Università degli Studi di Torino
Doctoral School of Sciences and Innovative Technologies
Ph.D. Program in Pharmaceutical and Biomolecular Science
Cycle XXXIII

Italian olive oil sensory quality: Advanced analytical strategies for volatile organic compounds chemical fingerprinting

PhD Candidate
Federico Stilo

Supervisor
Chiara Emilia Irma Cordero

Università degli Studi di Torino

Ph.D. Program in Pharmaceutical and Biomolecular Sciences
Department of Drug Science and Technology
Cycle XXXIII

Thesis title - Italian olive oil sensory quality: Advanced analytical strategies for volatile organic compounds chemical fingerprinting

Ph.D. Candidate - Dott. Federico Stilo

Tutor - Prof. Chiara Emilia Irma Cordero

Ph.D. Coordinator - Prof. Gianmario Martra, Prof.ssa Roberta Cavalli

Academic years 2017/2018, 2018/2019, 2019/2020

Scientific field of application - Food Chemistry – SSD CHIM/10

Thesis reviewers - Prof. Peter Quinto Tranchida, Prof. Jean-Marie Dimandja

Jury members - Prof. Luigi Mondello, Prof. Jean-Marie Dimandja, Prof. Giorgia Purcaro

Abstract

The present Doctoral Thesis deals with the in-depth chemical characterization of the Italian olive oil volatile fraction contributing to the quality features of this unique product, with emphasis on compounds correlated to sensory perception. With this objective in mind, research efforts have been directed to the development of analytical strategies for accurate profiling and effective fingerprinting of volatile organic compounds (VOCs). Through a systematic design, sample preparation by headspace techniques has been exploited and its performances examined in light of the molecular resolution potential of multidimensional gas chromatography coupled to mass spectrometry and flame ionization detector. Moreover, the challenging step of multidimensional signals data processing, has been also investigated for its flexibility and reliability when multiple analytical batches or signals derived from different analytical platforms have to be explored. The outcomes of the PhD project research have the ambition to propose straightforward analytical solutions that achieve molecular resolution characterization of olive oil VOCs (*i.e.*, *profiling*) with all the advantages of *fingerprinting* effectiveness in samples *identidation* and discrimination. By these analytical tools, that include computer vision strategies and *artificial intelligence smelling machines*, food quality objectification makes a step ahead as advanced analytical chemistry enters in our daily life.

Extra virgin olive oils (EVOOs) analyzed during a timeframe of 3 years were supplied by Progetto Ager, with the grant ID "Valorization of Italian OLive products through INnovative analytical tools- VIOLIN". EVOOs were selected within top-quality products often awarded with extra-certifications as protected designation of origin – PDO, protected geographical indication – PGI, or as Organic. Samples represent a snapshot of the impressively complex heritage of olives genetic varieties typical of the Italian production (*i.e.,* about 540 different registered cultivars and 46 PDO EVOOs) and of the impact of the *terroir* due to the local pedoclimatic related to the different geographical locations (*i.e.,* Regions).

EVOOs is recognized as the most valuable product among the edible oil, and the reason of the increasing demand for high-quality EVOO is related not only to its nutritional and healthy values, due to the presence of antioxidants and high oleic acid content, but also to its peculiar sensory characteristics. For this reason, EVOO is listed among the top foods subjected to adulteration (*e.g.,* addition of external cheap products, mislabeling practices etc.), with quality control issues that are mainly related to a lack of sufficiently powerful analytical methods able to provide an objective classification of olive oils (OOs). In particular, OO is to date the only food product whose classification is officially regulated by an evaluation of the sensory properties based on a Panel Test assessment, and the scientific community is searching for analytical methods able to support it.

When the aim is to provide a full *identitation* of an OO, *i.e.*, to define the identity of a particular OO based on the characteristic features that make it singular or unique, multidimensional analytical (MDA) platforms are required. They are characterized by the combination of a physicochemical separation (*e.g.*, chromatography, electromigration, size exclusion, etc.) with spectroscopic techniques (*e.g.*, Mass Spectrometry - MS, Nuclear Magnetic Resonance - NMR, Infrared - IR, Ultraviolet - UV, etc.), having the potential to discriminate, identify and quantify sample constituents, providing a solid foundation for generating hypothesis driven studies.

In particular, the core of the research activity of this PhD project is represented by MDA platforms combining headspace sampling (HS), multidimensional separation, and different detection approaches often combined with "parallel detection" solutions. Each configured platform was optimized and/or adapted to match for sample's compositional complexity and dynamic range of concentrations to cover.

Headspace sampling was explored as the most suitable sample preparation approach to cover highly-to-semi volatiles within the EVOO volatilome. A series of high concentration capacity (HCC) techniques were tested including: (a) solid phase microextraction (SPME), (b) headspace sorptive extraction (HSSE) with either a single-material coating (polydimethylsiloxane—PDMS) or a dual-phase coating that combines PDMS/Carbopack and PDMS/EG (ethyleneglycol); (c) monolithic material sorptive extraction (MMSE), using octa-decyl silica combined with graphite carbon (ODS/CB); and dynamic headspace (d) with either PDMS foam, operating in partition mode, or Tenax TA™, operating in adsorption mode. Among them, SPME was found to be the most suitable for both accurate quantitative profiling and comprehensive fingerprinting due to the possibility of full automation and integration with the available analytical platforms. Sampling parameters by SPME (*i.e.,* amount of sample, sampling temperature and sampling time) were then optimized matching for headspace linearity conditions, to provide a meaningful picture of sample's components and a realistic quantitative reference for quantitative investigations.

The separation step was by comprehensive two-dimensional gas chromatography (GC×GC), selected for its separation power, sensitivity, and for the intrinsic potential of being also an effective fingerprinting tool. In view of a possible routine adoption of GC×GC screening, the translation of analytical methods from a platform equipped with a loop-type thermal modulator to a system equipped with a differential-flow modulator (DFM) was therefore investigated and operational parameters optimized.

The hyphenation of the GC×GC with a mass spectrometry (MS) detection adds another level of specificity while providing essential information for analyte identification, thanks to the spectra produced by electron ionization (EI). The potential of a Tandem Ionization™ system, able to provide variable energy electron ionization, was also explored.

The large amount of information provided by the HS-SPME-GC×GC-MS requires for dedicated data elaborations to exploit the informative levels encrypted by samples components distribution. In this context, advanced fingerprinting approaches, *i.e.,* combined *Untargeted* and *Targeted (UT) fingerprinting*, were exploited to allow a comprehensive cross-sample comparison by considering both the distribution of known and unknown volatiles. Moreover, a workflow for chromatographic fingerprinting by template matching based on pattern recognition algorithms was defined, optimized and adapted through data re-alignment tools to address 2D-peak patterns variations occurring in long-term studies, as to cover different harvest years of shelf-life modifications.

The optimized analytical workflow was then applied on a large selection of EVOOs, providing validation of ripening markers and quality indicators and solid foundation for samples *identitation*

and discrimination based on production regions. Finally, the *UT fingerprinting* process was combined with an accurate quantitative profiling by multiple headspace solid phase microextraction (MHS-SPME) to quantify a large selection of key-odorants and quality markers by external calibration and flame ionization detection (FID) predicted relative response factors (RRFs). By accurate amounts of odorants responsible of positive and negative attributes, odor activity values (OAVs) were defined while extending the concept of sensomics based expert system to a fully automated platform acting as an *artificial intelligence smelling machine* that delineates EVOOs aroma blueprint.

Outline

The Doctoral Thesis is organized in five chapters covering the main topics of the research activities conducted over the three years of the project.

- **Chapter 1 – The olive oil**. This chapter proposes an overview on olive (*Olea europaea L.*) oil production chain by introducing olive tree botanical features, followed by a brief history of the product with a focus on the Italian heritage. Then, the process of transformation of olives into edible oil is described, the legislation framework for olive oil classification is introduced highlighting the peculiar condition for which the olive oil is, to date, the only food product with commercial classification officially regulated by a sensory evaluation (*i.e.,* Panel Test). Olive oil composition is then described, while a specific section is dedicated to the volatile fraction, object of this thesis.
- **Chapter 2 – Comprehensive two-dimensional gas chromatography**. In this chapter, the brief introduction on the origin of multidimensional gas chromatography (MDGC) is followed by the conceptualization of the term "multidimensionality" and the description of the two general categories of MDGC, the heart-cut MDGC (H/C MDGC) and the GC×GC, which is the core analytical platform of this PhD project. Columns' configurations and related properties (e.g., orthogonality, selectivity and separation logic) are introduced. A focus on the modulation process and on the evolution of modulators technology complete the technological overview.
- **Chapter 3 – Principles of chemometrics.** This chapter provides an overview of the most common data processing and data mining approaches suitable to explore the great complexity of GC×GC patterns; they are introduced according to a rational workflow from the design of experiment (DoE) to the creation of classification and regression models. Principles of DoE are explained, together with the main designs – grouped in designs for screening and response surface designs – and the parameters considered to evaluate the model. Then, raw chromatographic data preprocessing steps, directed to the correct acquisition, interpretation, and alignment of GC×GC analyses, are described. Data processing includes pertinent algorithms adopted for the purpose of this thesis. Multivariate analysis completes the chapter by describing exploratory analysis algorithms (*i.e.,* principal component analysis – PCA, and cluster analysis – CA) and supervised techniques (*i.e.,* linear discriminant analysis – LDA, k-nearest neighbor – k-NN, classification and regression tree – CART, soft independent modeling of class analogy – SIMCA, principal component regression – PCR, partial least square regression – PLS, and partial least square discriminant analysis – PLS-DA).

- **Chapter 4 – Published reviews.** It is dedicated to reviews published within the PhD thesis period; they are:
    - Chromatographic fingerprinting by comprehensive two-dimensional chromatography: fundamentals and tools.
    - Untargeted approaches in food-omics: the potential of comprehensive two-dimensional gas chromatography/mass spectrometry
    - Comprehensive two-dimensional gas chromatography as a boosting technology in food-omics investigations.
- **Chapter 5 – Published research papers.** This last chapter includes the published research papers dealing with technical advancements related to the characterization of the EVOO; they are:
    - Highly informative fingerprinting of extra-virgin olive oil volatiles: The role of high concentration-capacity sampling in combination with comprehensive two-dimensional gas chromatography.
    - Chromatographic fingerprinting by template matching for data collected by comprehensive two-dimensional gas chromatography.
    - Untargeted and Targeted Fingerprinting of Extra Virgin Olive Oil Volatiles by Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry: Challenges in Long-Term Studies.
    - A step forward in the equivalence between thermal and differential-flow modulated comprehensive two-dimensional gas chromatography methods.
    - Exploring the extra-virgin olive oil volatilome by adding extra dimensions to comprehensive two-dimensional gas chromatography and time of flight mass spectrometry featuring tandem ionization: validation of ripening markers in headspace linearity conditions.
    - Chromatographic fingerprinting enables effective discrimination and *identitation* of high-quality Italian extra-virgin olive oils.
    - Delineating the extra-virgin olive oil aroma blueprint by multiple headspace solid phase microextraction and differential-flow modulated comprehensive two-dimensional gas chromatography.

Chapter 1

The olive oil

1.1 *Olea europaea* L.

1.1.1 Botanical introduction

The olive tree (*Olea europaea* L.) belongs to the botanical family of Oleaceae,[1] and two subspecies can be found: *Olea europaea* L. ssp. *sativa*, or *olive*, namely the cultivated olive tree, and *Olea europaea* L. ssp. *sylvestris*, or *oleaster*, namely the wild olive tree.[2]

Olea europaea L. is native of the Mediterranean basin and Middle East, cultivated after a long process of domestication and selection. It is a typical Mediterranean plant, that can face low temperatures in winter, but needs of warm springs and hot summers. Olive tree demand for water is not very high, although non-stratified and moderately fine textured soils, providing aeration for root growth and having a high water holding capacity are prefeable.[3]

Olea europaea L. var *sativa* is an evergreen slow-growing small tree (maximum height of 8-10 meters) characterized by a greyish cylindrical trunk until the tenth year of age, after which it becomes knotty, contorted and with a darker colour.[4]

Leaves grow from spring to autumn and are shed every two years; they are opposite, lanceolate and persistent, 6-7 cm long and 1-2 cm broad, dark green in the upper side and silver in the lower side, depending on the presence of trichomes.[2,5] Flowers, growing on axillary branches, are white, with a persistent quadridentate calyx and a gamopetalous tetramer corolla (4-5 mm).[1,3]

The fruit is produced through a phenomenon known as *induction*: undifferentiated tissue is transformed into vegetative and/or reproductive buds by endogenous metabolic factors and, in general, a year of low production is followed by a year of high vegetative activity and *viceversa*.[4,6] The olive is an oval or spheroid drupe 1-3.5 cm long and 1-7 g weight, composed by different parts (see **Figure 1.1**):

- Epicarp or skin (1.5-3%), with a color that changes from green to purple-black during the ripening and covered by a protective waxy layer.
- Mesocarp or pulp (75-85%), characterized by a high oil content.
- Endocarp or kernel (13-23%), a woody shell containing the seed (2-4%), itself constituted by an external membrane named episperm enveloping the endosperm, finally including the embryo, containing a small amount (1-1.5%) of oil hardly extractable.[6]

The olive chemical composition is influenced by several factors, including cultivar, ripening stage, pedoclimatic conditions etc. However, a ripe olive is on average composed by a 40-50 % of water, 15-35% of oil and 10-20 % of carbohydrates, with a remaining part which is mostly raw fiber and ash.[2,6,7]

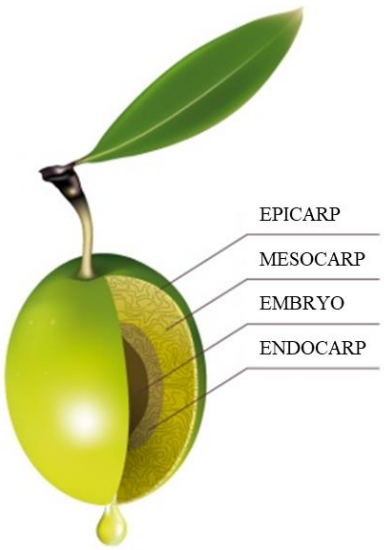

Figure 1.1. Olive structure. Adapted from[7]

1.1.2 History

Olive tree is one of the oldest plant in the world and it seems to have coexisted with humans since the early bronze age, 6000-7000 years ago,[4,8] having its origin in the region corresponding to Persia and Mesopotamia.[9] The olive plant later spread from this area to nearby territories on the eastern Mediterranean Coast, corresponding to present-day Syria, Palestine and Lebanon.[8,9] After the domestication in this area it spread from east to west thanks to Phoenicians, nation of traders and navigators, to Egypt, Crete and Greek (**Figure 1.2**), where its cultivation was favored by extremely suitable climate and soils.[9,10]

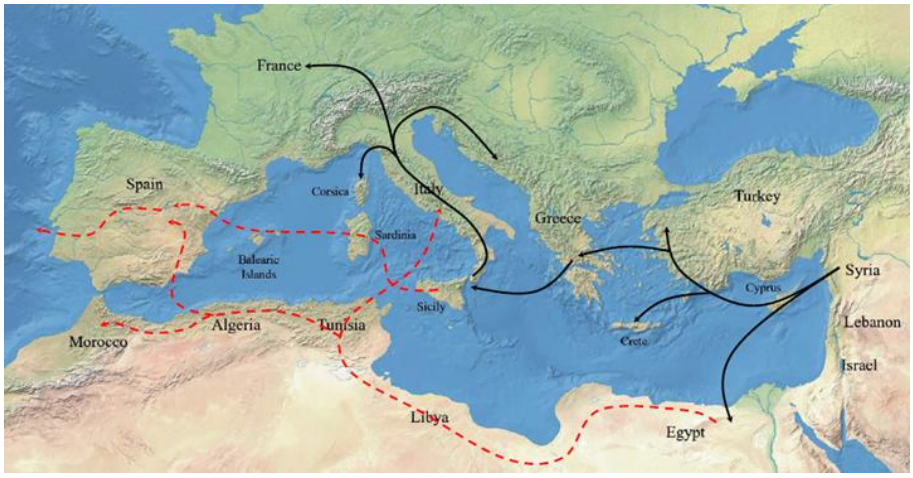

Figure 1.2. Diffusion of the olive tree cultivation in the Mediterranean area, distinguished in first (black lines) and second wave (red dotted line). [4]

In Italy, the olive tree was known since the period of the Etruscan civilization, however the regular cultivation was adopted after the Greek's colonization of the area called Magna Grecia, corresponding to the present Sicilia, Calabria, Campania, Puglia and Basilicata regions.[9] Romans learned olive cultivation through their contacts with Greeks colonies in Italy, and contributed themselves to propagate olive diffusion in all the Roman Empire, expanding their dominion into North Africa and Iberian Peninsula.[9,11]

In summary, all ancient civilizations grown around the Mediterranean basin and in the Middle east evolved, in different moment, the olive cultivation, leaving clear evidence of olive cultivation and oil production activities (**Figure 1.3**). The olive was a key element of the Mediterranean culture however, interestingly, the edible use was not the most common and/or the main one.[4,9,12] Olive oil was primarily a lamp fuel[8,13], in the ancient Egypt was used in religious ceremonies and as an ointment[9], as in Greece, where it was used as pharmaceutical ointment to cure disease, to make skin and hair appear healthier and during athletics ceremonies.[8]

Figure 1.3. Archaeological find, in the Knossos Palace, Crete: vases (pithoi) for olive oil storage.[12]

The key role of the olive in the ancient cultures is finally testified by myths and religions. In Greek mythology, in the conflict for the possession of the Attica region, Poseidon and Athena offers to people different gifts: Poseidon used his trident to create a seawater spring on land, claiming Athenians would rule the waves (another version has Poseidon offering the first horse, symbolizing war and power), while Athena created the first olive tree, symbol of food, lighting, medicine and cosmetics. People preferred Athena's gift, giving to the capital of Attica the name of Athens, in her honor.[4,13] In the Old Testament instead, in the narration of the Flood, Noah released a dove that returned with an olive branch, symbol of the new peace between God and humans.[13]

1.1.3 Italian cultivars

A *cultivar* is a group of similar of plants selected for one or more desirable characteristics that are maintained during propagation, and it is worthy of mention the origin of this term, which probably derives from a combination of the words "cultivated" and "variety".[14]

The Italian olive genetic assets is the widest in the world, with 538 registered cultivars against the 262 registered for Spain, which is the first world producer of olive oil.[14,15] However, a limited number of cultivars is intensively used: about 80 cultivars account for the 90% of the total olive-growing surface area in Italy and, in particular, 24 cultivars account for about the 60% of the total olive growing surface area.[14] They are: Coratina, Ogliarola salentina, Cellina di Nardò, Carolea, Frantoio, Leccino, Ogliarola barese, Moraiolo, Bosana, Cima di Mola, Dolce di Rossano, Ogliarola messinese, Ottobratica, Sinopolese, Nocellara del Belice, Canino, Carboncella, Itrana, Moresca, Rotondella, Taggiasca, Tondina, Grossa di Gerace, Nocellara Etnea.

Not all cultivars are widespread throughout Italy, but each region has its own typical ones, that in many cases have been genetically characterized. **Figure 1.4** shows the regional map of the principal and most used Italian Olive cultivars, each one characterized by typical flavor perceptions, from the marked strong *bitter* and *spicy* notes distinctive of Coratina (Puglia) to the well balanced taste, with *almond* and *fresh grass* notes characterizing Frantoio (Toscana).[16,17] **Figure 1.4** reports the region of origin and traditional cultivation for major cultivars from Italy. Some of them, as Leccino and Frantoio, are widespread in all the Italian soil, and in the latest decades they were transplanted and cultivated also in Piemonte and Val d'Aosta, thanks to their adaptability and resistance to lower temperatures.[18]

1 – Val d'Aosta
Frantoio, Leccino

2 – Piemonte
Frantoio, Moraiolo, Leccino

3 – Liguria
Taggiasca, Razzola, Pignola, Rossese, Colombaia

4 – Lombardia
Casaliva, Grignano, Favarol, Gargnà

5 – Trentino Alto Adige
Casaliva, Raza, Favarol

6 – Friuli Venezia Giulia
Bianchera

7 – Veneto
Casaliva, Grignan, Favarol, Perlarol, Trepp

8 – Emilia Romagna
Correggiolo, Ghiacciolo, Nostrana di Brisighella

9 – Toscana
Frantoio, Moraiolo, Leccino, Maurino, Pendolino, Olivastra Saggianese

10 – Marche
Ascolana, Carboncella, Canino, Mignola, Rosciola

11 – Umbria
Borgiona, Bianchella di Umbertide, Correggiolo, Dolce Agagia, Raia

12 – Lazio
Caninese, Itrana, Carboncella, Raia

13 – Abruzzo
Dritta, Gentile di Chieti, Tortiglione, Castiglionese, Intosso, Cucco, Nebbio

14 – Molise
Aurina di Venafro, Cellina di Rotello, Gentile di Larino, Oliva di Colletorto

15 – Campania
Asprinia, Biancolella, Carpellese, Minucciola, Ortice, Ortolana, Ravece, Pisciottana, Rotondella, Pisciottana

16 – Puglia
Coratina, Cima di Bitonto, Cellina di Nardò, Ogliarola Barese/Salentina, Peranzana Provenzale, Pizzuta di Massafra

17 – Basilicata
Ogliarola, Coratina, Maiatica

18 – Calabria
Carolea, Sinopolese, Ottobratica, Cassanese, Grossa di Gerace

20 – Sardegna
Bosana, Semidana, Nera di Villacidro, Tonda di Cagliari, Pizz E Carroga

19 – Sicilia
Cerasuola, Biancolilla, Nocellara del Belice, Nocellara Etnea, Tonda Iblea, Ogliarola Messinese

Figure 1.4. Regional map of the principal Italian Olive cultivars.

1.2 From olives to olive oil

1.2.1 Ripening stages and maturity index

The decision of the proper "when" for the harvesting of olives influences organoleptic characteristics, color, shelf life and yield of the final olive oil produced. The harvest time depends on a series of factors (*i.e.,* cultivar, pedoclimatic conditions as temperatures and sunlight, irrigation) influencing the ripening process, which results in a change of the olives' chemical composition: decrease of the total polyphenols content, increase of the unsaturated fatty acids fraction, especially oleic acid, and changes in tocopherols and pigments levels are the most relevant.[6,19–21]

Ripening stages of olives are essentially three: (a) green immature and quite firm fruit, with a high content in polyphenols and chlorophyll resulting in a quite green oil, characterized by marked *bitter* and *pungent* notes; (b) veraison fruit, having a skin that turns in red-purple color and with a still high polyphenol content, resulting in some *bitterness* and *pungency*, although balanced by the developing of some ripe-fruity characteristics; (c) black mature fruit (even if some varieties never turn totally black), characterized by a low content of polyphenols and chlorophyll and an increased content of carotenoid, with the resulting oil more golden in color and described as *sweeter*, less *bitter* and less *pungent*.[20] The oil yield increases from green to black olives because of an easier destruction of vacuoles containing oils, while the shelf life is shorter for oils produced from black mature fruits.

Although high-quality oils are usually produced from olives picked at least at the veraison stage, harvest timing depends also on the desired oil aroma and shelf life characteristic and should take place when most of the olives are at the desired stage. To help producers in their evaluation and in comparing oils from year to year, a maturity index (MI) was developed in 1975 by Uceda and Frias[22] and it is still considered the official method by the International Olive Council (IOC). The maturity index depends on the color of skin and flesh of one-hundred olives randomly picked out from a representative sample of the olives in a crop and each olive is assigned to a score from 0 to 7 (**Figure 1.5**): 0) skin color deep green, 1) skin color yellow-green, 2) skin color with less than half the fruit surface turning red, purple or black, 3) skin color with more than half the fruit surface turning red, purple or black, 4) skin color all purple or black with all white or green flesh, 5) skin color all purple or black with less than half the flesh turning purple, 6) skin color all purple or black with more than half the flesh turning purple, 7) skin color all purple or black with all the flesh purple to the pit.

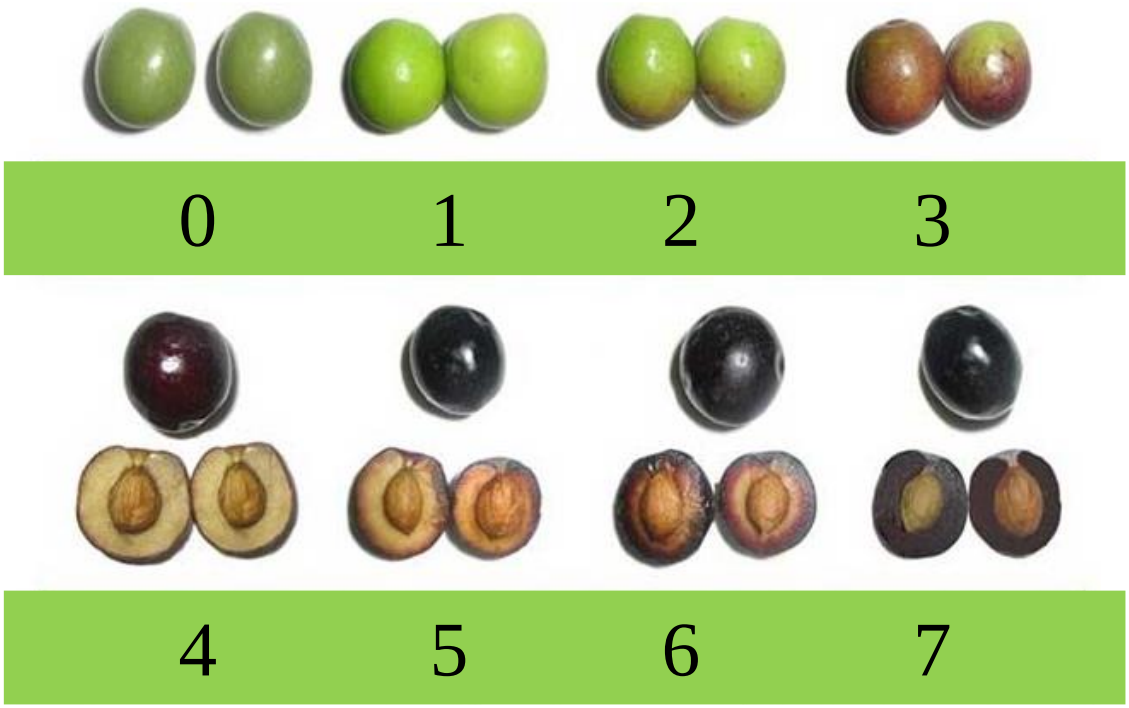

Figure 1.5. The eight score classes used to calculate the maturity index of olives. Modified from [21].

At this point, by following **Equation 1.1**, the number of the olives in each group is counted and the olive MI is calculated as the sum of the number of olives in each group (n) multiplied by the relative score, and finally divided by 100.

Equation 1.1 $$MI = \frac{(n\times0)+(n\times1)+(n\times2)+(n\times3)+(n\times4)+(n\times5)+(n\times6)+(n\times7)}{100}$$

1.2.2 Harvest

Olives are generally harvested between the end of October and the beginning of March, depending of the climate, the cultivar and the harvest method.[6]

Harvesting method are referable to three main groups:[5,6,23]

- Hand-picking: it is the best method, with olives directly collected into a basket, however it is also the less efficient and consequently the most expensive.
- Stripping: it is faster and less expensive, developed through (mechanical) devices, however it is stressful for the trees and olives could be damaged. Most common devices, with an increasing degree of automation, are: (a) plastic rakes and long wood sticks, (b) electric vibrators or shaker sticks, (c) machinery which shakes the entire tree.
- Picking after the fall, from nets placed under the tree crown or directly from the ground: in this case olives are contaminated with vegetal (*i.e.*, leaves and small branches) and mineral (*i.e.*, soil, dust, small stones) impurities.

The three method are listed in decreasing order of costs and time, but also of resulting quality: stripping and fall could damage olives, and usually fallen olives have a degree of maturation higher that the desired, with a consequent increase of the free acidity and a reduced shelf-life.[6]

1.2.3 Extraction

The olive oil extraction is a process composed by different technological steps.[24] The entire procedure can be in continuous or not, and the phase separation can be conducted by using various technological approaches; however, before the final separative step, cleaning, crushing and malaxation are required to turn the entire olives into a homogeneous paste.

1.2.3.1 Cleaning

The first step in the oil extraction process is the removal of extraneous matter that can affect the final quality of the olive oil.[23]

Before washing, leaves are removed: they are responsible of an increase in green color of the oil, depending on a higher quantity of chlorophyll pigment, and of a stronger perception of *green*, *leaf-like* and *bitter* notes (depending on an increased level of *(E)*-2-hexenal*)*, that may not be agreeable for consumers.[13,23]

Figure 1.6. Leaves removal and olives washing.[25]

After leaves removal, olive washing is performed to remove dust, sand, soil, small stones, and any metallic contaminant (**Figure 1.6**).[25] They are removed both for hygienic reasons and because they are abrasive and could damage the metallic parts of a hammer mill.[6,23] Moreover, to produce a top-quality extra virgin olive oil (EVOO), it is appropriate to perform a selection of the olives, by excluding wrinkled and rotten ones; usually washing machines are provided with grills intended for the cut-off selection.[6,13,23]

1.2.3.2 Crushing

The second step consists in olives crushing, which cause the oil droplets release by breaking up the cellular structure (specifically, the vacuoles) and smashing the endocarp. The result of this process is an olive paste, composed by oil, water and solids.[6,13]

The traditional method consists in olive crushing conducted by 2-6 granite mill-stones for 20-40 minutes (**Figure 1.7A**), while modern machines are high rotation centrifugation system equipped with metallic crushers, as disks (**Figure 1.7B**) or hammers (**Figure 1.7C**), usually integrated into a continuous process.[6,23]

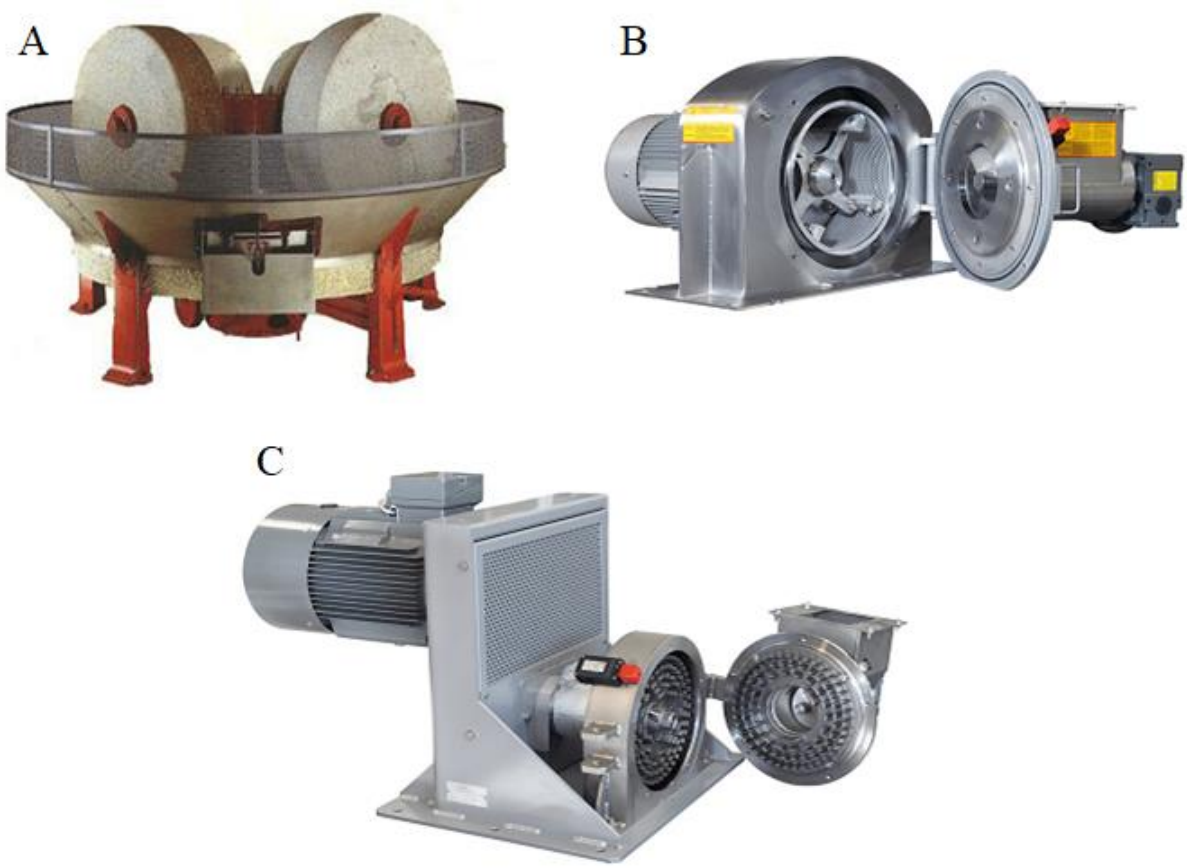

Figure 1.7. (**1.7A**) Traditional stones mill, (**1.7B**) Disks crusher, (**1.7C**) Hammer crusher.[26]

The milling method affects the quality and the amount of the olive oil produced. In particular metallic crushers result in higher extraction yields, reduced process' time (depending also on the fact that traditional mills are non in continuous) and a higher content of polyphenols resulting in a higher intensity of *bitter* and *pungent* notes.[23,24]

Moreover, it is important to monitor the olive paste heating depending on the transformation of kinetic energy into thermal, to avoid olive paste degradation. The use of metallic crushers, characterized by a high speed, results in an increase of 13-15°C of the olive paste, while by using low speed stone mill the temperature increase is lower (*i.e.*, 4-5°C).[23,24,27]

1.2.3.3 Malaxation

The last step before the phase separation it is a process named malaxation. It consists of a continuous slow kneading of the freshly crushed olive paste to break the oil-water emulsion and allow the concentration of the microscopic oil droplets into larger ones, to increase the extraction yield.[6,23,24,28] In particular, after the malaxation, about 80-85% of oil droplets in olive paste had a diameter greater than 30 microns.[23]

Malaxation is carried out in vats with different shape and size, however they are usually semicylindrical, with a horizontal shaft and spiral mixing blades realized in stainless steel (**Figure 1.8**).[13,23,24,29]

Figure 1.8. Detail of the malaxation process.[29]

The olive paste is stirred at approximately 20-30 rpm and the container is equipped with a heating jacket where hot water is circulated to heat olive paste. In recent years, many efforts were spent to increase the overall heat transfer coefficient improving the ratio between the surface area of the malaxing vat and the volume of the olive paste.[23,24,30] A temperature increase is needed to lower the olive paste viscosity and enhance the olive oil yield, however temperature also affects the resulting olive oil sensory profile therefore the time-temperature ratio has to be carefully controlled.[24,27] Temperature lower than 30-32 °C are recommended to avoid an increase in acidity and, in case of 30-35°C temperature, the malaxation time has to be reduced to 10-15 minutes to avoid the development of sensorial defects. However, to obtain an increased oil yield, more time is required, and a low temperature of 25°C allows olive paste to be malaxed for 30-60 minutes, with a concurrent decrease in polyphenols content and *bitter* perception.[13,23,24,28]

The malaxation step is fundamental especially for olive pastes obtained from crushers because, when tradition granite millstones are used for crushing, from one side the low rotation speed reduces the emulsification of oil with water, from the other the slow movement of olive paste during the process is equivalent to a partial malaxation. For this reason, when the olive paste is obtained from a tradition mill operating with a pressure system, the malaxation step is usually reduced to 10-20 minutes with a temperature of 20-25 °C.[23]

1.2.3.4 Phase separation

The separation of oil from the other constituents of olive paste (solid and water liquid phases) can be obtained by using only mechanical methods.[23] It particular, the extraction processes are mainly three: pressure, centrifugation and selective filtration (or percolation).[13,24]

- Pressure process. It is the oldest and most widespread system, at least in Italy, to obtain virgin olive oil (VOO); traditional mills have been unchanged during millennia, until the XX century when they were upgraded with the invention of the hydraulic press.[23] In this process, the olive paste is placed in 2-3 cm thick layers on oil diaphragms, which are then placed in moving units with central shaft.[13] A metal tray is placed after every 3-4 diaphragms to obtain uniform pressure application and a stable loading. Then pressure is applied and the liquid phase (a mixture of oil and water) run through the "olive cake", while the residual solid fraction stacking on diaphragms is named pomace.[6,13] The liquid phase is then centrifugated to separate oil from vegetation water, avoiding oxidation phenomena. VOO extracted by pressure is known to be especially rich in polyphenols, it is cheap, simple and it requires a low energy consumption; however it is a discontinuous process and sometimes it is affected by diaphragms contamination.[6]

- Centrifugation. It is a worldwide spread continuous system developed in the 1980s and based on the difference density characterizing the olive paste constituents.[13,23] The separation is performed through an horizontal 3-phases centrifuge, named decanter, where, to facilitate the separation, warm water (30°C) can be added to the olive paste (40-60 l/100 kg of olive paste).[6,23] A rotational speed of 6000-7000 rpm is needed to separate solid, water and oil; each decanter has a maximum hourly capacity that usually ranges from 1.5 to 7 tons per hour. 3-Phases decanters are fast, easy cleaning, automatized and continuous, also allowing a high separation yield; conversely, the energy demand is higher, and the addition of warm water to olive paste causes a lower content of natural antioxidants in the oil and an increase in the volume of vegetable water. To solve part of the drawbacks of 3-phases decanters, in the early 1990s a 2-phase decanter was launched on the market. [23] It works without adding water or only adding a minimal amount of water in the olive paste has a moisture amount < 50%, to facilitate separation. Main advantages of the 2-phase decanters are a lower water and energy consumption, reduced volume of vegetable water and lower operating costs, moreover they give and olive oil with a higher polyphenols content. [23,24] The only drawback is represented by the higher moisture percentage in the pomace. [23]
- Selective filtration or percolation. This is another method still used and it is based on the different surface tensions of the liquid phases in the paste: a steel plates is plunged into olive paste and it is coated with oil that then drips off, creating a low of oil. [13,24] The process is repeated many times to recover most of the oil from the paste, however the extraction yield reaches usually a maximum of 80%. [23,24] For this reason, this method is usually followed by a centrifugation, to extract the residual part of oil contained in the olive paste. [13,24]

Finally, an optional step suggested by IOC after the separation is represented by filtration, useful to make the oil clearer excluding all the suspended particles. The same result can be obtained spontaneously leaving to stand the VOO for a while, preferably in a conical bottom vessel.[6]

1.2.4 Bottling and labelling

Olive oil exposition to oxygen, light and heat is the main reason of oxidative degradation, so a proper bottling is needed and usually an extra virgin olive oil, depending also on its quality, should be consumed within two years. Marketing standards for olive oil are ruled by EU Regulation n° 29/2012 of 13 January 2012, which indicate in 5 liters the maximum capacity of a packaging presented to the final consumer, while, in the case of oils intended for consumption in restaurants or collective establishments, the maximum capacity of 5 liters can be exceeded. [6,31]

Moreover, a series of information on olive oil labelling are provided; the mandatory indications, in clear and indelible lettering, are:

- Sales denomination that, depending on the olive oil classification, is (a) "*superior category olive oil obtained directly from olives and solely by mechanical means*" for extra virgin olive oil; (b) "*olive oil obtained directly from olives and solely by mechanical means*", for virgin olive oil; (c) "*oil comprising exclusively olive oils that have undergone refining and oils obtained directly from olives*" for olive oil composed of refined olive oils and virgin olive oils; (d) "*oil comprising exclusively oils obtained by treating the product obtained after the extraction of olive oil and oils obtained directly from olives*" or "*oil comprising exclusively oils obtained by processing olive pomace and oils obtained directly from olive*s", for olive-pomace oil.

- Designations of origin for extra virgin and virgin olive oil. It must refer to the geographical area in which the olive oil was obtained, which is usually also the area in which the oil was extracted from the olives; if not, the information regarding the harvesting place should be stated on the packaging. Depending on the case, the information regarding the designation of origin. shall only consist of:
  (a) a reference to the Member State, to the Union or to third country, as appropriate, in the case of olive oil originating from one Member State or third country.
  (b) the indication "*blend of olive oils of European Union origin*" or a reference to the Union; "*blend of olive oils not of European Union origin*" or a reference to origin outside the Union; "*blend of olive oils of European Union origin and not of European Union origin*" or a reference to origin within the Union and outside the Union; in the case of blend of olive oils originating from more than one Member State or third country.
  (c) the indication *"(extra) virgin olive oil obtained in (the Union or the name of the Member State concerned) from olives harvested in (the Union or the name of the Member State or third country concerned)",* in the case of olives harvested in a Member State or third country other than that in which the mill where the oil was extracted from the olives is situated.
- Net amount: not exceeding the volume of 5 liters when directly sold to the final consumer.
- Storage conditions: "*Store in a dry place, protected from light and heat sources*".

In addition to mandatory indications, some optional information can be added:

- "*first cold pressing*" allowed only for extra virgin or virgin olive oils obtained at a temperature below 27 °C from a first mechanical pressing of the olive paste by a traditional extraction system using hydraulic presses.
- "*cold extraction*" allowed only for extra virgin or virgin olive oils obtained at a temperature below 27 °C by percolation or centrifugation of the olive paste.
- Indication of acidity or maximum acidity that, if present, must be accompanied by an indication, in lettering of the same size and in the same visual field, of the peroxide value, the wax content and the ultraviolet absorption.

1.3 Olive oil classification

The virgin olive oil is protected and normed by a comprehensive and detailed regulation including laws and standards from three organizations[6,32]:

- EU, whose Regulation is effective for the EU member states, which are the most important olive oil producers in the world.
- IOC, which is the only international organization dedicated specifically to olives and olive oils.
- FAO/OMS that, in agreement with all the member states, produces the Codex Alimentarius, a set of guidelines regarding all foods and, within them, olive oil.

In the past, regulations produced from the three organization not always agreed; however, from 2003, an adequate harmonization between them has been reached.[33–36]

Current definitions, in accordance with the last EU Regulation updates – Regulation (EU) 2013/1308 of 17 December 2013 and Regulation (EU) 2019/1604 of 27 September 2019 – are the following[33,36]:

- Virgin Olive Oils. They are "oils obtained from the fruit of the olive tree solely by mechanical or other physical means under conditions that do not lead to alterations in the oil, which have not undergone any treatment other than washing, decantation, centrifugation or filtration, to the exclusion of oils obtained using solvents or using adjuvants having a chemical or biochemical action, or by re-esterification process and any mixture with oils of other kinds." Virgin olive oils are then classified in three subgroups:
    a) Extra Virgin Olive Oil: It is "virgin olive oil having a maximum free acidity in terms of oleic acid, of 0.8 g per 100 g, the other characteristics of which comply with those laid down by the Commission in accordance with Article 75(2) for this category";
    b) Virgin Olive Oil: It is "virgin olive oil having a maximum free acidity in terms of oleic acid, of 2 g per 100 g, the other characteristics of which comply with those laid down by the Commission in accordance with Article 75(2) for this category;
    c) Lampante Olive Oil: "Lampante olive oil means virgin olive oil having a free acidity in terms of oleic acid, of more than 2 g per 100 g, and/or the other characteristics of which comply with those laid down by the Commission in accordance with Article 75(2) for this category.
- Refined Olive Oil. It is an "olive oil obtained by refining virgin olive oil, having a free acidity content, expressed as oleic acid, of not more than 0.3 g per 100 g, and the other characteristics of which comply with those laid down by the Commission in accordance with Article 75(2) for this category".
- Olive Oil – Composed of Refined Olive Oils and Virgin Olive Oils. It is "olive oil obtained by blending refined olive oil and virgin olive oil other than lampante olive oil, having a free acidity content, expressed as oleic acid, of not more than 1 g per 100 g, and the other characteristics of which comply with those laid down by the Commission in accordance with Article 75(2) for this category."
- Crude Olive-Pomace Oil. It is "oil obtained from olive pomace by treatment with solvents or by physical means or oil corresponding to lampante olive oil, except for certain specified characteristics, excluding oil obtained by means of re-esterification and mixtures with other types of oils, and the other characteristics of which comply with those laid down by the Commission in accordance with Article 75(2) for this category EN L 347/820 Official Journal of the European Union 20.12.2013".
- Refined Olive-Pomace Oil. It is "oil obtained by refining crude olive-pomace oil, having free acidity content, expressed as oleic acid, of not more than 0.3 g per 100 g, and the other characteristics of which comply with those laid down by the Commission in accordance with Article 75(2) for this category".
- Olive-Pomace Oil. It is "oil obtained by blending refined olive-pomace oil and virgin olive oil other than lampante olive oil, having a free acidity content, expressed as oleic acid, of not more than 1 g per 100 g, and the other characteristics of which comply with those laid down by the Commission in accordance with Article 75(2) for this category EN 20.12.2013 Official Journal of the European Union L 347/821".

In addition to acidity, indicating the percentage of free fatty acids developed from triglyceride enzymatic hydrolysis, other analytical parameters are used to evaluate olive oil quality and authenticity:

- Peroxide value: it is expressed in active O_2 milliequivalents per oil kg and it is index of the oil oxidation, reporting a bad storage.

- K_{232} and K_{270}, respectively the UV absorbance at 232 nm and at 270 nm, and ΔK: they indicate the presence of conjugated diene and triene because of refining or oxidation processes.
- Fatty acid ethyl esters: a high concentration depends on fermentative processes caused by bad storage or harvesting after olives' falling.
- A series of parameters indicating a contamination with seeds oil: C14:0 (EU limit: 0.05%), C18:3 (EU limit: 0.9%), C20:0 (EU limit: 0.6%), C20:1 (EU limit: 0.4%), C22:0 (EU limit: 0.2%), C24:0 (EU limit 0.2%).

Table 1.1 reports limit values of these parameters for each class previously defined.

Table 1.1. Limit values of the analytical parameters used to evaluate the oil quality[36]

	Acidity	Peroxide value (mEq O_2/Kg)	K_{232}	K_{270}	ΔK	Fatty acid ethyl esters (mg/kg)
Extra virgin olive oil	≤ 0.80	≤ 20.0	≤ 2.5	≤ 0.22	≤ 0.01	≤ 35
Virgin olive oil	≤ 2.0	≤ 20.0	≤ 2.6	≤ 0.25	≤ 0.01	/
Lampante olive oil	> 2.0	/	/	/	/	/
Refined olive oil	≤ 0.30	≤ 5.0	/	≤ 1.25	≤ 0.16	/
Olive oil composed of refined olive oil and virgin olive oil	≤ 1.00	≤ 15.0	/	≤ 1.15	≤ 0.15	/
Crude olive-pomace oil	/	/	/	/	/	/
Refined olive-pomace oil	≤ 0.30	≤ 5.0	/	≤ 2.00	≤ 0.20	/
Olive-pomace oil	≤ 1.00	≤ 15.0	/	≤ 1.70	≤ 0.18	/

Finally, depending on the different quality, Regulation (EU) 2019/1604 of 27 September 2019 establishes the following price's reference threshold regarding virgin olive oils: (a) 1779 €/ton for extra virgin olive oil; (b) 1710 €/ton for extra virgin olive oil and (c) 1 524 €/ton for lampante olive oil with two degrees of free acidity, this amount being reduced by 36.70 €/ton for each additional degree of acidity.

1.3.1 Sensory analysis

The classification of virgin olive oil has a peculiar and unique characteristic: in addition to analytical parameters, it is mandatory to collect information from a sensory evaluation, introduced by Regulation (ECC) 2586/1991.

The last update of the method for the organoleptic assessment of virgin olive oil is published in 2018 by IOC[35] and the method consists in the determination, by a group of trained tasters named *panel*, of the intensity of the *fruitiness* and of the defects perceived in the virgin olive oil. [6,35,37]

A panel is a group of 8-12 expert tasters coordinated by a panel leader and test conditions are strictly normed:[37,38]

- The oil samples shall be presented in standardized tasting glass, which are dark to avoid an influence of the color on the final evaluation. The glasses are marked with a random code of digits and letters, and shall contain 14-16 ml of oil or 12.8-14.6 g if weighted.
- The oil samples must be kept in the glasses at 28°C ± 2°C and the test room must be at a temperature between 20°C and 25°C.
- It is recommended to hold the tasting sessions between 10.00 a.m. and 12.00 noon.

- Tasters shall not smoke or drink coffee at least 30 minutes before the time set for the test and they must not have used any fragrance, cosmetic or perfumed soap 1 hour before the tasting.
- It is recommended the evaluation of four samples at the most in each session, with a maximum of three sessions per day.
- The tasting technique consists in different phases. Tasters shall rotate the glass and warm up it with hands. Then they smell the sample, taking slow deep breaths, for a maximum of 30 s; if necessary, after a short rest, the operation is tried again. At this point tasters shall evaluate the buccal sensation (retro-nasal olfactory, gustatory and tactile sensations) with small sip of approximately 3 ml of oil: it is important to distribute the oil throughout the whole of the mouth cavity and to take short and successive breaths that, drawing in air through the mouth, enables the taster both to spread the sample over the whole of the mouth and to perceive the volatile aromatic compounds via the back of the nose by forcing the use of this channel.

After the tasting, each taster of the panel shall enter the intensity of each perceived negative and positive attributes on a scale in the profile sheet provided.

At this point, the coordinator catches all the evaluations and calculates the median of each attribute and, as result of the Panel Test, virgin olive oil can be classified in one of the following classes:

- Extra virgin olive oil: the median of the defects is 0.0 and the median of the *fruity* attribute is above 0.0.
- Virgin olive oil: the median of the defects is above 0.0 but not more than 3.5 and the median of the *fruity* attribute is above 0.0.
- Ordinary virgin olive oil: the median of the defects is above 3.5 but not more than 6.0, or the median of the defects is not more than 3.5 and the median of the *fruity* attribute is 0.0.
- Lampante virgin olive oil: the median of the defects is above 6.0.

1.4 Olive oil composition

The olive oil composition is well-known, and compounds can be classified in two main groups: saponifiable and unsaponifiable fraction.

- The saponifiable fraction is the main one (98-99%), and it mainly consists of esters of glycerol with fatty acids: they are mostly triacylglycerols or triglycerides (97-98%), but small amounts of diacylglycerols or diglycerides (1-3%) and monoacylglycerols or monoglycerides (0.1-0.2%) are also present.[6,39]
- The unsaponifiable fraction (1-2%), including a series of minor compounds as hydrocarbons, aliphatic and aromatic alcohols, fat-soluble vitamins, polyphenols, pigment, phytosterols and volatile compounds. This fraction, although minor in terms of abundance, strongly affects both organoleptic and nutritive characteristic of an olive oil.[6,19,20,39]

1.4.1 Triacylglycerols or triglycerides

Triacylglycerols (TAG) are esters derived from glycerol and three fatty acids. TAG can be simple (55%) if all the alcoholic functions of the glycerol are esterified by the same fatty acid, or mixed (45%) if fatty acids are different.[6] The most prevalent triacylglycerol in olive oil is the oleic-oleic-oleic – OOO (40-59%) followed, in order of incidence, by palmitic-oleic-oleic – POO (12-20%), oleic-oleic-linoleic – OOL (12.5-20%), palmitic-oleic-linoleic – POL (5.5-7%) and stearic-oleic-oleic – SOO (3-7%).[6,39]

Therefore, oleic acid is the most plentiful fatty acid in olive oil followed by palmitic and linoleic acids,[40,41] and fatty acid compositional limits adopted in the most recent editions of the International Olive Council are given in **Table 1.2**.[42]

Table 1.2. Fatty acids composition as determined by Gas Chromatography (% m/m methyl esters): average values. * Limit raised to <0.3 for olive-pomace oils.

Fatty acid	IOC
Myristic C14:0	< 0.05
Palmitic C16:0	7.5-20.0
Palmitoleic C16:1	0.3-3.5
Heptadecanoic C17:0	≤ 0.4
Heptadecenoic C17:1	≤ 0.6
Stearic C18:0	0.5-5.0
Oleic C18:1	55.0-83.0
Linoleic C18:2	2.5-21.0
Linolenic C18:3	≤ 1.0
Arachidic C20:0	≤ 0.6
Eicosenoic C20:1	≤ 0.5
Behenic C22:0	≤ 0.2*
Lignoceric C24:0	≤ 0.2

Fatty acid composition is influenced by cultivar, climate, ripening stage etc. and, in particular, a prolonged stay of the olive on its tree causes a progressive decrease of the palmitic acid and the oleic acid with a contemporary increase of the linoleic acid.[5,6] A good quality olive oil usually contains less than 73% of oleic acid, while the percentage of linoleic acid is usually lower than 10%, with a ratio oleic/linoleic ≥ 7. The abundance of oleic acid is the reason of the higher stability and resistance to oxidation of the olive oil, if compared to other vegetable oils, usually richer in linoleic and linolenic acids, which are polyunsaturated.[6]

1.4.2 Partial glycerides

The presence of partial glycerides in olive oil is due either to incomplete triacylglycerol biosynthesis or hydrolytic reactions.[39] Diacylglycerols concentration range from 1 to 3% and their composition gives information about age of the oil and storage conditions because 1,2-diacylglycerols present in fresh oil tend to isomerize to the more stable 1,3-diacylglycerols; for this reason the ratio 1,3-DAG/1,2-DAG is usually considered a quality control criterion.[6,39]

Monoacylglycerols are present in smaller amounts, usually less than 0.25%, and 1-species are significantly higher than 2-species.[39]

1.4.3 Hydrocarbons

Hydrocarbons constitute the 30-50% of the unsaponifiable fraction and the two hydrocarbons most abundant are squalene and β-carotene.[5,39] The latter is discussed in the pigment section, while squalene, the last metabolite preceding sterol ring formation, represent itself more than 90% of the hydrocarbon fraction. In addition to them, diterpene and triterpene hydrocarbons, polyolefins and paraffins have been individuated.[39]

1.4.4 Aliphatic alcohols, aromatic alcohols, and waxes

Aliphatic and aromatic alcohols present in olive oil are found in free and esterified form and, when constituted of less than twelve carbon atoms, are part of the volatile fraction.[39] In the liquid phase, the most important are fatty alcohols and diterpene alcohols.

The main fatty alcohols are docosanol, tetracosanol, hexacosanol and octacosanol, while fatty alcohols with odd carbon atoms may be present in trace amounts.[39]

In the group of diterpene alcohols, phytol and geranylgeraniol are the main compounds. Phytol, probably originating from chlorophyll, is present in monovarietal virgin olive oils at levels ranging from 25 to 595 mg/kg, while geranylgeraniol has been found in virgin olive oil at levels lower than 50 mg/kg.[39]

Waxes, which are esters of fatty alcohols with fatty acids, are important minor olive oil constituents, and their content is officially used by the IOC for the distinction between extra virgin olive oil (waxes content < 150 mg/kg), virgin olive oil (waxes content < 200 mg/kg), lampante olive oil (waxes content < 300 mg/kg) and refined olive oil (waxes content < 350 mg/kg).[39,42]

1.4.5 Fat-soluble vitamins

Tocopherol, or vitamin E, is the main fat-soluble vitamin found in virgin olive oil, although cultivar, ripening step and technological process strongly impacts on its level, and it is important for its in vivo antioxidant properties. α-tocopherol is the main homologue with an average concentration significantly higher than 100 mg/kg, while amounts of β-tocopherol, δ-tocopherol and γ-tocopherol are usually lower than 10-20 mg/kg.[6,39]

1.4.6 Polyphenols

Polyphenols constitute the 2-3% of the unsaponifiable fraction and are specialized plant metabolites chemically characterized by the presence of one or more aromatic rings with one or more hydroxyl substituents.[6,41]

Polyphenols includes over 8000 compounds and in olive oil it is mainly composed by phenolic acids, phenolic alcohols, secoiridoids and flavonoids.[6,41] Hydroxytyrosol (HT) is the most abundant phenolic compound in olive oil and it is enclosed in complex structures, named secoiridoids, as the oleuropein. Oleuropein, due to the action of hydrolytic enzyme during the mechanical processing of the olives, is hydrolyzed into free HT and aglycon structures, which are HT precursors. During human ingestion, the precursors are themselves hydrolyzed leading to an increase in free HT.[41] Polyphenols are known for their remarkable antioxidant power and strongly contributes to the

stability of olive oil; however, HT is the only one backed by a European Food Safety Authority (EFSA) health claim, which focuses on the protection provided by HT and its derivatives, *i.e.*, oleuropein and tyrosol, to LDL-C particles against oxidative damage.[39–41,43]

Moreover, as previously mentioned in **Chapter 1.2**, the technological process strongly impacts on polyphenols content, and their abundance is directly correlated with the intensity of bitter and pungent notes.[43]

1.4.7 Pigments

Pigments responsible of green and yellow shades of virgin olive oil belongs to two groups of fat-soluble pigments: chlorophylls and carotenoids.[39] Chlorophylls in olive oil are principally pheophytin α and chlorophylls α, more abundant in fresh oils, while β forms are present in minor amounts; chlorophylls have prooxidant activity in the presence of light and antioxidant activity in the dark.[6,39] The most important carotenoids, instead, are lutein and β-carotene.[39]

1.4.8 Phytosterols

Phytosterols are important constituents related to the quality of the oil and used for checking its genuineness: total sterol content, expressed in mg/kg, is normed by IOC legislation, with refined olive oils usually containing lower levels (losses of phytosterols during the refining proves may be as high as 25/30%).[39,42] The main component is β-sitosterol, which constitutes between 75% and 90% of the total sterol fraction; other important phytosterol are campesterol and stigmasterol, constituting about 4% and 2%, respectively, of the total fraction.[39]

1.5 Volatile compounds

Volatile compounds are low-molecular weight (< 300 Da) non polar compounds released from the olive oil at room temperature.[44] A subgroup of them, after reaching the olfactory epithelium of both orthonasal and retronasal olfaction, dissolves into the mucus and bonds with olfactory receptors to give an odor sensation.

1.5.1 Origin and composition of the volatile fraction

Volatiles organic compounds (VOCs) characterizing VOO mainly originate from three pathways: the lipoxygenase (LOX) pathway, the autooxidation and photooxidation of fatty acids, and the microbiological activity.[44,45]

1.5.1.1 Lipoxygenase pathway

The LOX pathway involves several enzymes and produces C5 and C6 volatile aldehydes, alcohols, and esters, which are the most important fraction of volatile compounds of high quality virgin olive oils; they are responsible for the *green* and *fruity* notes perception.[44–46]

The pathway starts with the production of 9- and 13-hydroperoxides of linoleic and linolenic acids, which are subsequently cleaved by hydroperoxide lyases leading to C6 aldehydes (*i.e.,* hexanal, *(Z)*-3-hexenal and *(E)*-2-hexenal, with the unsaturated ones that can isomerize from cis-3 to trans-2 form, more stable).[46] At this point, through the mediation of alcohol dehydrogenase, C6 aldehydes are reduced to the corresponding alcohols (*i.e.,* 1-hexanol, *(Z)*-3-hexen-1-ol and *(E)*-2-hexen-1-ol) ,

which are themselves substrate for the production of esters (*i.e.,* hexyl acetate, *(Z)*-3-hexenyl acetate, *(E)*-2-hexenyl acetate) because of the catalytic activity of alcohol acetyl transferases.[45,46]

When the substrate is linolenic acid, an addition branch of the LOX pathway is active: hydroperoxides are cleaved through alkoxy radical producing stabilized 1,3-pentene radicals that can dimerize to C10 hydrocarbons or couple with a hydroxy radical producing C5 alcohols, which can be enzymatically oxidated to the corresponding C5 carbonyl compounds (*e.g.,* 1-penten-3-ol, 1-penten-3- one etc.).[45,46]

1.5.1.2 Fatty acids oxidation

Although VOO is one of the edible oil most resistant to auto-oxidation and photo-oxidation process, thanks to the abundance of antioxidants and the prevalence of oleic acid, if oxidation is poorly prevented it can cause the development of compounds responsible for off-flavors perception.[44–46]

The autoxidation process is a radical-induced chain reaction, involving unsaturated fatty acids, that triggers the formation of free radicals and hydroperoxides; these latter decompose in secondary products among which there are aroma active compounds responsible for off-flavors, in particular *rancidity*.[45] Autoxidation starts with the formation of a lipid alkyl radical through the removal of a hydrogen atom from the fatty acid chain and the subsequent quick reaction with singlet O_2 to form a lipid peroxyl radical. The initiation step is favored by heat, metal catalysts, and UV-Vis light; in particular, the effect of light in the presence of sensitizers such as chlorophylls is particularly strong and is known as the photo-oxidation.

VOCs usually associated with oxidation process and rancid perception are mainly aldehydes (*i.e.,* pentanal, *(E)*-2-heptenal, *(E,E)*-2,4-heptadienal, nonanal, *(Z)*-2-nonenal, *(E,Z)*-2,4-decadienal and *(E,E)*-2,4- decadienal), but also short chain fatty acids (*e.g.,* acetic, propanoic and butanoic acids), octane, lactones, and furans have been associated with oxidation.[45] The role of hexanal is instead ambiguous, being it produced both linoleic acid oxidation and the LOX pathway. Finally C8 compounds, usually associated with musty perception, are associated to a photo-oxidation process.[45]

1.5.1.3 Microbiological activity

Another possible source of VOC responsible for off-flavors is represented by enzymatic activities of mold, yeasts and other microorganisms that may depend on both improper and prolonged olive storage, and prolonged storage of unfiltered VOOs.[44–46]

Compounds responsible for the *fusty* defect derives from the conversion of some amino acids: 2-methyl-butanal, 2-methyl-1-butanol, 2-methyl- butanoic acid from isoleucine, and 3-methyl-butanal, 3-methyl-1- butanol, 3-methyl-butanoic acid from leucine. Molds belonging to the *Penicillium* and *Aspergillus* species lead to an increase of C8 compounds (*e.g.*, 1-octen-3-ol, 1-octen-3- one) and methyl ketones (*i.e.,* 2-heptanone, 2-octanone and 2-nonanone), while *acetobacters* are responsible for the oxidation of ethanol (produced by yeast through the fermentation of sugars) in acetic acid.[45]

Other VOCs usually associated with the presence of molds, yeasts and/or bacteria are benzyl alcohol, phenylethyl alcohol, 2- and 3-methyl-1-butanol, 1-pentanol, 2-heptanol, butanone, and methyl-butyl acetate. Moreover, the presence of microorganisms is often associated with an overall

decrease of the VOCs belonging to the LOX pathway, probably inhibited because of an increase of acidity and the competition among enzymes of these species and the LOX pathway.[45]

1.5.2 Minor components of the volatile fraction

In addition to the major chemical classes mentioned above (*i.e.,* saturated and unsaturated aldehydes, ketones, esters and alcohols), together with their origin pathway, other minor classes compose the volatile fraction of VOO. They include either native compounds or compounds originated from minor pathways.

