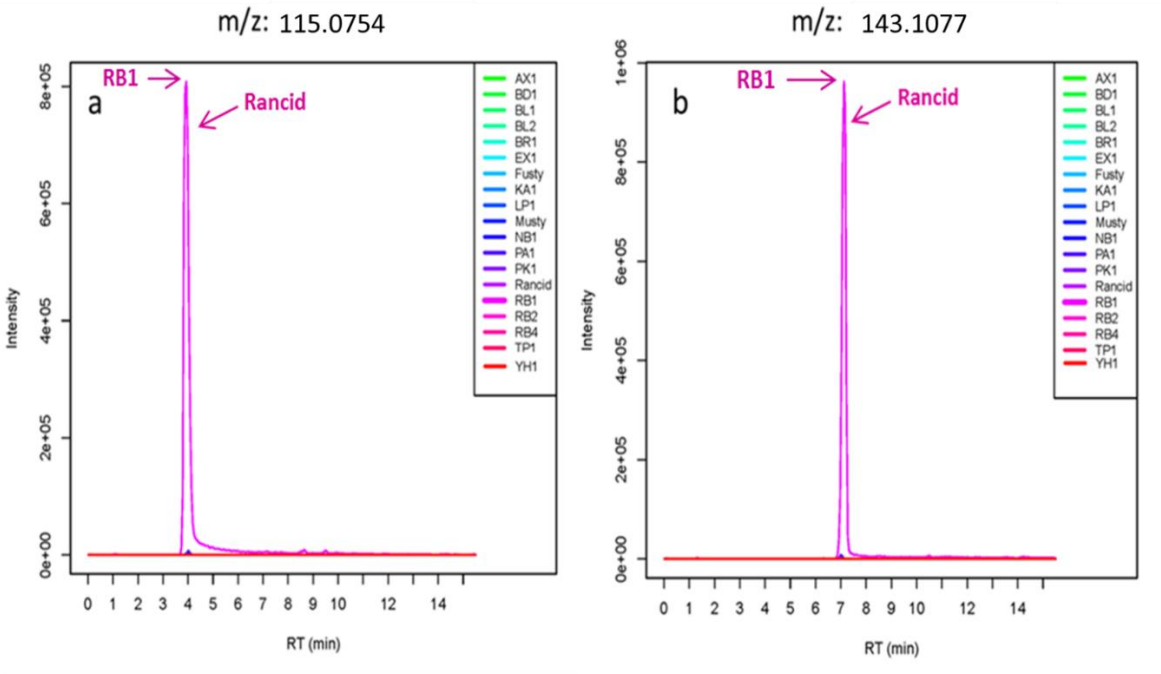


Figure 3.10: (a) EIC of hexanoic acid in the analyzed samples; (b) EIC of octanoic acid in the analyzed samples. Both acids are markers for the defective olive oils

Following the non-target procedure, cinnamic acid and quinic acid were tentatively identified at level 2a. The predicted t_R was very close to the experimental for both compounds (cinnamic acid: experimental t_R= 6.37 min and predicted t_R=6.38 min; quinic acid: experimental t_R= 1.12 min and predicted t_R=1.37 min) and the MS/MS fragments were matched and confirmed using MassBank [92] records. **Figure 3.11** presents the EIC and MS/MS spectrum of quinic acid.

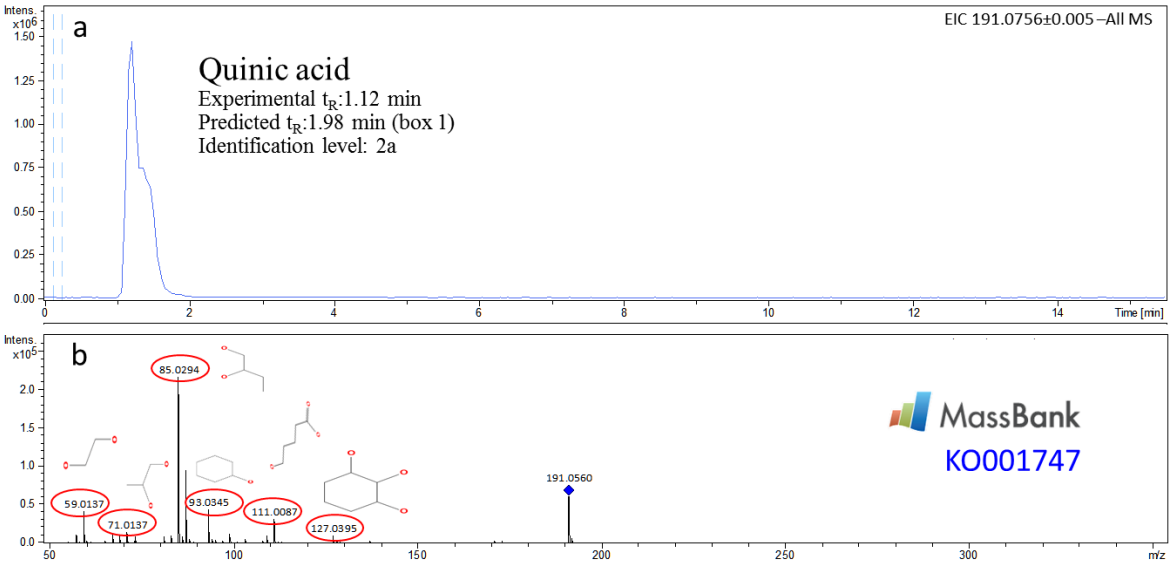


Figure 3.11: (a) EIC of quinic acid, marker in EVOOs; (b) MS/MS spectrum of quinic acid, the fragments were matched with MassBank (record: KO001747)

In addition, sinapic acid and acetosyringone were identified at level 2b. **Table 3.8** summarizes the identification results for the compounds that were identified at level 1 and level 2, including the identification levels, the experimental and predicted t_R as well as the list of fragments and tentative structures. For the rest of the features selected by VIP, an unequivocal molecular formula was assigned (identified at level 4) for 24 compounds. In these cases, the MS/MS spectra were not informative enough to proceed in further identification. The identification attempts for the compounds identified at level 4 are presented in the **ESM I, Table S4.**

Table 3.8: Identification of non-target compounds at levels of identification 1 and 2

m/z	Ident. Level	Unequivocal Molecular Formula	t_R (min)	Pred. t_R (min)	ChemSpider Hits	Fragments m/z	Tentative Structure SMILES	Peaks explained (m/z), Elemental Formula
255.2334	1	$C_{16}H_{32}O_2$ **Palmitic acid**	13.81	11.62	329	59.0132 83.0502 236.2147 237.2225	CCCCCCCCCCCCCCCC(=O)O	59.0132, $C_2H_3O_2$ 83.0502, C_5H_7O 236.2147, $C_{16}H_{28}O$ 237.2225, $C_{16}H_{29}O$
115.0754	1	$C_6H_{12}O_2$ **Hexanoic acid**	4.12	5.90	465	44.9982 59.0139	CCCCCC(=O)O	44.9982, CHO_2 59.0139, $C_2H_3O_2$
143.1077	1	$C_8H_{16}O_2$ **Octanoic acid**	7.18	10.03	905	44.9982 59.0139 71.0138 205.1963 233.2273	CCCCCCCC(=O)O	44.9982, CHO_2 59.0139, $C_2H_3O_2$ 71.0139, $C_3H_3O_2$
277.2280	1	$C_{18}H_{30}O_2$ **α-Linolenic acid**	12.88	14.17	673	59.0140 71.0138 205.1963 233.2273	CC/C=C\C/C=C\C/C=C\CCCCCCCC(=O)O	59.0140, $C_2H_3O_2$ 71.0138, $C_3H_3O_2$ 205.1963, $C_{15}H_{25}$ 233.2273, $C_{17}H_{29}$
279.2331	1	$C_{18}H_{32}O_2$ **α-Linoleic acid**	13.51	14.50	551	59.0139 71.0139 205.1962 233.2275 261.2224	CCCCC/C=C/C/C=C/CCCCCCCC(=O)O	59.0139, $C_2H_3O_2$ 71.0139, $C_3H_3O_2$ 205.1962, $C_{15}H_{25}$ 233.2275, $C_{17}H_{29}$ 261.2224, $C_{18}H_{29}O$

m/z	Ident. Level	Unequivocal Molecular Formula	t_R (min)	Pred. t_R (min)	ChemSpider Hits	Fragments m/z	Tentative Structure SMILES	Peaks explained (m/z), Elemental Formula
281.2485	1	$C_{18}H_{34}O_2$ **Oleic acid**	13.98	15.86	396	59.0140 71.0139 83.0502 97.0659 99.0452 111.0815 125.0972 127.0765 182.1311	CCCCC/C=C/C/C=C/CCCCCCCC(=O)O	59.0140, $C_2H_3O_2$ 71.0139, $C_3H_3O_2$ 83.0502, C_5H_7O 97.0659, C_6H_9O 99.0452, $C_5H_7O_2$ 111.0815, $C_7H_{11}O$ 125.0972, $C_8H_{13}O$ 127.0765, $C_7H_{11}O_2$ 182.1311, $C_{11}H_{18}O_2$
311.3046	1	$C_{20}H_{40}O_2$ **Arachidic acid**	15.10	17.38	135	59.0139 139.1492	CCCCCCCCCCCCCCCCCCCC(=O)O	59.0139, $C_2H_3O_2$ 139.1492, $C_{10}H_{19}$
147.0452	2a **(MassBank Record :KO000401)**	$C_9H_8O_2$ **Cinnamic acid**	6.37	6.48	242	103.0553 129.0346	c1ccc(cc1)/C=C/C(=O)O	103.0553, C_8H_7 129.0346, C_9H_5O

m/z	Ident. Level	Unequivocal Molecular Formula	t_R (min)	Pred. t_R (min)	ChemSpider Hits	Fragments m/z	Tentative Structure SMILES	Peaks explained (m/z), Elemental Formula
191.0556	2a **(MassBank Record :KO001747)**	$C_7H_{12}O_6$ **Quinic acid**	1.12	1.37	104	59.0137 71.0137 85.0298 93.0345 111.0091 127.0395	C1[C@@](C[C@H]([C@@H]([C@@H]1O)O)O)(O)C(=O)O	59.0137, $C_2H_3O_2$ 71.0137, $C_3H_3O_2$ 85.0298, $C_4H_5O_2$ 93.0345, C_6H_5O 111.0091, $C_5H_3O_3$ 127.0395, $C_6H_7O_3$
195.0743	2b	$C_{10}H_{12}O_4$ **Acetosyringone**	5.60	5.57	842	151.0401 181.0508	CC(=O)c1cc(c(c(c1)OC)O)OC	151.0401, $C_8H_7O_3$ 181.0508, $C_9H_9O_4$
223.0678	2b	$C_{11}H_{12}O_5$ **Sinapic acid**	5.08	6.13	487	55.0190 166.0635	COc1cc(cc(c1O)OC)/C=C/C(=O)O	55.0190, C_3H_3O 166.0635, $C_9H_{10}O_3$

Level Ident.: Level of identification (Level 1 corresponds to confirmed structures where a reference standard is available; level 2a: evidence by spectra matching from literature or library and level 2b: diagnostic evidence where no other structure fits the experimental MS/MS information)

3.3.4 Retrospective Analysis

Upon the successful non-target identification of the free fatty acids reported above, retrospective analysis was also performed to search the possible presence of all the free fatty acids encountered in oil. 19 fatty acids were identified: hexanoic acid, octanoic, dodecanoic, myristic, pentadecanoic, palmitic, palmitoleic, heptadecanoic, heptadecenoic, stearic, oleic, **α-**linoleic, **α-**linolenic, arachidic, cis-eicosenoic, heneicosanoic, docosanoic, tricosanoic and lignoceric acid. The distribution of the VIPs of the identified free fatty acids are presented in the **ESM I, Table S5**. Their identification proves that they were extracted together with the phenolic constituents during the single extraction.

3.3.5 Prediction models and classification

Totally 19 samples were used to study the discrimination between samples. 15 samples used to train PLS-DA [120, 149] in which these samples consisted of 9 and 6 samples belonging to good and defect samples of olive, respectively. 4 samples (Musty, RB1, YH1, BL1) were used to evaluate the external accuracy of developed PLS-DA model. The accuracy of the model was assessed internally and externally by Receiver Operating Characteristics (ROC) curve. More information about the optimization of the PLS-DA model and the interpretation of results can be found in the **ESM A, Section S1A**. According to **Figure 3.12**, all 19 samples were classified with high accuracy into two groups, EVOOs or defectives. The developed model was evaluated with 4 samples and their classes were calculated by the developed model. In conclusion, the developed model is robust and can be applied to unknown samples to understand their sensory profile with high accuracy.

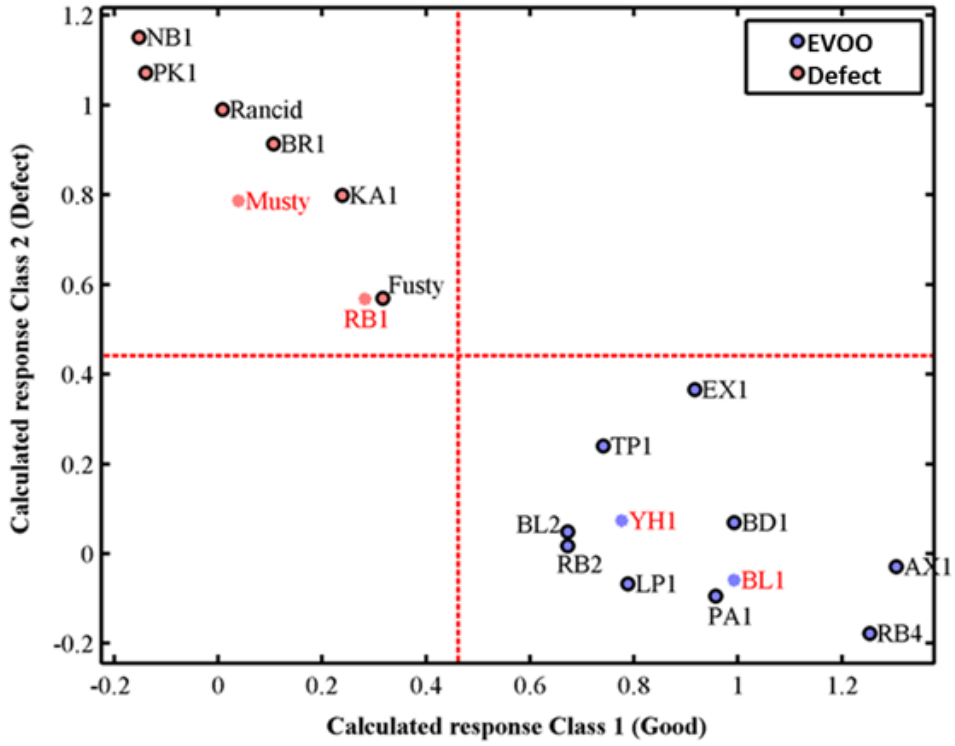


Figure 3.12: Sample distribution based on PLS-DA model. The samples with red color belonged to test set

In addition to PLS-DA, Counter Propagation Artificial Neural Networks (CP-ANNs) [147-149] were used to build a classification model. To develop this model without over fitting issues, the number of neurons (or size of map) and epochs were optimized using genetic algorithms. The procedures for the optimization of CP-ANNs can be found in the **ESM A, Section S2A**. The final map, using Self-organizing maps (SOMs), was proposed based on neuron size of 6×6, frequency of 0.3 and the number of epochs of 300 and is presented in **Figure 3.13.** In this figure, the blue and red neurons are representing the EVOOs and defective samples, respectively. The samples of external test set are shown with black color and their classes were predicted with high accuracy.

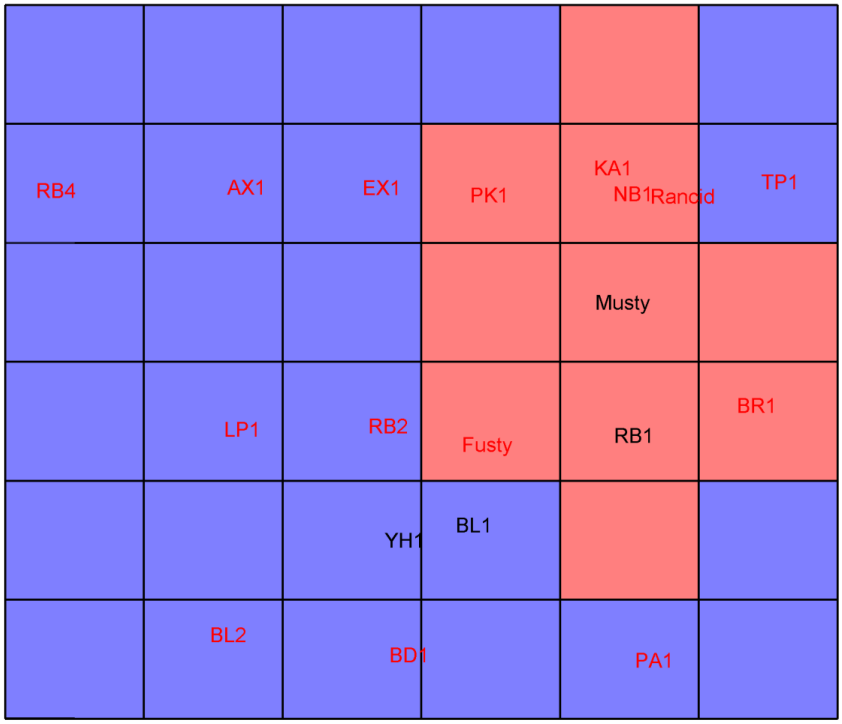


Figure 3.13: Mapping of samples using SOMs of the developed CP-ANNs model. The samples of external test set are shown with black color. Neurons in blue represent EVOOs and neurons in red represent defective samples

3.4 Conclusions

This study has made progress towards the organoleptic profiling of extra virgin olive oil demonstrating the prospects of a novel RP-UHPLC-ESI-QTOF-MS/MS analytical method. The use of target, suspect and non-target screening strategies in combination with supervised classification techniques, PLS-DA and CP-ANNs, constitutes a powerful tool that can be successfully applied in the investigation of extra virgin olive oil's authenticity.

The use of target and suspect screening resulted in the determination of 14 target compounds and 26 suspects. Using non-target screening, 11 compounds were newly identified as responsible for the discrimination between EVOOs and defective olive oils, after data processing with R-language and the XCMS package, following the VIP method for the suggestion of the most important features. Overall, 51 compounds are suggested as markers responsible for olive oil's organoleptic characteristics. Interestingly, a clear increase existed for

hexanoic acid and octanoic acid levels in defective oils. Similarly, a clear increase was observed in the concentration of palmitic acid, linolenic, linoleic, oleic and arachidic acid in the EVOOs. Nevertheless, the detection of those free fatty acids demonstrates that the proposed method can be applied in the identification of phenolic compounds and free fatty acids simultaneously, since they are both extracted within a single LLME extraction and detected in the same analytical run.

Furthermore, two robust classification models using PLS-DA and CP-ANNs were built based on all the features detected and they can classify olive oils into two groups, defectives and EVOOs, with high accuracy.

CHAPTER 4

INVESTIGATING THE ORGANIC AND CONVENTIONAL PRODUCTION TYPE OF OLIVE OIL WITH TARGET AND SUSPECT SCREENING BY LC-QTOF-MS, A NOVEL SEMI-QUANTIFICATION METHOD USING CHEMICAL SIMILARITY AND ADVANCED CHEMOMETRICS

4.1 Introduction

The term “organic food” denotes products that have been produced in accordance with the principles and practices of organic agriculture [36]. Organic farming is a production system which avoids the use of synthetically compounded fertilizers, pesticides and growth regulators. Organic farming practices are based on the idea that each part of the farm operation augments the other parts to form an efficient and sustainable food production system, offering many advantages, such as minimizing all forms of pollution and producing food of high quality [150]. For these reasons, there is considerable increase in consumers demand for organic foods, in a global scale.

In respect to consumer’s needs, many attempts have been made in order to differentiate between organic and conventional products, but the results were controversial [150]. However, new results arising from foodomics and metabolomics studies have detected differences in minor food components, such as polyphenols and other bioactive compounds [127]. The term foodomics describes the discipline that studies the food and nutrition domain through the application of advanced omics technologies to improve consumer’s well-being, health and confidence [8]. Among other applications, foodomics can explore the effect of the agronomic environment on the metabolite profile of food. In this field, a lot of studies have been conducted [151-154]. Koh et al. [152] compared the chemical composition of organic and conventional spinach using liquid chromatography-electrospray ionization-tandem mass spectrometry (LC-ESI-MS/MS) and analysis of variance (ANOVA), concluding that the content of flavonoids is higher in organic products. These findings reported to affect the carbon/nutrient balance theory and growth rate as well as growth-differentiation

balance hypothesis, indicating the allocation of plant metabolism toward higher carbon-containing components (like flavonoids) [152]. In addition, Ren et al. [153] as well as Vallverdu-Qurealt et al. [154] observed the same effects in vegetables by means of target screening in LC-MS. However, the studies on this subject for olive oil are scarce in the literature, trying to investigate the correlation between olive oil contents and production types, using metabolomics and chemometrics [150]. Nonetheless, the comparison between environmental factors and harvesting years has been neglected [11]. Anastasopoulos et al. [21] and Rosati et al. [60] measured total phenolic content with Folin Ciocalteu. They observed that the phenolic content was higher in organic production type. The above mentioned studies [9-12] denote that there is an existing connection between the phenolic content and the production type. Thus, the development of high resolution analytical methodologies, with higher identification confidence, that enable the identification of phenolic compounds in such cases is of high interest.

While dealing with analysis of the multi class of compounds having different polarities (polar compounds such as alcohols and acids; and less polar ones, like secoiridoids, lignans and flavonoids), there is an emerging need to derive the optimum experimental conditions. Chemometric methods have been frequently applied to optimize analytical methods, introducing several advantages such as reduction in the number of experiments, reagent consumption and less laboratory work [155]. Moreover, these methods can reveal the significance of the factors, their effects and interaction effects. Factorial design, one factor design (OFD), central composite designs (CCD) or Box-Behnken design are some of the widely used methods for design of experiments [155, 156]. These methods often couple to response surface methodology (RSM) to derive the optimal conditions for any property under study [157, 158]. After the optimization of the analytical methodology and the identification of the target and suspect compounds, the semi-quantification of the suspect remains a challenge since in most cases, there are no reference standards commercially available.

Usefulness and reliability of the semi-quantification can be more well-established using the most relevant standard [159]. To such an end, chemical similarity analysis can be applied to rank the standards for semi-quantification purposes.

Chemical similarity has been subject of nearly a decade, trying to find the correct and meaningful similarity assignment between compounds [157, 160-163]. From chemical perspective, similar compounds should have similar functional groups or fragments [162]. The scoring function and a scale that can describe the chemical space edge are vital [162]. Such a score can be easily developed as chemical space is subjected to understand the correlation between a property and chemical descriptors [121]. Application of such chemical descriptors have made a breakthrough in terms of identification of similar chemical structures in a large-scale database [162]. Introduction of chemical fingerprints with a suitable similarity metrics (Euclidean or Tanimoto) [157, 162] could also help to assign an accurate chemical similarity score.

Following the optimization of analytical method and selection of appropriate standards for semi-quantification purposes, a robust model should be developed to discriminate between organic and conventional extra virgin olive oils (EVOOs). Although Partial least Squares Discriminant Analysis (PLS-DA) can score the MS features and select markers [60], the interpretation of results might be complex when the explained variances are too low giving little discriminative power to the PLS-DA model. In such a case, models must be inspected by cross-validation analysis and an external test set to verify that the model is capable of correct class assignment [106]. Moreover, it is of great need to set a threshold for the suggested markers that define the olive oil production types. PLS-DA is not capable of setting such threshold. In order to have a discriminative method applicable and reliable, the threshold derived for each marker should be evaluated throughout the changes in environmental conditions between different harvesting years. Building a correlation between the interaction factors and phenolic content may help to understand whether there are significant differences between organic and conventional olive oils and if olive oils of different harvesting years are comparable or not. This reveals that the markers identified are extremely relevant to build the discriminative models.

The main aim of this study is to develop an optimized reversed-phase ultra-high performance liquid chromatography-electrospray ionization quadrupole time of flight tandem mass spectrometric method (RP-UHPLC-ESI-QTOF-MS), using

target and suspect screening workflows combined with advanced chemometrics to reveal the correlation between the phenolic compounds and the production type. Secondly, to identify the markers responsible for the discrimination in a two-years study. The method was optimized by OFD-RSM to derive the optimal conditions for the extraction of the phenolic compounds, the appropriate internal standard and its concentration. The method was applied in 52 EVOOs of Kolovi variety from Lesvos, both organic and conventional that were harvested during the years 2014-2015 and 2015-2016, for the determination of 13 target phenolic compounds and suspect screening was followed for the identification of 96 suspect phenolic compounds. The target phenolic compounds were quantified and a novel semi-quantitation strategy is introduced based on chemical similarity analysis. Then, Ant Colony Optimization-Random Forest (ACO-RF) was employed to investigate alterations between organic and conventional olive oils and introduce one or more markers, suggesting a concentration threshold and discriminate between organic and conventional EVOOs.

4.2 Experimental section

4.2.1 Chemicals and standards

All standards and reagents were of high-purity grade (>95%). Methanol (MeOH) as well as acetonitrile (ACN) of LC-MS grade and sodium hydroxide (>99%) were purchased from Merck (Darmstadt, Germany). Ammonium acetate (≥99.0%) for HPLC and formic acid (LC-MS Ultra) were purchased from Fluka (Buchs, Switzerland). Isopropanol was purchased from Fisher Scientific (Geel, Belgium). Distilled water was provided by a Milli-Q purification apparatus (Millipore Direct-Q UV, Bedford, MA, USA). For the analytical method validation the following reagents were used: syringic acid 95% was purchased from Extrasynthèse (Genay, France), gallic acid 98 %, ferulic acid 98%, epicatechin 97%, p-coumaric (4-hydroxycinnamic acid) 98%, homovanillic acid 97%, as well as oleuropein 98% and pinoresinol 95% were obtained from Sigma-Aldrich (Steinheim, Germany), hydroxytyrosol 98% and luteolin 98% were acquired from Santa Cruz Biotechnologies. Vanillin 99%, ethyl vanillin 98%, apigenin (4, 5, 7

trihydroxyflavone) 97% and tyrosol (2-(4-hydroxyphenyl) ethanol) 98% were acquired from Alfa Aesar (Karlsruhe, Germany). Caffeic acid 99% and syringaldehyde 98% (internal standards) were purchased from Sigma- Aldrich (Steinheim, Germany). Stock standard solutions of individual compounds (1000 mg L^{-1}) were solubilized in MeOH and stored at -20 °C in dark brown glass. All intermediate standard solutions containing the analytes were prepared by dilution of the stock solutions in MeOH.

4.2.2 Olive oil samples

Overall, 52 monovarietal EVOOs were acquired from the Island of Lesvos for a two years study. 41 EVOOs of Kolovi variety produced from olives cultivated over the harvesting period 2015-2016, consisting of 17 organic and 24 conventional olive oils. Moreover, 11 extra virgin olive oils of the same variety produced during the harvesting period 2014-2015 (2 organic and 9 conventional) were also included in the current research, as a test set to evaluate the successful applicability of the proposed discrimination models in previous harvesting periods. **Figure 4.1** presents the geographical distribution of the monovarietal organic and conventional extra virgin olive oils that were produced during the harvesting periods 2014-2015 and 2015-2016. In this Figure, all samples that are in italic relate to the harvesting period 2014-2015 and all samples in bold relate to the harvesting period 2015-2016. Moreover, samples labeled as organic are marked with "*". More information regarding the harvesting and production details of the EVOOs can be found in the **ESM II, Table S1**. All samples were protected from light and humidity and were preserved as it has already been reported by Kalogiouri et al. [164].

Figure 4.1: Geographical distribution of EVOOs selected from Lesvos Island

4.2.3 Instrumental analysis

A UHPLC system with an HPG-3400 pump (Dionex UltiMate 3000 RSLC, Thermo Fisher Scientific, Germany) was used for RP analysis, interfaced to a Q-TOF mass spectrometer (Maxis Impact, Bruker Daltonics, Bremen, Germany), in negative electrospray ionization mode. Separation was carried out using an Acclaim RSLC C18 column (2.1 × 100 mm, 2.2 μm) purchased from Thermo Fisher Scientific (Driesch, Germany) with a pre-column of ACQUITY UPLC BEH C18 (1.7 μm, VanGuard Pre-Column, Waters (Ireland)). Column temperature was set at 30°C. The solvents used consisted of: (A) 90% H_2O, 10% MeOH and 5 mM CH_3COONH_4, (B) 100% MeOH and 5 mM CH_3COONH_4. The adopted elution gradient started with 1% of organic phase B with flow rate 0.2 mL min^{-1} during one minute, gradually increasing to 39 % for the next 2 minutes, and then increasing to 99.9% and flow rate 0.4 mL min^{-1} for the following 11 minutes. These almost pure organic conditions were kept constant for 2 minutes (flow rate 0.48 mL min^{-1}) and then initial conditions (1% B - 99% A) were restored within 0.1 minute (flow rate decreased to 0.2 mL min^{-1}) to re-equilibrate the column for the next injection.

The Q-TOF MS system was equipped with an electrospray ionization interface (ESI), operating in negative mode with the following settings: capillary voltage of

3500 V; end plate offset of 500 V; nebulizer pressure of 2 bar (N_2); drying gas of 8 L min^{-1} (N_2); and drying temperature of 200 °C. A Q-TOF external calibration was daily performed with sodium formate (cluster solution), and a segment (0.1−0.25 min) in every chromatogram was used for internal calibration, using calibrant injection at the beginning of each run. The sodium formate calibration mixture consisted of 10 mM sodium formate in a mixture of water/isopropanol (1:1). Full scan mass spectra were recorded over the range of 50-1000 m/z, with a scan rate of 2 Hz. MS/MS experiments were conducted using AutoMS data dependent acquisition mode based on the fragmentation of the five most abundant precursor ions per scan. The instrument provided a typical resolving power (FWHM) between 36,000-40,000 at m/z 226.1593, 430.9137 and 702.8636.

4.2.4 Screening methodology

Target and suspect screening methodologies were followed, as it has already been described in Chapter 3. The identification workflow incorporated strict filtering steps, interpretation of MS/MS spectra and retention time prediction. A target list was created including 13 phenolic compounds, including phenolic acids, secoiridoids, flavonoids and lignans (gallic acid, p-coumaric acid, ferulic acid, syringic acid, homovanillic acid, tyrosol, hydroxytyrosol, pinoresinol, apigenin, oleuropein, vanillin, ethyl vanillin and epicatechin) that have already been identified in extra virgin olive oils of Kolovi variety in a previous study [164] (Chapter 3). An updated suspect list of 96 bioactive constituents was generated from literature including all the bioactive constituents and mainly the phenolic compounds that have been identified in olive oils, drupes and leaves. The initial suspect list is presented in the **ESM II, Table S2**.

The software packages Target Analysis 1.3 and Data Analysis 4.1 (Bruker Daltonics, Bremen, Germany) along with the tools of these packages, Bruker Compass Isotope Pattern and SmartFormula Manually were in the target screening workflow. Extracted ion chromatograms (EICs) were obtained using the function Find Compounds-Chromatogram in Target Analysis Software. Mass accuracy was set at 5 ppm, mSigma was below or equal to 50, signal to noise

threshold of 3, minimum area threshold of 800, and minimum intensity threshold of 200. The relative tolerance of the retention time window was set lower than ±0.2 min. The target compounds were identified on the basis of mass accuracy, isotope pattern, retention time (t_R), and MS/MS fragments [164].

In suspect screening, the EICs were created using Target Analysis Software 1.3 and the following parameters were set: mass accuracy threshold of 5 ppm, isotopic fit below or equal to 50, ion intensity of more than 800, peak area threshold of 2000 and peak score (area/intensity ratio) between 4-38 [164]. The EICs were studied using Data Analysis 4.1 software to confirm that the peak represents the suspect compound. The MS/MS fragments were compared and interpreted with the use of Metfrag [91] and FooDB [93]. The retention time of each suspect compound was predicted and compared with the experimental retention time with the use of quantitative structure-retention relationship model (QSRR) [94].

As for the level of confidence achieved in the identification of the suspect compounds, compounds are identified at level 1 when the structures are confirmed with available reference standards. In the cases that there are no standards commercially available, level 2 corresponds to probable structures (level 2a, MS/MS fragments were verified with spectral libraries or literature; level 2b, diagnostic evidence where no other structure fits the experimental MS/MS information) and level 3 corresponds to tentative candidates [142].

4.2.5 Optimization of experimental conditions

The initial design consisted of three main factors (one numeric and two categorical variables) within one block. The design model was selected quadratic to cover the multilevel limits for parameters intended to be optimized. Extraction (which was a categorical factor) was set at three levels (MeOH, MeOH:H_2O (80:20, v/v) and acetonitrile). The second factor was the internal standard (caffeic acid and syringaldehyde) and the final factor was the concentration of the internal standard, set within the range of 0.5 up to 1.5 mg L^{-1}. In the case of quadric model, 5 levels (0.5, 0.75, 1.00, 1.25, and 1.5 mg L^{-1}) for one numeric factor (concentration) are required with some replicates points. This design could be

duplicated for every combination of categorical factor levels. The optimization task was performed to minimize the relative standard deviation (%RSD) values of the peak areas of each spiked standard. The combination of all these factors required a set of 42 experiments. These experimental plans based on OFD method coupled to RSM, along with the %RSD value for each spiked standard as response (n=3), are presented in the **ESM II, Table S3**. The design of experiments and all statistical assessments were calculated by Design-Expert software version 7 [165].

4.2.6 Method Validation

The optimized RP-UHPLC-ESI-MS method was validated to ensure that it is suitable for identification and quantification purposes. Standard addition curves were constructed for all the analytes. All the compounds were spiked in real EVOO samples. Gallic acid, p-coumaric acid, ferulic acid, syringic acid, homovanillic acid, pinoresinol, apigenin, vanillin, ethyl vanillin, epicatechin and luteolin were spiked at concentrations between 0.02-10 mg kg^{-1} (14 calibration levels with 3 replicates at each level). Tyrosol, hydroxytyrosol and oleuropein calibration curves were constructed over the range of 0.02-100 mg kg^{-1} (20 calibration levels with 3 replicates at each level). Calibration curves were constructed with the use of the peak area of the spiked analyte subtracted by the peak area of a neat sample and divided by the peak area of the internal standard. Limits of detection (LODs) and limits of quantification (LOQs) were calculated at the lowest concentration range of the analytes (0.02-1 mg kg^{-1}), by the equations (1) and (2) described in **Section 3.2.9**.

The accuracy of the method was estimated using recoveries, at 2 mg kg^{-1} concentration level, calculated as described by equation (3) in **Section 3.2.9**. To evaluate the matrix effect, the matrix factor was calculated at 2 mg kg^{-1} concentration level according to equation (4), **Section 3.2.9**. For the calculation of ME, 1 was subtracted by of the quotient (4) and multiplied by 100, so that the negative result indicates suppression and the positive result indicates enhancement of the analyte signal. The precision of the method was demonstrated in terms of repeatability (intra-day precision) and intra-laboratory

reproducibility (inter-day precision). Repeatability was expressed as the $\%RSD_r$ values of 6 replicate analyses (n = 6) in the same day. Reproducibility experiments expressed as the $\%RSD_R$ value of 3 replicates of three consecutive days ($n \times k = 3 \times 3 = 9$). Finally, lack-of-fit F-test was applied to ensure that the calibration curves can be used for quantification purposes. For this scope, all 3 replicates of each concentration level were used and the number of data points (concentration levels) was 20 for oleuropein, tyrosol and hydroxytyrosol, and 14 for the rest of the analytes.

4.2.7 Chemical similarity analysis

Three standards including tyrosol, hydroxytyrosol and oleuropein were used as a main scheme for semi-quantification to define the chemical space boundaries (chemical space edge) and their similarity distance from 14 secoiridoids (10-hydroxy-10 methyl oleuropein aglycone, methyl oleuropein aglycone, 10-hydroxy oleuropein aglycone, oleoside, oleuropein aglycone, oleomissional, lingstroside aglycone, oleokoronal, 10-hydroxy decarboxymethyl oleuropein aglycone, decarboxymethyl oleuropein aglycone, decarboxymethyl lingstroside aglycone, hydroxylated form of elenolic acid, elenolic acid and hydroxytyrosol acetate). All structures of chemicals used here were drawn and their geometries were constructed by searching between conformers with lowest energy using Balloon [166]. The chemical similarity matrix for these compounds was then built based on the molecular descriptors. These molecular descriptors consisted of logD (at pH=6.2) (measure of hydrophobicity for ionizable compounds), constitutional descriptors, topological descriptors, walk and path counts, connectivity indices, information indices, 2D autocorrelation, edge adjacency indices, burden eigenvalues, topological charge indices, eigenvalue-based indices, Randic molecular profiles, geometrical descriptors, radial distribution function descriptors (RDF), 3D molecular representation of structure based on electron diffraction descriptors (3D-MoRSE), weighted holistic invariant molecular descriptors (WHIM), GETAWAY (geometry, topology and atoms-weighted assembly) descriptors, functional group counts, atom-centred fragments, charge descriptors, molecular properties [167-170].

Concerning the above descriptors, they encode the atomic or molecular properties, overall molecular connectivity, molecular geometry and their size and shape [171]. These descriptors were calculated by E-dragon [172, 173]. LogD was calculated using the ChemAxon package [174] (the calculated molecular descriptors can be found in the **ESM II, Table S4**. Afterwards, the calculated molecular descriptors were pre-treated in order to remove the constant and near constant descriptors. Molecular descriptors with inter-correlation above 0.95 were also removed using variable reduction method adapted from space-filling designs (V-WSP) as an unsupervised variable reduction method [100] (the survived molecular descriptors can be found in the **ESM II, Table S5**. Euclidean based similarity metric was used to measure the chemical similarity between the compounds. To define the chemical space, tyrosol, hydroxytyrosol and oleuropein were used as a main scheme and the other 14 compounds were measured against these three standards. The chemical space edge was also achieved by normalizing the mean distance score for the three standards (these values range from 0 to 1 where 0.0 is least diverse, and 1.0 is the most diverse compound). Then, the normalized mean distance scores for the rest of compounds were calculated, and those test compounds, which were scored outside of 0.0 to 1.0 range, were defined to be outside of the chemical space edge. Therefore, this method could define the most appropriate standard for semi-quantification. Similarity analysis and V-WSP calculation were done in MATLAB 8.5 (MathWorks) program.

4.2.8 Prioritizing MS features and modelling strategies

Overall, a matrix containing quantified and semi-quantified results (expressed in mg kg^{-1}) for 30 compounds was generated for 52 extra virgin olive oil samples. These samples were split into a training and a test set based on their harvesting year (to evaluate whether the discrimination achieved is applicable to previous years or not) to build the discrimination models and then evaluate the accuracy of the discrimination model for the external set of samples. ACO [121, 122] was used to prioritize compounds and rank them by their importance and contribution in increasing the accuracy of discrimination model. The fitness function (a

measure of error for the discrimination model) was set based on the error of miss-discrimination in cross-validation leave-one-out analysis. ACO was then coupled to discrimination modelling techniques to evaluate the internal and external accuracy of models every time by inclusion of new features. Using feature selection coupled with discrimination model such as Linear Discriminative Analysis (LDA) [175-177] or RF [109] can prevent over-fitting issues and can introduce more accuracy to a discrimination problem. In RF, variables and their contributions can be ranked based on a measure of variable importance and the modeling can be followed based on the highly important predictors [178]. Therefore, the introduction of a features prioritizing method (ACO) might not be so important. The following fitness function was used to measure the error of discrimination in leave one out cross validation analysis and to decrease it using ACO:

$$F = \sum \frac{Class \sim Pred.Class}{n} \quad (5)$$

Where *F* is the objective function (discrimination error measure), *Class* is the observed group for a case (here is each sample), *Pred. Class* is the predicted group by the modelling technique and n is the number of samples used to build the discrimination model. The entire data processing step was done in a homemade program, called ChemoTrAMS, written in MATLAB environment.

4.2.9 Validation procedure of the models

The initial parameter used to evaluate the internal accuracy of the models was the error rate of miss-discrimination in training set and cross validation analysis. Leaving-one-out, cross validation was also performed during the training step to understand the error rate by excluding a certain sample from the rest of the training set. The predictive power of the proposed discrimination model was evaluated independently using a set of external samples that were not part of the initial training set and confusion matrix was calculated to derive error rate, class specificity and sensitivity [106]. Moreover, Receiver Operating Characteristics (ROC) was calculated to check the discrimination capability of the models. ROC curves were calculated for each class by plotting the sensitivity versus 1-

specificity for a binary case study (organic or conventional). A perfect discrimination model would yield a point in the upper left corner of the ROC area, representing maximum sensitivity and specificity, while a random discrimination causes points to be along the diagonal line from the left bottom to the top right corner [106].

4.3 Results and discussion

4.3.1 Optimization of the method

The evaluation of the best extraction conditions and the selection of the appropriate internal standard took place using OFD. The goal of OFD-RSM was to optimize these three factors at a point which low %RSD values of the peak areas of the spiked standard compounds would be achieved. It was found that ACN has the lowest desirability [51] for all compounds under study and it shows high error (high %RSD). The interaction map and effect of extractions are shown in **Figure 4.2.** As it can be derived from **Figure 4.2(a)** and **Figure 4.2(b)**, MeOH:H_2O (80:20, v/v) has the highest desirability (lowest %RSD) among other extractions. Since the desirability observed for MeOH and MeOH:H_2O (80:20, v/v) is close, the interaction maps were investigated.

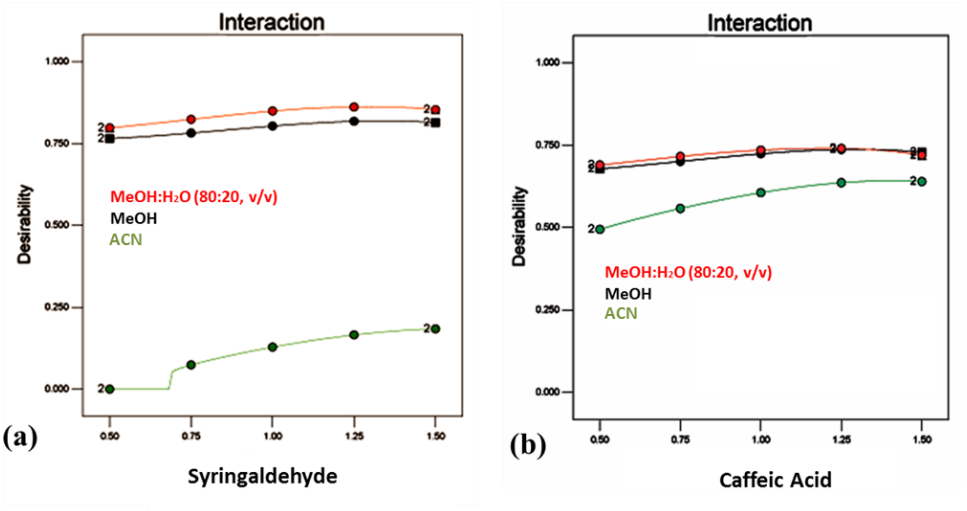


Figure 4.2: Desirability of different extractions while using (a) syringaldehyde and (b) caffeic acid as internal standard

Table 4.1 summarizes the results of the desirability plots for MeOH and MeOH:H_2O (80:20, v/v) using syringaldehyde as an internal standard as well as for MeOH:H_2O (80:20, v/v) using caffeic acid as internal standard, based on the set of 42 experiments**.** For most of the spiked standard compounds, MeOH:H_2O (80:20, v/v) presented the lowest %RSD values and was selected as the optimum extractor. Comparing MeOH with MeOH:H_2O (80:20, v/v) when syringaldehyde was the as internal standard, both extraction solvents demonstrate close %RSD values for all the spiked standards, except for syringic acid which has highest %RSD while using solely MeOH. The good performance of MeOH:H_2O (80:20, v/v) is clearly demonstrated in **Table 4.1** showing higher desirability values (especially for syringic acid) compared to pure MeOH.

Table 4.1: Desirability values of all the spiked standard compounds

Extracting solvent	**MeOH**	**MeOH:H_2O (80:20, v/v)**	**MeOH:H_2O (80:20, v/v)**
Internal Standard	**Syringaldehyde**	**Syringaldehyde**	**Caffeic acid**
Vanillin	0.811019	0.893997	0.737738
Apigenin	0.849736	0.912608	0.377854
Epicatechin	0.926173	0.778114	0.541533
Ethyl Vanillin	0.972953	0.962772	0.433490
Ferulic acid	0.899410	0.867664	0.926501
Gallic acid	0.769335	0.810612	1
Homovanillic acid	0.704925	0.770156	0.986127
Hydroxytyrosol	0.954376	1	0.925576
p-Coumaric acid	0.793496	0.833178	0.988094
Syringic acid	0.484745	0.701904	0.975974
Tyrosol	0.910728	0.933066	0.785103
Pinoresinol	0.895791	0.849412	0.632015
Oleuropein	0.825102	0.938799	0.758019
Combined	0.819218	0.861535	0.740338

Moreover, the comparison between the desibarility plots where MeOH:H_2O (80:20, v/v) is the extracting solvent and the internal standard used is syringaldehyde in **Figure 4.3(a)** and caffeic acid in **Figure 4.3(b),** reveal that all the spiked standard compounds of the phenolic acids class presented higher desirability in the case that caffeic acid was used as an internal standard.

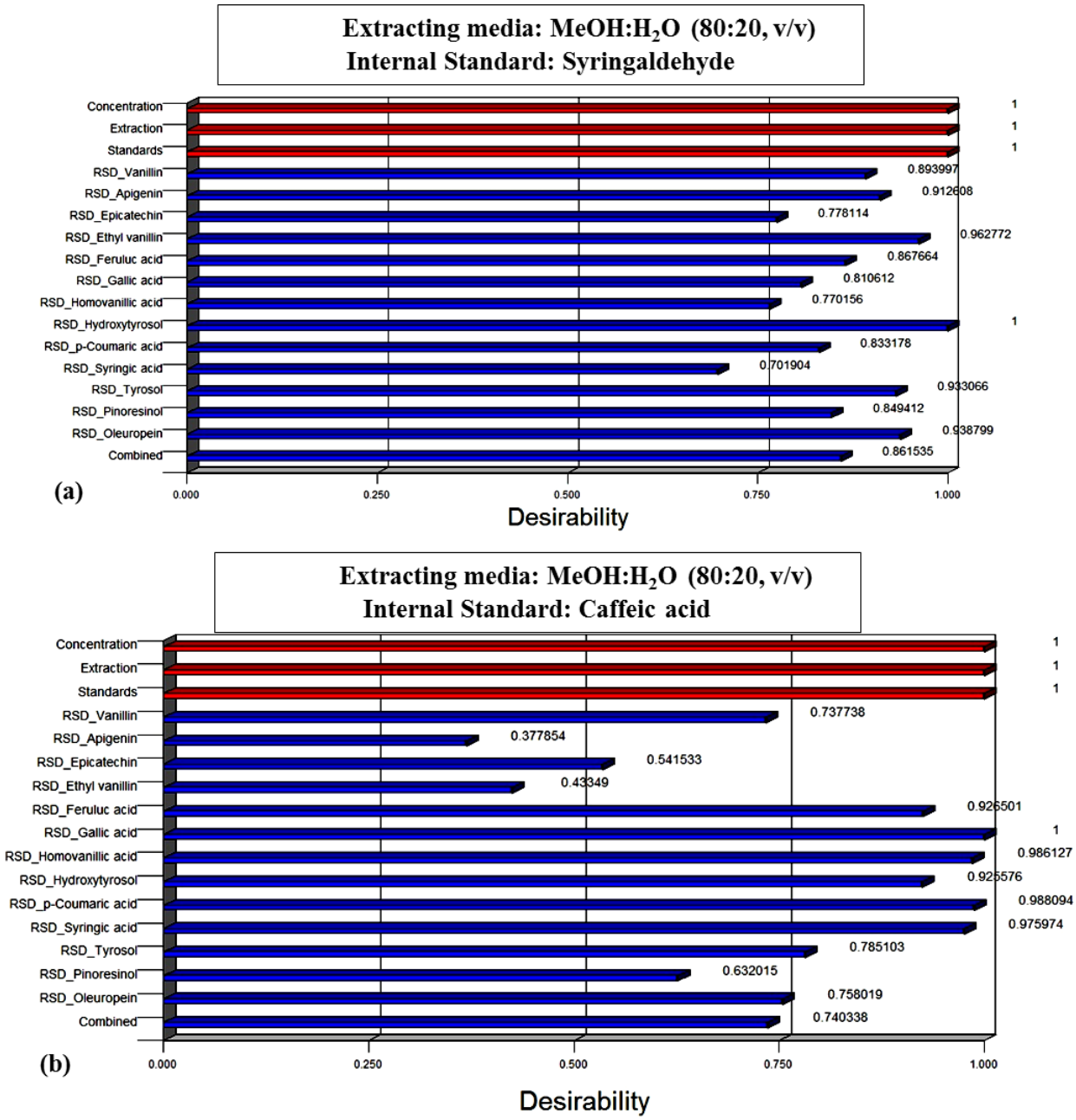


Figure 4.3: Desirability plots for MeOH:H_2O (80:20, v/v) using (a) syringaldehyde; (b) caffeic acid as internal standards

In a further step, OFD-RSM was applied in order to derive the optimum conditions and select the appropriate internal standards at the optimal concentration level, by providing prediction results of the %RSD values of the peak areas for the 13 spiked standard compounds (n=3). It generated predicted %RSD values in the case that MeOH:H_2O (80:20, v/v) was the extracting media and compared the desirability of both internal standards; syringaldehyde and caffeic acid at 1.30 mg L^{-1} and 1.20 mg L^{-1}, respectively **(ESM II, Table S6)**. The predicted results revealed that the optimum conditions are derived when syringaldehyde is the internal standard at 1.30 mg L^{-1}. The experimental factors suggested by OFD-RSM were applied and the experimental %RSD values were in accordance with the predicted.