Lactones have been detected in variables amounts depending on cultivar, and their role in defining the aroma of VOO is still unclear, even if some authors reported them as contributing to determine *fruity* notes.[45,47] The most common identified are butyrolactone, δ-valerolactone γ-caprolactone, γ-nonalactone, γ-decalactone, δ-decalactone, and γ-dodecalactone.[45]

Terpenes have been identified in the volatile fraction of VOOs and the composition of this chemical class is strongly influenced by the cultivar and the pedoclimatic conditions. δ-limonene, β-ocimene, linalool, α-copaene, *(E,E)*-α-farnesene, and δ-3-carene are the most common and reports as contributing to define *green* and *floral* notes.[45,48,49]

Several saturated and unsaturated hydrocarbons are present, however they are characterized by high odor thresholds and their contribution to the overall aroma is poor, although synergistic effects may affect it.[45,50] However, in addition to well-known hydrocarbons derived from endogenous mechanisms (*e.g.*, pentene dimers formed in a branch of the LOX pathway), some hydrocarbons derived from exogenous contamination (*i.e.*, environmental contaminants) have been occasionally found: acenaphthene, acenaphthylene, anthracene, some benzene derivatives, fluorene, fluoranthene, phthalates, and pyrene.[45]

Finally, some minor classes, as furans, identified at low concentration and probably resulting from decomposition of hydroperoxides during oil storage,[51] and volatile phenols, as guaiacol, 4-ethylguaiacol, 4-ethylphenol, and 4-vinylphenol, contributing to define *fusty* and *musty* defects.[45]

1.6 Olive oil consumption: current challenges

In the last 25 years the EVOO, symbol of the Mediterranean diet, has increased its worldwide diffusion and consumption.[40] The increased number of producer and consumer countries has stressed the request of a quality and authenticity control, and the concept of traceability has become an issue of great interest for regulators, suppliers, and consumers, in both traditional and emerging markets.[52]

Traceability consists of documented proof of the identity of a product and it has two main goals: (a) juridical, in case of physical or economical damage to the consumers because of wrongdoing or misleading information; (b) technical, with the identification of the causes of loss or spoilage, in order to apply an appropriate correction and individuate the step of the production where improper handling or fraud took place. [14] For this reason, traceability is based on documenting material balances through management and discrete batch monitoring, where a batch is intended as a portion of a given material having a specific identity and composition.[14]

However, despite the increased relevance of the traceability concept, first criterion of choice for most consumers is still price, with the current market scenario characterized by increasing levels of competition based on cost reduction.[14,53] This strategy depends on consumers' difficulties in evaluating the quality of olive oils and on a lack in the olive oil culture, but the price war cannot last

forever, and it is important to develop persuasive communication strategies to increase consumers' awareness and, consequently, their willingness to pay an higher price to guarantee a proper income for high quality producers.[53] In particular, the adoption of health claims is needed, being EVOO known for its nutraceutical value and a lot of studies confirm and investigate the beneficial effects, especially on cardiovascular system, of a regular EVOO consumption; these positive effects are mainly driven by phenolic constituents, especially hydroxytyrosol and oleuropein. Polyphenols' concentration and composition is not only related to health effects, but it also affects the sensory perception of an EVOO, being directly correlated with the *bitter* perception that, together with *fruity* and *pungent*, is the main positive attributes considered by panelists.[53,54] However, as reported in literature, on average consumers still don't have the ability to discriminate between an high quality and a low quality olive oil[53,55] or, even, they prefer a "neutral taste" instead of a higher *bitterness* and *pungency*.[54]

Currently, the emerging need to characterize the EVOOs quality, meant as sensory properties, traceability, safety, nutritional value and biological activity led to the development of advanced analytical methodologies applying "omics" strategies to investigate the EVOOs chemical fingerprints.[9,56–58] In particular, multidimensional analytical (MDA) platforms combining physicochemical separation techniques (*e.g.*, chromatography, electromigration, size exclusion, etc.) with spectroscopic techniques (*e.g.*, Mass Spectrometry - MS, Nuclear Magnetic Resonance - NMR, Infrared - IR, Ultraviolet - UV, etc.) have the potential to discriminate, identify and quantify sample constituents, providing a solid foundation for generating hypothesis-driven studies.

In the next chapter, principles and evolution of the comprehensive two-dimensional gas chromatography (GC×GC) are introduced, being this technique, usually coupled to MS, the technical core of the research activity of this PhD project.

References

(1) Pignatti, S. Flora d'Italia. 1982.

(2) Maugini, E.; Maleci Bini, L.; Mariotti Lippi, L. *Botanica Farmaceutica*, IX Edizion.; Piccin, 2014.

(3) Poli, F. *Biologia Farmaceutica*; Pearson, 2019.

(4) Tamasi, G.; Bonechi, C.; Belyakova, A.; Pardini, A.; Rossi, C. The Olive Tree, a Source of Antioxidant Compounds. *J. Siena Acad. Sci.* **2017**, *8* (1). https://doi.org/10.4081/jsas.2016.6952.

(5) Cabras, P.; Martelli, A. *Chimica Degli Alimenti*; Piccin, Ed.; 2004.

(6) Cappelli, P.; Vannucchi, V. *Principi Di Chimica Degli Alimenti*; Zanichelli, 2016.

(7) Olvigelato. L'olio di oliva https://www.olvigelato.com/l-olio-di-oliva.html.

(8) Vossen, P. Olive Oil: History, Production, and Characteristics of the World's Classic Oils. *HortScience* **2007**, *42* (5), 1093–1100. https://doi.org/10.21273/HORTSCI.42.5.1093.

(9) Aparicio, R.; Harwood, J. *Handbook of Olive Oil. Analysis and Properties. Second Edition*; Springer, 2013.

(10) Riley, F. R. Olive Oil Production on Bronze Age Crete: Nutritional Properties, Processing Methods and Storage Life of Minoan Olive Oil. *Oxford J. Archaeol.* **2002**, *21* (1), 63–75. https://doi.org/10.1111/1468-0092.00149.

(11) Connor, D. J.; Gómez-del-Campo, M.; Rousseaux, M. C.; Searles, P. S. Structure, Management and Productivity of Hedgerow Olive Orchards: A Review. *Sci. Hortic. (Amsterdam).* **2014**, *169*, 71–93. https://doi.org/10.1016/j.scienta.2014.02.010.

(12) Raddato, C. Minoan Storage Jars at the Palace of Knossos https://www.ancient.eu/image/10598/minoan-storage-jars-at-the-palace-of-knossos/.

(13) Kapellakis, I. E.; Tsagarakis, K. P.; Crowther, J. C. Olive Oil History, Production and by-Product Management. *Rev. Environ. Sci. Bio/Technology* **2008**, *7* (1), 1–26. https://doi.org/10.1007/s11157-007-9120-9.

(14) Peri, C. *The Extra Virgin Olive Oil Handbook*; Wiley-Blackwell, 2014.

(15) Aprea, E.; Gasperi, F.; Betta, E.; Sani, G.; Cantini, C. Variability in Volatile Compounds from Lipoxygenase Pathway in Extra Virgin Olive Oils from Tuscan Olive Germoplasm by Quantitative SPME/GC-MS. *J. Mass Spectrom.* **2018**, *53* (9), 824–832. https://doi.org/10.1002/jms.4274.

(16) Extraevo. Italian olive cultivars https://www.extraevo.com/italian-olive-cultivars/.

(17) Our extra virgin olive oil. A guide to the most common olive cultivars in Europe https://ouroliveoil.com/post/a-guide-to-the-most-common-olive-cultivars-in-europe?category=olive-oil-facts.

(18) Torino, P. *L'olivicoltura in Provincia Di Torino - Un Censimento Ragionato*; 2008.

(19) Sohaimy, A. A. S. El; Sheikh, H. M. E.-; Refaay, M. T.; Zaytoun, A. M. M. Effect of Harvesting in Different Ripening Stages on Olive (Olea Europea) Oil Quality. *Am. J. Food Technol.* **2015**, *11* (1–2), 1–11. https://doi.org/10.3923/ajft.2016.1.11.

(20) Sibbett, S. G.; Ferguson, L. *The Olive Oil Production Manual*, Second edi.; Agriculture and

Natural Resources, 2004.

(21) Montaño, A.; Zambrano, M.; Lázaro-Madrea, A.; Martínez, B. Monitorización del grado de maduración de la aceituna: nuevos parámetros para la variedad Arbequina Title https://www.interempresas.net/Produccion-Aceite/Articulos/209789-Monitorizacion-grado-maduracion-aceituna-parametros-variedad-Arbequina.html.

(22) Uceda, M.; Frias, L. Harvest Dates: Evolution of the Fruit Oil Content, Oil Composition and Oil Quality. *Proc. 2nd Semin. Oleic. Int.* **1975**, 125–128.

(23) Di Giovacchino, L.; Sestili, S.; Di Vincenzo, D. Influence of Olive Processing on Virgin Olive Oil Quality. *Eur. J. Lipid Sci. Technol.* **2002**, *104* (9–10), 587–601. https://doi.org/10.1002/1438-9312(200210)104:9/10<587::AID-EJLT587>3.0.CO;2-M.

(24) Postma-botha, M. The Extraction , Quantification and Application of High-Value Biological Compounds from Olive Oil Processing Waste The Extraction , Quantification and Application of High-Value Biological Compounds from Olive Oil Processing Waste. **2018**.

(25) orsomarsoblues. TEMPO DELLE OLIVE: Fasi della lavorazione nel frantoio https://www.orsomarsoblues.it/2019/11/tempo-delle-olive-fasi-della-lavorazione-nel-frantoio/.

(26) Addison, A. Stone mills make better oil: it's fiction https://www.oliveoilsource.com/article/stone-mills-make-better-oil-its-fiction.

(27) Servili, M.; Esposto, S.; Taticchi, A.; Urbani, S.; Selvaggini, R.; Di Maio, I.; Veneziani, G. Innovation in Extraction Technology for Improved Virgin Olive Oil Quality and By-Product Valorisation. *Acta Hortic.* **2011**, *888*, 303–316.

(28) Angerosa, F.; Mostallino, R.; Basti, C.; Vito, R. Influence of Malaxation Temperature and Time on the Quality of Virgin Olive Oils. *Food Chem.* **2001**, *72* (1), 19–28. https://doi.org/10.1016/S0308-8146(00)00194-1.

(29) Teatro Naturale. Basse rese in avvio di campagna olearia in Spagna https://www.teatronaturale.it/tracce/mondo/26795-basse-rese-in-avvio-di-campagna-olearia-in-spagna.htm.

(30) Clodoveo, M. L. An Overview of Emerging Techniques in Virgin Olive Oil Extraction Process: Strategies in the Development of Innovative Plants. *J. Agric. Eng.* **2013**, *44* (2s), 297–305. https://doi.org/10.4081/jae.2013.s2.e60.

(31) European Commission. Regulation (EU) No 29/2012 of 13 January 2012 on Marketing Standards for Olive Oil (Codification). *Off. J. Eur. Union* **2012**, No. 12, 14–21.

(32) H.-D.Belitz; W.Grosch; P.Schieberle. *Food Chemistry-4th Revised and Extended Edition*; 2013; Vol. 53. https://doi.org/10.1017/CBO9781107415324.004.

(33) European Parliament; Council of the European Union. Regulation (EU) No 1308/2013 of 17 December 2013. *Off. Journals Eur. Union* **2013**, *2008* (1308), 184.

(34) FAO; WHO. Standard for Olive Oils and Olive Pomace Oils. *Codex Aliment.* **2015**.

(35) IOC. Sensory Analysis of Olive Oil - Method for the Organoleptic Assessment of Virgin Olive Oil. *Int. Olive Counc.* **2018**, No. 15, Madrid, Spain.

(36) European Commission. Regulation (EU) No 1604/2019 of 27 September 2019. *Off. Journals Eur. Union* **2019**, *2019* (2568).

(37) IOC. Method for the Organoleptic Assessment of Extra Virgin Olive Oil Applying To Use a Designation of Origin. **2005**, No. 22, 1–29.

(38) IOC. International Olive Council. *Newsl. No 144 DECEMBER 2019* **2019**, No. December.

(39) Boskou, D.; Blekas, G.; Tsimidou, M. *Olive Oil Composition*; 2006. https://doi.org/10.1016/B978-1-893997-88-2.50008-0.

(40) Tome-Carneiro, J.; Crespo, M. C.; de las Hazas, M. C. L.; Visioli, F.; Davalos, A. Olive Oil Consumption and Its Repercussions on Lipid Metabolism. *Nutr. Rev.* **2020**, *78* (11), 952–968. https://doi.org/10.1093/nutrit/nuaa014.

(41) Visioli, F.; Davalos, A.; López de las Hazas, M. C.; Crespo, M. C.; Tomé-Carneiro, J. An Overview of the Pharmacology of Olive Oil and Its Active Ingredients. *Br. J. Pharmacol.* **2020**, *177* (6), 1316–1330. https://doi.org/10.1111/bph.14782.

(42) IOC. International Trade Standard Applying To Olive Oils and Olive-Pomace Oils. **2015**, No. 3, 1–17.

(43) Rigacci, S.; Stefani, M. Nutraceutical Properties of Olive Oil Polyphenols. An Itinerary from Cultured Cells through Animal Models to Humans. *Int. J. Mol. Sci.* **2016**, *17* (6), 843. https://doi.org/10.3390/ijms17060843.

(44) Kalua, C. M.; Allen, M. S.; Bedgood, D. R.; Bishop, A. G.; Prenzler, P. D.; Robards, K. Olive Oil Volatile Compounds, Flavour Development and Quality: A Critical Review. *Food Chem.* **2007**, *100* (1), 273–286. https://doi.org/10.1016/j.foodchem.2005.09.059.

(45) Cecchi, L.; Migliorini, M.; Mulinacci, N. Virgin Olive Oil Volatile Compounds: Composition, Sensory Characteristics, Analytical Approaches, Quality Control, and Authentication. *J. Agric. Food Chem.* **2021**. https://doi.org/10.1021/acs.jafc.0c07744.

(46) Angerosa, F.; Servili, M.; Selvaggini, R.; Taticchi, A.; Esposto, S.; Montedoro, G. Volatile Compounds in Virgin Olive Oil: Occurrence and Their Relationship with the Quality. *J. Chromatogr. A* **2004**, *1054* (1–2), 17–31. https://doi.org/10.1016/j.chroma.2004.07.093.

(47) Kesen, S.; Kelebek, H.; Selli, S. Characterization of the Key Aroma Compounds in Turkish Olive Oils from Different Geographic Origins by Application of Aroma Extract Dilution Analysis (AEDA). *J. Agric. Food Chem.* **2014**, *62* (2), 391–401. https://doi.org/10.1021/jf4045167.

(48) Hachicha Hbaieb, R.; Kotti, F.; Gargouri, M.; Msallem, M.; Vichi, S. Ripening and Storage Conditions of Chétoui and Arbequina Olives: Part I. Effect on Olive Oils Volatiles Profile. *Food Chem.* **2016**, *203*, 548–558. https://doi.org/10.1016/j.foodchem.2016.01.089.

(49) Mascrez, S.; Psillakis, E.; Purcaro, G. A Multifaceted Investigation on the Effect of Vacuum on the Headspace Solid-Phase Microextraction of Extra-Virgin Olive Oil. *Anal. Chim. Acta* **2020**, *1103* (xxxx), 106–114. https://doi.org/10.1016/j.aca.2019.12.053.

(50) Angerosa, F.; Camera, L.; D'Alessandro, N.; Mellerio, G. Characterization of Seven New Hydrocarbon Compounds Present in the Aroma of Virgin Olive Oils. *J. Agric. Food Chem.* **1998**, *46* (2), 648–653. https://doi.org/10.1021/jf970352y.

(51) Angerosa, F. Influence of Volatile Compounds on Virgin Olive Oil Quality Evaluated by Analytical Approaches and Sensor Panels. *Eur. J. Lipid Sci. Technol.* **2002**, *104* (9–10), 639–660. https://doi.org/https://doi.org/10.1002/1438-9312(200210)104:9/10<639::AID-EJLT639>3.0.CO;2-U.

(52) Bajoub, A.; Bendini, A.; Fernández-Gutiérrez, A.; Carrasco-Pancorbo, A. Olive Oil Authentication: A Comparative Analysis of Regulatory Frameworks with Especial Emphasis on Quality and Authenticity Indices, and Recent Analytical Techniques Developed for Their Assessment. A Review. *Crit. Rev. Food Sci. Nutr.* **2018**, *58* (5), 832–857. https://doi.org/10.1080/10408398.2016.1225666.

(53) Roselli, L.; Clodoveo, M. L.; Corbo, F.; De Gennaro, B. Are Health Claims a Useful Tool to Segment the Category of Extra-Virgin Olive Oil? Threats and Opportunities for the Italian Olive Oil Supply Chain. *Trends Food Sci. Technol.* **2017**, *68*, 176–181. https://doi.org/10.1016/j.tifs.2017.08.008.

(54) Cavallo, C.; Caracciolo, F.; Cicia, G.; Del Giudice, T. Extra-Virgin Olive Oil: Are Consumers Provided with the Sensory Quality They Want? A Hedonic Price Model with Sensory Attributes. *J. Sci. Food Agric.* **2018**, *98* (4), 1591–1598. https://doi.org/10.1002/jsfa.8633.

(55) Salazar-Ordóñez, M.; Schuberth, F.; Cabrera, E. R.; Arriaza, M.; Rodríguez-Entrena, M. The Effects of Person-Related and Environmental Factors on Consumers' Decision-Making in Agri-Food Markets: The Case of Olive Oils. *Food Res. Int.* **2018**, *112*, 412–424. https://doi.org/10.1016/j.foodres.2018.06.031.

(56) Kalogiouri, N. P.; Aalizadeh, R.; Dasenaki, M. E.; Thomaidis, N. S. Application of High Resolution Mass Spectrometric Methods Coupled with Chemometric Techniques in Olive Oil Authenticity Studies - A Review. *Anal. Chim. Acta* **2020**, *1134*, 150–173. https://doi.org/10.1016/j.aca.2020.07.029.

(57) Sebastiani, L.; Busconi, M. Recent Developments in Olive (Olea Europaea L.) Genetics and Genomics: Applications in Taxonomy, Varietal Identification, Traceability and Breeding. *Plant Cell Rep.* **2017**, *36* (9), 1345–1360. https://doi.org/10.1007/s00299-017-2145-9.

(58) Stilo, F.; Bicchi, C.; Robbat, A.; Reichenbach, S. E.; Cordero, C. Untargeted Approaches in Food-Omics: The Potential of Comprehensive Two-Dimensional Gas Chromatography/Mass Spectrometry. *TrAC Trends Anal. Chem.* **2021**, *135*, 116162. https://doi.org/10.1016/j.trac.2020.116162.

Chapter 2

Comprehensive two-dimensional gas chromatography

2.1 Introduction

As anticipated in **Chapter 1**, EVOO's volatile fraction, and more in general foods' volatile fraction, consists of a large number of chemical classes having a wide molar mass range, variable polarity, and a broad concentration range.[1] Over the years, great efforts have been done to develop improved analytical methods for foods analysis, and nowadays MDA platform combining physicochemical separation with spectroscopic techniques are considered the golden standard to discriminate, identify and quantify foods constituents. This chapter is focused on the "core" of the MDA platform used during this PhD project: the separation step, which is often the limiting factor in a complete analysis of complex chemical mixtures.[2] An introduction to the history of multidimensional gas chromatography (MDGC) resulting in the development of comprehensive two-dimensional gas chromatography (GC×GC) is here proposed, together with some fundamental and theoretical aspects.

2.1.1 The origin of Chromatography and Gas Chromatography (GC)

The first idea of chromatography was reported by Professor Friedrich Goppelsröder during the Congress of "Naturforschenden Gesellschaft" held in Basel in 1861 and the successive paper "Note sur une méthode nouvelle propre à déterminer la nature d'un mélange de principes colorants" (Note on a new method for determining the nature of a mixture of dyes) published in 1862.[3,4] The paper reported that: "*A strip of filter paper (is dipped) several millimetres into an aqueous solution of blue litmus. The solution is seen to rise quickly above the level of the liquid by capillary aspiration. (…) No separation between solvent and dissolved substance is observed. A different situation arises when the experiment is repeated with litmus coloured with sulphuric acid. In this case, three zones are formed on the paper above the liquid. The first (i.e., the highest) contains only water, the second is due to dilute sulphuric acid, and the third contains water, acid, and dye.*

Clearly, a partial separation of the three components in the mixture has occurred and is due to their different abilities to rise within the pores of the paper." Goppelsröder, at the end of the paper, added the following consideration: "*I saw in these observations the key to a new analytical method*".[3,4]

Goppelsröder's observations were, around 40 years later, taken in account by Mikhail Semenovich Tswett, which is usually considered the inventor of chromatography.[4] Tswett described the method in 1901 during the XI Congress of Naturalists and Physicians in St. Petersburg. Then, in 1906, published a paper presenting a separation of herbal pigments in colored bands by using calcium carbonate as a solid absorber and petroleum ether with a small amount of alcohol as the mobile phase, through the instrumentation shown in **Figure 2.1**.[5]

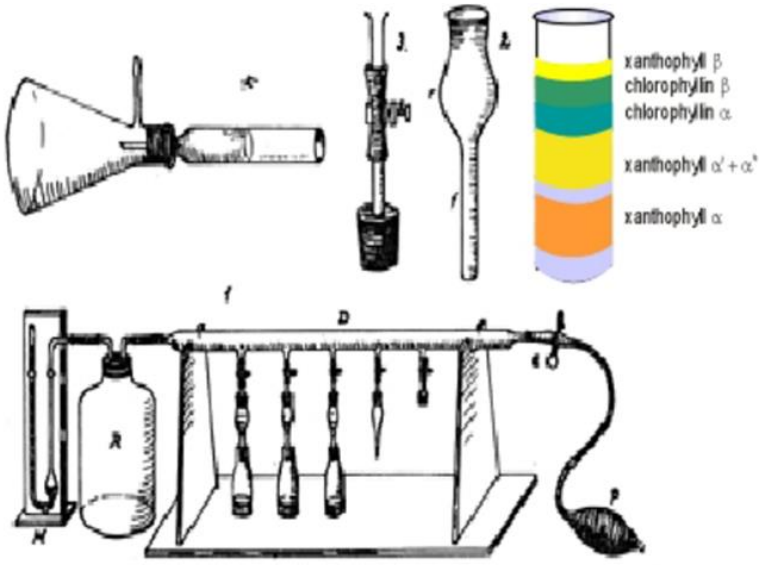

Figure 2.1. The first chromatographic instrumentation, presented by Twsett in 1901.

In this paper he termed for the first time this technique as "*chromatography*", derived from the Greek words χρωμα, meaning color, and γραφι, to write.[4,5]

Tswett's research was the starting point for a century, the 20th, dominated by a growing interest on the exploration of nature's complex mixtures.[6] Each decade brought new innovations in the chromatography area and, at the end of the 20th century, chromatography has become the most widely used separation technique in chemistry and biochemistry.[7]

In the 1930s chromatographic separation by continuous adsorption/desorption on open columns was applied on various plant extracts.[8,9] In 1941 Martin and Synge replaced countercurrent liquid-liquid extraction by partition chromatography for the analysis of amino acids contained in wool.[10] Martin again, together with James, developed gas chromatography (GC) in 1951 by using a gas, rather than a liquid, as mobile phase.[11,12] A turning point is represented by the work of Marcel Golay, in 1958, which showed how a tortuous path along a packed bed could be replaced by a straighter path through a narrow open tube.[13] Metal and glass capillaries were soon fabricated, however applications on capillary columns were minor, until the revolutionary introduction of silica column in 1979 and the successive development of a variety of stationary phases.[6]

Since its introduction in 1951, GC demonstrated to be suitable for food analysis and it has been successfully adopted for different applications: flavor and aroma characterization; investigation of volatile organic compounds (VOCs) mixtures from vegetable matrices, essential oils and extracts;

fat analysis and characterization; residues and contaminants determination, food packaging migration products; food origin authentication etc.[14]

2.1.2 From GC to MultiDimensional Gas Chromatography (MDGC)

Over time, however, the complexity of the applications increased, with a growing interest for the determination of non-targeted compounds, and this process reached its apex in the late 1990s, with the dawn of the "-omic" sciences.[15] The impact of these increasingly sophisticated applications opens to the challenge to improve the peak capacity and the resolution power of GC.

Peak capacity, introduced by Giddings in 1967,[16] is defined as the maximum number of peaks that can be separated on a given column with a defined resolution in a defined retention time window, *e.g.,* starting from the first peak (hold-up time) up to the last peak (retention time or retention factor of the last peak).[16,17] It can be described as the number of individual components that can be placed, side by side, as single entities, within the separation space.[2] Peak capacity (n_C), calculated as indicated in **Equation 2.1**, is correlated with peak width and it is a metric that informs about column efficiency.[17]

Equation 2.1. $$n_c = 1 + \frac{\sqrt{N}}{4R_S} * ln(\frac{t_{R(max)}}{t_M})$$

where N is the plate number, R_S peaks resolution, t_M hold-up time, $t_{R(max)}$ the retention time of the last peak.

However, peak capacity is a theoretical value assuming that the peaks are evenly distributed across the chromatograms, but unfortunately it seldom happens in real separations. Indeed, Davies and Giddings demonstrated that peak resolution (*i.e.,* the degree of separation for an adjacent peak pair) is affected if the number of solutes exceeds 37% of the peak capacity and that, in order to resolve 98% of the components, the peak capacity must exceed the number of components by a factor of 100.[18,19] The most common approach to improve peak capacity matching for the analysis of very complex mixtures is to use a longer column; however, a more effective way to enhance separation power is to work on technology implementation.[20]

Matching for new challenging applications, during decades, two different path of technology implementation were engaged. On one side the improvement of the resolution power of chromatographic columns, as done with the invention of the capillary columns by Golay. On the other side, the route of the multidimensional chromatography (MDC) and the multidimensional gas chromatography (MDGC) was explored, with many significant developments in the last 60 years.[15]

2.2 MDGC and the development of comprehensive two-dimensional gas chromatography (GC×GC)

2.2.1 Concept of multidimensionality and multidimensional gas chromatography (MDGC)

Multidimensional separations have the advantage of providing a greatly increased peak capacity due to the combination of multiple separation steps, or dimensions, within a single analysis.[1,19] Moreover, MDC and MDGC add further advantages related to the possibility of exploring sample chemical dimensionality, a parameter that strongly influences component resolution. Sample

dimensionality (s) is defined by Giddings as "*the number of independent variables that must be specified to identify the components of the sample. It is assumed (as part of this definition) that the properties of the components, including chromatographic retention parameters, vary in some systematic way with the s variables.*".[19]

The example in **Figure 2.2A** can be helpful to understand this concept. The components of a sample are distributed in a 3D space defined, on the three axes, by volatility, polarity, and molecular weight. Different volatility of the components of a sample is the main separation mechanisms in GC, and it is captured by different parameters as boiling point or Kovats/Van den Dool retention index.[21,22] Polarity is the second most exploited separation mechanisms as it results from several interactions mechanisms, including hydrogen bonding, dipole-dipole forces, and dipole-induced dipole forces.[15,17] Finally molecular weight, defining the range of molecules composing the sample.[15] In a simple mixture consisting of 60 analytes (from one to three order of magnitude lower than the components of a natural mixture) a mono-dimensional separation (*e.g.,* mono-dimensional gas chromatography - 1D-GC) corresponds to a linear probe (**Figure 2.2B**) exploring only of the three dimensions previously defined. A multidimensional separation, instead, allows the exploration of the sample dimensionality by using a planar probe (**Figure 2.2C**).[15]

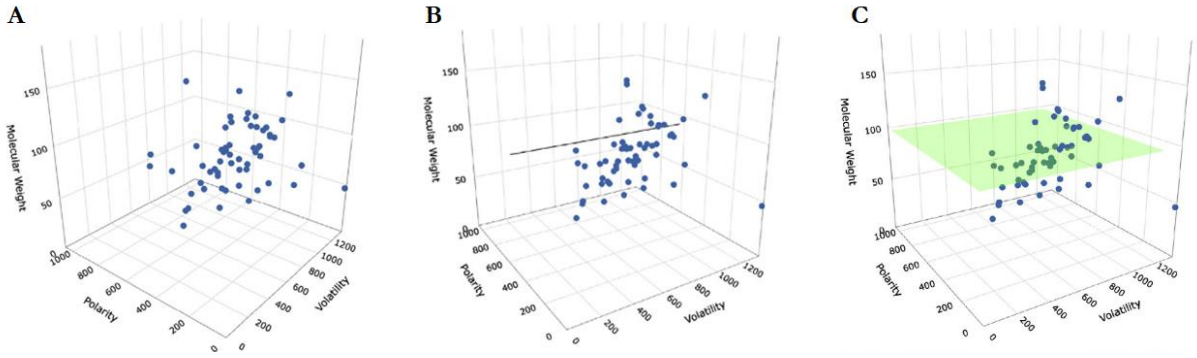

Figure 2.2B. A sample described by three dimensions (**2.2A**) and explored by a mono-dimensional separative technique (**2.2B**) or a two-dimensional separative technique (**2.2C**)

MDC has therefore emerged has a probe able to better investigate sample composition and to obtain higher and further levels of information. More recently a general definition of MDC was proposed as "*n-dimensional analysis that generates n-dimensional displacement information*".[14] A variety of combinations of different separation mechanisms can be used to create multidimensional separation as long as, as discussed by Giddings in 1990,[23] two conditions are verified.[2] The first one is that the two (or more) separations steps should be governed by orthogonal discrimination principles; the second is that the separation produced during the first separation step should be kept in the following step, so that the resolving power of the composite separation exceeds that of each individual stage.[14,24,25]

When gas-phase separations are considered, MDGC has been defined by Marriott as "*the process of selecting a (limited) region or zone of eluted compounds issuing from the end of one GC column, and subsequently subjecting the zone to a further GC displacement*".[26,27]

The large majority of the MDGC separations use two columns, therefore they are classified as two-dimensional gas chromatography (2D-GC).[25] **Figure 2.3** shows the schematic diagram of a 2D-GC system, which can be classified into two main categories: heart-cut (H/C or H/C MDGC or GC-GC) and comprehensive two-dimensional gas chromatography (GC×GC).[1,20,25,28,29]

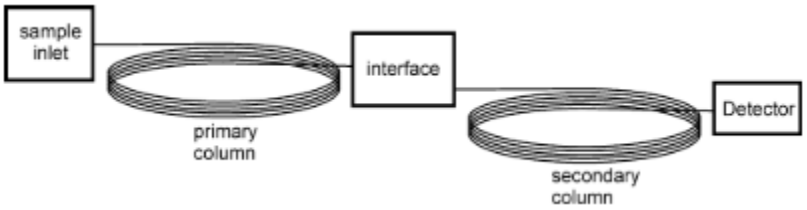

Figure 2.3. Scheme of a 2D-GC. Differences between H/C (or GC-GC) and GC×GC depends on the operative principles and dynamics of the interface.[25]

GC-GC or heart-cut, where a single fraction (or a few specific fractions) of the primary column eluate is transferred to the secondary column for further separation, with the aim to improve separation by combining different discrimination principles.[14,20] The theoretical peak capacity of such a system is the sum of peak capacities of the first and second dimension, the latter multiplied by the number of heart-cuts performed (**Figure 2.4A**).[20] This method proved to be effective in target analysis, where the focus is on individual fraction(s) and not on the entire sample.[2,27] If the interest is extended to the whole sample components, multiple analytical runs are required to enable multiple yet comprehensive cuts over the entire volatility/polarity range.

The solution for those interested in separating the entire sample in two dimensions is to extend comprehensively the heart-cut over the entire analytical run as it is done by GC×GC.[20] The first separation through a GC×GC platform was documented by Liu and Phillips 30 years ago, in 1991:[30] it was characterized by a fast and continuous heart-cutting, named modulation.[14,27] The interface between the two columns in GC×GC is the "modulator" and is considered the "heart" of the GC×GC system. The modulator works as an inter-column injector that serially traps fractions eluting from the primary column before their re-injection/introduction into the secondary column. These operations occur at a fixed time intervals called modulation period (P_M).[15] In theory, by GC×GC, the system peak capacity is approximately the product of the peak capacities of the two separation dimensions (**Figure 2.4B**).[20] However, Blumberg et al. [2,20] proved that it approaches the theoretical one just when optimal chromatographic conditions are applied to both analytical dimensions and modulation efficiency does not produce excessive band-broadening in space.[31]

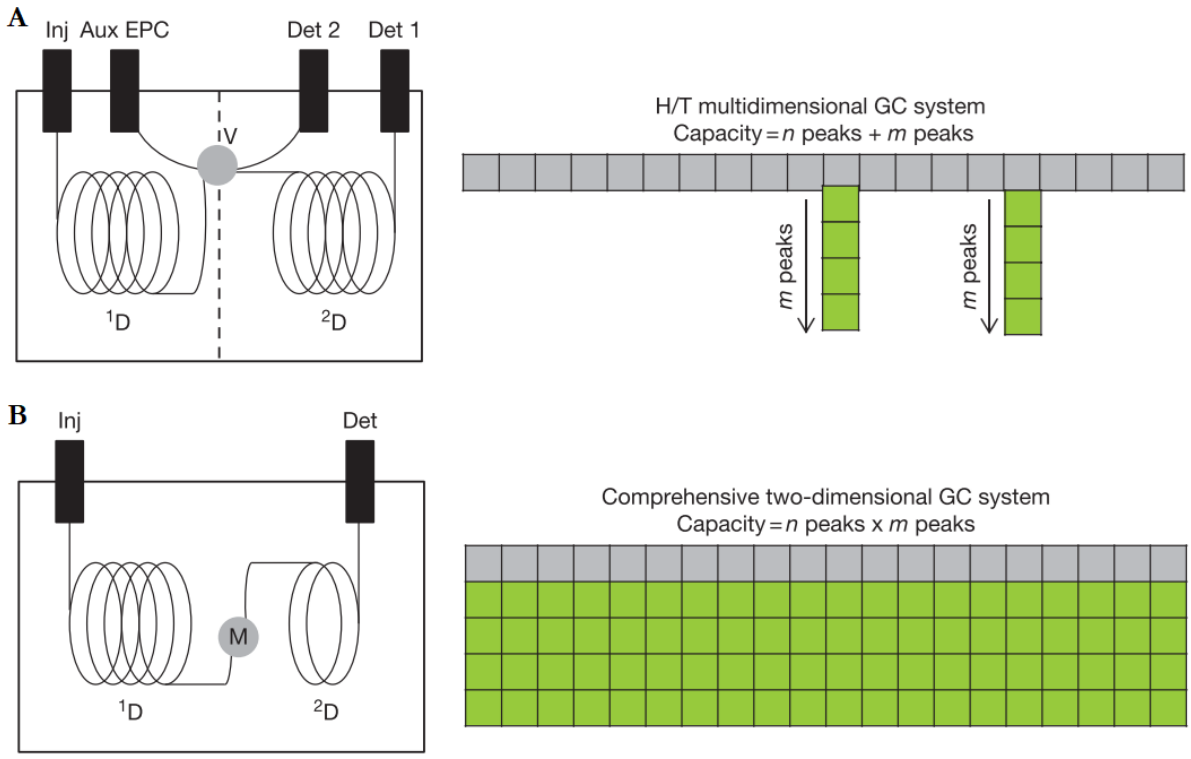

Figure 2.4. Schematic visualization of multidimensional GC instrumental configurations and their actual peak capacity: **2.4A** shows H/C MDGC system, while **2.4B** shows a GC×GC system. Here the list of abbreviation used: Inj - injector; aux EPC - auxiliary gas electron pressure controller; Det 1 - primary detector; Det 2 - secondary detector; V - switching interface; M - modulator; ^{1}D - first dimension column; ^{2}D - second dimension column; n - ^{1}D peak capacity; m - ^{2}D peak capacity. Adapted from[14]

2.2.2 Heart-cut

H/C MDGC is based on a heart-cutting process, which allows the transfer of one or more selected portions of the primary column eluate into the secondary column for additional separation.[15] The interface of a H/C MDGC system (letter *V* in **Figure 2.4A**)allows efficient transfer of fractions from the primary column (^{1}D) into the secondary column (^{2}D). The sample is initially injected into the ^{1}D then, before the elution of the fraction of interest, the interface is switched in the *transfer* position for a fixed time. Once the fraction is completely loaded onto the ^{2}D, the interface is switched back to its original or *idle* state, completing the heart-cutting process.[25]

An important milestone for the development of an efficient H/C procedure was the invention of a switching device named Deans Switch (from the name of the author, David Deans) in 1968, that solved some issues related to the functioning of mechanical valves in coupled-column systems.[15,32] The first H/C system was developed by Schomburg in the 1980s and the first applications were presented for petroleomics.[15,33]

H/C was appreciated for the increased peak capacity and the possibility to resolve challenging separations; however, it has an important limitation, represented by the increased total runtime when the sequential separation has to be applied to multiple cuts. Indeed, having first and second column similar dimensions, even if different stationary phases, the time scale of the secondary separations is approximately equal to that of the primary. It means that each second-dimension

analysis will easily add 30-60 minutes to the total runtime and, if more than one or two "cuts" are necessary the total runtime will be drastically increased.[24,25]

2.2.3 GC×GC

Comprehensive two-dimensional gas chromatography, namely GC×GC, is the natural progression of the H/C MDGC. H/C MDGC is developed through a dedicated time-programmable interface that, automatically and online, transfers selected eluting fractions from the ^{1}D to the ^{2}D; conversely, in GC×GC, each fraction eluting from the ^{1}D is trapped, focused and released into ^{2}D for further separation. The full/complete transfer of the ^{1}D column effluent into the ^{2}D is the reason for the introduction of the term "comprehensive" to describe this technique.[4,14,34,35] The principle is to subject small chromatography bands, potentially containing co-eluted and non-resolved compounds, to an additional separation in a continuous and sequential manner.[36]

The operation of trapping, focusing and re-injecting in the second separative dimension is done by a modulator within a fixed time (usually 2-8 s), which is ideally also the time allowed for the analysis in the ^{2}D.[14,36] It means that GC×GC, thanks to the fast separation in the second dimension, allows to perform a comprehensive multidimensional separation within the time of a 1D-GC separation.[24,29]

In addition to the possibility of keeping reasonable analysis times by combining two-separation dimensions, GC×GC offers a series of additional advantages, if compared to 1D-GC: (a) the effective in-space band-focusing (especially in case of thermal modulation) toward the ^{2}D resulting in an increased overall peak capacity; (b) a signal-to-noise ratio (S/N) increase, resulting in a sensitivity gain of one order of magnitude (especially in case of thermal modulation); and (c) formation and identification of group-type patterns based on homologous series of compounds that have a retention logic over the 2D separation space.[36,37]

The augmented informative potential of a GC×GC analysis, in addition to its effective total analysis time if compared to the H/C MDGC, brought to a remarkable increase of applications and published research based on this technique over the years. **Figure 2.5** shows the number of

publications per year (years 1991-2020) recorded on Scopus database using the keywords: "comprehensive two-dimensional gas chromatography".

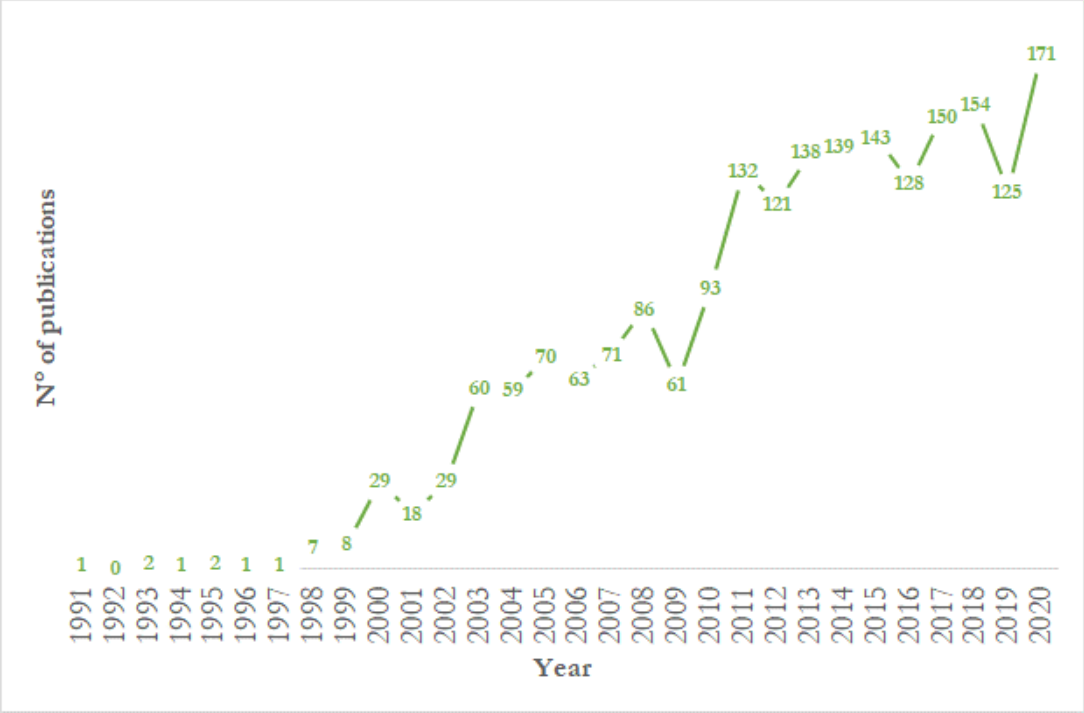

Figure 2.5. Number of publications per year recorded on Scopus database with the keywords: "comprehensive two-dimensional gas chromatography", years 1991-2020.

During the 1990s applications were mainly focused on petroleomics Flame Ionization Detection (FID) as elective detection technique. In 1999 appeared the first paper reporting the adoption of Mass Spectrometry (MS) and, since that time, the powerful alliance of comprehensive two-dimensional gas chromatography with mass spectrometry (GC×GC-MS) allowed to address the needs of many different application areas.[34,36,38] In 2002 the first application in the food area was by Shellie and Marriott, studying enantiomer separations in a bergamot essential oil. At present, GC×GC-MS is used in several fields: petrochemicals and fuels; food, flavor and fragrance; forensic; environmental; biological; metabolomics and volatile organic compounds (VOCs) profiling.[15,36,39]

Method optimization in GC×GC is generally addressed to maximize method separation power and sensitivity.[20] The main parameters that impacts on them involve the modulation step (*i.e.,* selection of the appropriate modulator and P_M), columns configuration (*i.e.,* the choice of suitable columns' dimensions and combination of stationary phases), carrier gas flow, temperature programming and detector settings.[20,29]

In the next sections, relevant aspects related to the column selection, the modulation step and the detectors are further investigated.

2.3 Columns configuration

In GC×GC, the effluent of the first column is focused in narrow fractions at regular and short time intervals, and then injected into the secondary column.[4,40] Therefore, the resulting chromatogram is composed by two axes corresponding to the retention time of each of the two columns and, being each peak eluting from the first dimension "cut" into at least four or five

fractions, the separation of the fractions of the second column must be completed within a few seconds. **Figure 2.6** shows an example of a second dimension chromatogram resulting from the cut of a first dimension peak in GC×GC.[4]

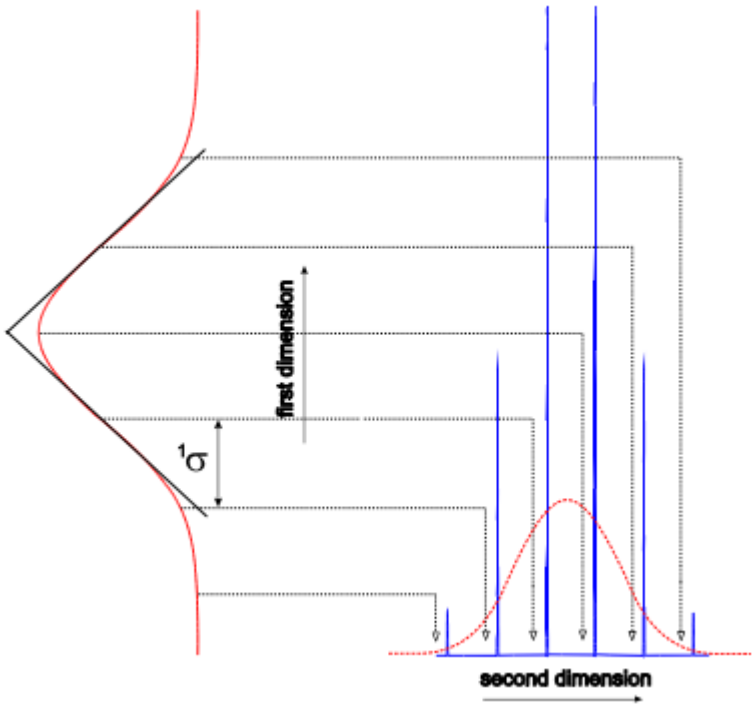

Figure 2.6. Cut of a first dimension peak in six fractions into the second column in a GC×GC system.[4]

To ensure a fast separation in the ^{2}D, the column is usually a short segment, while the primary column is usually a 15-60 m GC column with conventional internal diameter (ID) (0.25 – 0.32 mm ID).[34] Moreover, matching for an appropriate orthogonality of the separative process, the two columns are generally characterized by different stationary phases, that typically consist of a non-polar ^{1}D combined to a mid-polar/polar in the ^{2}D.[20] Their operating temperatures (*i.e.*, minimum and maximum temperatures of usage) often vary from each other as well as the relative retention of analytes, therefore it is a preferable option to place them in two separate GC ovens (*i.e.,* primary and secondary oven) that enable temperature off-set along the chromatographic run.[35,41]

2.3.1 Columns' dimension

As anticipated in the introduction, ^{1}D and ^{2}D columns are characterized by different dimension: the ^{1}D column is a typical GC "normal-bore" usually long 15–60 m, with 0.25 mm ID and a film thickness of 0.18–1.0 μm, while the ^{2}D column is shorter and often with a narrow-bore, typically 0.5–1.5 m × 0.1–0.25 mm ID, and a film thickness of 0.1–0.25 μm.[4,34,42] The different dimensions affect the n_C of the two columns, with the ^{2}D column that may have a modest n_C due to its short length and/or very fast elution: approximate typical values of n_C might be 500 and 20 for the ^{1}D and ^{2}D column, respectively.[43]

This design is necessary to ensure, as in any hyphenated technique, that the ^{2}D works at much higher speed than the ^{1}D, to produce a series of many hundreds of fast analyses (2-8 s), during the period of one ^{1}D analysis.[34,40] The optimum relative speed of the two columns must generate an

appropriate number of secondary chromatograms during the elution of each peak from the ^{1}D: a too slow relative ^{2}D separation degrades the ^{1}D separation, as there are not enough analyzed samples (at least 3-4) across each peak to preserve the resolution obtained; conversely, a too fast relative ^{2}D separation results in a low resolution in the ^{2}D.[40]

Consequently, a typical choice in many GC×GC systems is to slow down the ^{1}D separation, with an oven temperature rate in the order of 1-4°C/min, to provide ^{1}D peaks having a width of tens of second (*i.e.,* 30-40 s).[4,35] It means that, the second column operates very rapidly in comparison with the first column and in comparison with the temperature gradient in the column oven: the result is that the temperature of the second column increases only of a few tens of a degree per separation, and thus ^{2}D separations are basically isothermal separations.[4]

However, the result of the optimization of the relative speed among the two columns is that each single column works in conditions different from the optimal, and especially the ^{2}D works at a linear velocity that is far from the ideal, with column overloading that could be an issue. To reduce this risk many strategies have been tested and adopted.[4,35,42] Broader ^{1}D peaks help to produce smaller fractions of the analytes in the ^{2}D, reducing the chance of ^{2}D overloading. One possibility to generate broader peaks, in addition to a slow temperature rate, is to increase the film thickness in the ^{1}D. 1 μm film thickness over 0.25-0.32 mm ID, provide optimal resolution in both dimensions, but at the expense of a much longer analysis time, while 0.1 μm over 0.25-0.32 mm ID is inadequate for GC×GC separations. Intermediate phase ratios values can be considered good compromises.[4] Another possibility is to split the ^{1}D effluent in two parallel secondary columns behind the modulation step. It was first described by Seeley, and the dual secondary columns GC×GC (GC×2GC) was demonstrated to be effective to increase the resolution of the two-dimensional separation.[44] Harynuk *et al.*, instead, studied the influence the ^{2}D column ID on peak-width, concluding that narrow-bore columns in the ^{2}D might not provide as great advantages as is commonly thought.[45] When high chromatographic resolution is desired, a thicker film ^{1}D column coupled to a larger diameter (*i.e.,* 0.18 mm) ^{2}D column could be a better combination. Similar results were obtained by Cordero *et al.*, testing a series of OV1-OV1701 column sets, in which the two dimensions differ in ID and/or film thickness. Results showed that 0.25 mm homologous ID column combination, reduced ^{2}D column overloading effects due to the increased ^{2}D mass loadability, thus facilitating the analysis of mixtures whose components differ significantly in relative abundance.[46] Conversely, if the goal is to prioritize the speed of the separation at the cost of chromatographic resolution, a standard 0.25 mm ID ^{1}D column can be coupled to a 0.05 mm ID ^{2}D column, as applied by Adahchour *et al.*[47] Other expedients to optimize linear velocity in GC×GC separations are: the reduction of the head pressure that, however, results in a loss of resolution in the ^{1}D column, and the use of longer ^{2}D column, requiring an increase in the modulation time.[31,35,48]

2.3.2 Orthogonality and stationary phases

As anticipated in the introduction of this chapter, the separations performed in the two dimensions are properly tuned matching for the orthogonality of the system; it means that the separation mechanisms of the two columns must produce a low-degree of retention correlation in the two dimensions, to avoid the risk of a one-dimensional separation with peaks distributed along the diagonal. In particular, the orthogonality is achieved by varying the retention of the ^{2}D as a function of the ^{1}D separation.[4,40]

Figure 2.7 shows an example of ideal orthogonal separation where two uncorrelated mechanisms, volatility and polarity, are combined. Graphically it is represented by a Cartesian plane with perpendicular axes corresponding to the two separation principles.[4] To note, this situation is

not realistic since in any GC separation the relative volatility is driving components separation/discrimination. Moreover, as discussed by Marriott *et al.*, stationary phases selectivity and system orthogonality can be tuned by playing with columns dimensions and temperature programming even by adopting the same stationary phase in both analytical dimensions.[49]

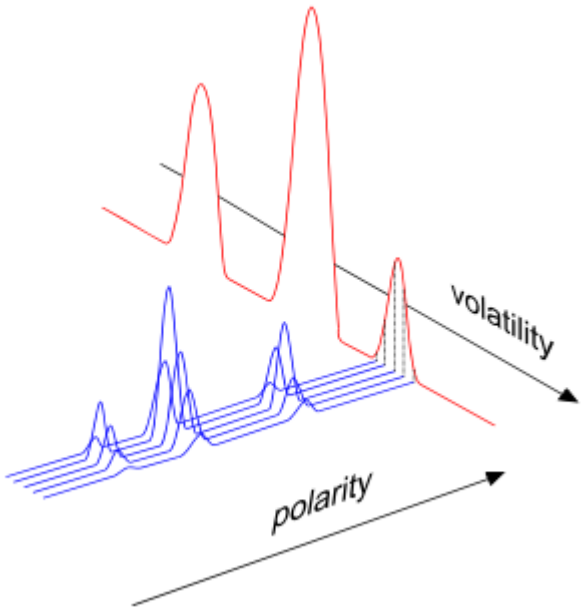

Figure 2.7. Orthogonality of GC×GC because of the application of two different separative principles.[4]

Moreover, system orthogonality enables the generation of structured elution patterns as a function of the differential selectivity of the two separation dimensions. When properly combined, homologous series result distributed over the separation space with a stringent logic and group-type analysis is possible.[42,50] It can be easily explained by following the visual example in **Figure 2.8**, inspired to the concept of sample dimensionality introduced by Giddings in 1995.[19] A hypothetical sample whose components differ in shape, color and size has a dimensionality of three, and there is no chance of separating all of them using a 1D system. However, if the proper combination of orthogonal separative principles (*i.e.,* size and shape in this case) is applied, not only the complete resolution of all components is achieved by using the total separation space, but also components are ordered on the ^{2}D depending on their shape.[19]

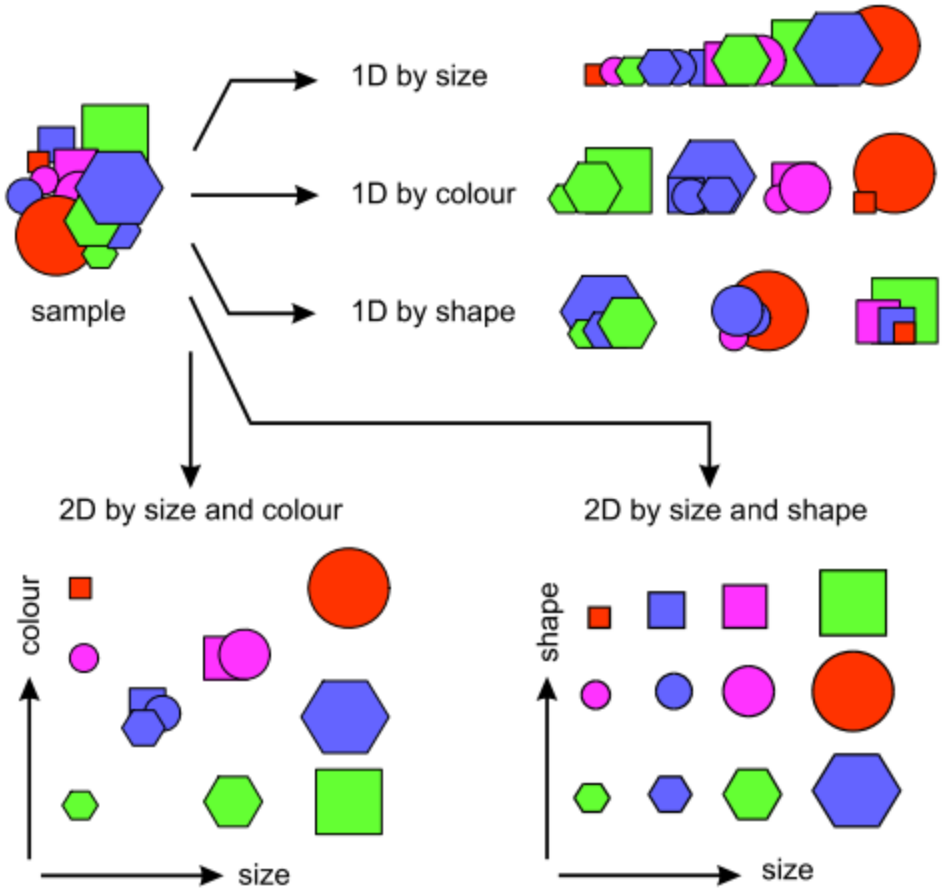

Figure 2.8. Sample separation through two orthogonal separation systems.[4]

It implies that, in GC×GC, components' identification is potentially more reliable being each compound described by two retention times, and the formation of structured patterns in the two dimension is a further identification source.[4]

In principle, all kinds of stationary phases can be used in the ^{1}D of a GC×GC system; however, the most common column set is a non-polar/low-polarity column as ^{1}D (*e.g.,* 100% dimethyl polysiloxane, or 5% diphenyl/95% dimethyl polysiloxane) combined with a medium-polarity or polar column as ^{2}D (*e.g.,* polyethylene glycol, or 50% phenyl/50% methyl polysiloxane).[20,29,35,51] This configuration provides a separation of the analytes in the ^{1}D based on their vapor pressures; while the relative retention in the ^{2}D is mostly related to the presence of polar functions.[20,42]

However, in several applications, the polar × apolar combination is particularly effective. Flavours and fragrances, with a predominance of polar analytes, as well as some petroleum cuts take benefits from a "reversed" stationary phase combination[20,42]

Figure 2.9 summarizes the most used column combinations for GC×GC as it results from published papers until 2018: as anticipated, the "apolar × polar configuration is the most common.[42]

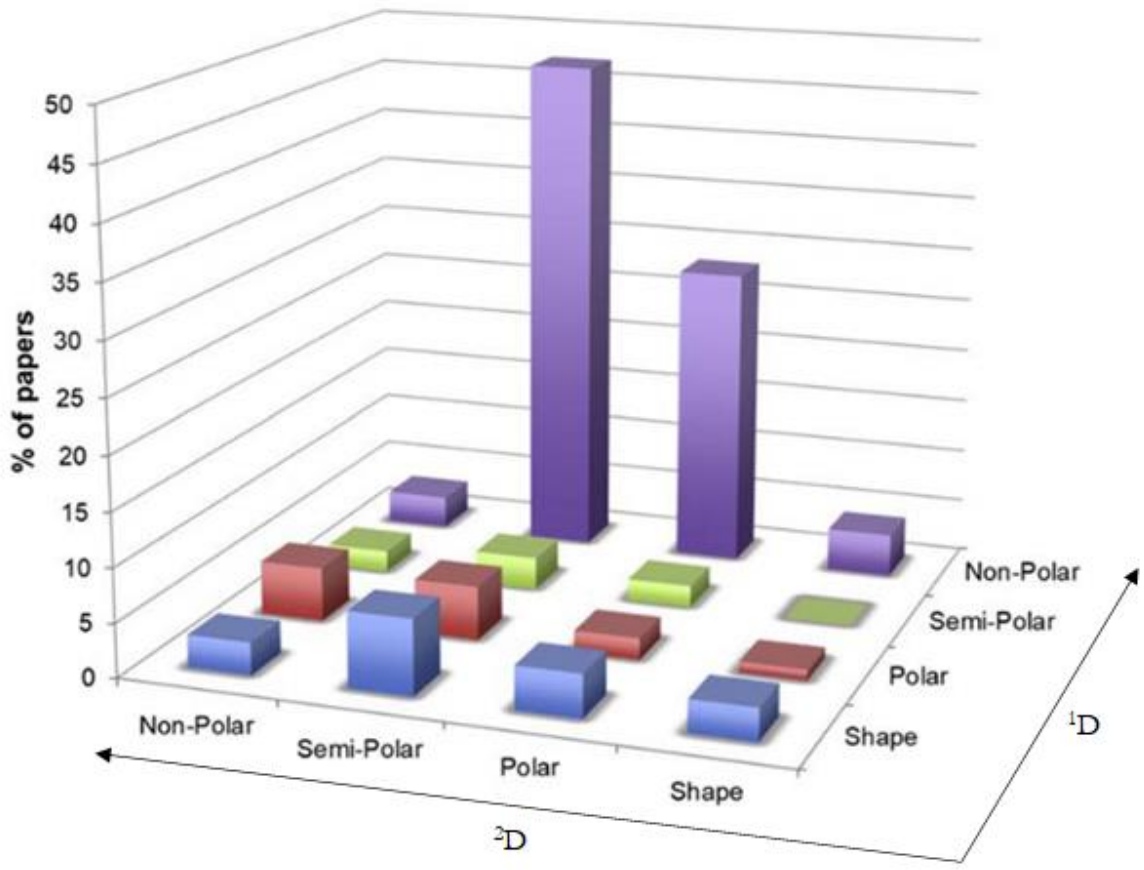

Figure 2.9. Relative proportions of column combinations used in GC×GC published papers until 2018.[42]

2.4 Modulation and modulators

2.4.1 Modulation process and visualization

The modulation is the key-aspect of GC×GC process.[43] The term modulation was adapted from the telecommunication industry, where audio and video signals' frequency can be modulated to facilitate their transmission over long distances to the signal receiver. In chromatography, as in telecommunication industry, the modulation of the signal does not carry new or extra information, however it led to the possibility of a different elaboration of the raw signal. **Figure 2.10** shows how the mono dimensional chemical signal, when combined to the injection frequency corresponding to the P_M, produces in the ^{2}D a composite chemical signal with the form of the modulation (*i.e.,* composed by a series of "pulsed" peaks, or "subpeaks", or "peaklets") but containing the information of the analytical signal.[15,43]

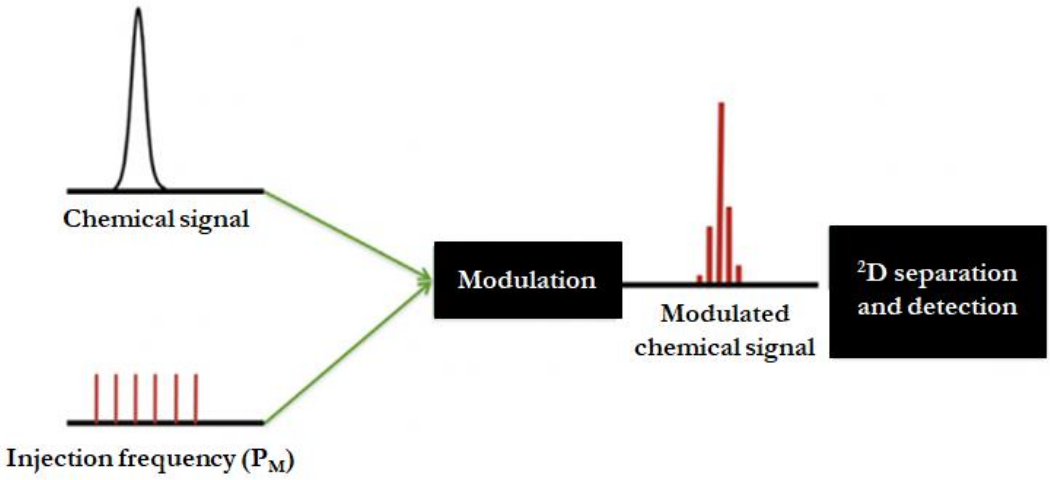

Figure 2.10. Transformation of the mono-dimensional chemical signal in a modulated signal. Adapted from[15]

The modulated chromatogram is so divided into sequential modulation events, each of a given P_M.[43] The appropriate resulting visualization is the next step: it is implemented through rasterization by arranging data values acquired during a single modulation cycle as a column of pixels, so that the ordinate corresponds to the elapsed time for the ^{2}D separation, while the abscissa axes corresponds to the elapsed time for the ^{1}D separation.[14] **Figure 2.11A** shows an example of two coeluted peaks modulated with a P_M of 4 s. Retention times in both dimensions are then calculated by peak integration for each modulate peak (**Figure 2.11B)**. In particular, ^1t_R predicted as the apex position of the fitted peak, while the intensity is calculated as sum of all modulated sub-peaks intensities. At this point, a 3D plot is obtained by changing the position of the successive modulation from subsequent to side-by-side, by placing each modulated peak in the Cartesian coordinate system defined by ^1t_R and ^2t_R (**Figure 2.11C)**. The intensity is represented by the peak height however, in the common top-down contour plot visualization, height is substituted by a color-scale.[14,43]

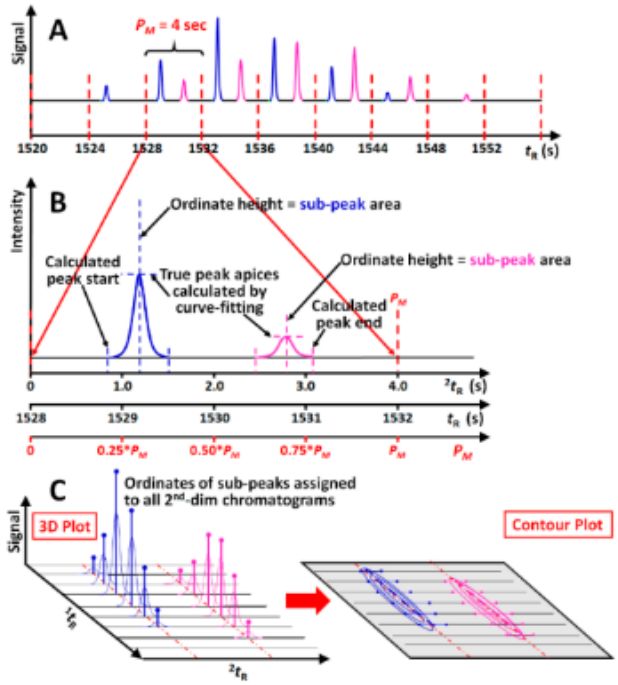

Figure 2.11. (**2.11A**) An example of modulated chromatogram containing coeluting peaks modulated with a P_M of 4s; (**Figure 2.11B**) Calculation of the retention times in both dimensions are then calculated by peak integration for each modulate peak; (**Figure 2.11C**) 3D plot and Contour plot visualization.[43]

The modulation process is characterized by the P_M. To maintain the separation produced in the ^{1}D, the effluent must be sampled frequently enough to prevent the coelution of already separated components. As suggested by Murphy *et al.*, each chromatographic peak should be sampled 3-to-4 times to meet this condition.[52,53] The choice of a correct P_M, is fundamental to obtain a proper ^{2}D separation, however this parameter can be better described by the modulation ratio (M_R), defining the effective number of times that the chosen P_M will modulate a given ^{1}D peak.[27] It is calculated as indicated in **Equation 2.2**, where the peak width at the base (w_b), or 1.6985 times the peak width at half height of the peak (w_h), is divided by the P_M.