Moreover, the recoveries were calculated for all the spiked standard compounds in the three different extraction solvents (MeOH, MeOH:H_2O (80:20, v/v), ACN) with syringaldehyde at 1.30 mg L^{-1} in order to further investigate the adequacy of MeOH:H_2O (80:20, v/v), and are illustrated as the mean values (n=3) in the heatmap in **Figure 4.4**.

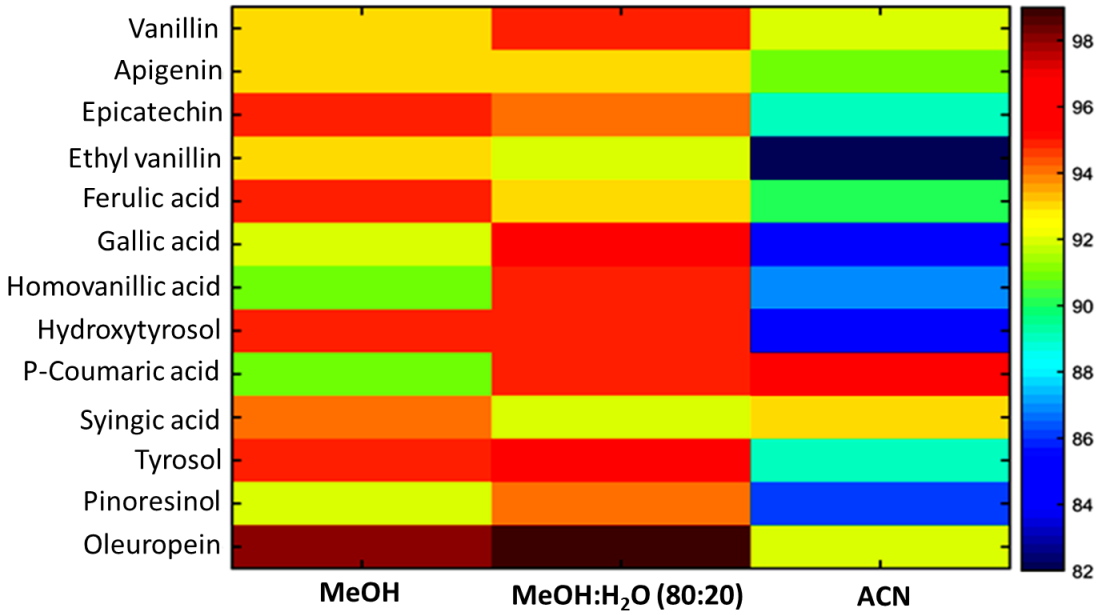


Figure 4.4: Heatmap of the calculated recoveries (mean values, n=3) for all the spiked standard compounds in different extraction solvents (MeOH, MeOH:H_2O (80:20, v/v), ACN) and syringaldehyde at 1.30 mg L^{-1} as an internal standard.

MeOH:H_2O (80:20, v/v) is a better extracting media than both pure MeOH and ACN. The use of syringaldehyde as an internal standard at 1.30 mg L^{-1} presents higher desirability compared to caffeic acid 1.20 mg L^{-1}. **Figure 4.5** illustrates these optimal experimental conditions.

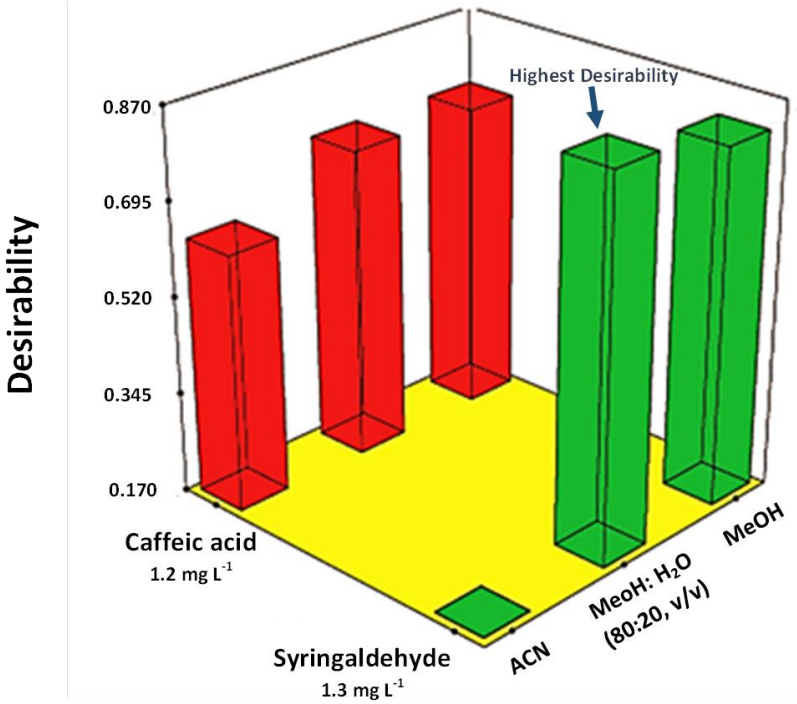


Figure 4.5: Derived optimal experimental conditions

These optimal conditions were implemented and a liquid-liquid microextraction (LLME) method was developed and validated in order to isolate all the phenolic compounds from the olive oil samples. For this, 0.5 g of each sample was weighted and spiked with 1.30 mg L^{-1} syringaldehyde and in a further step 0.5 mL of MeOH:H_2O (80:20, v/v) was added to 2-mL Eppendorf tubes. Then, the mixture was vortexed for 2 min and centrifuged for 5 min at 13,400 rpm. Additionally, the upper phase was collected and filtered through membrane syringe filters of regenerated cellulose (CHROMAFIL® RC) (15-mm diameter, 0.22-μm pore size, purchased by Macherey-Nagel, Düren, Germany). Finally, 5 μL of this solution was injected into the chromatographic system. Procedural blanks were prepared and processed in the chromatographic system to detect any potentional contamination. Quality control samples were prepared to confirm that the analytical system has been stabilized before the batch of samples and to evaluate its performance. The quality control sample was prepared by mixing all

aliquots of the samples. Then, it was spiked with 50 μL of a standard solution mix (including all the target compounds: vanillin, apigenin, epicatechin, ethyl vanillin, ferulic acid, gallic acid, homovanillic acid, hydroxytyrosol, oleuropein, p-coumaric acid, pinoresinol, syringic acid, tyrosol and luteolin, at a final concentration of 1 mg L^{-1}). It was injected at the beginning of the analysis (five times for conditioning) and afterward it was injected at regular intervals (every ten sample injections). The calculated %RSDs for the retention time (t_R) and the peak areas as well as mass errors (Δ*m)* are presented in the **ESM II,Table S7** demonstrating the good performance of the analytical system; (n=8).

4.3.2 Target screening results

After the optimization of the experimental conditions, a data-dependent method was used to scan the presence of the target compounds in real olive oil samples. All the target phenolic compounds that belonged in the initial target list such as gallic acid, p-coumaric acid, ferulic acid, syringic acid, homovanillic acid, tyrosol, hydroxytyrosol, pinoresinol, apigenin, luteolin, oleuropein, vanillin, ethyl vanillin and epicatechin were determined. The mass accuracies of the precursor ions and the qualifier ions of the detected compounds were less than 5 ppm compared with the standard solutions and the isotopic fit was calculated less than 50 mSigma in all cases. Moreover, the retention time shift was less than 0.05 min for all the detected target compounds. The most abundant fragments provided by the AutoMS spectra were verified with MS/MS records in literature [164]. Target screening results are summarized in the **ESM II, Table S8**.

All validation parameters including LODs and LOQs, calculated recoveries (RE), regression equations, regression coefficient (r^2), the lack-of-fit test, method precision expressed as intraday and interday precision, as well as the matrix factor (MF) and matrix effect (ME) are summarized in **Table 4.2**.

Table 4.2: Validation results

Compound	**LOD (mg kg^{-1})**	**LOQ (mg kg^{-1})**	**Equation $y=(a\pm S_a) + (b\pm S_b)x$** (linear range: 0.02-10 mg kg^{-1}; *0.02-100 mg kg^{-1}*)	**r^2**	**Lack of fit** F_{calc}	F_{tab}	**Intraday precision RSD_r, (%) (n=6)**	**Interday precision RSD_R, (%) (n×k=3×3)**	**RE %**	**MF**	**ME %**
Gallic acid	0.005	0.014	y=(0.79±0.68)+(7.73±0.16)x	0.995	1.510	2.118	2.1	5.1	98.4	0.94	-5.89
p-Coumaric acid	0.024	0.073	y=(0.06±0.05)+(0.51±0.01)x	0.994	1.772	2.118	1.9	4.1	96.8	0.94	-6.13
Ferulic acid	0.011	0.032	y=(0.017±0.008)+(0.687±0.004)x	0.999	0.056	2.118	2.2	3.5	95.2	0.96	-4.31
Syringic acid	0.003	0.008	y=(0.06±0.04)+(0.48±0.01)x	0.995	1.740	2.118	1.8	3.6	97.6	0.95	-4.90
Homovanillic acid	0.017	0.051	y=(0.02±0.09)+(1.49±0.02)x	0.998	0.726	2.118	1.9	5.4	97.5	0.93	-7.27
Tyrosol	0.028	0.086	*y=(-4.34±2.83)+(4.99±0.08)x*	0.996	1.493	1.868	1.3	2.6	96.9	0.92	-7.78
Hydroxytyrosol	0.010	0.031	*y=(-0.43±1.07)+(3.77±0.03)x*	0.999	0.420	1.868	1.8	2.8	96.8	0.94	-5.80
Pinoresinol	0.023	0.068	y=(0.01±0.02)+(0.297±0.005)x	0.996	0.405	2.118	1.5	2.9	98.9	0.93	-7.12
Apigenin	0.024	0.074	y=(0.08±1.28)+(12.81±0.31)x	0.993	1.910	2.118	2.3	3.1	92.4	0.92	-7.60
Oleuropein	0.003	0.008	*y=(0.91±1.84)+(4.58±0.05)x*	0.998	1.762	1.868	0.7	1.4	97.7	0.96	-4.16
Vanillin	0.018	0.054	y=(-0.46±0.36)+(4.09±0.09)x	0.995	0.176	2.118	2.2	3.1	95.2	0.92	-7.59

Compound	**LOD (mg kg^{-1})**	**LOQ (mg kg^{-1})**	**Equation $y=(a\pm S_a) + (b\pm S_b)x$** (linear range: 0.02-10 mg kg^{-1}; *0.02-100 mg kg^{-1}*)	**r^2**	**Lack of fit**		**Intraday precision RSD_r, (%) (n=6)**	**Interday precision RSD_R, (%) (n×k=3×3)**	**RE %**	**MF**	**ME %**
					F_{calc}	**F_{tab}**					
Ethyl vanillin	0.013	0.039	y=(0.51±0.23)+(2.04±0.06)x	0.991	1.940	2.118	1.6	2.8	95.3	0.94	-6.30
Epicatechin	0.019	0.059	y=(-0.18±0.21)+(3.67±0.05)x	0.998	0.330	2.118	1.8	3.2	94.2	0.96	-4.47
Luteolin	0.002	0.007	y=(0.53±0.40)+(3.64±0.10)x	0.992	1.793	2.118	2.1	3.4	96.9	0.94	-6.34

LOD: limit of detection, LOQ: limit of quantification, r^2: regression coefficient, F_{tab}: $F_{tabulated}$, F_{calc}: $F_{calculated}$, RSD: relative standard deviation, RE: recovery, MF: matrix factor, ME: matrix effect

Linear range for oleuropein, tyrosol and hydroxytyrosol: 0.02-100 mg kg^{-1}; for the rest of the analytes: 0.02-10 mg kg^{-1}

The analytes presented satisfying recovery efficiency (92-99%). The precision limit ranged between 0.7-2.2% for intraday experiments and between 1.4-5.4% for interday experiments, demonstarting the good precision of the optimized method. The method demonstrated low LODs over the range of 0.002 mg kg^{-1} (luteolin) and 0.028 (tyrosol) and adequate LOQs over the range of 0.007 mg kg^{-1} (luteolin) and 0.086 mg kg^{-1} (tyrosol). The analytical curves presented an adequate fit when submitted to the lack-of-fit test ($F_{calculated}$ was less than $F_{tabulated}$ in all cases), with r^2 above 0.99, proving that they can be used for the quantification of the phenolic compounds. The matrix factor ranged between 0.92-0.96 and low matrix supression was observed for all the analytes, up to 7.78%.

In a further step, the 14 target compounds detected were quantified in all EVOO samples on the basis of their reference standards, using syringaldehyde as the internal standard. Quantitative results for the target compounds were expressed as mg kg^{-1} and can be found in the **ESM II, Table S9**.

4.3.3 Suspect screening

In suspect screening, 24 phenolic compounds were tentatively identified in real olive oil samples of Kolovi variety with ion intensities above 800 and peak areas of more than 2000, in all cases. The results presented high mass accuracy (less than 5 ppm) and acceptable isotopic fit values (less than 50 mSigma). The peak score (peak area/peak intensity ratio) ranged between 10 and 22 for all the suspect compounds. MS/MS spectra were examined with Metfrag [91], FooDB [93] and literature records. The list of the fragments of all the identified phenolic compounds were compared and verified with those reported in literature [164] (Chapter 3). The suspect screening results are summarized in the **ESM II, Table S10**, providing information about the identification criteria and the level of identification of each compound.

Scanning for oleuropein aglycone and lingstroside aglycone showed four different peaks. This is not unexpected because oleuropein and lingstroside are stabilized by the presence of the lingstroside residue. The removal of the glucose exposes

the labile hemiacetal carbon that undergoes ring opening and a series of subsequents transformations can occur [29]. **Figure 4.6** summarizes these transformations which can occur physiologically through biotransformations during fruit maturation or olive oil processing. Thus, the single biophenols: oleuropein and lingstroside can give rise to many compounds in olive oil.

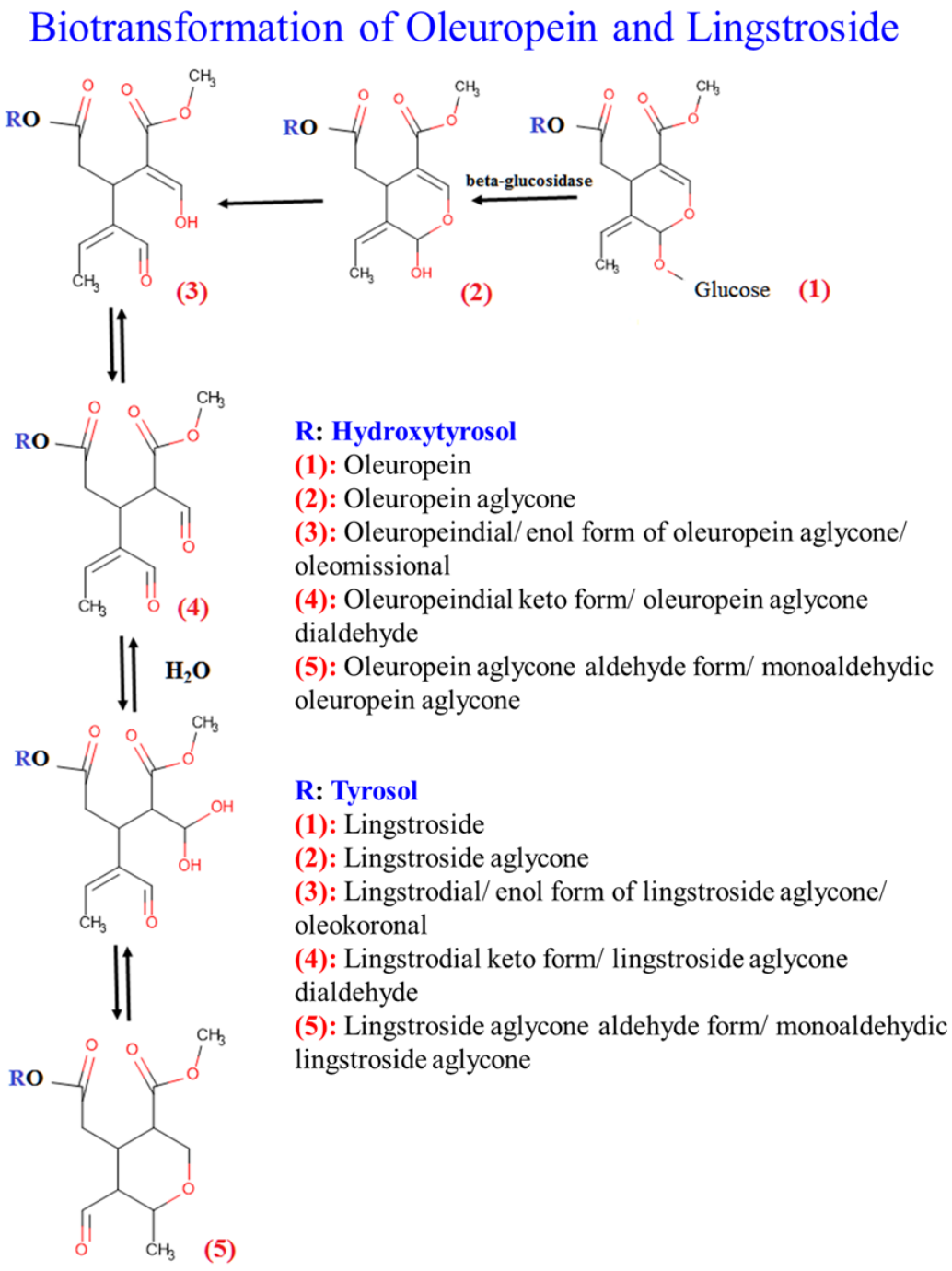


Figure 4.6: Biotransformation of Oleuropein (R: Hydroxytyrosol) and Lingstroside (R: Tyrosol)

The derivatives of oleuropein aglycone, oleuropein aglycone monoaldehydic form, oleuropein aglycone dialdehydic form as well as the enol form of oleuropein aglycone, known as oleomissional [43], were identified at level 3. The qualifier ions of oleuropein aglycone monoaldehydic form (t_R=7.43 min) were detected at m/z: 69.0345, 99.0088, 121.0294, 127.0400, 135.0453, 151.0401 and 163.0400 correspond to C_4H_5O, $C_4H_3O_3$, $C_7H_5O_2$, $C_6H_7O_3$, $C_8H_7O_2$, $C_8H_7O_3$ and $C_9H_7O_3$, respectively. The MS/MS spectrum of the dialdehydic form of oleuropein aglycone (t_R=7.61 min) shows peaks at m/z 59.0139, 67.0187, 95.0138, 123.0453, 128.0478, 153.0558 and 195.0662 that correspond to $C_2H_3O_2$, C_4H_3O, $C_5H_3O_2$, $C_7H_7O_2$, $C_6H_8O_3$, $C_8H_9O_3$ and $C_{10}H_{11}O_4$. Oleomissional elutes at 7.75 min and shows two qualifier ions at m/z: 101.0245 and 163.0400, corresponding to $C_4H_5O_3$ and $C_9H_7O_3$, respectively. More information concerning the proposing transformation of oleuropein and lingstroside, along with the MS/MS spectra of oleuropein aglycone derivatives and lingstroside aglycone EICs can be found in **ESM B**.

The Extracted Ion Chromatogram (EIC) at m/z: 361.1291 presented four different peaks. It has been suggested in literature that lingstroside aglycone eluted at 6.63 min [164]. In the present study, it eluted at 6.65 min, the MS/MS spectra was compared and verified with previously reported fragments [164]. It is possible that the other three peaks with retention times (t_R) 7.81, 8.13 and 8.34 belong to lingstroside agycone monoaldehydic form, the dialdehydic form of oleuropein aglycone and the enol form of lingstroside aglycone, named oleokoronal [43], respectively. These three isomers of lingstroside aglycone were identified at level 3. The MS/MS spectrum of lingstroside aglycone monoaldehydic form shows two qualifier ions at m/z: 137.0608 and 241.0718, corresponding to $C_8H_9O_2$ and $C_{11}H_{13}O_6$, respectively. Next, oleokoronal presents two characteristic fragments at m/z: 195.0663 and 291.0874, corresponding to $C_{10}H_{11}O_4$ and $C_{15}H_{15}O_6$, respectively. In the MS/MS spectra of the dialdehydic form of oleuropein aglycone, the fragments at m/z: 69.0346, 101.0244 and 259.0976 correspond to C_4H_5O, $C_4H_5O_3$ and $C_{15}H_{15}O_4$, respectively.

4.3.4 QSRR applicability domain study

The QSRR model [94] was used for the prediction of the possible retention time of oleuropein aglycone and lingstroside isomers, since there were no reference standards available. The difference of the experimental retention time and the predicted was less than 1 min for all the suspect isomers which were inside the applicability domain of the model. The bubble plot for the screening of oleuropein aglycone and its three isomers, lingstroside aglycone and its three isomers is presented in **Figure 4.7**. Oleuropein aglycone and the isomers oleuropein aglycone monoaldehydic from as well as oleuropein aglycone dialdehydic form are found in box 1, while oleomissional is located in box2. Lingstroside aglycone and its lingstroside aglycone monoaldehydic form, lingstroside aglycone dialdehydic form as well as oleokoronal were within box 1. Therefore, all 8 suspect compounds belong to the applicability domain of the model and the predicted retention time results are highly reliable.

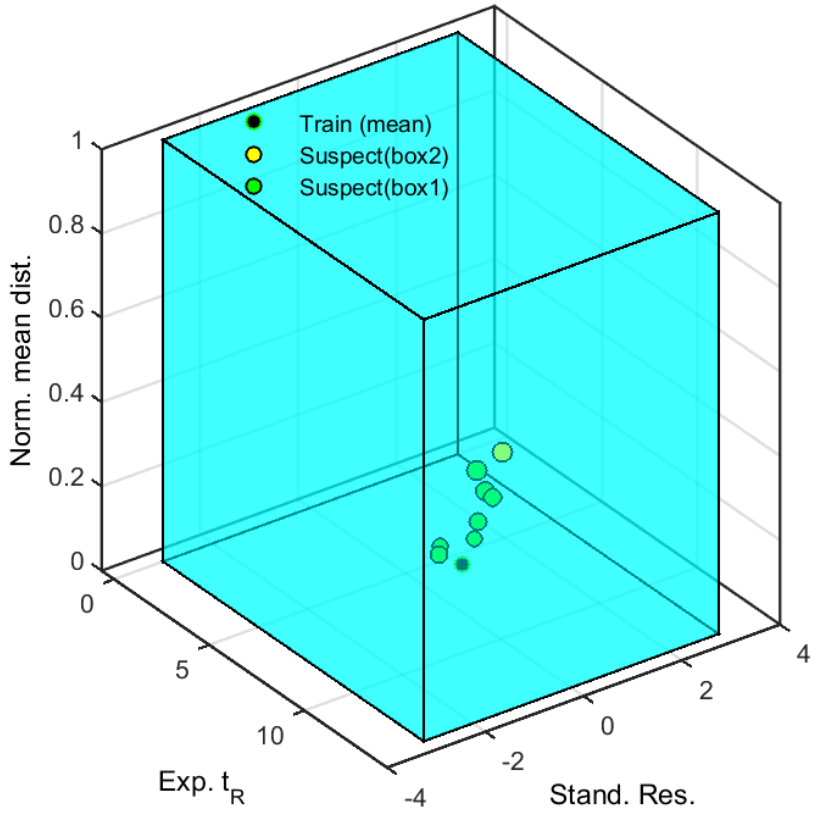


Figure 4.7: The applicability domain study for the predicted retention time of the studied suspect compounds

In a further step, all the suspect compounds were semi-quantified. The lignans syringaresinol, 1-hydroxypinoresinol and the isomer of 1-hydroxypinoresinol as well as 1-acetoxypinoresinol were semi-quantified with the use of pinoresinol calibration curve. The suspect compounds which belong to the class of secoiridoids were semiquantified on the basis of target compounds having similar structure (oleuropein, tyrosol or hydroxytyrosol), after measuring similarity with chemometric tools, as it is described in the following section “Semi-quantification and similarity measurement”.