Equation 2.2

$$M_R = \frac{4\sigma}{P_M} = \frac{w_b}{P_M} = \frac{w_h \times 1.6985}{P_M}$$

Moreover, in case of modulation performed through low-duty cycle modulators (see **Chapter 2.4.2**), M_R value can be used as indicator to guide the choice of the correct P_M: a value of at least 3 should be used for quantitative measurements of trace compounds, while for qualitative analysis a value of about 1.5 is considerate adequate.[27]

The correct reconstruction and the degree of resolution of a peak depends on the phase of modulation. This is defined as the difference between the center of the ^{1}D peak and the mean of the peak region sampled by the modulator.[43,53] The two limiting scenarios, represented in **Figure 2.12**, are respectively in phase (**Figure 2.12A**), having the single maximum peak centered at the peak apex of the ^{1}D peak, and 180° out of the phase modulation (**Figure 2.12B**), where two equally tall symmetric maxima are equidistant from the ^{1}D peak maximum. **Figure 2.12C**, instead, represents a phase between the two limiting cases.[53]

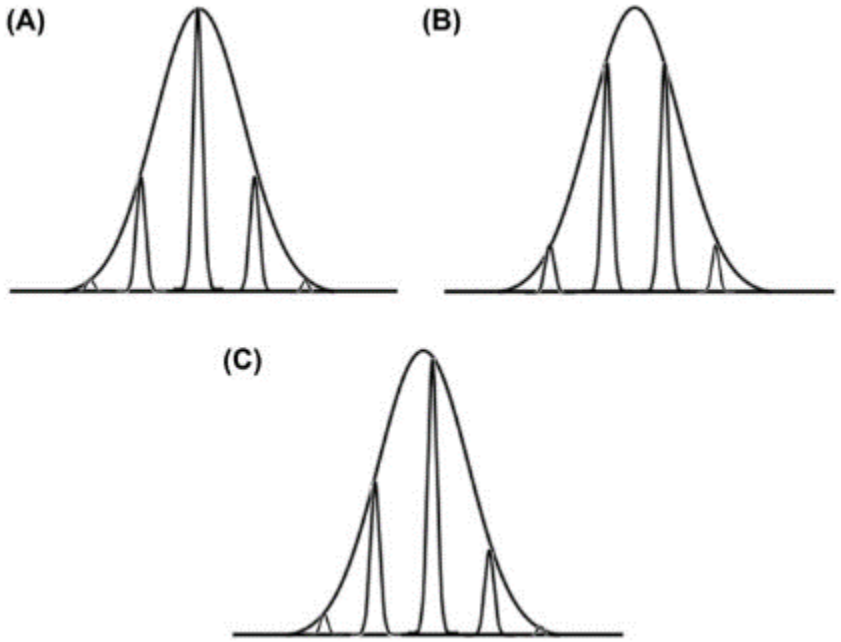

Figure 2.12. Different phases of a GC×GC peak pulse: (**2.12A**) in-phase modulation, (**2.12B**) 180° out-of-phase modulation and (**2.12C**) an intermediate phase modulation.[53]

Finally, the peak wrap-around is a phenomenon occurring when the ^{2}D retention of an analyte is longer than the P_M.[27,53] It happens when an analyte has a stronger retention in the ^{2}D column, eluting in the successive P_M (**Figure 2.13A).** Peak wrap-around leads to potential coelutions between components belonging to subsequent fractions and can disturb the structure of the chromatogram, making method development and data interpretation more complicate (**Figure 2.13B**).[20,28,53]

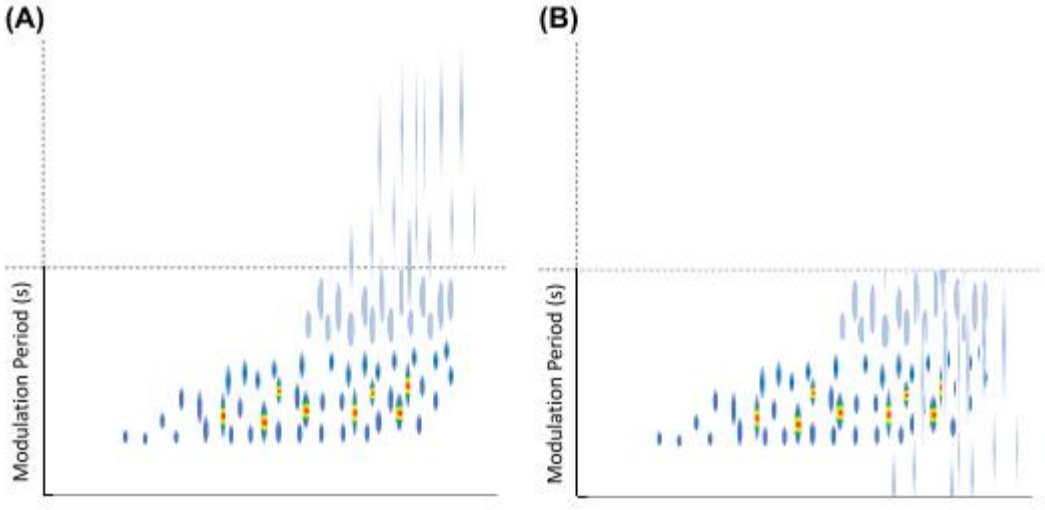

Figure 2.13. Example of peaks wraparound. (**2.13A**) Analytes exceed the duration of a single modulation period (indicated by the dotted line) and (**2.13B**) they elute overlapping with separated peaks in the subsequent modulation period.

2.4.2 Modulators

As previously anticipated, the modulator is the core of a GC×GC system being responsible for the correct transferring of the analyte from the ^{1}D column to the ^{2}D column, through its action of cut, re-concentration, and re-injection of portions of ^{1}D column effluent into the ^{2}D column in a sequential and continuous way.[54,55]

A modulator is required to create very narrow chromatographic bands during the re-injection into the ^{2}D to preserve peak capacity and improve method sensitivity. It should be compatible, in terms of efficiency, to the MW range of the analytes under study, should be robust, and cost-effective.[54–56] The first modulator adopted for GC×GC was that proposed by Liu and Phillips in 1991.[30] It consisted of a 15 cm segment of gold-coated thick-film capillary column, looped outside the GC oven and position between the ^{1}D column and the ^{2}D column; the modulation was produced by periodic resistive heating of the gold-painted trap.[20] Since its invention a wide variety of other modulators have been developed and, nowadays, they can be classified in two main groups based on the principle used to obtain the modulation:

- Thermal modulators, that use a positive and/or negative temperature difference, compared to GC oven, to transfer fractions from the ^{1}D to the ^{2}D.[14,34] Thermal modulators can be further subdivided into: (a) heater-based, collecting analytes bands eluting from the ^{1}D at the oven temperature (or slightly below it) and releasing them through an increase in temperature; and (b) cooling-based, usually cryogenic, collecting analytes at very low temperatures obtained through cryogens (usually liquid N_2 or CO_2) and releasing them at the oven temperature (or above it).[53] Thermal modulators are characterized by a duty cicle of 1, it means that they completely transfers all of the ^{1}D eluate into the ^{2}D column, without losses.[14,54]
- Pneumatic modulators, adopt valves connected either in-line or out-line through a sample loop (or channel) with the column set, for band transfer.[14,29,34,56] Pneumatic modulators can be further subdivided into valve-based modulators, where the flow rates for the ^{1}D and the ^{2}D are not coupled, and flow-based modulators, where the two columns flows are coupled.[29] Pneumatic modulators can have a duty cycle ≤ 1, depending on their design. When the duty cycle is ≤ 0.5, they are named low duty cycle modulators.[54]

2.4.3 Thermal modulators

2.4.3.1 Heater-based modulators

Heater-based modulators, as mentioned above, trap primary column effluent at ambient temperature (or slightly below it) usually with the help of thick stationary phases, while active heating is used for the rapid desorption of the trapped analytes.[2,35,53]

2.4.3.1.1 Thermal desorption modulator (TDM)

The first modulator, the *thermal desorption modulator (TDM)* introduced by Liu and Phillips in 1991,[30] used a 15 cm segment thick-film capillary column externally gold-coated and allowing for electrical heating. It was originally developed as a single-stage modulator, but it rapidly evolved in a dual-stage modulator to solve the problem band broadening and analytes breakthrough.[2,35] **Figure 2.14** shows how it worked the dual-stage modulation:

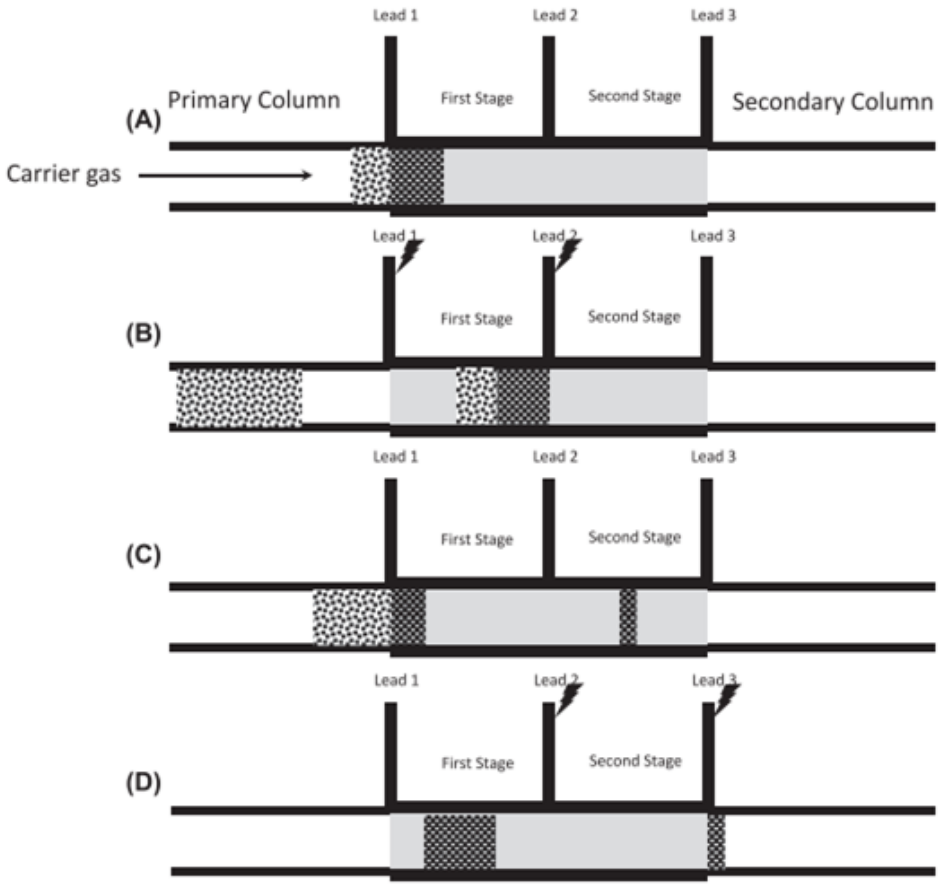

Figure 2.14. (**2.14A**) Trapping and focusing of analytes eluting from the ^{1}D column. (**2.14B-C**) Remobilization of the analytes and trapping in the second stage, while the next fraction of analytes reached the first stage of modulation. (**2.14D**) Re-injection in the ^{2}D column.[53]

Analytes eluting from the ^{1}D column were trapper and focused by the stationary phase coating of the ^{2}D column (**Figure 2.14A**). At this point, the direct resistive heating through a 20-ms electrical pulse of the first stage of the modulation process remobilized the analytes (**Figure 2.14B**), which were then trapped and focused in the second stage while first step cooled down to trap the next fraction of analytes (**Figure 2.14C**). The second stage was finally resistively heated and trapped analytes were injected into the ^{2}D column (**Figure 2.14D**).[53,55] The use of a dual-stage modulation allowed the injection of the analytes in the ^{2}D column in narrow band, if compared to those produced by a single-stage modulation, rapidly abandoned. However, this first modulator presented many drawbacks, and the design was not robust enough for routine applications, due to frequent capillary burnouts and the delicate nature of the thin conductive film.[35,53]

2.4.3.1.2 Rotating thermal modulator (RTM) – "sweeper"

The first commercially available device was the *rotating thermal modulator (RTM),* or *"sweeper",* introduced by Zoex (Zoex Corporation, Lincoln, NE, USA). It was theoretically described by Ledford and Phillips in 1996, but the final version was presented in 1999, after over 2 years of development and optimization.[57] The *RTM*, commonly referred to as the *"sweeper"*, worked in 4 steps (*i.e.,* accumulate, cut, focus and launch) by using a slotted heater rotating around a shaft to heat the modulator capillary, as described in **Figure 2.15**. [2,53,55]

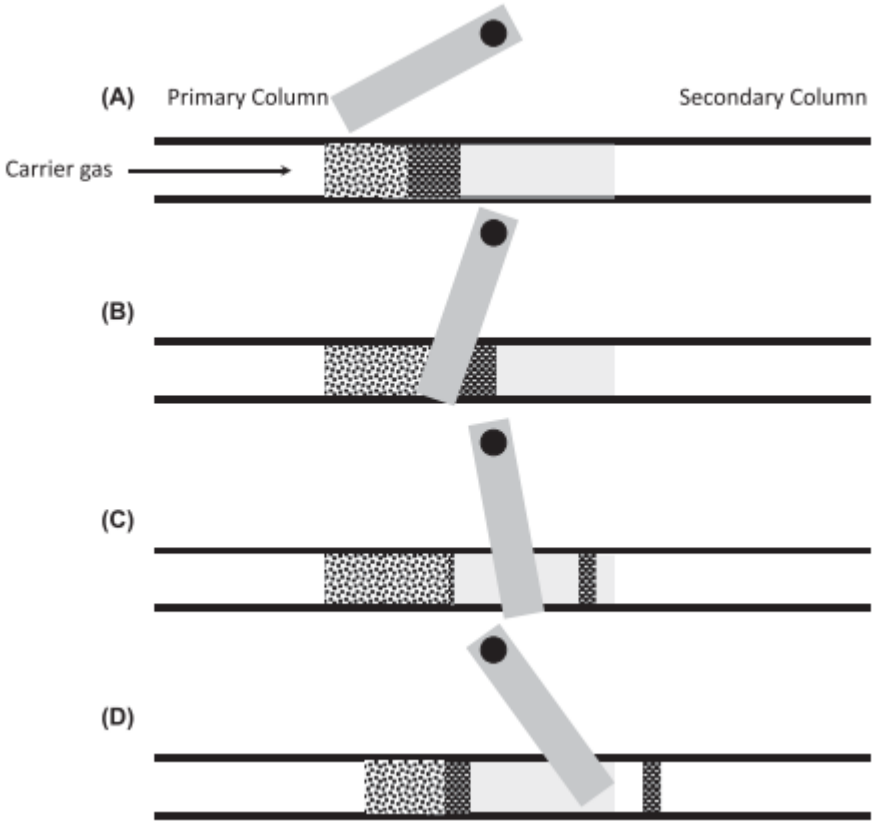

Figure 2.15. Modulation process of the sweeper in 4 main steps: accumulation (**2.15A**), cut (**2.15B**), focus (**2.15C**), re-injection (**2.15D**).[53]

Analytes from the ^{1}D column accumulate at the head of the modulator capillary coated with a thick film of the stationary phase (**Figure 2.15A**) until the rotation of the slotted heater remobilize analytes trapped in the upstream portion of the modulator (**Figure 2.15B**). Remobilized analytes are then refocused in the downstream portion of the modulator (**Figure 2.15C**) and finally injected into the ^{2}D column (**Figure 2.15D**). To produce sufficiently narrow modulated peaks, it is required a temperature difference of at least 100°C between the heater and the modulator capillary: this limited the maximum GC oven temperature thus reducing the volatility range of compounds that could be modulated.[2,53]

2.4.3.2 Cooling-based modulators

Cooling based modulators are the most common thermal modulators: they use cold temperatures to trap and focus compounds, that are then released through temperatures equal or above the oven temperature. In particular, temperatures required to trap compounds from C_4 to C_{40} on a deactivated modulator column were determined: to trap and focus any compound in this range, a temperature of 120-140°C lower than its average elution temperature is required while, for their release, temperatures approximately 40°C above the elution temperature are needed.[58]

The focusing is usually operated by using cryogens (*i.e.,* liquid N_2 or CO_2), and the two main design are developed through a longitude movable trap or a jet trap. Although the use of cryogens added a consumable cost to the system, it provided the best performance and overcome the temperature limitations experienced using heater-based thermal modulators.[53] Band focusing is operated between ^{1}D column and ^{2}D column, and it can be done on a portion of the ^{1}D column,

in a deactivated column of suitable inner diameter (usually a narrow bore-capillary 0.10 mm ID), or at the head of the ^{2}D column.[14]

2.4.3.2.1 Longitudinal modulated cryogenic system (LMCS)

The *longitudinal modulated cryogenic system (LMCS)* was introduced by Kinghorn and Marriott in 1998, in a short communication describing the separation of a kerosene sample.[2,55,59]

Figure 2.16 shows the scheme of the LMCS using a trap, cooled through a flow of liquid CO_2, that moved longitudinally along the head of the ^{2}D column in less than 10 ms, through a pneumatic arm electrically controlled.[2,53] It works in three main steps: (a) in the first step analytes were trapped at the segment of the column cooled by the trap in the top position; (b) then the longitudinal movement of the trap to the bottom position exposed the cooled segment of the column to the GC oven temperature, allowing the remobilization of the focused band and the successive re-focusing at the second trap position; (c) finally, the trap moved back to the original position, allowing the trapped band to remobilize again entering the ^{2}D column.[53]

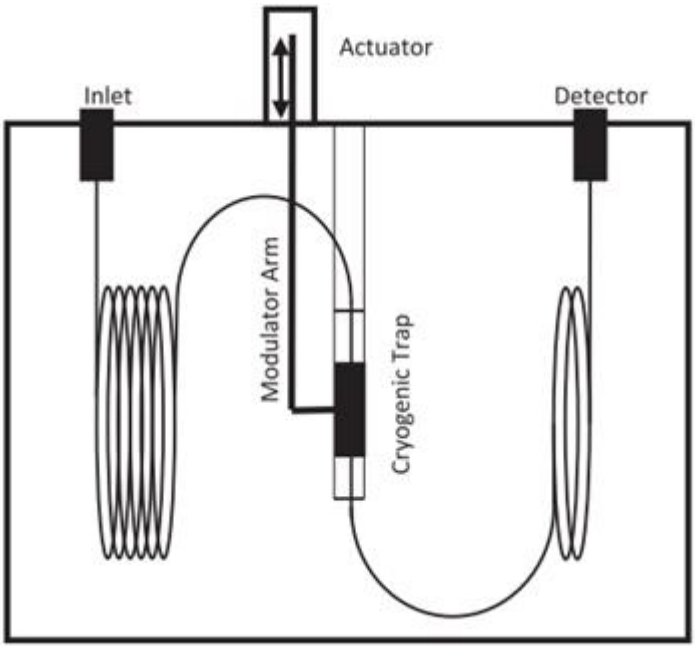

Figure 2.16. Diagram of the longitudinal modulated cryogenic system (LMCS).[53]

The *LCMS* was a straightforward innovation in modulation technology since the *RTM* developed by Phillips.[20] LCMS had its own limitations: very volatile analytes were difficulty trapped, ice buildup in the trap could cause column breakage and the mechanically moving trap is potentially problematic.[20,35] However, the successive introduction of gaseous N_2 as cryogen solved part of these problems and it was estimated that by 2004, cryogenic modulators had likely replaced almost all of the TDMs in use.[2,53]

2.4.3.2.2 Dual-stage quad-jet modulator

In 2000, Ledford introduced a new modulator, based on a static cryogenic modulation, to overcome possible problems arising from the use of mechanically moving parts.[60] The *dual-stage quad-jet modulator* utilized two cold jets cooled through liquid CO_2 (subsequently replaced by N_2) and two hot gas jets, and its operation mode is showed in **Figure 2.17**.[35,56]

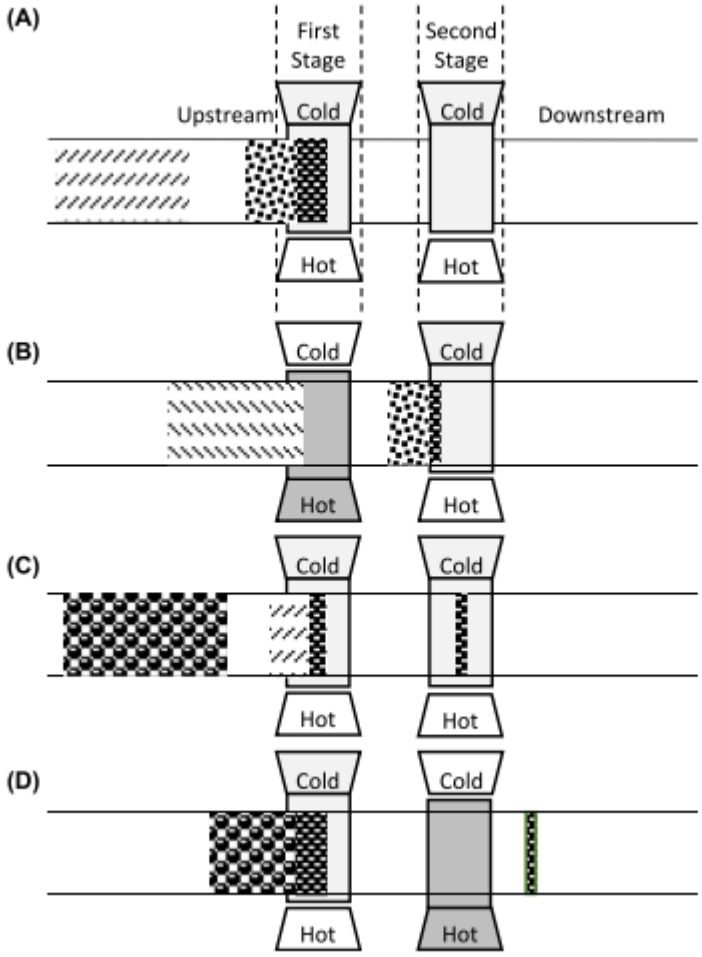

Figure 2.17. Dual-stage quad-jet modulator modulation process. (**2.17A**) Activation of the upstream cold-jet to trap and focus analytes eluting from the ^{1}D column, (**2.17B**) Remobilization of the trapped analytes through the activation of the upstream hot-jet and subsequent focusing by the downstream cold-jet, (**2.17C**) focusing of the next fraction of analytes by the upstream cold-jet, and (**2.17D**) injection of the band in the second stage into the ^{2}D column by activating the downstream hot-jet.[53]

Analytes eluting from the ^{1}D column were trapped and focused through the upstream cold-jet (**Figure 2.17A**), then the upstream hot-jet was pulsed to remobilize the trapped analytes towards the second stage, where they are refocused by the downstream cold-jet to prevent any breakthrough (**Figure 2.17B**).[2,20,53] The upstream cold-jet started again to trap the next fraction of analytes (**Figure 2.17C**), and the refocused band in the second stage was injected into the ^{2}D column by activating the downstream hot-jet, while the upstream cold-jet still stayed on to prevent breakthrough (**Figure 2.17D**).[2,53]

Since all jets are placed close into a small chamber, hot-jet and cold-jet may influence each other, therefore the duration of the heating and cooling periods needs proper optimization. A too long hot-jet period, or a too high temperature, will result in a breakthrough of a part of the trapped fraction or on a thermal degradation of the column's stationary phase and consequent bleeding. On the other hand, when the temperature of the cryogen is too low, or the cold-jet period is too long, the analytes will not be properly remobilized.[4]

The *dual-stage quad-jet modulator* is one of the most appreciated and used design of thermal modulator. It was commercialized by Zoex as the KT2001 modulator, while currently is available from the LECO (LECO Corporation, St. Joseph, MI, USA) in their Pegasus lineup.

2.4.3.2.3 Dual-stage loop modulator

A couple of year later the introduction of the *dual-stage quad-jet modulator*, in 2002 Ledford and coworkers presented the *dual-stage loop modulator*, capable of providing a dual-stage modulation although using a single couple of hot and cold-jets.[20,61]

Figure 2.18 shows how each gas-jet passed through two-segments of a looped capillary. The cold-jet ran continuously creating two cold-spots, in the upstream and in the downstream portion of the loop (**Figure 2.18A**). The hot-jet instead, was activated periodically, diverting the cold-jet from the loop, and heating the cold spots to remobilize the trapped analytes in both spots (**Figure 2.18B**).[29,53]

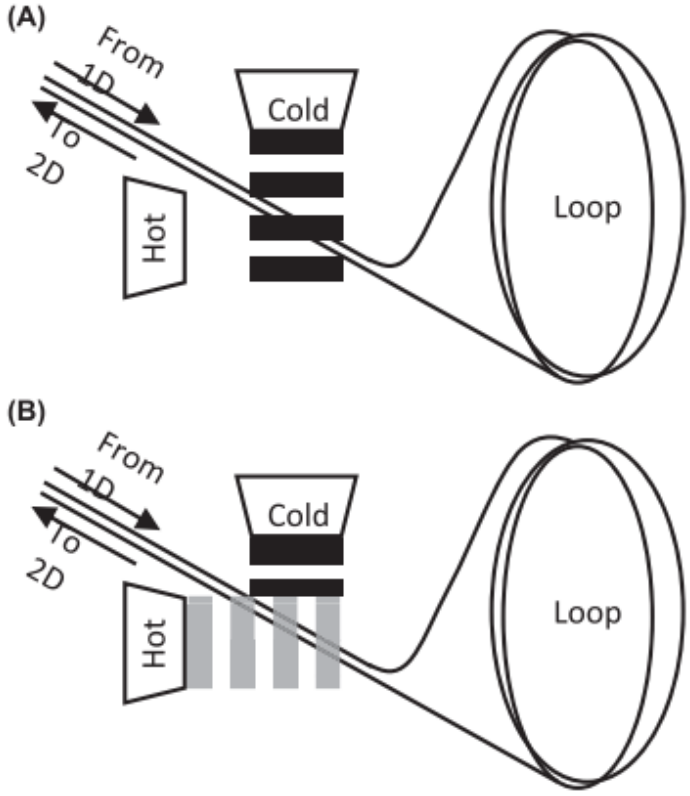

Figure 2.18. Modulation process of the dual-stage loop modulator: (**2.18A**) the cold-jet creates two cold spots in the upstream and downstream portion of the loop and (**2.18B**) the subsequent hot-jet, periodically activated, remobilize the trapped analytes in each section of the column.

It means that analytes eluted from the ^{1}D column are trapped by the cold-jet at the upstream spot and then the activation of the hot-jet remobilizes them into the loop capillary. As the analytes traveled through the loop the hot-jet was deactivated allowing the formation of new cold spots to through the cold-jet: analytes in the loop are trapped in the downstream cold spot and, at the same time, the next fraction of the analytes eluting from the ^{1}D column is trapped in the upstream spot. The subsequent activation of the hot-jet injected the refocused band of analytes from the downstream spot into the ^{2}D column and released the trapped analytes from the upstream cold spot into the loop.[35,53]

As for the *dual-stage quad-jet modulator*, the *dual-stage loop modulator* needs an optimization of the temperatures to avoid breakthrough while matching for an efficient reinjection of the trapped analytes. Moreover, the length of the delay loop requires an optimization, and typically its length is about 1 m. If the loop is too short, the risk is that the band travelling through it might reach the second spot at a time when it is not still sufficiently cold, allowing breakthrough to occur; conversely, multiple injections from the first cold-spot could be present within the loop simultaneously, if the loop is too long.[20] For this reason, a model to optimize the length of a loop capillary in GC × GC systems adopting a *dual-stage loop modulator* was proposed by Harynuk and Gorecki in 2005.[62]

The dual-stage loop modulator is today on of the most used and it is commercially available from Zoex.[53]

2.4.3.2.4 Cryogen-free thermal modulators

Cryogen free thermal modulators were designed with the purpose to eliminate cryogens, and thus reducing costs, without significant compromises to the performance and robustness.

For both the *dual-stage quad-jet modulator* and *dual-stage loop modulator* are available consumable free variants using a chiller instead of liquid N_2 to cool the heat exchanger. However, while cryogen modulators are able to modulate analytes in the range C_4-C_{40}, cryogen-free thermal modulator are unable to trap highly volatile species under C_7 or C_8.[29] Nevertheless, in addition to those variants of cryogenic modulators, original designs for cryogen-free thermal modulators were developed:

- *Single-stage consumable-free modulator*. It was introduced by the Gorecki group in 2015 and utilized a specially coated stainless steel capillary trap that could be passively or actively cooled.[63] Analytes were focused on the trap thanks to two ceramic cooling blocks that were cooled – passively or actively – through a copper heat transfer conduit. Then, the focused band of analytes was injected into the 2D column by resistively heating through a capacitive discharge.[53]
- *Solid-state modulator*. It was introduced from Luong et al in 2016 as *thermal independent modulator (TiM)* and produced by J&X Technologies (J&X Technologies, Shanghai, China).[64] The *TiM* operated outside the GC oven with thermally independent heating and cooling stages: heated areas used micathermic heaters to heat the aluminum chambers from ambient temperature to over 350°C, while cooling zones can be programmed from -50°C to +50°C through a pair of three-stage thermoelectric coolers.[53,64] Dual-stage modulation was obtained by mechanically moving the modulator column back and forth: analytes from the 1D column were trapped in correspondence of the cooled zone and remobilization was accomplished by moving the column toward the heated entry zone. This movement simultaneously exposed the downstream segment of the modulator column to the cooling units, while the newly cooled zone for the second stage of trapping is reached by the remobilized analytes. Finally, moving the column toward the exit zone, trapped analytes are once again exposed to high temperature and injected into the 2D column, with a simultaneous exposition of the upstream segment of the column to the first stage of trapping. The modulation range depends on the modulator column installed (e.g., C_2-C_{14}, C_9-C_{40} etc.), while a replacement of the moving parts has to be done after a defined number of modulations: 100,000 for the column guides and 1,000,000 for the modulator column.[53,64]

2.4.4 Pneumatic modulators

The growing interest for cryogen-free modulators led to the development of modulators based on a different modulation principle. While thermal modulators trap effluent from the ^{1}D column using temperature differentials, pneumatic modulators adopt valves connected either in-line or out-line through a sample loop (or channel) with the column set, for band transfers.[14,29,34,56]

Different designs have been developed and optimized matching for an effectiveness comparable to that achieved by thermal modulators that, to date, are still considered the "golden standard". In particular, two major families of pneumatic modulators exist: valve-based modulators, where the flow rates for the ^{1}D and the ^{2}D are not coupled and that are usually low-duty cycle modulators, and flow-based modulators, where the two columns' flows are coupled and the full (or almost full) transfer of the ^{1}D effluent into the ^{2}D column results in a duty cycle = 1 (or close to 1).[14,29,34,53,56] However, in order to achieve a full transfer modulation, high ^{2}D column flows are necessary, leading to an increased difficulty of use when coupled to under vacuum detectors, such as mass spectrometers. [53]

2.4.4.1 Diaphragm valve modulator (DVM)

The *diaphragm valve modulator (DVM)* was the first pneumatic modulator, particularly a valve-based one, successfully applied to GC×GC.[2] It was described by Bruckner et al in 1998 as alternative to thermal modulators for the analysis of VOCs.[65]

The DVM, shown in **Figure 2.19**, utilized four ports of a gas-actuated diaphragm valve with low dead-volume fittings as interface between the two columns within the oven; gas supply, instead, was by a three-port solenoid valve. The diaphragm valve would be activated twice per second for a very short time, allowing a small portion of the ^{1}D column effluent to enter the ^{2}D column (**Figure 2.19A**) then, while the separation was occurring, all residual effluent from the ^{1}D column would be vented to the atmosphere (**Figure 2.19B**).[53] It means that this modulator is not mass conservative and it is characterized by a low duty cycle, with only a 2-10% of the effluent from the ^{1}D column that is transferred to the ^{2}D column.[14,35] Relatively narrow ^{2}D peak widths were obtained with interesting results in term of robustness, with an excellent $^{2}t_R$ reproducibility.[29] However, because of the significant loss in mass transfer, this modulator is less sensitive than a thermal modulator and is not suitable for trace analysis. Moreover, the valve could be operated at a maximum temperature of 175°C, restricting analysis to very volatiles VOCs and thus limiting the modulator applicability.[2,53]

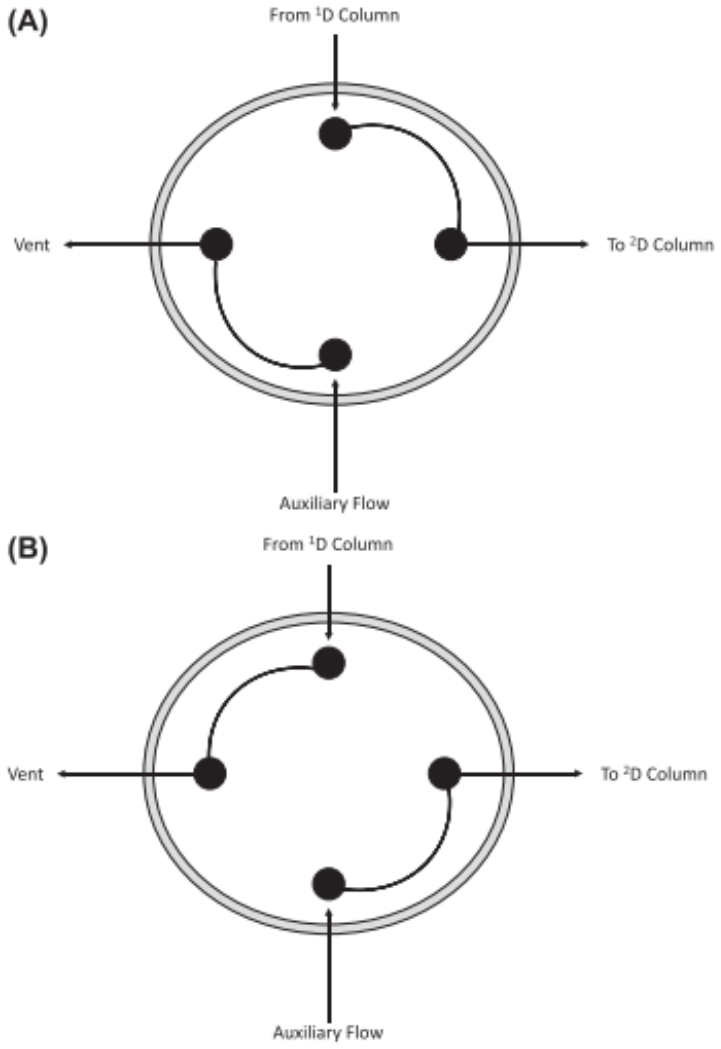

Figure 2.19. Scheme of the DVM showing (**2.19A**) the primary column effluent entering the ^{2}D column and (**2.19B**) the primary effluent being vented to the atmosphere during the second separation.[53]

This first design of *DVM* for modulation in GC×GC has continued to improve in the next few years mainly by Synovec's group, with two aims: (a) the improvement of a higher duty cycle to enhance detector sensitivity, and (b) the overcoming of the temperature limits.

The first goal was matched with the work of Mohler et al in 2006, which introduced a total-transfer valve based modulator bringing the fraction of the ^{1}D column effluent transferred to the ^{2}D column from 5-10% to 100%.[66] As shown in **Figure 2.20**, a high-speed six-port diaphragm valve was used and a sample loop implemented, based on the original idea by Seeley developed on *differential flow modulator (DFM)* in 2000 (see **Chapter 2.4.4.3** for details).

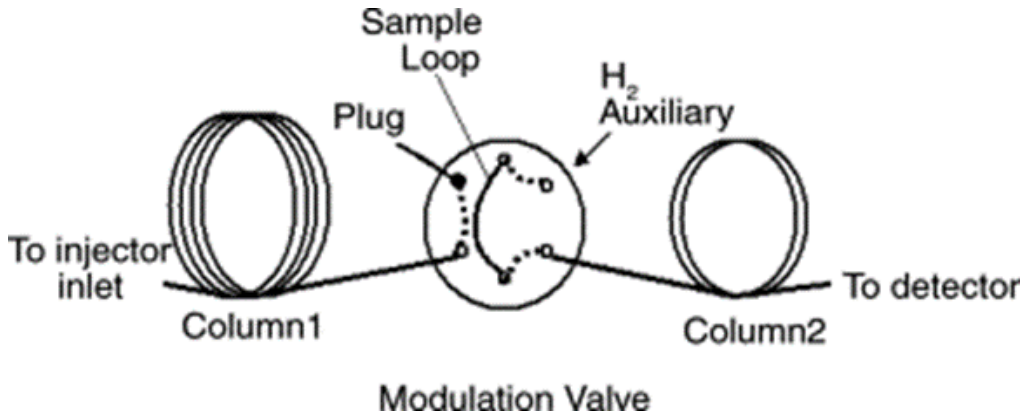

Figure 2.20 Diagram of the total-transfer diaphragm valve-based GC×GC with the dotted lines showing the port connection in the column 2 inject mode.[66]

An early solution to temperature limit, instead, was to have the valve face mounted external to the oven: this approach extended the maximum temperature of usage to 250°C, enabling application to higher boiling point analytes.[67] However, a successive technical solution was found by Freye et al.,[68] in 2015, by replacing the temperature sensitive O-ring with a perfluoroelastomer-based O-ring, allowing reliable function up to 325°C. The functioning of the updated 6-port modulator was investigated, showing the narrow and reproducible ^2D peak widths were produced, with a high reproducible ^2t_R and a detection sensitivity 8-fold higher than 1D-GC due to zone compression.[68]

2.4.4.2 Flow-diversion or Microfluidic Deans Switch modulator

The *flow-diversion modulator* was introduced by Seeley in 2007.[69] As the first design of *DVM*, it was a low-duty cycle modulator sampling the ^1D effluent for only a small portion of the P_M. However, Seeley here replaced multiport valves that cannot operate at elevated temperatures with a *Deans Switch* - used for decades in H/C systems - as a *flow diversion modulator*, that have been successfully tested at temperatures as high as 350°C. Nevertheless, it must be considered that the time scale for H/C is about 1 order of magnitude greater than the time scale required for successful GC×GC modulation, thus *Deans* Switch should not be used under high-speed separation conditions, when high precision is desired. Indeed, it has been reported that the RSD of the fraction of material transferred from the ^1D to the ^2D is less than 1% if the M_R is greater than 2.5, while increases rapidly when the M_R is decreased below 2.5.[69]

The modulator was created by implementing a three-port solenoid valve, placed outside the oven, with an Agilent microfluidic Deans Switch five port manifold, placed inside the oven. **Figure 2.21** illustrate the functioning of the modulator: the two-state fluidic device directed the ^1D effluent to the ^2D column of a flow restrictor, which was a deactivated fused silica column with the same dimensions as that of the ^2D column. The state of the modulator is then determined by the direction of the auxiliary carrier gas – whose flow was significantly greater than ^1D column flow – provided by the solenoid valve: (**Figure 2.21A**) in the "bypass" state the ^1D column effluent was directed to the flow restrictor, while (**Figure 2.21B**) in the "inject" state, the ^1D column passed onto the ^2D column.[53,69] Modulation continued by alternating between the two states of the Deans Switch throughout the entire chromatographic run.

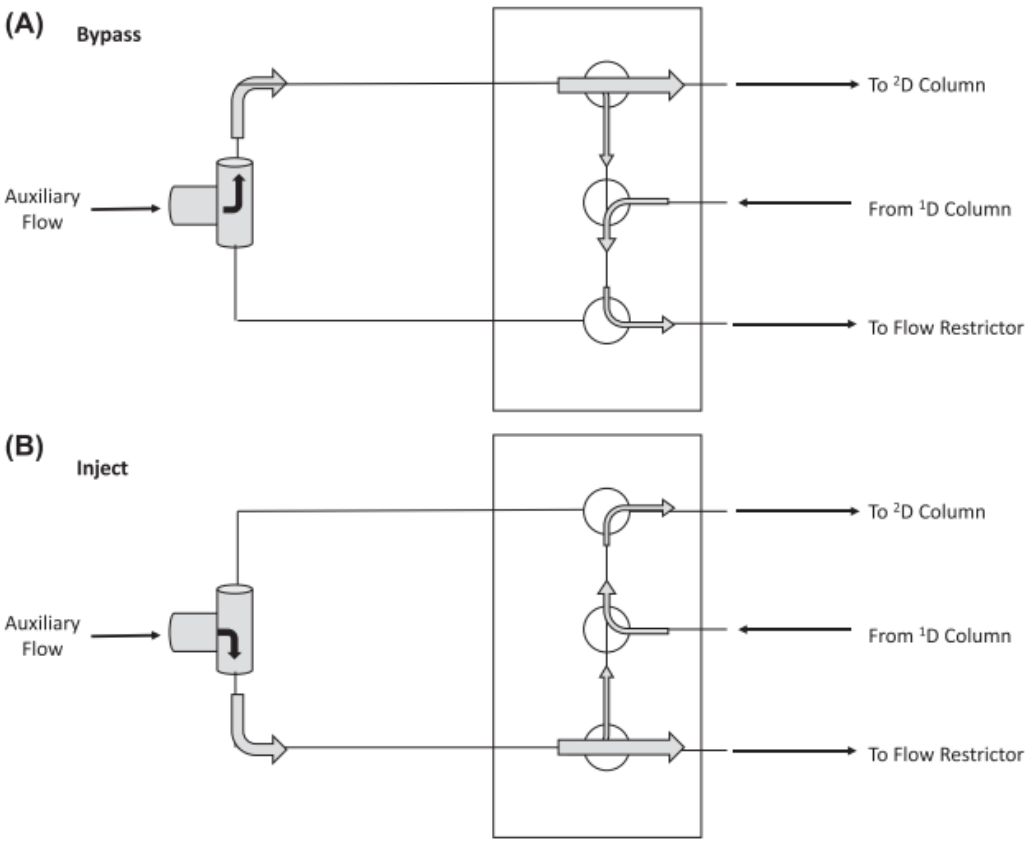

Figure 2.21. Flow diversion/Deans Switch modulator. (**2.21A**) Bypass state, showing the ^{1}D column effluent directed toward the flow restrictor; and (**2.21B**) Inject state, showing the auxiliary flow directing the ^{1}D column effluent into the ^{2}D column.[53]

Today this modulator is commercialized as *Flux™ Flow Modulator* by LECO, that implement it on the Pegasus® BT platform.

2.4.4.3 Differential flow-modulator (DFM)

*DFM*was introduced by Seeley in 2000 to overcome the weaknesses characterizing the first *DVM* design,[70] but new developments occurred in the successive years and a key event for the wider adoption of flow modulation was Agilent Technologies' introduction of the capillary flow technology (CFT) modulator in 2008, based on the modulator originally introduced by Seeley et al. in 2006.[71]

The first design of *DFM*, named "*Forward-Fill/Forward-Flush Modulator*" is shown in **Figure 2.22** indicating how, through a valve and a collection channel (or sample loop), the modulation process is realized in two main steps: (**Figure 2.22A**) "fill" and (**Figure 2.22B**) "flush". When the valve is switched into the "fill" mode, the ^{1}D eluate flows into the collection channel and an auxiliary flow of carrier gas enters the ^{2}D column; the auxiliary gas is supplied by a pressure control module (PCM) through a solenoid valve that can be switched.[14] At the end of the collection period, the collection capillary is flushed for 100–200 ms by a very high gas flow (typically 20 ml min^{-1}) generated by switching the valve to the flush position, allowing the entering of the eluate into the ^{2}D column. The flow in the sample loop is generally higher than that of the primary column and this design of modulation allows the entire volume of the ^{1}D column to be diverted into the ^{2}D column: it means that it produces a true "comprehensive" GC×GC, differently from low-duty cycle

pneumatic modulators.[35,56] This simple design resulted effective, even if susceptible to breakthrough if modulator channel is overfilled.[72]

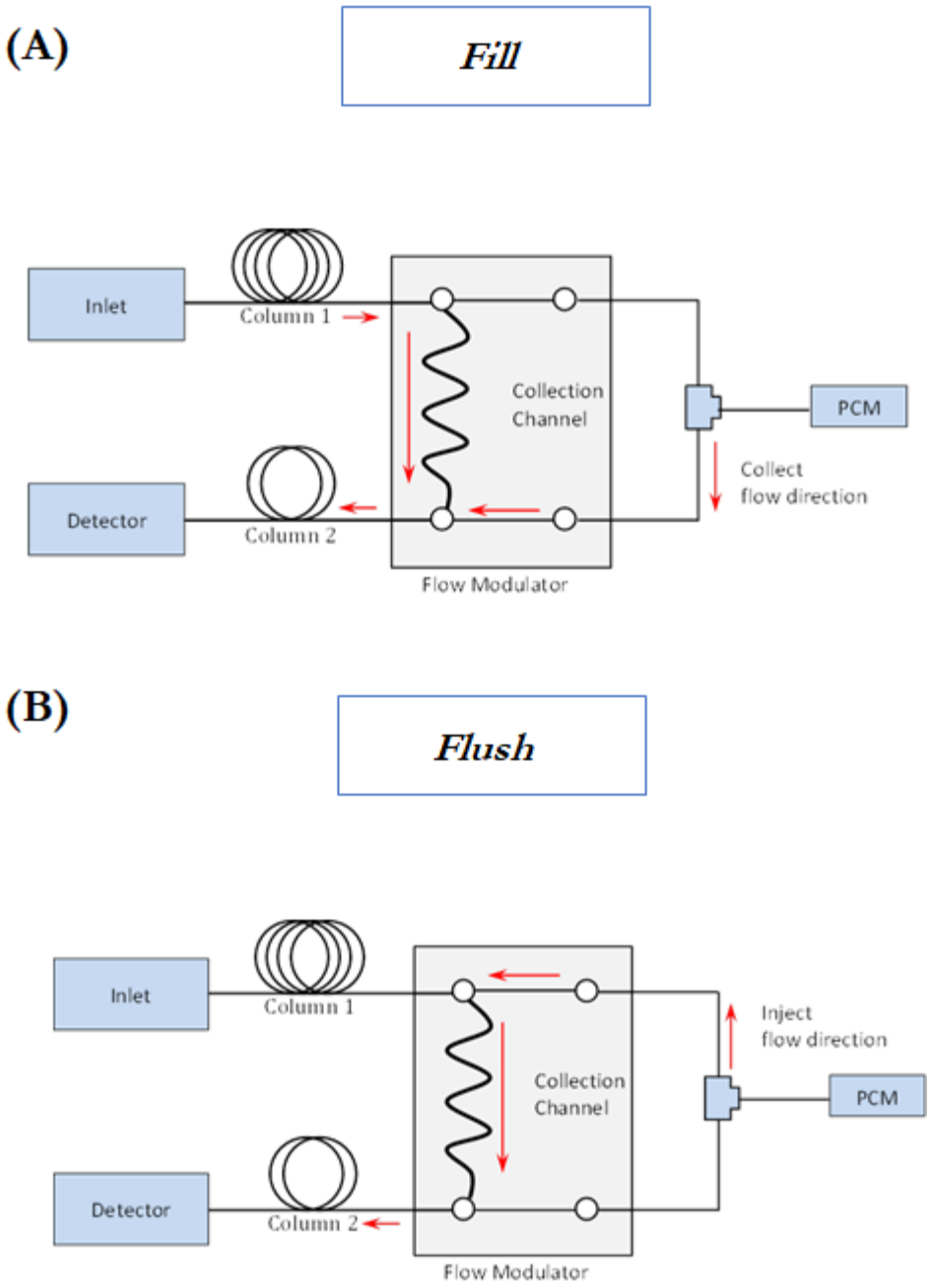

Figure 2.22. "Forward-Fill/Forward-Flush Modulator": (**2.22A**) fill mode and (**2.22B**) flush mode. Abbreviation: PCM - pressure control module. Adapted from[72]

Another important advancement resulted from the introduction of a new design combining 2 CTF by Griffith et al. in 2012 and successively commercialized by Agilent Technologies in a single CFT, with the addition of a vent restrictor: the "*Forward-Fill/Reverse-Flush Modulator*".[25,73] As shown in **Figure 2.23A** in the first step the collection channel is forward filled in the direction of a vent restrictor (usually a narrow-bore capillary) that, if opportunely set, avoid the collection channel

overloading without loss of effluent from the ^{1}D column. The flush (**Figure 2.23B**) instead, differently from the first design, is in reverse mode, resulting in a tailing of the peak at the base reduced 10-20 fold.[73] Excellent precision has been demonstrated as well as the capacity to handle significant overloading without loss of resolution in the ^{2}D dimension, resulting in an increase of the effective peak capacity of flow modulated GC×GC.[25,73] Moreover, the maximum operational temperature reported to be up to 300°C for both *DFM* design allows an extension of the possible applications, if compared to *DVM*.[56]

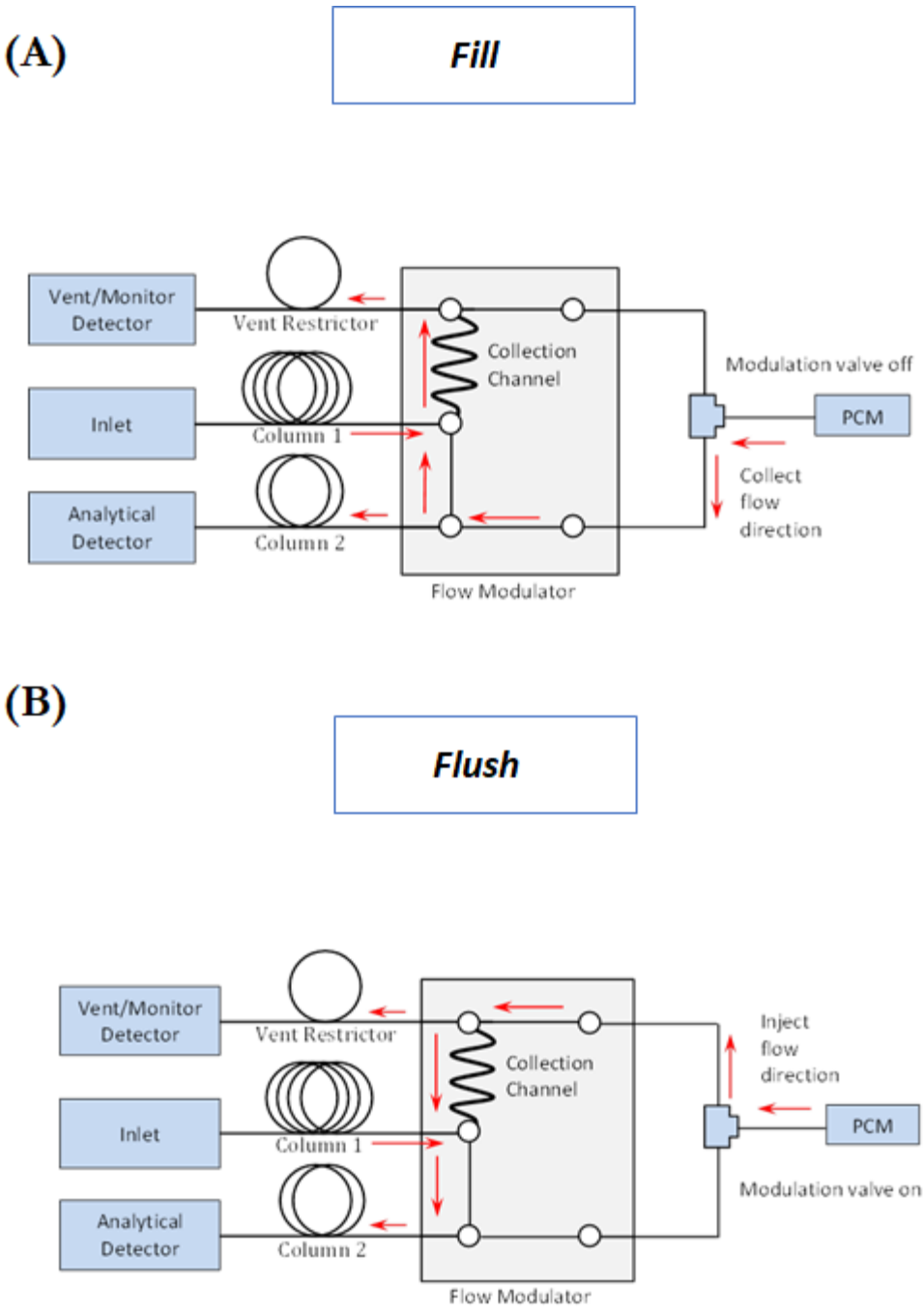

Figure 2.23. "Forward-Fill/Reverse-Flush Modulator": (**2.23A**) fill mode and (**2.23B**) flush mode. Abbreviation: PCM - pressure control module. Adapted from[72]

Other manufacturers have subsequently introduced their own commercial version of the DFM as SepSolve, that has recently introduced the INSIGHT flow modulator (**Figure 2.24**), which is based on a reverse fill/flush modulation.[29]

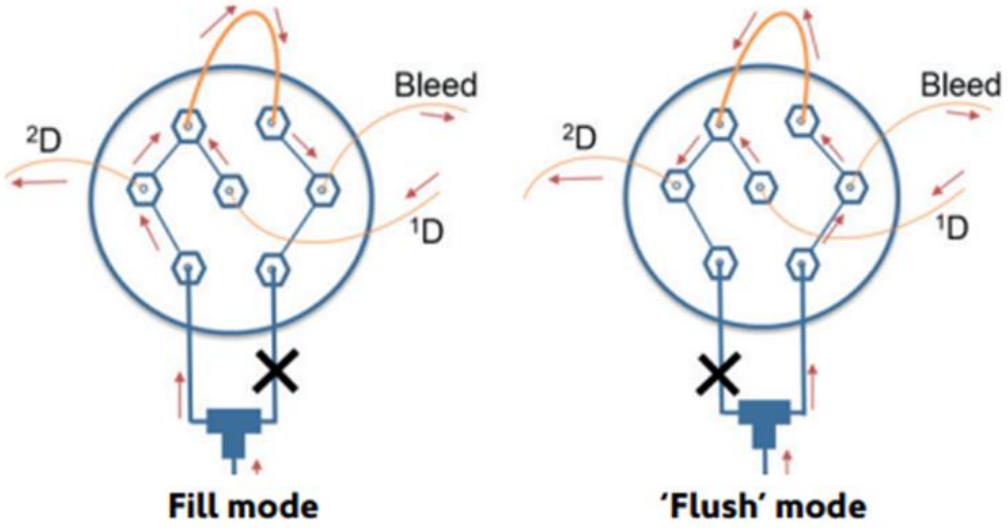

Figure 2.24. INSIGHT flow modulator from Sepsolve.[74]

Another version of the DFM was proposed by Tranchida et al. in 2011.[75] It was based on a seven port valve with a flexible loop between ports to collect the sample and it offers the possibility to optimize flows using a waste branch bridging the interface and a needle valve (**Figure 2.25**).

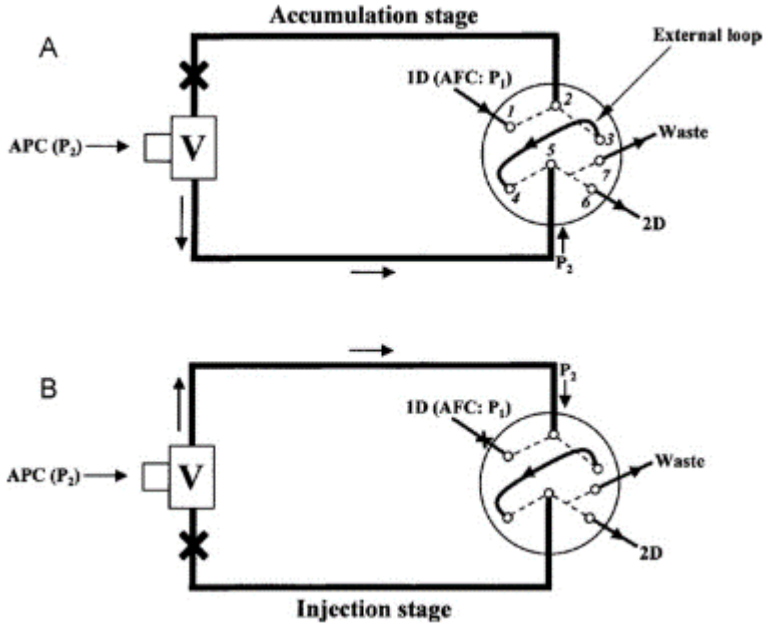

Figure 2.25. Scheme of the DFM accumulation (**2.25A**) and injection modes (**2.25B**). Abbreviations: V: 2-way solenoid valve; APC: advanced pressure controller.

The interface comprises a metallic disc and internal rectangular channels, with two metallic branches connecting the valve to the interface in positions 2 and 5, and the ^{1}D and ^{2}D columns linked to positions 1 and 6, respectively.[75] The size of the loop is chosen considering P_M, ^{1}D and ^{2}D columns flow and dimensions, with the flow exiting the loop that it divided between the channels linked to ports 6 and 7. The operating mechanism is related to the Seeley design discussed above, with fill and flush steps here named as accumulation and injection stages.

2.4.4.4 Stop-flow modulator

A particular design of modulator, combining valves and cryogens, is the *Stop-flow modulator* proposed by Harynuk and Gorecki in 2004:[76] It stops the carrier gas in the ^{1}D column to allow the separation of the effluent in the ^{2}D column to reach completion. The scheme of this modulator is shown in **Figure 2.26**: the interface utilized an air-actuated six-port valve within a heated chamber and a single-stage N_2 jet was placed downstream of the valve to trap and focus the ^{1}D effluent before the subsequent injection into the ^{2}D column. An uncoated fused silica capillary having the same dimensions of the ^{1}D column is used to supply auxiliary carrier gas and a separate transfer line of deactivated fused silica was placed between the valve and the cryotrap. In the sample position (**Figure 2.26A**), the 1D column effluent passed through the valve to the cryotrap where it was trapped and focused by the liquid N_2 jet. After a determine time, the valve was switched to the stop position, the cryo-jet was switched off and the warm-jet activated. The ^{1}D column flow was then stopped and the carrier gas directed to allow the flush of the focused analytes into the ^{2}D column (**Figure 2.26B**). Finally, after the completion of the 2D separation, valve was re-activated, and cryo-jet was employed. The main advantage of the *Stop-flow modulator* consists of the independent use of the two dimensions, therefore overcoming the necessity of operating in "fast" conditions in the second dimension; on the other hand, the overall analysis time was extended by the total stop time.[35,53]

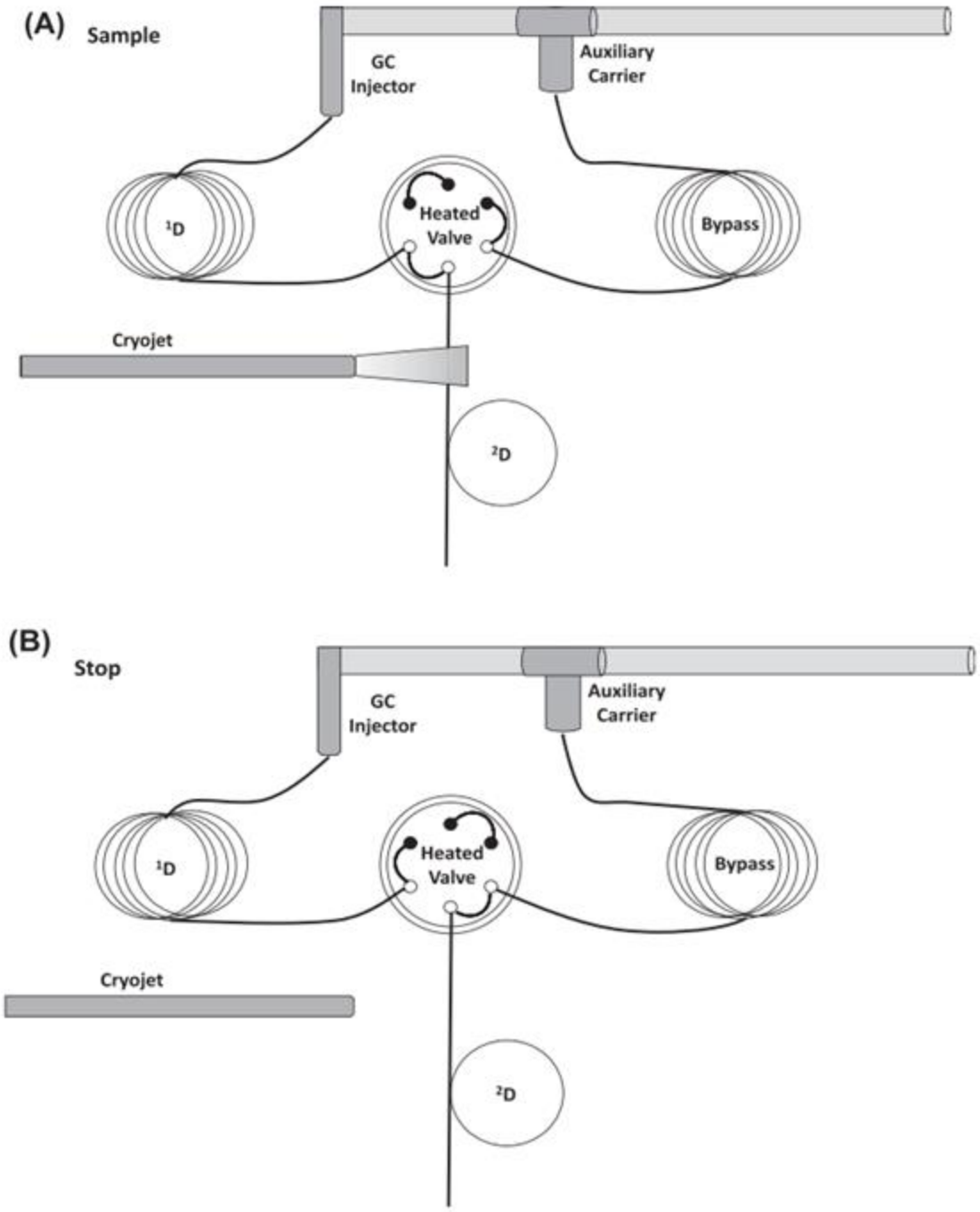

Figure 2.26. Scheme of the *Stop-flow modulator* showing the two positions: (**2.26A**) sample and (**2.26B**) stop. The black dots correspond to the plugged ports, while the open dots correspond to the three ports that were utilized.[53]

2.4.5 Perspectives

Nowadays thermal modulation is still considered the "golden standard" because of superior performances in terms of sensitivity and overall peak capacity. However, the perspective of a cryogen-free modulation is of growing interest, leading to a significant reduction of the associated setup and running costs, in addition to the decrease of safety issues.[34] If from one side research is progressing in the field of cooling-based cryogen-free modulators, an exponentially increasing

interest is directed towards the world of flow modulators: Zanella et al. reported that, while only 7% of the studies making used of flow modulators until 2017, this number was increased to 16% in 2020.[34]

New applications are several, and an accurate optimization of the DFM setup is bringing to a performance level that is increasingly closer to that produced by cryogen modulators. Recent studies from Aloisi *et al.*[77] and Stilo *et al.*[78] demonstrated that is still difficult to simultaneously match for sensitivity and separation power to those obtained by a cryogen modulated platform, however an equivalent separation power can be reached at the cost of sensitivity or, an almost equivalent sensitivity can be achieved at the cost of ~20% separation power.

An interesting recent innovation in this field was that proposed by Seeley et al in 2018 with the introduction of the *multi-mode modulator (MMM)*.[79] It is a fluidic device proposed with the ambition to cover the entire range of possibilities included in the world of MDGC: it can be used as a traditional heart-cutting device, a low duty cycle GC×GC modulator, and a full transfer GC×GC modulator. It is characterized by a deactivated metal joining capillary linked on the ^{2}D column side to a T union and on the ^{1}D side to a cross union, while an auxiliary pressure unit supplies a solenoid valve connected to the unions through two metal capillaries. When the valve flow is directed to the T union, the ^{1}D gas flow can be directed to the restrictor (low duty cycle mode) or accumulated in the joining capillary (high duty cycle mode) while, when the valve flow is directed to the cross union (normally in closed position), the ^{1}D effluent is directed to the ^{2}D column.[79,80]

The latest advancement proposed in the field of pneumatic modulator was by Aloisi et al. in 2021, presenting the use of the Flux™ Flow Modulator under conditions capable of providing a higher duty cycle (0.04), with an acceptable level of analyte transfer from the ^{1}D to the ^{2}D column. It was implemented with a columns set-up using a 10 m×0.25 mm×0.25 μm as ^{1}D column, and a 1 m×0.10 mm×0.10μm as ^{2}D column, and a P_M of 700 ms, with a re-injection period of 80 ms was used.[80]

Finally, the latest frontier in the field of MDGC was explored by Synovec's group since 2007, with the introduction of comprehensive three- dimensional (3D) gas chromatography (GC^3). However, interesting advancements have been done with the introduction of the high-temperature diaphragm valve based on the perfluoroelastomer-based O-ring, which was recently implemented on a GC×GC×GC-TOF MS platform combining the Diaphragm valve and a *Dual-stage quad-jet modulator*.[81,82]

2.5 Detectors for GC×GC

Peaks eluted from the ^{2}D are the result of a focusing and reinjection process operated by the modulator, and they are characterized by a narrow width (*i.e.*, 100-300 ms).[14,27,29] For this reason, the main requirement of a GC×GC detector is a suitably fast acquisition rate (at least 50-100 Hz, to provide a minimum of 10 scans per peak).[25,27,29,35]

Detectors provide a conversion from analog-to-digital of the chromatographic signal collected and they are classified in univariate or single channel detectors, producing a single data point for each time point of the chromatogram, and multivariate or multichannel detectors, producing multiple data points for each time point (typically over a spectral range).[14,29] Data collected are then stored under proprietary data file format that can be converted to text by using the ASCII format comma separated values (CSV) or the ASTM format analytical data interchange (ANDI), which is a standard for chromatography and MS data.[14]

2.5.1 Single channel detectors

Flame ionization detector (FID) was the first detector coupled to GC×GC. It detects ions formed from the combustion of organic compounds by a hydrogen flame, giving a response that is mostly proportional to the number of carbons present in the molecules. It is simple, robust, with a small internal volume, and it is characterized by a high acquisition frequency (50-300 Hz). FID is particularly suitable for quantitative purposes and it is highly adopted in studies of the petroleum area[29,35]

Electron capture detector (ECD) is usually characterized by a slow acquisition rate, being theoretically not suitable for GC×GC analysis. However, being an electron emitter attracting high electron affinity molecules, it is sensitive to halogenated compounds and thus used for specific applications (*e.g.,* pesticides, herbicides).[14,27,35]

Nitrogen-chemiluminescence detectors (NCD) and sulfur-chemiluminescence detectors (SCD) are other element-selective detectors used for specific applications. They have higher acquisition speed than ECD, comparable to that of FID. [14,27,35]

Nitrogen–phosphorous detector (NPD), more correctly named Thermoionic Detector (TID) because of its ionization mechanism. It is very similar to the FID, however here the source is not a flame, but an alkali salt inside a ceramic cement matrix.[35]

2.5.2 Multichannel detectors

Multichannel detectors have the ability to produce spectral data used to identify the analyte based on known spectral databases.[29]

The most used multichannel detectors are mass spectrometers (MS) and, particularly, the most common implementation of MS within a GC×GC system is time-of-flight-MS (TOF-MS), due to its high acquisition frequency (50-250 Hz).[25,29,35] TOF-MS is highly selective and it is based on the simple principle according to which, if the ionized species start from the same point at the same time, their acceleration and velocity are directly correlated to their mass-to-charge ratio (m/z), and they will reach the detector at a time depending on their masses.[35] Recently, high-resolution TOF-MS platforms (*e.g.,* Pegasus GC-HRT+ 4D) have offered an increased informative potential through an accurate mass detection, at the cost of a reduced acquisition frequency.[14,35] Tandem ionization TOF-MS (*e.g.,* BenchTOF-Select TOFMS by Markes International) are able instead to alternate the defined ionization energy in the range 10-70 eV (at the cost of a halved acquisition frequency – 50 Hz per channel), producing twin data streams that bring the complementary information derived from the hard and the milder electron ionization energies (EI).[29]

Less effective are quadrupole MS (q-MS) because of their relatively low acquisition frequency (*i.e.,* 20-50 Hz). However, they represent a good alternative to high-end MS because of their lower costs and robustness.[14,25,27]

Finally, an alternative multichannel detector is the Vacuum Ultraviolet (VUV). Modern VUV detectors (*e.g.,* VGA-100 and VGA-101 by VUV Analytics) reach an acquisition frequency of 90 Hz; moreover, they provide a linear response and no need for calibration, they are excellent for isomer differentiation and robust, with minimal maintenance requirements.[29]

2.6 GC×GC and olive oil

This short chapter is intended to be a guide including the current state-of-the-art regarding the application of GC×GC to olive oil characterization. GC×GC has been applied to investigate the complex fractions of volatiles and semi-volatiles composing olive oil, by exploiting both untargeted fingerprinting based on 2D patterns and detailed profiling of known compounds (*i.e.,* targeted analytes). Applications related to olive oil geographical origin and variety discrimination, technological issues, identification of ripening indicators and characterization of minor fractions are collected in **Table 2.6.1**, listed by alphabetical order by author.

Table 2.6.1. It includes the list of GC×GC papers related to olive oil before July 2021. Papers are listed in alphabetical order by author

Author	Title	Year	Journal	DOI
Adahchour M. et al.	Twin comprehensive two-dimensional gas chromatographic system: concept and applications	2005	Journal of Chromatography A	10.1016/j.chroma.2004.12.021
Aloisi I. et al.	Fingerprinting of the Unsaponifiable Fraction of Vegetable Oils by Using Cryogenically Modulated Comprehensive Two-Dimensional Gas Chromatography-High Resolution Time-of-Flight Mass Spectrometry	2020	Food Analytical Methods	10.1007/s12161-020-01773-9
Barp L. et al.	In-pipette solid-phase extraction prior to flow-modulation comprehensive two-dimensional gas chromatography with dual detection for the determination of minor components in vegetable oils	2017	Talanta	10.1016/j.talanta.2017.01.009
Cajka T. et al.	Traceability of olive oil based on volatiles pattern and multivariate analysis	2010	Food Chemistry	10.1016/j.foodchem.2009.12.011
Da Ros A. et al.	Complementary Untargeted and Targeted Metabolomics for Differentiation of Extra Virgin Olive Oils of Different Origin of Purchase Based on Volatile and Phenolic Composition and Sensory Quality	2019	Molecules	10.3390/molecules24162896
Fiori F. et al	Enhanced Profile Characterization of Virgin Olive Oil Minor Polar Compound Extracts by Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometric Detection	2016	Polish Journal of Applied Sciences	
Janssen H.-G. et al.	Comprehensive two-dimensional liquid chromatography × gas chromatography: evaluation of the applicability for the analysis of edible oils and fats	2003	Journal of Chromatography A	10.1016 /S0021-9673(02)02058-7
Lukic I. et al.	Combined targeted and untargeted profiling of volatile aroma compounds with comprehensive two-dimensional gas chromatography for differentiation of virgin olive oils according to variety and geographical origin	2019	Food Chemistry	10.1016/j.foodchem.2018.07.133
Magagna F. et al.	Combined untargeted and targeted fingerprinting with comprehensive two-dimensional chromatography for volatiles and ripening indicators in olive oil	2016	Analytica Chimica Acta	10.1016/j.aca.2016.07.005

Mondello L. et al.	Evaluation of fast gas chromatography and gas chromatography–mass spectrometry in the analysis of lipids	2004	Journal of Chromatography A	10.1016/j.chroma.2004.02.058
Purcaro G. et al.	Determination of polycyclic aromatic hydrocarbons in vegetable oils using solid-phase microextraction–comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry	2007	Journal of Chromatography A	10.1016/j.chroma.2007.05.105
Purcaro G. et al.	Fingerprinting of vegetable oil minor components by multidimensional comprehensive gas chromatography with dual detection	2015	Analytical and Bioanalytical Chemistry	10.1007/s00216-014-8140-x
Purcaro G. et al.	Characterisation of minor components in vegetable oil by comprehensive gas chromatography with dual detection	2016	Food Chemistry	10.1016/j.foodchem.2016.06.048
Purcaro G. et al.	Toward a definition of blueprint of virgin olive oil by comprehensive two-dimensional gas chromatography	2014	Journal of Chromatography A	10.1016/j.chroma.2014.01.067
Stilo F. et al.	Untargeted and Targeted Fingerprinting of Extra Virgin Olive Oil Volatiles by Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry: Challenges in Long-Term Studies	2019	Journal of Agricultural and Food Chemistry	10.1021/acs.jafc.9b01661
Stilo F. et al.	Highly Informative Fingerprinting of Extra-Virgin Olive Oil Volatiles: The Role of High Concentration-Capacity Sampling in Combination with Comprehensive Two-Dimensional Gas Chromatography	2019	Separations	10.3390/separations6030034
Stilo F. et al.	Chromatographic Fingerprinting by Template Matching for Data Collected by Comprehensive Two-Dimensional Gas Chromatography	2020	Journal of Visual Experiment	10.3791/61529
Stilo F. et al.	Exploring the extra-virgin olive oil volatilome by adding extra dimensions to comprehensive two-dimensional gas chromatography and time of flight mass spectrometry featuring tandem ionization: validation of ripening markers in headspace linearity conditions	2020	Journal of AOAC International	10.1093/jaoacint/qsaa095
Tranchida P.Q. et al.	Elucidation of fatty acid profiles in vegetable oils exploiting group-type patterning and enhanced sensitivity of comprehensive two-dimensional gas chromatography	2008	Journal of Separation Science	10.1002/jssc.200800002
Vaz-Freire L.T. et al.	Comprehensive two-dimensional gas chromatography for fingerprint pattern recognition in olive oils produced by two different techniques in Portuguese olive varieties Galega Vulgar, Cobranc¸ osa e Carrasquenha	2009	Analytica Chimica Acta	10.1016/j.aca.2008.11.057
Vyviurska O. et al.	Comprehensive Two-Dimensional Gas Chromatography–Mass Spectrometry Analysis of Different Types of Vegetable Oils	2015	Journal of the American Oil Chemists' Society	10.1007/s11746-015-2635-2

References

(1) Nolvachai, Y.; Kulsing, C.; Marriott, P. J. Multidimensional Gas Chromatography in Food Analysis. *TrAC - Trends Anal. Chem.* **2017**, *96*, 124–137. https://doi.org/10.1016/j.trac.2017.05.001.

(2) Edwards, M.; Mostafa, A.; Górecki, T. Modulation in Comprehensive Two-Dimensional Gas Chromatography: 20 Years of Innovation. *Anal. Bioanal. Chem.* **2011**, *401* (8), 2335–2349. https://doi.org/10.1007/s00216-011-5100-6.

(3) Goppelsröder, F. Note Sur Une Méthode Nouvelle Propre à Déterminer La Nature d'un Mélange de Principes Colorants. *Bull. Soc. Ind. Mulhouse* **1862**, *XXXII*.

(4) Beens, J. *Comprehensive Two-Dimensional Gas Chromatography: The State-of-Separation-Arts*; 2009.

(5) Tswett, M. S. Physikalisch-Chemische Studien Über Das Chlorophyll. Die Absorptionen. *Ber. deut. Bot. Ges.* **1906**, *24*, 316–384.

(6) Mondello, L.; C. Lewis, A.; D. Bartle, K. *Multidimensional Chromatography*; John Wiley & Sons, Ltd: Chichester, UK, 2001. https://doi.org/10.1002/0470845775.

(7) Ettre, L. S. Chromatography: The Separation Technique of the 20th Century. *Chromatographia* **2000**, *51* (1–2), 7–17. https://doi.org/10.1007/BF02490689.

(8) Karrer, P. Purity and Activity of Vitamin A. *Helv. Chim. Acta* **1939**, *22*.

(9) Swab, G. M.; Jockers, K. Inorganic Chromatography I. *Angew. Chem.* **1937**, *50*.

(10) Martin, A. J. P.; Synge, R. L. M. A New Form of Chromatogram Employing Two Liquid Phases. I: A Theory of Chromatogaphy. *Biochem. J.* **1941**, *35* (1158).

(11) James, A. T.; Martin, A. J. P. The Separation and Micro Estimation of Volatile Fatty Acids. *Biochem. J.* **1952**, *50*.

(12) Phillips, C. G. S. The Chromatography of Gases and Vapours. *Disc. Faraday. Soc.* **1949**, *67*.

(13) Golay, M. J. E. Theory of Chromatography in Open and Coated Tubular Columns with Round and Rectangular Cross-Section. *Butterworths Sci. Publ.* **1958**, 36–55.

(14) Cordero, C.; Cagliero, C.; Liberto, E.; Sgorbini, B.; Rubiolo, P.; Bicchi, C. Chromatography: Focus on Multidimensional GC. In *Encyclopedia of Food and Health*; Elsevier, 2016; Vol. 2, pp 85–92. https://doi.org/10.1016/B978-0-12-384947-2.00157-4.

(15) Dimandja, J. M. D. *Introduction and Historical Background: The "inside" Story of Comprehensive Two-Dimensional Gas Chromatography*; Elsevier, 2020; Vol. 12. https://doi.org/10.1016/B978-0-12-813745-1.00001-5.

(16) Giddings, J. C. Maximum Number of Components Resolvable by Gel Filtration and Other Elution Chromatographic Methods. *Anal. Chem.* **1967**, *39* (8), 1027–1028.

(17) Dettmer-Wilde, K.; Engewald, W. *Practical Gas Chromatography - A Comprehensive Reference*; Springer, 2014.

(18) Davis, J. M.; Giddings, J. C. Statistical Theory of Component Overlap in Multicomponent Chromatograms. *Anal Chem* **1983**, *55* (3), 418–424.

(19) Giddings, J. C. Sample Dimensionality: A Predictor of Order-Disorder in Component Peak Distribution in Multidimensional Separation. *J. Chromatogr. A* **1995**, *703* (1–2), 3–15.

https://doi.org/10.1016/0021-9673(95)00249-M.

(20) Mostafa, A.; Edwards, M.; Górecki, T. Optimization Aspects of Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2012**, *1255*, 38–55. https://doi.org/10.1016/j.chroma.2012.02.064.

(21) Kovats, E. Gas-Chromatographische Charakterisierung Organischer Verbindungen. Teil 1: Retentionsindices Aliphastischer Halogenide, Alkohole, Aldehyde Und Ketone. *Helv. Chim. Acta* **1958**, *41* (7), 1915–1932.

(22) Van den Dool, H.; Kratz, P. D. A Generalization of the Retention Index System Including Linear Temperature Programmed Gas-Liquid Partition Chromatography. *J. Chromatogr* **1963**, *11*, 463–471.

(23) Giddings, J. C. *Multi- Dimensional Chromatography. Techniques and Applications*, Chromatogr.; Dekker, M., Ed.; New York, 1990.

(24) Ramos, L. *Comprehensive Two Dimensional Gas Chromatography*; Wilson & Wilson's, 2009.

(25) Seeley, J. V.; Seeley, S. K. Multidimensional Gas Chromatography: Fundamental Advances and New Applications. *Anal. Chem.* **2013**, *85* (2), 557–578. https://doi.org/10.1021/ac303195u.

(26) Marriott, P. J. *Encyclopedia of Analytical Science*; Worsfold, P., Townshend, A., Poole, C., Eds.; Elsevier, 2005.

(27) Marriott, P. J.; Chin, S. T.; Maikhunthod, B.; Schmarr, H. G.; Bieri, S. Multidimensional Gas Chromatography. *TrAC - Trends in Analytical Chemistry*. 2012, pp 1–20. https://doi.org/10.1016/j.trac.2011.10.013.

(28) Marriott, P. J.; Schoenmakers, P.; Wu, Z. Y. Nomenclature and Conventions in Comprehensive Multidimensional Chromatography- an Update. *LC GC Eur.* **2012**, *25* (5), 1–7.

(29) Prebihalo, S. E.; Berrier, K. L.; Freye, C. E.; Bahaghighat, H. D.; Moore, N. R.; Pinkerton, D. K.; Synovec, R. E. Multidimensional Gas Chromatography: Advances in Instrumentation, Chemometrics, and Applications. *Anal. Chem.* **2018**, *90* (1), 505–532. https://doi.org/10.1021/acs.analchem.7b04226.

(30) Liu, Z.; Phillips, J. B. Comprehensive Two-Dimensional Gas Chromatography Using an on-Column Thermal Modulator Interface. *J. Chromatogr. Sci.* **1991**, *29* (5), 227–231. https://doi.org/10.1093/chromsci/29.6.227.

(31) Klee, M. S.; Cochran, J.; Merrick, M.; Blumberg, L. M. Evaluation of Conditions of Comprehensive Two-Dimensional Gas Chromatography That Yield a near-Theoretical Maximum in Peak Capacity Gain. *J. Chromatogr. A* **2015**, *1383*, 151–159. https://doi.org/10.1016/j.chroma.2015.01.031.

(32) Deans, D. R. A New Technique for Heart Cutting in Gas Chromatography. *Chromatographia* **1968**, *1*, 18–22.

(33) Schomburg, G.; Weeke, F.; Muller, F.; Oreans, M. Multidimensional Gas Chromatography (MDC) in Capillary Columns Using Double Oven Instruments and a Newly Designed Coupling Piece for Monitoring Detection after Pre-Separation. *Chromatographia* **1982**, *16*, 87–91.

(34) Zanella, D.; Focant, J.; Franchina, F. A. 30 Th Anniversary of Comprehensive Two-

dimensional Gas Chromatography: Latest Advances . *Anal. Sci. Adv.* **2021**, *2* (3–4), 213–224. https://doi.org/10.1002/ansa.202000142.

(35) L. Mondello. Fundamental Principles of Comprehensive 2D GC. *GC×GC Handb.* **2012**, 1–29.

(36) Tranchida, P. Q.; Purcaro, G.; Maimone, M.; Mondello, L. Impact of Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry on Food Analysis. *J. Sep. Sci.* **2016**, *39* (1), 149–161. https://doi.org/10.1002/jssc.201500379.

(37) Stilo, F.; Gabetti, E.; Bicchi, C.; Carretta, A.; Peroni, D.; Reichenbach, S. E.; Cordero, C.; McCurry, J. A Step Forward in the Equivalence between Thermal and Differential-Flow Modulated Comprehensive Two-Dimensional Gas Chromatography Methods. *J. Chromatogr. A* **2020**, *1627*, 461396. https://doi.org/10.1016/j.chroma.2020.461396.

(38) Frysinger, G. S.; Gaines, R. B. Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometric Detection (GC × GC/MS) Applied to the Analysis of Petroleum. *J. High Resolut. Chromatogr.* **1999**, *22* (5), 251–255. https://doi.org/10.1002/(SICI)1521-4168(19990501)22:5<251::AID-JHRC251>3.0.CO;2-V.

(39) Shellie, R.; Marriott, P. J. Comprehensive Two-Dimensional Gas Chromatography with Fast Enantioseparation. *Anal. Chem.* **2002**, *74* (20), 5426–5430. https://doi.org/10.1021/ac025803e.

(40) Phillips, J. B.; Beens, J. Comprehensive Two-Dimensional Gas Chromatography: A Hyphenated Method with Strong Coupling between the Two Dimensions. *J. Chromatogr. A* **1999**, *856* (1–2), 331–347. https://doi.org/10.1016/S0021-9673(99)00815-8.

(41) Amaral, M. S. S.; Nolvachai, Y.; Marriott, P. J. Comprehensive Two-Dimensional Gas Chromatography Advances in Technology and Applications: Biennial Update. *Anal. Chem.* **2020**, *92* (1), 85–104. https://doi.org/10.1021/acs.analchem.9b05412.

(42) Stefanuto, P. H.; Focant, J. F. Columns and Column Configurations. *Sep. Sci. Technol. (New York)* **2020**, *12*, 69–88. https://doi.org/10.1016/B978-0-12-813745-1.00003-9.

(43) Nolvachai, Y.; McGregor, L.; Spadafora, N. D.; Bukowski, N. P.; Marriott, P. J. Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry: Toward a Super-Resolved Separation Technique. *Anal. Chem.* **2020**, *92* (18), 12572–12578. https://doi.org/10.1021/acs.analchem.0c02522.

(44) Seeley, J. V.; Kramp, F. J.; Sharpe, K. S. A Dual-Secondary Column Comprehensive Two-Dimensional Gas Chromatograph for the Analysis of Volatile Organic Compound Mixtures. *J. Sep. Sci.* **2001**, *24* (6), 444–450. https://doi.org/10.1002/1615-9314(20010601)24:6<444::AID-JSSC444>3.0.CO;2-5.

(45) Harynuk, J.; Górecki, T.; Zeeuw, J. de. Overloading of the Second-Dimension Column in Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2005**, *1071* (1–2), 21–27. https://doi.org/10.1016/j.chroma.2004.11.049.

(46) Cordero, C.; Bicchi, C.; Galli, M.; Galli, S.; Rubiolo, P. Evaluation of Different Internal-Diameter Column Combinations in Comprehensive Two-Dimensional Gas Chromatography in Flavour and Fragrance Analysis. *J. Sep. Sci.* **2008**, *31* (19), 3437–3450. https://doi.org/10.1002/jssc.200800280.

(47) Adahchour, M.; Tazöz, A.; Beens, J.; Vreuls, R. J. J.; Batenburg, A. M.; Brinkman, U. A. T. Fast Comprehensive Two-Dimensional Gas Chromatography (GC×GC) Using 50-Mm ID

Second-Dimension Columns. *J. Sep. Sci.* **2003**, *26* (9–10), 753–760. https://doi.org/10.1002/jssc.200301332.

(48) Peroni, D.; Sampat, A. A. S.; van Egmond, W.; de Koning, S.; Cochran, J.; Lautamo, R.; Janssen, H.-G. Comprehensive Two-Dimensional Gas Chromatography with a Multi-Capillary Second Dimension: A New Column-Set Format for Simultaneous Optimum Linear Velocity Operation. *J. Chromatogr. A* **2013**, *1317*, 3–11. https://doi.org/10.1016/j.chroma.2013.07.097.

(49) Nolvachai, Y.; Kulsing, C.; Marriott, P. J. Thermally Sensitive Behavior Explanation for Unusual Orthogonality Observed in Comprehensive Two-Dimensional Gas Chromatography Comprising a Single Ionic Liquid Stationary Phase. *Anal. Chem.* **2015**, *87* (1), 538–544. https://doi.org/10.1021/ac5030039.

(50) Franchina, F. Novel Developments in the Field of Cryogenic and Flow "Loop-Type" Modulators for Comprehensive Two-Dimensional Gas Chromatography Applied to Complex Food Analysis, 2013.

(51) Cordero, C.; Rubiolo, P.; Sgorbini, B.; Galli, M.; Bicchi, C. Comprehensive Two-Dimensional Gas Chromatography in the Analysis of Volatile Samples of Natural Origin: A Multidisciplinary Approach to Evaluate the Influence of Second Dimension Column Coated with Mixed Stationary Phases on System Orthogonality. *J. Chromatogr. A* **2006**, *1132* (1–2), 268–279. https://doi.org/10.1016/j.chroma.2006.07.067.

(52) Murphy, R. E.; Schure, M. R.; Foley, J. P. Effect of Sampling Rate on Resolution Comprehensive Two- Dimensional Liquid Chromatography. *Anal Chem* **1998**, *70*, 1585–1594.

(53) Boswell, H.; Chow, H. Y.; Gorecki, T. *Modulators*; Elsevier, 2020; Vol. 12. https://doi.org/10.1016/B978-0-12-813745-1.00004-0.

(54) Bahaghighat, H. D.; Freye, C. E.; Synovec, R. E. Recent Advances in Modulator Technology for Comprehensive Two Dimensional Gas Chromatography. *TrAC - Trends Anal. Chem.* **2019**, *113*, 379–391. https://doi.org/10.1016/j.trac.2018.04.016.

(55) Tranchida, P. Q.; Purcaro, G.; Dugo, P.; Mondello, L.; Purcaro, G. Modulators for Comprehensive Two-Dimensional Gas Chromatography. *TrAC Trends Anal. Chem.* **2011**, *30* (9), 1437–1461. https://doi.org/10.1016/j.trac.2011.06.010.

(56) Tranchida, P. Q. Comprehensive Two-Dimensional Gas Chromatography: A Perspective on Processes of Modulation. *J. Chromatogr. A* **2018**, *1536*, 2–5. https://doi.org/10.1016/j.chroma.2017.04.039.

(57) Phillips, J. B.; Gaines, R. B.; Blomberg, J.; van der Wielen, F. W. M.; Dimandja, J.-M.; Green, V.; Granger, J.; Patterson, D.; Racovalis, L.; de Geus, H.-J.; de Boer, J.; Haglund, P.; Lipsky, J.; Sinha, V.; Ledford, E. B. A Robust Thermal Modulator for Comprehensive Two-Dimensional Gas Chromatography. *J. High Resolut. Chromatogr.* **1999**, *22* (1), 3–10. https://doi.org/10.1002/(SICI)1521-4168(19990101)22:1<3::AID-JHRC3>3.0.CO;2-U.

(58) Gaines, R. B.; Frysinger, G. S. Temperature Requirements for Thermal Modulation in Comprehensive Two-Dimensional Gas Chromatography. *J. Sep. Sci.* **2004**, *27* (5–6), 380–388. https://doi.org/10.1002/jssc.200301651.

(59) Kinghorn, R. M.; Marriott, P. J. Enhancement of Signal-to-Noise Ratios in Capillary Gas Chromatography by Using a Longitudinally Modulated Cryogenic System. *J. High Resolut.*

Chromatogr. **1998**, *21* (1), 32–38. https://doi.org/10.1002/(SICI)1521-4168(19980101)21:1<32::AID-JHRC32>3.0.CO;2-0.

(60) Ledford, E. B. Contribution Presented at the 23rd International Symposium on Capillary Chromatography - PL.20; 2000.

(61) Ledford, E. B.; Billesbach, C.; TerMaat, J. Contribution Presented at PITTCON 2002 - 2262P; 2002.

(62) Harynuk, J.; Górecki, T. Flow Model for Coupled-Column Gas Chromatography Systems. *J. Chromatogr. A* **2005**, *1086* (1–2), 135–140. https://doi.org/10.1016/j.chroma.2005.06.008.

(63) Muscalu, A. M.; Edwards, M.; Górecki, T.; Reiner, E. J. Evaluation of a Single-Stage Consumable-Free Modulator for Comprehensive Two-Dimensional Gas Chromatography: Analysis of Polychlorinated Biphenyls, Organochlorine Pesticides and Chlorobenzenes. *J. Chromatogr. A* **2015**, *1391* (1), 93–101. https://doi.org/10.1016/j.chroma.2015.02.074.

(64) Luong, J.; Guan, X.; Xu, S.; Gras, R.; Shellie, R. A. Thermal Independent Modulator for Comprehensive Two-Dimensional Gas Chromatography. *Anal. Chem.* **2016**, *88* (17), 8428–8432. https://doi.org/10.1021/acs.analchem.6b02525.

(65) Bruckner, C. A.; Prazen, B. J.; Synovec, R. E. Comprehensive Two-Dimensional High-Speed Gas Chromatography with Chemometric Analysis. *Anal. Chem.* **1998**, *70* (14), 2796–2804. https://doi.org/10.1021/ac980164m.

(66) Mohler, R. E.; Prazen, B. J.; Synovec, R. E. Total-Transfer, Valve-Based Comprehensive Two-Dimensional Gas Chromatography. *Anal. Chim. Acta* **2006**, *555*, 68–74. https://doi.org/10.1016/j.aca.2005.08.072.

(67) Sinha, A. E.; Prazen, B. J.; Fraga, C. G.; Synovec, R. E. Valve-Based Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometric Detection : Instrumentation and Figures-of-Merit. *J. Chromatogr. A* **2003**, *1019*, 79–87. https://doi.org/10.1016/j.chroma.2003.08.047.

(68) Freye, C. E.; Synovec, R. E. High Temperature Diaphragm Valve-Based Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry. *Talanta* **2015**, *161*, 675–680. https://doi.org/10.1016/j.talanta.2016.09.002.

(69) Seeley, J. V.; Micyus, N. J.; Bandurski, S. V.; Seeley, S. K.; McCurry, J. D. Microfluidic Deans Switch for Comprehensive Two-Dimensional Gas Chromatography. *Anal. Chem.* **2007**, *79* (5), 1840–1847. https://doi.org/10.1021/ac061881g.

(70) Seeley, J. V.; Kramp, F.; Hicks, C. J. Comprehensive Two-Dimensional Gas Chromatography via Differential Flow Modulation. *Anal. Chem.* **2000**, *72* (18), 4346–4352. https://doi.org/10.1021/ac000249z.

(71) Seeley, J. V.; Micyus, N. J.; McCurry, J. D.; Seeley, S. K. Comprehensive Two-Dimensional Gas Chromatography with a Simple Fluidic Modulator. *American Laboratory*. 2006, pp 24–26.

(72) Krupčík, J.; Gorovenko, R.; Špánik, I.; Sandra, P.; Giardina, M. Comparison of the Performance of Forward Fill/Flush and Reverse Fill/Flush Flow Modulation in Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2016**, *1466*, 113–128. https://doi.org/10.1016/j.chroma.2016.08.032.

(73) Griffith, J. F.; Winniford, W. L.; Sun, K.; Edam, R.; Luong, J. C. A Reversed-Flow Differential Flow Modulator for Comprehensive Two-Dimensional Gas Chromatography.

J. Chromatogr. A **2012**, *1226*, 116–123. https://doi.org/10.1016/j.chroma.2011.11.036.

(74) Sepsolve. INSIGHT ™ Flow Modulator; 2021.

(75) Tranchida, P. Q.; Purcaro, G.; Visco, A.; Conte, L.; Dugo, P.; Dawes, P.; Mondello, L. A Flexible Loop-Type Flow Modulator for Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2011**, *1218* (21), 3140–3145. https://doi.org/10.1016/j.chroma.2010.11.082.

(76) Harynuk, J.; Górecki, T. Comprehensive Two-Dimensional Gas Chromatography in Stop-Flow Mode. *J. Sep. Sci.* **2004**, *27* (5–6), 431–441. https://doi.org/10.1002/jssc.200301649.

(77) Aloisi, I.; Schena, T.; Giocastro, B.; Zoccali, M.; Tranchida, P. Q.; Caramão, E. B.; Mondello, L. Towards the Determination of an Equivalent Standard Column Set between Cryogenic and Flow-Modulated Comprehensive Two-Dimensional Gas Chromatography. *Anal. Chim. Acta* **2020**, *1105*, 231–236. https://doi.org/10.1016/j.aca.2020.01.040.

(78) Stilo, F.; Gabetti, E.; Bicchi, C.; Carretta, A.; Peroni, D.; Reichenbach, S. E.; Cordero, C.; McCurry, J. A Step Forward in the Equivalence between Thermal and Differential-Flow Modulated Comprehensive Two-Dimensional Gas Chromatography Methods. *J. Chromatogr. A* **2020**, *1627*, 461396. https://doi.org/10.1016/j.chroma.2020.461396.

(79) Seeley, J. V.; Schimmel, N. E.; Seeley, S. K. The Multi-Mode Modulator: A Versatile Fluidic Device for Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2018**, *1536*, 6–15. https://doi.org/10.1016/j.chroma.2017.06.030.

(80) Aloisi, I.; Giocastro, B.; Ferracane, A.; Salerno, T. M. G.; Zoccali, M.; Tranchida, P. Q.; Mondello, L. Preliminary Observations on the Use of a Novel Low Duty Cycle Flow Modulator for Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2021**, *1643*, 462076. https://doi.org/10.1016/j.chroma.2021.462076.

(81) Watson, N. E.; Bahaghighat, H. D.; Cui, K.; Synovec, R. E. Comprehensive Three-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry. *Anal. Chem.* **2017**, *89* (3), 1793–1800. https://doi.org/10.1021/acs.analchem.6b04112.

(82) Watson, N. E.; Prebihalo, S. E.; Synovec, R. E. Targeted Analyte Deconvolution and Identification by Four-Way Parallel Factor Analysis Using Three-Dimensional Gas Chromatography with Mass Spectrometry Data. *Anal. Chim. Acta* **2017**, *983*, 67–75. https://doi.org/10.1016/j.aca.2017.06.017.

Chapter 3

Principles of chemometrics

3.1 Introduction

It is difficult to define a proper date for chemometrics' birthday: modern chemometrics is the result of a gradual approach of the chemistry to sciences as mathematical and statistical to select or design optimal measurement procedures, while providing maximum relevant chemical information by analyzing chemical data.[1]

One of the fathers of chemometrics is the food chemist William Sealy Gosset. Since the company where he was employed did not allow him to publish the outcomes of his researches, he assumed the pseudonym Student, considering himself as a modest contributor in the field.[2] He is mostly known for his work, published in 1908, describing the probability distribution commonly known as Student's distribution.[2]

However, despite the important role played by some fathers of the discipline in the early XXth century, the modern chemometrics is a younger discipline: Svante Wold, a Swedish organic chemist, used the term chemometrics in a grant application for the first time in 1972 and, together with the US analytical chemistry Bruce Kowalski, sent a letter to Analytical Chemistry in 1974 proposing to officially use this term to identify a scientific discipline aiming to extract useful chemical information from complex chemical data.[2–4] Mathematics and statistics offer to chemistry a toolbox including both univariate methods, considering one variable at a time, and multivariate methods, taking into account the inter-correlation between chemical variables, and allowing a more complete exploration and interpretation of data structures.[2] These tools assist the analyst in select the right experiments to perform, explore chemical data, build optimal models and interpret the final models,[4,5] always remembering that a chemometrician is first of all a chemist, and the data analyst should always have a complete understanding of the nature of the analytical data and a deep awareness of the problem under study.[2] Different chemical data require in fact different processing strategies.[2]

In the late 1980s and 1990s, the diffusion of more user-friendly computers lead to the incorporation of complex algorithms into packaged software (*e.g.,* PLS-Toolbox, Pirouette,

XLSTAT, SIMCA etc.) and the word chemometrics widespread.[3] That moment is considered, by "hard" chemometricians, as a crucial point: more people heard about chemometrics and more wanted to use these methods, but without spend time in learn about computational and statistical basis of chemometrics methods. Chemometrics has become domain of people who want this methods immediately, through a fast and easy learning route and the rapid expansion of chemometrics packages coincided with a slow decline in dedicated chemometrics expertise.[3]

For this reason, this chapter of introduction to chemometric is included to provide an overview of the most common methods used within the GC×GC and food-omics field, ordered in a rational workflow from the design of experiment to the creation of classification and regression models.

Indeed, GC×GC(-TOF MS) generate high/multi-dimensional data files that, combined with an ideally large number of samples, lead GC×GC users to adopt robust chemometrics tools to access and extract reliable information.[6] Chemometrics can be used for different purposes within GC×GC-TOF MS foodomics dataset, including: (a) design of experiments, necessary to generate high quality-raw data; (b) preprocessing, intended both in terms as GC×GC signal preprocessing and data preprocessing; (c) showing simple trends by univariate methods; (d) exploring complex data matrices by exploratory analysis and pattern recognition; and (e) modelling classification and discrimination by supervised methods.[6–8]

3.2 Design of Experiment

3.2.1 Introduction

Design of experiment (DOE), or experimental design, is a multivariate approach that aims at maximizing the quality of information extracted from a chemical system/process while minimizing the experimental efforts.[9] This concept was introduced in 1935 by Ronald A. Fisher,[10] which implemented the concept of a rational planning and selection of the experiments to perform, in order to obtain the best knowledge of the system under study.[9]

This strategy is based on the idea that each experimental system (*e.g.,* chemical reaction, instrumental analysis, extraction procedure etc.) is characterized by variables (*i.e.,* factors that can be changed and set at specific values) and responses (*i.e.*, measurable quantities indicating the result of an experiments).[9] Variables can be qualitative or quantitative, and usually they are not totally independent from each other, making it necessary an approach (*i.e.,* DOE) able to consider and study interaction among variables.[9,11,12] This optimization approach is opposed to the one-variable-at-a-time (OVAT) strategy, which optimize a system changing one variable at a time while keeping constant all the other, in order to study separately the effect on each variable. The problem of this approach is that interactions among variables are totally missed and only a small part of the experimental domain is explored although, with an increasing number of variables, the number of required experiments grows fast.[9,11,12]

The usual workflow followed during an experimental design is the following:

- Define the goal of the experiment.
- Define the variables/factors to investigate. All variables having an effect have to be selected, while variables that are not selected have to be kept constant ad a fixed level in all experiments; however, in case of doubt about some variables, it is always better to include a few extra variables at the beginning than adding one variable later.[13]

- Plan the experiment. It is necessary to define the range of investigation for each variable and the model to be applied, depending on the scope of the study (screening or optimization).[9,11,12]
- Perform the experiment.
- Analyze results. Information extracted from the model are transformed into logical conclusions. If a single experimental design does not lead to the solution of a problem, information obtained are used to reformulate the problem (by removing non-significant variables, redefine the experimental domain etc.) and to perform a new design of experiment.[11]

Experimental designs can be divided into two main categories: (a) experimental designs for independent variables, where each variable can be set at any value inside its range regardless of values takes by other variables; and (b) mixture design, used for mixture components, where the implicit constraint is that the sum of all the component/variables must be 1 (*i.e.,* 100%) and what matter is the proportion of each component in relation to others.[11,13] In this chapter mixture designs are not discussed, while the focus will be on experimental designs for independent variables, by distinguishing between screening designs and optimization designs.

3.2.2 Screening designs

Screening designs are usually used to detect the relevant variables in a system, or to remove the non-relevant ones; for this reason, if a variable is already known to be relevant, it should be not included in this study, to not "cover" the effect of the others.[9,11]

Common screening designs are (a) Plackett-Burman design, used to screen the linear effects of a large number of factors with a limited number of experiments without including interactions between variables; and (b) factorial design (full or fractional), which is not properly a screening design, but it is simple and therefore often used for this purpose.[9,11]

3.2.2.1 Plackett-Burman design

Plackett-Burman design, introduced in 1946,[14] is the most efficient approach when the aim is to determine the main effect of a large number of variables with a small number of experiments.[9] The linear terms of the postulated model are computed as shown by **Equation 3.1**, where Y is the response, k the number of variables, b_0 is the constant term and b_i are the coefficients of the variables.

Equation 3.1 $$Y = b_0 + \sum_{i=1}^{k} b_i X_i$$

Two levels are set for each variable, coded as -1 and +1 (or simply – and +), and the number of experiments (N) is equal to the first multiple of 4 greater than the number of variables (k): it means that, differently from other designs where a precise number of variables correspond to a defined number of experiments, here N experiments can be used for a different number of variables, up to N-1.[9]

The building of the experimental matrix for a Plackett-Burman design is quite simple: the first row of the experimental matrix can be found in literature (**Table 3.1**) while the second row is built taking the last element of the first row and making it the first of the second row, then copy all the remaining elements of the first row shifted by one position to the right. Third row, fourth row, etc. are built by repeating the same procedure, starting from the previous row.[9,11] The final model matrix is obtained by adding a first column of + to compute the constant term b_0.

Table 3.1. First row of the experimental matrix for Plackett-Burman designs of 8,12,16,20 and 24 experiments.[9]

N	First row																						
8	+	+	+	-	+	-	-																
12	+	+	-	+	+	+	-	-	-	+	-												
16	+	+	+	+	-	+	-	+	+	-	-	+	-	-	-								
20	+	+	-	-	+	+	+	+	-	+	-	+	-	-	-	-	+	+	-				
24	+	+	+	+	+	-	+	-	+	+	-	-	+	+	-	-	+	-	+	-	-	-	-

Although Plackett-Burman design allows to investigate N-1 variables with only N experiments, for the experimental design it is recommended to add some replicated experiments to estimate experimental variability and validate the model.[9,11] Performing additional experiments in replicates, introduces more degrees of freedom (DoF), that are calculated as indicated in **Equation 3.2**, where *N* is the number of experiment and *p* the number of model coefficients .

Equation 3.2 $$DoF = N - p$$

Moreover it is a good practice to perform the experiments in a random order to avoid systematic errors or drifts.[9]

Finally, it is worthy of mention that this simple and rapid approach, although useful to select the most important variables to investigate, does not detect interactions among them. Therefore, it might happen to draw misleading conclusions if a main effect is masked by an interaction.[9]

3.2.2.2 Full Factorial design

The simplest type of experimental design that allows to understand the effect of the variables and their interaction on the response is the Full Factorial design.[9] It requires a number of experiments equal to 2^k, where *k* is the number of variables under study.[12] Variables can be both quantitative and qualitative, and they are explored at two levels as in the Plackett-Burman design.[11] From a geometrical point of view, as shown in **Figure 3.1**, a Full Factorial design explores the corners of a square, a cube of a hypercube, depending on the number of variables *k*.[11,12] However, it is recommended to include three to four center experiments to minimize the risk of missing a non-linear relationships in the middle of the intervals and to determine confidence intervals.[13]

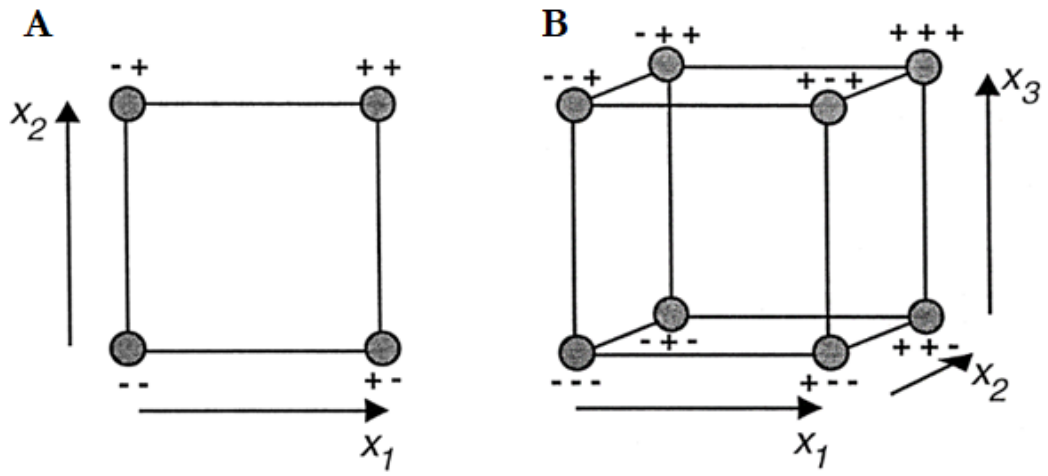

Figure 3.1. The experiments in a Full factorial design (A) with two variables and (B) with three variables.[13]

A general equation for a model obtained by a two-level Full Factorial design is **Equation 3.3**, where *Y* is the response, *k* the number of variables, b_0 is the constant term, b_i is the coefficient of the variables and b_{ij} is the coefficient of the binary interactions.

Equation 3.3 $$Y = b_0 + \sum_{i=1}^{k} b_i x_i + \sum_{1 \le i < j}^{k} b_{ij} x_i x_j$$

Full factorial design is simple and useful, but the number of experiments exponentially grows with the number of factors (*e.g.*, 5 factors investigated mean 32 experiments, 6 factors mean 64 experiments etc.). An option to reduce the number of experiments is to apply a Fractional Factorial design.[9]

3.2.2.3 Fractional Factorial design

In a Fraction Factorial design, the number of experiments is equal to 2^{k-p}, where *k* is the number of the variables and *p* the size of the fraction: the fraction is defined by the formula $(\frac{1}{2})^p$, therefore p=1 means $\frac{1}{2}$, p=2 means $\frac{1}{4}$, p=3 means $\frac{1}{8}$ and so on.[9,13] A graphical visualization of a 2^{3-1} design is given by **Figure 3.2**, where the number of experiments has been reduced by half. The figure shows how the Full Fractional selects a limited number of experiments, without investigating the interactions between all the variables, while covering as much as possible of the experimental domain, in this case with experiments having the form of a tetrahedron.[13]

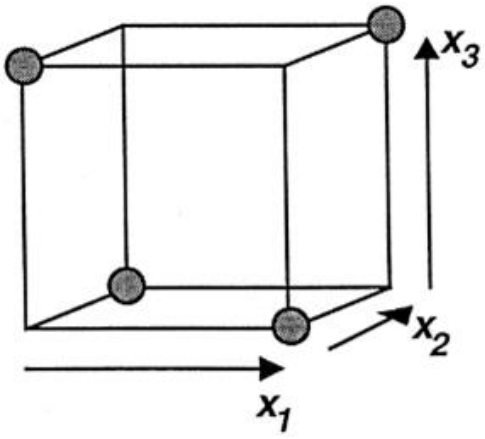

Figure 3.2. Distribution of the experiments in a 2^{3-1} Fractional factorial design.[13]

Fractional Factorial designs investigate many variables with a reduced number of experiments: less information is gained compared to full factorial designs, at the price of a contamination of the main effects by the interactions' effects.

3.2.3 Response surface designs

Response surface methodologies are the most used approaches to optimize a process and determine an optimum: for this purpose, it is necessary that their polynomial function contains quadratic terms.[13] Response surface designs, moreover, are a good way to graphically illustrate the relation between couple of experimental variables and the response.[13]

Theoretically, a three level (*i.e.*, +1, 0, -1) Full Factorial design is the best for this purpose but, being for this model the number of experiment (N) equal to 3^k, it is useless if applied to a process including more than 3 variables. For this reason, the most common response surface design is the Box-Behnken, the Central Composite and the Doehlert. They are incomplete versions of a full factorial design differing for the selection of the experimental points and fitting experimental data to a second-degree polynomial model defined by **Equation 3.4**[9]

Equation 3.4. $$Y = b_0 + \sum_{i=1}^{k} b_i x_i + \sum_{i=1}^{k} b_{ii} x_i^2 + \sum_{i<j}^{k} \sum b_{ij} x_i x_j + \varepsilon$$

where Y is the response, k the number of variables, b_0 is the constant term, b_i is the coefficients of the variables, b_{ij} is the coefficient of the quadratic effect, b_{ij} are the coefficients of the binary interactions and ε is the residual term.[13]

3.2.3.1 Box-Behnken design

The Box-Behnken design is characterized by rotatability (*i.e.*, the precision of the response estimation is equal in all directions) and it is defined as quasi-orthogonal, being zero the covariance among the coefficient for most terms and almost zero for the others.[13] The number of experiment required is defined by **Equation 3.5**.[15]

Equation 3.5. $$N = 2^k(k-1) + C_0$$

where k is number of factors and C_0 is the number of replicates at the central point.

Figure 3.3 shows the actual distribution of the experimental points in a design with 3 variables: the extreme points of the experimental domain are not explored.[13,15] It means that Box-Behnken design is not totally suitable in case the optimum falls in extreme regions.[13,15]

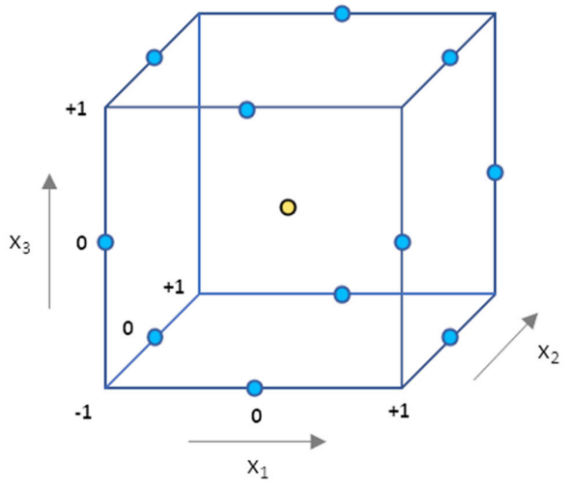

Figure 3.3. Experimental points of a Box-Behnken design with 3 factors.[9]

3.2.3.2 Central Composite design

The Central composite design is characterized by rotatability and orthogonality and it is the most used for fitting quadratic response models.

The number of experiments required is defined by **Equation 3.6**.[9]

Equation 3.6. $$N = 2^k + 2k + C_0$$

where k is number of factors and C_0 is the number of replicates at the central point.

The Central Composite design explore a cubical domain and that it is the result of the combination between a Factorial Design, exploring the extreme point of the domain, and a Star Design, exploring the central points.[11,12] The two most common type of Central Composite design are the Face Centered design (**Figure 3.4A**) and the Circumscribed Central Composite design (**Figure 3.4B**), where the "arms" of the star are longer and correspond to the square root of the variables number.[11,12] In the latest, all points are equidistant from the center, and the domain explored is spherical.[12]

Both designs include (a) levels -1 and +1 typical of a Factorial design; (b) central points, having coordinates 0,0..0, providing an estimation of pure error and contributing to estimate the quadratic terms; (c) star points, exploring different combination of the three levels -1,0,+1 in the Face Centered design, while including two additional levels (-α and +α, named *axial points*), whose value corresponds to the square root of the variables, in the Circumscribed Central Composite design.[11–13]

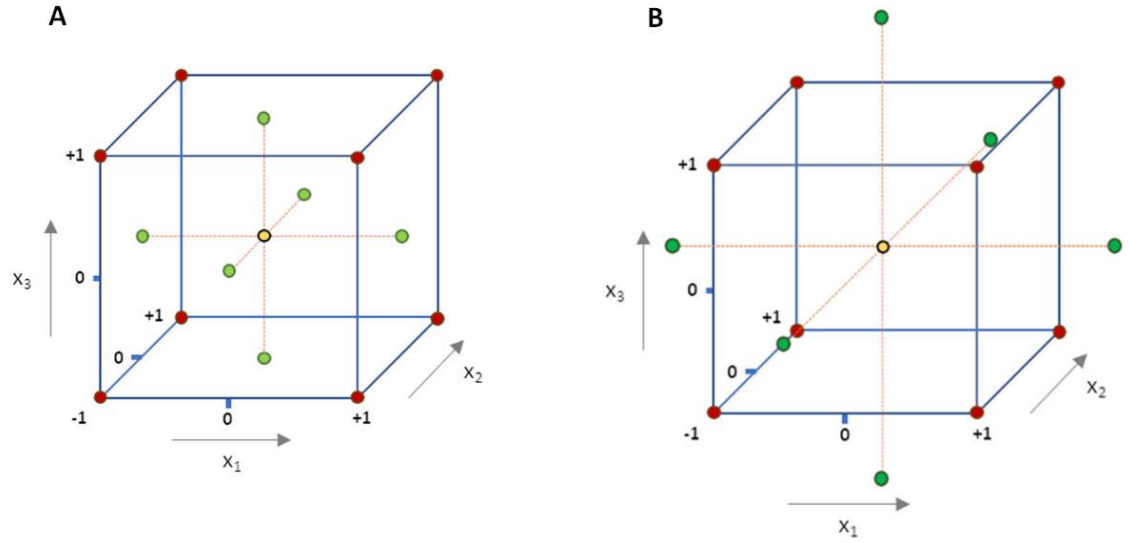

Figure 3.4. Experimental points of (**3.4A**) a faced centered and (**3.4B**) a circumscribed central composite design for 3 factors. Green circles represent the star points. [9]

3.2.3.3 Doehlert design

The Doehlert design satisfies the rotatability only for k=2, and it is not completely orthogonal, although the covariance among the coefficients of the model is minimal.[9]

The number of experiments required is defined by **Equation 3.7**.[9]

Equation 3.7. $$N = k^2 + k + C_0$$

where k is number of factors and C_0 is the number of replicates at the central point.

Interestingly, the number of experiments is smaller than for a Box-Behnken design or a Central Composite design, nevertheless the Doehlert design too contains coefficients for linear terms, interactions and quadratic terms.[11,12] It explores an hexagonal domain: **Figure 3.5** shows the actual distribution of the experimental points in a design with two variables (**Figure 3.5A**) and three variables (**Figure 3.5B**).[13] **Figure 3.5** also highlights an attractive feature of Doehlert design that makes it versatile: it can be extended both in terms of variables and levels, adding some experiments to the already performed ones when the choice of factors or the domain is unclear.[9,13]

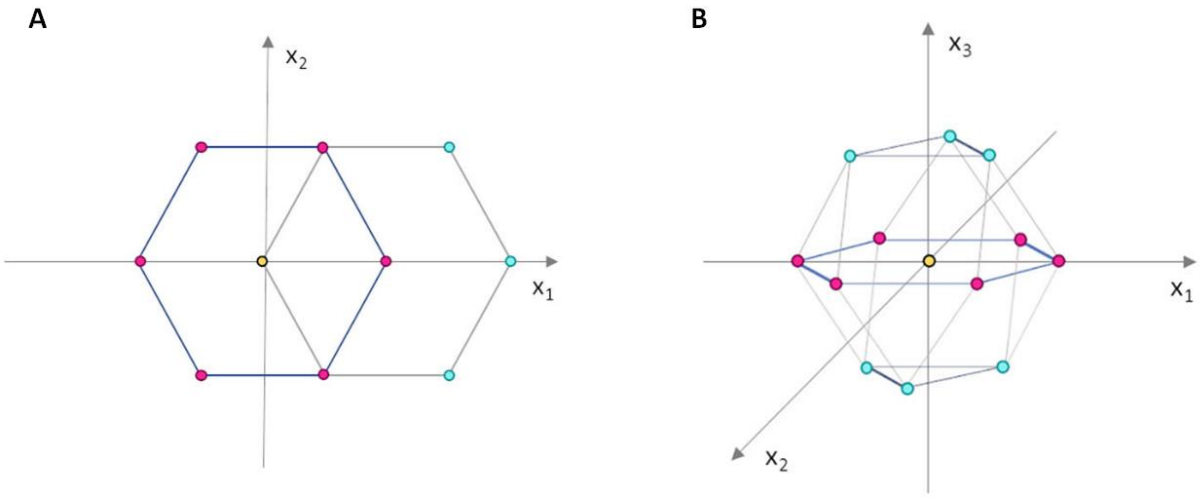

Figure 3.5. Experimental points for (**3.5A**) a 2 factors Doehlert design and (**3.5B**) a 3 factors Doehlert design. Pink circles represent the starting 2-factor design, while light-blue circles are the extensions. [9]

3.2.4 Evaluation of the model

The evaluation of the model can be made with different approaches: residual analysis, analysis of variance (ANOVA), cross validation or by validating the model with an external test set.[9,13]

Here the first approach is examined: residuals represent the difference between the experimental and the fitted values, and the parameters considered to evaluate the model are the explained variance and the significance of coefficients.

- Explained variance. It corresponds to R^2 (*i.e.*, determination coefficient of the regression model normalized to its degree of freedom) and indicates the percentage of data variability explained by the model.[9] The acceptable value for R^2 depends on the aim of the optimization, but usually values ≥ 0.8 are considered good for chemical data, while for biological data values > 0.7 are considered acceptable.[13]
- Predicted variance. It corresponds to Q^2 and it corresponds to the fraction of the total variation of the response that can be predicted in the model. As for the explained variance the acceptable value depends on the nature of the data, values ≥ 0.5 are considered acceptable for chemical data, while values > 0.4 are considered sufficient for biological data.[13]
- Significance of coefficients. It indicates the statistically significant variables and the level of confidence. A p value is associated to each coefficient: smaller the p value higher the probability that the coefficient is significant (*e.g.*, $p<0.05$ corresponds to a confidence

level of 95%, $p<0.01$ corresponds to a confidence level of 99%, $p<0.001$ corresponds to a confidence level of 99.9%).[9]

3.3 Preprocessing

The optimization of the experimental conditions is followed by the analytical experiments, whose final step is data acquisition and then results interpretation. However, to move from raw acquired data to cleaned data ready for data elaboration, the application of some preprocessing action is often an important (or even mandatory) step.

However, the term preprocessing can be referred to two different categories of treatments, depending on the object: (a) signal preprocessing, typical of the technique of analysis (*i.e.*, GC×GC-MS) and in this case mainly directed to the correct acquisition, interpretation, and alignment of the analyses; and (b) data preprocessing, mostly aiming to correct signals' intensity.

3.3.1 Signal Preprocessing and Template creation

3.3.1.1 Signal preprocessing

Signal preprocessing is a step with a medium-to-low impact on data interpretation time. GC×GC-MS commercial software implements a basic signal preprocessing package with many functions performed automatically/semi-automatically and most users usually apply the default settings.[6] The most common steps of signal preprocessing are here listed and described.

Phase correction. During the rasterization of GC×GC data is common that the initial datapoint of each ^{2}D chromatogram in the image corresponds to the time that the modulator released into the ^{2}D column, and the vertical axis of the image indicates the retention time in the ^{2}D column. However, when the initial data point is not synchronized as desired with the modulation cycle, the phase correction is necessary. This operation consists in a shift of the data to align the start of each modulation cycle with the initial part of each image column or, when wrap-around phenomena occur, the synchronization is preferably based on the hold-up time of the compound with the shortest ^{2}D retention time, to easily individuate peaks eluting during the void time of the following modulation.[16]

Baseline correction. Baseline, under controlled conditions, consists primarily of the steady-state standing-current baseline of the detector and unresolved complex mixture, and baseline fluctuations are generally due to low frequency detection noises and systems fluctuations (*i.e.*, temperature, pressure etc.).[6] It means that, in a typical GC×GC analysis, each datapoint value results from the sum of (a) the signal due to the presence of the detected compound(s); (b) a non-negative baseline value present even where there is no sample compound detected; and (c) the signal due to random noise fluctuations.[16] A baseline correction is thus required since it may affect chemometric analysis or, quantification, and has the purpose of separate analyte(s) signals from noise signals and baseline.[7,8]

Two general methods exist to estimate the baseline: the estimation of the baseline around each individual peak and the estimation of the baseline across the data comprehensively. Schmarr and others have estimated the baseline around each datapoint by using the rolling ball algorithm, moving a sphere across the underside of the 3D surface defined by the datapoint values.[17,18] When the baseline is comprehensively estimated, instead, the two main approaches are: (a) curve fitting methods, where a polynomial function is fitted to the baseline, and its contribution is consequently subtracted from the original signal, as asymmetric least squares (ALS) method; and (b) baseline

modelling by multivariate resolution methods, as multivariate curve resolution-alternating least squares (MCR-ALS) and parallel factor analysis (PARAFAC).[6,7,19]

Peak detection. Peak detection and reconstruction is based on retention times, S/N and MS library matches, and it is usually associated with a deconvolution step, which aims to identify isotopic peaks corresponding to the same compound, to remove redundant information and create a simplified data matrix.[6,8]

Indeed, the association of each single peak to a single "object" defined by the detection algorithm (e.g., GC Image™ software names these objects as blobs) is not obvious: *over-segmentation* is the error defined by the detection multiple objects that should be detected as a single peak, while *under-segmentation* occurs when multiple analyte peaks are detected as one single object, requiring the unmixing or deconvolution, previously mentioned.[20] The creation of objects, ideally corresponding to single peaks, allows the computation of important statistical features named *metadata* associated to them, as response measurement, association to unique coordinates (*i.e.,* ^1t_R and ^2t_R), measurement of the peak-width, ratio of the tailing and fronting half-widths to evaluate symmetry and so on.[16]

Analyte identification. An important metadata associated, when possible, to the objects composing an analysis is analyte identification. It is automated or semi-automated in many software and it is the result of a double channel of identification:

- Identification by retention index (RI). RI are used to convert retention times into system-independent constants, and they are the result of the interpolation of a chemical compound retention time between adjacent *n*-alkanes. RI were introduced by Kovats in 1958 and they are sufficiently independent from chromatographic parameters (*e.g.,* column length, film thickness, diameter, inlet pressure), while dependent from the stationary phase of the capillary columns used.[21] Kovats RI however, is only suitable for isothermal analyses: in 1963 it was updated into a more general form to include also temperature programmed analyses by Van den Dool and Kratz.[22] This generalized RI is calculated as indicated in **Equation 3.8**.

  Equation 3.8. $$I_i = 100 * (n + \frac{t_i - t_n}{t_{n+1} - t_n})$$

- Identification through spectral data. It is possible in case of a multichannel detector coupled to GC×GC, the most common of which is MS. In this case the identification depends on the comparison between the analyte(s) electron ionization (EI) mass spectra with those loaded on libraries through an automatic algorithm (*e.g.*, NIST).[6,8,19]

3.3.1.2 Data alignment and Template creation

Data alignment is a crucial step necessary to eliminate retention times shifts caused by system fluctuations that can affect subsequent chemometric analysis.[7] Alignment methods use different approaches, the main ones are: (a) time warping and parametric time warping algorithms, that stretch and shift each chromatogram in order to match a reference chromatogram (ChromaTOF™ software); (b) tile-based approach, based on the definition of tiles, which cut chromatograms into sub-pieces matched across chromatograms (see Synovec's group researches[23–25]); (c) peak features and peak region features based alignment, where metadata (*i.e.*, ^1t_R, ^2t_R, I^T, MS) from each feature are included in a template, which is the applied over all chromatograms searching for positive matches and eventually updated following polynomial geometrical transformations (GC Image™ software).[6,8] A detailed discussion about features used in GC×GC-MS investigations, and their role in affecting signal preprocessing, is included in the review at **Chapter 4.1**.

In particular, the straightforward concept of template matching based on pattern recognition algorithms to align GC×GC data is implemented by GC Image™ software package.[26] Template matching is a powerful extension of the traditional approaches of retention time windows and marker peaks to identify patterns of peaks in multidimensional separations. A template includes patterns of blobs (*i.e.,* objects, corresponding to peaks) with their complete set of metadata resulting from one or several analyzed chromatograms, and template matching algorithm establishes as many correspondences as possible between objects in the template and 2D-peaks in the selected analysis.[26] After matching, all *metadata* are copied from the template to the analysis, consequently, all the matched objects in the analysis are identified through the template.

3.3.2 Data Preprocessing

After data acquisition and signal preprocessing, data are organized in a matrix form *Xij*, where column vectors (j) are called variables and row vectors (i) are called objects or samples.[1] Data preprocessing is an important step necessary to correct signal intensity, however it is important to remember that each preprocessing method influences the information contained in the data and the successive elaborations.[4,6]

Moreover, chemometric elaborations require to have a complete matrix, without missing data or zero values. The best way to substitute missing data is to use column means or partial column means, while measurement below the limit of detection (LOD) – in many cases erroneously reported as zero – can be substituted with a fixed fraction of the LOD (*e.g.,* half of the LOD).[1]

Data preprocessing methods belong to three main categories: normalization, transformation, and scaling.