4.3.5 Semi-quantification and similarity measurement

Similarity indices were performed over 14 compounds so that they could be semi-quantified with the most appropriate standard (**Figure 4.8**). It was found that oleuropein is the most appropriate standard to be used to semi-quantify 10-hydroxy oleuropein aglycone, oleuropein aglycone, lingstroside aglycone, methyl oleuropein aglycone, 10-hydroxy-10methyl oleuropein aglycone, oleomissional and oleoside. Moreover, hydroxylated form of elenolic acid, 10-hydroxy decarboxymethyl oleuropein aglycone, decarboxymethyl oleuropein aglycone, decarboxymethyl lingstroside aglycone, elenolic acid and hydroxytyrosol acetate can be semi-quantified with both tyrosol and hydroxytyrosol. However, the degree of similarity indices is more close to tyrosol. Oleokoronal can also be quantified based on tyrosol as its similarity indices is in the middle of oleuropein and tyrosol.

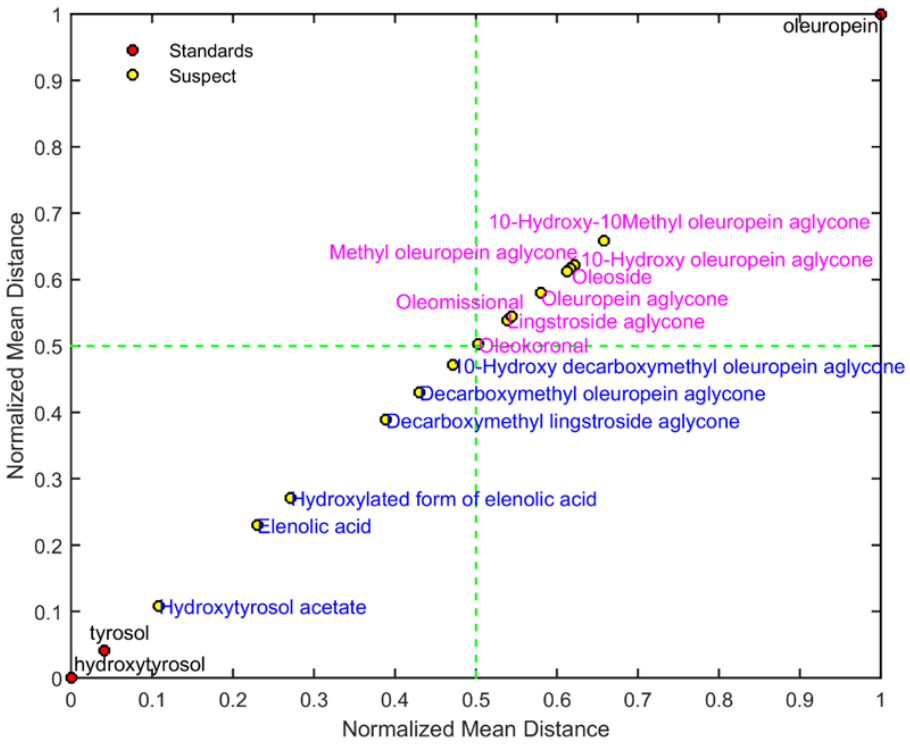


Figure 4.8: Similarity indices calculated between compounds to be semi-quantified

The suspect secoiridoids were semi-quantified, as suggested above, and the semi-quantification concentrations of all the suspect compounds are presented in the **ESM II, Table S11.**

4.3.6 Ant Colony Optimization-Linear Discriminant Analysis (ACO-LDA)

The discrimination between organic and conventional EVOOs was based on the quantification and semi-quantification results of the target and suspect compounds, respectively. LDA tries to separate the samples by increasing the variance between the classes and decreasing the variance within class. For an optimal discrimination, it is essential to know the class posteriors probability. The probability of each sample belonging to the corresponding classes along with the predicted classes (organic and conventional production type of olive oil) is given in the **ESM II, Table S12**. The LDA model was built only based on luteolin as ACO selected it as the most important feature, causing discrimination between classes. The validation criteria were met for ACO-LDA and the miss-discrimination error for training, cross-validation analysis as well as external samples was zero.

4.3.7 Ant Colony Optimization-Random Forest/Random Forest (ACORF/RF)

Using ACO-LDA selects the appropriate compound to discriminate between two classes but it cannot justify at which threshold this discrimination is achieved. Therefore, it is needed to create a tree with defined threshold. After using ACO as a variable selection tool to select the most appropriate compounds (independent variable), luteolin was selected due to the higher contribution to the discrimination problem in RF. Here, using ACO coupled with RF did not affect the outcome, comparing to using RF alone. This was expected since RF has a capability to neglect inclusion of extra variables if the miss-discrimination rate achieved to a minimum value with a single variable. RF generated a simple tree to justify how production type (organic and conventional) of olive oil can be predicted, using the concentration (mg kg^{-1}) of luteolin in a sample. An EVOO is organic if the concentration (mg kg^{-1}) of luteolin is more than 4.16 mg Kg^{-1}, otherwise the EVOO is conventional (**Figure 4.9**).

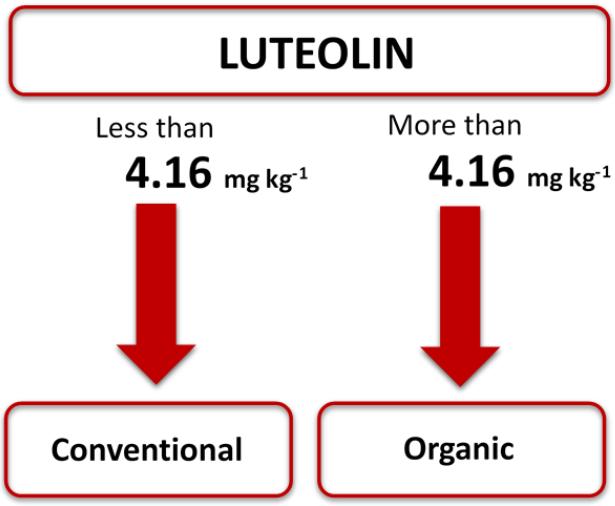


Figure 4.9: Discrimination results of organic and conventional EVOOs using ACO-RF/RF

The validation results of the proposed decision tree suggested that the ACO-RF/RF model shows both internal and external accuracy. The miss-discrimination rate was obtained zero for both training and the external samples. Leave-one-out cross-validation analysis was also indicated that there is not any sample that its removal could affect the outcome of model substantially (the miss-discrimination error was zero). ROC curves were also calculated for both classes and the

results indicated that the discrimination model was well-established with specificity and sensitivity equals to 1.

A well-established discrimination model should show sensitivity towards the changes in the environmental conditions for a two year study. This could be studied by setting EVOOs produced in different years into different evaluation sets. Here, we applied the discrimination models trained on the olive oils produced during the harvesting period of 2015-2016 and tested with those that were produced during the harvesting period 2014-2015. The results show that the marker which is responsible for the discrimination between organic and conventional EVOOs is the same for those that were produced in the previous harvesting period as well.

Luteolin is a predominant flavonoid in olive oil, originating from the glucoside that is present in the drupe, and its concentration highly depends on the geographical area, season and environmental conditions [52]. High luteolin content is crucial due to its antioxidant and other health related activities [179]. Kessen et al. [52] has reported that the concentration of luteolin ranged between 1.51 to 7.57 mg kg^{-1}, showing variations among olive oils of different geographical regions and harvest years. In this research, luteolin exhibited higher values for organic EVOOs compared to conventional, but showed no difference in the content range between the previous harvesting year. Luteolin ranged between 4.16 to 7.03 mg kg^{-1} for organic EVOOs, both harvested in 2014-2015 and 2015-2016. Therefore, it can be a good indicator for the discrimination of organic and conventional EVOOs in different harvesting years, as it is not affected by climate changes. The feature selection algorithm found luteolin as the top compound to make this discrimination possible and predict if a sample is organic or conventional. ACO-RF/RF found that there is a threshold for luteolin in EVOOs of different production types and calculated this threshold at 4.16 mg kg^{-1}. EVOOs with higher concentrations of luteolin are organic and those with lower than 4.16 mg kg^{-1} are predicted as conventional.

4.4 Conclusions

This study contributes to the field of food authenticity by discriminating the organic and conventional EVOOs using an optimized LLME-UHPLC-QTOFMS method. Target and suspect screening quantification results together with ACO-RF established a discrimination model that could reveal markers responsible for the discrimination of production type in EVOOs.

The optimum extraction condition and the selection of the appropriate internal standard were achieved using OFD-RSM, as experimental design and optimization technique. The results showed that the extraction with MeOH:H_2O (80:20, v/v) presents the lowest %RSD values and showed that syringaldehyde 1.30 mg L^{-1} is the most appropriate internal standard.

The proposed method was successfully applied in 52 EVOOs of Kolovi variety produced during the harvesting periods of 2014-2015 and 2015-2016. Totally, 14 target and 24 suspect phenolic compounds were identified. All target compounds were quantified based on their commercially available reference standards, while the identified suspect compounds were semi-quantified according to a novel strategy that incorporates the chemical structure similarity.

A robust discrimination model was established by RF to prioritize the target and suspect phenolic compounds quantification and semi-quantification results, respectively, according to their importance in discrimination task. However, coupling ACO to RF did not change the initial results of RF. Eventually, the flavonoid luteolin was found to be responsible for the discrimination; and if its content is higher than 4.16 mg kg^{-1}, the EVOO is organic; while if it is less than 4.16 mg kg^{-1}, the EVOO should be characterized conventional. The proposed discrimination model, based on 52 samples of Kolovi variety from Lesvos Island within a two year study, is robust showing high internal and external accuracy and thus, it could be sufficiently employed for the discrimination between organic and conventional EVOOs, and further guarantee quality and authenticity.

CHAPTER 5
CLASSIFICATION OF GREEK OLIVE OIL VARIETIES WITH NON-TARGET UHPLC-QTOF-MS AND ADVANCED CHEMOMETRICS

5.1 Introduction

The importance of EVOO phenolic compounds is related to their antioxidant activity and to their contribution to health benefits associated with EVOO consumption [180]. EVOO composition determines its intrinsic quality and could be influenced by several factors, including agronomical and technological factors, such as olive cultivar [46], the climate [45], the degree of maturation [47], crop season [69] and the production process [30, 33, 69]. However, geographical area is greatly responsible for the specific characteristics of VOOs and EVOOs [181]. Olive cultivars, the geographical region along with environmental factors have been reported as the main parameters affecting the chemical profile of EVOOs dominantly [23, 85].

The olive tree (*Olea Europaea L.*) has diverged naturally to many cultivars and is cultivated mainly in the Mediterranean region; Spain, Italy, Greece, Tunisia, Turkey, Morocco and Algeria [9]. The cultivar defines the quality of the drupe and the olive oil [10]. Greece is among the leading olive producing countries of the world, ranked third after Spain and Italy. The number of Greek cultivars is greater than 40 and more than 90% of the territory is cultivated with 20 cultivars [12]. Olive oil produced in Greece has excellent quality and this is because of the local climatic and soil conditions. According to the International Olive Oil Council [13], 70% of Greek production is categorized as EVOO while almost the 35% is exported. Thus, it is imperative for Greece to characterize and authenticate EVOOs based on cultivar and geographical origin in an effort to establish a brand name in the international market.

The European Union has adopted a series of regulations providing guidelines to maintain the Protected Designation of Origin (PDO) and Protected Geographical

Indication (PGI). These include characterization of olive oils based on cultivar and geographical origin to reassure that the quality of the product is closely linked to its territorial and botanical origin and consequently to increase its market value (EC 510/2006) [182]. This regulation states that there is an economic basis for the identification of markers that distinguish PDO EVOOs. Thus, there is an understanding need to enforce the above regulation and develop analytical methods for the authentication of EVOOs and to reassure that the product is closely linked to its territorial origin.

During the past decade, there have been intensive studies for the determination of the cultivar of EVOOs on the basis of different olive oil constituents, including fatty acids, triacylglycerols, sterols, volatiles and phenolic compounds with different analytical methodologies, such as LC-MS [9] LC-TOFMS [23, 25, 84, 183], GC coupled to FID [34, 184, 185] and MS [186, 187], NMR [188] and HPLC coupled to UV [189, 190]. Most of these methods combined chemometrics, as complimentary tool. Recently, Bajoub et al. [191] made an interesting effort applying Knn, PLS-DA and SIMCA. It was concluded that the application of LC coupled to chemometrics, for data treatment, can define the EVOOs varieties with acceptable accuracy. This study, however, could not address the important markers and their contribution behind the classification models. In this field, the literature survey indicates gaps in information, which should be filled in the near future.

LC coupled with HRMS followed by non-target screening strategies and chemometrics could fulfill this gap. LC-HRMS has been widely applied to analyze complex mixtures with wide polarities owing to its high separation efficiency and sensitivity in the identification of compounds at low concentration levels. Recently, this method was successfully applied in two authenticity studies of EVOOs concerning the organoleptic profile and the production type, suggesting markers [85]. One important step in non-target HRMS methods is the prioritization of MS features. In Chapters 3 and 4, two different prioritization tools were introduced to prioritize the MS features and extract the markers that could discriminate between defective olive oils and EVOOs and predict the conventional or organic cultivar. Moreover, a simple and robust decision tree was

established and could define production type and guarantee EVOOs authenticity. The decision tree based approach coupled to HRMS could open new horizons in authenticity studies in the foodomics field.

The objective of this study was to apply an RP-UHPLC-ESI-QTOF-MS method, that has already been developed and optimized in Chapter 4, and combine non-target screening with chemometrics to suggest markers and guarantee the cultivar origin of Greek EVOOs. For this purpose, 51 Greek EVOOs labelled as Amfissis, Chalkidikis, Kolovi, Koroneiki, Ladoelia and Manaki, produced during the harvesting period 2015-2016 were acquired from different regions in Greece. ACO was applied for feature selection with a fitness function representing only the meaningful masses (m/z) that contribute in the EVOOs classification problem. Then, the non-target screening identification workflow was followed including strict rule based filtering steps with deep interpretation of MS/MS and retention time prediction. In addition, a local database consisting of all the natural compounds commonly occurring in olives and olive oils was compiled from FooDB [93] and chemically curated to accelerate the identification of the unknown masses. Finally, RF was employed to classify the EVOOs according to their cultivars.

5.2 Experimental Section

5.2.1 Chemicals and standards

All standards and reagents were of high-purity grade (>95%). MeOH of LC-MS grade and sodium hydroxide (>99%) were purchased from Merck (Darmstadt, Germany). Ammonium acetate (≥99.0%) for HPLC and formic acid (LC-MS Ultra) were purchased from Fluka (Buchs, Switzerland). Isopropanol was acquired from Fisher Scientific (Geel, Belgium). Distilled water was provided by a Milli-Q purification apparatus (Millipore Direct-Q UV, Bedford, MA, USA). Syringaldehyde 98% was acquired from Sigma-Aldrich (Stenheim, Germany) and used as an internal standard. Oleuropein 98%, vanillic acid 97% and pinoresinol 95% were purchased from Sigma-Aldrich (Stenheim, Germany) and luteolin 98% was

purchased from Santa Cruz Biotecnologies. Apigenin (4,5,7 trihydroxyflavone) 97% and tyrosol (2-(4-hydroxyphenyl)ethanol) 98% were purchased from Alfa Aesar (Karlsruche, Germany). Stock standard solutions of individual compounds (1000 mg L^{-1}) were solubilized in methanol and stored at -20 °C in dark brown glass. All intermediate standard solutions containing the analytes were prepared by dilution of the stock solutions in methanol.

5.2.2 Olive oil samples

51 monovarietal EVOOs belonging to five different cultivars were collected from local producers from various regions in Greece. Taking into consideration that the harvest period and the production processing affects the phenolic profile of the EVOOs, all EVOOs under study were collected between December and January 2015-2016. All the samples acquired were cultivated with conventional type of farming, and they were processed with three phase centrifugation technique. Grinding mills were used for grinding in all cases, and the malaxation time was between 45-60 min. In total, 11 samples of the variety Kolovi from Lesvos Island, 9 samples of Chontrolia Chalikidikis, 5 samples of Amfissis, 17 of Koroneiki (8 samples were acquired from Crete and 9 of Peloponnese) 4 samples of Ladoelia and 5 samples of Manaki. **Table 5.1** summarizes the geographical origin of the samples. All samples were collected and stored in dark glass bottles, protected from light and humidity. Nitrogen was inserted as an inert gas to better preserve olive oils and increase the resistance to autoxidation [137].

Table 5.1: Geographical Origin of the monovarietal Greek EVOOs

Samples	Class	Variety	Region	Prefecture	Area
A1	1	Manaki	Central Greece	Fokida	Amfissa
A2	1	Manaki	Central Greece	Fokida	Delfoi
A3	1	Manaki	Central Greece	Fokida	Itea
A4	1	Manaki	Central Greece	Viotia	Arachova
A5	1	Manaki	Central Greece	Fokida	Itea
DL1	2	Ladoelia	Peloponese	Argolis	Argos
DL2	2	Ladoelia	Peloponese	Achaia	Patra
DL3	2	Ladoelia	Peloponese	Korinthia	Korinthos
DL4	2	Ladoelia	Peloponese	Achaia	Patra
X1	3	Chalkidikis	Macedonia	Chalkidiki	Poligiros

Samples	Class	Variety	Region	Prefecture	Area
X2	3	Chalkidikis	Macedonia	Chalkidiki	Poligiros
X3	3	Chalkidikis	Macedonia	Chalkidiki	Sıkia
X4	3	Chalkidikis	Macedonia	Chalkidiki	Kalives
X5	3	Chalkidikis	Macedonia	Chalkidiki	Poligiros
X6	3	Chalkidikis	Macedonia	Chalkidiki	Kassandria
X8	3	Chalkidikis	Macedonia	Chalkidiki	Kassandria
X9	3	Chalkidikis	Macedonia	Chalkidiki	Kalives
K1	4	Koroneiki	Crete	Lasithi (Crete)	Siteia
K2	4	Koroneiki	Crete	Rethimno (Crete)	Rethimno
K3	4	Koroneiki	Crete	Rethimno (Crete)	Rethimno
K4	4	Koroneiki	Crete	Xania (Crete)	Xania
K5	4	Koroneiki	Crete	Rethimno (Crete)	Rethimno
K6	4	Koroneiki	Crete	Rethimno (Crete)	Rethimno
K7	4	Koroneiki	Crete	Xania (Crete)	Xania
K8	4	Koroneiki	Crete	Lasithi (Crete)	Siteia
P1	4	Koroneiki	Peloponese	Ilia (Peloponese)	Zacharo
P2	4	Koroneiki	Peloponese	Ilia (Peloponese)	Zacharo
P3	4	Koroneiki	Peloponese	Lakonia	Neapolis Voion
P4	4	Koroneiki	Peloponese	Messinia	Kalamata
P5	4	Koroneiki	Peloponese	Ilia	Pirgos
P6	4	Koroneiki	Peloponese	Ilia	Amaliada
P7	4	Koroneiki	Peloponese	Messinia	Kalamata
P8	4	Koroneiki	Peloponese	Argolida	Argos
P9	4	Koroneiki	Peloponese	Arkadia	Megalopoli
L1	5	Kolovi	North Aegean	Lesvos	Gera
L2	5	Kolovi	North Aegean	Lesvos	Akrasi
L3	5	Kolovi	North Aegean	Lesvos	Mantamados
L4	5	Kolovi	North Aegean	Lesvos	Plomari
L5	5	Kolovi	North Aegean	Lesvos	Agiasos
L6	5	Kolovi	North Aegean	Lesvos	Gera
L7	5	Kolovi	North Aegean	Lesvos	Agiasos
L8	5	Kolovi	North Aegean	Lesvos	Akrasi
L9	5	Kolovi	North Aegean	Lesvos	Gera
L10	5	Kolovi	North Aegean	Lesvos	Agia Paraskeui
L11	5	Kolovi	North Aegean	Lesvos	Palaiokipos
S1	6	Amfissis	Central Greece	Fthiotida	Stilida
S2	6	Amfissis	Central Greece	Aitoloakarnania	Agrinio
S3	6	Amfissis	Central Greece	Evia	Evia
S4	6	Amfissis	Central Greece	Fthiotida	Stilida
S5	6	Amfissis	Central Greece	Fokida	Amfissa

5.2.3 Sample extraction

Sample preparation was carried out using LLME as it has been previously described in Chapter 3 and Chapter 4. First, 0.5 g of each sample was weighted in 2-mL Eppendorf tubes and, then, it was spiked with 1.30 mg L^{-1}. For the extraction, 0.5 mL of MeOH:H_2O (80:20, *v*/*v*,) was added to the Eppendorf tube. Then, the mixture was vortexed for 2 min and centrifuged for 5 min at 13,400 rpm. Additionally, the supernatant was collected and filtered through membrane syringe filters of regenerated cellulose (CHROMAFIL® RC) (15-mm diameter, 0.22-µm pore size, purchased by Macherey-Nagel, Düren, Germany). Finally, 5 µL of this solution was injected into the chromatographic system. Procedural blanks were prepared and processed in the chromatographic system to detect any potential contamination.

5.2.4 Quality Control

Quality control samples were prepared to confirm that the analytical system has been stabilized before the batch of samples and to evaluate its performance. The quality control sample was prepared by mixing EVOO aliquots and was spiked with a standard solution mix (2 mg L^{-1}) that comprised vanillic acid, oleuropein, luteolin and pinoresinol, so that the final concentration of the QC sample was 1 mg L^{-1}. It was injected at the beginning of the analysis (five times for conditioning), and afterwards, it was injected at regular intervals (every ten sample injections). The %RSDs for the peak areas of the standard compounds were less than 5% (n=10). The retention time shift was in the range 0.09-0.23% RSD (n=10) and mass error was less than 0.28 ppm (n=10), confirming the good performance of the analytical system. The quality control results are summarized in **Table 5.2**.

Table 5.2: Quality Control results

Spiked Standard	**%RSD of Peak Area n=10**	**%RSD of $t_{R\ (min)}$ n=10**	**Δm (±error, ppm) n=10**
Ferulic acid	4.13	0.23	0.20
Oleuropein	4.46	0.09	-0.19

Spiked Standard	%RSD of Peak Area n=10	%RSD of $t_{R\ (min)}$ n=10	Δm (±error, ppm) n=10
Luteolin	4.53	0.06	0.25
Pinoresinol	3.77	0.09	0.28

5.2.5 Instrumental analysis

RP chromatographic analysis was performed using a UHPLC system with an HPG-3400 pump (Dionex UltiMate 3000 RSLC, Thermo Fisher Scientific, Germany) interfaced to a Q-TOF mass spectrometer (Maxis Impact, Bruker Daltonics, Bremen, Germany) in negative electrospray ionization mode. Separation was carried out using an Acclaim RSLC C18 column (2.1 × 100 mm, 2.2 µm) purchased from Thermo Fisher Scientific (Driesch, Germany) with a pre-column of ACQUITY UPLC BEH C18 (1.7 µm, VanGuard Pre-Column, Waters (Ireland)). The separation was operated at column temperature of 30°C. The solvents used consisted of: (A) 90% H_2O, 10% MeOH and 5 mM CH_3COONH_4, (B) 100% MeOH and 5 mM CH_3COONH_4. The adopted elution gradient started with 1% of organic phase B (flow rate 0.2 mL min^{-1}) during one minute, gradually increasing to 39 % for the next 2 minutes, and then increasing to 99.9 % (flow rate 0.4 mL min^{-1}) in the following 11 minutes. These almost pure organic conditions were kept constant for 2 minutes (flow rate 0.48 mL min^{-1}) and then initial conditions (1% B - 99% A) were restored within 0.1 minute (flow rate decreased to 0.2 mL min^{-1}) to re-equilibrate the column for the next injection.