3.3.2.1 Normalization

Normalization in chromatographic analysis is a necessary step to react to unexpected signal's variations corresponding to a fluctuation of the peaks' response. The most common approach is the normalization on an internal standard,[6,27] where samples are spiked with known quantity of a non-native compound. This approach is effective for targeted studies, while it is more complex for untargeted analysis.[28,29]

Another common possibility, alternative or coupled with internal standard normalization, is the total area normalization. It consists in dividing the response of each peak by the total response measured in the chromatograms. The effectiveness of this normalization strategy, also used for untargeted studies, depends on samples composition, because it assumes that each sample contains the same number of analytes spanning the same range of concentrations.[6,19]

A more recent approach, introduced in the last years, is the probabilistic quotient normalization (PQN).[30] This methods evaluate the distribution of quotients to determine the most probable normalization factor, and it doesn't affect neither the variable correlation nor the data structure.[6,19,30]

3.3.2.2 Transformation

Transformations of variables are used for different purposes: stabilize the variance, normalize the distribution, realize more robust models.[4,6]

The most common transformations are:

- Logarithmic transformation. It is used to linearize the behavior of variables having a multiplicative effect.[4] It is defined by **Equation 3.9**.

  Equation 3.9. $$x'_{ij} = \log(x_{ij}) \;\; or \;\; x'_{ij} = \log(1 + x_{ij})$$

- Square root transformation. It is used to correct data belonging to Poisson distribution, where variances are proportional to mean values (common for biological data).[4] It is defined by **Equation 3.10**.

  Equation 3.10. $$x'_{ij} = \sqrt{x_{ij} + 0.5} \;\; or \;\; x'_{ij} = \sqrt{x_{ij} + \frac{3}{8}}$$

Other common transformations are arcsine transformation, inverse transformation, hyperbolic tangent transformation.[4]

3.3.2.3 Scaling

Scaling methods are used to make variables comparable in terms of order of magnitude and variance.[4] For this reason, they are extremely useful when successive elaborations include also models influenced by the variables variance (e.g., principal component analysis – PCA) and where variables with highest variance dominate the projection. [4,6]

The most used scaling methods are here listed.

- **Mean centering**. It corresponds to the subtraction of the mean value of each variable, namely the columns in the data matrix, from each measured variable, as indicated in **Equation 3.11**.[1,4,6,19]

  Equation 3.11. $$x'_{ij} = x_{ij} - \bar{x}_j$$

  After mean centering, the mean value of each variable is zero instead of the mean. It corrects differences between high and low abundance compounds, without modifying data variance.[4] It is usually used together with other scaling approaches.[6]

- **Range scaling**. It includes the maximum column value and the minimum column value, scaling each variable as indicated in **Equation 3.12**.[4,6]

  Equation 3.12. $$x'_{ij} = \frac{x_{ij} - min_i(x_{ij})}{max_i(x_{ij}) - min_i(x_{ij})}$$

  After range scaling all variables range between 0 and 1.

- **Autoscaling**. It is one of the most used scaling method, and it consists of a mean centering together with the use of the standard deviation as scaling factor, as indicated in **Equation 3.13**.[2,4,6,19]

  Equation 3.13. $$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$$

  Autoscaling results in a mean value equal to 0 and a variance equal to 1, making all variables having the same importance in the projection.

Other common scaling methods are maximum scaling, block scaling, unit variance scaling, pareto scaling, logarithmic scaling and logarithmic double centering.[1,4,19]

3.4 From univariate to multivariate methods

After data preprocessing, the data matrix is ready to be converted into information. The simpler methods are univariate ones, approaching one variable at a time; they are still the most used in many cases, although they usually offer a limited information potential, if compared to multivariate analysis.[2]

The most common univariate methods are graphical tools, as histograms, radar-charts, point-charts etc. They are easy to understand, simple to be performed and offer an immediate visualization. However, this short section only includes some tools that are useful as preliminary step and/or coupled to multivariate analysis, in particular normality test and Analysis of variance – ANOVA.

3.4.1 Normality test

The normal distribution is a basic requirement that must be verified since it is a precondition for the application of many multivariate methods. Frequency distribution can be evaluated visually by examining histograms shapes, but the basis of the most used group of normality tests (*i.e.,* Kolmogorov-Smirnov tests) is the evaluation of the cumulative empirical frequency distributions.[2]

One of the most effective method is the Lilliefors test.[31] It is used to assess the fitting between the empirical distribution and the theoretical one: the null hypothesis (H_0) is that, at a given significance level, they are not significantly different, while the alternative hypothesis (H_1) is that the empirical distribution is not compatible with the theoretical normal distribution.[2,31]

The test consists in ordering the values of the variable to be tested and scaling them by *autoscaling* (or Student's transformation), as indicated in **Equation 3.13**. Then the theoretical probability distribution is estimated for the values obtained and it is compared with the empirical distribution, at a given significance level, to determine the acceptance of the rejection of the null hypothesis.[2,31]

3.4.2 Analysis of variance – ANOVA

Analysis of variance (ANOVA) is a statistical method based on Fisher's F tests generally used to verify the existence/absence of significant differences between groups of data.[2,4,6]

The null hypothesis (H_0) is that all data derive from the same population, namely there is no significant difference between the groups considered, while the alternative hypothesis (H_1) is that the data belongs to different populations.[2] Consequently, the Fisher ratio is calculated as the ratio between the between-class variance and the within-class variance, and the resulting valued is compared to the F critical value at a given significance level.[2,6]

The simplest case of ANOVA test is the one-way ANOVA, when a single variable is explored; however it can be applied also when the effect of two variability sources has to be verified, in this case is called two-way ANOVA.[2] To note that this test is sensitive to the normal distribution of data, which must be verified before running an ANOVA test.

3.4.3 Parametric *vs.* non-parametric tests

ANOVA belongs to the group of parametric tests, which assume underlying statistical distribution (usually normal distribution) in the data.[32] Non-parametric tests, instead, do not rely on any distribution and thus are also called distribution-free tests.[32,33]

Parametric and non-parametric tests are like two sides of the same coin, since parametric tests often have nonparametric equivalents, as indicated in **Table 3.2,** which includes the most common tests.

Table 3.2. Common parametric and non-parametric tests.

Parametric tests	Non-parametric test
1-sample t test	1-sample Sign, 1-sample Wilcoxon
2-sample t test	Mann-Whitney test
One-Way ANOVA	Kruskal-Wallis, Mood's median test

In addition to their dependence or not on a particular distribution, parametric and non-parametric tests present different characteristics and areas of application:

- Parametric tests work on group means, while non-parametric tests work on group medians; the choice of the correct test depends on whether the center of the distribution and the area of study are better represented by a mean or a median value.[33]
- Non-parametric tests are more suitable when the sample size is very small, while parametric test can perform quite well even if the distribution is non-normal provided that sample size guidelines are achieved (*i.e.*, sample size > 20 for 1-sample t test, group size > 15 for 2-sample t test, while for ANOVA each group must be greater than 15 if the number of groups is 2-9, while the groups size requirement increase for an higher number of groups).[33]
- Parametric tests have a higher statistical power (*i.e.*, the p-value associated to a parametric test is usually lower to that associated to the equivalent non-parametric test run on the same dataset), while non-parametric tests are more robust, having fewer limitation and a broader range of validity.[32]

3.5 Exploratory analysis

Exploratory data analysis represents the first step for chemometric processing. They provide, without the need of a pre-formulated hypothesis, an overview of the data, allowing pattern recognition, outliers diagnosis and variable importance in describing the system.[8]

The most important and common exploratory analysis are principal component analysis (PCA) and cluster analysis (CA).

3.5.1 Principal Component Analysis – PCA

PCA's origin is attributable to the work of K. Pearson in 1901, even if its fundamental ideas were based on approaches named eigen-analysis well known to physicists and mathematicians even before the XX century.[2,34]

PCA is the most used explorative analysis technique and, broadly, the most used multivariate analysis approach.[19] It is based on the principle that multivariate data having high variability (*i.e.*, high variance value) underline an amount of information that can be more easily explored through a process of data reduction.[2,35] PCA works through the definition of a small number of linearly uncorrelated variables named principal components (PCs), able to replace all correlated original variables of a data matrix, thus retaining most of the information and explaining the majority of the variation in the original system.[19,34]

PCA algorithm builds, in the multidimensional space of the original data, hyperplanes that are linear combinations of the original variables, and then describe the hyperplanes through the definition of PCs.[2,34] The first principal component (PC1) is built on the maximum variance direction, preferably passing through the data centroid, while the second principal component (PC2) must be orthogonal to the first one in order to explain the larger amount of the remaining variance.[19,34] The other components and the defined likewise until the explanation of the total variance included in the system. **Figure 3.6** shows the easier example, corresponding to a system described by only two original variables.

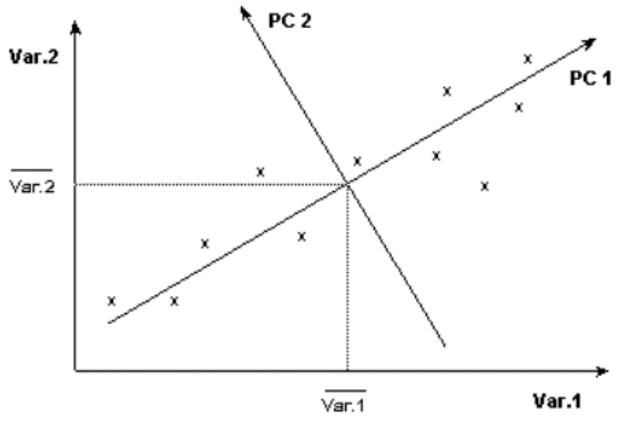

Figure 3.6. PCs definition when the system is described by two variables.

From a mathematical point of view, given a data matrix $\mathbf{X}_{O,V}$ (**Figure 3.7**), where o =1:O represents samples/objects along the rows and v=1:V represents variables along the columns, the principal components ($\mathbf{s}_{o,p}$) are calculated as linear combination of the original variables (**Equation 3.14**).

Equation 3.14

$$s_{o,1} = x_{o,1} \cdot l_{1,1} + x_{o,2} \cdot l_{2,1} + x_{o,3} \cdot l_{3,1} + ... + x_{o,v} \cdot l_{v,1} = \mathbf{x}_{o,V} \cdot \mathbf{l}_{V,1} \rightarrow \mathbf{s}_{o,1} = \mathbf{X}_{o,V} \cdot \mathbf{l}_{V,1}$$

$$s_{o,2} = x_{o,2} \cdot l_{1,2} + x_{o,2} \cdot l_{2,2} + x_{o,2} \cdot l_{3,2} + ... + x_{o,v} \cdot l_{v,2} = \mathbf{x}_{o,V} \cdot \mathbf{l}_{V,2} \rightarrow \mathbf{s}_{o,2} = \mathbf{X}_{o,V} \cdot \mathbf{l}_{V,2}$$

where $\mathbf{s}_{O,1}$ e $\mathbf{s}_{O,2}$ are new vectors resulting from the product of the original data ($\mathbf{X}_{O,V}$) and the relative coefficients $\mathbf{l}_{V,1}$, $\mathbf{l}_{V,2}$.[36]

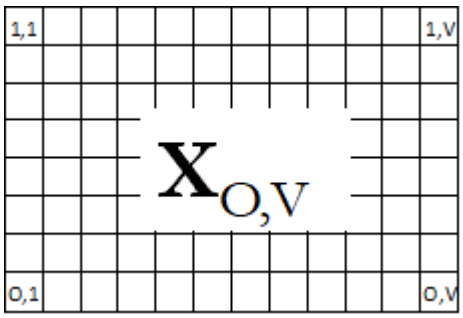

Figure 3.7. Data matrix structure.[37]

Equation 3.15 is the result of the extension of this calculation to all PCs, which is also represented in **Figure 3.8**.

Equation 3.15 $\boldsymbol{S_{O,P} = X_{O,V} \cdot L_{V,P}}$

where $\mathbf{S}_{O,P}$ is the called scores matrix e $L_{V,P}$ is called loadings matrix.[1,2,37]

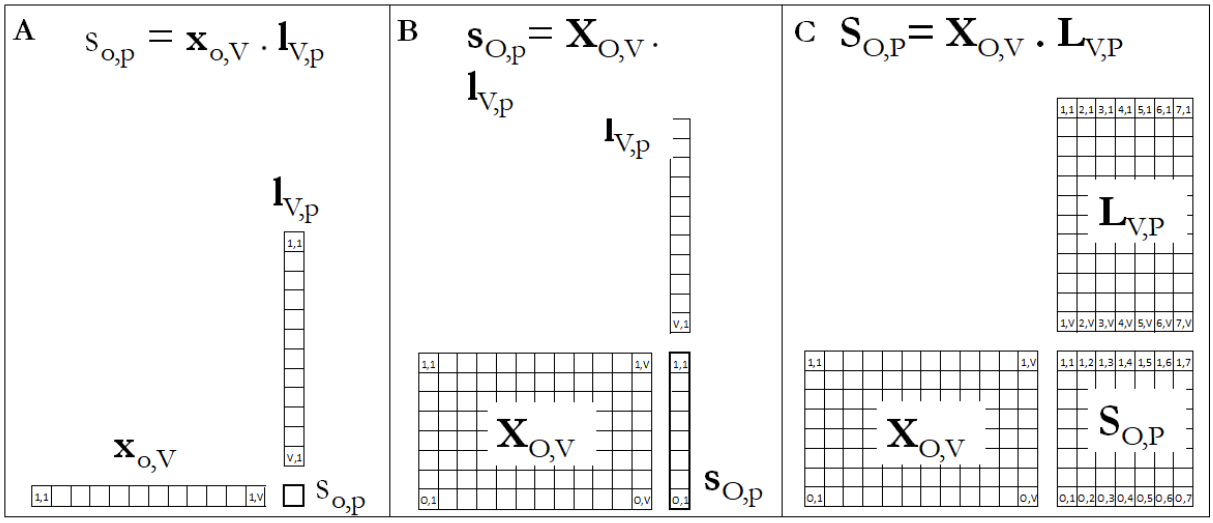

Figure 3.8. Matrix creation for (**3.8A**) one sample and one component, (**3.8B**) all samples and one component, (**3.8C**) all samples and all components.[37]

Scores, corresponding to samples (or objects or observations), are expressed in the same unit of the original data when data are mean-centred while they are dimensionless when data are autoscaled. They are visualized in a score plot, a 2D or 3D plot containing all observations represented within the plan defined by the selected PCs; the score plot shows samples distribution, their similarity and the eventual formation of natural groupings.[2,4]

Loadings, geometrically speaking, correspond to the cosines of the angles between each original variable and each component. For this reason, when the angle between a variable and a component is equal to 0, the loading is equal to 1 while when the angle is 180°, the loading is equal to -1; orthogonal directions (90°) instead, have loadings equal to 0.[19,37] In the same way, two compounds are positively correlated if the angle is small, while an angle close to 180° or 90° indicates an anti-correlation or absence of correlation, respectively.[19] A loading plot, showing the position of the original variables in the new plane defined by PCs, indicates the correlation of each variable with the selected PCs and, consequently, the correlation between each variable and each sample or clusters of samples.

Figure 3.9 shows an example of score and loadings plot resulting from a data matrix composed by 14 observations/samples and 7 variables.

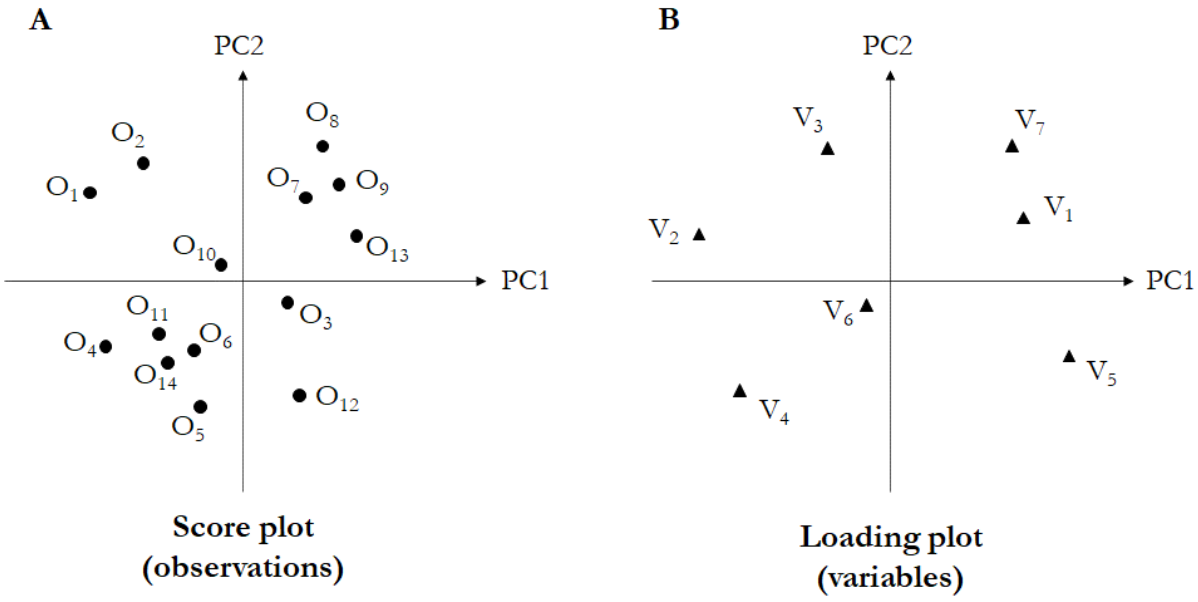

Figure 3.9. Examples of a (**3.9A**) score plot and a (**3.9B**) loading plot.

How to decide how many PCs must be considered? Different approaches exist, for example it may stop after reaching a specific value of total explained variance (*i.e.*, the sum of the variance explained by each PC considered), but this value is strongly affected by the nature of the data, their distribution and the size of the data matrix.[2,4,5] The most common approach consists in the visual analysis of the scree plot, showing on the x-axis the number of the components and on the y-axis the aggregate value of cumulative variability: usually PCs are considered until the addition of a new PC doesn't result in a significant increase the cumulative variability.

3.5.2 Cluster Analysis – CA

Cluster analysis methods are used to individuate, within the dataset, natural groupings of samples (*i.e.*, clusters) that are more similar to each other than to those in other groups/clusters.

The two types of CA are hard clustering and fuzzy clustering: in hard clustering, each object belongs to only one cluster, while in fuzzy clustering, an object can belong to one or more clusters with a specific probability.[38] Only hard clustering is matter of discussion of this chapter, and it includes hierarchical cluster analysis (HCA) and non-hierarchical (or partitional) cluster analysis.[38]

CA methods are based on the concept of similarity between objects and this similarity is measured by using different types of distances, the most common are Euclidean, Pearson, Manhattan, Mahalanobis and Minkowski.[1,4,19] Moreover, different linkage rules, used to calculate the distance among groups, have been defined: simple, average, complete and Ward's method are the most common.[1]

3.5.2.1 Hierarchical Cluster Analysis – HCA

HCA includes different methods to link the objects, as single linkage, average linkage, complete linkage, Ward's method etc.[1,4,34] They can be divided in two main categories: (a) divisive or top-down methods, less used, where all observation starts in one cluster, followed by division into smaller clusters as the hierarchy moves down; (b) agglomerative methods, more common, where clustering begins with single-object clusters and, at each step, the most similar cluster pairs are combined until all objects are included in a single cluster.[4,8,38]

The result of this process is presented in the form of a dendrogram (or tree diagram), which allows an easy and rapid detection of clusters and similar objects.[4,19,34,38] In the example shown in **Figure 3.10**, object B and E are easily identified as the most similar, followed A and H, while the final agglomeration involves the two main clusters [A,H,B,E,D,G,C] and [F,J,I].[4]

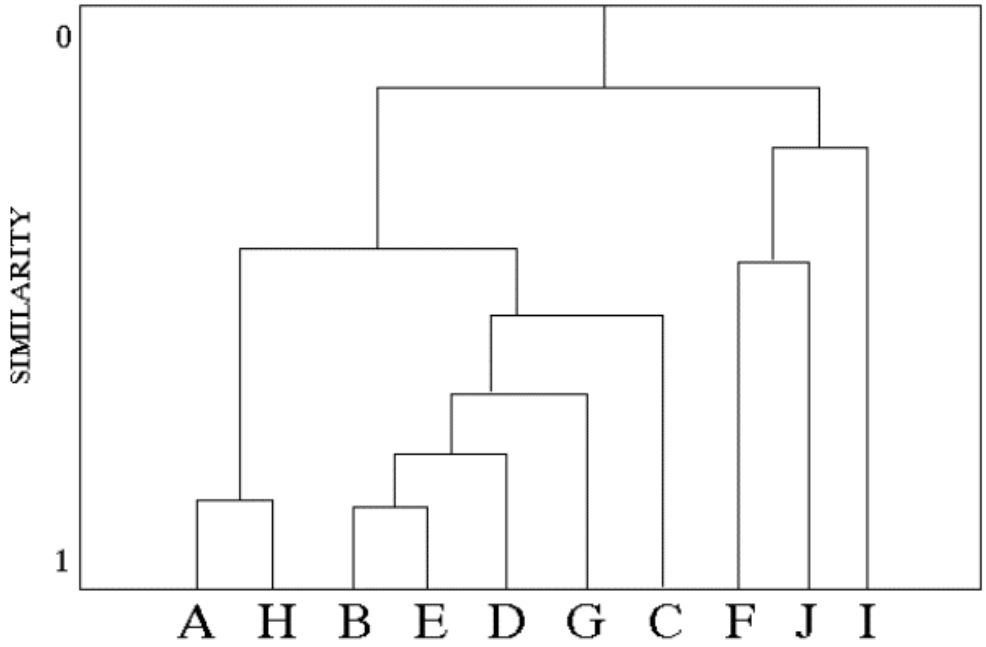

Figure 3.10. Example of dendrogram resulting from an HCA.[4]

HCA, due to its simplicity, is the most widely used clustering approach.[19] However, it works better with smaller data sets, while it is not recommended for larger ones.[38] Moreover, its main disadvantage is that it does not provide information about the compounds responsible for the resulting clustering.[19] For this reason, it is often associated to an Heatmap visualization showing the composition of the object or to a non-hierarchical clustering.

3.5.2.2 Non-hierarchical/Partitional cluster analysis

Partitional clustering includes methods are based on different techniques and they are very efficient (compared to HCA) when applied to big data set.[38]

The most common partitional clustering method is the k-means clustering. This algorithm is characterized by the following steps:

1. Random selection of the centroids, whose number (*k*) is a priori decided, representing clusters.
2. Each object of the dataset is assigned to the closer centroid by measuring its distance between all the selected centroids.
3. The mean of each cluster is calculated defining new centroids, and the assignment is reconsidered until the clustering does not change anymore; at this point the quality of the clustering is assessed by summing up the variation within each cluster.
4. Points 1-3 are repeated with a completely new random selection of the centroids and the comparison between the total variation within each class provides the best clustering.[4,19,38]

The algorithm is so repeated several times until the optimal set of centroids (*i.e.*, the one providing the lower total variation within each class) is found.[19]

Moreover, a method is used to set the optimal number (*k*) of centroids (point 1), and it is similar to the scree plot method used to define the number of PCs considered in a PCA. Usually, as *k* increases, the average variation within each class decreases, being objects grouped in a higher number of clusters. However, at a certain point, the addition of another cluster doesn't result in a significant decrease of the average variation within each class: this is the best value of k to use.[38]

3.6 Supervised techniques

Supervised techniques, usually performed after exploratory analysis, are based on the construction of a model having predictive capabilities for new data.[7,8]

Methods and algorithms belonging to this group of supervised techniques can be classified following different criteria:

- Classification *vs.* regression models.
  The application of classification or regression models depends on the type of variable that one wants to predict, discrete or continuous respectively.[8]
- Class modelling techniques *vs.* discriminant classification techniques.
  Class modelling techniques works to build class(es), defined as group(s) of individuals having one or more properties in common, described by continuous or discrete variables.[39] Class modelling techniques define independently an enclosed class space (see **Figure 3.11A**) for each class: for this reason, they can be used both when the focus is on one class and the aim is to verify samples' compliance to this class (*e.g.*, verification of food authenticity) and when the aim to distinguish samples from two or more classes.[39,40] Boundary position and shape are determined exclusively on the basis of the samples of the modeled class; it means that classes' spaces may both overlap, producing a region in which samples are recognized as compatible with more than one class, and, conversely, not cover a region of the global domain, resulting in a non-assignment for some samples.[40] For this reason, class modelling techniques are also classified as soft/fuzzy classification methods, assigning to each sample the probability of being part of one or more classes or, even, leaving some samples without assignation. The most common class modelling techniques are soft independent modeling of class analogy (SIMCA) and unequal dispersed classes (UNEQ).[39,40]
  Discriminant classification techniques delineate a delimiter between two or more classes. The main difference with modelling approaches is that discriminant techniques use the contribution of samples from all of the classes considered to create delimiter(s)

among classes.[40] The result of this process is that, as indicated in **Figure 3.11B** and **Figure 3.11C**, the different composition of one class affects the region delimiting other classes, while the modelling class space stay unvaried.[40] Moreover, in discriminant classification techniques, also known as hard classification methods, all samples are assigned to only one class and the entire domain is covered, thus all samples have an assignation to a class.[34,39,40]During time, modified discriminant strategies have been proposed, defining a maximum allowed distance for each class centroids and excluding samples too far, so helping the individuation of potential outliers.[39]

Despite their limits, because of historical and convenience reasons, discriminant classification techniques are more common than class modelling ones. The most used algorithms are linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), k-nearest neighbors (k-NN), classification and regression trees (CART), Random forest (RF), artificial neural networks (ANNs), support vector machines (SVM) and partial least squares discriminant analysis (PLS-DA).[7,34,39,40]

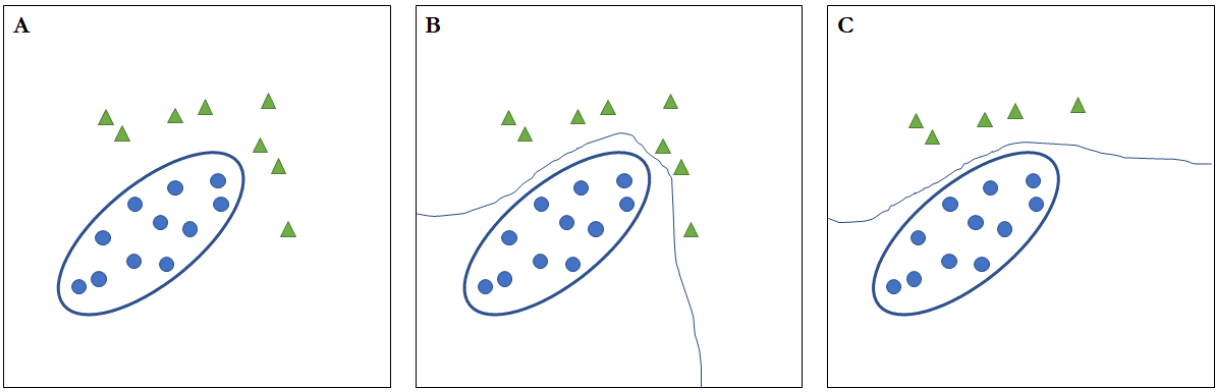

Figure 3.11. Examples of (**3.11A**) the definition of an enclosed class space by using a classification modelling approach and (**3.11B** and **3.11C**) the effect of a class composition in defining the region deliming other class(es)

- Parametric *vs.* non-parametric classification algorithms.
  Parametric, or linear, algorithms offer simple interpretation of the results, fewer parameters to optimize and fast calculation, some examples are LDA, PLS-DA and SIMCA. Non-parametric, or non-linear, algorithms instead, as k-NN, CART, RF, SVM and ANN, are more complex but usually powerful and more accurate in the prediction.[7,8,19]

Although many different algorithms are available, a common strategy can be applied for classification purposes. It could be stepped as follows:[34]

1. Build the training set. The training set is composed by all analyzed known samples used to build the model. It is fundamental to have a training set as larger as possible, and representative of the real composition of the entire population
2. Validate the model. The quality of the predictions arising from the model built with the training set must be validated by using and independent test set. A focus on validation is the object of **Chapter 3.5.1**.
3. Model application. Once a satisfactory model is obtained, it can be applied to unknown new samples to make predictions.

3.6.1 Model performance evaluation

The performance of a class model can be evaluated through different parameters, the most important of which are here described.

- Sensitivity. It is defined as the fraction of samples belonging to the modeled class that is correctly accepted by the respective model.[40] It is calculated following **Equation 3.16**, where TP stays for true positive (*i.e.*, samples modeled within the correct class), and FN stays for false negative (*i.e.*, samples belonging to a class, but modeled outside).

  Equation 3.16. $Sensitivity = \frac{TP}{TP+FN}$

- Specificity. It is defined as the fraction of samples not belonging to the modeled class that is correctly rejected by the respective model.[40] It is calculated following **Equation 3.17,** where TN stays for true negative (*i.e.*, samples correctly modeled outside a class), and FP stays for false positive (*i.e.*, samples not belonging to a class, but found within its boundaries).

  Equation 3.17. $Specificity = \frac{TN}{TN+FP}$

- Precision. It indicates the probability that a positive decision is correct and it is calculated following **Equation 3.18**.[40]

  Equation 3.18. $Precision = \frac{TP}{TP+FP}$

- Efficiency. It is the parameter used to measure the ability of the model to avoid errors during classification.[41] A useful tool to evaluate this parameter is represented by receiver operating curve (ROC). They are usually obtained by plotting the sensitivity value *vs.* the (1-specificity) value, as shown in **Figure 3.12**.[39,40]

  The red dotted diagonal line from the lower left corner to the upper right corner represents a model that provide a random assignment of samples, while curves that tend towards the upper left corner indicate efficient models.[40] The area under the curve (AUC) is often used as indicator of a model performance.[40]

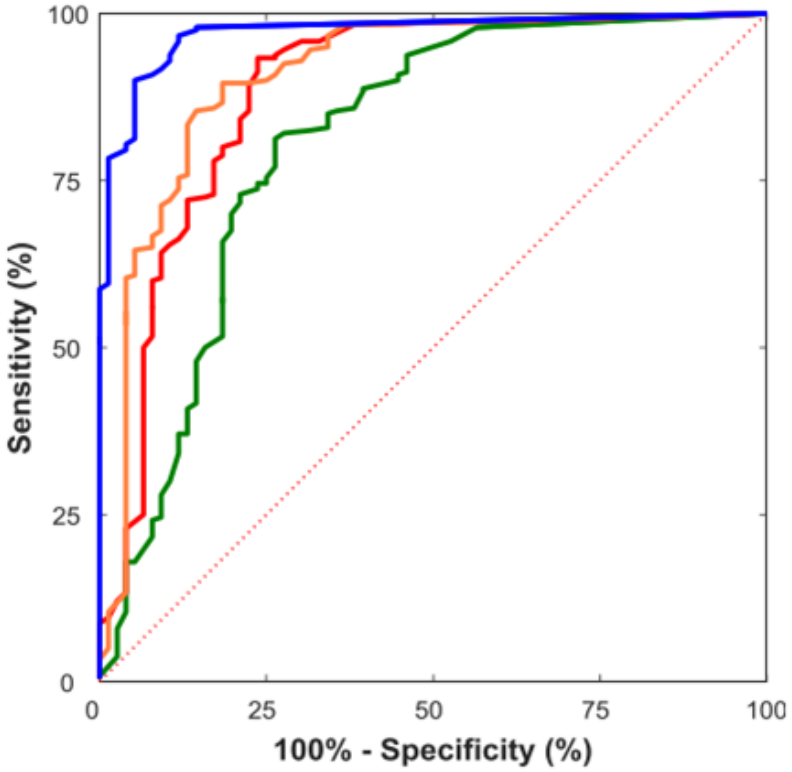

Figure 3.12. Example of ROC curves (solid lines) used to evaluate the performance of different class models. Efficiency of models associated to the curves decreases from blue to green, while the red dotted diagonal line from the lower left corner to the upper right corner represents a model that provide a random assignment of samples.[39]

3.6.2 Model validation

The validation is a crucial part of data analysis, used to evaluate performances and robustness of a model when applied to new data.[19] Indeed, it is necessary to avoid the overfitting: a situation in which a model fits well on data used to build it, being vice versa unable to perform well when it is used for predictive purposes on a new dataset.[4,6,8]

The most common validation methods are here explained:

- Training and validation set. The optimal approach, when the number of samples is adequate, data are divided into a training set, used to train the supervised method, a validation/test set, to optimize the parameters, and an independent test, to evaluate the performance of a model and its predictive power.[6,19] Validation set is built by selecting a fraction corresponding to the 10-50% of the available samples (usually 20-30%) and the selection can be random (it requires multiple runs to ensure different scenarios) or guided (by selecting a set representing of the whole population variability).[4,6,19,42]
- Cross validation (CV). It is a very common validation strategy, used as alternative for the first one especially when the number of samples is not so high.[6,8] CV splits the *N* rows of a data matrix (*i.e.*, samples) into *C* cancellation groups following a predetermined scheme, and the model is computed *C* times, using each time one of the cancellation groups as test set and the remaining samples as training set.[39] The number of cancellation groups *C* is usually between 4 and 8, but it can theoretically range from 2 to N. In this latest case CV correspond to a leave-one-out validation.

- Leave-one-out. It is the extreme case of a CV, where the dataset is divided into as many cancellation groups as the samples number.[19] Leave-one-out validation is usually characterized by optimistic results, with an over estimation of the predictive power of the supervised model.[19,39]

Other types of validation commonly used are the double CV or nested CV, and the bootstrap method.[4,6,39]

3.6.3 Linear Discriminant Analysis – LDA

Linear discriminant analysis (LDA) is the first multivariate classification technique, introduced by Fisher in 1936.[40] LDA is a linear discriminative classification technique that searches for a linear delimiter between classes of objects and use it to assign unknown samples to a class.[8] LDA can be used when more than two classes have to be discriminated and can be only applied when the number of samples is larger than the number of measured compounds.[19] Another limitation to the application of LDA, in addition to the requirement of a normal distribution, is that LDA considers that all classes have the same dispersion, so the variance-covariance matrix.[8,40] For this reason, when groups have different variance structures, it is more appropriate to use non-linear methods.[40]

The classification can be expressed both in the Bayesian form, based on the probability that an objects belong to the probability density distribution of a class, and using Mahalanobis distance.[4,8] The Mahalanobis distance is calculated for all samples to the class centers, and samples are assigned to the lowest class distance.[8] The delimiter between two classes, instead, is created by connecting the intersection points of each couple of corresponding class ellipses, having equal eccentricity and axis orientation, with a straight line.[40]

3.6.4 K-Nearest Neighbor – k-NN

K-NN is a non-parametric discriminant classification technique, and it is one of the simplest approaches for classification.[2] It is a distance-based method and the algorithm works by calculating the distance of an unknown compounds to all members of the training set.[4,34] At this point, a number *k* is chosen, and the unknown object is assigned to the class better represented among the closer *k* known objects.[34] It is useful to perform *k*-NN analysis using different values of *k*, searching for the model able to provide the best performance.[4] In some cases, it is possible that, for objects close to the delimiter, the assignation of an unknown compound could change with a different *k* selection, as shown in **Figure 3.13**. In this case the assignation of the unknown object is blue if *k* value is 3 (**Figure 3.13A**) and green if *k* value is 5 (**Figure 3.13B**).

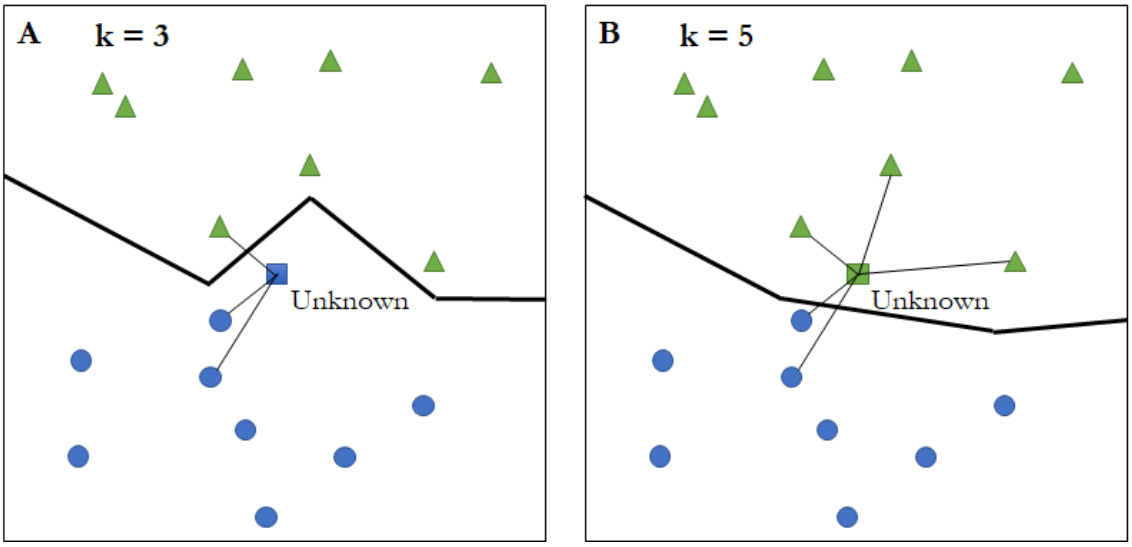

Figure 3.13. Different assignation of an unknown sample when *k* value is 3 (**3.13A**) or 5 (**3.13B**)

Distance is usually calculated by using Euclidean distance, but also Mahalanobis distance can be used.[2] *k*-NN works well in a lot of situation, it is simple, very efficient in describing complex nonlinear boundaries between the groups while it not considers the variance in the classes.[19,34] However it is influenced by the numerosity of the classes, that should be approximately equal to avoid a bias towards the classes highly represented, and, being a distance-based method, it is sensitive to eventual scaling procedures applied and to the measurement unit.[19,34]

3.6.5 Classification And Regression Tree – CART

CART is a method proposed by Breiman in 1984,[43] and it is a non-parametric discriminant classification method not requiring nor a normal distribution of data neither the equality of the within-class variance.[44] The outcome of the CART is represent by a tree, built by dividing samples from a parent node into two child nodes by following a rule base on one compound.[19] Each node represents a single compounds and branches arise from a binary answer (i.e., present/absent) or a multi-interval answer (i.e., depending on different levels of abundance of a compound in the different classes).[19]

Figure 3.14 shows the simpler example of a CART: a binary tree where objects are assigned to class 1 or 2 (*i.e.*, C_1 and C_2) depending on the value of variables x_1 and x_2. Here each node uses a different compounds but, in more complex problems resulting in branched trees, the same compound may appear more than once.[19]

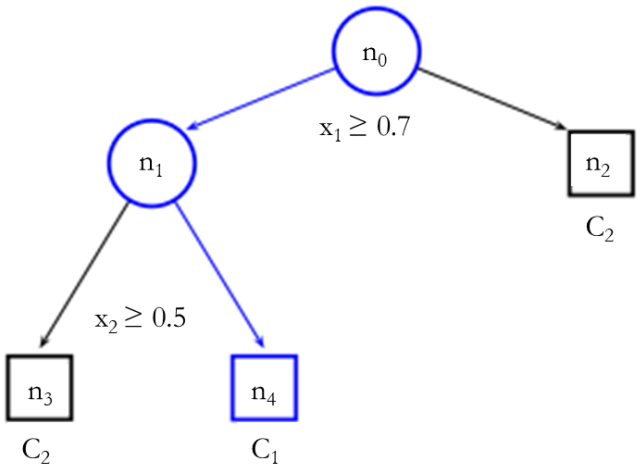

Figure 3.14. A decision tree built for a binary classification problem.[45]

CART has many advantages: it is a simple method, where each step is characterized by only one variable, it is robust to outliers presence and the visual output also provides a selection of the most important variables in the discrimination process.[4] Conversely, this method is characterized by a strong tendency of overfitting results.[19]

To address this drawback, the random forest (RF) method was proposed by Breiman in 2001[46]: RF builds a large collection of uncorrelated trees resulting from a random selection of subsets of the original samples.[19] RF is an approach more robust than CART, especially when the number of variables is much larger than the number of samples.[19]

3.6.6 Soft Independent Modeling of Class Analogy - SIMCA

SIMCA is the first class modeling technique introduced into chemometrics by Svante Wold in 1977.[47] It is a non-parametric and non-probabilistic distance-based method, where each class of the training set is modeled independently and, being a soft modeling approach, each object could belong to multiple (or no one) classes simultaneously.[4,34,40]

SIMCA models are based on the PCs, by definition the directions of maximum variance in a multivariate data space, of a PCA performed using only the samples of the category studied.[2,39,40] Then it is evaluated the number of significant PCs through a CV, and this number defines the so-called SIMCA inner space that could corresponds to a rectangle (two PCs), a parallelepiped (three PCs) or to hyper-parallelepipeds (in case of more than three PCs).[39,40] Finally, residuals (*i.e.*, the distances between each sample and the model) are evaluated in the full dimensional space, including also the so called SIMCA outer space, defined by non-significant PCs.[40] The distance from sample s to class C ($d_{s,C}$) is calculated as indicated in **Equation 3.19**, where ID is the distance in the score inner space and OD is the distance from the class model.[39]

Equation 3.19. $$d_{s,C} = \sqrt{ID_{s,C}^2 + OD_{s,C}^2}$$

The critical value of this distance, determining acceptation/rejection of a new sample, is defined by a critical value of Fisher statistics at a predetermined confidence level.[39]

SIMCA defines an enclosed space for each class, and a common method to visualize the results of SIMCA classification is represented by Coomans plots (**Figure 3.15**). In these plots the coordinates of each object correspond to the ratio of their distances from the two classes (in the simplest case) to the critical distance corresponding to the boundary of each class, which are represented by the two straight lines parallel to axes.[39] The plot area is divided in four section, respectively containing: samples accepted by class 1 (upper left rectangle); samples accepted by class 2 (lower right rectangle); samples accepted by both of the models (lower left square) and samples rejected by both of the models (upper right square).[39]

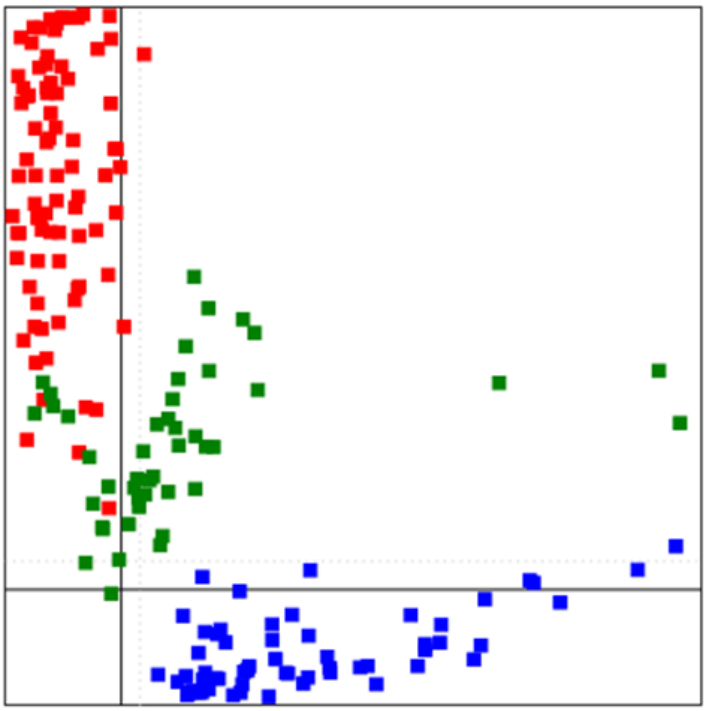

Figure 3.15. Example of Coomans plot.

SIMCA algorithm can be also used as alternative discriminant classification technique, by calculating the delimiter (usually non-linear) as the locus of points having the same distance from the models of the two classes.[40]

3.6.7 Regression methods

As for classification methods, regression methods are used for prediction, but their difference is about the type of variable that they predict: while classification algorithm are used to predict discrete values/label/classes, regression methods are about predicting continuous values and quantities.

A conventional regression problem, in terms of matrices, includes a X_{IJ} predictor matrix and a Y_{IK} response matrix, where K is equal to the number of classes.[19] Traditionally this modelling of Y is done using multiple linear regression (MLR) based on Ordinary Least Squares (OLS). The least squares solution it is calculated by following **Equation 3.20**, where b is the column vector of the regression coefficient, X the independent variable array for the calibration set objects, and **y** is the vector containing he reference responses for the same objects.[2]

Equation 3.20. $b = (X'X)^{-1}X'y$

MLR works well until X-variables are fewer than the number of samples analyzed and sufficiently uncorrelated.[2,48] However, when X contains variables resulting from multidimensional analytical (MDA) platforms, X-variables are usually a lot and often strongly correlated (especially in the case of continuous analytical signals used as predictor variables).[48]

To address this limitation, regression techniques based on latent variables (LVs) that reduce the space of the original predictors to a limited number of orthogonal components, were introduced. They are Principal Component Regression (PCR) and Partial Least Squares (PLS) regression.

3.6.7.1 Principal Component Regression – PCR

PCR satisfies the requirement for the number of objects higher than the number of predictors by using PCs as LVs: the matrix of predictors X is projected into the space defined by PCs and the significant PCs are selected.[2] Thus, although not all the original information is considered, the information contained in the system is not deformed, being PCs a linear combination of the original predictors. This transformation results in a reduction in the number of predictors, moreover the new variables are orthogonal to each other and therefore non-correlated.

PCR is a technique very efficient in many cases however, not always the directions which explain the highest variance amount are the most important in predicting a response variable.[2] For this reason a refined approach was then introduced and it is the PLS regression.

3.6.7.2 Partial Least Square regression – PLS and Partial Least Square Discriminant Analysis – PLS-DA

PLS is the most important regression method, and although historically it was proposed to handle continuous variables it can be adapted to solve classification tasks, by using categorical output variables.[19,49] In this case it is named Partial Least Square Discriminant Analysis (PLS-DA).

PLS is based on PCs as PCR but, instead of maximizing variance in the data it is focused on capturing most of the information in the data related to a response/class vector Y, in a linear way.[6] Thus, the first LV is the direction characterized by the maximum covariance with the selected response variable. The information related to the first LV is then subtracted from both the original predictors and the response, and the second LV, orthogonal to the first one, is the direction of maximum covariance between the residuals of the predictors and the residuals of the response. This approach continues for the subsequent LVs.[2]

A modification of PLS is Orthogonal-PLS (O-PLS).[6] It is based on splitting the overall variation in the data into response predictive (*i.e.*, linearly related to the class/response vector) and orthogonal (*i.e.,* uncorrelated to the response). The main benefit of O-PLS is the simpler interpretation of the model, being relevant information captured by the first LV, while irrelevant information is directly filtered out. However, it is important to specify that O-PLS and PLS have comparable prediction power.[6]

Results of a PLS can be graphically visualized on a score plot, while variable importance can be obtained from the regression coefficient.[6] Different approaches are available to select relevant variables, however Variables Importance in the Projection (VIPs) score is probably the most used.[6] This score, usually represented in form of bar chart (**Figure 3.16**) is a summary of the importance of each variable in the model.

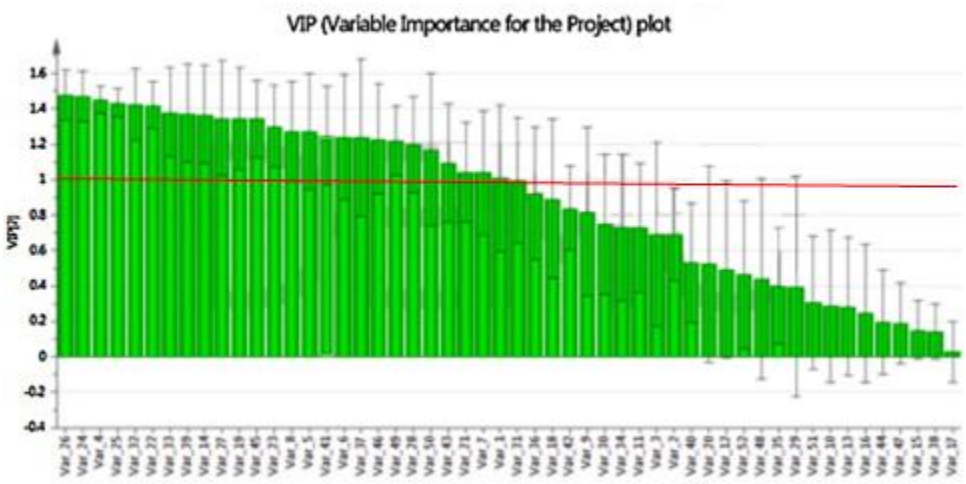

Figure 3.16. Variables Importance in the Projection (VIPs) score represented in form of bar chart.

Usually variables with VIP score (± standard deviation) ≥ 1 are considered significant and, among them, higher the VIP score higher the importance of the variable.[19]

References

(1) Héberger, K. Chemoinformatics—Multivariate Mathematical–Statistical Methods for Data Evaluation. In *Medical Applications of Mass Spectrometry*; Elsevier, 2008; pp 141–169. https://doi.org/10.1016/B978-044451980-1.50009-4.

(2) Oliveri, P.; Forina, M. Data Analysis and Chemometrics. In *Chemical Analysis of Food: Techniques and Applications*; Elsevier, 2012; pp 25–57. https://doi.org/10.1016/B978-0-12-384862-8.00002-9.

(3) Brereton, R. G. The Evolution of Chemometrics. *Anal. Methods* **2013**, *5* (16), 3785. https://doi.org/10.1039/c3ay90051g.

(4) Todeschini, R. Introduzione Alla Chemiometria. *EdiSES, Napoli* **1998**.

(5) Kjeldahl, K.; Bro, R. Some Common Misunderstandings in Chemometrics. *J. Chemom.* **2010**, *24* (7–8), 558–564. https://doi.org/10.1002/cem.1346.

(6) Stefanuto, P. H.; Smolinska, A.; Focant, J. F. Advanced Chemometric and Data Handling Tools for GC×GC-TOF-MS: Application of Chemometrics and Related Advanced Data Handling in Chemical Separations. *TrAC - Trends Anal. Chem.* **2021**, *139*, 116251. https://doi.org/10.1016/j.trac.2021.116251.

(7) Feizi, N.; Hashemi-Nasab, F. S.; Golpelichi, F.; Sabouruh, N.; Parastar, H. Recent Trends in Application of Chemometric Methods for GC-MS and GC×GC-MS-Based Metabolomic Studies. *TrAC - Trends Anal. Chem.* **2021**, *138*, 116239. https://doi.org/10.1016/j.trac.2021.116239.

(8) Peris-Díaz, M. D.; Krężel, A. A Guide to Good Practice in Chemometric Methods for Vibrational Spectroscopy, Electrochemistry, and Hyphenated Mass Spectrometry. *TrAC - Trends Anal. Chem.* **2021**, *135*. https://doi.org/10.1016/j.trac.2020.116157.

(9) Benedetti, B.; Caponigro, V.; Ardini, F. Experimental Design Step by Step: A Practical Guide for Beginners. *Crit. Rev. Anal. Chem.* **2020**, *0* (0), 1–14. https://doi.org/10.1080/10408347.2020.1848517.

(10) Fisher, R. A. The Design of Experiments. *Hafner Press* **1935**.

(11) Leardi, R. Experimental Design. In *Data Handling in Science and Technology*; Copyright © 2013 Elsevier B.V. All rights reserved., 2013; Vol. 28, pp 9–53. https://doi.org/10.1016/B978-0-444-59528-7.00002-8.

(12) Leardi, R. Experimental Design in Chemistry: A Tutorial. *Anal. Chim. Acta* **2009**, *652* (1–2), 161–172. https://doi.org/10.1016/j.aca.2009.06.015.

(13) Lundstedt, T.; Seifert, E.; Abramo, L.; Thelin, B.; Nyström, Å.; Pettersen, J.; Bergman, R. Experimental Design and Optimization. *Chemom. Intell. Lab. Syst.* **1998**, *42* (1–2), 3–40. https://doi.org/10.1016/S0169-7439(98)00065-3.

(14) Plackett, R. L.; Burman, J. P. The Design of Optimum Multifactorial Experiments. *Biometrika* **1946**, *33* (4), 305–325. https://doi.org/10.1093/biomet/33.4.305.

(15) Ferreira, S. L. C.; Bruns, R. E.; Ferreira, H. S.; Matos, G. D.; David, J. M.; Brandão, G. C.; da Silva, E. G. P.; Portugal, L. A.; dos Reis, P. S.; Souza, A. S.; dos Santos, W. N. L. Box-Behnken Design: An Alternative for the Optimization of Analytical Methods. *Anal. Chim. Acta* **2007**, *597* (2), 179–186. https://doi.org/10.1016/j.aca.2007.07.011.

(16) GC Image™. *GC Image GCxGC Edition Users' Guide*; 2021.

(17) Sternberg, R. S. Biomedical Image Processing. *Computer (Long. Beach. Calif).* **1983**, *16*, 22–23. https://doi.org/10.1109/MC.1983.1654163.

(18) Schmarr, H.-G.; Bernhardt, J. Profiling Analysis of Volatile Compounds from Fruits Using Comprehensive Two-Dimensional Gas Chromatography and Image Processing Techniques. *J. Chromatogr. A* **2010**, *1217* (4), 565–574. https://doi.org/10.1016/j.chroma.2009.11.063.

(19) Smolinska, A.; Hauschild, A.-C.; Fijten, R. R. R.; Dallinga, J. W.; Baumbach, J.; van Schooten, F. J. Current Breathomics—a Review on Data Pre-Processing Techniques and Machine Learning in Metabolomics Breath Analysis. *J. Breath Res.* **2014**, *8* (2), 027105. https://doi.org/10.1088/1752-7155/8/2/027105.

(20) Reichenbach, S. E.; Ni, M.; Kottapalli, V.; Visvanathan, A. Information Technologies for Comprehensive Two-Dimensional Gas Chromatography. *Chemom. Intell. Lab. Syst.* **2004**, *71*, 107–120. https://doi.org/10.1016/j.chemolab.2003.12.009.

(21) Kovats, E. Gas-Chromatographische Charakterisierung Organischer Verbindungen. Teil 1: Retentionsindices Aliphastischer Halogenide, Alkohole, Aldehyde Und Ketone. *Helv. Chim. Acta* **1958**, *41* (7), 1915–1932.

(22) Van den Dool, H.; Kratz, P. D. A Generalization of the Retention Index System Including Linear Temperature Programmed Gas-Liquid Partition Chromatography. *J. Chromatogr* **1963**, *11*, 463–471.

(23) Freye, C. E.; Moore, N. R.; Synovec, R. E. Enhancing the Chemical Selectivity in Discovery-Based Analysis with Tandem Ionization Time-of-Flight Mass Spectrometry Detection for Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2018**, *1537*, 99–108. https://doi.org/10.1016/j.chroma.2018.01.008.

(24) Watson, N. E.; Parsons, B. A.; Synovec, R. E. Performance Evaluation of Tile-Based Fisher Ratio Analysis Using a Benchmark Yeast Metabolome Dataset. *J. Chromatogr. A* **2016**, *1459*, 101–111. https://doi.org/10.1016/j.chroma.2016.06.067.

(25) Marney, L. C.; Christopher Siegler, W.; Parsons, B. A.; Hoggard, J. C.; Wright, B. W.; Synovec, R. E. Tile-Based Fisher-Ratio Software for Improved Feature Selection Analysis of Comprehensive Two-Dimensional Gas Chromatography-Time-of-Flight Mass Spectrometry Data. *Talanta*. 2013, pp 887–895. https://doi.org/10.1016/j.talanta.2013.06.038.

(26) Reichenbach, S. E.; Carr, P. W.; Stoll, D. R.; Tao, Q. Smart Templates for Peak Pattern Matching with Comprehensive Two-Dimensional Liquid Chromatography. *J. Chromatogr. A* **2009**, *1216* (16), 3458–3466. https://doi.org/10.1016/j.chroma.2008.09.058.

(27) Wang, Y.; O'Reilly, J.; Chen, Y.; Pawliszyn, J. Equilibrium In-Fibre Standardisation Technique for Solid-Phase Microextraction. *J. Chromatogr. A* **2005**, *1072* (1), 13–17. https://doi.org/10.1016/j.chroma.2004.12.084.

(28) Stilo, F.; Liberto, E.; Reichenbach, S. E.; Tao, Q.; Bicchi, C.; Cordero, C. Untargeted and Targeted Fingerprinting of Extra Virgin Olive Oil Volatiles by Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry: Challenges in Long-Term Studies. *J. Agric. Food Chem.* **2019**, *67* (18), 5289–5302. https://doi.org/10.1021/acs.jafc.9b01661.

(29) Stilo, F.; Cordero, C.; Sgorbini, B.; Bicchi, C.; Liberto, E. Highly Informative Fingerprinting

of Extra-Virgin Olive Oil Volatiles: The Role of High Concentration-Capacity Sampling in Combination with Comprehensive Two-Dimensional Gas Chromatography. *Separations* **2019**, *6* (3). https://doi.org/10.3390/separations6030034.

(30) Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H. Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application In1H NMR Metabonomics. *Anal. Chem.* **2006**, *78* (13), 4281–4290. https://doi.org/10.1021/ac051632c.

(31) Lilliefors, H. W. On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *J. Am. Stat. Assoc.* **1967**, *Volume 62* (Issue 318), 399–402.

(32) XLSTAT. What is the difference between a parametric and a nonparametric test? https://help.xlstat.com/s/article/what-is-the-difference-between-a-parametric-and-a-nonparametric-test?language=en_US.

(33) Minitab. Choosing Between a Nonparametric Test and a Parametric Test https://blog.minitab.com/en/adventures-in-statistics-2/choosing-between-a-nonparametric-test-and-a-parametric-test.

(34) Brereton, R. G. *Chemometrics Data Analysis for the Laboratory and Chemical Plant*; Wiley, 2003; Vol. 24.

(35) Bressanello, D. Multidisciplinary Approaches in Foodomic Studies: Analytical and Chemometric Relationships between Sensory Evaluation and Chemical Composition., 2017.

(36) Bro, R.; Smilde, A. K. Principal Component Analysis. *Anal. Methods* **2014**, *6* (9), 2812–2831. https://doi.org/10.1039/C3AY41907J.

(37) Malegori, C. What Is PCA ? Principal Component Analysis; 2019.

(38) Akman, O.; Comar, T.; Hrozencik, D.; Gonzales, J. Data Clustering and Self-Organizing Maps in Biology. In *Algebraic and Combinatorial Computational Biology*; Elsevier, 2019; pp 351–374. https://doi.org/10.1016/B978-0-12-814066-6.00011-8.

(39) Oliveri, P. Class-Modelling in Food Analytical Chemistry: Development, Sampling, Optimisation and Validation Issues – A Tutorial. *Anal. Chim. Acta* **2017**, *982*, 9–19. https://doi.org/10.1016/j.aca.2017.05.013.

(40) Oliveri, P.; Downey, G. Multivariate Class Modeling for the Verification of Food-Authenticity Claims. *TrAC Trends Anal. Chem.* **2012**, *35*, 74–86. https://doi.org/10.1016/j.trac.2012.02.005.

(41) Cuadros-Rodríguez, L.; Pérez-Castaño, E.; Ruiz-Samblás, C. Quality Performance Metrics in Multivariate Classification Methods for Qualitative Analysis. *TrAC Trends Anal. Chem.* **2016**, *80*, 612–624. https://doi.org/10.1016/j.trac.2016.04.021.

(42) Brereton, R. G.; Jansen, J.; Lopes, J.; Marini, F.; Pomerantsev, A.; Rodionova, O. Chemometrics in Analytical Chemistry — Part II : Modeling , Validation , and Applications. **2018**, 6691–6704.

(43) Breiman, L.; Friedman, J.; Stone, C. S.; Olshen, R. A. *Classification and Regression Trees*; 1984.

(44) Stilo, F.; Liberto, E.; Spigolon, N.; Genova, G.; Rosso, G.; Fontana, M.; Reichenbach, S. E.; Bicchi, C.; Cordero, C. An Effective Chromatographic Fingerprinting Workflow Based on Comprehensive Two-Dimensional Gas Chromatography - Mass Spectrometry to Establish Volatiles Patterns Discriminative of Spoiled Hazelnuts (Corylus Avellana L.). *Food Chem.*

2020, *340* (September 2020), 128135. https://doi.org/10.1016/j.foodchem.2020.128135.

(45) Louppe, G. Understanding Random Forests: From Theory to Practice, 2014.

(46) Breiman, L. Random Forest. In *Machine Learning*; CRC Press: First. | Boca Raton : CRC Press, 2019., 2001; pp 5–32.

(47) Wold, S.; Sjostrom, M. SIMCA: A Method for Analyzing Chemical Data in Terms of Similarity and Analogy. **1977**, 243–282. https://doi.org/10.1021/bk-1977-0052.ch012.

(48) Wold, S.; Sjöström, M.; Eriksson, L. PLS-Regression: A Basic Tool of Chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58* (2), 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1.

(49) Lee, L. C.; Liong, C. Y.; Jemain, A. A. Partial Least Squares-Discriminant Analysis (PLS-DA) for Classification of High-Dimensional (HD) Data: A Review of Contemporary Practice Strategies and Knowledge Gaps. *Analyst* **2018**, *143* (15), 3526–3539. https://doi.org/10.1039/c8an00599k.

Chapter 4
Published reviews

4.1 Chromatographic fingerprinting by comprehensive two-dimensional chromatography: fundamentals and tools

Federico Stilo[1], Carlo Bicchi[1], Ana M. Jimenez-Carvelo[2], Luis Cuadros-Rodriguez[2], Stephen E. Reichenbach[3,4]* and Chiara Cordero[1]*

[1] Dipartimento di Scienza e Tecnologia del Farmaco, Università degli Studi di Torino, Via Pietro Giuria 9, I-10125 Torino, Italy

[2] Department of Analytical Chemistry, Faculty of Science, University of Granada, Avenida de la Fuente Nueva S/N, E-18071 Granada, Spain

[3] Computer Science and Engineering Department, University of Nebraska – Lincoln 260 Avery Hall, Lincoln, NE 68588-0115, USA

[4] GC Image, PO Box 57403, Lincoln, NE 68505-7403, USA

* Corresponding author:

Prof. Dr. Chiara Cordero - Dipartimento di Scienza e Tecnologia del Farmaco, Università di Torino; Via Pietro Giuria 9, I-10125 Torino, Italy – e-mail: chiara.cordero@unito.it; phone: +39 011 6702197

Prof. Stephen E Reichenbach - Computer Science and Engineering Department, University of Nebraska – Lincoln; 260 Avery Hall, Lincoln, NE 68588-0115, USA - email: reich@unl.edu; phone +1 402 4722401

Available online: 28 November 2020

DOI: https://doi.org/10.1016/j.trac.2020.116133

JSS, Volume 134, January 2021, 116133

4.1.1 Abstract

This contribution reviews state-of-the approaches for chromatographic fingerprinting of 2D peak patterns. Concepts of sample's fingerprint and profile, as established in metabolomics, are conceptually translated to comprehensive two-dimensional chromatography (C2DC) separations embracing the principles of biometric fingerprinting.

Approaches founded on this principle - referred to as *chromatographic fingerprinting* - are described and discussed for their information potential and limitations for providing a higher level of information about sample composition. The different type of features (*i.e.*, datapoint, region, peak, and peak-region) are discussed and insights on processing tools and advances in the development of new algorithms are provided. Selected examples cover the most relevant application fields of GC×GC. Challenging scenarios with severe chromatographic misalignment, parallel detection, and translation of methods from thermal to differential-flow modulated GC×GC are also considered for their relevance in specific applications. Machine learning/chemometrics tools are briefly introduced, highlighting their fundamental role in supporting fingerprinting workflows.

Key words

chromatographic fingerprinting; comprehensive two-dimensional gas chromatography; multidimensional analytical platforms; peak features; peak-region features; machine learning; chemometrics; profiling vs. fingerprinting; fingerprinting workflows; GC×GC data processing

4.1.2 Introduction

The terms 'profiling' and 'fingerprinting' have been adopted for metabolomics[1,2] to refer to distinct analytical approaches capable of informing about compositional differences between samples. For profiling, analytical platforms are set to provide detailed information (retention, mass spectrum, detector response, etc.) on qualitative and/or quantitative distributions of samples' components. Profiling can be conducted on a targeted basis,[3] if analytes of interest are defined *a priori* and monitored across all samples. However, if the analytical process is capable of generating individual yet distinctive features for all components, the process can be conceptually extended toward a comprehensive evaluation of all detected constituents and referred to as "untargeted profiling".[4,5] Fingerprinting, as defined by Fiehn,[2] is a high-throughput process capable of unravelling compositional differences between samples, not necessarily achieving accurate quantitative data or compound identifications for all individual constituents. A fingerprint provides a comprehensive set of features ideally corresponding to all chemical constituents and aims to extract the non-evident chemical information included in the whole signal acquired from an analytical instrumental technique. This information mining process is carried out by application of statistical-mathematic tools of chemometric multivariate analysis. Note that chemometrics does not work magic. The information of concern to be mined must be previously embedded in the analytical signal, even if it is hidden to an observer, and the analytical methods to obtain that signal must be specifically designed and optimized keeping this crucial fact in mind.

Fingerprinting methodology can be effectively performed by different approaches:

1. The fingerprint is directly obtained from the sample in its natural state without any pre-treatment except, if applicable, dissolution.
2. The fingerprint is recorded from a particular fraction or family of compounds after a separation or fractionation step. Thus, the fingerprints would be specific of a compound family (*e.g.*, the volatile organic compounds).
3. The fingerprint is obtained after a chemical reaction step (*e.g.*, derivatization), so that there is an alteration of the initial chemical composition of the sample and new compounds are produced (*e.g.*, the fatty acids methyl esters).

A sample's fingerprint can be considered as a totally unspecific signal when the first approach is applied and a partially specific signal when the second and third approaches are employed.

In this sense, signals from spectroscopic techniques fit well with this definition; and nuclear magnetic resonance (NMR), chromatography, mass spectrometry (MS), and Fourier transform infrared spectroscopy (FT-IR) spectra are in fact the most popular fingerprinting methods in metabolomics.[6] Fingerprinting and related concepts have been extended to other fields, *e.g.*: foodomics,[7] sensomics,[8,9] nutrimetabolomics,[10] and petroleomics.[11] With the rapid evolution of analytical techniques, more stable and informative multidimensional platforms now are readily available, offering further possibilities to develop the concept of fingerprinting.

4.1.2.1 Nomenclature for working data in the area of chemical fingerprinting

Regarding analytical signals recorded by each analytical technique, there is a proper nomenclature for the different working data [6], based on the instrumental signals with different measuring setups, namely with different detection systems. In order to fully understand and to extract the relevant information of a signal, it is important to define and clarify the meaning of the terms usually employed during the step of treatment of data: *dimension*, *way*, *order*, *vector*, *matrix*, *cube*,

tensor and *array*. The terms *dimension*, *way* and *order* refer to the type of signal acquired by the analytical instrument. Each analytical signal is described by a main *dimension* or *way* that is related with the signal intensity and one or more complementary *dimensions* or *ways* which characterize the position scores of each intensity value into the signal. The number of complementary dimensions defines the data *order*. A conventional chromatogram (*e.g.*, 1D GC-FID) is an instance of a *two-way* signal (retention times and detector intensities) and constitutes a *first-order* data. Note that in the particular case of signals defined by two chromatographic dimensions with two retention times, the term 2D chromatogram is then applied which in turn is a *three-way* signal. The terms *vector, matrix* and *cube* are usually employed to name a mathematical layout where the working data are arranged once the acquired signal is exported from the instrument. For example, a *vector* denotes a *first-order* data (*two-way* signal), a *matrix* containing a *second-order* data (*three-way* signal), and a *cube* is used for *third-order* data. The term *tensor* is used to name collectively all of these. Finally, the term *array* should refer to a structure consisting of a set of *tensors* including the working data from a group of samples. Every *array* has an additional dimension, *i.e.*, the number of samples, with regard to the dimensionality of each sample data.

Usually, the raw chromatographic signal exported from the instrument consists of several thousand intensity values and could be used as a whole to apply fingerprinting. However, the number of elements may be reduced by applying mathematical methods (*e.g.*, resampling) or scientific-technical operations (*e.g.*, computing peak areas). This strategy is typical of profiling. The reduction of the number of elements may reduce the dimensionality, *e.g.*, obtaining a peak-response vector (*first-order* data) from a 2D chromatogram (*second-order* data), although this is not always applicable. A tutorial on analytical chromatographic fingerprinting is provided by Cuadros *et al.*[6]

4.1.2.2 Conceptual translation to 2D patterns fingerprinting

Most multidimensional analytical (MDA) platforms, provide physico-chemical discrimination of a sample's constituents by chromatographic processes, *e.g.*, gas chromatography (GC) and liquid chromatography (LC), accompanied by spectroscopic processes, *e.g.,* MS, to achieve suitable specificity and selectivity thereby expanding discrimination potentials. When chromatography is conducted by comprehensively coupling two separation dimensions, as in the case of comprehensive two-dimensional chromatography (C2DC), the analytical output requires suitable processing to enable data visualization and interpretation.

In particular, in C2DC (*e.g.,* GC×GC, LC×LC, or SFC×SFC), two columns are serially connected and components eluting from the first-dimension (^{1}D) column are periodically trapped and on-line re-injected into a second-dimension (^{2}D) column. In GC×GC, this operation is governed by a modulator, *e.g.*, a thermal or valve-based focusing interface with a brief modulation time-period (P_M), typically between 0.5 to 8 s. The detector, connected to the end of the ^{2}D column, produces sequential data values that vary as a function of the quality/identity and amount of eluting analytes. An analog-to-digital (A/D) converter collects the signal output at a certain frequency and in a sequential order. Two-dimensional chromatogram visualization therefore is rendered by arranging data values from single modulation period (or cycle) as a column of pixels (picture elements) where each pixel corresponds to a single detector event. This process is known as *rasterization*. Pixel columns are sequenced along the abscissa (X-axis, left-to-right) according to ^{1}D separation time and 2D data is presented in a right-handed Cartesian coordinate system, where the ordinate (Y-axis, bottom-to-top) corresponds to the ^{2}D separation elapsed time.[12]

2D peak patterns generated by C2DC can be treated as sample's unique fingerprint with detected compounds providing *minutiae* features to be used for effective cross-comparative analysis.

The term minutiae derives from fingerprint recognition technology, exploited in forensic applications, where the term corresponds to ridge endings and ridge bifurcations on fingertips. Automatic biometric fingerprint verification systems localize and extract a set of minutiae from inked impressions, or detailed images of human fingertips, for cross-matching with stored templates [13].

By translating the concept of biometric fingerprinting into C2DC, any process that detects, re-aligns, and compares minutiae features extracted from 2D peaks patterns across a series of 2D chromatograms, can be classified as fingerprinting. Moreover, because, at the processing level, the 2D chromatographic fingerprint "*contains unspecific and non-evident information which should be extracted by chemometric tools*",[6] such an approach can be deemed "chromatographic fingerprinting". This is in keeping with established views that chromatographic fingerprints refer "*to the entire chromatogram from a certain test material which is distinctive of its composition*" and that "*chromatograms provide a specific and differentiating tool, as an identity card, which could be used in order to 'identitate' or identify a certain material*".[6] **Figure 4.1.1** illustrates how chromatographic signals can be processed according to fingerprinting or profiling principles to achieve a high level of information. The types of features available will be introduced at **Section 4.1.4**.

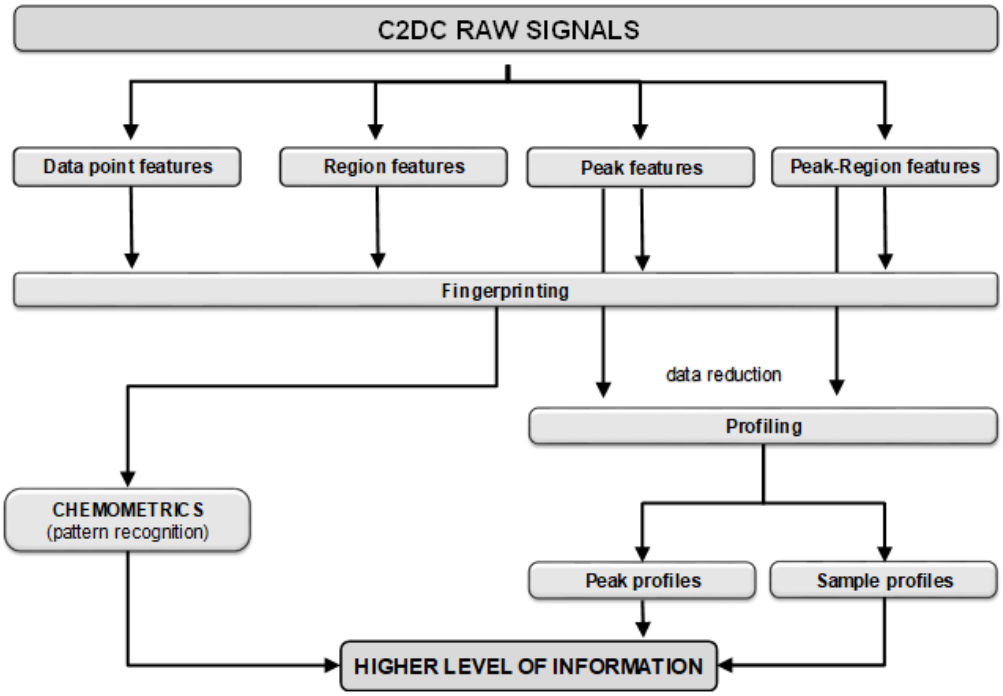

Figure 4.1.1. Types of features for C2DC data processing according to fingerprinting and/or profiling methodologies. Adapted from Ref [6]

In this review, by following this conceptual track, data processing approaches and workflows that comply with the above-mentioned definition are presented, illustrated by selected applications, and critically discussed in view of their capabilities to provide higher levels of information. If 2D chromatographic signals,[6] together with all their metadata, informing about components' identity and physico-chemical characteristics (retention times, detector response, spectral signatures, etc.), are subjected to chromatographic fingerprinting, the overall process achieves a truly comprehensive meaning.

4.1.3 Analytical platforms, dimensions of information available and fingerprinting specificity.

To maximize the information achievable by 2D chromatographic fingerprinting, the analytical platform must be appropriately configured and sample preparation, the zeroth dimension of the system,[14] should be tuned to avoid biases that compromise investigational meanings. Moreover, as stated by Fiehn,[2] to access hidden information in metabolomics, fingerprinting should take into consideration that *the resolution of the analytical devices must be high enough to handle critical information*".

For GC×GC, the separation power, a system characteristic that relates to separation efficiency and resolution, is the product of two separation dimensions and operating in optimized conditions can be very close to its theoretical limit.[15,16] This separation power improvement produces detailed 2D chromatographic fingerprints where co-elution issues and hidden analytes are relatively limited. Based on sample's dimensionality [17], compositional complexity, and components' concentration dynamic range, the choice of the columns' configuration (*i.e.*, columns' dimensions and stationary phases) is a key aspect. In petroleomics, samples' compositional characteristics enable rational ordering of chemical classes and homologous series over the chromatographic space.[18,19]

Figure 4.1.2A shows the image of a diesel fuel analyzed by an apolar × semi-polar column combination producing a high peak capacity.[18] This study was conducted on a ^{1}D column BPX-1 (60 m × 0.25 mm × 0.25 μm) combined to a ^{2}D column BPX-50 (3 m × 0.1 mm × 0.1 μm) and thermal modulation by quad-jet dual-stage modulator (LECO Corp., St. Joseph MI, USA). Chemical classes and groups show relative retentions that follow the discrimination principles of the two dimensions. *n*-Alkanes have the lowest retention along the ^{2}D, while ^{1}D separation is according to carbon number. The π−π interaction, provided by BPX-50 with 50% phenyl substitution, provides separation for unsaturated compounds and, within them, an increasing relative retention based on the presence of multiple aromatic rings. To achieve a high-resolution 2D pattern, a long 2D column was used, with an independent temperature program to avoid pattern distortion (*e.g.*, wrap-around phenomena).

In this example,[18] detection was by time-of-flight MS (TOF MS) operated at nominal mass resolution and 200 Hz acquisition frequency. Multi-channel detection, *e.g.*, by MS, Vacuum Ultraviolet (VUV) analyser, or diode array detection (DAD), provides an additional dimension to the data matrix and superior discrimination to make 2D chromatographic fingerprinting more specific. The specificity relies on the possibility to actively use, for example, MS information for 2D patterns exploration or to open for further data visualization opportunities with scripting. In particular, by Visual Basic Scripting (VBS),[18] on the fragmentation pattern obtained by electron ionization (EI) at 70 eV, rules were derived to enable selective visualization of compounds and compound classes. The application of scripting based on common fragmentations of monoaromatics, enables selective visualization (**Figure 4.1.2B**) of alkylated benzenes from the samples whose complete detectable pattern is shown in **Figure 4.1.2A**. Scripting tools, implemented in most of the commercial software for C2DC, are of help in various application fields, *e.g.*, food[20,21] and environmental applications.[22]

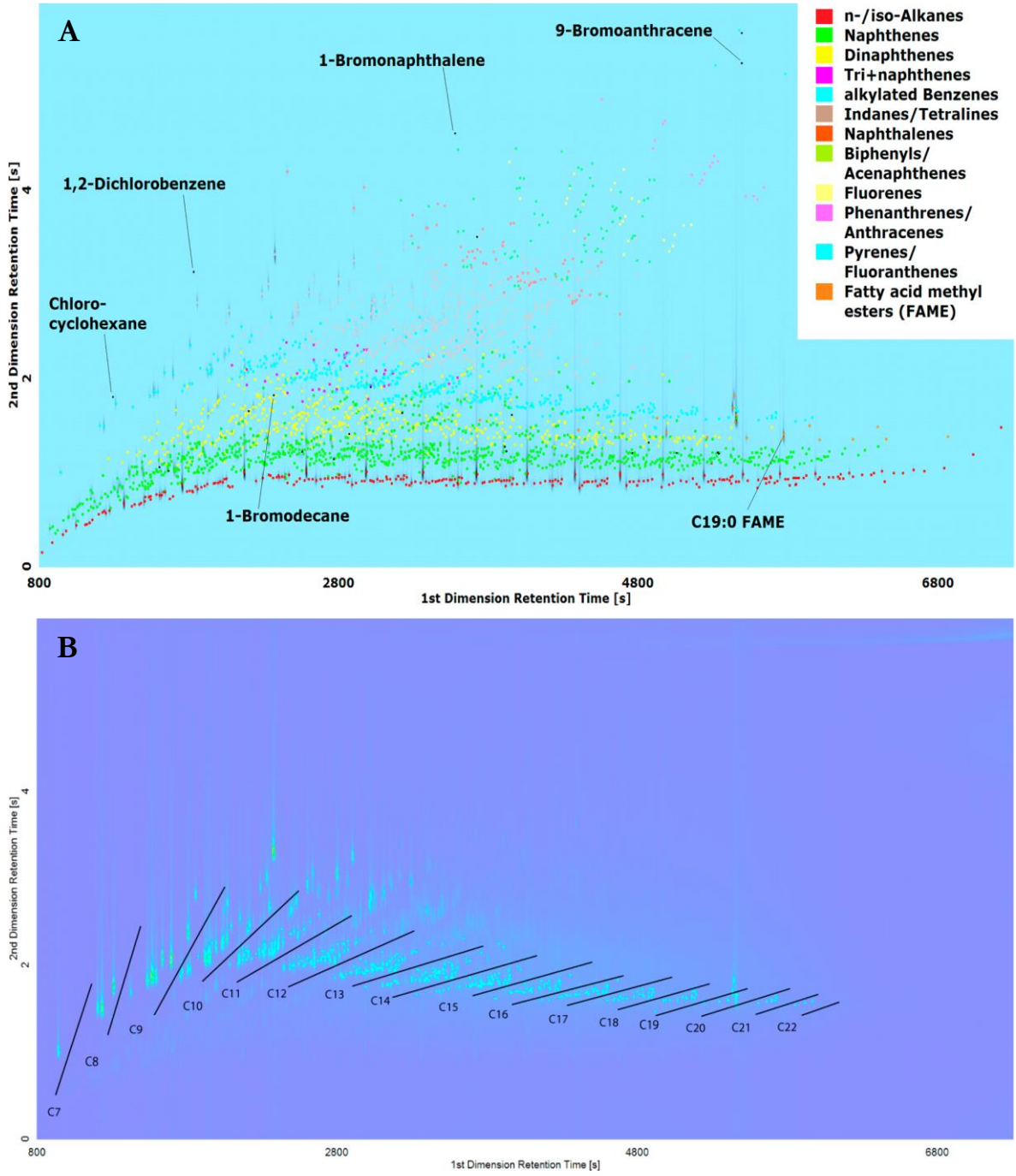

Figure 4.1.2. Complete 2D chromatogram (**4.1.2A**) of a common diesel fuel. Internal standards are named, while colours indicate different chemical classes and groups. In **4.1.2B** the application of the script base on common fragmentation of aromatics, enabling the selective visualization of alkylated benzenes by carbon number. The column set was: 1D column BPX-1 (60 m × 0.25 mm × 0.25 µm) and 2D column BPX-50 (3 m × 0.1 mm × 0.1 µm). Thermal modulation by quad-jet dual-stage modulator. Primary oven program: 60 °C to 220 °C, at 2.0 °C/min; secondary oven program: 60 °C to 140°C at 2.0 °C/min, and 140-310°C at 2.4 °C/min. PM 6s, 0.6 s hot-jet. From Jennerwein et al.[103]

Single-channel detectors, *e.g.*, flame ionization detector (FID), electron capture detector (ECD), sulphur chemiluminescence detector (SCD), ultraviolet (UV) detector, etc., produce single values for each time sample, a first-order data. In this case, fingerprinting specificity is somewhat limited but chromatographic efficacy and separation pattern logic may support particular applications. To

approach a detailed analysis of hydrocarbons in middle distillates type light cycle oil (LCO), Semard *et al.*[23] optimized the column set-up in a differential-flow modulated (FM) platform with FID detection. The application needs suitable resolution and efficiency that provides coherent chemical groups' patterns logic. The authors achieved a global peak capacity ($n_{GC\times GC}$) of 5,148 and filled 56% of the available separation space. The resulting method was suitable for high-throughput screening and qualification of LCO in a cryogen-free GC×GC platform.

Thermal modulation (TM) achieves band-compression in space, which improves method sensitivity by one order of magnitude over conventional 1D GC.[16] This characteristic opens the possibility to acquire data with parallel detectors without compromising method information potential. MS and FID are the most commonly paired detectors; they can be connected with a 2D column post-splitting device[24] or directly to two-parallel second dimension columns [25,26] with some advantages on separation performances. 2D patterns generated by parallel detectors can be explored by various chromatographic fingerprinting approaches (as in **Section 4.1.4**) although their alignment is sometimes challenging due to the different operative pressures of the detectors (vacuum for MS *vs.* ambient pressure for FID). However, the complementary nature of these two detectors offer further possibilities to extend quantitation by response factors (RF)[18,24,27] and predicted RF.[28,29]

Novel MS technologies providing variable energy electron ionization (*viz.*, Tandem Ionization™[30]) by rapidly switching between two pre-selected energies, offer intriguing possibilities to explore chromatographic fingerprints from tandem.[31–35] Details of chromatographic fingerprinting from parallel detector and tandem signals are provided in **Section 4.1.6**.

By configuring the analytical platform, the analyst can optimize and tune chromatographic fingerprint information potential; key aspects are: system geometry, and resulting separation power;[15] system orthogonality;[36–38] and resulting logic of the separation patterns; and detection options, and resulting specificity. What follows the analytical process/data acquisition, *i.e.*, data processing, should extract and treat features in a way to achieve a high level of information: discrimination, classification, prediction, and information about composition.[39]

4.1.4 Data processing principles and tools

Here, our discussion of data processing focuses on feature extraction for pattern recognition (PR), but these data-analysis steps may require preprocessing such as for rasterization, modulation-phase adjustment, baseline correction, retention time alignment, and peak detection. Some recent developments in these areas are discussed here as they relate to feature extraction and analysis, but several reviews discuss methodologies in these areas more comprehensively.[12,40–44]

Reichenbach *et al.*[45] described five types of C2DC features for cross-sample analyses: visual images, datapoints, peaks, regions, and peak-regions. Visual images present chromatograms using various methods such as pseudo-colorization and three-dimensional projections. Datapoint analyses treat each datapoint as a feature. Peak-based approaches attempt to separately integrate multiple datapoints induced by each individual analyte. Regional features aggregate datapoints in separate regions of the two-dimensional chromatographic plane. Peak-region methods attempt to define a region for each individual analyte. **Sections 4.1.4.1-4.1.4.4** discuss the four quantitative approaches to C2DC features (*i.e.*, all except visual images).

After corresponding features are established across chromatograms, PR and machine learning (ML) methods can be used for various cross-sample problems, including classification, chemical

fingerprinting, trend monitoring, unsupervised clustering, and chemical marker discovery. **Section 4.1.6** introduces some of the PR and ML methods that have been used for C2DC analyses.

4.1.4.1 Datapoint features

One of the earliest approach for comparing C2DC chromatograms used the datapoints themselves as separate features: *e.g.*, FID intensities with Fisher's *f* ratio (FR) and principal components analysis (PCA),[46] and, later, MS spectra with PCA for selected channels[47,48] and with pointwise summed *f* ratios across MS spectral channels.[49] In 2015, Pierce *et al.* surveyed pixel-level GC×GC data analyses, including fingerprinting and PR.[50]

Datapoint features are comprehensive of untargeted analytes and provide the highest precision for chromatographic fingerprints, but: (a) there are many duplicative features per analyte (*i.e.*, multiple spectra per analyte), which results in greater computational complexity, and (b) retention time variations confound feature matching between chromatograms. Various methods for aligning chromatograms have been proposed, but none guarantee accuracy to within a datapoint.

4.1.4.2 Region features

Marney *et al.*[51] used small rectangular tiles to generate features, summing multiple datapoints to reduce the number of features and diminish adverse effects of chromatographic misalignment. Results indicated that this approach enhanced PR of true positives and reduced the likelihood of false positives compared to datapoint features. Although this approach requires less computation and is less susceptible to chromatographic misalignment, it is not fully selective, *i.e.*, there may be multiple analytes per tile and analyte peaks may be split across tiles. Still, informative patterns may be recognizable in such tiles if discriminative analytes are not dominated by other analytes in the tiles. Regions features are well-suited for group-type investigations in which there is a region for each chemical group and the total response for analytes eluting within each region is computed and adopted for classification and qualification. This is a popular approach for petroleomics.[11,19]

4.1.4.3 Peak features

Peak features, *i.e.*, summed data for each analyte peak with associated metadata, aim to profile a sample's constituents. Cross-comparative analyses using peak features require matching peak features across multiple chromatograms. For targeted analyses, this approach requires detection of the peak for each target analyte, using its known characteristics, in each chromatogram. Then, the peaks for each target analyte are explicitly matched across chromatograms by the target name. Untargeted cross-comparative analyses with peak features are more difficult because there is no explicit matching of peaks across chromatograms. So, before performing cross-comparisons of untargeted analytes using peak features, the software must match unidentified (or putatively identified) peaks across chromatograms — a process usually referred to as peak matching or peak tracking.

Ideally, untargeted peak matching results in a list of all analytes in each chromatogram, but in practice the results tend to be incomplete and/or error-prone for large sets of chromatograms with many detected peaks. Ambiguities arise from variations in retention time and spectral signatures across chromatograms and among multiple analytes with similar (or co-eluting) retention times and similar spectral signatures, especially if some of those analytes are not detected in some chromatograms. For cross-comparisons of many complex chromatograms, matching invariably

results in unmatched and mismatched peaks. Because of these challenges, the matching process typically yields a list of matched peaks that is not comprehensive of all peaks in all chromatograms and therefore of all analytes. However, many automated methods have been proposed and demonstrated and this approach in LECO's Statistical Compare (SC) software (LECO Corp., St. Joseph MI, USA) is used widely.

Cross-matched peak lists profile each sample with respect to the same analytes. However, with reference to the previous definitions, these profiles provide only partial fingerprints if some peaks are discarded in the matching process. The ultimate goal of research aimed at improving peak matching (some of which is surveyed below) is to produce more accurate and complete profiles and fingerprints.

Our previous literature review of C2DC peak features for comparative analysis [45] described early work from 1999 through early 2011. More recently, other reviews covering methods with comparative peak features include those by Pierce *et al.*,[50] Seeley and Seeley,[52] Chin and Marriott,[53] Tranchida *et al.*,[54] Reaser *et al.*,[55] and Berrier *et al.*[56]

Since 2011, Almstetter *et al.*[57] developed a retention time correction and data alignment tool, Integrative Normalization and Comparative Analysis (INCA), and compared it to SC, noting advantages and disadvantages of each tool. Castillo *et al.*[58] developed the GUINEU software for GC×GC data, including a tool for constructing sequential "paths" containing matching peaks in successive chromatograms.

The research group of Xiang Zhang at the University of Louisville has published several papers on peak matching. Kim *et al.*[59] proposed a similarity measure computed as a mixture of the peak distance and the spectra similarity and implemented peak alignment algorithms in a software package named mSPA. Subsequently, Kim and Zhang[60] published a comparative analysis of MS similarity measures on peak alignment. Kim *et al.*[61] also developed peak alignment algorithms using Smith-Waterman (SW) local alignment[62] and, in experiments with two experimental mixtures each analysed with varying chromatographic conditions, the SW-based method out-performed the DISCO algorithm.[63] Wang *et al.*[64] developed DISCO2 to improve upon the DISCO algorithm. DISCO2 has two stages: full alignment, to find peaks present in all chromatograms, and partial alignment, to align the remaining peaks. Jeong *et al.*[65] developed a model-based peak alignment method using different distance and MS similarity measures. Wei *et al.*[66] developed the Metabolomics Profiling Pipeline (MetPP) that performs peak-list alignment in the two steps, full alignment and partial alignment (as in DISCO2). MetPP results improved on LECO's SC and compared favourably to DISCO and GUINEU.

Deng *et al.*[67] developed four variations of a point matching algorithm but did not compare their performance with other methods. Hoffmann *et al.*[68] developed bidirectional best-hit peak assignment and clique extension for 2D chromatograms (BiPACE 2D) to match peaks across a large number of chromatograms based on comparing peak mass spectra and retention times. Their results indicated that BiPACE 2D outperformed mSPA, SW, and GUINEU for precision, recall, and F1 on data sets acquired under homogeneous conditions but lagged mSPA and GUINEU for data sets acquired with heterogeneous temperature programs. (F1 is the product of precision and recall divided by their average.) Bean *et al.*[69] developed a work-flow to improve SC with filters for peak-picking and alignment. Egert *et al.*[70] developed an alignment strategy called SquareDance, but did not present comparative results with other methods. Barcaru *et al.*[71] developed two variants of an FID peak tracking method that ranks candidates based on Bayesian statistics. Pirok *et al.*[72] developed an algorithm for peak tracking between pairs of LC×LC chromatograms. Titaley *et al.*[73] developed a method to improve the output of SC by post-processing. Li *et al.*[74] developed four

variations of a global peak alignment algorithm based on coherent peak drift and presented results in which CPD achieved better F1 scores than mSPA, SW, and BiPACE 2D.

4.1.4.4 Peak-region features

Peak-region features combine the ideas of regions and peaks by attempting to define one region in the 2D retention time plane for each analyte peak. Peak-regions seek to achieve one-feature-to-one-analyte selectivity, as do peak features, but with implicit matching of the same peak-region across multiple chromatograms, which peak features do not provide. Due to the intrinsic properties of peak-regions, the process of their re-alignment across chromatograms provides comprehensive chromatographic-fingerprinting.

Reichenbach *et al.*[75–77] and Schmarr *et al.*[78,79] described similar peak-region approaches. After pre-processing, the source chromatograms are aligned then combined (*e.g.*, simply by addition or other fusion operations[80]) to form a single composite chromatogram that comprises peaks for all analytes in all samples. Then, a region for each peak detected in the composite chromatogram is recorded in a *feature template*. Finally, each source chromatogram is analyzed by geometrically remapping the feature template back to each source chromatogram.

Ideally, peak-region features are comprehensively selective, accounting for each and every analyte, and feature matching is implicitly performed by mapping each region of the template to each chromatogram. In practice, co-elutions and variable retention times complicate the definition and application of peak-regions, just as they do for peak-based features. For MS data, co-elutions within a peak-region can be addressed with quantitative ions in some cases, but deconvolution (which can be incorporated into peak features) is required for co-eluted analytes with similar spectral signatures. Variations in retention time are addressed by geometric transformations in the retention time plane, as described below. Peak-regions are supported by GC Image GC×GC and LC×LC software (GC Image, LLC, Lincoln NE, USA) in its Smart Templates™, which also support peak features for target compounds and other reliably matched peaks.

Chromatograms must be aligned to construct a composite chromatogram from which the feature template is extracted, and the feature template must be realigned to analyse each chromatogram. (Note that the transform to align a chromatogram for compositing can be inverted to map the composite feature template back to that chromatogram.) The task of chromatographic alignment for peak-region features is related to the problem of peak matching for peak features; in fact, the most common approach to chromatographic alignment is to parameterize a transformation model using a set of matched peaks.

Early work on GC×GC retention time variations[81] found that simple global, affine transformation model can largely account for retention time variability, but subsequent work has explored more powerful methods. Gros *et al* [82] developed a method that uses piecewise linear interpolation/extrapolation for the first dimension and Sibson natural-neighbor interpolation[83] for the second dimension. On three data sets, their method generally outperformed methods proposed by Pierce *et al.*[84] and Zhang *et al.*[85] De Boer and Lankelma[86] developed Curfit2D, an extension of semi-parametric warping[87] to two-dimensions, but did not present comparative results with other methods. Reichenbach *et al.*[88] demonstrated that global, low-degree polynomial transforms could effectively align GC×2GC chromatograms (with dual second columns for FID and MS detectors). Rempe *et al.*[89] presented results for global, low-degree polynomial alignment transformations indicating that performance was highly dependent on the size of the matching-points set and that with enough alignment points they can outperform Gros' method. Zushi *et al.*[90] extended Gros' method to high-resolution MS data. Couprie[91] 2017 developed an alignment method called

BARCHAN that uses a non-rigid transformation in the first dimension and a rigid transformation in the second dimension. For one of the two data sets analyzed, BARCHAN out-performed rigid transforms in both dimensions.

4.1.5 Applications of chromatographic fingerprinting based on various features

4.1.5.1 Chromatographic fingerprinting Visual images and datapoints features

Comparative visualization is a chromatographic fingerprinting approach that enables prompt and intuitive evidence of compositional differences between samples pairs. It could be classified within datapoint features approaches since chromatograms pairs are compared pixel-by-pixel with or without pattern re-alignment or transformation. It has been applied to reveal differences in petrochemical applications,[92–94] food,[78,95–98] body fluids metabolites composition[99–101] and in plant metabolomics.[102,103]

An effective workflow was recently designed to reveal patterns of volatiles related to spoiled hazelnuts showing sensory defects (*mouldy*, *rancid, solvent-like, stale*, and *general unpleasant notes)* while compensating for the compositional variability due to the presence of confounding variables (cultivar, geographical origin, shelf-life and storage) that dominate volatiles composition.[95] The workflow generates composite class-images from samples grouped by sensory qualification of spoilage. Composite chromatograms were obtained by combining 2D chromatograms from samples of the same class into a single composite chromatogram or image. The procedure was made more robust by effective re-alignment of 2D chromatograms through registration peaks reliably matched across chromatogram patterns, and made informative by analytes targeting by untargeted/targeted (*UT*) fingerprinting.[104,105] The combination of two chromatographic fingerprinting approaches in one single workflow (*i.e.*, visual/datapoint features and peak-regions features extended to untargeted and targeted compounds) enabled not only the prompt delineation of fingerprint features correlated to spoilage but also the identification of specific chemical markers to be used in a classification tree with suitable sensitivity and good specificity.

The challenging scenario of multitrophic interactions between plants, insects, and microbes was the object of a study by Pizzolante *et al.*[102] In particular, they investigated the association among *Mentha aquatica*, its specific herbivore *Chrysolina herbacea* (mint bug), and insects' gut microbial community. Comparative visualization was applied to reveal metabolites patterns deriving from oxidative metabolism of gut microbial communities on mint leaf terpenoids. **Figure 4.1.3** shows the comparison between volatiles from *C. herbacea* feces (frass) and those from *M. aquatica* leaf (reference image); **Figure 4.1.3A** relates to female population feeding on *M. aquatica* leaf while **Figure 4.1.3B** is for male population. Colour rendering was by "colorized fuzzy ratio"[106] revealing compositional differences based on normalized responses (over the total image response) and not on absolute detector signal. Red colorization locates features whose normalized response was larger in the reference image (*M. aquatica* leaf), green colour locates features with larger response in female (**Figure 4.1.3A**) or male (**Figure 4.1.3B**) frass. Yellow circles highlight features further investigated to understand biotransformation and/or degradation metabolism while pink circles indicate gender-specific analytes, exclusively detected in male population.

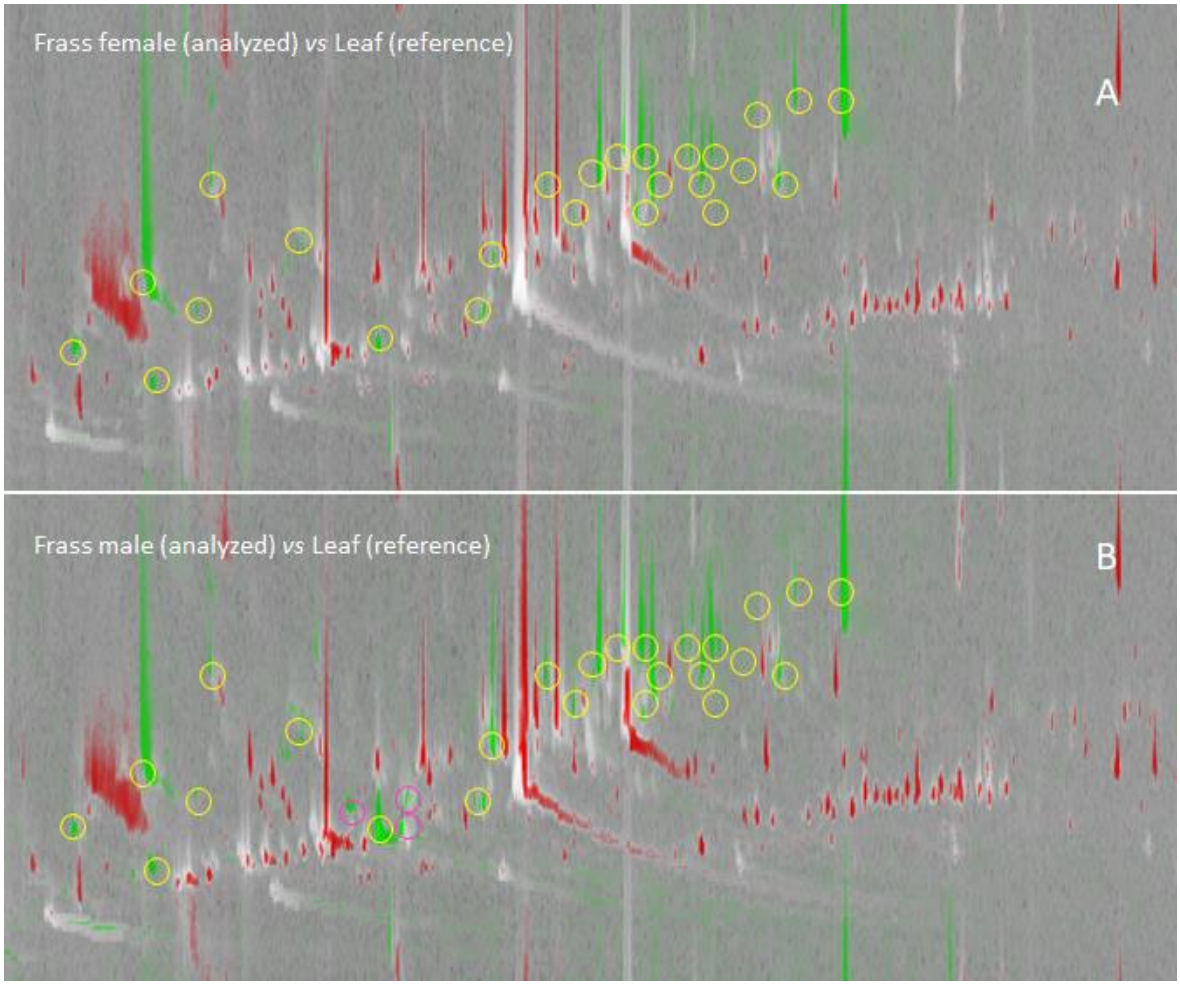

Figure 4.1.3. Comparative visualization between volatiles patterns from C. herbacea feces (analyzed image) and those from M. aquatica leaf (reference image) in a female (**4.1.3A**) or male (**4.1.3B**) population. Colour rendering is by colorized fuzzy ratio (details provided in the text). The analyses were performed by a column set consisting of a ^{1}D SE52 (30 m × 0.25 mm ID, 0.25 μm) and a ^{2}D OV-1701 (0.1 m × 0.1 mm ID, 0.1 μm). Thermal modulation was by loop-type modulator. Modulator capillary: 1 m × 0.10 mm ID of deactivated fused silica capillary. Oven program: 45 °C (1 min) to 260 °C (5 min), rate 2.5 °C/min, P_M 4s, hot-jet 0.25s. Adapted from Pizzolante et al.[93]

Examples provided here suggest how strategic is the combination of different chromatographic fingerprinting strategies, or the combination of fingerprinting with profiling, to access higher level information (**Figure 4.1.1**) encrypted within the chemical code. By visual/datapoint fingerprinting, samples are discriminated based on their 2D signatures'/patterns' quali-quantitative compositions. Behind datapoints, if metadata are tracked, *ex-post* investigation enables identification of potential markers while deepening the knowledge of sample composition and on the phenomenon under study.

4.1.5.2 Chromatographic fingerprinting by regions features

Region features that collect information for multiple analytes belonging to a series of homologues or a chemical group are adopted in the quantitative chromatographic fingerprinting of petrochemical samples[18,92,107–110] and for safety assessments of food suspected of mineral oil hydrocarbon (MOH) contamination.[111]

Sample dimensionality[17] in petrochemical samples enables effective fingerprinting by region features and differential quantification of chemical classes. **Figure 4.1.4** shows GC×GC-(TOF)MS plots of crude oil samples (named B06, B08, B07) analysed by a ^{1}D DB-17 (Agilent Technologies, Wilmington, DE, USA), 50%-phenyl-50%-dimethyl-siloxane (30 m, 0.25 mm i.d., 0.25 μm df) coupled to ^{2}D DB-5 (Agilent Technologies, Wilmington, DE, USA), 5%-phenyl-95%-dimethyl-siloxane (1.2 m, 0.18 mm i.d., 0.18 μm df) and quad-jet thermal modulator (LECO Corp., St. Joseph MI, USA). By this stationary phase combination, saturated *n*-alkanes have longer retention in the ^{2}D while aromatics are less retained. Samples were compared by their differential amounts of group classes (*n*-alkanes, branched alkanes, monocyclic, bicyclic and polycyclic hydrocarbons, alkyl-benzenes, alkyl-naphthalenes, alkyl-phenanthrenes and alkyl-9H-fluorenes) by the semi-quantification by total ion chromatogram (SQTIC) method.[18,112] The class semi-quantification was by summing all peak areas from each class before their normalization to that of the class specific internal standard (*n*-hexadecane-D34 for saturated hydrocarbons up to 20 carbon atoms and pyrene-D10 for aromatic hydrocarbons) at a known concentration. Authors applied the group-type semi-quantification to a selection of Brazilian crude oils (n=12) ranging between 14.8 and 44.5°API. By direct analysis of a single oil drop, and without any fractionation or clean up, effective discrimination was achieved with rational explanation of samples' behaviours.

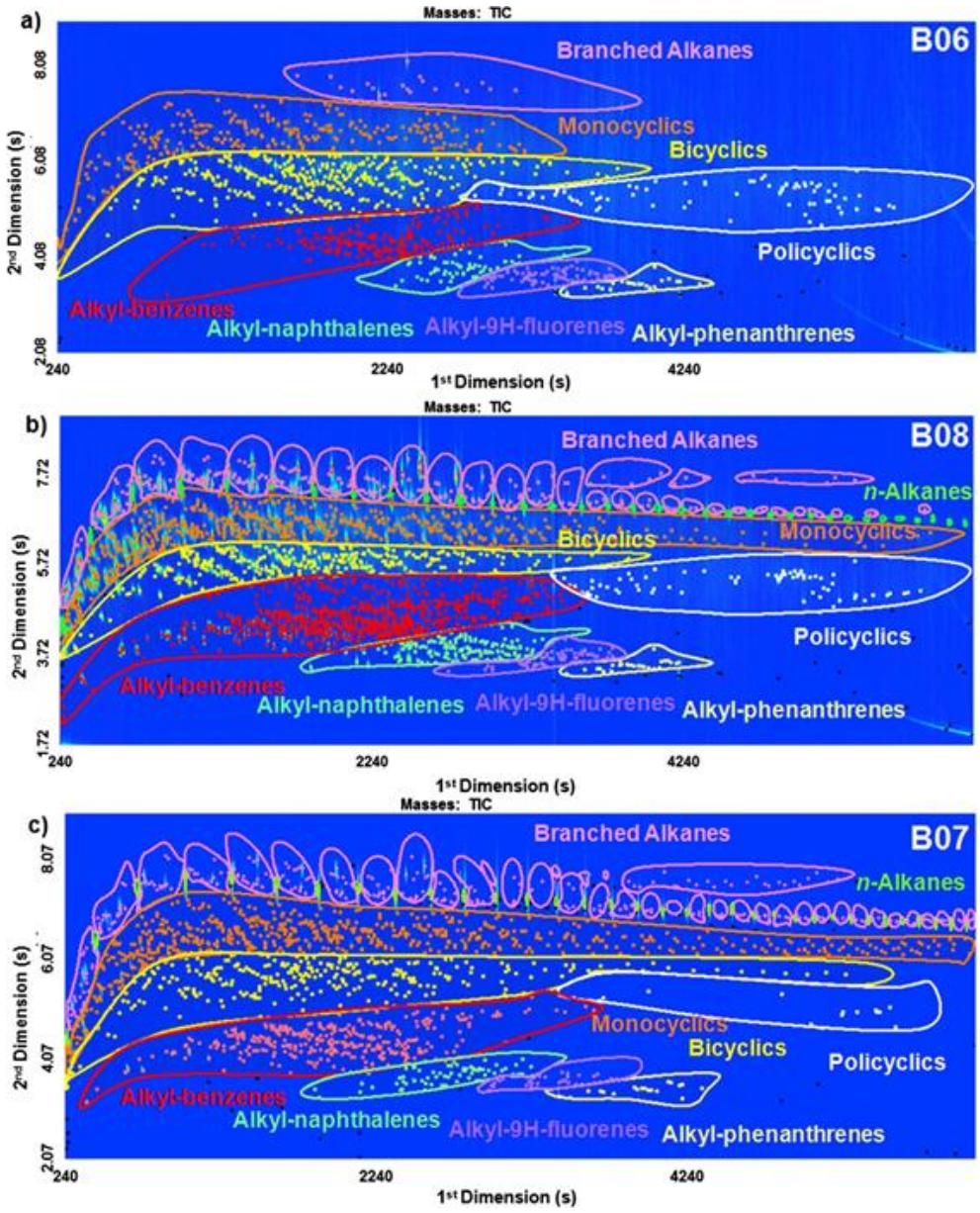

Figure 4.1.4. Contour plot of crude oil samples with different API° values. Group classes, enclosed in region features, are shown with different colours: n-alkanes, branched alkanes, monocyclic, bicyclic and policyclic hydrocarbons, alkyl-benzenes, alkyl-naphthalenes, alkyl-phenanthrenes and alkyl-9H-fluorenes. The column set was: ^{1}D column DB-17 (30 m × 0.25 mm × 0.25 μm) and ^{2}D column DB-5 (1.2 m × 0.17 mm × 0.17 μm). Thermal modulation by quad-jet dual-stage modulator. Oven program: 40 °C (5 min) to 330 °C, at 3.0 °C/min; secondary oven 45 °C (5 min) to 335 °C, at 3.0 °C/min. P_M 9s, 2.25s hot-jet. From Coutinho et al.[99]

MOHs are a complex mixture of isomers including mineral oil saturated hydrocarbons (MOSHs) and mineral oil aromatic hydrocarbons (MOAHs); their occurrence in food might be related to contamination by exogenous sources. The MOSH fraction includes aliphatic hydrocarbons (linear -alkanes, branched – isoalkanes and cyclic compounds – cycloalkanes or naphthenes) with possible substitution. The MOAH fraction includes aromatic derivatives with one or more benzene rings and extensive alkylation. Within this class, known toxic compounds are present, exerting mutagenic activity and tumour promotion. To date, an official confirmatory method aimed at accurately quantifying exogenous MOH in food with the concurrent possibility of establishing the origin of this contamination [24,111,113] is not yet available.

The analytical challenge, faced by established methods for MOSH/MOAH quantification in food and based on liquid chromatography (LC) off-line/on-line coupled to GC-FID,[114] relies on the presence of several interfering compounds not effectively isolated by sample preparation and/or by LC pre-fractionation.[115,116] The unresolved complex mixture (UCM) profile, also called a "hump", produced by the LC-GC approach does not guarantee accurate quantification, and overestimation is an issue. To comply with the need for adding a confirmatory step to the analytical procedure, as required by EU Commission Decision 657/2002,[117] MS is fundamental although not sufficiently selective to isolate unique and distinctive signals from exogenous MOSHs and MOAHs.

In this complex scenario, the potentials of GC×GC were immediately clear; Grob and collaborators proposed a fingerprinting approach based on a medium polar × apolar column combination (OV-17 in the ^{1}D and PS-255-dimethyl polysiloxane in the ^{2}D) capable of separating compound classes with a clear logic based on their physico-chemical properties while enabling a fingerprinting based survey to define the source of contamination.[118] **Figure 4.1.5** shows on the left LC-GC-FID chromatograms of MOSH and MOAH fractions and, on the right, the corresponding GC×GC-MS 2D plots from Asian rice samples suspected for MOH contamination of unknown origin.

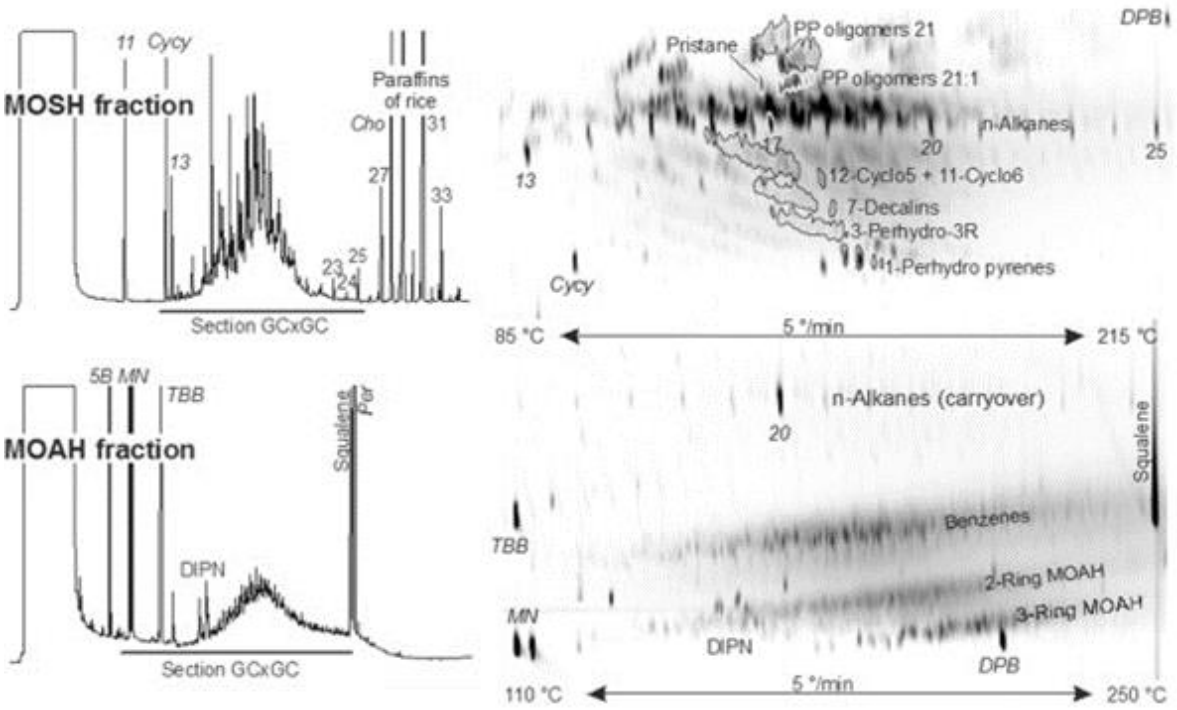

Figure 4.1.5. LC-GC-FID chromatograms (left side) of MOSH and MOAH fractions and the corresponding GC×GC-TOF MS contour plots (right side) from Asian rice samples. The column set was: ^{1}D column OV-17 (15 m × 0.25 mm × 0.15 μm) and ^{2}D column PS-255 (2.5 m × 0.15 mm × 0.055 μm). Thermal modulation by loop-type modulator. Oven program: 70 °C (3 min) to 310 °C, at 5.0 °C/min; P_M 6s. Acronyms are explicated in the text. From Biedermann and Grob.[109]

Samples images in **Figure 4.1.5** show, for the MOSH fraction, the ordered elution pattern of *n*-alkanes ranging from *n*-C13 (internal standard) to *n*-C25 mostly present in uncontaminated rice. Branched paraffines, are ordered along the elution line of *n*-alkanes but with a slightly higher retention in the second dimension. Oligomeric polyolefins (POSH) (C21 oligomers) from polypropylene (PP) packaging have distinctive elution regions. Between the POSH and the *n*-alkanes, monounsaturated PP oligomers were detected. Naphthenes, cyclic hydrocarbons, have a lower retention along the ^{2}D while forming slanted elution bands. Monocyclic C17 naphthenes

(dodecyl cyclopentanes, 12-Cyclo5, and undecyl cyclohexanes, 11-Cyclo6) are highlighted in the pattern.

The MOAH fraction (lower plots) produces on the LC-GC a hump with unresolved hydrocarbons; by structuring analytes elution over the 2D space retention logic enables the separation of diisopropyl naphthalenes (DIPN), selective indicators for recycled paperboard, suggests that MOAHs likely belong to such a contamination source. Moreover, the 2-ring and 3-ring MOAHs show clear elution bands with increasing retention based on the degree of alkylation.

The application of GC×GC-MS to MOSH-MOAH determination brought to truly multidimensional solutions by, for example, combining, on-line the LC pre-fractionation and parallel detection by FID/MS provide room for the development of a methodology capable of qualifying and quantifying MOH contamination with the potential of fingerprinting to accurately identify contamination source(s).[119]

4.1.5.3 Chromatographic fingerprinting by peak features

Chromatographic fingerprinting by peak features is challenging and, as previously discussed, it often results in unmatched and mismatched peaks that impact the comprehensiveness of the process and may hide a sample's traits of relevance. Well established software for untargeted cross-comparative analysis of GC×GC-(TOF)MS data (*e.g.*, LECO SC) might be adopted for chromatographic fingerprinting although careful tuning of processing parameters (S/N detection thresholds, MS matching thresholds, deconvolution function parameters etc.) is mandatory to obtain comprehensive yet distinctive peak lists. Recent review papers cover application fields where this approach was successful,[120–125] although in most cases data reduction with various criteria reduces the profiles' comprehensiveness.

Among recent contributions, untargeted investigations on bacterial volatile organic compounds (VOCs) are worthy of mention[69,121,126–131] for the attempts to capture diagnostic fingerprints from specific bacteria strains. Franchina *et al.*[130] investigated bacterial VOCs signatures after their sampling from *in vitro* cultures. *Staphylococcus aureus*, *Pseudomonas aeruginosa*, and *Escherichia coli* were selected within bacterial species causing infections and sepsis in the respiratory tract. In their study, in view of clinical translation of the methodology, the authors tested dynamic headspace extraction (DHS) followed by thermal desorption with sorbent tubes (TDTs) packed with various materials (Carbopack Y/X/Carboxen1000, Carbopack Y, Tenax TA and Carbopack B/X). Raw 2D chromatograms were processed to extract comprehensive yet consistent untargeted peak features across all samples. In particular, processing parameters were set to discard peaks below a S/N threshold of 100, while re-aligning features based on retention times windows of 6 s and 0.1 s in the ^{1}D and ^{2}D respectively. Inter-chromatogram spectral match threshold, to confirm positive correspondence between features generated by the same component, was set at 70% while a *post hoc* S/N threshold of 50 was defined to recover features lost by the initial peak finding.

The data processing/analysis workflow included a first step of chromatographic fingerprinting followed by data reduction and feature selection (*i.e.*, profiling). Operations were as follows: (a) chromatograms alignment (n=40); (b) artefacts removal; (c) inclusion/exclusion criteria to retain or discard features; (d) response normalization; and (e) discriminatory feature selection by random forest. Of the 317 untargeted peak features, 67 were selected for their discrimination potential and, from these, a few targeted analytes were identified and interrelated to existing knowledge on bacterial VOCs emission.

Focant's group at the University of Liege developed a methodology based on GC×GC-(TOF)MS for the discrimination of cannabis species (*Cannabis indica*, *Cannabis sativa* and hybrids). Cannabis volatiles were extracted by stirbar sorptive extraction (SBSE) and analysed on a system equipped with a differential-flow modulator based on the Bueno and Seeley design[132] and a column set consisting of an apolar ^{1}D Rxi-5MS (5% diphenyl-95% dimethylpolysiloxane phase) (30 m × 0.25 mm i.d. × 0.25 µm df) and a medium polar 2D Rxi-17Sil MS (equivalent to a 50% diphenyl-50% dimethylpolysiloxane phase) (5.0 m × 0.25 mm i.d. × 0.25 µm df (Restek Corporation). **Figure 4.1.6A** shows the 2D plot of a representative sample of cannabis inflorescence volatiles. The fingerprinting by peak features was accompanied with classification regions coherent with the chemical classes characterizing the sample. Untargeted features were annotated with additional information about chemical groups (*i.e.*, monoterpene and sesquiterpene hydrocarbons, sesquiterpene alcohols, fatty acids and cannabinoids) while enabling informative clustering, shown in **Figure 4.1.6B**, and classification of samples based on different chemotypes. Furthermore, the intrinsic profiling power of the technique, enabled the identification of exogenous components (*i.e.*, pesticides and plasticizers) with relevance for product safety.

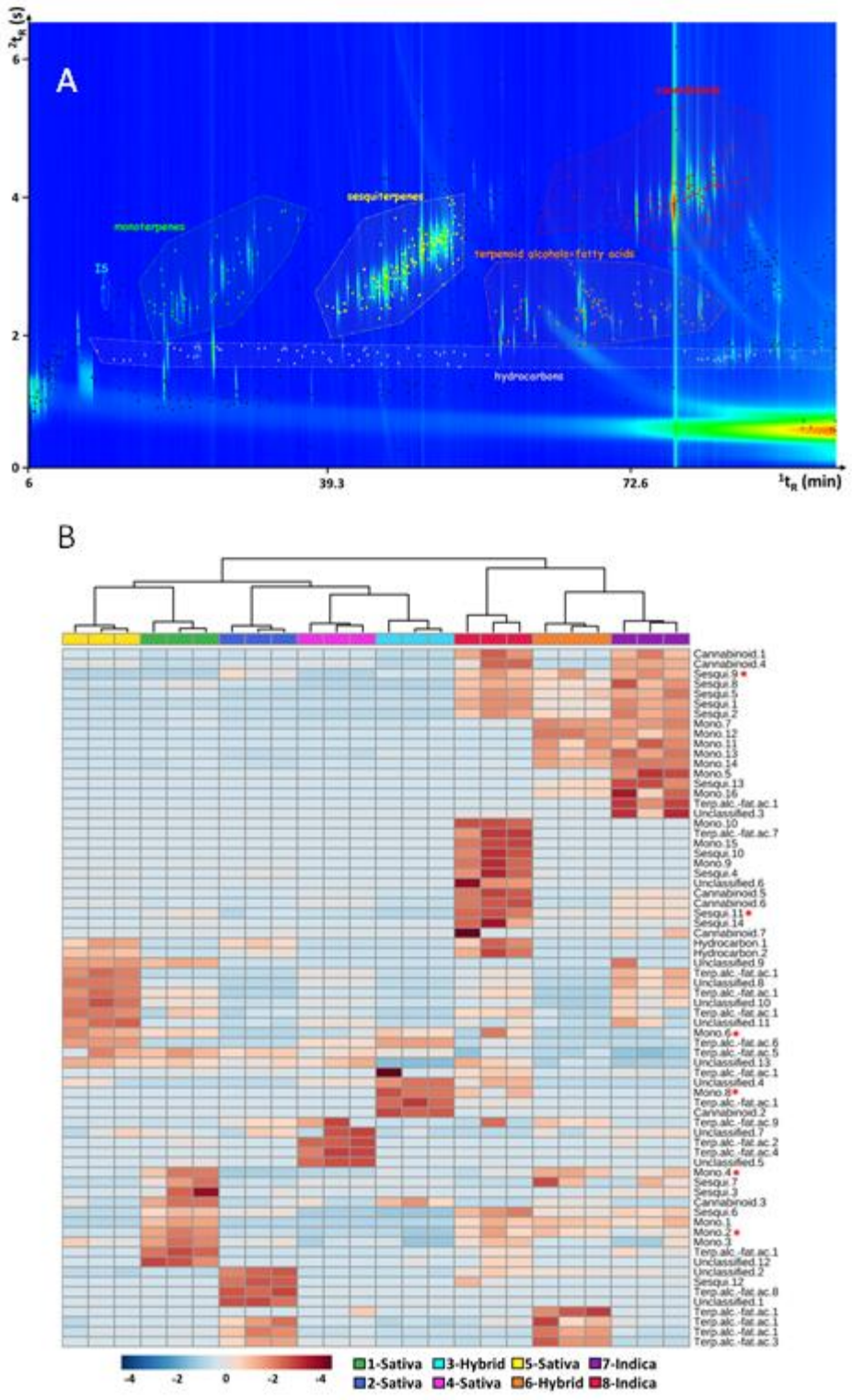

Figure 4.1.6. Contour plot of cannabis volatiles (**4.1.6A**) analyzed after stirbar sorptive extraction with a column set consisting of: ^{1}D column Rxi-5MS (30 m × 0.25 mm × 0.25 µm) and ^{2}D column Rxi-17Sil MS (5.0 m × 0.25 mm × 0.25 µm). Differential-flow modulation based on Bueno and Seeley design [123]. Oven program: 50 °C (5 min) to 330 °C, at 3.0 °C/min; P_M 6.6s. 1D flow 0.7 mL/min; 2D flow 7 mL/min. Heat-map (**4.1.6B**) and hierarchical clustering of 70 peak features with meaningful differences between samples. Features connoted by an asterisk were confirmed by reference standard analysis. Adapted from Franchina et al.[155]

4.1.5.4 Chromatographic fingerprinting by peak-region features

Chromatographic fingerprinting based on peak-region features has been adopted in many fields including metabolomics,[33,77,88,100,133] forensics[125,134,135] bio-oils[136,137] and food applications.[96,98,104,105,138–142]

The reliability of peak-region features fingerprinting was validated against a peak features approach, based on template matching, in a study aimed at delineating distinctive patterns of volatiles in roasted hazelnuts from nine different geographical areas.[96] Data processing was conducted by: (a) locating peak-region features over the GC×GC chromatographic plane and extracting information from analytes distribution; and (b) considering a subset of 2D peaks, together with all available metadata, for features re-alignment. The latter, approaching the chromatographic profiling concept.

The first step was by delineating peak-region features from a cumulative chromatogram formed by summing all of the chromatograms of the set after retention times alignment. Then, peak-regions were detected and collected in a "consensus template", *i.e.,* the collection of minutiae features from the sample set accounting for 411 peak-regions. From the consensus template, peak-region features were copied into individual 2D chromatograms with retention time transformation by geometric scaling and translation parameterized through peak matching. Results, expressed as % of matching between consensus template peak-regions and those delineated for each sample of the set, were interpreted as similarity indicator. *Piemonte* hazelnuts (considered as gold standard for confectionery industry) had the lowest similarity (68%) *vs.* the complete set of minutiae features investigated. The second step, validation, was by template matching of individual 2D peaks across all chromatograms. A consensus template of peak features was built by adding MS similarity constraints to positive matches and by template transformation to compensate for retention time shift. The resulting consensus template had 422 peaks with a subset of 196 reliable peaks that matched in all chromatograms. By this comprehensive peak features template, *Piemonte* hazelnuts confirmed their lowest similarity to the reference fingerprint. Moreover, peak features with the highest response and meaningful variations within the set of samples were putatively identified. With the exception of three features (out of twenty), the two chromatographic fingerprinting approaches gave univocal results. Of the most discriminant variables: acetic acid, 3-methylbutanal, 5-methyl-(*E*)-2-hepten-4-one, and octanal were also key-aroma compounds informing about peculiar sensory characteristics of samples.

A large set of wine samples (n=127; 254 chromatograms), from three main grape producing regions of Brazil, were analysed for their volatiles profiles by *UT* fingerprinting followed by ML. The challenge was to effectively re-align peaks and peak-regions from a large dataset with incidental retention time fluctuations. The *UT* workflow identified 53 reliable peaks, well distributed over the chromatographic space and 793 peak-regions. Response data from peak-regions were explored to answer many different questions about samples characteristics. Grape variety, vintage year, growing region, and winery were explored by various ML methods while accounting for computation time for training, classification, and cross validation. The study demonstrates the importance of a reliable and almost comprehensive mapping of targeted and untargeted features. Whatever the question of interest, if the answer has foundation in the examined fraction and its chemical dimensions, comprehensive chromatographic fingerprinting might help provide consistent answers by achieving higher level information.

4.1.6 Challenging scenarios

Chromatographic fingerprinting faces several challenges when severe misalignment occurs between the chromatograms of a set. As previously discussed, by template matching fingerprinting,

retention times variations can be compensated by applying suitable transformations (see **Section 4.1.4.4**). However, severe misalignment might need analyst supervision in setting critical processing parameters. Stilo *et al.*[140] tackled pattern misalignment and detection inconsistencies, such as those occurring in long-term studies, by examining the effect of processing parameters for *UT* fingerprinting of extra-virgin olive oil volatiles. The authors induced severe distortion by changing chromatographic conditions (carrier gas linear velocity across columns and modulation period) and impacted on detector performances by applying different tuning to the TOF MS.