The Q-TOF MS system was equipped with an electrospray ionization interface (ESI), operating in negative mode with the following settings: capillary voltage of 3500 V; end plate offset of 500 V; nebulizer pressure of 2 bar (N_2); drying gas of 8 L min^{-1} (N_2); and drying temperature of 200 °C. A Q-TOF external calibration was daily performed with a sodium formate cluster solution, and a segment (0.1−0.25 min) in every chromatogram was used for internal calibration, using a calibrant injection at the beginning of each run. The sodium formate calibration mixture consisted of 10 mM sodium formate in a mixture of water/isopropanol (1:1). Full scan mass spectra were recorded over the range of 50-1000 m/z, with a scan rate of 2 Hz. MS/MS experiments were conducted using AutoMS data

dependent acquisition mode based on the fragmentation of the five most abundant precursor ions per scan. For certain masses of interest, if the intensity of the m/z was low, a second analysis including the list of the selected precursor ions was carried out in AutoMS (data dependent acquisition) mode. The instrument provided a typical resolving power (FWHM) between 36,000-40,000 at m/z: 226.1593, 430.9137 and 702.8636.

5.2.6 Method Validation

The validation procedure of the RP-UHPLC-ESI-MS method has already been described in Chapter 4. Standard addition curve was constructed for the quantification of vanillic acid. The standard compound was spiked in real EVOO samples at concentrations between 0.02 and 10 mg kg^{-1} (10 calibration levels with 3 replicates at each level). The calibration curve was constructed with the use of the peak area of the spiked analyte subtracted by the peak area of a neat sample and divided by the peak area of the internal standard (syringaldehyde 1.3 mg L^{-1}). LOD and LOQ were calculated at the lowest concentration range of the analytes (0.02-1 mg kg^{-1}), according to the following equations:

$$LOD = \frac{3.3 \times S_a}{b} \quad (1)$$

$$LOQ = \frac{10 \times S_a}{b} \quad (2)$$

Where: S_a is the standard error of the intercept (*a)* and *b* is the slope of the calibration curve. For vanillic acid, LOD was calculated 0.031 mg kg^{-1} and LOQ 0.095 mg kg^{-1}.

5.2.7 Non-target screening protocol

All 51 samples analyzed were converted to mzXML files using ProteoWizard. Further on, these files transferred to R environment to perform peak picking. Among the peak picking workflows [192, 193], XCMS has been widely used in LC-HRMS data processing, owing to high efficiency of *centWave* algorithm,

which is proved to have high performance due to its robust and sensitive detection of potential region-of-interesting mass traces (ROIs). Moreover, noise and baseline correction can be estimated locally for each ROIs offering high F-score (combined measure of recall and precision, calculated from the ground truth features). XCMS has three main internal parameters of *ppm* (which is the tolerated mass deviation), minimum and maximum chromatographic peak width, and *snthresh* ratio which defines the chromatographic signal-to-noise threshold. Optionally, prefilter (intensity threshold (*k,I*)) can be applied to discard lower intense peaks in detected ROIs. Therefore, only those ROI will be retained that contain at least *k* consecutive values with intensity ≥ *I*. A general peak picking workflow also needs a step of retention time correction and alignment (here we used the non-linear retention time alignment wrapping algorithm by loess) [98] as well as peaks group across samples. Filling any missing peaks across samples and the also annotation of extract m/z features are highly needed to prevent adducts/isotopic peaks to cofound with their molecular ions. Here we optimized ppm (23.3), minimum (17.5) and maximum (40) peak width using IPO package in R environment [146, 194]. Signal-to-noise threshold was also set at default value of 10, and prefilter was adjusted at 3-1000. The response surface of these parameters can be found in **Figure 5.1**. Annotations of selected peaks were also done using CAMERA package [145]. A matrix of 51 samples and 287 features (m/z) was generated based on the optimized XCMS object and proceeded to identification and development of classification model. For further identification of these peaks, a new prioritization tool so called Ant Colony Optimization (ACO) was used to limit the searching space of m/zs from 287 to least 4 ones. Afterwards, the non-target screening workflow was applied [164].

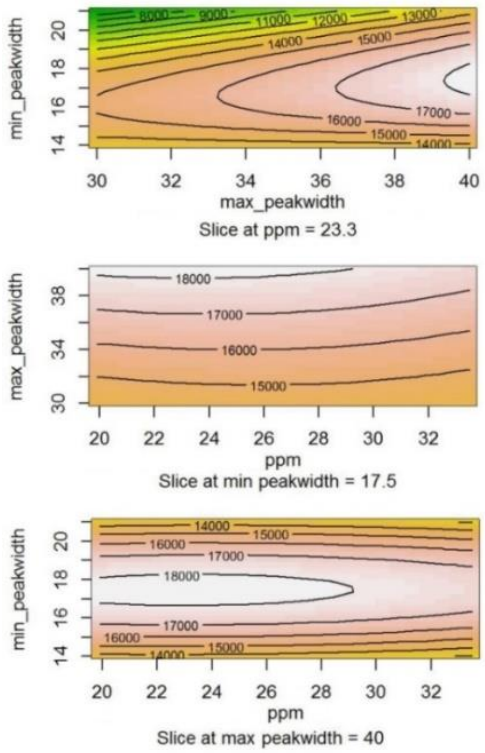


Figure 5.1: Results of response surface methodology for optimizing the internal parameter of XCMS using IPO package

This workflow involves the identification of the selected peaks according to mass accuracy (less than 5 ppm) and isotopic pattern of the precursor ion (less than 100 mSigma), their fragmentation pattern, and the retention time of the chromatographic peak. Extracted ion chromatograms (EICs) were obtained and MS/MS spectra were examined and interpreted. “SmartFormula Manually” tool was applied in Data Analysis 4.1 (Bruker Daltonics, Bremen, Germany) to assign plausible molecular formula(s) to the mass of interest and suggest elemental compositions of the precursor and fragment ions. Then, the prepared local database consisting only the metabolites and natural products that commonly occur in olives or olive oil was uploaded in Metfrag [91]. Then, the exact mass and molecular formula were inserted and the mass error for searching the chemical database was set to 5 ppm. Moreover, the MS/MS fragments with relative intensity were added to elucidate the best candidate(s). Further on, the chromatographic retention time of the tentative candidates was predicted using an in silico approach that is based on quantitative structure retention relationships (QSRR) [94].

The level of confidence achieved in the identification of the detected compounds was established according to Schymanski et al. [142] to ease the communication of identification confidence. Initially, a mass (m/z) of interest corresponding to an unknown compound starts at Level 5 (exact mass of interest). If it is possible to unambiguously assign a molecular formula to this m/z, then it will be upgraded to level 4 (unequivocal molecular formula). If there is sufficient MS (exact mass, isotope or adducts) and experimental information (eg. t_R), non-target components can gain in confidence through level 3 (Tentative Candidate). This level indicates that evidence exists for one or more possible structure(s), but insufficient information is available to eliminate other possible structural candidates (isomers etc.). Nonetheless, if there is a spectral library match for one single structure or if diagnostic evidence is present to exclude all other possible structures from consideration, the compound can reach level 2 (probable structure). Level 2 includes two sublevels; level 2a eg. evidence by matching MS/MS information with literature or spectral library and level 2b denotes diagnostic evidence, such as agreement between predicted and experimental t_R. Finally, if the structure can be confirmed via appropriate measurement of a reference standard with MS, MS/MS fragments and t_R matching, level of identification is 1.

5.2.8 Database preparation

A relevant database is of need to be used to search unknown masses via their experimental MS/MS and in silico fragmentation MS/MS, and confirm their identification according to Schymanski et al. [142]. To this end, a database consisting of compounds commonly occurring in olive/olive oil such as fatty acids, phenolic compounds, amino acids, enzymes, alkaloids, etc. was compiled from FooDB. This list can be found in the **ESM III, Table S1**, including chemical identifiers, predicted retention time and MS information. This list was chemically curated (removing the duplicates, metals, salts, solvents and ambiguous bonding between atoms), following eight main steps [195]: (1) the initially retrieved chemical identifiers (CAS number or SMILES) were unified into InChI; (2) 2D structures of the InChI were created and the dative bonds (*e.g.* nitro group) were standardized using Open Babel (http://openbabel.org/docs/current/) [196]; (3)

salts, metals and solvents were removed from the chemical structure; (4) the octect number was fixed and hydrogens were added; (5) 2D structures were created using Open Babel and 3D structures were obtained out of various tautomer forms (the tautomer with the lowest energy was retained to get one structure out of different forms of a duplicate entries) using Balloon [197]; (6) A SDF file with optimized 3D structures for all entries was created; (7) optimized InChI chemical identifier were derived from the SDF file; (8) duplicates were identified and removed by comparing their optimized InChI from the SDF file. This list can be directly used in MetFrag [91] to elucidate the structure more appropriately and relevantly than other databases, by limiting the searching space to compounds that occurring in olive/olive oil. The compiled database includes the monoisotopic mass, $[M+H]^+$ and $[M-H]^-$, predicted retention time (t_R), molecular formula and chemical identifiers together with reference.

5.2.9 Data processing

Using only annotation results created by CAMERA package and excluding molecular ion adducts may not be as effective as removing them based on their intercorrelation profile. In other words, retaining adducts in the final list of m/z features and keeping track of their respective molecular ion is better if the adducts give reasonably high intensity. To further prevent the highly cofounded features prior performing the classification model, V-WSP algorithm was used as an unsupervised variable reduction method [100]. This method allows the selection of a representative set of variables based on linear correlation (here we set the correlation threshold to 0.8), so that multicollinearity and redundant information in the data can be reduced. Using V-WSP, we reduced the total features from 287 to 250.

5.2.10 Ant Colony Optimization (ACO)

Using a feature selection technique with a fitness function (in this case, it is the misclassification in 51 olive oil samples) can efficiently reveal only the meaningful m/z that contribute in the classification. Further prioritization of m/z was done by

ACO. ACO is a swarm intelligence algorithm that is based on the behavior of the ants searching for the food resources using pheromone deposition [121, 122]. This enables ants to be adoptable to any environmental changes, and find a new shortest path to the resources [122]. ACO is preferably a good method to handle features selection related problems because ants can derive the best combination of subsets that has the minimum fitness objective (here is the miss-classification error). In a typical ACO based features selection case, the algorithm begins with the generation of certain number of ants (here we set this at 100 ants) placed randomly on the graph, which represents the possible combinations of every m/z. Thus, each node (in a graph) relates to a m/z, and each edge shows the traversal of an ant from one m/z to another. The number of artificial pheromone [0,1] for an edge is associated with the popularity of the particular traversal by previous ants. Therefore, ants could make probabilistic decisions to stay at which node and select which edge, based on the artificial pheromone and related traversal degree. This will continue until the minimum degree for the mis-classification error has been reached, otherwise all process will be iterated again [121, 122]. The maximum number of iteration was set to 100 and the desired number of features was set up to 7 features. Evaporation Rate (ER) was also set to 0.05 (this value is kept constant during performing ACO and generally is low value (0.01-0.05)) [122]. ER causes uniformly decrease in all the pheromone values. From a practical point of view, pheromone evaporation is required to prevent a rapid convergence of the algorithm towards a sub-optimal space. ACO algorithm was written and performed in MATLAB.

5.2.11 Random Forest (RF)

RF was used for the classification of EVOOs based on their cultivar. RF was introduced by Breiman [109] and applied for both regression and classification problems. In this regard, the classification is achieved by constructing an ensemble of randomized classification trees. RF starts with creation of several bootstrapped samples (subsamples) from the original matrix of samples and features. This step permits the estimation of the error of the left out samples (those that were not involved during each time of training the classification tree).

Bagging (bootstrap aggregation) is one of the famous methods based on RF, in which several binary decision trees can be fit to a bootstrapped data (subsampled data) [109, 110]. In RF, random samples are replaced ***n*** times from the data (bootstrap sample) to be used as tree seeds. A large portion of these samples is used to form the training set, and the rest of samples are being used as test set; known as out of bag samples (OOB). In this context, apart from the test set, since each tree was fitted by random number of samples, there are some observations that their samples are out of bag in contrast to other trees. These samples are used to calculate OOB error which is referred as prediction error of the model. For each seed and node of a tree, random number of predictors and the decision can be set to finally construct a model with lower OOB error [109, 111]. Finally, a model with large number of trees and the lowest OOB error can be selected from the different prediction trees and with the following variable importance measure [110]:

$$VI(X^f) = \frac{1}{ntree}\sum_{t}\left(\widetilde{OOB}_t^f\, error - OOB_t error\right) \quad (3)$$

Where for each tree (*t*) of a forest: OOB_t is associated with the data, which are not included in the bootstrap samples, to construct *t*. OOB_t*error* is the mean square error (MSE) of a single tree on OOB_t. $\widetilde{OOB}_t^f\, error$ is the error of perturbed sample created by randomly permuting the values of X^f (variable) in OOB_t. RF internally acts as feature selection and only selects features that explain the less OOB error. However, it may avoid exploring other features (m/z) and their contributions to the overall true classification. This is why ACO was coupled with RF to rank each m/z by their importance.

Classification model based on ACO-RF was achieved using miss-classification error in leave-one-out cross validation as fitness. The predictive power of the proposed classification model was evaluated independently using a set of 11 external samples that were not part of the initial training set and confusion matrix was calculated to derive error rate, class specificity and sensitivity [106]. The division into training and test set was achieved by Kennard-Stone algorithm [198]. Kennard-Stone algorithm starts by selecting the pair of points (i samples and m/z

features) that are the furthest apart. The selected samples were assigned to the training sets and removed from the list of samples. Then, the next pair of samples, which are furthest apart, are assigned to the test set. In a third step, the procedure assigns each remaining sample alternatively to the training and test sets based on the distance to the previously selected sample. The distance function used is Euclidean distance. Moreover, Receiver Operating Characteristics (ROC) was calculated to control the accuracy and error rate of proposed model. ROC curves were derived for each class by plotting the sensitivity versus 1-specificity in six cultivars. A reliable classification model would yield a point in the upper left corner of the ROC area, representing maximum sensitivity and specificity, while a random one causes points to be along the diagonal line from the left bottom to the top right corner [106].

The usefulness of feature selection was addressed using unsupervised classification method like Principal Component Analysis (PCA). PCA was applied before and after performing feature selection to investigate whether the covariance explained by PCs increases or not. All the data processing, pretreatment and classification were performed by a homemade program so called ChemoTrAMS in MATLAB environment.

5.3 Results and discussion

5.3.1 Non-target identification

Using 40 EVOOs and 250 features in the training set along with leave-one-out cross validation analysis as fitness function to identify potential m/z features, ACO selected the four most relevant m/zs that could explain the distribution of samples based on their varieties. These selected m/z features could also create a final classification model with miss-classification error of zero for each class. These features were m/z: 167.0345/$t_{R=}$ 2.43, m/z: 299.0561/$t_{R=}$ 8.11, m/z: 269.0456/$t_{R=}$ 8.04 and m/z: 303.1237/$t_{R=}$ 6.50. In an attempt to identify these masses, an inclusion list was created and QTOF system operated in Auto MS/MS mode to obtain the MS/MS spectra of the unknown analytes. Following the non-target screening workflow, EICs were generated in Data Analysis and the most

plausible molecular formulas were determined showing high mass accuracy (less than 2.97 ppm) and acceptable isotopic fit values (less than 15.9 mSigma). The determined molecular formulas were elucidated to certain chemical structures with mass accuracy of ±0.001 ppm.

Specifically, for the mass detected at m/z: 167.0345, the molecular formula $C_8H_8O_4$ was assigned to it using “SmartFormula Manually”, according to the criteria of mass accuracy (2.4 ppm) and isotopic fit (5.9 mSigma). In a further step, the prepared local database search, as introduced in **5.2.8**, was loaded in Metfrag [91]. The MS/MS spectra was examined and verified with MetFrag [91] resulting in 4 candidate compounds. Only 1 tentative compound was scored with 1.0 in Metfrag [91] with all 7 fragments explained, vanillic acid. Predicted t_R with QSRR (3.97 min) was close to the experimental. The corresponding standard was purchased and the presence of vanillic acid in the samples was verified. Vanillic acid is an antioxidant with antioxibacterial, antimicrobial and antifungal activity [35]. The EIC and MS/MS spectrum of vanillic acid is shown in **Figure 5.2** (Identification level: 1).

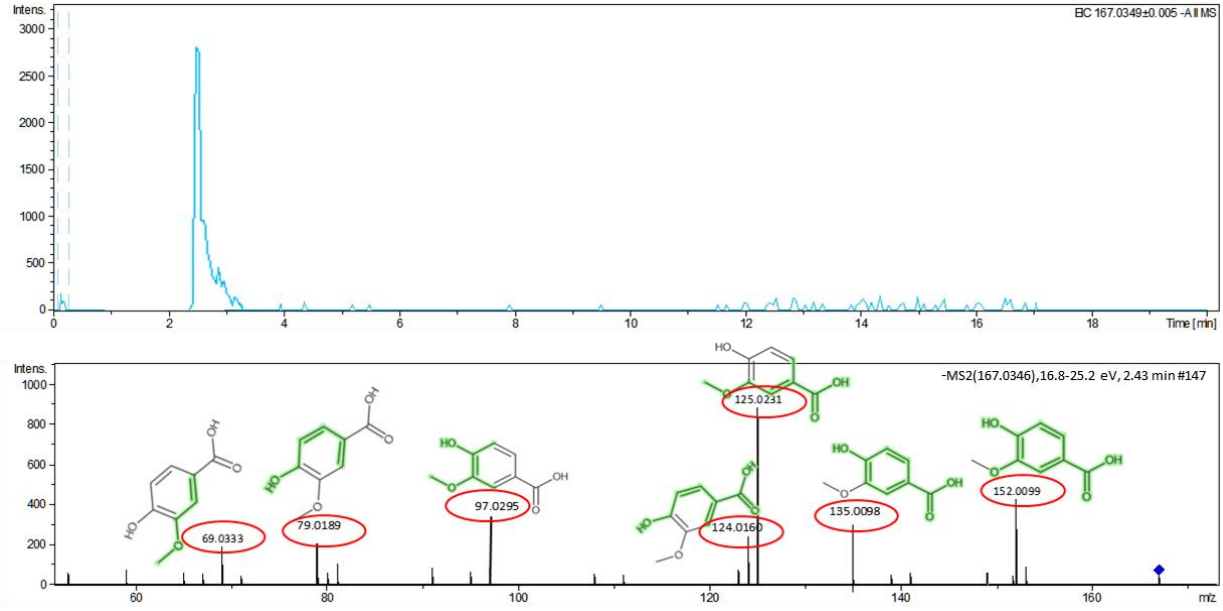


Figure 5.2: EIC and MS/MS spectra with 7 explained fragments of vanillic acid

For the mass with m/z: 269.0456, the molecular formula $C_{15}H_{10}O_5$ was assigned to it using “SmartFormula Manually”, with mass accuracy: 2.97 and isotopic fit:

15.9 mSigma). The local database search resulted in three candidate compounds for that molecular formula. Performing in silico fragmentation with Metfrag [91] using the molecular formula and measured MS/MS revealed 1 tentative candidate with high score (1.0) and all fragments explained, apigenin. The predicted t_R for apigenin with QSRR (6.99 min) was close to the experimental. Finally, the identity of apigenin in the samples was confirmed with a standard. The EIC and MS/MS spectrum of apigenin is shown in **Figure 5.3** (Identification level: 1).

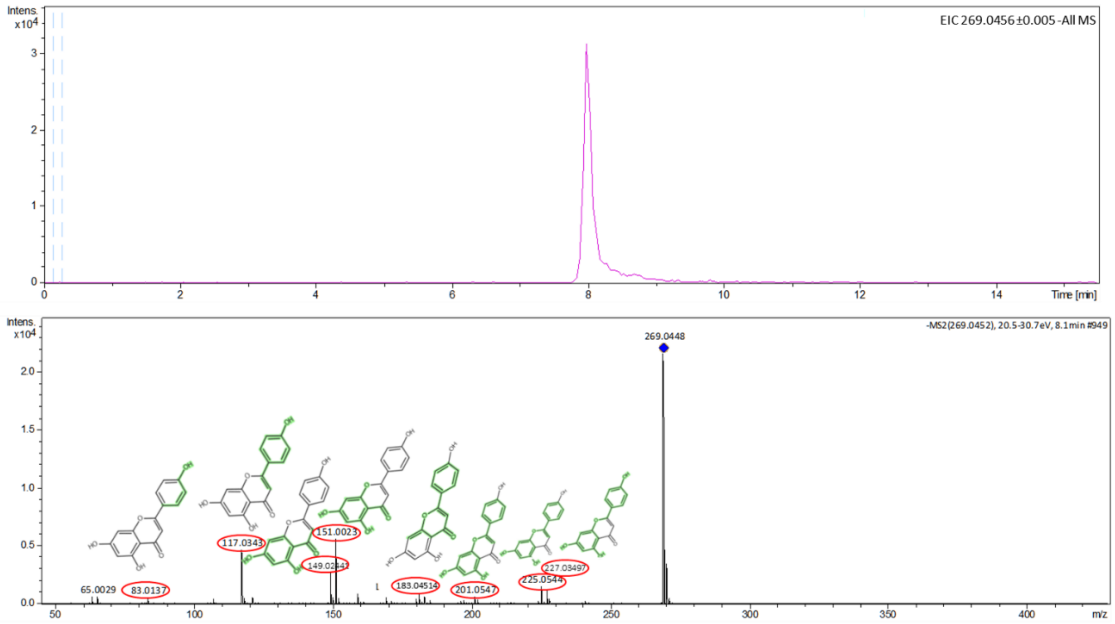


Figure 5.3: EIC and MS/MS spectrum with 8 explained fragments of apigenin

A peak corresponding to m/z: 299.0561 was detected in the EVOO samples. After applying mass accuracy and isotopic fit filters (mass accuracy: 1.08 and isotopic fit: 9.3 Sigma), the molecular formula $C_{15}H_{12}O_6$ was assigned to it. The local database provides 4 possible compounds for this molecular formula. These 4 substances were able to explain all the fragments found in the MS/MS spectrum. In this case, QSRR provided the tentative candidate, luteolin 7-methyl ether, the predicted t_R with QSRR (7.01 min) was close to the experimental. **Figure 5.4** illustrates the EIC and MS/MS spectrum of luteolin 7-methyl ether (Identification level: 2b).

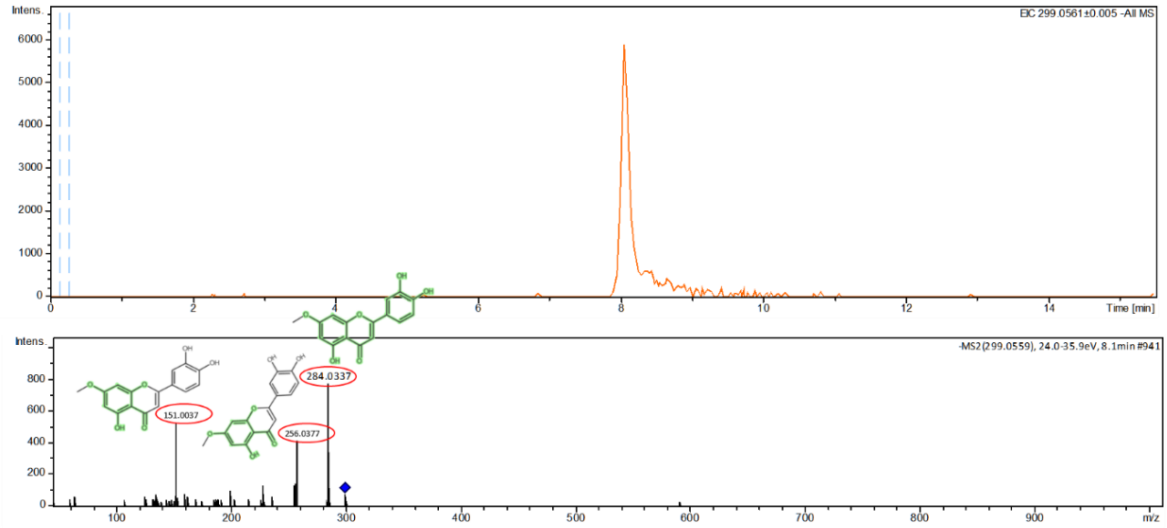


Figure 5.4: EIC and MS/MS spectrum with 3 explained fragments of luteolin 7-methyl ether

For the mass with m/z: 303.1237, the molecular formula $C_{17}H_{20}O_5$ was assigned to it using "SmartFormula Manually", with mass accuracy: 2.64 and isotopic fit: 6.8 mSigma). The local database search resulted in only one candidate compound for that molecular formula, oleocanthal. Performing in silico fragmentation with Metfrag [91] using the molecular formula and measured MS/MS all the fragments were explained. The predicted t_R for oleocanthal with QSRR (7.17 min) was close to the experimental. The peak at m/z: 165.0556 corresponding to $C_9H_9O_3$ has been reported by Kanakis et al. [31] and Dierkes et al. [76]. In addition, the peak at m/z: 183.0663 corresponding to $C_9H_{11}O_4$ has been reported by Dierkes et al. [76] and Bajoub et al. [85]. Oleocanthal shares unique perceptual and anti-inflammatory characteristics with Ibuprofen [42, 199], and acts beneficially against Alzheimer's and Parkinson's diseases [200]. The EIC and MS/MS spectrum of oleocanthal is shown in **Figure 5.5** (Identification level: 2a).

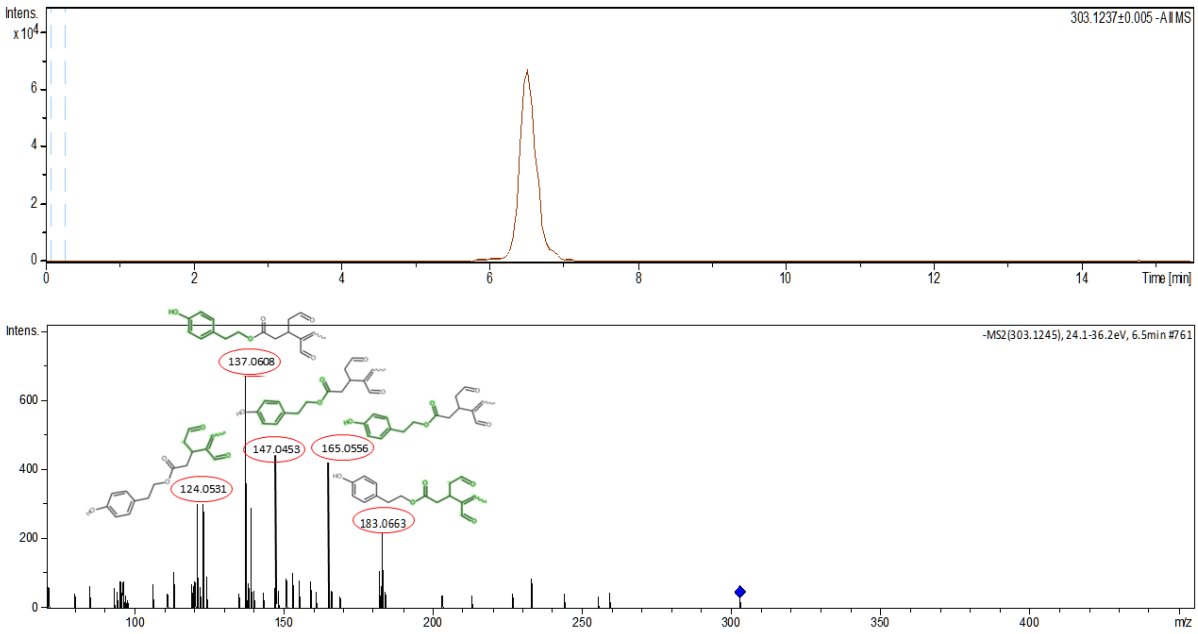


Figure 5.5: EIC and MS/MS spectrum with 3 explained fragments of oleocanthal

5.3.2 Applicability domain

The bubble plot showing the applicability domain of the QSRR model, as it was previously described in Chapter 1, for the 4 predicted compounds can be found in **Figure 5.6**. Apigenin, luteolin 7-methyl ether and oleocanthal are found in box 1, whereas vanillic acid is located in box2. Therefore, all 4 candidate compounds belong to the applicability domain of the model and the predicted retention time results are highly reliable.

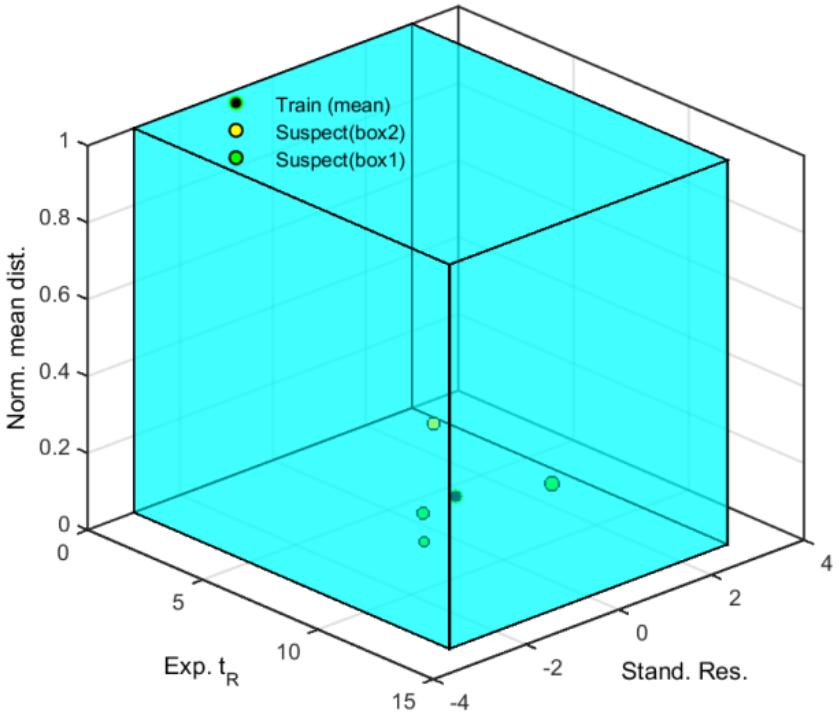


Figure 5.6: The applicability domain study of the QSRR model

5.3.3 Quantification and semi-quantification results

Standard addition calibration curves were constructed for the quantification and semi-quantification of the results. All standard addition calibration curves were constructed with the use of the peak area of the spiked analyte subtracted by the peak area of a neat sample and divided by the peak area of the internal standard (syringaldehyde 1.3 mg L^{-1}). Vanillic acid and apigenin were quantified based on the standard addition curves of their commercial standards. These two standards were spiked in real EVOO samples at concentrations between 0.02 and 10 mg kg^{-1} (10 calibration levels with 3 replicates at each level) and the equations of the curves were: y= [(-0.08±0.07)+(0.73±0.02)x] and y= [(0.37±1.21)+(12.05±0.26)x], respectively; the linear range was 0.02-10 mg kg^{-1} in both cases. For the semi-quantification of luteolin 7-methyl ether, luteolin was spiked in real EVOO samples at concentrations between 0.1-20 mg kg^{-1} (10 calibration levels with 3 replicates at each level). The standard addition curve of luteolin was: y= [(0.69±0.57) + (4.28±0.29)x], and the linear range: 0.1-20 mg kg^{-1}. Oleocanthal was found to have structural similarity with tyrosol in Chapter 4. For the semi-

quantification of oleocanthal, standard addition calibration curve of tyrosol was constructed over the range 1-100 mg kg^{-1} and the equation was: [y = (-2.17±0.03) + (4.41±0.07)x]; linear range: 1-100 mg kg^{-1}.

The quantification and semi-quantification results of the identified markers are presented as mean values mg kg^{-1} (n=3) in the **ESM III, Table S2.**

5.3.4 Principal Component Analysis (PCA)

All in all, two PCs explained 59% of variance and showed appropriate grouping of samples belonging to Manaki, Amfissis and Chalkidikis EVOOs variety. These results are shown in **Figure 5.7** This plot is generated by XCMS online and is based on the intensity of the MS selected by "centWave" algorithm with the same parameters used in peak picking step. It is clear that PCA is not capable of separating and grouping the samples based on their varieties using all MS features. Surprisingly, a significant increase in variance (80.8%) is observed after the selection of four features, followed by their identification and quantification. According to **Figure 5.8**, it can be seen that even though PCA is not a method creating discrimination between samples and it is just showing the distribution of samples and their grouping, samples belonging to Ladoelia, Amfissis, Kolovi and Koroneiki EVOOs varieties could be grouped together. This proves the requirement of feature selection tool to avoid adding false positive MS features inside the loading variables.

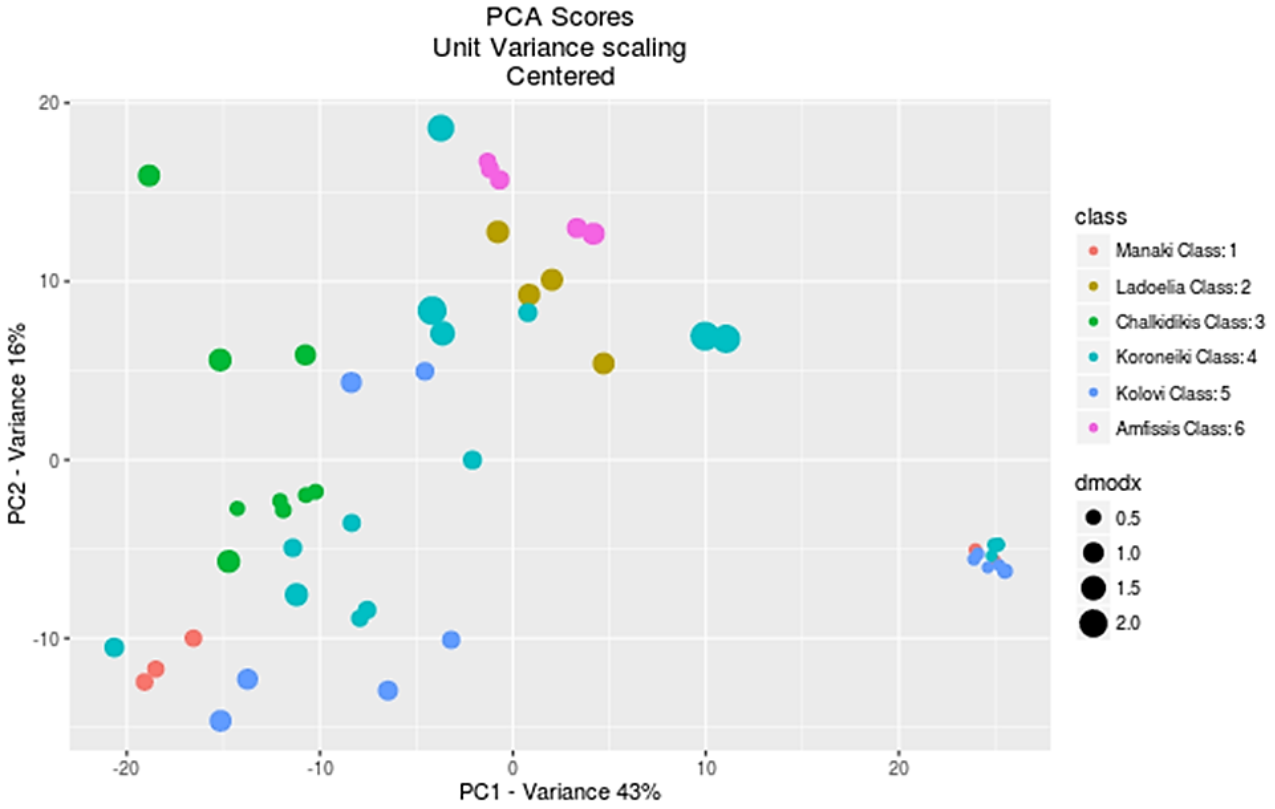


Figure 5.7: PCA with color shows the varietal before MS features prioritization

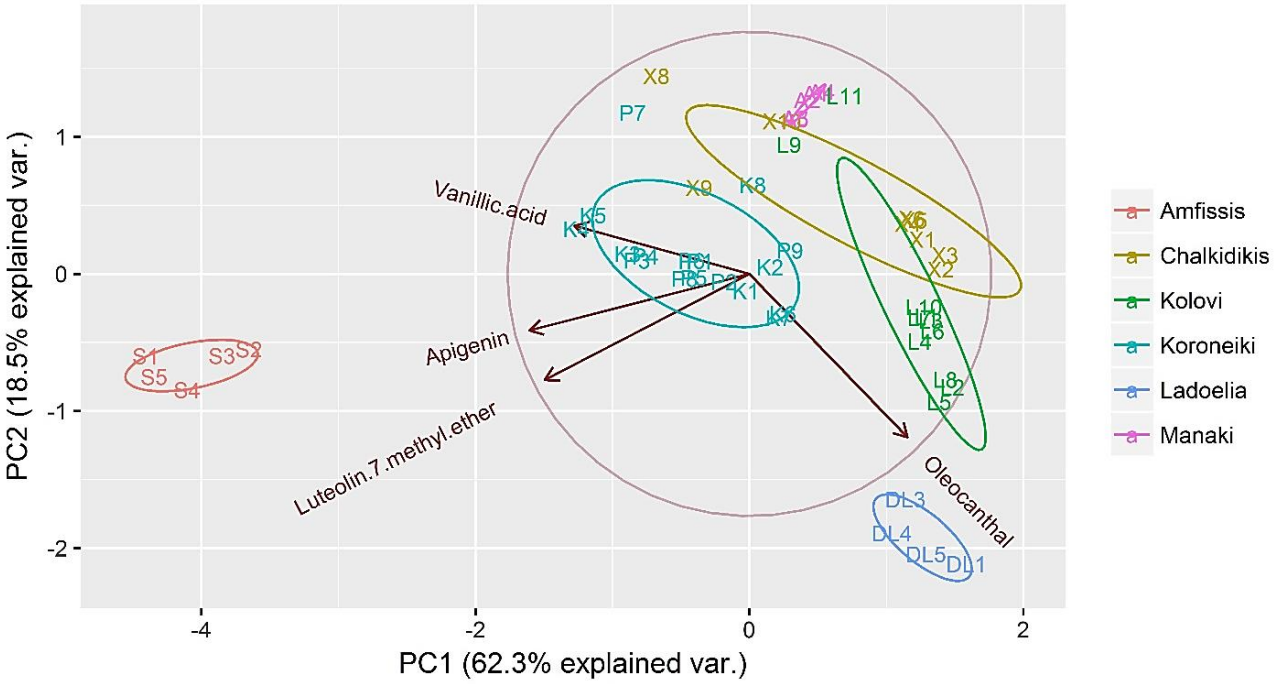


Figure 5.8: PCA with color shows the varietal after MS features prioritization

5.3.5 Ant Colony Optimization-Random Forest (ACO-RF)

Having identified all the selected m/z by ACO, their quantification results were used to build the decision tree using RF. ACO-RF as a validated classification approach generated a graph with a threshold for each identified compound. The validation was done using ROC curve showing the accuracy, specificity and selectivity for each variety along with the error associated with leave-one-out and k-fold cross validation results [106]. K-fold cross validation is a good parameter to judge validity and over-fitting of a classification model as instead of 1 sample per analysis, it excludes several samples out and tries to calculate the error associated with classification model. The number of k was set to 10 and cross validation was performed. The errors after leave-one-out and k-fold cross validation were calculated and found to be very low at 0.075 and 0.175, respectively. ROC curves also described area under curve (AUC), specificity and selectivity at 1.00, which together with cross validation show that the classification model is not over-fitted and can be applied to a suspect external sample. The decision tree developed for the classification of EVOOs according to their cultivar is presented in **Figure 5.9**.

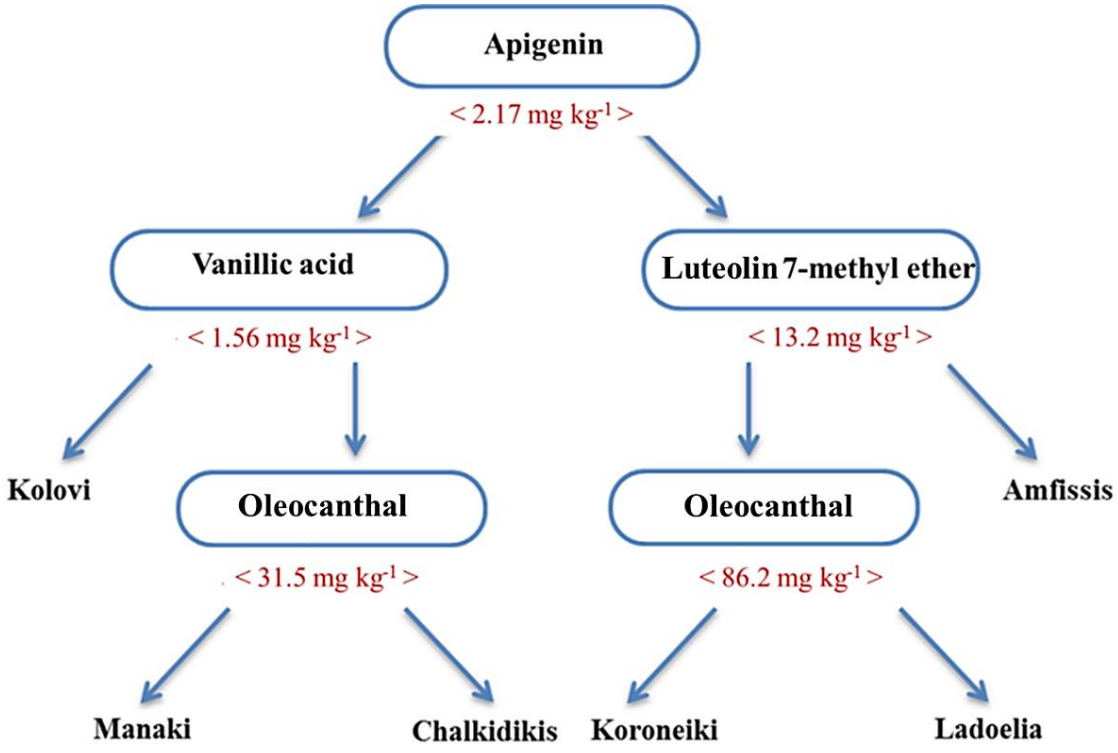


Figure 5.9: Varietal classification of EVOOs according to ACO-RF decision tree

According to this decision tree achieved by ACO-RF, oleocanthal and apigenin play dominant roles. Oleocanthal is important for the discrimination of EVOOs labeled as Manaki or Chalkidikis and Ladoelia or Koroneiki after justifying if they have high or low content of the flavonoid apigenin. Therefore, if the concentration of apigenin is higher than 2.16 mg kg^{-1}, it belongs to the cultivars of Ladoelia, Koroneiki or Amfissis; and if its concentration of apigenin is less than 2.16 mg kg^{-1}, then, it belongs to Manaki, Chalkidikis or Kolovi. Interestingly, vanillic acid was found at lowest concentration (below 1.56 mg kg^{-1}) in EVOOs of Kolovi. On the other hand, when the concentration of apigenin is above 2.16 mg kg^{-1} and the concentration of luteolin 7-methyl ether is above 13.2 mg kg^{-1}, the EVOOs belongs to the variety of Amfissis. This is also observed from PCA (**Figure 5.8**) where the loading plot showed high content of luteolin 7-methyl ether and apigenin, causing Amfissis EVOOs to group together. None of the EVOOs belonging to the other varieties showed similarly high content of luteolin 7-methyl ether. It is also observed that EVOOs with higher content of oleocanthal and apigenin, but lower content of luteolin 7-methyl ether belong to Ladoelia variety, otherwise (if the concentration of oleocanthal is less than 86.2 mg kg^{-1}) they would be classified as Koroneiki. In addition. The PCA loading plot (**Figure 5.8**) showed that Ladoelia EVOOs grouped together, presenting high concentration valued for oleocanthal. EVOOs with higher concentrations of vanillic acid (more than 1.56 mg kg^{-1}), but lower concentration of apigenin (less than 2.16 mg kg^{-1}) belong to either Manaki or Chalkidikis. To discriminate between Manaki and Chalkidikis, the decision tree used again oleocanthal and indicated that EVOOs of the Manaki cultivar have lower oleocanthal content containing (less than 31.5 mg kg^{-1}).

Therefore, the decision tree could simply and easily apply discrimination rule to understand how the EVOO varieties correspond to the chemical profile, while PCA as a commonly used chemomterics tool failed to distribute all the EVOOs based on their varieties.

5.4 Conclusions

This study contributes in the field of food authenticity and guarantees the classification of Greek PDO EVOOs with the application of a non-target screening LLME-UHPLC-QTOFMS method combined with ACO-RF. The proposed method was successfully applied in 51 EVOOs of the Greek cultivars: Amfissis, Chalkidikis, Kolovi, Koroneiki, Ladoelia and Manaki that were produced during the harvesting year 2015-2016.

A peak list consisting of 280 features was generated using the XCMS package, and was processed with chemometrics. PCA failed to distribute the samples based on their cultivars according to the initial non-target list, showing that further m/z prioritization is needed to prevent the incorporation of false positive features, which affect negatively the distribution of the samples. After the implication of ACO and the selection of the 4 most important features, PCA exhibited higher variance and better sample distribution.

Non-target screening workflow was applied in order to identify these 4 markers. In order to accelerate the identification task, a local database consisting of more than 1600 compounds commonly occurring in olive matrices (fatty acids, phenolic compounds, flavonoids, amino acids, enzymes, alkaloids, etc.) was compiled.

Finally, RF established a robust classification that could successfully classify Greek EVOOs, harvested in 2015-2016, into 6 Greek cultivars, setting a concentration threshold for each selected marker. This tree was based on the selection of 4 markers that were identified as apigenin, vanillic acid, luteolin 7-methyl ether and oleocanthal. Based on ACO-RF, it was concluded that the concentration of oleocanthal changes dramatically across Greek olive oil cultivars and has distinguished quantification threshold between Ladoelia (containing more than 86.2 mg kg^{-1}) and Manaki (containing less than 31.5 mg kg^{-1}). Interestingly, apigenin was found to play a crucial role in the prediction of the cultivars. This method set several concentration thresholds (based on the quantification results) over the markers identified, making the authentication task simple.

CHAPTER 6
CONCLUSIONS

Food analysis is continuously requiring the development of more robust, efficient and cost-effective food authentication analytical methods to guarantee the safety, quality, and traceability of food commodities with respect to legislation and consumer demands. This doctoral thesis described the development and optimization of a novel reversed-phase ultra high performance liquid chromatography-electrospray ionization quadrupole time of flight tandem mass spectrometry (RP-UHPLC-ESI-QTOF-MS/MS) analytical method for the investigation of the authenticity of Greek extra virgin olive oils (EVOOs). Integrated target, suspect and non-target screening workflows were applied, followed by advanced data processing, comprehensive data mining and predictive modelling tools for the investigation of critical olive oil authenticity issues that were related to the sensory profile, the production type of the cultivar and the variety.

In the first study, 14 phenolic compounds were identified with target scrrening and 26 with suspect screening. Non-target screening molecular features were obtained and grouped within 19 samples of olive oils using centwave peak picking. These features were then processed with advanced chemometric techniques to build Partial Least Square Discriminant Analysis (PLS-DA) and Counter Propagation Artificial Neural Networks (CP-ANNs) models that classify olive oil into two groups, defective or extra virgin olive oils with high accuracy. A comprehensive workflow was established for the development and validation of the models. The classification models were validated by external evaluation set with known experimental profile.The proposed method suggested a simple, yet effective, approach for studying authenticity (as the organoleptic profile) of olive oils.

In the second study, the RP-UHPLC-ESI-QTOF-MS/MS method was optimized by One Factor Design and Response Surface Methodology to derive the optimal experimental conditions. Target and suspect screening strategies were applied. After the identification, target phenolic compounds were quantified and suspect

compounds were semi-quantified according to a newly developed strategy that incorporates chemical structure similarity with existing standards. Based on the quantification and semi-quantification results of the target and suspect phenolic compounds, a robust discrimination model was established by coupling Ant Colony Optimization (ACO) to Random Forest (RF) in order to differentiate organic olive oils from conventional ones. The model discriminated 52 extra virgin olive oil samples with high accuracy. 41 olive oils produced over the harvesting period 2015-2016 were used as training set to develop the model, and 11 samples produced over the harvesting period 2014-2015 were used as test set. The results showed that the marker, which is responsible for the discrimination between organic and conventional extra virgin olive oils (setting a concentration threshold) is the same for the samples that were produced in the previous harvesting period as well. This work proposed an optimized LLME-LC-QTOF-MS method that enables the identification of phenolic compounds in a wide range of polarities through target and suspect screening. The novel and reliable semi-quantification strategy was developed to define which is the most appropriate reference standard to be used for the semiquantification of suspect compounds, based on chemical structure similarity, considering that, for most phenolic compounds, there are no reference standards commercially available. A robust discrimination model was built in relation to quantification and semi-quantification results, proposing only one marker (luteolin) and its concentration threshold for the discrimination between organic and conventional olive oils.