Signal-to-noise ratio detection threshold, reference spectra choice (2D-peak or highest modulation), and similarity match factor threshold for matching constraints critically impacted the false-negative matches' rate. Moreover, distance thresholds, in the two retention dimensions, and the second-order polynomial transform were key parameters to effectively re-align template data. For targeted analytes, template matching executed by full supervision of the analyst achieved a 97% of accuracy (*i.e.*, 97 % true-positive matches) when replicated samples of the same batch were analysed, while a 92.5% of accuracy was achieved for severely misaligned patterns. By UT fingerprinting, a fully unsupervised procedure, accuracy was of 99.7 % in replicated analyses (same batch) and 97.9 % for misaligned data.

Another context where chromatographic fingerprinting might be challenging is when 2D patterns need re-alignment between platforms operating with different modulator principles. Cordero and co-workers examined the feasibility of translating chromatographic methods from TM GC×GC to differential-flow modulated (FM) GC×GC[143–146] by preserving analytes elution order, relative retention in the two-chromatographic dimensions, and limit of quantification (LOQ). Method translation[147] offers the opportunity of transferring applications from TM to FM platforms with benefits for laboratory operational costs. However, if this task is conducted without considering the chromatographic parameters governing analytes' elution, the resulting 2D patterns may be distorted and difficult to re-align. In practice, method translation attempts to replicate a sample's chromatographic fingerprint without distortions so as to be effectively used for discrimination and qualification.

Figure 4.1.7 shows the GC×GC images of volatiles from a roasted cocoa sample analysed by a reference methodology with TM GC×GC (**Figure 4.1.7A**) combining a ^{1}D PEG SolGelWax™ (30 m, 0.25mm d_c, 0.25 μm d_f) (Trajan Scientific and Medical, Ringwood, Victoria, Australia) to a ^{2}D OV1701 (1.0 m, 0.10 mm d_c, 0.10 μm d_f) (J&W, Wilmington, DE, USA) of by FM GC×GC with translated chromatographic parameters (**Figure 4.1.7B**) on a system equipped with a reverse-inject differential-flow modulator (Agilent Technologies, Wilmington, DE, USA) and a column set consisting of a ^{1}D PEG SolGelWax™ (10 m, 0.10 mm d_c, 0.10 μm d_f) (Trajan Scientific and Medical, Ringwood, Victoria, Australia) and two-parallel ^{2}Ds columns OV1701 (1.3 m, 0.10 mm d_c, 0.10 μm d_f) (J&W, Wilmington, DE, USA). Pink circles indicate targeted peak features whose relative positions were kept coherent between the two analytical platforms. The consistency of the 2D patterns enabled effective transfer of all metadata collected for peak-region features (red areas in **Figure 4.1.7C**) in the original method. Metadata included: target peaks chemical names, reference MS, qualifier and quantifier ions, and MS similarity thresholds. The template of peaks and peak-regions was transformed[88,89] and then matched over the FM GC×GC sample's pattern with minimal supervision. With the translated method, the total analysis time was reduced by a factor of 2, from 60 to 30 minutes of analysis, while fingerprinting information potential was maintained, enabling coherent classification of cocoa samples based on cultivar/origin and processing step.

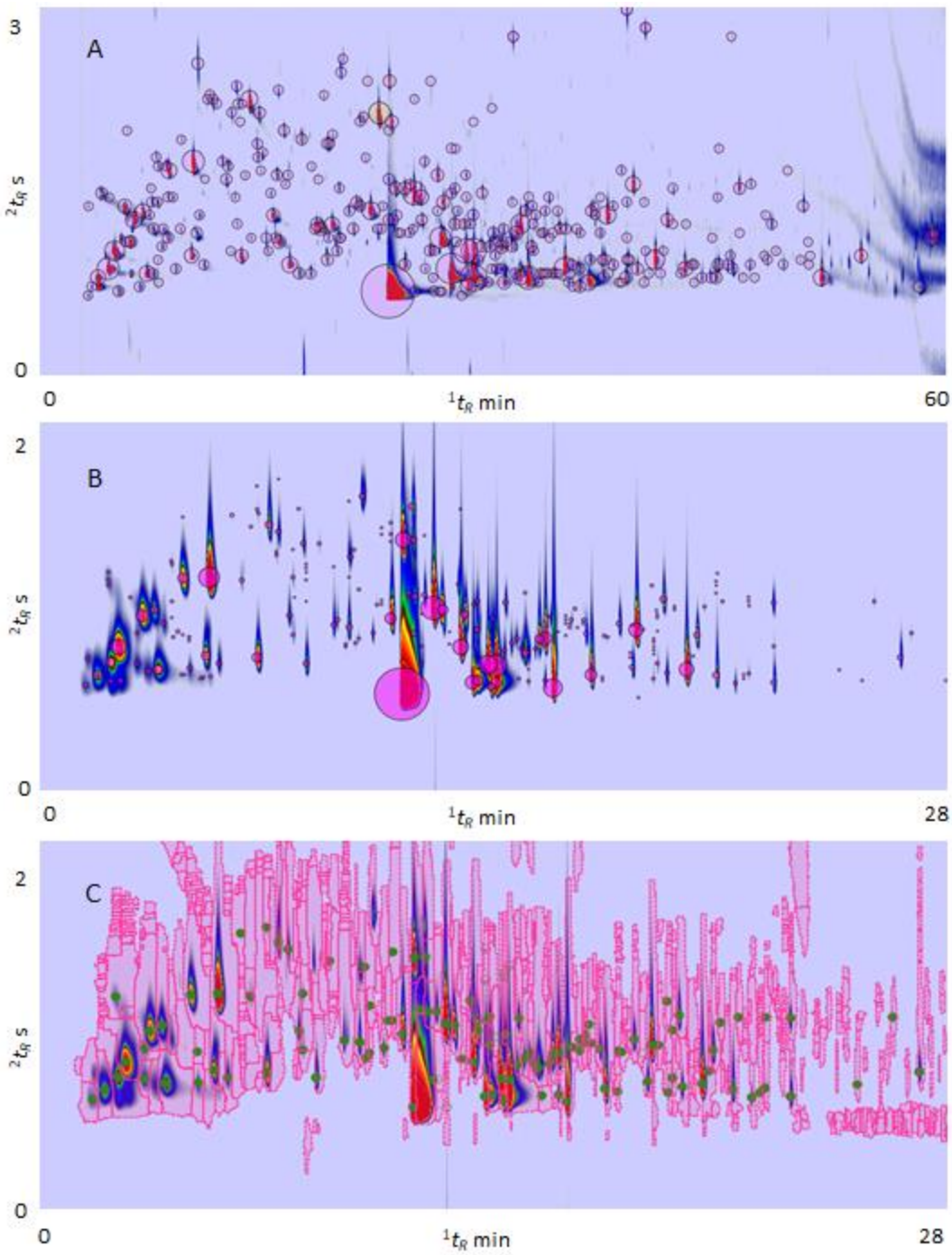

Figure 4.1.7. Contour plots of a roasted cocoa sample analyzed by loop-type TM-GC×GC-TOF MS (**4.1.7A**) and in translated conditions by reverse-inject FM GC×GC-qMS (**4.1.7B**). **Figure 4.1.7C** shows peak-region features template (red areas) after its transform from the reference pattern obtained by TM-GC×GC. In the original method (**4.1.7A**) the column set was by ^{1}D column SolGelWax (30 m × 0.25 mm × 0.25 µm) and ^{2}D column OV-1701 (1 m × 0.1 mm × 0.1 µm). Thermal modulation by loop-type thermal modulator. Oven program: 40 °C (1 min) to 190°C (10 min) at 3.0 °C/min. P_M 3s, 0.25 s hot-jet. In the translated method (**4.1.7B** and **4.1.7C**) the column set was by: ^{1}D column SolGelWax (10 m × 0.1 mm × 0.1 µm) and ^{2}D column OV-1701 (1.3 m × 0.1 mm × 0.1 µm). Modulation by a reverse-inject differential-flow modulator. Oven program: 40 °C (0.43 min) to 190°C at 6.94 °C/min. P_M 2s, 0.11 s pulse time. From Magagna et al.[137]

Parallel detection by MS and FID is commonly adopted in those applications where fingerprinting is followed by, or accompanied by, profiling, identification, and quantitation of marker analytes. FID, in fact, offers the opportunity to accurately quantify analytes based on RFs or predicted RFs[148]. However, the re-alignment between tandem signals from two detectors can be challenging. For MS and FID, the different operative pressures (*i.e.*, vacuum and ambient pressure respectively) generate a carrier gas linear velocity gradient along the ^{2}D column that impacts on ^{2}D retention times (^2t_R).

Parallel detectors pattern re-alignment was the object of contribution by Reichenbach and co-workers[88] who adopted a two-parallel second dimension columns and detection (GC×2GC-MS/FID) to capture complementary fingerprint from urine samples of diabetic (type-2 diabetes) patients from IDES2 cohort study [149]. To achieve a consistent data fusion from the two detectors, chromatographic features (peaks and peak-regions) were aligned by mapping retention times from one detector chromatogram to those of the other detector. **Figure 4.1.8** shows misalignment vectors for a selection of 2D peak pairs (n=156) between FID and MS and resulting re-alignment by applying affine or polynomial transforms. Low-degree polynomial mapping functions out-performed affine transformation (as measured by root-mean-square residuals for matched peaks) and achieved good performance close to a lower-bound benchmark of inherent variability. The transformation grids are illustrated in **Figure 4.1.8** along with the results obtained by their application to pattern re-alignment. The authors examined experimental results and concluded that third-degree transforms out-performed the second-degree polynomials, although second-degree polynomials might be preferred for computational simplicity, number of parameters to be set, and overall robustness.

More recently, the issue of parallel detection re-alignment was tackled by Byrne *et al.*[150] who adopted stencils regions around 2D peaks of interest for effective pattern re-alignment between qMS and FID data from kava (*Piper methysticum*) volatiles fingerprinting. Another application relates to MOSH-MOAH contaminant analysis where regions of the chromatogram where these contaminants elutes need re-alignment between (TOF)MS and FID for accurate group quantitation by RF.[119]

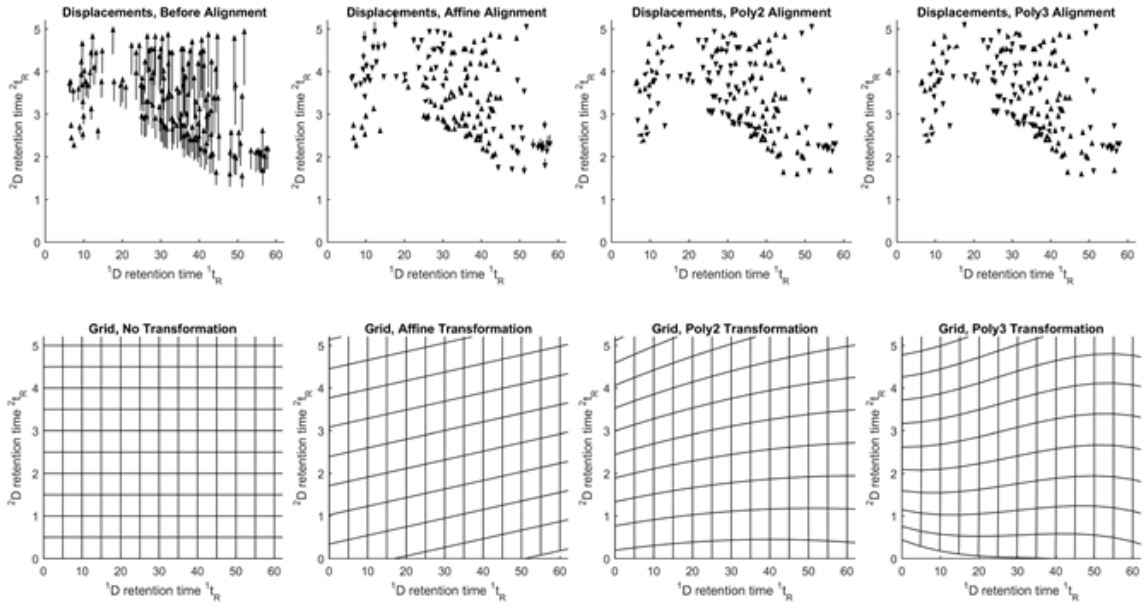

Figure 4.1.8. Misalignment vectors (from FID to MS) for 156 peak pairs in chromatograms from GC × 2GC analysis of a urine sample. Columns from left to right are for the four alignment transformations which grid is illustrated in the second-row images: None (f0), affine (f1), second-degree polynomial (f2), and third-degree polynomial (f3). Adapted from Reichenbach et al.[79]

4.1.7 Machine learning for effective data exploration

Generally, PR with C2DC has proceeded with established methods rather than developing new methods. A fundamental division of PR is between supervised and unsupervised problems. For supervised PR, a training set of feature vectors with class labels (*e.g.*, healthy or unhealthy) are provided; then, the training set is used to develop a method(s) to discern differences between classes. For unsupervised PR, methods must discern both natural groupings/clusters and differences between those clusters.

PR can be conducted with complete fingerprints that provide comprehensive feature sets corresponding to all (*i.e.*, untargeted) chemical constituents or with partial fingerprints (*e.g.*, features extracted for targeted compounds or from non-comprehensive profiles). C2DC typically generates many features, especially for untargeted analyses, so it can be useful to reduce the number of feature variables by either (*a*) projecting feature vectors into a reduced-dimension feature space or (*b*) selecting a subset of the most relevant features. However, it should be noted that any feature reduction and/or selection may decrease the information available for PR (*e.g.*, per Shannon's information theory[151]). Multivariate analysis for which there are many features relative to the number of samples (*i.e.,* hyper-variate) pose the risk of over-fitting models to limited sample data. Cross-validation is a method for evaluating over-fitting by dividing the labelled samples into training and testing sets.

4.1.7.1 Popular unsupervised pattern recognition methods

PCA computes a set of orthonormal basis vectors with maximal energy packing (*i.e.*, the *i*th vector is the best fit of the data while being orthogonal to the first *i*-1 vectors). PCA can reveal natural clusters if those clusters are well separated by the features with greatest variance. PCA also can be used to reduce features by capturing feature correlations. In the C2DC literature, PCA has

been used frequently for unsupervised PR even when class labels are available, but that may be due to familiarity with PCA and lack of familiarity with methods for supervised PR.

Hierarchical cluster analysis (HCA) builds a hierarchy of clusters, with one cluster with all samples at the pinnacle and one cluster for each sample at the base. Two approaches for HCA are *agglomerative* which builds the hierarchy from the bottom up by combining clusters from a lower level to create clusters for the next higher level, and *divisive* which builds the hierarch from the top down by dividing clusters from a higher level to create clusters for the next lower level. The operations (to combine or to divide) attempt to create the densest clusters with the greatest separation between clusters. Various measures for the distance between items and the distance between clusters have been proposed.[152]

4.1.7.2 Popular supervised pattern recognition methods – Qualitative and quantitative approaches

The most commonly used classification methods apply discriminant analysis and voting-based decision rules. Discriminant analysis consists of the establishment of boundaries among the different classes defined by the training set and voting-based decision rules divide the samples into subsets based on the value of certain variables, and this process is repeated on each derived subset of samples. In both cases probabilistic approaches are applies for classification since the belonging of each sample to one class or another is established determining distances or voting. The distances employed are based on Euclidian, Manhattan, or Mahalanobis, among others.

The *F*-test or Fisher ratio (FR) is a popular method for assessing discriminative efficacy with the ratio of between-class variance to within-class variance; a large between-class variance relative to the within-class variance indicates discriminative power. The *F*-test assumes normal distributions and sample independence. Linear discriminant analysis (LDA) determines the linear combination of features (*i.e.*, projection) with the maximal FR, assuming identical class covariances. Subject to its assumptions, LDA is the Bayes-optimal classifier. Quadratic discriminant analysis (QDA) relaxes the constraint that class covariances are equal. Regularized discriminant analysis (RDA) employs two parametric regularization methods to generalize both LDA and QDA.

K nearest neighbors (*k*NN) classifies a sample by a vote of the *k* most similar samples in the training set. The value of the number *k* is a parameter and there are various alternative methods for measuring nearness (*i.e.*, similarity or distance) and for weighted voting. *k*NN classifiers don't posit any distribution or model for the data and so can be especially effective for complex distributions that would be difficult to model.

Decision tree learning builds a tree that organizes a system of rules (or tests) that sequentially drive the method toward its conclusion (*e.g.,* a sample's class). Decision trees are simple to understand and relatable to the features; however, they tend to be not robust, which is a danger for small training sets with many features. In order to cope with such problems, ensemble methods build multiple decision trees with boosting (emphasizing mis-modelled samples) or bagging (resampling training data). Random forests (RF) utilize multiple decision trees to render a decision.

Partial least-squares discriminant analysis (PLS-DA) is a variant for categorical predicted variables. This involves building a PLS regression model to establish class limits and then, performs a discriminant analysis (DA) to achieve classification. Orthogonal projection to latent structures (OPLS) and OPLS-DA are related methods with orthogonal projections.

Support vector machines (SVM) construct a hyperplane in a high-dimensional space for classification, regression, and other PR problems. Various kernel functions can be specified for the

decision function, which provides flexibility. SVM can perform relatively well even for large feature spaces. Deep learning refers to a broad family of methods based on artificial neural networks (ANNs) and is one of the hottest ML topics. However, deep learning is best suited for problems with large data sets, which, unfortunately, is atypical for C2DC, and so has not yet been much used for C2DC analyses.

Quantitative multivariate methods are focused on determining the functional relationships between the analytical signal acquired from a set of samples and a characteristic feature of such samples such as their composition. The advantages compared to using univariate quantitative methods are to perform the quantification in the presence of interfering substances and to quantify materials (not analytes), *e.g.*, the proportion of particular vegetal oil in a blend of vegetal oils or the quantification of mineral fuel in biodiesel.[153] The most widely used is PLS, which consists of building a linear model projecting observable (predictors) and predicted (responses) variables to a new space maximizing covariance.

4.1.7.3 C2DC pattern recognition research

Some previous work from the past decade on pattern analysis with C2DC data is summarized here. In 2010, Humston *et al.*[138] used GC×GC to assess progressive moisture damage to cacoa beans, employing PCA for dimensionality reduction and visualization and a suite of ten regression methods to model progressive damage with 29 analytes selected by FR analysis. Reichenbach *et al.*[76] used *k*-NN, SVM, and PCA with SVM with data from urine samples analysed by LC×LC to classify samples by individual, before/after procedure, and concentration with leave-one-out and replicate *K*-fold cross-validation. In 2012, Caldeira *et al.*[154] used Partial Least Squares-Discriminant Analysis in tandem with Monte Carlo Cross Validation with data from exhaled breath samples analysed by GC×GC-(TOF)MS to assess the predictive power and aid interpretation of recovered compounds possibly related to oxidative stress, inflammation processes, or other cellular processes characteristic of asthma.

In 2015, Steingass *et al.*[155] used GC×GC to assess ripening-dependent changes for pineapple volatiles, employing Analysis of Variance (ANOVA) for selection of 477 features for unsupervised HCA and PCA and for supervised Partial Least Squares Discriminant Analysis (PLS-DA) and Partial Least Squares (PLS) regression. In 2016, Strozier *et al.*[156] used GC×GC to classify chemical agents, employing RF to distinguish different formulations of three types of organophosphate pesticides (OPPs); and Zhu *et al.*[157] used GC×GC to analyse aroma components of teas, employing orthogonal projection to latent structures discriminant analysis (OPLS-DA) and HCA with 478 features to determine compounds with significant across group differences. In 2017, Dubois *et al.*[158] used GC×GC to assess changes in volatile profiles of human blood, employing FR and ANOVA to filter features and PCA and HCA for dimensionality reduction and visualization; and Miyazaki *et al.*[159] used GC×GC to analyse serum metabolites of neonatal calves, employing HCA, PCA, and PLS regression, and correlation coefficients.

In recent years, the Hill Lab (Bean,[131] Rees,[160] Purcaro,[161] Beccaria,[162] *et al.*) has used GC×GC extensively to investigate various issues of respiratory diseases, including fungi, bacteria, viruses, and human breath; employing HCA, PCA for dimensionality reduction, RF, linear SVM, and PLS-DA. In 2018, Reichenbach *et al.*[105] evaluated 22 different ML methods (including decision trees, discriminant analysis, SVM, k-NN, and ensemble methods) for seven different classification problems with GC×GC-MS chromatograms for 127 different Brazilian wines; and Shi *et al.*[163] used GC×GC to analyze 39 dark teas by group, employing PCA for dimensionality reduction and HCA. In 2019, Lukić *et al.*[164] used GC×GC for differentiation of olive oils by variety and geographic origin,

employing ANOVA and FR for feature ranking, PCA for dimensionality reduction, HCA, and forward stepwise linear discriminant analysis (SLDA).

4.1.8 Conclusions

Chromatographic fingerprinting by C2DC is undoubtedly a profitable strategy for cross-comparative analysis of large set of samples with an almost comprehensive coverage of their constituent components. Dedicated data processing on instrumental fingerprint is necessary to extract meaningful high-level information from different types of features, while tackling issues related to retention times misalignment and MS detection inconsistencies.

Multidimensional analytical platforms combining orthogonal information dimensions enable highly accurate and specific chromatographic fingerprinting, giving access to components' profiles (MS spectra) that brings the investigation toward profiling knowledge. By machine learning, fingerprinting allows sample classification, discrimination, and even identification, while the combination with profiling gives access to a higher level of information corroborated by a deep knowledge on sample constitutive elements.

Algorithms and workflows for effective chromatographic fingerprinting are available within most of the commercial software packages dedicated to GC×GC and other C2DC platforms. Besides the clear advantages of automatic procedures, analyst supervision and critical approaches can make the difference and guides to the selection of the most appropriate strategy to achieve investigation goals.

References

(1) Horning, E. C.; Horning, M. G. Human Metabolic Profiles Obtained by Gc and Gc/Ms. *J. Chromatogr. Sci.* **1971**, *9* (3), 129–140. https://doi.org/10.1093/chromsci/9.3.129.

(2) Fiehn, O. Combining Genomics, Metabolome Analysis, and Biochemical Modelling to Understand Metabolic Networks. *Comp. Funct. Genomics* **2001**, *2* (3), 155–168. https://doi.org/10.1002/cfg.82.

(3) Beale, D. J.; Pinu, F. R.; Kouremenos, K. A.; Poojary, M. M.; Narayana, V. K.; Boughton, B. A.; Kanojia, K.; Dayalan, S.; Jones, O. A. H.; Dias, D. A. *Review of Recent Developments in GC–MS Approaches to Metabolomics-Based Research*; Springer US, 2018; Vol. 14. https://doi.org/10.1007/s11306-018-1449-2.

(4) Wolfender, J. L.; Marti, G.; Thomas, A.; Bertrand, S. Current Approaches and Challenges for the Metabolite Profiling of Complex Natural Extracts. *J. Chromatogr. A* **2015**, *1382*, 136–164. https://doi.org/10.1016/j.chroma.2014.10.091.

(5) Alonso, A.; Marsal, S.; Julià, A. Analytical Methods in Untargeted Metabolomics: State of the Art in 2015 . *Frontiers in Bioengineering and Biotechnology* . 2015, p 23.

(6) Cuadros-Rodríguez, L.; Ruiz-Samblás, C.; Valverde-Som, L.; Pérez-Castaño, E.; González-Casado, A. Chromatographic Fingerprinting: An Innovative Approach for Food "identitation" and Food Authentication - A Tutorial. *Anal. Chim. Acta* **2016**, *909*, 9–23. https://doi.org/10.1016/j.aca.2015.12.042.

(7) Cifuentes, A. Foodomics, Foodome and Modern Food Analysis. *TrAC Trends Anal. Chem.* **2017**, *96* (September), 1. https://doi.org/10.1016/j.trac.2017.09.001.

(8) Cordero, C.; Kiefl, J.; Schieberle, P.; Reichenbach, S. E.; Bicchi, C. Comprehensive Two-Dimensional Gas Chromatography and Food Sensory Properties: Potential and Challenges. *Anal. Bioanal. Chem.* **2015**, *407* (1), 169–191. https://doi.org/10.1007/s00216-014-8248-z.

(9) Cordero, C.; Kiefl, J.; Reichenbach, S. E.; Bicchi, C. Characterization of Odorant Patterns by Comprehensive Two-Dimensional Gas Chromatography: A Challenge in Omic Studies. *TrAC Trends Anal. Chem.* **2019**, *113*, 364–378. https://doi.org/10.1016/j.trac.2018.06.005.

(10) Ulaszewska, M. M.; Weinert, C. H.; Trimigno, A.; Portmann, R.; Andres Lacueva, C.; Badertscher, R.; Brennan, L.; Brunius, C.; Bub, A.; Capozzi, F.; Cialiè Rosso, M.; Cordero, C. E.; Daniel, H.; Durand, S.; Egert, B.; Ferrario, P. G.; Feskens, E. J. M.; Franceschi, P.; Garcia-Aloy, M.; Giacomoni, F.; Giesbertz, P.; González-Domínguez, R.; Hanhineva, K.; Hemeryck, L. Y.; Kopka, J.; Kulling, S. E.; Llorach, R.; Manach, C.; Mattivi, F.; Migné, C.; Münger, L. H.; Ott, B.; Picone, G.; Pimentel, G.; Pujos-Guillot, E.; Riccadonna, S.; Rist, M. J.; Rombouts, C.; Rubert, J.; Skurk, T.; Sri Harsha, P. S. C.; Van Meulebroek, L.; Vanhaecke, L.; Vázquez-Fresno, R.; Wishart, D.; Vergères, G. Nutrimetabolomics: An Integrative Action for Metabolomic Analyses in Human Nutritional Studies. *Mol. Nutr. Food Res.* **2019**, *63* (1), 1–38. https://doi.org/10.1002/mnfr.201800384.

(11) Marshall, A. G.; Rodgers, R. P. Petroleomics: The Next Grand Challenge for Chemical Analysis. *Acc. Chem. Res.* **2004**, *37* (1), 53–59. https://doi.org/10.1021/ar020177t.

(12) Maltoni, D.; Cappelli, R.; Meuwly, D. Challenges for Fingerprint Recognition Spoofing, Skin Diseases, and Environmental Effects. *Handb. Biometrics Forensic Sci.* **2017**, No. February 2018, 63–83. https://doi.org/10.1007/978-3-319-50673-9.

(13) Cordero, C.; Schmarr, H.-G.; Reichenbach, S. E.; Bicchi, C. Current Developments in Analyzing Food Volatiles by Multidimensional Gas Chromatographic Techniques. *J. Agric. Food Chem.* **2018**, *66* (10), 2226–2236. https://doi.org/10.1021/acs.jafc.6b04997.

(14) Harynuk, J.; Górecki, T. Comparison of Comprehensive Two-Dimensional Gas Chromatography in Conventional and Stop-Flow Modes. *J. Chromatogr. A* **2006**, *1105* (1-2 SPEC. ISS.), 159–167. https://doi.org/10.1016/j.chroma.2005.09.046.

(15) Klee, M. S.; Cochran, J.; Merrick, M.; Blumberg, L. M. Evaluation of Conditions of Comprehensive Two-Dimensional Gas Chromatography That Yield a near-Theoretical Maximum in Peak Capacity Gain. *J. Chromatogr. A* **2015**, *1383*, 151–159. https://doi.org/10.1016/j.chroma.2015.01.031.

(16) Giddings, J. C. Sample Dimensionality: A Predictor of Order-Disorder in Component Peak Distribution in Multidimensional Separation. *J. Chromatogr. A* **1995**, *703* (1–2), 3–15. https://doi.org/10.1016/0021-9673(95)00249-M.

(17) Jennerwein, M. K.; Eschner, M.; Gröger, T.; Wilharm, T.; Zimmermann, R. Complete Group-Type Quantification of Petroleum Middle Distillates Based on Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry (GCxGC-TOFMS) and Visual Basic Scripting. *Energy and Fuels* **2014**, *28* (9), 5670–5681. https://doi.org/10.1021/ef501247h.

(18) Vendeuvre, C.; Ruiz-Guerrero, R.; Bertoncini, F.; Duval, L.; Thiébaut, D. Comprehensive Two-Dimensional Gas Chromatography for Detailed Characterisation of Petroleum Products. *Oil Gas Sci. Technol.* **2007**, *62* (1), 43–55. https://doi.org/10.2516/ogst:2007004.

(19) Cordero, C.; Cagliero, C.; Liberto, E.; Nicolotti, L.; Rubiolo, P.; Sgorbini, B.; Bicchi, C. High Concentration Capacity Sample Preparation Techniques to Improve the Informative Potential of Two-Dimensional Comprehensive Gas Chromatography-Mass Spectrometry: Application to Sensomics. *J. Chromatogr. A* **2013**, *1318*, 1–11. https://doi.org/10.1016/j.chroma.2013.09.065.

(20) Magagna, F.; Cordero, C.; Cagliero, C.; Liberto, E.; Rubiolo, P.; Sgorbini, B.; Bicchi, C. Black Tea Volatiles Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography – Mass Spectrometry Combined with High Concentration Capacity Sample Preparation Techniques: Toward a Fully Automated Sensomic Assessment. *Food Chem.* **2017**, *225*, 276–287. https://doi.org/10.1016/j.foodchem.2017.01.003.

(21) Van De Weghe, H.; Vanermen, G.; Gemoets, J.; Lookman, R.; Bertels, D. Application of Comprehensive Two-Dimensional Gas Chromatography for the Assessment of Oil Contaminated Soils. *J. Chromatogr. A* **2006**, *1137* (1), 91–100. https://doi.org/10.1016/j.chroma.2006.10.014.

(22) Semard, G.; Gouin, C.; Bourdet, J.; Bord, N.; Livadaris, V. Comparative Study of Differential Flow and Cryogenic Modulators Comprehensive Two-Dimensional Gas Chromatography Systems for the Detailed Analysis of Light Cycle Oil. *J. Chromatogr. A* **2011**, *1218* (21), 3146–3152. https://doi.org/10.1016/j.chroma.2010.08.082.

(23) Zoccali, M.; Barp, L.; Beccaria, M.; Sciarrone, D.; Purcaro, G.; Mondello, L. Improvement of Mineral Oil Saturated and Aromatic Hydrocarbons Determination in Edible Oil by Liquid-Liquid-Gas Chromatography with Dual Detection. *J. Sep. Sci.* **2016**, *39* (3), 623–631. https://doi.org/10.1002/jssc.201501247.

(24) Nicolotti, L.; Cordero, C.; Bressanello, D.; Cagliero, C.; Liberto, E.; Magagna, F.; Rubiolo,

P.; Sgorbini, B.; Bicchi, C. Parallel Dual Secondary Column-Dual Detection: A Further Way of Enhancing the Informative Potential of Two-Dimensional Comprehensive Gas Chromatography. *J. Chromatogr. A* **2014**, *1360*, 264–274. https://doi.org/10.1016/j.chroma.2014.07.081.

(25) Mommers, J.; Ritzen, E.; Dutriez, T.; van der Wal, S. A Procedure for Comprehensive Two-Dimensional Gas Chromatography Retention Time Locked Dual Detection. *J. Chromatogr. A* **2016**, *1461*, 153–160. https://doi.org/10.1016/j.chroma.2016.07.052.

(26) Filippi, J.-J.; Belhassen, E.; Baldovini, N.; Brevard, H.; Meierhenrich, U. J. Qualitative and Quantitative Analysis of Vetiver Essential Oils by Comprehensive Two-Dimensional Gas Chromatography and Comprehensive Two-Dimensional Gas Chromatography/Mass Spectrometry. *J. Chromatogr. A* **2013**, *1288*, 127–148. https://doi.org/10.1016/j.chroma.2013.03.002.

(27) Belhassen, E.; Bressanello, D.; Merle, P.; Raynaud, E.; Bicchi, C.; Chaintreau, A.; Cordero, C. Routine Quantification of 54 Allergens in Fragrances Using Comprehensive Two-Dimensional Gas Chromatography-Quadrupole Mass Spectrometry with Dual Parallel Secondary Columns. Part I: Method Development. *Flavour Fragr. J.* **2018**, *33* (1), 63–74. https://doi.org/10.1002/ffj.3416.

(28) Sgorbini, B.; Cagliero, C.; Boggia, L.; Liberto, E.; Reichenbach, S. E.; Rubiolo, P.; Cordero, C.; Bicchi, C. Parallel Dual Secondary-Column-Dual Detection Comprehensive Two-Dimensional Gas Chromatography: A Flexible and Reliable Analytical Tool for Essential Oils Quantitative Profiling. *Flavour Fragr. J.* **2015**, *30* (5), 366–380. https://doi.org/10.1002/ffj.3255.

(29) Markes International. Select-EV: The next Generation of Ion Source Technology. *Technical Note*. 2016.

(30) Freye, C. E.; Moore, N. R.; Synovec, R. E. Enhancing the Chemical Selectivity in Discovery-Based Analysis with Tandem Ionization Time-of-Flight Mass Spectrometry Detection for Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2018**, *1537*, 99–108. https://doi.org/10.1016/j.chroma.2018.01.008.

(31) Cordero, C.; Guglielmetti, A.; Bicchi, C.; Liberto, E.; Baroux, L.; Merle, P.; Tao, Q.; Reichenbach, S. E. Comprehensive Two-Dimensional Gas Chromatography Coupled with Time of Flight Mass Spectrometry Featuring Tandem Ionization: Challenges and Opportunities for Accurate Fingerprinting Studies. *J. Chromatogr. A* **2019**, *1597*, 132–141. https://doi.org/10.1016/j.chroma.2019.03.025.

(32) Rosso, M. C.; Mazzucotelli, M.; Bicchi, C.; Charron, M.; Manini, F.; Menta, R.; Fontana, M.; Reichenbach, S. E.; Cordero, C. Adding Extra-Dimensions to Hazelnuts Primary Metabolome Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography Combined with Time-of-Flight Mass Spectrometry Featuring Tandem Ionization: Insights on the Aroma Potential. *J. Chromatogr. A* **2020**, *1614*. https://doi.org/10.1016/j.chroma.2019.460739.

(33) Kulsing, C.; Nolvachai, Y.; Marriott, P. J. Concepts, Selectivity Options and Experimental Design Approaches in Multidimensional and Comprehensive Two-Dimensional Gas Chromatography. *TrAC - Trends Anal. Chem.* **2020**, *130*, 115995. https://doi.org/10.1016/j.trac.2020.115995.

(34) Venkatramani, C. J.; Xu, J.; Phillips, J. B. Separation Orthogonality in Temperature-

Programmed Comprehensive Two-Dimensional Gas Chromatography. *Anal. Chem.* **1996**, *68* (9), 1486–1492. https://doi.org/10.1021/ac951048b.

(35) Nolvachai, Y.; Kulsing, C.; Marriott, P. J. Thermally Sensitive Behavior Explanation for Unusual Orthogonality Observed in Comprehensive Two-Dimensional Gas Chromatography Comprising a Single Ionic Liquid Stationary Phase. *Anal. Chem.* **2015**, *87* (1), 538–544. https://doi.org/10.1021/ac5030039.

(36) Cevallos-Cevallos, J. M.; Reyes-De-Corcuera, J. I.; Etxeberria, E.; Danyluk, M. D.; Rodrick, G. E. Metabolomic Analysis in Food Science: A Review. *Trends in Food Science and Technology*. 2009, pp 557–566. https://doi.org/10.1016/j.tifs.2009.07.002.

(37) Reichenbach, S. E. Chapter 4 Data Acquisition, Visualization, and Analysis. *Comprehensive Analytical Chemistry*. 2009, pp 77–106. https://doi.org/10.1016/S0166-526X(09)05504-4.

(38) Matos, J. T. V.; Duarte, R. M. B. O.; Duarte, A. C. Trends in Data Processing of Comprehensive Two-Dimensional Chromatography: State of the Art. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*. 2012, pp 31–45. https://doi.org/10.1016/j.jchromb.2012.06.039.

(39) Pierce, K. M.; Kehimkar, B.; Marney, L. C.; Hoggard, J. C.; Synovec, R. E. Review of Chemometric Analysis Techniques for Comprehensive Two Dimensional Separations Data. *J. Chromatogr. A* **2012**, *1255*, 3–11. https://doi.org/10.1016/j.chroma.2012.05.050.

(40) Hoffmann, N. Computational Methods for High-Throughput Metabolomics, Bielefeld University, 2014.

(41) Bos, T. S.; Knol, W. C.; Molenaar, S. R. A.; Niezen, L. E.; Schoenmakers, P. J.; Somsen, G. W.; Pirok, B. W. J. Recent Applications of Chemometrics in One- and Two-Dimensional Chromatography. *Journal of Separation Science*. 2020, pp 1678–1727. https://doi.org/10.1002/jssc.202000011.

(42) Reaser, B. C.; Watson, Nathanial E. Prebihalo, S. E.; Pinkerton, D. K.; Skogerboe, K. J.; Synovec, R. E. Management and Interpretation of Capillary Chromatographymass Spectrometry Data. In *Hyphenations of Capillary Chromatography with Mass Spectrometry*; Tranchida, P., Mondello, L., Eds.; Elsevier: Amsterdam, 2019; pp 449–480.

(43) Reichenbach, S. E.; Tian, X.; Cordero, C.; Tao, Q. Features for Non-Targeted Cross-Sample Analysis with Comprehensive Two-Dimensional Chromatography. *J. Chromatogr. A* **2012**, *1226*, 140–148. https://doi.org/10.1016/j.chroma.2011.07.046.

(44) Johnson, K. J.; Synovec, R. E. Pattern Recognition of Jet Fuels: Comprehensive GC × GC with ANOVA-Based Feature Selection and Principal Component Analysis. In *Chemometrics and Intelligent Laboratory Systems*; 2002; Vol. 60, pp 225–237. https://doi.org/10.1016/S0169-7439(01)00198-8.

(45) Mohler, R. E.; Dombek, K. M.; Hoggard, J. C.; Young, E. T.; Synovec, R. E. Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Analysis of Metabolites in Fermenting and Respiring Yeast Cells. *Anal. Chem.* **2006**, *78* (8), 2700–2709. https://doi.org/10.1021/ac052106o.

(46) Pierce, K. M.; Hope, J. L.; Hoggard, J. C.; Synovec, R. E. A Principal Component Analysis Based Method to Discover Chemical Differences in Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry (GC × GC-TOFMS) Separations of Metabolites in Plant Samples. *Talanta* **2006**, *70* (4), 797–804.

https://doi.org/10.1016/j.talanta.2006.01.038.

(47) Pierce, K. M.; Hoggard, J. C.; Hope, J. L.; Rainey, P. M.; Hoofnagle, A. N.; Jack, R. M.; Wright, B. W.; Synovec, R. E. Fisher Ratio Method Applied to Third-Order Separation Data to Identify Significant Chemical Components of Metabolite Extracts. *Anal. Chem.* **2006**, *78* (14), 5068–5075. https://doi.org/10.1021/ac0602625.

(48) Pierce, K. M.; Parsons, B. A.; Synovec, R. E. Pixel-Level Data Analysis Methods for Comprehensive Two-Dimensional Chromatography. In *Data Handling in Science and Technology*; Elsevier B.V., 2015; Vol. 29, pp 427–463. https://doi.org/10.1016/B978-0-444-63527-3.00010-2.

(49) Marney, L. C.; Christopher Siegler, W.; Parsons, B. A.; Hoggard, J. C.; Wright, B. W.; Synovec, R. E. Tile-Based Fisher-Ratio Software for Improved Feature Selection Analysis of Comprehensive Two-Dimensional Gas Chromatography-Time-of-Flight Mass Spectrometry Data. *Talanta*. 2013, pp 887–895. https://doi.org/10.1016/j.talanta.2013.06.038.

(50) Kim, S.; Fang, A.; Wang, B.; Jeong, J.; Zhang, X. An Optimal Peak Alignment for Comprehensive Two-Dimensional Gas Chromatography Mass Spectrometry Using Mixture Similarity Measure. *Bioinformatics* **2011**. https://doi.org/10.1093/bioinformatics/btr188.

(51) Kim, S.; Zhang, X. Comparative Analysis of Mass Spectral Similarity Measures on Peak Alignment for Comprehensive Two-Dimensional Gas Chromatography Mass Spectrometry. *Comput. Math. Methods Med.* **2013**. https://doi.org/10.1155/2013/509761.

(52) Kim, S.; Koo, I.; Fang, A.; Zhang, X. Smith-Waterman Peak Alignment for Comprehensive Two-Dimensional Gas Chromatography-Mass Spectrometry. *BMC Bioinformatics* **2011**. https://doi.org/10.1186/1471-2105-12-235.

(53) Smith, T. F.; Waterman, M. S. Identification of Common Molecular Subsequences. *J. Mol. Biol.* **1981**, *147*, 195–197.

(54) Wang, B.; Fang, A.; Heim, J.; Bogdanov, B.; Pugh, S.; Libardoni, M.; Zhang, X. DISCO: Distance and Spectrum Correlation Optimization Alignment for Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry-Based Metabolomics. *Anal. Chem.* **2010**. https://doi.org/10.1021/ac100064b.

(55) Wang, B.; Fang, A.; Shi, X.; Kim, S. H.; Zhang, X. DISCO2: A Comprehensive Peak Alignment Algorithm for Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; 2011. https://doi.org/10.1007/978-3-642-24553-4_64.

(56) Jeong, J.; Shi, X.; Zhang, X.; Kim, S.; Shen, C. Model-Based Peak Alignment of Metabolomic Profiling from Comprehensive Two-Dimensional Gas Chromatography Mass Spectrometry. *BMC Bioinformatics* **2012**. https://doi.org/10.1186/1471-2105-13-27.

(57) Wei, X.; Shi, X.; Koo, I.; Kim, S.; Schmidt, R. H.; Arteel, G. E.; Watson, W. H.; McClain, C.; Zhang, X. MetPP: A Computational Platform for Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry-Based Metabolomics. *Bioinformatics* **2013**. https://doi.org/10.1093/bioinformatics/btt275.

(58) Deng, B.; Kim, S.; Li, H.; Heath, E.; Zhang, X. Global Peak Alignment for Comprehensive Two-Dimensional Gas Chromatography Mass Spectrometry Using Point Matching

Algorithms. In *Journal of Bioinformatics and Computational Biology*; 2016. https://doi.org/10.1142/S0219720016500323.

(59) Hoffmann, N.; Wilhelm, M.; Doebbe, A.; Niehaus, K.; Stoye, J. BiPACE 2D-Graph-Based Multiple Alignment for Comprehensive 2D Gas Chromatography-Mass Spectrometry. *Bioinformatics* **2014**. https://doi.org/10.1093/bioinformatics/btt738.

(60) Bean, H. D.; Hill, J. E.; Dimandja, J. M. D. Improving the Quality of Biomarker Candidates in Untargeted Metabolomics via Peak Table-Based Alignment of Comprehensive Two-Dimensional Gas Chromatography-Mass Spectrometry Data. *J. Chromatogr. A* **2015**, *1394*, 111–117. https://doi.org/10.1016/j.chroma.2015.03.001.

(61) Egert, B.; Weinert, C. H.; Kulling, S. E. A Peaklet-Based Generic Strategy for the Untargeted Analysis of Comprehensive Two-Dimensional Gas Chromatography Mass Spectrometry Data Sets. *J. Chromatogr. A* **2015**, *1405*, 168–177. https://doi.org/10.1016/j.chroma.2015.05.056.

(62) Barcaru, A.; Derks, E.; Vivó-Truyols, G. Bayesian Peak Tracking: A Novel Probabilistic Approach to Match GCxGC Chromatograms. *Anal. Chim. Acta* **2016**. https://doi.org/10.1016/j.aca.2016.09.001.

(63) Pirok, B. W. J.; Molenaar, S. R. A.; Roca, L. S.; Schoenmakers, P. J. Peak-Tracking Algorithm for Use in Automated Interpretive Method-Development Tools in Liquid Chromatography. *Anal. Chem.* **2018**. https://doi.org/10.1021/acs.analchem.8b03929.

(64) Titaley, I. A.; Ogba, O. M.; Chibwe, L.; Hoh, E.; Cheong, P. H.-Y.; Simonich, S. L. M. Automating Data Analysis for Two-Dimensional Gas Chromatography/Time-of-Flight Mass Spectrometry Non-targeted Analysis of Comparative Samples. *J. Chromatogr. A* **2018**, *1541*, 57–62. https://doi.org/10.1016/J.CHROMA.2018.02.016.

(65) Li, Z.; Kim, S.; Zhong, S.; Zhong, Z.; Kato, I.; Zhang, X. Coherent Point Drift Peak Alignment Algorithms Using Distance and Similarity Measures for Two-Dimensional Gas Chromatography Mass Spectrometry Data. *J. Chemom.* **2020**. https://doi.org/10.1002/cem.3236.

(66) Cordero, C.; Liberto, E.; Bicchi, C.; Rubiolo, P.; Reichenbach, S. E.; Tian, X.; Tao, Q. Targeted and Non-Targeted Approaches for Complex Natural Sample Profiling by GC×GC-QMS. *J. Chromatogr. Sci.* **2010**, *48* (4), 251–261. https://doi.org/10.1093/chromsci/48.4.251.

(67) Reichenbach, S. E.; Tian, X.; Tao, Q.; Stoll, D. R.; Carr, P. W. Comprehensive Feature Analysis for Sample Classification with Comprehensive Two-Dimensional LC. *J. Sep. Sci.* **2010**, *33* (10), 1365–1374. https://doi.org/10.1002/jssc.200900859.

(68) Reichenbach, S. E.; Tian, X.; Tao, Q.; Ledford, E. B.; Wu, Z.; Fiehn, O. Informatics for Cross-Sample Analysis with Comprehensive Two-Dimensional Gas Chromatography and High-Resolution Mass Spectrometry (GCxGC-HRMS). *Talanta* **2011**, *83* (4), 1279–1288. https://doi.org/10.1016/j.talanta.2010.09.057.

(69) Schmarr, H. G.; Bernhardt, J.; Fischer, U.; Stephan, A.; Müller, P.; Durner, D. Two-Dimensional Gas Chromatographic Profiling as a Tool for a Rapid Screening of the Changes in Volatile Composition Occurring Due to Microoxygenation of Red Wines. *Anal. Chim. Acta* **2010**. https://doi.org/10.1016/j.aca.2010.05.002.

(70) Schmarr, H.-G.; Bernhardt, J. Profiling Analysis of Volatile Compounds from Fruits Using

Comprehensive Two-Dimensional Gas Chromatography and Image Processing Techniques. *J. Chromatogr. A* **2010**, *1217* (4), 565–574. https://doi.org/10.1016/j.chroma.2009.11.063.

(71) Luhn, S.; Berth, M.; Hecker, M.; Bernhard, J. Using Standard Positions and Image Fusion to Create Proteome Maps from Collections of Two-Dimensional Gel Electrophoresis Images. In *Proteomics*; 2003. https://doi.org/10.1002/pmic.200300433.

(72) Ni, M.; Reichenbach, S. E.; Visvanathan, A.; TerMaat, J.; Ledford, E. B.; Ledford Jr., E. B. Peak Pattern Variations Related to Comprehensive Two-Dimensional Gas Chromatography Acquisition. In *Journal of Chromatography A*; 2005; Vol. 1086, pp 165–170. https://doi.org/10.1016/j.chroma.2005.06.033.

(73) Gros, J.; Nabi, D.; Dimitriou-Christidis, P.; Rutler, R.; Arey, J. S. Robust Algorithm for Aligning Two-Dimensional Chromatograms. *Anal. Chem.* **2012**. https://doi.org/10.1021/ac301367s.

(74) Sibson, R. A Vector Identity for the Dirichlet Tessellation. *Math. Proc. Cambridge Philos. Soc.* **1980**. https://doi.org/10.1017/S0305004100056589.

(75) Pierce, K. M.; Hope, J. L.; Johnson, K. J.; Wright, B. W.; Synovec, R. E. Classification of Gasoline Data Obtained by Gas Chromatography Using a Piecewise Alignment Algorithm Combined with Feature Selection and Principal Component Analysis. *J. Chromatogr. A* **2005**. https://doi.org/10.1016/j.chroma.2005.04.078.

(76) Zhang, D.; Huang, X.; Regnier, F. E.; Zhang, M. Two-Dimensional Correlation Optimized Warping Algorithm for Aligning GCxGC-MS Data. *Anal. Chem.* **2008**. https://doi.org/10.1021/ac7024317.

(77) De Boer, W. P. H.; Lankelma, J. Two-Dimensional Semi-Parametric Alignment of Chromatograms. *J. Chromatogr. A* **2014**. https://doi.org/10.1016/j.chroma.2014.04.034.

(78) van Nederkassel, A. M.; Daszykowski, M.; Eilers, P. H. C.; Heyden, Y. Vander. A Comparison of Three Algorithms for Chromatograms Alignment. *J. Chromatogr. A* **2006**. https://doi.org/10.1016/j.chroma.2006.03.114.

(79) Reichenbach, S. E.; Rempe, D. W.; Tao, Q.; Bressanello, D.; Liberto, E.; Bicchi, C.; Balducci, S.; Cordero, C. Alignment for Comprehensive Two-Dimensional Gas Chromatography with Dual Secondary Columns and Detectors. *Anal. Chem.* **2015**, *87* (19), 10056–10063. https://doi.org/10.1021/acs.analchem.5b02718.

(80) Rempe, D. W.; Reichenbach, S. E.; Tao, Q.; Cordero, C.; Rathbun, W. E.; Zini, C. A. Effectiveness of Global, Low-Degree Polynomial Transformations for GCxGC Data Alignment. *Anal. Chem.* **2016**, *88* (20), 10028–10035. https://doi.org/10.1021/acs.analchem.6b02254.

(81) Zushi, Y.; Gros, J.; Tao, Q.; Reichenbach, S. E.; Hashimoto, S.; Arey, J. S. Pixel-by-Pixel Correction of Retention Time Shifts in Chromatograms from Comprehensive Two-Dimensional Gas Chromatography Coupled to High Resolution Time-of-Flight Mass Spectrometry. *J. Chromatogr. A* **2017**, *1508*. https://doi.org/10.1016/j.chroma.2017.05.065.

(82) Couprie, C.; Duval, L.; Moreaud, M.; Hénon, S.; Tebib, M.; Souchon, V. BARCHAN: Blob Alignment for Robust CHromatographic ANalysis. *J. Chromatogr. A* **2017**, *1484*, 65–72. https://doi.org/10.1016/j.chroma.2017.01.003.

(83) Nizio, K. D.; McGinitie, T. M.; Harynuk, J. J. Comprehensive Multidimensional Separations for the Analysis of Petroleum. *J. Chromatogr. A* **2012**, *1255*, 12–23.

https://doi.org/10.1016/j.chroma.2012.01.078.

(84) Ghasemi Damavandi, H.; Sen Gupta, A.; Nelson, R. K.; Reddy, C. M. Interpreting Comprehensive Two-Dimensional Gas Chromatography Using Peak Topography Maps with Application to Petroleum Forensics. *Chem. Cent. J.* **2016**, *10* (1), 1–14. https://doi.org/10.1186/s13065-016-0211-y.

(85) Hilaire, F.; Basset, E.; Bayard, R.; Gallardo, M.; Thiebaut, D.; Vial, J. Comprehensive Two-Dimensional Gas Chromatography for Biogas and Biomethane Analysis. *J. Chromatogr. A* **2017**, *1524*, 222–232. https://doi.org/10.1016/j.chroma.2017.09.071.

(86) Cordero, C.; Liberto, E.; Bicchi, C.; Rubiolo, P.; Schieberle, P.; Reichenbach, S. E.; Tao, Q. Profiling Food Volatiles by Comprehensive Two-Dimensional Ga Schromatography Coupled with Mass Spectrometry: Advanced Fingerprinting Approaches for Comparative Analysis of the Volatile Fraction of Roasted Hazelnuts (Corylus Avellana L.) from Different Ori. *J. Chromatogr. A* **2010**, *1217* (37), 5848–5858. https://doi.org/10.1016/j.chroma.2010.07.006.

(87) Purcaro, G.; Cordero, C.; Liberto, E.; Bicchi, C.; Conte, L. S. Toward a Definition of Blueprint of Virgin Olive Oil by Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2014**, *1334*, 101–111. https://doi.org/10.1016/j.chroma.2014.01.067.

(88) Stilo, F.; Liberto, E.; Spigolon, N.; Genova, G.; Rosso, G.; Fontana, M.; Reichenbach, S. E.; Bicchi, C.; Cordero, C. An Effective Chromatographic Fingerprinting Workflow Based on Comprehensive Two-Dimensional Gas Chromatography - Mass Spectrometry to Establish Volatiles Patterns Discriminative of Spoiled Hazelnuts (Corylus Avellana L.). *Food Chem.* **2020**, *340* (September 2020), 128135. https://doi.org/10.1016/j.foodchem.2020.128135.

(89) Magagna, F.; Guglielmetti, A.; Liberto, E.; Reichenbach, S. E.; Allegrucci, E.; Gobino, G.; Bicchi, C.; Cordero, C. Comprehensive Chemical Fingerprinting of High-Quality Cocoa at Early Stages of Processing: Effectiveness of Combined Untargeted and Targeted Approaches for Classification and Discrimination. *J. Agric. Food Chem.* **2017**, *65* (30), 6329–6341. https://doi.org/10.1021/acs.jafc.7b02167.

(90) Bressanello, D.; Liberto, E.; Collino, M.; Reichenbach, S. E.; Benetti, E.; Chiazza, F.; Bicchi, C.; Cordero, C. Urinary Metabolic Fingerprinting of Mice with Diet-Induced Metabolic Derangements by Parallel Dual Secondary Column-Dual Detection Two-Dimensional Comprehensive Gas Chromatography. *J. Chromatogr. A* **2014**, *1361*, 265–276. https://doi.org/10.1016/j.chroma.2014.08.015.

(91) Bressanello, D.; Liberto, E.; Collino, M.; Chiazza, F.; Mastrocola, R.; Reichenbach, S. E.; Bicchi, C.; Cordero, C. Combined Untargeted and Targeted Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography: Revealing Fructose-Induced Changes in Mice Urinary Metabolic Signatures. *Anal. Bioanal. Chem.* **2018**, *410* (11), 2723–2737. https://doi.org/10.1007/s00216-018-0950-9.

(92) Collotta, D.; Cordero, C.; Gerlach, J. Q.; Liberto, E.; Chiazza, F.; Cialiè Rosso, M.; Mele, C.; Mai, S.; Reichenbach, S. E.; Tao, Q.; Le Berre, M.; Mastrocola, R.; Aragno, M.; Joshi, L.; Bicchi, C.; Marzullo, P.; Aimaretti, G.; Collino, M. Pilot Study on Comparative Profiling of Biofluids (Plasma, Urine and Saliva) from Metabolically Healthy and Metabolically Unhealthy Obese Subject. In *Proceedings of the 39° Congresso Nazionale della Società Italiana di Farmacologia Firenze, 20-23 Novembre, 2019*; 2019.

(93) Pizzolante, G.; Cordero, C.; Tredici, S. M.; Vergara, D.; Pontieri, P.; Del Giudice, L.;

Capuzzo, A.; Rubiolo, P.; Kanchiswamy, C. N.; Zebelo, S. A.; Bicchi, C.; Maffei, M. E.; Alifano, P. Cultivable Gut Bacteria Provide a Pathway for Adaptation of Chrysolina Herbacea to Mentha Aquatica Volatiles. *BMC Plant Biol.* **2017**, *17* (1), 1–20. https://doi.org/10.1186/s12870-017-0986-6.

(94) Cordero, C.; Zebelo, S. A.; Gnavi, G.; Griglione, A.; Bicchi, C.; Maffei, M. E.; Rubiolo, P. HS-SPME-GC×GC-QMS Volatile Metabolite Profiling of Chrysolina Herbacea Frass and Mentha Spp. Leaves. *Anal. Bioanal. Chem.* **2012**, *402* (5), 1941–1952. https://doi.org/10.1007/s00216-011-5600-4.

(95) Magagna, F.; Valverde-Som, L.; Ruíz-Samblás, C.; Cuadros-Rodríguez, L.; Reichenbach, S. E.; Bicchi, C.; Cordero, C. Combined Untargeted and Targeted Fingerprinting with Comprehensive Two-Dimensional Chromatography for Volatiles and Ripening Indicators in Olive Oil. *Anal. Chim. Acta* **2016**, *936*, 245–258. https://doi.org/10.1016/j.aca.2016.07.005.

(96) Reichenbach, S. E.; Zini, C. A.; Nicolli, K. P.; Welke, J. E.; Cordero, C.; Tao, Q. Benchmarking Machine Learning Methods for Comprehensive Chemical Fingerprinting and Pattern Recognition. *J. Chromatogr. A* **2019**, *1595*, 158–167. https://doi.org/10.1016/j.chroma.2019.02.027.

(97) Striebich, R. C.; Shafer, L. M.; West, Z. J.; Adams, R. K.; Zabarnick, S. Hydrocarbon Group-Type Analysis of Current and Future Aviation Fuels: Comparing ASTM D2425 to GCXGC. In *12th International Conference on Stability, Handling and Use of Liquid Fuels 2011*; University of Dayton Research Institute (UDRI), 300 College Park, Dayton, OH 45469-0043, United States, 2011; Vol. 1, pp 835–859.

(98) Dutriez, T.; Thiébaut, D.; Courtiade, M.; Dulot, H.; Bertoncini, F.; Hennion, M. C. Application to SFC-GC × GC to Heavy Petroleum Fractions Analysis. *Fuel* **2013**, *104*, 583–592. https://doi.org/10.1016/j.fuel.2012.04.048.

(99) Coutinho, D. M.; França, D.; Vanini, G.; Mendes, L. A. N.; Gomes, A. O.; Pereira, V. B.; Ávila, B. M. F.; Azevedo, D. A. Rapid Hydrocarbon Group-Type Semi-Quantification in Crude Oils by Comprehensive Two-Dimensional Gas Chromatography. *Fuel* **2018**, *220* (February), 379–388. https://doi.org/10.1016/j.fuel.2018.02.009.

(100) Ávila, B. M. F.; Pereira, R.; Gomes, A. O.; Azevedo, D. A. Chemical Characterization of Aromatic Compounds in Extra Heavy Gas Oil by Comprehensive Two-Dimensional Gas Chromatography Coupled to Time-of-Flight Mass Spectrometry. *J. Chromatogr. A* **2011**, *1218* (21), 3208–3216. https://doi.org/10.1016/j.chroma.2010.09.051.

(101) Biedermann, M.; Grob, K. Advantages of Comprehensive Two-Dimensional Gas Chromatography for Comprehensive Analysis of Potential Migrants from Food Contact Materials. *Anal. Chim. Acta* **2019**, *1057*, 11–17. https://doi.org/10.1016/j.aca.2018.10.046.

(102) Weggler, B. A.; Gröger, T.; Zimmermann, R. Advanced Scripting for the Automated Profiling of Two-Dimensional Gas Chromatography-Time-of-Flight Mass Spectrometry Data from Combustion Aerosol. *J. Chromatogr. A* **2014**, *1364*, 241–248. https://doi.org/10.1016/j.chroma.2014.08.091.

(103) Biedermann, M.; Munoz, C.; Grob, K. Epoxidation for the Analysis of the Mineral Oil Aromatic Hydrocarbons in Food. An Update. *J. Chromatogr. A* **2020**, *1624*. https://doi.org/10.1016/j.chroma.2020.461236.

(104) Weber, S.; Schrag, K.; Mildau, G.; Kuballa, T.; Walch, S. G.; Lachenmeier, D. W. Analytical

Methods for the Determination of Mineral Oil Saturated Hydrocarbons (MOSH) and Mineral Oil Aromatic Hydrocarbons (MOAH)—A Short Review. *Anal. Chem. Insights* **2018**, *13*, 117739011877775. https://doi.org/10.1177/1177390118777757.

(105) Biedermann, M.; Grob, K. On-Line Coupled High Performance Liquid Chromatography–Gas Chromatography for the Analysis of Contamination by Mineral Oil. Part 2: Migration from Paperboard into Dry Foods: Interpretation of Chromatograms. *J. Chromatogr. A* **2012**, *1255*, 76–99. https://doi.org/10.1016/j.chroma.2012.05.096.

(106) Biedermann, M.; Munoz, C.; Grob, K. Update of On-Line Coupled Liquid Chromatography – Gas Chromatography for the Analysis of Mineral Oil Hydrocarbons in Foods and Cosmetics. *J. Chromatogr. A* **2017**, *1521*, 140–149. https://doi.org/10.1016/j.chroma.2017.09.028.

(107) Commission, E. Commission Decision 2002/657/EC Implementing Council Directive 96/23/EC Concerning the Performance of Analytical Methods and the Interpretation of Results. *Off. J. Eur. Union* **2002**, *L221* (23 May 1996), 8–36.

(108) Biedermann, M.; Grob, K. Comprehensive Two-Dimensional Gas Chromatography for Characterizing Mineral Oils in Foods and Distinguishing Them from Synthetic Hydrocarbons. *J. Chromatogr. A* **2015**, *1375*, 146–153. https://doi.org/10.1016/j.chroma.2014.11.064.

(109) Pantò, S.; Collard, M.; Purcaro, G. Comprehensive Gas Chromatography Coupled to Simultaneous Dual Detection (TOF-MS/FID) as a Confirmatory Method for MOSH and MOAH Determination in Food. *Curr. Trends Mass Spectrom.* **2020**, No. July 2020, 15–20.

(110) Tranchida, P. Q.; Purcaro, G.; Maimone, M.; Mondello, L. Impact of Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry on Food Analysis. *J. Sep. Sci.* **2016**, *39* (1), 149–161. https://doi.org/10.1002/jssc.201500379.

(111) Higgins Keppler, E. A.; Jenkins, C. L.; Davis, T. J.; Bean, H. D. Advances in the Application of Comprehensive Two-Dimensional Gas Chromatography in Metabolomics. *TrAC - Trends Anal. Chem.* **2018**, *109*, 275–286. https://doi.org/10.1016/j.trac.2018.10.015.

(112) Nolvachai, Y.; Kulsing, C.; Marriott, P. J. Multidimensional Gas Chromatography in Food Analysis. *TrAC - Trends Anal. Chem.* **2017**, *96*, 124–137. https://doi.org/10.1016/j.trac.2017.05.001.

(113) Prebihalo, S. E.; Berrier, K. L.; Freye, C. E.; Bahaghighat, H. D.; Moore, N. R.; Pinkerton, D. K.; Synovec, R. E. Multidimensional Gas Chromatography: Advances in Instrumentation, Chemometrics, and Applications. *Anal. Chem.* **2018**, *90* (1), 505–532. https://doi.org/10.1021/acs.analchem.7b04226.

(114) Tranchida, P. Q.; Franchina, F. A.; Dugo, P.; Mondello, L. Comprehensive Two-Dimensional Gas Chromatography-Mass Spectrometry: Recent Evolution and Current Trends. *Mass Spectrom. Rev.* **2016**, *35* (4), 524–534. https://doi.org/10.1002/mas.21443.

(115) Gruber, B.; Weggler, B. A.; Jaramillo, R.; Murrell, K. A.; Piotrowski, P. K.; Dorman, F. L. Comprehensive Two-Dimensional Gas Chromatography in Forensic Science: A Critical Review of Recent Trends. *TrAC - Trends Anal. Chem.* **2018**, *105*, 292–301. https://doi.org/10.1016/j.trac.2018.05.017.

(116) Beccaria, M.; Bobak, C.; Maitshotlo, B.; Mellors, T. R.; Purcaro, G.; Franchina, F. A.; Rees, C. A.; Nasir, M.; Black, A.; Hill, J. E. Exhaled Human Breath Analysis in Active Pulmonary

Tuberculosis Diagnostics by Comprehensive Gas Chromatography-Mass Spectrometry and Chemometric Techniques. *J. Breath Res.* **2019**, *13* (1). https://doi.org/10.1088/1752-7163/aae80e.