In the third study, the whole metabolome of 51 Greek EVOOs from 6 different varieties was studied via non-target screening. Peak picking was carried out using the XCMS package and centwave algorithm, resulting in 280 feautures. V-WSP algorithm was used as an unsupervised variable reduction method to decrease the features from 280 to 250. ACO, a significant feature selection technique, could efficiently reveal only the meaningful m/z that contributed in the classification of EVOOs according to their variety. The further couplement of ACO to RF established a robust classification model for the differentiation of EVOOs based on the following varieties: Manaki, Amfissis, Chalkidikis, Kolovi, Koroneiki and Ladoelia, suggesting four markers, with concentration thresholds, as responsible for the classification. The proposed model was robust, showing high

internal and external accuracy, and thus, could sufficiently guarantee olive oil variety and origin.

These studies have made progress towards the fingerprinting of Greek EVOOs, contributing to and increasing the existing knowledge of HRMS-based foodomics and can assure olive oil authenticity, while they can be expanded to different cases of food authenticity, as well.

ACRONYMS AND ABBREVIATIONS

ACN	Acetonitrile
ACO	Ant Colony Optimization
APCI	Atmospheric Pressure Chemical Ionization
API	Atmospheric Pressure Ionization
bbcid	Broad Band Collision Induced Dissociation
CAS	Numerical identifier assigned by Chemical Abstracts Service
CCD	Central Composite Designs
CE	Capillary Electrophoresis
CP-ANNs	Counter Propagation Artificial Neural Networks
CT	Classification trees
DAD	Diode Array
EIC	Extracted Ion Chromatogram
ER	Evaporation Rate
ESI	Electrospray
EVOO	Extra Virgin Olive Oil
Fcalc	Lack of fit calculated
FID	Flame Ionization Detector
Ftab	Lack of fit tabulated
GC	Gas Chromatography
HPLC	High Performance Liquid Chromatography
HR-MS	High Resolution Mass Spectrometers
InChI	International Chemical Identifier
IOC	International Olive Council
IR	Infrared absorption spectroscopy
IT	Ion Trap
k-NN	k-Nearest Neighborhood
LC	Liquid Chromatogrphy
LDA	Linear Discriminant Analysis
LLE	Liquid Liquid Extraction
LOD	Limit of Detection
LOQ	Limit of Quantification

ME	Matrix Effect
MeOH	Methanol
MF	Matrix Factor
NMR	Nuclear Magnetic Resonance
OFD	One Factor Design
PC	Pricipal Component
PCA	Principal Component Analysis
PDO	Protected Designation of Origin
PGI	Protected Geographical Indication
PLS-DA	Partial Least Square Discriminant Analysis
QC	Quality Control
QDA	Quadratic Discriminant Analysis
QqQ	triple quadrupole
QSRR	Quantitative structure-retention relationship
QTOF	Quadropole-Time-of-Flight
RE	Recovery
RF	Random Forest
ROC	Receiver Operating Characteristics
RP	Reversed-Phase
RSD	Relative Standard Deviation
SD	Standard Deviation
SIMCA	Soft Independent Modeling of Class Analogies
SMILES	Simplified Molecular -Input Line- Entry System
SOMs	Self-Organizing Maps
SPE	Solid Phase Extraction
TSG	Traditional Speciality Guaranteed
UHPLC	ultra-high performance liquid chromatography
VIP	Variance in Projection
VOO	Virgin Olive Oil

REFERENCES

1. L. Cuadros-Rodríguez et al., Chromatographic fingerprinting: An innovative approach for food "identitation" and food authentication – A tutorial, *Analytica Chimica Acta*, vol. 909, 2016, pp. 9-23.

2. N.P. Kalogiouri and N.S. Thomaidis, Screening and High-Throughput Multi-Contaminants Methods, *Food Authentication: Management, Analysis and Regulation*, C.A. Georgiou and G.P. Danezis, eds., John Wiley & Sons Ltd., Chichester, UK, 2017, pp. 453-470.

3. E. Pascu, The authenticity and traceability of food - consumers protection form, *Annals of Faculty of Economics*, vol. 1, no. 1, 2013, pp. 658-662.

4. G. P. Danezis, A.S. Tsagkaris, F. Camin, V. Brusic, and C.A. Georgiou, Food authentication: Techniques, trends & emerging approaches, *Trends in Analytical Chemistry*, vol. 85, 2016, pp. 123-132.

5. S. Esslinger, J. Riedl, and C. Fauhl-Hassek, Potential and limitations of non-targeted fingerprinting for authentication of food in official control, *Food Research International*, vol. 60, 2014, pp. 189-204.

6. G.R. Takeoka and S.E. Ebeler, Progress in Authentication of Food and Wine, *Progress in Authentication of Food and Wine, ACS Symposium Series*, vol. 1081, S.E. Ebeler, G.R. Takeoka, P. Winterhalte, eds., American Chemical Society, Washington, DC, 2011, pp. 3-11.

7. J. Dennis, Recent developments in food authentication, *Analyst*, vol. 123, no. 9, 1998, pp. 151-156.

8. A. Cifuentes, Food analysis and foodomics, *Journal of Chromatography A*, vol. 1216, no. 43, 2009, pp. 7109.

9. A. Bakhouche et al., Phenolic characterization and geographical classification of commercial Arbequina extra-virgin olive oils produced in

southern Catalonia, *Food Research International*, vol. 50, no. 1, 2013, pp. 401-408.

10. I. Kosma et al., Characterization and Classification of Extra Virgin Olive Oil from Five Less Well-Known Greek Olive Cultivars, *Journal of the American Oil Chemists' Society*, vol. 93, no. 6, 2016, pp. 837-848.

11. J.-L. Barjol, Introduction, *Handbook of Olive Oil: Analysis and Properties*, R. Aparicio and J. Harwood, eds., Springer, New York 2013, pp. 1-15.

12. N. Kalogeropoulos and M.Z. Tsimidou, Antioxidants in Greek Virgin Olive Oils, *Antioxidants*, vol. 3, no. 2, 2014, pp. 387-413.

13. International Olive Oil Council, *Mission Statement*, 23 Oct. 2015; http://www.internationaloliveoil.org/estaticos/view/100-mission-statement/.

14. International Olive Oil Council, *Trade standard applying to olive oils and olive-pomace oils COI/T.15/NC No 3/Rev. 8*, 2015, pp. 1-16.

15. Commission Regulation (EU) No 1151/2012, of the european parliament and of the council, on quality schemes for agricultural products and foodstuffs, *Official Journal of the European Union*, L343, 2012, pp.1-29.

16. Commission Regulation (EC) No 182/2009, amending the Annex to Regulation (EC) No 1019/2002 on marketing standards for olive oil, *Official Journal of the European Union*, L63, 2009, pp. 6-8.

17. R. Aparicio, L.S. Conte, and H.J. Fiebig, Olive Oil Authentication, *Handbook of Olive Oil: Analysis and Properties, 2nd edition*, R. Aparicio and J. Harwood, eds., Second ed., New York, Springer, 2013, pp. 589-654.

18. Comission Regulation (EC) No 432/2012, establishing a list of permitted health claims, made on foods, other than those referring to the reduction of disease risk and to childrens' development and health, *Official Journal of the European Union*, L-136, 2012, pp. 1-40.

19. J.S. Perona and K.M. Botham, Olive oil as a Functional Food: Nutritional and Health Benefits, *Handbook of Olive Oil: Analysis and Properties*, R. Aparicio and R. Harwood, eds., Second ed., Springer, New York, 2013, pp. 677-714.

20. H.R. Adhami et al., Preparative isolation of oleocanthal, tyrosol, and hydroxytyrosol from olive oil by HPCCC, *Food Chemistry*, vol. 170, 2015, pp. 154-159.

21. E. Anastasopoulos, N. Kalogeropoulos, A.C. Kaliora, A. Kountouri, and N.K. Andrikopoulos, The influence of ripening and crop year on quality indices, polyphenols, terpenic acids, squalene, fatty acid profile, and sterols in virgin olive oil (Koroneiki cv.) produced by organic versus non-organic cultivation method, *International Journal of Food Science & Technology*, vol. 46, no. 1, 2011, pp. 170-178.

22. F. Angerosa, R. Mostallino, C. Basti, and R. Vito, Virgin olive oil odour notes: Their relationships with volatile compounds from the lipoxygenase pathway and secoiridoid compounds, *Food Chemistry*, vol. 68, no. 3, 2000, pp. 283-287.

23. C.A. Ballus et al., Profile of phenolic compounds of Brazilian virgin olive oils by rapid resolution liquid chromatography coupled to electrospray ionisation time-of-flight mass spectrometry (RRLC-ESI-TOF-MS), *Food Chemistry*, vol. 170, 2015, pp. 366-377.

24. S. Cicerale, L. Lucas, and R. Keast, Biological activities of phenolic compounds present in virgin olive oil, *International Journa of Molecular Science*, vol. 11, no. 2, 2010, pp. 458-479.

25. Y. Ouni et al., Characterisation and quantification of phenolic compounds of extra-virgin olive oils according to their geographical origin by a rapid and resolutive LC-ESI-TOF MS method, *Food Chemistry*, vol. 127, no. 3, 2011, pp. 1263-1267.

26. P. Reboredo-Rodriguez, C. Gonzalez-Barreiro, B. Cancho-Grande, and J. Simal-Gandara, Quality of extra virgin olive oils produced in an emerging olive growing area in north-western Spain, *Food Chemistry*, vol. 164, 2014, pp. 418-426.

27. D. Boskou, Olive Fruit, Table Olives, and Olive Oil Bioactive Constituents, *Olive and Olive Oil Bioactive Constituents*, D. Boskou, ed., AOCS Press, Urbana, 2015, pp. 1-30.

28. T.J. Klen, A.G. Wondra, U. Vrhovsek, and B.M. Vodopivec, Phenolic Profiling of Olives and Olive Oil Process-Derived Matrices Using UPLC-DAD-ESI-QTOF-HRMS Analysis, *Journal of Agricultural and Food Chemistry*, vol. 63, no. 15, 2015, pp. 3859-3872.

29. H.K. Obied, D.R. Bedgood, P.D. Prenzler, and K. Robards, Chemical screening of olive biophenol extracts by hyphenated liquid chromatography, *Analytica Chimica Acta*, vol. 603, 2007, pp. 176-189.

30. A. Bendini et al., Phenolic molecules in virgin olive oils: a survey of their sensory properties, health effects, antioxidant activity and analytical methods. An overview of the last decade, *Molecules*, vol. 12, 2007, pp. 1679-1719.

31. P. Kanakis et al., From olive drupes to olive oil. An HPLC-orbitrap-based qualitative and quantitative exploration of olive key metabolites, *Planta Medica*, vol. 79, no. 16, 2013, pp. 1576-1587.

32. A. Termentzi, M. Halabalaki, and A. L. Skaltsounis, From Drupes to Olive Oil: An Exploration of Olive Key Metabolites, *Olive and Olive Oil Bioactive Constituents*, D. Boskou, ed., AOCS Press, Urbana, 2015, pp. 147-177.

33. J. Lozano-Sánchez et al., Prediction of Extra Virgin Olive Oil Varieties through Their Phenolic Profile. Potential Cytotoxic Activity against Human Breast Cancer Cells, *Journal of Agricultural and Food Chemistry*, vol. 58, 2010, pp. 9942-9955.

34. E. Katsoyannos et al., Quality parameters of olive oil from stoned and nonstoned Koroneiki and Megaritiki Greek olive varieties at different maturity levels, *Grasas y Aceites*, vol. 66, no. 1, 2015, doi:/10.3989/gya.0711142.

35. H.K. Obied et al., Bioactivity and analysis of biophenols recovered from olive mill waste, *Journal of Agricultural and Food Chemistry*, vol. 53, no. 4, 2005, pp. 823-837.

36. D. Bourn and J. Prescott, A comparison of the nutritional value, sensory qualities, and food safety of organically and conventionally produced foods, *Critical Reviews in Food Science and Nutrition*, vol. 42, no. 1, 2002, pp. 1-34.

37. A. Cardeno, M. Sanchez-Hidalgo, and C. Alarcon-de-la-Lastra, An update of olive oil phenols in inflammation and cancer: molecular mechanisms and clinical implications. *Current Medicinal Chemistry*, vol. 20, 2013,pp. 4758-4776.

38. L. Grinder-Pedersen et al., Effect of Diets Based on Foods from Conventional versus Organic Production on Intake and Excretion of Flavonoids and Markers of Antioxidative Defense in Humans, *Journal of Agricultural and Food Chemistry*, vol. 51, no. 19, 2003, pp. 5671-5676.

39. S.H. Omar, Oleuropein in olive and its pharmacological effects, *Scientia Pharmaceutica*, vol. 78, no. 2, 2010, pp. 133-154.

40. S. Bulotta et al., Beneficial effects of the olive oil phenolic components oleuropein and hydroxytyrosol: focus on protection against cardiovascular and metabolic diseases, *Journal of Translational Medicine*, vol. 12, no. 219, 2014, doi:10.1186/s12967-014-0219-9

41. E. Antonini et al., Phenolic compounds and quality parameters of family farming versus protected designation of origin (PDO) extra-virgin olive oils, *Journal of Food Composition and Analysis*, vol. 43, 2015, pp. 75-81.

42. G. K. Beauchamp et al., Phytochemistry: Ibuprofen-like activity in extra-virgin olive oil, *Nature*, vol. 437, 2005, pp. 45-46.

43. P. Diamantakos et al., Oleokoronal and oleomissional: new major phenolic ingredients of extra virgin olive oil, *OLIVAE*, vol. 122, 2015, pp. 22-33.

44. A. Taamalli et al., Classification of 'Chemlali' accessions according to the geographical area using chemometric methods of phenolic profiles analysed by HPLC–ESI-TOF–MS, *Food Chemistry*, vol. 132, no. 1, 2012, pp. 561-566.

45. O. Baccouri et al., Chemical composition and oxidative stability of Tunisian monovarietal virgin olive oils with regard to fruit ripening, *Food Chemistry*, vol. 109, no. 4, 2008, pp. 743-754.

46. D. Tura et al., Influence of cultivar and site of cultivation on levels of lipophilic and hydrophilic antioxidants in virgin olive oils (Olea Europea L.) and correlations with oxidative stability, *Scientia Horticulturae*, vol. 112, no. 1, 2007, pp. 108-119.

47. L. Cerretani et al., Preliminary characterisation of virgin olive oils obtained from different cultivars in Sardinia, *European Food Research and Technology*, vol. 222, no. 3-4, 2005, pp. 354-361.

48. M. Servili et al., Phenolic compounds in olive oil: antioxidant, health and organoleptic activities according to their chemical structure, *Inflammopharmacology*, vol. 17, 2009, pp. 76-84.

49. A. de Torres, F. Espínola, M. Moya, and E. Castro, Composition of secoiridoid derivatives from Picual virgin olive oil using response surface methodology with regard to malaxation conditions, fruit ripening, and irrigation management, *European Food Research and Technology*, vol. 242, no. 10, 2016, pp. 1709-1718.

50. Y. Ouni, G. Flamini, and M. Zarrouk, The Chemical Properties and Volatile Compounds of Virgin Olive Oil from Oueslati Variety: Influence of Maturity

Stages in Olives, *Journal of the American Oil Chemists' Society*, vol. 93, no. 9, 2016, pp. 1265-1273.

51. A. Bajoub et al., Quality and chemical profiles of monovarietal north Moroccan olive oils from "Picholine Marocaine" cultivar: registration database development and geographical discrimination, *Food Chemistry*, vol. 179, 2015, pp. 127-136.

52. S. Kesen, H. Kelebek, and S. Selli, LC–ESI–MS Characterization of Phenolic Profiles Turkish Olive Oils as Influenced by Geographic Origin and Harvest Year, *Journal of the American Oil Chemists' Society*, vol. 91, no. 3, 2014, pp. 385-394.

53. R. Aparicio et al., Authenticity of olive oil: Mapping and comparing official methods and promising alternatives, *Food Research International*, vol. 54, no. 2, 2013, pp. 2025-2038.

54. R. Briante et al., Changes in phenolic and enzymatic activities content during fruit ripening in two Italian cultivars of Olea europaea L., *Plant Science*, vol. 162, 2002, pp. 791-798.

55. P. Inglese et al., Factors affecting Extra-Virgin Olive Oil composition, *Horticultural Reviews*, vol. 38, J. Janik, ed., John Wiley & Sons, Hoboken, New Jersey, USA, 2011, pp. 83-148.

56. A. Piscopo and M. Poiana, Packaging and Storage of Olive Oil, *Olive Germplasm-The Olive cultivation, Table Olive and Olive Oil Industry in Italy*, I. Muzzalupo, ed., 2012, doi:10.5772/51827

57. M.I. Alarcón Flores, R. Romero-González, A. Garrido Frenich, and J.L. Martínez Vidal, Analysis of phenolic compounds in olive oil by solid-phase extraction and ultra high performance liquid chromatography-tandem mass spectrometry, *Food Chemistry*, vol. 134, no. 4, 2012, pp. 2465-2472.

58. M.Z. Tsimidou, Analytical Methodologies: Phenolic Compounds Related to Olive Oil Taste Issues, *Handbook of Olive Oil: Analysis and Properties*, R. Aparicio and J. Harwood, eds., Springer, New York 2013, pp. 311-334.

59. I. Lauri et al., Application of "magnetic tongue" to the sensory evaluation of extra virgin olive oil, *Food Chemistry*, vol. 140, no. 4, 2013, pp. 692-699.

60. A. Rosati et al., Effect of agronomical practices on carpology, fruit and oil composition, and oil sensory properties, in olive (Olea europaea L.), *Food Chemistry*, vol. 159, 2014, pp. 236-243.

61. M.E. Mora-Ruiz et al., Assessment of polar phenolic compounds of virgin olive oil by NIR and mid-IR spectroscopy and their impact on quality, *European Journal of Lipid Science and Technology*, vol. 119, no. 1, 2017, pp. 1-7.

62. A. Hirri, M. Bassbasi, S. Souhassou, F. Kzaiber, and A. Oussama Prediction of Polyphenol Fraction in Virgin Olive Oil Using Mid-Infrared Attenuated Total Reflectance Attenuated Total Reflectance Accessory-Mid-Infrared Coupled with Partial Least Squares Regression, *International Journal of Food Properties*, vol. 19, no. 7, 2016, pp. 1504-1512

63. I. Vulcano, M. Halabalaki, L. Skaltsounis, and M. Ganzera, Quantitative analysis of pungent and anti-inflammatory phenolic compounds in olive oil by capillary electrophoresis, *Food Chemistry*, vol. 169, 2015, pp. 381-386.

64. M.A. de Fernandez, V.C. Sotovargas, and M.F. Silva, Phenolic Compounds and Antioxidant Capacity of Monovarietal Olive Oils Produced in Argentina, *Journal of the American Oil Chemists' Society*, vol. 91, no. 12, 2014, pp. 2021-2033.

65. G. Purcaro et al., Evaluation of total hydroxytyrosol and tyrosol in extra virgin olive oils, *European Food Research and Technology*, vol. 116, no. 7, 2014, pp. 805-811.

66. R. García-Villalba et al., Gas chromatography-atmospheric pressure chemical ionization-time of flight mass spectrometry for profiling of phenolic compounds in extra virgin olive oil., *Journal of Chromatography A*, vol. 1218, no. 7, 2011, pp. 959-971.

67. I. Romero et al., Characterization of Virgin Olive Oils with Two Kinds of "frostbitten Olives" Sensory Defect, *Journal of Agricultural and Food Chemistry* vol. 64, no. 27, 2016, pp. 5590-5597.

68. M.D.P. Godoy-Caballero, T. Galeano-Díaz, and M. Isabel Acedo-Valenzuela, Simple and fast determination of phenolic compounds from different varieties of olive oil by nonaqueous capillary electrophoresis with UV-visible and fluorescence detection, *Journal of Separation Science*, vol. 35, no. 24, 2012, pp. 3529-3539.

69. D. Alkan, F. Tokatli, and B. Ozen, Phenolic Characterization and Geographical Classification of Commercial Extra Virgin Olive Oils Produced in Turkey, *Journal of the American Oil Chemists' Society*, vol. 89, no. 2, 2011, pp. 261-268.

70. M. Tasioula-Margari and E. Tsabolatidou, Extraction, Separation, and Identification of Phenolic Compounds in Virgin Olive Oil by HPLC-DAD and HPLC-MS, *Antioxidants*, vol. 4, no. 3, 2015, pp. 548-562.

71. M. Becerra-Herrera, M. Sánchez-Astudillo, R. Beltrán, and A. Sayago, Determination of phenolic compounds in olive oil: New method based on liquid–liquid micro extraction and ultra high performance liquid chromatography-triple–quadrupole mass spectrometry, *LWT - Food Science and Technology*, vol. 57, no. 1, 2014, pp. 49-57.

72. V. Sanchez de Medina, F. Priego-Capote, and M.D.L. de Castro, Characterization of monovarietal virgin olive oils by phenols profiling, *Talanta*, vol. 132, 2015, pp. 424-432.

73. F. Mazzotti et al., Assay of tyrosol and hydroxytyrosol in olive oil by tandem mass spectrometry and isotope dilution method, *Food Chemistry*, vol. 135, no. 3, 2012, pp. 1006-1010.

74. D. Caruso et al., Rapid evaluation of phenolic component profile and analysis of oleuropein aglyeon in olive oil by atmospheric pressure

chemical ionization- mass spectrometry (APCI-MS), *Journal of Agricultural and Food Chemistry*, vol. 48, no. 4, 2000, pp. 1182-1185.

75. R. García-Villalba et al., Characterization and quantification of phenolic compounds of extra-virgin olive oils with anticancer properties by a rapid and resolutive LC-ESI-TOF MS method, *Journal of Pharmaceutical and Biomedical Analysis*, vol. 51, no. 2, 2010, pp. 416-429.

76. G. Dierkes et al., High-performance liquid chromatography-mass spectrometry profiling of phenolic compounds for evaluation of olive oil bitterness and pungency, *Journal of Agricultural and Food Chemistry*, vol. 60, no. 31, 2012, pp. 7597-7606.

77. S. Vichi, N. Cortes-Francisco, and J. Caixach, Insight into virgin olive oil secoiridoids characterization by high-resolution mass spectrometry and accurate mass measurements, *Journal of Chromatography A*, vol. 1301, 2013, pp. 48-59.

78. S. Fu et al., Characterization of isomers of oleuropein aglycon in olive oils by rapid-resolution liquid chromatography coupled to electrospray time-of-flight and ion trap tandem mass spectrometry, *Rapid Communications in Mass Spectrometry*, vol. 23, 2009, pp. 51-53.

79. T. Michel et al., UHPLC-DAD-FLD and UHPLC-HRMS/MS based metabolic profiling and characterization of different Olea europaea organs of Koroneiki and Chetoui varieties, *Phytochemistry Letters*, vol. 11, 2015, pp. 424-439.

80. M. Krauss, H. Singer, and J. Hollender, LC-high resolution MS in environmental analysis: from target screening to the identification of unknowns, *Analytical and Bioanalytical Chemistry*, vol. 397, no. 3, 2010, pp. 943-951.

81. P. Gago-Ferrero et al., Extended Suspect and Non-Target Strategies to Characterize Emerging Polar Organic Contaminants in Raw Wastewater

with LC-HRMS/MS, *Environmental Science and Technology*, vol. 49, no. 20, 2015, pp. 12333-12341.

82. A. Bajoub et al., Comprehensive 3-year study of the phenolic profile of Moroccan monovarietal virgin olive oils from the Meknes region, *Journal of Agricultural and Food Chemistry*, vol. 63, no. 17, 2015, pp. 4376-4385.

83. A.L. Capriotti et al., Comparison of extraction methods for the identification and quantification of polyphenols in virgin olive oil by ultra-HPLC-QToF mass spectrometry, *Food Chemistry*, vol. 158, 2014, pp. 392-400.

84. B. Gilbert-López et al., Determination of Polyphenols in Commercial Extra Virgin Olive Oils from Different Origins (Mediterranean and South American Countries) by Liquid Chromatography–Electrospray Time-of-Flight Mass Spectrometry, *Food Analytical Methods*, vol. 7, no. 9, 2014, pp. 1824-1833.

85. A. Bajoub et al., Comparing two metabolic profiling approaches (liquid chromatography and gas chromatography coupled to mass spectrometry) for extra-virgin olive oil phenolic compounds analysis: A botanical classification perspective, *Journal of Chromatography A*, vol. 1428, 2016, pp. 267-279.

86. A. Bakhouche et al., A new extraction approach to correct the effect of apparent increase in the secoiridoid content after filtration of virgin olive oil, *Talanta*, vol. 127, 2014, pp. 18-25.

87. A. Bakhouche et al., Time course of Algerian Azeradj extra-virgin olive oil quality during olive ripening, *European Journal of Lipid Science and Technology*, vol. 117, no. 3, 2015, pp. 389-397.

88. A. Loubiri et al., Usefulness of phenolic profile in the classification of extra virgin olive oils from autochthonous and introduced cultivars in Tunisia, *European Food Research and Technology*, vol. 243, no. 3, 2016, pp. 467-479.

89. J. Lozano-Sánchez et al., Monitoring the bioactive compounds status of extra-virgin olive oil and storage by-products over the shelf life, *Food Control*, vol. 30, no. 2, 2013, pp. 606-615.

90. A.A. Bletsou et al., Targeted and non-targeted liquid chromatography-mass spectrometric workflows for identification of transformation products of emerging pollutants in the aquatic environment, *Trends in Analytical Chemistry*, vol. 66, 2015, pp. 32-44.

91. S. Wolf, S. Schmidt, M. Muller-Hannemann, and S. Neumann, In silico fragmentation for computer assisted identification of metabolite mass spectra, *BMC Bioinformatics*, vol. 11, 2010, pp. 148.

92. H. Horai et al., MassBank: a public repository for sharing mass spectral data for life sciences, *Journal of Mass Spectrometry*, vol. 45, no. 7, 2010, pp. 703-714.

93. FooDB, The Food Components Database, 2016; http://foodb.ca/.

94. R. Aalizadeh, N.S. Thomaidis, A.A. Bletsou, and P. Gago-Ferrero, Quantitative Structure-Retention Relationship Models To Support Nontarget High-Resolution Mass Spectrometric Screening of Emerging Contaminants in Environmental Samples, *Journal of Chemical Information and Modeling*, vol. 56, no. 7, 2016, pp. 1384-1398.

95. T. Kind and O. Fiehn, Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry, *BMC Bioinformatics*, vol. 8, no. 105, 2007, pp. 1-20.

96. ChemSpider, The free Chemical Database, 2001; http://www.chemspider.com/.

97. PubChem, National Center for Biotechnology Information, PubChem BioAssay Database, 2002; https://pubchem.ncbi.nlm.nih.gov/.

98. R. Tautenhahn, C. Böttcher, and S. Neumann, Highly sensitive feature detection for high resolution LC/MS, *BMC Bioinformatics*, vol. 9, 2008, pp. 504.

99. T. Pluskal, S. Castillo, A. Villar-Briones, and M. Orešič, MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data, *BMC Bioinformatics*, vol. 11, no. 395, 2010, doi:10.1186/1471-2105-11-395

100. D. Ballabio et al., A novel variable reduction method adapted from space-filling designs, *Chemometrics and Intelligent Laboratory Systems*, vol. 136, 2014, pp. 147-154.

101. N.P. Kalogiouri, R. Aalizadeh, and N.S. Thomaidis, Investigating the organic and conventional production type of olive oil with target and suspect screening by LC-QTOF-MS, a novel semi-quantification method using chemical similarity and advanced chemometrics, *Analytical and Bioanalytical Chemistry*, 2017, doi: 10.1007/s00216-017-0395-6.

102. D.L. Massart, ed., *Handbook of Chemometrics and Qualimetrics-data Handling in Science and Technology*, vol. 20A, first ed., Elsevier, Amsterdam, Netherlands, 1997.

103. H. Abdi and L.J. Williams, Principal component analysis, *WIREs Computational Statistics*, vol. 2, no. 4, 2010, pp. 433-459.

104. A. M. Martinez and A.C. Kak, PCA versus LDA, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, 2001, pp. 228-233.

105. S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.R. Mullers, Fisher discriminant analysis with kernels, *Neural Networks for Signal Processing -- Proceedings of the IEEE Workshop,* IEEE, Piscataway, New Jersey, USA, 2002, pp. 41-48.

106. D. Ballabio and V. Consonni, Classification tools in chemistry. Part 1: linear models. PLS-DA, *Analytical Methods*, vol. 5, no. 16, 2013, pp. 3790-3798.

107. M. Barker and W. Rayens, Partial least squares for discrimination, *Journal of Chemometrics*, vol. 17, no. 3, 2003, pp. 166-173.

108. R. G. Brereton and G. R. Lloyd, Partial least squares discriminant analysis: taking the magic away, *Journal of Chemometrics*, vol. 28, 2014, pp.213-225.

109. L. Breiman, Random Forests, *Machine Learning*, vol. 45, no. 1, 2001, pp. 5-32.

110. A. Liew and M. Wiener, Classification and Regression by randomForest, *R News*, vol. 2, 2002, pp. 18-22.

111. A. Hapfelmeier and K. Ulm, Variable selection by Random Forests using data with missing values, *Computational Statistics & Data Analysis*, vol. 80, 2014, pp. 129-139.

112. T. Kohonen, Self-Organizing Feature Maps, *Self-Organization and Associative Memory*, Springer, Heidelberg, Berlin, 1989, pp. 119-157.

113. F. Marini, Artificial neural networks in foodstuff analyses: Trends and perspectives A review, *Analytica Chimica Acta*, vol. 635, no. 2, 2009, pp. 121-131.

114. D. Ballabio and M. Vasighi, A MATLAB toolbox for Self Organizing Maps and supervised neural network learning strategies, *Chemometrics and Intelligent Laboratory Systems*, vol. 118, 2012, pp. 24-32.

115. I. Kuzmanovski and M. Novič, Counter-propagation neural networks in Matlab, *Chemometrics and Intelligent Laboratory Systems*, vol. 90, no. 1, 2008, pp. 84-91.

116. A.P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms, *Pattern Recognition*, vol. 30, no. 7, 1997, pp. 1145-1159.

117. J.A. Hanley and B.J. McNeil, The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Radiology*, vol. 143, no. 1, 1982, pp. 29-36.

118. M.H. Zweig and G. Campbell, Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine, *Clinical Chemistry*, vol. 39, no. 4, 1993, pp. 561-577.

119. T. Fawcett, An introduction to ROC analysis, *Pattern Recognition Letters*, vol. 27, no. 8, 2006, pp. 861-874.

120. R. Gosselin, D. Rodrigue, and C. Duchesne, A Bootstrap-VIP approach for selecting wavelength intervals in spectral imaging applications, *Chemometrics and Intelligent Laboratory Systems*, vol. 100, no. 1, 2010, pp. 12-21.

121. M. Dorigo, M. Birattari, and T. Stützle, Ant colony optimization artificial ants as a computational intelligence technique, *IEEE Computational Intelligence Magazine*, vol. 1, no. 4, 2006, pp. 28-39.

122. M. Dorigo and C. Blum, Ant colony optimization theory: A survey, *Theoretical Computer Science*, vol. 344, no. 2-3, 2005, pp. 243-278.

123. K. Romdhane, Food authenticity and fraud, *Chemical analysis of food: techniques and applications*, Y. Picó, ed., Academic Press, New York, 2012, pp. 499-517.

124. International Olive Oil Council, *Sensory analysis of olive oil. Method for the organoleptic assessment of virgin olive oil*, 2015, pp. 1-20.

125. International Olive Oil Council, *Sensory analysis of olive oil standard guide for the selection, training and monitoring of skilled virgin olive oil tasters*. 2013, pp. 1-14.

126. N. Tena et al., An in-depth assessment of analytical methods for olive oil purity, safety and quality characterization, *Journal of Agricultural and Food Chemistry*, vol. 63, no. 18, 2015, pp. 4509-4526.

127. D.L. García-González and R. Aparicio, Research in olive oil: challenges for the near future, *Journal of Agricultural and Food Chemistry*, vol. 58, 2010, pp. 12569-12577.

128. M.E. Escuderos, Olive Oil Aroma Evaluation By Gas Chromatographic Method: A Critical Review, *Critical Reviews in Analytical Chemistry*, vol. 41, no. 1, 2011, pp. 70-80.

129. G. Montedoro, M. Bertuccioli, and F. Anichini, Aroma analysis of virgin olive oil by head space volatiles extraction techniques, *Flavor of foods and beverages*, G. Charalampous and G. Inglet, eds., Academic Press, New York, 1978, pp. 247-281.

130. A. Cifuentes, Food Analysis: Present, Future, and Foodomics, *ISRN Analytical Chemistry*, vol. 2012, 2012, pp. 1-16.

131. C. Ibanez et al., The role of direct high-resolution mass spectrometry in foodomics, *Analytical and Bioanalytical Chemistry*, vol. 407, no. 21, 2015, pp. 6275-6287.

132. F. Favati, N. Condelli, F. Galgano, and M.C. Caruso, Extra virgin olive oil bitterness evaluation by sensory and chemical analyses, *Food Chemistry*, vol. 139, no. 1-4, 2013, pp. 949-954.

133. P. Andrewes et al., Sensory properties of virgin olive oil polyphenols: identification of deacetoxy-ligstroside aglycon as a key contributor to pungency, *Journal of Agricultural and Food Chemistry*, vol. 51, no. 5, 2003, pp. 1415-1420.

134. V. García-Cañas et al., Present and Future Challenges in Food Analysis: Foodomics, *Analytical Chemistry*, vol. 84, 2012, pp. 10150-10159.

135. A.M. Gómez-Caravaca, R.M. Maggio, and L. Cerretani, Chemometric applications to assess quality and critical parameters of virgin and extra-virgin olive oil. A review, *Analytica Chimica Acta*, vol. 913, 2016, pp. 1-21.

136. Comission Implementing Regulation (EC) No 1348/2013, amending the Annex to Regulation (EEC) No 2568/91 on the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis, *Official Journal of the European Union*, L338, 2013, pp.34-67.

137. E. Stefanoudaki, M. Williams, and J. Harwood, Changes in virgin olive oil characteristics during different storage conditions, *European Journal of Lipid Science and Technology*, vol. 112, no. 8, 2010, pp. 906-914.

138. E.J. Want et al., Global metabolic profiling procedures for urine using UPLC-MS, *Nature Protocols*, vol. 5, no. 6, 2010, pp. 1005-1018.

139. M.A. Hashmi, M. Hanif, O. Farooq, and S. Perveen, Traditional Uses, Phytochemistry, and Pharmacology of Olea europaea (Olive), *Evidence Based Complementary and Alternative Medicine*, vol. 2015, 2015, pp. 541-591.

140. F. Hernández et al., Current use of high-resolution mass spectrometry in the environmental sciences., *Analytical and Bioanalytical Chemistry*, vol. 403, no. 5, 2012, pp. 1251-1264.

141. E.L. Schymanski et al., Strategies to characterize polar organic contamination in wastewater: exploring the capability of high resolution mass spectrometry, *Environmental Science and Technology*, vol. 48, no. 3, 2014, pp. 1811-1818.

142. E.L. Schymanski et al., Identifying small molecules via high resolution mass spectrometry: communicating confidence, *Environmental Science and Technology*, vol. 48, no. 4, 2014, pp. 2097-2098.

143. M.B. Chambers et al., A cross-platform toolkit for mass spectrometry and proteomics, *Nature Biotechnology*, vol. 30, no. 10, 2012, pp. 918-920.

144. C.A Smith, J.A. Want, G. O'Maille, R. Abagyan, and G. Siuzdak, Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification, *Analytical Chemistry,* vol. 78, no. 3, 2006, pp. 779-787.

145. C. Kuhl et al., CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets, *Analytical Chemistry*, vol. 84, no. 1, 2012, pp. 283-289.

146. G. Libiseller et al., IPO: a tool for automated optimization of XCMS parameters, *BMC Bioinformatics*, vol. 16, no. 118, 2015, doi: 10.1186/s12859-015-0562-8

147. T. Mehmood, K. H. Liland, L. Snipen, and S. Sæbø, A review of variable selection methods in Partial Least Squares Regression, *Chemometrics and Intelligent Laboratory Systems*, vol. 118, 2012, pp. 62-69.

148. D. Ballabio, V. Consonni, and R. Todeschini, The Kohonen and CP-ANN toolbox: A collection of MATLAB modules for Self Organizing Maps and Counterpropagation Artificial Neural Networks., *Chemometrics and Intelligent Laboratory Systems*, vol. 98, no. 2, 2009, pp. 115-122.

149. I.G. Chong and C.H. Jun, Performance of some variable selection methods when multicollinearity is present, *Chemometrics and Intelligent Laboratory Systems*, vol. 78, no. 1-2, 2005, pp. 103-112.

150. A. Vallverdú-Queralt and R.M. Lamuela-Raventós, Foodomics: A new tool to differentiate between organic and conventional foods, *Electrophoresis*, vol. 37, 2016, pp. 1784-1794.

151. L.V.T. Shepherd et al., Effect of agricultural production systems on the potato metabolome, *Metabolomics*, vol. 10, no. 2, 2014, pp. 212-224.

152. E. Koh, S. Charoenprasert, and A.E. Mitchell, Effect of Organic and Conventional Cropping Systems on Ascorbic Acid, Vitamin C, Flavonoids, Nitrate, and Oxalate in 27 Varieties of Spinach (Spinacia oleracea L.),

Journal of Agricultural and Food Chemistry, vol. 60, no. 12, 2012, pp. 3144-3150.

153. H. Ren, H. Endo, and T. Hayashi, Antioxidative and antimutagenic activities and polyphenol content of pesticide-free and organically cultivated green vegetables using water-soluble chitosan as a soil modifier and leaf surface spray, *Journal of the Science of Food and Agriculture*, vol. 81, no. 15, 2001, pp. 1426-1432.

154. A. Vallverdú-Queralt, A. Medina-Remón, I. Casals-Ribes, and R.M. Lamuela-Raventos, Is there any difference between the phenolic content of organic and conventional tomato juices?, *Food Chemistry*, vol. 130, no. 1, 2012, pp. 222-227.

155. S.L. Ferreira et al., Box-Behnken design: an alternative for the optimization of analytical methods, *Analytica Chimica Acta*, vol. 597, no. 2, 2007, pp. 179-186.

156. N.C. Maragou, N.S. Thomaidis, and M.A. Koupparis, Optimization and Comparison of ESI and APCI LC-MS/MS Methods: A Case Study of Irgarol 1051, Diuron, and their Degradation Products in Environmental Samples, *Journal of The American Society for Mass Spectrometry*, vol. 22, no. 10, 2011, pp. 1826-1838.

157. D. Baş and I.H. Boyacı, Modeling and optimization I: Usability of response surface methodology, *Journal of Food Engineering*, vol. 78, no. 3, 2007, pp. 836-845.

158. M.A. Bezerra et al., Response surface methodology (RSM) as a tool for optimization in analytical chemistry, *Talanta*, vol. 76, no. 5, 2008, pp. 965-977.

159. J. Yang et al., A chemical profiling strategy for semi-quantitative analysis of flavonoids in Ginkgo extracts, *Journal of Pharmaceutical and Biomedical Analysis*, vol. 123, 2016, pp. 147-154.

160. J.D. Holliday, N. Salim, and P. Willett, On the Magnitudes of Coefficient Values in the Calculation of Chemical Similarity and Dissimilarity, *Chemometrics and Chemoinformatics*, vol. 894, 2005, pp. 77-95.

161. G. Cincilla, M. Thormann, and M. Pons, Structuring Chemical Space: Similarity-Based Characterization of the PubChem Database, *Molecular Informatics*, vol. 29, no. 1-2, 2010, pp. 37-49.

162. G. Maggiora, M. Vogt, D. Stumpfe, and J. Bajorath, Molecular similarity in medicinal chemistry, *Journal of Medicinal Chemistry*, vol. 57, no. 8, 2014, pp. 3186-3204.

163. H.P. Singer, A.E. Wossner, C.S. McArdell, and K. Fenner, Rapid Screening for Exposure to "Non-Target" Pharmaceuticals from Wastewater Effluents by Combining HRMS-Based Suspect Screening and Exposure Modeling, *Environmental Science and Technology*, vol. 50, no. 13, 2016, pp. 6698-6707.

164. N.P. Kalogiouri, N.A. Alygizakis, R. Aalizadeh, and N.S. Thomaidis, Olive oil authenticity studies by target and nontarget LC-QTOF-MS combined with advanced chemometric techniques, *Analytical and Bioanalytical Chemistry*, vol. 408, no. 28, 2016, pp. 7955-7970.

165. Stat-Ease, *Design-Expert*, 7.0.0. version, June 2015; http://www.statease.com/dx10.html

166. P. Liu, D.K. Agrafiotis, and D.L. Theobald, Fast determination of the optimal rotational matrix for macromolecular superpositions, *Journal of Computational Chemistry*, vol. 31, no. 7, 2010, pp. 1561-1563.

167. R. Todeschini, M. Lasagni, and E. Marengo, New molecular descriptors for 2D and 3D structures. Theory, *Journal of Chemometrics*, vol. 8, no. 4, 1994, pp. 263-272.

168. R. Todeschini et al., Modeling and prediction of molecular properties. Theory of grid-weighted holistic invariant molecular (G-WHIM) descriptors,

Chemometrics and Intelligent Laboratory Systems, vol. 36, no. 1, 1997, pp. 65-73.

169. R. Todeschini, R. Cazar, and E. Collina, The chemical meaning of topological indices, *Chemometrics and Intelligent Laboratory Systems*, vol. 15, no. 1, 1992, pp. 51-59.

170. R. Todeschini, P. Gramatica, R. Provenzani, and E. Marengo, Weighted holistic invariant molecular descriptors. Part 2. Theory development and applications on modeling physicochemical properties of polyaromatic hydrocarbons, *Chemometrics and Intelligent Laboratory Systems*, vol. 27, no. 2, 1995, pp. 221-229.

171. R. Todeschini and V. Consonni, Handbook of molecular descriptors, Weinheim, *Methods and Principals in Medicinal Chemistr*y, vol. 11, R. Mannhold, H. Kubinyi, H. Timmerman, eds., Wiley-VCH, Weinheim, Germany, 2000, doi:10.1002/9783527613106.ch1d.

172. I.V. Tetko et al., Virtual Computational Chemistry Laboratory – Design and Description, *Journal of Computer-Aided Molecular Design*, vol. 19, no. 6, 2005, pp. 453-463.

173. Virtual Computational Chemistry Laboratory, 10 Sept. 2016; http://www.vcclab.org.

174. Partitioning (logD) plugin was used for the calculation of logD, ChemAxon; 17 Oct. 2016; http://www.chemaxon.com.

175. K.S. Kim, H.H. Choi, C.S. Moon, and C.W. Mun, Comparison of k-nearest neighbor, quadratic discriminant and linear discriminant analysis in classification of electromyogram signals based on the wrist-motion directions, *Current Applied Physics*, vol. 11, no. 3, 2011, pp. 740-745.

176. G. McLachlan, ed., *Discriminant Analysis and Statistical Pattern Recognition*, New York, Wiley, 2004.

177. T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* second ed., United States, Springer, 2009.

178. R. Genuer, J.M. Poggi, and C. Tuleau-Malot, Variable selection using random forests, *Pattern Recognition Letters*, vol. 31, no. 14, 2010, pp. 2225-2236.

179. S. Martín-Peláez, M.I. Covas, M. Fitó, A. Kušar, and I. Pravst, Health effects of olive oil polyphenols: Recent advances and possibilities for the use of health claims, *Molecular Nutrition and Food Research*, vol. 57, 2013, pp. 760-771.

180. R. Ghanbari et al., Valuable Nutrients and Functional Bioactives in Different Parts of Olive (Olea europaea L.) - A Review, *International Journal of Molecular Sciences*, vol. 13, no. 3, 2012, pp. 3291.

181. M. Casale, C. Cazolino, P. Olivieri, and M. Forina, The potential of coupling information using three analytical techniques for identifying the geographical origin of Liguria extra virgin olive oil, *Food Chemistry*, vol. 118, 2010, pp. 163.170.

182. Council Regulation (EC) No 510/2006, on the protection of geographical indications and designations of origin for agricultural products and foodstuffs, *Official Journal of the European Union*, L93, 2006, pp.12-25.

183. A. Bajoub et al., Potential of LC-MS phenolic profiling combined with multivariate analysis as an approach for the determination of the geographical origin of north Moroccan virgin olive oils, *Food Chemistry*, vol. 166, 2015, pp. 292-300.

184. I. Kosma et al., Differentiation of Greek extra virgin olive oils according to cultivar based on volatile compound analysis and fatty acid composition, *European Journal of Lipid Science and Technology*, vol. 118, no. 6, 2016, pp. 849-861.

185. I. Karabagias et al., Classification of Western Greek virgin olive oils according to geographical origin based on chromatographic, spectroscopic, conventional and chemometric analyses, *Food Research International*, vol. 54, no. 2, 2013, pp. 1950-1958.

186. E. Pouliarekou et al., Characterization and classification of Western Greek olive oils according to cultivar and geographical origin based on volatile compounds, *Journal of Chromatography A*, vol. 1218, no. 42, 2011, pp. 7534-7542.

187. F. Longobardi et al., Characterisation of the geographical origin of Western Greek virgin olive oils based on instrumental and multivariate statistical analysis, *Food Chemistry*, vol. 133, no. 1, 2012, pp. 169-175.

188. P.V. Petrakis et al., Geographical Characterization of Greek Virgin Olive Oils (Cv. Koroneiki) Using 1H and 31P NMR Fingerprinting with Canonical Discriminant Analysis and Classification Binary Trees, *Journal of Agricultural and Food Chemistry*, vol. 56, no. 9, 2008, pp. 3200-3207.

189. A. Allalout et al., Characterization of virgin olive oil from Super Intensive Spanish and Greek varieties grown in northern Tunisia, *Scientia Horticulturae*, vol. 120, no. 1, 2009, pp. 77-83.

190. D. Ocakoglu, F. Tokatli, B. Ozen, and F. Korel, Distribution of simple phenols, phenolic acids and flavonoids in Turkish monovarietal extra virgin olive oils for two harvest years, *Food Chemistry*, vol. 113, no. 2, 2009, pp. 401-410.

191. A. Bajoub et al., Assessing the varietal origin of extra-virgin olive oil using liquid chromatography fingerprints of phenolic compound, data fusion and chemometrics, *Food Chemistry*, vol. 215, 2017, pp. 245-255.

192. E. Tengstrand, J. Lindberg, and K.M. Åberg, TracMass 2—A Modular Suite of Tools for Processing Chromatography-Full Scan Mass Spectrometry Data, *Analytical Chemistry*, vol. 86, no. 7, 2014, pp. 3435-3442.

193. H. Gowda, et al., Interactive XCMS Online: Simplifying Advanced Metabolomic Data Processing and Subsequent Statistical Analyses, *Analytical Chemistry*, vol. 86, no. 14, 2014, pp. 6931-6939.

194. S. Bertrand, Y. Guitton, and C. Roullier, Success and pitfalls in automated dereplication strategy using liquid chromatography coupled to mass spectrometry data: A CASMI 2016 experience, *Phytochemistry Letters*, 2017, doi:/10.1016/j.phytol.2016.12.025.

195. R. Aalizadeh, P.C. von der Ohe, and N.S. Thomaidis, Prediction of acute toxicity of emerging contaminants on the water flea Daphnia magna by Ant Colony Optimization-Support Vector Machine QSTR models, *Environmental Sciences: Processes and Impacts*, vol. 19, no. 3, 2017, pp. 438-448.

196. N.M. O'Boyle et al., Open Babel: An open chemical toolbox, *Journal of Cheminformatics*, vol. 3, no. 33, 2011, doi:10.1186/1758-2946-3-33

197. M.J. Vainio and M.S. Johnson, Generating Conformer Ensembles Using a Multiobjective Genetic Algorithm, *Journal of Chemical Information and Modeling*, vol. 47, no. 6, 2007, pp. 2462-2474.

198. R.W. Kennard and L.A. Stone, Computer Aided Design of Experiments, *Technometrics*, vol. 11, no. 1, 1969, pp. 137-148.

199. L. Parkinson and R. Keast, Oleocanthal, a Phenolic Derived from Virgin Olive Oil: A Review of the Beneficial Effects on Inflammatory Disease, *International Journal of Molecular Sciences*, vol. 15, no. 7, 2014, pp. 12323-12334.

200. F. Casameti and M. Stefani, Olive polyphenols: new promising agents to combat aging-associated neurodegeneration, *Expert Review of Neurotherapeutics*, vol. 17, no. 4, pp. 345-358.