(117) Purcaro, G.; Nasir, M.; Franchina, F. A.; Rees, C. A.; Aliyeva, M.; Daphtary, N.; Wargo, M. J.; Lundblad, L. K. A.; Hill, J. E. Breath Metabolome of Mice Infected with Pseudomonas Aeruginosa. *Metabolomics* **2019**, *15* (1), 1–10. https://doi.org/10.1007/s11306-018-1461-6.

(118) Mellors, T. R.; Rees, C. A.; Franchina, F. A.; Burklund, A.; Patel, C.; Hathaway, L. J.; Hill, J. E. The Volatile Molecular Profiles of Seven Streptococcus Pneumoniae Serotypes. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **2018**, *1096* (February), 208–213. https://doi.org/10.1016/j.jchromb.2018.08.032.

(119) Mellors, T. R.; Nasir, M.; Franchina, F. A.; Smolinska, A.; Blanchet, L.; Flynn, J. L.; Tomko, J.; O'Malley, M.; Scanga, C. A.; Lin, P. L.; Wagner, J.; Hill, J. E. Identification of Mycobacterium Tuberculosis Using Volatile Biomarkers in Culture and Exhaled Breath. *J. Breath Res.* **2019**, *13* (1). https://doi.org/10.1088/1752-7163/aacd18.

(120) Franchina, F. A.; Purcaro, G.; Burklund, A.; Beccaria, M.; Hill, J. E. Evaluation of Different Adsorbent Materials for the Untargeted and Targeted Bacterial VOC Analysis Using GC×GC-MS. *Anal. Chim. Acta* **2019**, *1066*, 146–153. https://doi.org/10.1016/j.aca.2019.03.027.

(121) Bean, H. D.; Rees, C. A.; Hill, J. E. Comparative Analysis of the Volatile Metabolomes of Pseudomonas Aeruginosa Clinical Isolates. *J. Breath Res.* **2016**, *10* (4). https://doi.org/10.1088/1752-7155/10/4/047102.

(122) Bueno, P. A.; Seeley, J. V. Flow-Switching Device for Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2004**, *1027* (1–2), 3–10. https://doi.org/10.1016/j.chroma.2003.10.033.

(123) Di Giovanni, N.; Meuwis, M. A.; Louis, E.; Focant, J. F. Untargeted Serum Metabolic Profiling by Comprehensive Two-Dimensional Gas Chromatography-High-Resolution Time-of-Flight Mass Spectrometry. *J. Proteome Res.* **2020**, *19* (3), 1013–1028. https://doi.org/10.1021/acs.jproteome.9b00535.

(124) Verheggen, F.; Perrault, K. A.; Megido, R. C.; Dubois, L. M.; Francis, F.; Haubruge, E.; Forbes, S. L.; Focant, J.-F.; Stefanuto, P.-H. The Odor of Death: An Overview of Current Knowledge on Characterization and Applications. *Bioscience* **2017**, *67* (7), 600–613. https://doi.org/10.1093/biosci/bix046.

(125) Cnuts, D.; Perrault, K. A.; Stefanuto, P.-H.; Dubois, L. M.; Focant, J.-F.; Rots, V. Fingerprinting Glues Using HS-SPME GC×GC–HRTOFMS: A New Powerful Method Allows Tracking Glues Back in Time. *Archaeometry* **2018**, *60* (6), 1361–1376. https://doi.org/10.1111/arcm.12364.

(126) Reichenbach, S. E.; Tian, X.; Boateng, A. A.; Mullen, C. A.; Cordero, C.; Tao, Q. Reliable Peak Selection for Multisample Analysis with Comprehensive Two-Dimensional Chromatography. *Anal. Chem.* **2013**, *85* (10), 4974–4981. https://doi.org/10.1021/ac303773v.

(127) Mullen, C. A.; Boateng, A. A.; Reichenbach, S. E. Hydrotreating of Fast Pyrolysis Oils from Protein-Rich Pennycress Seed Presscake. *Fuel* **2013**, *111*, 797–804. https://doi.org/10.1016/j.fuel.2013.04.075.

(128) Magagna, F.; Cordero, C.; Cagliero, C.; Liberto, E.; Rubiolo, P.; Sgorbini, B.; Bicchi, C. Black Tea Volatiles Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography – Mass Spectrometry Combined with High Concentration Capacity Sample Preparation Techniques: Toward a Fully Automated Sensomic Assessment. *Food Chem.* **2017**, *225*, 276–287. https://doi.org/10.1016/j.foodchem.2017.01.003.

(129) Humston, E. M.; Knowles, J. D.; McShea, A.; Synovec, R. E. Quantitative Assessment of Moisture Damage for Cacao Bean Quality Using Two-Dimensional Gas Chromatography Combined with Time-of-Flight Mass Spectrometry and Chemometrics. *Journal of Chromatography A*. 2010, pp 1963–1970. https://doi.org/10.1016/j.chroma.2010.01.069.

(130) Humston, E. M.; Zhang, Y.; Brabeck, G. F.; McShea, A.; Synovec, R. E. Development of a GCxGC-TOFMS Method Using SPME to Determine Volatile Compounds in Cacao Beans. *J. Sep. Sci.* **2009**, *32* (13), 2289–2295. https://doi.org/10.1002/jssc.200900143.

(131) Stilo, F.; Liberto, E.; Reichenbach, S. E.; Tao, Q.; Bicchi, C.; Cordero, C. Untargeted and Targeted Fingerprinting of Extra Virgin Olive Oil Volatiles by Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry: Challenges in Long-Term Studies. *J. Agric. Food Chem.* **2019**, *67* (18), 5289–5302. https://doi.org/10.1021/acs.jafc.9b01661.

(132) Morimoto, J.; Rosso, M. C.; Kfoury, N.; Bicchi, C.; Cordero, C.; Robbat, A. Untargeted/Targeted 2D Gas Chromatography/Mass Spectrometry Detection of the Total Volatile Tea Metabolome. *Molecules* **2019**, *24* (20), 1–14. https://doi.org/10.3390/molecules24203757.

(133) Stilo, F.; Tredici, G.; Bicchi, C.; Robbat, A.; Morimoto, J.; Cordero, C. Climate and Processing Effects on Tea (Camellia Sinensis L. Kuntze) Metabolome: Accurate Profiling and Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography/Time-of-Flight Mass Spectrometry. *Molecules* **2020**, *25* (10), 2447. https://doi.org/10.3390/molecules25102447.

(134) Cordero, C.; Rubiolo, P.; Cobelli, L.; Stani, G.; Miliazza, A.; Giardina, M.; Firor, R.; Bicchi, C. Potential of the Reversed-Inject Differential Flow Modulator for Comprehensive Two-Dimensional Gas Chromatography in the Quantitative Profiling and Fingerprinting of Essential Oils of Different Complexity. *J. Chromatogr. A* **2015**, *1417*, 79–95. https://doi.org/10.1016/j.chroma.2015.09.027.

(135) Cordero, C.; Rubiolo, P.; Reichenbach, S. E.; Carretta, A.; Cobelli, L.; Giardina, M.; Bicchi, C. Method Translation and Full Metadata Transfer from Thermal to Differential Flow Modulated Comprehensive Two Dimensional Gas Chromatography: Profiling of Suspected Fragrance Allergens. *J. Chromatogr. A* **2017**, *1480*, 70–82. https://doi.org/10.1016/j.chroma.2016.12.011.

(136) Magagna, F.; Liberto, E.; Reichenbach, S. E.; Tao, Q.; Carretta, A.; Cobelli, L.; Giardina, M.; Bicchi, C.; Cordero, C. Advanced Fingerprinting of High-Quality Cocoa: Challenges in Transferring Methods from Thermal to Differential-Flow Modulated Comprehensive Two Dimensional Gas Chromatography. *J. Chromatogr. A* **2018**, *1536*, 122–136. https://doi.org/10.1016/j.chroma.2017.07.014.

(137) Stilo, F.; Gabetti, E.; Bicchi, C.; Carretta, A.; Peroni, D.; Reichenbach, S. E.; Cordero, C.; McCurry, J. A Step Forward in the Equivalence between Thermal and Differential-Flow Modulated Comprehensive Two-Dimensional Gas Chromatography Methods. *J. Chromatogr.*

A **2020**, *1627*, 461396. https://doi.org/10.1016/j.chroma.2020.461396.

(138) Klee, M. S.; Blumberg, L. M. Theoretical and Practical Aspects of Fast Gas Chromatography and Method Translation. *J. Chromatogr. Sci.* **2002**, *40* (5), 234–247. https://doi.org/10.1093/chromsci/40.5.234.

(139) De Saint Laumer, J. Y.; Cicchetti, E.; Merle, P.; Egger, J.; Chaintreau, A. Quantification in Gas Chromatography: Prediction of Flame Ionization Detector Response Factors from Combustion Enthalpies and Molecular Structures. *Anal. Chem.* **2010**, *82* (15), 6457–6462. https://doi.org/10.1021/ac1006574.

(140) Benetti, E.; Liberto, E.; Bressanello, D.; Bordano, V.; Rosa, A. C.; Miglio, G.; Haxhi, J.; Pugliese, G.; Balducci, S.; Cordero, C. Sedentariness and Urinary Metabolite Profile in Type 2 Diabetic Patients, a Cross-Sectional Study. *Metabolites* **2020**, *10* (5). https://doi.org/10.3390/metabo10050205.

(141) Byrne, J. M.; Dubois, L. M.; Baker, J. D.; Focant, J. F.; Perrault, K. A. A Non-Targeted Data Processing Workflow for Volatile Organic Compound Data Acquired Using Comprehensive Two-Dimensional Gas Chromatography with Dual Channel Detection. *MethodsX* **2020**, *7*. https://doi.org/10.1016/j.mex.2020.101009.

(142) Altermann, M.; Kuhn, D. Vergleich Der Bodensystematischen Einheiten Der Ehemaligen DDR Mit Denen Der Bundesrepublik Deutschland. *Zeitschrift fur Angew. Geol.* **1994**, *40* (1), 1–11. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

(143) Caldeira, M.; Perestrelo, R.; Barros, A. S.; Bilelo, M. J.; Morête, A.; Câmara, J. S.; Rocha, S. M. Allergic Asthma Exhaled Breath Metabolome: A Challenge for Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2012**, *1254*, 87–97. https://doi.org/10.1016/j.chroma.2012.07.023.

(144) Steingass, C. B.; Jutzi, M.; Müller, J.; Carle, R.; Schmarr, H. G. Ripening-Dependent Metabolic Changes in the Volatiles of Pineapple (Ananas Comosus (L.) Merr.) Fruit: II. Multivariate Statistical Profiling of Pineapple Aroma Compounds Based on Comprehensive Two-Dimensional Gas Chromatography-Mass Spectrometry. *Anal. Bioanal. Chem.* **2015**, *407* (9), 2609–2624. https://doi.org/10.1007/s00216-015-8475-y.

(145) Strozier, E. D.; Mooney, D. D.; Friedenberg, D. A.; Klupinski, T. P.; Triplett, C. A. Use of Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometric Detection and Random Forest Pattern Recognition Techniques for Classifying Chemical Threat Agents and Detecting Chemical Attribution Signatures. *Anal. Chem.* **2016**, *88* (14), 7068–7075. https://doi.org/10.1021/acs.analchem.6b00725.

(146) Zhu, Y.; Lv, H. P.; Dai, W. D.; Guo, L.; Tan, J. F.; Zhang, Y.; Yu, F. L.; Shao, C. Y.; Peng, Q. H.; Lin, Z. Separation of Aroma Components in Xihu Longjing Tea Using Simultaneous Distillation Extraction with Comprehensive Two-Dimensional Gas Chromatography-Time-of-Flight Mass Spectrometry. *Sep. Purif. Technol.* **2016**, *164*, 146–154. https://doi.org/10.1016/j.seppur.2016.03.028.

(147) Dubois, L. M.; Perrault, K. A.; Stefanuto, P. H.; Koschinski, S.; Edwards, M.; McGregor, L.; Focant, J. F. Thermal Desorption Comprehensive Two-Dimensional Gas Chromatography Coupled to Variable-Energy Electron Ionization Time-of-Flight Mass Spectrometry for Monitoring Subtle Changes in Volatile Organic Compound Profiles of Human Blood. *J. Chromatogr. A* **2017**, *1501*, 117–127. https://doi.org/10.1016/j.chroma.2017.04.026.

(148) Miyazaki, T.; Okada, K.; Yamashita, T.; Miyazaki, M. Two-Dimensional Gas

Chromatography Time-of-Flight Mass Spectrometry-Based Serum Metabolic Fingerprints of Neonatal Calves before and after First Colostrum Ingestion. *J. Dairy Sci.* **2017**, *100* (6), 4354–4364. https://doi.org/10.3168/jds.2017-12557.

(149) Rees, C. A.; Stefanuto, P. H.; Beattie, S. R.; Bultman, K. M.; Cramer, R. A.; Hill, J. E. Sniffing out the Hypoxia Volatile Metabolic Signature of Aspergillus Fumigatus. *J. Breath Res.* **2017**, *11* (3). https://doi.org/10.1088/1752-7163/aa7b3e.

(150) Purcaro, G.; Rees, C. A.; Melvin, J. A.; Bomberger, J. M.; Hill, J. E. Volatile Fingerprinting of Pseudomonas Aeruginosa and Respiratory Syncytial Virus Infection in an in Vitro Cystic Fibrosis Co-Infection Model. *J. Breath Res.* **2018**, *12* (4). https://doi.org/10.1088/1752-7163/aac2f1.

(151) Beccaria, M.; Bobak, C.; Maitshotlo, B.; Mellors, T. R.; Purcaro, G.; Franchina, F. A.; Rees, C. A.; Nasir, M.; Black, A.; Hill, J. E. Exhaled Human Breath Analysis in Active Pulmonary Tuberculosis Diagnostics by Comprehensive Gas Chromatography-Mass Spectrometry and Chemometric Techniques. *J. Breath Res.* **2019**, *13* (1), 016005. https://doi.org/10.1088/1752-7163/aae80e.

(152) Shi, J.; Zhu, Y.; Zhang, Y.; Lin, Z.; Lv, H. P. Volatile Composition of Fu-Brick Tea and Pu-Erh Tea Analyzed by Comprehensive Two-Dimensional Gas Chromatography-Time-of-Flight Mass Spectrometry. *LWT* **2019**, *103* (December), 27–33. https://doi.org/10.1016/j.lwt.2018.12.075.

(153) Lukić, I.; Carlin, S.; Horvat, I.; Vrhovsek, U. Combined Targeted and Untargeted Profiling of Volatile Aroma Compounds with Comprehensive Two-Dimensional Gas Chromatography for Differentiation of Virgin Olive Oils According to Variety and Geographical Origin. *Food Chem.* **2019**, *270* (March 2018), 403–414. https://doi.org/10.1016/j.foodchem.2018.07.133.

4.2 Untargeted approaches in food-omics: the potential of comprehensive two-dimensional gas chromatography/mass spectrometry

Federico Stilo[1], Carlo Bicchi[1], Albert Robbat Jr.[2], Stephen E Reichenbach[3,4]* and Chiara Cordero[1]*

[1] Dipartimento di Scienza e Tecnologia del Farmaco, Università degli Studi di Torino, Via Pietro Giuria 9, I-10125 Torino, Italy

[2] Department of Chemistry, Tufts University, 62 Talbot Avenue, Medford, MA 02155, USA

[3] Computer Science and Engineering Department, University of Nebraska – Lincoln, 104E Avery Hall, Lincoln, NE 68588-0115, USA

[4] GC Image, PO Box 57403, Lincoln, NE 68505-7403, USA

*Corresponding authors

Prof. Dr. Chiara Cordero - Dipartimento di Scienza e Tecnologia del Farmaco, Università di Torino; Via Pietro Giuria 9, I-10125 Torino, Italy – e-mail: chiara.cordero@unito.it; phone: +39 011 6702197

Prof. Stephen E Reichenbach - Computer Science and Engineering Department, University of Nebraska – Lincoln; 260 Avery Hall, Lincoln, NE 68588-0115, USA - email: reich@unl.edu; phone +1 402 4722401

Available online: 24 December 2020

DOI: https://doi.org/10.1016/j.trac.2020.116162

TrAC - Trends in Analytical Chemistry, Volume 135, February 2021, 116162

4.2.1 Abstract

The contribution focuses on untargeted data processing/analysis approaches that are currently adopted to explore the 4D-data matrices produced by comprehensive two-dimensional gas chromatography-mass spectrometry (GC×GC-MS) in food-omics. Strategies for untargeted explorations are rationalized through the type of features adopted (*i.e.*, visual, datapoint, peak, and peak-regions) at the data processing level, and then discussed through relevant applications and illustrative examples, selected over peer-reviewed literature. The role of MS, including high *vs.* low resolution MS, as an active probe for specific cross-comparative analysis, is critically discussed also in the context of spectral deconvolution and subtraction, well-established procedures for 1D GC-MS explorations. Moreover, the challenging task of post-targeting aimed at identifying "unknown – knowns", is examined in its potential, being the key to access a higher level of information. Selected examples emphasize the importance of reliable identification by retention indexing, retention pattern ordering, sensory evaluation (sensory analysis and olfactometry), and data mining.

Keywords

comprehensive two-dimensional gas chromatography; food-omics untargeted investigations; multidimensional analytical platforms; features for cross-comparative analysis; GC×GC data processing; spectral deconvolution; MS subtraction

4.2.2 Introduction

4.2.2.1 OMICs disciplines related to food

Inspired by systems biology,[1] hypothesis-generated investigations aimed at understanding the impact the food metabolome has on the human body (nutrimetabolomics),[2] the linkage between food composition with health and well-being (foodomics)[3] and its concomitant compositional linkages to sensory quality (sensomics and flavoromics).[4,5] Toward this end, modern approaches that provide a comprehensive analysis of food metabolome [6] are fundamental to integrate the multiple dimensions of information (sensory, biological, chemical, environmental) with a comprehensive knowledge of food chemical code.

The ability to correlate food chemical patterns with orthogonal properties such as sensory quality, traceability, safety, nutritional value, and biological activity requires analytical approaches that offer total detection of the metabolome. Multidimensional analytical (MDA) platforms that combine physicochemical separation techniques (chromatography, electromigration, size exclusion, etc.) with spectroscopic techniques (mass spectrometry (MS), nuclear magnetic resonance (NMR), infrared (IR), ultraviolet (UV), etc.) have the potential to discriminate, identify, and quantify sample constituents, over the long-term, while providing a solid foundation for generating hypothesis-driven studies.

MDA technologies support comprehensive untargeted investigations by providing a larger number of features with greater information for data processing/mining. An "untargeted" approach is a process in which as many features as possible are detected and monitored with the possibility of annotation and tracking across multiple samples.[7,8] In contrast, "targeted" approaches focus on identifying and quantifying a limited number of known analytes, defined *a priori*. Target analysis, in general, provides limited information with regard to the total composition of a sample and does not take full advantage of MDA platforms' information power.

In untargeted cross-sample analyses, various important goals can be achieved, for example: (*a*) classification, (*b*) chemical fingerprinting, (*c*) monitoring (within multiple samples or time-resolved changes), (*d*) clustering, and (*e*) marker discovery.[9] **Figure 4.2.1** summarizes the key characteristics of untargeted and targeted approaches as implemented in MDA platforms.

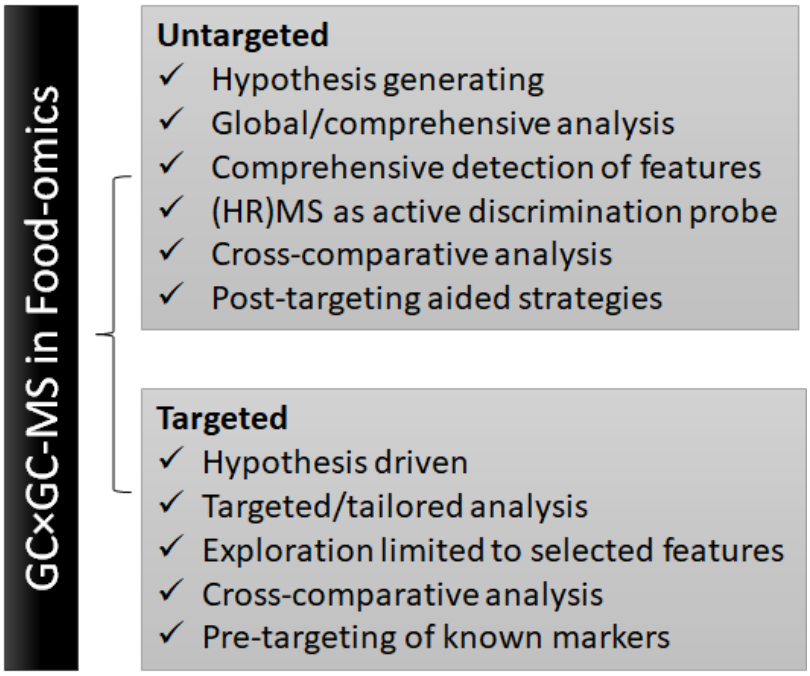

Figure 4.2.1. Key characteristics of untargeted and targeted investigations by GC×GC-MS in food-omics. Adapted from [8].

In the next section, we introduce some typical examples related to food, where untargeted characterization based on comprehensive two-dimensional gas chromatography (GC×GC) coupled to high/low resolution mass spectrometry [(HR)/(LR)-MS] has produced successful investigations. In particular, we discuss the information obtained by the analytical platform employed in light of its potential to inform on feature-identity during *ex-post* data analysis (*i.e.*, post-targeting).

4.2.2.2 Untargeted investigations and food compositional complexity: the role of GC×GC-MS

GC×GC with high or low resolution (HR)/(LR)MS has proven successful and capable of providing a high level of information[10] in untargeted investigations related to food. Applications include: (*a*) profiling and fingerprinting of volatile organic compounds (VOCs) to determine food sensory quality and spoilage, to understand the impact of environmental conditions on plant phenotype, or to discriminate samples' geographical origin,[11–20] (*b*) fatty acids methyl esters (FAMEs) profiling and fingerprinting as diagnostic markers of fat fraction quality and composition,[21,22] and (*c*) non-volatile small, primary metabolites as indicators of food quality[23] and harvesting and storage practices,[24–26] and for evaluating impacts of climate events on crops.[27,28] A few representative examples of untargeted strategies are provided in **Section 4.2.5**.

Data acquisition in untargeted explorations is fundamental to enable effective annotation and tracking of untargeted features across multiple samples. Data dependent acquisition (DDA)[8] is the most widely adopted strategy for GC×GC-MS explorations. DDA relies on full scan MS^1 to acquire individual features (accurate mass spectral signatures or fragmentation patterns) to allow specific cross-comparative analysis. Raw data is processed to produce relevant and meaningful features (datapoints or peak features in **Section 4.2.4**) that are further investigated to determine structure elucidation and identification. In contrast, data dependent acquisition (DIA) approaches integrate full MS^1 with tandem MS (MS/MS or MS^n) fragmentation to produce precursor ions, which are

either fragmented in a second stage simultaneously (MS^E)[29] or by *a priori* fixed mass ranges (sequential window acquisition of all theoretical fragment ion spectra - SWATH).[30] DIA methods are more extensively adopted in LC-MS/MS metabolomics, they produce complex fragmentation spectra that likely offer the potential to provide structure identification.[8]

In the case of MDA platforms, with comprehensive two-dimensional chromatography (GC×GC and LC×LC), untargeted investigations are corroborated by obtaining structured separation patterns for chemically correlated analytes. This key characteristic is the result of two concurrent yet independent properties. One is system orthogonality, which depends on the different separation mechanisms combined in the two analytical dimensions.[31] The second is an intrinsic property of the sample and relates to its *dimensionality* (*i.e.*, *s*),[32] a property that does not depend on the number of components (*i.e.*, *m*) in the sample, but refers to the number of independent variables that must be specified to identify the components in the sample. Briefly, the more chemical classes in the sample, the greater its dimensionality, and, consequently, the rationale for multidimensional separation.

Based on the compositional peculiarity (*i.e.*, sample dimensionality), the choice of stationary phase combination plays a fundamental role in providing group-type separations[33] or ordered homologues patterns.[31] Group-type separations also open simultaneous quantification of each chemical class as well as individual analytes within each class. For example, to assess mineral oil contamination in food and food packaging materials,[33,34] the saturated hydrocarbon fraction (MOSH) must be quantified independently from the aromatics (MOAH) and, within the latter, a few toxicologically relevant analytes need accurate quantitation by external calibration. **Figure 4.2.2** shows the chromatographic image of a mixture of mineral oils, to which were added the 16 polycyclic aromatic hydrocarbons selected by the U.S. Environmental Protection Agency (16 EPA PAH) and internal standards (Cycy, cyclohexyl cyclohexane; Cho, cholestane; MNs, 1- and 2-methyl naphthalene; Tbb, tri-tert butyl benzene; Per, perylene; DEHB, di(2- ethylhexyl) benzene; 18B, n-octadecylbenzene). The MOSH fraction (green-outlined area) is chromatographically separated from the MOAHs (blue-outlined area) due to the differential retention exerted by a medium polarity × apolar column combination.

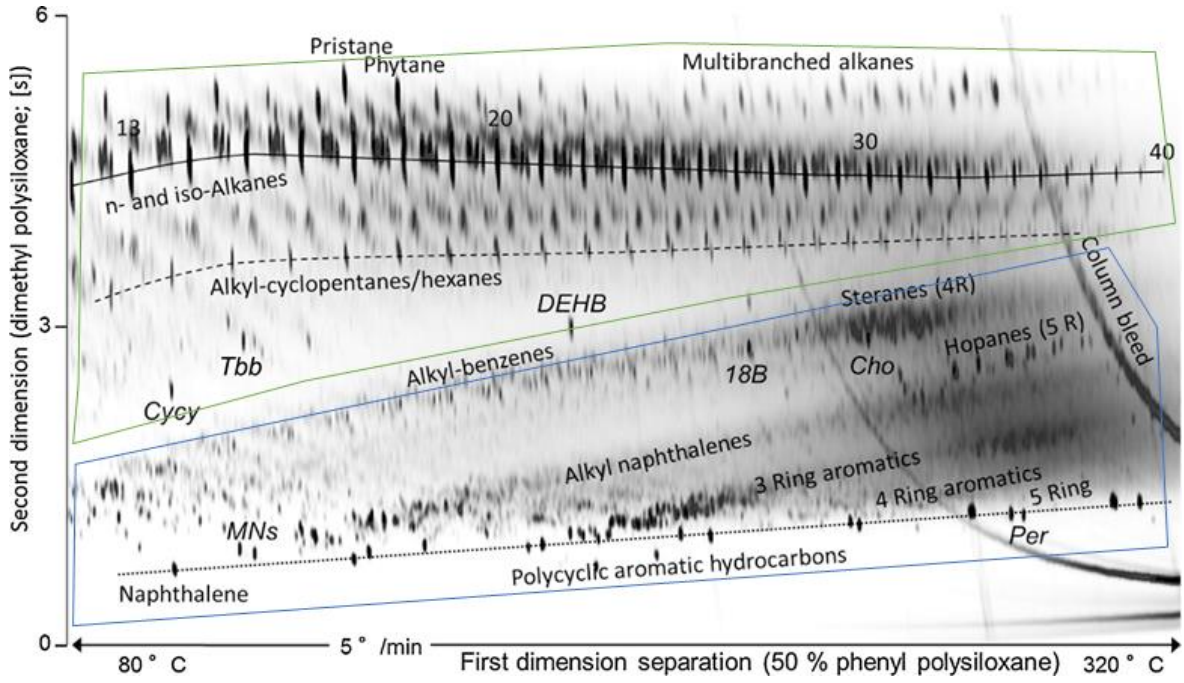

Figure 4.2.2. Contour plot of a mixture of mineral oils, to which were added the 16 polycyclic aromatic hydrocarbons selected by the U.S. Environmental Protection Agency (16 EPA PAH - EU Reg 1881/2006) and internal standards [Cycy, cyclohexyl cyclohexane; Cho, cholestane; MNs, 1- and 2-methyl naphthalene; Tbb, tri-tert butyl benzene; Per, perylene; DEHB, di(2-ethylhexyl) benzene; 18B, n-octadecylbenzene]. The green area includes the mineral oil saturated hydrocarbon fraction (MOSH) while the blue area highlights the mineral oil aromatic hydrocarbons (MOAH) elution area. The analysis was by a column set consisting of a ^{1}D OV-17 (15 m × 0.15 mm ID, 0.15 μm df) and a ^{2}D PS-255 (3.2 m × 0.15 mm ID, 0.055 μm df). Thermal modulation by loop-type modulator. Modulator capillary: 1 m of the ^{2}D column, leaving 2.2 m for the separation process. Oven temperature program: 50 °C (3 min) to 340 °C (1 min), rate 5 °C/min; P_M 6s. Modified from Biedermann and Grob.[33]

Moreover, ordered patterns are of great help in the post-targeting of relevant features; their relative retention within a homologue class and within different chemical classes helps to disambiguate, especially when linear retention data are not available or sufficiently consistent.[35–37]

4.2.3 The role of mass spectrometry in untargeted investigations

The role played by MS, the third dimension of the analytical system, is fundamental. It adds another level of specificity to cross-comparative investigations while providing essential information for analyte identification. The spectra produced by electron ionization (EI) is a third probe of discrimination, in addition to ^{1}D and ^{2}D retention data, also providing the means to identify co-eluting compounds by spectral deconvolution. EI produces stable and reproducible fragmentation patterns rendered as a compound's spectral signature. Moreover, at 70 eV, mass spectra are considered reproducible between MS instrument-types (*e.g.,* fast scanning quadrupole, time-of-flight, orbitrap, etc.)[8]. The standardization of EI ionization allows analysts to confidently use commercial or open-source spectral libraries, such as the GOLM Metabolome Database,[38] National Institute of Standards and Technology (NIST) Mass Spectral Database.[39] Wiley Registry of Mass Spectral Data,[40] and Adams' *Identification of Essential Oils Components by GC/MS*.[41]

It should also be mentioned that recently patented technologies that provide variable energy electron ionization[23,42,43] (viz., tandem ionization™) or "soft" ionization by laser photo-ionization[44] have been successfully used in untargeted GC×GC-MS studies. Their main drawback, however, is the limited potential for features identification due to the lack of available databases.

Spectra acquisition frequency, together with resolution and resolving power, are important characteristics that may have a role in the effectiveness of untargeted analysis. The highly efficient separation, correlated to the effective band compression in space provided by thermal modulation, produces very narrow modulated peaks. Peak-widths can, in fact, be as little as 100-300 ms at the baseline (w_b), thus requiring faster scanning quadrupoles and time-of-flight analyzers that achieve suitable acquisition rate for accurate peak definition.[45–47] Most TOFMS systems used for GC×GC operate at 100-200 Hz at unit mass resolution thereby providing adequate sampling density for effective spectra deconvolution.

Of the technologies capable of achieving high mass resolution, high mass accuracy, and very fast scan speeds, TOF analyzers are most often used, despite their limited application in the field of food-omics. A system commercialized by Agilent Technologies (Santa Clara, CA, USA) performs tandem mass spectrometry by coupling a low-resolution quadrupole filter to provide collision induced dissociation before the HR-TOF. Its application to a food related matrix, hop (*Humulus lupulus* L.) for beer production,[35] resulted in a wide coverage of different detectable analytes. Depending on the specific genotype, 210 to 306 unique compounds were detected, with 99 positively or tentatively identified. The authors described a workflow for consistent identification of untargeted candidates through a step-by-step procedure consisting of: (*a*) low-resolution MS fragmentation pattern search for similarity with candidates in commercial libraries; (*b*) congruency verification for molecular ion and base peaks mass accuracy (± 15 ppm) for candidates with a direct match factor ≥ 700; (*c*) ^{1}D linear retention index ($^1I^T$) consistency for isobaric compounds (± 20 units); and (*d*) position/relative retention over the 2D structure-retention pattern to complete the putative identification of unknowns.

The role of $^1I^T$ is critically important for disambiguating isobaric compounds that yield similar fragmentation patterns. Popular libraries, such as from NIST[39] and Wiley,[40] now incorporate experimental and estimated retention indices into compound records; and popular MS search software, such as the NIST MS Search software, include retention-index constraints. Conventionally, I^T values for analytes are computed from their retention times by piecewise linear or logarithmic interpolation from the retention times and indices of I^T calibration standards, *e.g.*, the Kovats or Van den Dool Indexes from *n*-alkanes. Recently, Reichenbach et al.[48] developed an ad hoc method for I^T computation that does not require I^T standards, but instead uses putative identifications returned by comprehensive, untargeted MS search based on the analytes detected by GC×GC. Based on the analysis of a terpenes' standard mixture, a consistent I^T calibration was derived from 15 of the 21 most concentrated analytes. (Three analytes were eliminated because their library record did not have an experimental I^T and three others eliminated because of inconsistent I^T, indicative of misidentification.)

4.2.4 Features and strategies for untargeted cross-comparative analysis

The goal of untargeted cross-comparative analytical strategies is to examine the distribution of all constituents in each and every sample of a set in a context where the chemical characteristics are not known *a priori*. Therefore, characteristic feature(s) must be comprehensively generated for each and every constituent. Within those features, mass spectral (total and/or selected ion) intensities provide the most comprehensive array of information for chemical identification and relative/absolute quantitation. As previously discussed, to achieve suitable method specificity, untargeted features must be: (*a*) comprehensive, to cover the full sample dimensionality; (*b*) selective, to avoid major components obscuring minor yet informative constituents; (*c*) matched

across all chromatograms, to avoid incorrect comparisons; and (*d*) accurate, to catch subtle differences between samples.

4.2.4.1 Approaches for feature generation

Different types of features have been used for untargeted investigations with GC×GC-MS, specifically: datapoints, tiles, and regions; peaks; and peak-regions. Datapoint features compare chromatograms based on the 2D arrays of GC×GC data, *i.e.*, each datapoint is a feature. For MS data, each datapoint is a mass spectrum. Tile and region features increase the granularity from single datapoints to subsume multiple datapoints per feature, with rectangular regions for tiles and more general shapes for regions. Peak features subsume multiple datapoints for each analyte peak detected in each chromatogram. Peak-region methods define a region feature for each analyte peak detected in a composite of all chromatograms.

Important considerations for choosing an approach include the difficulty of generating features, the ease of effective chemical interpretation of features, and the ability to comprehensively match features between two or more chromatograms for cross-sample analysis. Each approach has advantages and disadvantages, which are discussed at more length in the following sections, and each approach has been used for successful multi-sample analyses. In general:

- Datapoint, tile, and region features are relatively simple to generate and match comprehensively across samples but are not especially well suited for interpretive analysis of individual constituents. These approaches may be best suited for quickly evaluating chromatographic regions or groups of analytes for further investigation by other approaches.
- Peak features are, in principle, ideal for chemical interpretation because so long as a single peak is separately detected for each analyte, there is a one-to-one correspondence between analytes and features. However, in practice, peak detection is challenging and imperfect for complex samples, which effectively renders comprehensive matching across many samples incomplete and/or error prone. This approach may be best suited for analyses that yield easily interpreted but non-comprehensive results, *e.g.*, targeted analyses.
- Peak-region features require the most up-front work — datapoint alignment is required for creating a composite chromatogram and peak detection is required to define the peak-region features — but the resulting features are comprehensive, related to individual analytes, and inherently matched across chromatograms. This approach may be best suited for comprehensive, easily interpreted analyses provided that more up-front effort can be justified.

4.2.4.2 Datapoint, tile and region features

An early approach for comparing GC×GC chromatograms used datapoints features: first, for FID signals with Fisher's *f* ratio from analysis of variance (ANOVA) and principal components analysis (PCA),[49], and, later, for MS data with PCA for selected channels[50,51] and with point-wise summed *f* ratios for MS spectral channels.[52] In 2015, Pierce et al. surveyed pixel-level GC×GC data analysis, including fingerprinting and pattern recognition [53]. The Synovec Lab[54] pioneered and continues to demonstrate this approach.

Datapoint features require no extra computation to generate. They are comprehensive of all analytes and provide the highest-possible precision in the retention time plane, but there are many duplicative features (*i.e.*, multiple datapoints) per analyte, which means there is not a simple correspondence between analytes and features, and which requires substantially more computational complexity for cross-sample feature analysis. Matching of datapoint features across samples is straightforward, but retention time variations introduce matching errors unless alignment

is performed exactly. Various methods for aligning 2D chromatograms have been proposed, but none promise accuracy to within a datapoint.

Tile features, by subsuming rectangular regions of datapoints, are more computationally efficient for cross-sample analysis and less susceptible to retention time misalignment. Also, they reduce the number of features per analyte, but they do not yield one feature per analyte (nor one analyte per feature). Marney et al.[55] used tile features to reduce the feature set size and diminish adverse effects of chromatographic misalignment. Results indicated that this approach enhanced pattern recognition of true positives while simultaneously reducing the likelihood of detecting false positives. Although this approach is more computationally efficient and less susceptible to retention times misalignment, it is not fully selective, *i.e.*, there may be multiple analytes per tile. Still, it may recognize informative patterns in such tiles if the informative analyte is not dominated by other analytes in the tile.

Region features have the advantages of tile features vis a vis datapoint features, but region features can be defined to provide a better correspondence between analytes and features which makes chemical interpretation more straightforward. However, region features require a process to delineate regions. It can be relatively simple to define regions for group-type analyses, but, as discussed below for peak-region features, it is more complex to define regions that provide one-to-one correspondence with individual analytes like datapoint features, tile and region features are implicitly matched (after alignment) across chromatograms for multi-sample analyses. Honeywell UOP demonstrated region features in two standard methods for petroleum analyses with GC×GC-FID,[56,57] for which group-type analysis is important, but region features are less well suited for food-omics, for which analyses of individual compounds may be critical

4.2.4.3 Peak-based features

Peak features (*i.e.*, detected peaks with their metadata) provide a relatively straightforward basis for cross-comparative analyses of targeted analytes because data processing need only identify the detected peak for each target analyte, using its known characteristics, in each chromatogram. Then, detected peaks for each target analyte are explicitly matched across chromatograms. Untargeted cross-comparative analyses with peak features are more difficult because there is no explicit matching of peaks across chromatograms. So, before performing cross-comparisons of untargeted analytes using peak features, the software must match unidentified (or putatively identified) peaks across chromatograms — a process usually referred to as peak matching or peak tracking.

Peak features proceed from peak detection performed on each chromatogram. Ideally, peak-matching results in a list of all analytes in each chromatogram; however, due to coelutions (discussed in **Section 4.2.6**), chromatographic variability, and noise, the peak lists will have merged peaks (*i.e.*, multiple analytes per feature), fractured peaks (*i.e.*, multiple features per analyte), displaced peaks, undetected peaks, and spurious peaks. Moreover, some analytes might be detected in some samples but not in others. Thus, the peak lists vary among chromatograms, which makes matching of peaks across chromatograms extremely challenging. For untargeted analysis of complex samples, peak matching invariably results in many peaks that cannot be confidently matched across chromatograms, so the resulting analyses are not comprehensive and/or error prone. However, many automated methods have been proposed and demonstrated and the approach is used in LECO's Statistical Compare (SC) software (LECO Corp., St. Joseph MI, USA).

Our previous literature review of GC×GC peak features for comparative analysis[9] described pioneering work from 1999 through early 2011. Since then, other reviews covering methods with

comparative peak features include those by Pierce et al.,[58] Seeley and Seeley,[59] Chin and Marriott,[60] Tranchida et al.,[61] Reaser et al.,[62] and Berrier et al.[63]

4.2.4.4 Peak-regions features

Peak-region features attempt to define one region in the 2D retention time plane for each analyte peak. Peak-regions seek to achieve one-feature-to-one-analyte selectivity, as do peak features, but with implicit matching of the same peak-region across multiple chromatograms, which peak features do not provide. **Figure 4.2.3** shows the contour plot of a Yunnan tea extract submitted to oximation-silylation to map primary and specialized metabolites. **Figure 4.2.3A** shows the distribution of targeted 2D peaks (green circles) and connection lines with the internal standard (IS - 1,4-dibromobenzene) while **Figure 4.2.3B** shows untargeted–targeted (UT) peak-regions (red graphics) that comprehensively cover the chromatographic space. The enlarged area of **Figure 4.2.3C** highlights the position of the target L-Proline (green squared) and of the untargeted feature #245 with corresponding spectra.

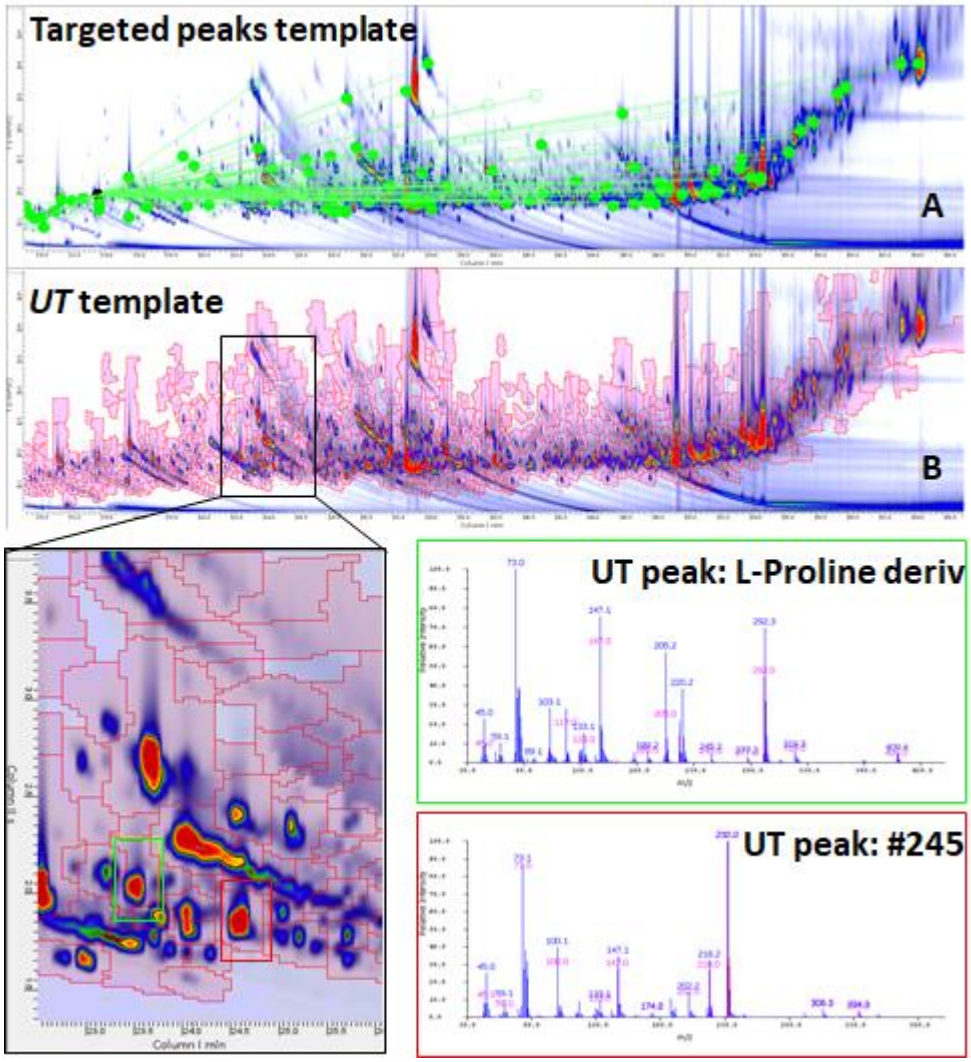

Figure 4.2.3. Contour plot of a *Yunnan* tea sample harvested after the monsoon season in 2014. (**4.2.3 A**) Shows targeted 2D peaks, highlighted with green circles, and connection lines with the internal standard. (**4.2.3 B**) Shows the distribution of UT peak-regions, identified by red graphics, comprehensively covering the chromatographic space. Enlarged area (**4.2.3 C**) highlights the position of the targeted peak L-Proline and of the untargeted peak-region #245, together with their corresponding spectra. The

column set was configured as follows: ^{1}D DB-5 (30 m × 0.25 mm ID, 0.25 μm df) and a ^{2}D OV-1701 (2 m × 0.1 mm ID, 0.1 μm df). Thermal modulation by loop-type modulator. Modulator capillary: 0.80 m of the ^{2}D column, leaving 1.2 m for the separation process. Oven program: 75 °C (1 min) to 290 °C (15 min), rate 4 °C/min; P_M 5s. From Stilo et al.[28]

Reichenbach et al.[64–66] and Schmarr et al.[26,67] described similar peak-region approaches. After preprocessing, the source chromatograms are aligned then combined (*e.g.*, simply by addition or other fusion operations)[68] to form a single composite chromatogram that comprises peaks for all analytes in all samples. Then, a region for each peak detected in the composite chromatogram is recorded in a feature template. Finally, each source chromatogram is analyzed by geometrically remapping the feature template back to each source chromatogram.

Ideally, peak-region features are comprehensively selective, accounting for each and every analyte, and feature matching is implicitly performed by mapping the template to each chromatogram. In practice, coelutions and variable retention times complicate the definition and application of peak-regions, just as for peak-based features. For GC×GC-MS datasets, coelutions within a peak-region can be addressed with quantitative ions in some cases, but deconvolution (discussed in **Section 4.2.4.5**) is required for coeluted analytes with similar spectral signatures. Variations in retention times are addressed by geometric transformations in the retention times plane, as described below. Peak-regions are supported by GC Image GC×GC and LC×LC software (GC Image, LLC, Lincoln NE, USA) in its Smart Templates™, which also support peak features for target compounds and other reliably matched peaks.

Chromatograms must be aligned to construct a composite chromatogram from which the feature template is extracted, and the feature template must be realigned to analyze each chromatogram. (Note that the transform to align a chromatogram for the composite can be inverted to map the composite feature template back to that chromatogram.) The task of chromatographic alignment for peak-region features is related to the problem of peak matching for peak features; in fact, the most common approach to chromatographic alignment is to parameterize a transformation model using a set of matched peaks.

4.2.4.5 Features based on spectral deconvolution and subtraction

Coelutions confound feature extraction regardless of the approach and type of feature analysis because the raw detector signals result from more than one analyte. Spectral deconvolution seeks to provide suitable specificity and the foundation for confident identification of relevant features. The most common algorithm for spectral deconvolution is AMDIS, developed by NIST, nowadays implemented in number of different vendor instruments. AMDIS extracts mass spectra, then fits a least-squares regression model to the compound's deconvolved ion current chromatogram. The process involves four sequential steps: 1) noise analysis, 2) component perception, 3) spectrum deconvolution, and 4) compound identification. However, AMDIS is incapable of unambiguously identifying low-concentration analytes in high-concentration matrixes unless three or more "clean" scans exist in the total ion current peak.[69,70]

LECO's (LECO Corp., St. Joseph, MI, USA) data analysis spectral deconvolution algorithm, is based on finding at least one unique target compound mass spectrum from peak scans in which chemical noise is minimal.[71,72] Once found, a system of simultaneous equations is used to deconvolve the remaining peak scans. This process is more akin to chromatographic deconvolution than "true" spectral deconvolution and it requires high-speed signal transmission of TOF filters to provide the necessary data density. Because chromatographic profiles change from sample-to-sample, even within the same sample-type, it is critically important that the data analysis be

independent of the matrix. Toward this end, the combination of both spectral deconvolution and mass spectral subtraction algorithms provides a synergistic approach resulting in the potential identification of all detectable compounds in complex samples analyzed by either 1D or 2DGC-MS.[73]

For example, Antle and co-workers[74] developed spectral deconvolution software for GC×GC-MS analysis of polycyclic aromatic hydrocarbons (PAH) and sulphur heterocycles (PASH) and their alkylated homologs in heavy hydrocarbon fractions (*i.e.*, crude oil and refined products, coal tar, and (Athabasca) tar sands). The authors used the deconvolution strategy to identify constituents and then modelled 2D component maps for these compounds to predict how they weathered in the environment.[73]

Also, Robbat and co-workers[75] developed optimized spectral deconvolution – MS subtraction algorithms to automate untargeted/targeted workflows. Once target compounds are identified by spectral deconvolution their spectra are subtracted from the TIC, resulting in peak signals that approximate the baseline or residual ion signals due to one or more coeluting compounds. If residual ion signals correspond to one compound, the peak maxima (or average) spectrum and retention time, if available, are compared to commercial libraries (NIST, Wiley, Adams, etc.). Although analyte peak areas are typically determined by deconvolving target compound signals (untargeted compounds become target compounds once input to the software), MS subtraction can also determine the peak intensities of unknowns. Annotations for identified compounds include chemical name, retention time or I^T, "clean" mass spectrum (for subtraction), relevant ions and their abundances (for deconvolution), relative peak area, etc. For compounds that cannot be identified, univocal identifiers are inputted associating retention times or I^T, spectra, relative peak areas, and sample information. If subtraction of the mass spectrum indicates multiple compound coelution, the above process repeats until residual and background signals approximate one another.

An example of how spectral deconvolution with MS subtraction was used to identify volatile specialized metabolites in tea leaves by GC×GC-MS[27,28] is shown in **Figure 4.2.4**. First, spectral deconvolution of target compounds 2-nonanone (228) and terpinolene (230) found in the TIC peak, followed by MS subtraction of each compound's mass spectrum, yields a residual signal that does not approximate background signal. Second, the library-building algorithms inspect the residual peak spectra, determining an invariant spectrum for p-cymenene (n=229 positively matched with NIST and Adams libraries). After the spectrum for this compound was subtracted from the TIC, the fragmentation pattern for unknown G48 (232) was invariant until signals from 231 and 232 caused subsequent spectra to differ. Because the methyl benzoate (231) spectrum dominated and was invariant between the tail of 232 and rising signals from 234 and 235, its spectrum, subtracted from the TIC, yielded the spectrum for undecane (233). Finally, linalool (234), a target compound, when subtracted from the TIC, yielded residual, whose peak scans were constant, unknowns G49 and G50. The figure shows the spectra produced by co-eluting compounds at specific times along the TIC peak prior to deconvolution and MS subtraction. Subtraction of the targeted/untargeted compound spectra from the TIC results in a signal that approximates background (black line).

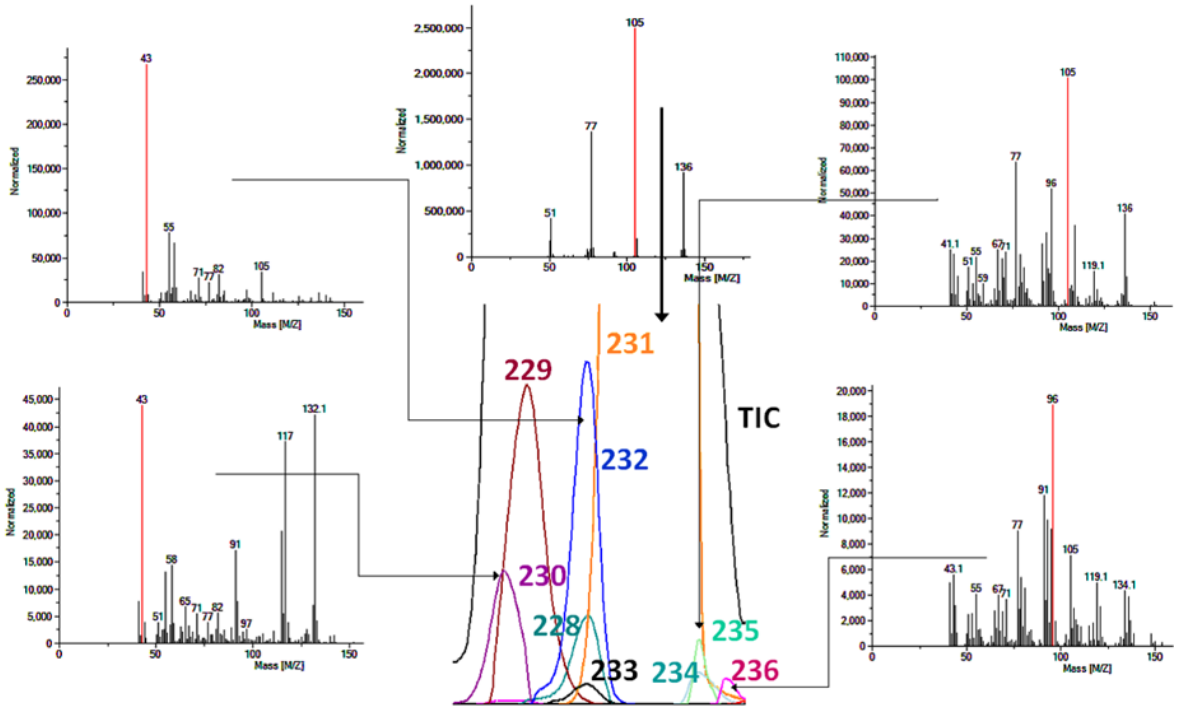

Figure 4.2.4. Illustrative example of the MS subtraction used to identify volatile specialized metabolites in tea leaves by GC×GC-MS.[27,28] The TIC trace (black-line) is subjected to sequential deconvolution and MS subtraction to isolate clean MS spectra for co-eluting analytes. The detailed description is provided in the text. Numerical coding for spectral features is as follows: #228 2-nonanone (dark green); #229 p-cymenene (red); #230 terpinolene (purple); #231 methyl benzoate (orange); #232 G48/unknown (blue); #233 n-undecane (black); #234 linalool (green); #235 G49/unknown (light green); #236 G50/unknown (pink).

4.2.5 Untargeted investigations in the food domain

4.2.5.1 Food volatilome and odorant patterns

Food volatiles are an interesting example of high complexity mixtures, resulting in thousands of detectable components that belong to different chemical classes. For this reason, thanks to its high dimensionality,[32] the food volatilome is often used as a test bench to develop new fingerprinting approaches in untargeted explorations.

In a study focused on roasted hazelnuts (*Corylus avellana* L.),[15] advanced untargeted fingerprinting methods were applied to investigate sample origin discrimination. The peak-regions features approach was adopted to build a *consensus* template composed by 422 peaks matched across the set of chromatograms. Pattern recognition by template matching using a least-squares-optimal retention time transformation, was set to establish 2D peaks correspondences while adding additional constraints on MS fragmentation pattern similarity. Within the 422 untargeted features, a sub-set of 79 analytes with known information potential (*e.g.*, indicators of thermal treatments, potent odorants or lipid oxidation products) were identified. The combined untargeted and targeted investigation was successful for both fast, untargeted origin authentication and qualification based on sensory characteristics.

Peak-regions were also adopted by Schmarr et al.[26] to study volatile patterns from apples, pears and quince fruit. Researchers designed a workflow based on 2D gel proteomics, as illustrated in **Figure 4.2.5**, where 2D images were fused and merged into a consensus pattern accounting for

more than 700 volatiles. This template pattern was then propagated to all images, by using *Delta 2D software*. The approach is comprehensive, since it computes the whole chromatographic information collected in all experiments, and relatively fast in terms of computational time. Data analysis by diverse pattern recognition tools allowed consistent clustering of similar samples with good prediction performance for unknowns. The main limit of this workflow is related to the identification of unknowns that is only possible with off-line operations back to the original acquisition software (*i.e.*, *HyperChrom* or *Xcalibur,* Thermo Fisher).

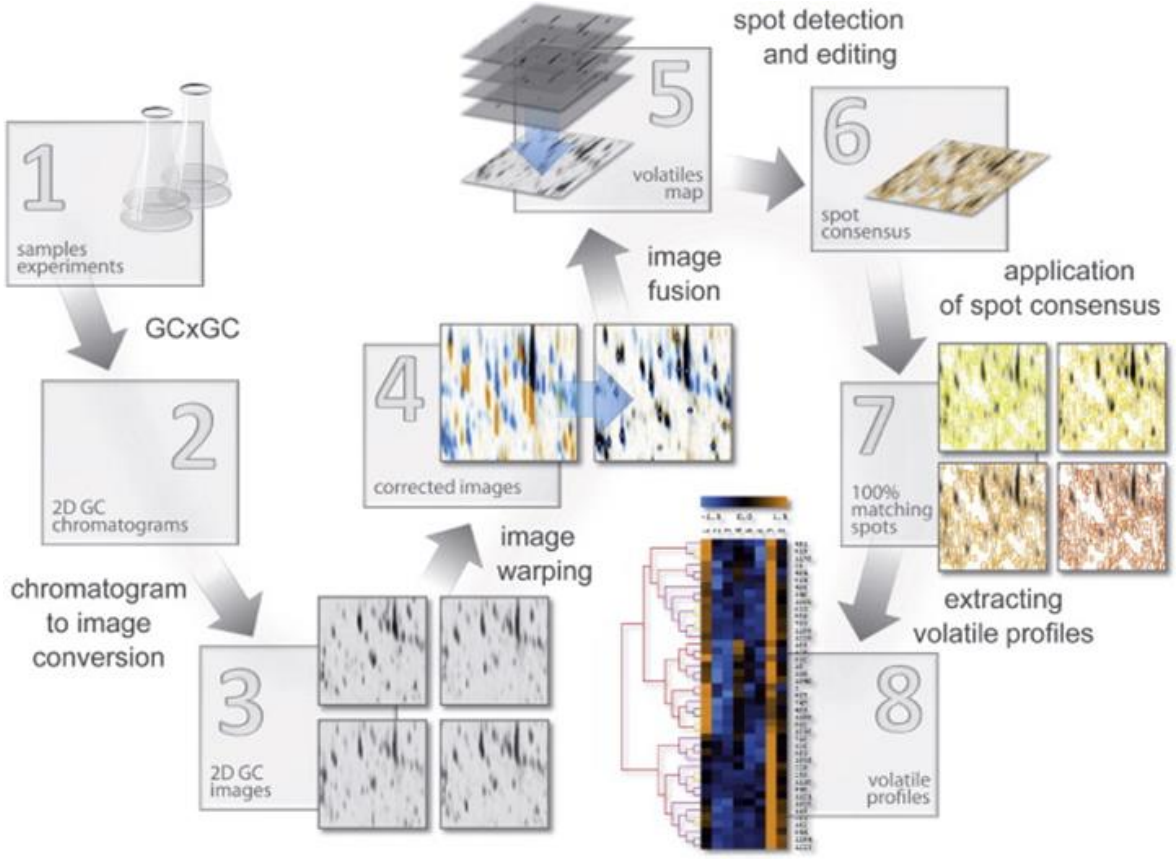

Figure 4.2.5. Workflow for the peak-region features untargeted fingerprinting proposed by Schmarr *et al*[26]. Steps: (1) samples preparation and analysis by HS-SPME-GC×GC-qMS; (2) transformation of the 2D chromatograms into 32-bit images; (3) storage of the 2D images in Delta2D software; (4) positional correction through warp vectors resulting in images congruency; (5) image fusion and volatiles mapping; (6) detection of spot consensus; (7) application of spot consensus boundaries to all 2D images for gray level integration; (8) extraction of volatiles profiles obtained by quantitative data resulting from integration. The analyses were performed through a column set consisting of a ^{1}D SolGel-Wax (30 m × 0.25 mm ID, 0.25 μm df) and a ^{2}D composed by OV-1701 (0.15 m × 0.15 mm ID, 0.15 μm df) and a BPX-5 (2 m × 0.15 mm ID, 0.15 μm df); column connections were made via press-fit connectors. Dual-jet thermal modulator. Oven program: 50 °C (5 min) to 235 °C, rate 6 °C/min (condition 1) or 50 °C (5 min) to 225 °C, rate 5 °C/min (condition 2); P_M 7s. From Schmarr et al.[26]

Long-term studies pose another challenging problem for effective untargeted fingerprinting because of shifting chromatography profiles, which makes analyte/pattern re-alignment difficult. Stilo et al.[16] simulated chromatographic misalignment by analysing volatiles from extra virgin olive oil (EVOO) samples with two different instrumental set-ups, characterized by different carrier gas linear velocities and with different modulation periods (P_M). The combined Untargeted and Targeted (UT) fingerprinting approach [76] was optimized by selecting key-processing parameters: signal-to-noise ratio (SNR) detection threshold, type of reference spectra for template construction, and similarity match factor thresholds. The computation was extended to about 1,500 peak-regions included in the template, of which 257 UT peaks were reliably matched across all-but-one chromatograms. The best performing realignment function was the second-order polynomial

transform that, combined with optimized thresholds, enabled an average cross-matching of 97.95% between *Picual* oils analyzed with severe misalignment.

EVOO volatiles were studied by Vaz-Freire et al.,[18] who adopted a datapoint features approach, based on image processing with dedicated software (*ImageJ),* National Institute of Mental Health, Bethesda, Maryland, USA) to discriminate EVOOs belonging to three olive cultivars (*i.e.*, *Galega*, *Carrasquenha* and *Cobrançosa*) obtained either with a metal hammer-decanter or a traditional metal hammer-press. The 2D contour plots were transformed into *JPEG* format digital images, converted into grey scale images, and then virtually divided in 12 quadrants (each 1,000s in the ^{1}D and 2s in the ^{2}D): the quantity of pixels for each quadrant reflected the presence of different compounds, and the values obtained were then submitted to chemometrics. Informative compounds capable of discriminating cultivars and extraction technology were highlighted and submitted to post-targeting in the acquisition software (*i.e.*, LECO's ChromaTOF software, LECO Corp., St. Joseph, MI, USA). Authors stated that a simple untargeted approach could be an effective alternative when fingerprinting is the main objective, and a full characterization of samples is not required.

Sensory analysis can guide post-targeting toward informative compounds with high discrimination power. In a study by Da Ros et al.,[77] monovarietal Italian EVOO and commercial-blended EVOO samples were screened for sensory profile by a panel test and then profiled for their volatiles by GC×GC-TOFMS. Monovarietal oils were found to be characterized by higher intensity of positive sensory attributes (apple, green grass/leaf, aromatic herbs) and absence of defects (muddy sediment, fusty, winey/vinegary, rancid, musty). An untargeted/targeted investigation based on peak features (LECO ChromaTOF software) was developed to differentiate the two groups of oils of different quality, subsequently searching for the analytes responsible of the different sensorial perceptions. Above all, the commercial-blended EVOOs were characterized by a larger number of volatile markers, including also those contributing to negative sensory notes (2-phenyl ethanol, isoamyl alcohol, isoamyl acetate, ethyl-2-methyl benzene, ethyl hexanoate, 3-methyl-2-buten-1-ol, 3-hydroxy-2-butanone), while some minor components, with positive odor qualities, were only found in the monovarietal EVOOs group.

Consistent with the above-mentioned study, some interesting applications used GC-Olfactometry (GC-O) to guide analyte post-targeting. Barbarà et al.[78] examined the impact of the grape maturation degree and maceration times on Syrah wine aroma. Untargeted peak features, resulting from GC×GC-TOFMS analyses, were submitted to supervised chemometrics to highlight discriminant analytes. Of them, 29 were identified and correlated to perceivable and pleasant sensory notes arising from GC-O-OSME (from $\theta\sigma\mu\varepsilon$, Greek word for odor). Results enabled determination of an optimal maturation degree at 19° Brix accompanied by a volatile fraction rich of compounds contributing to pleasant odors with higher persistence and intensity.

Chin et al.[79] adopted an integrated GC×GC/MDGC system with FID/O/MS detection to guide the identification of odor active analytes from Shiraz wine and coffee powder. In this study, the separation power of GC×GC and/or heart-cut 2D-GC (H/C-2DGC) are combined to the specificity of olfactometric detection that offers the unique opportunity of directing the post-targeting of unknowns toward sensory active analytes.

Peak features were used on an untargeted basis for authenticity confirmation of protected designation of origin (PDO) honeys labelled as "Corsica".[20] Honey samples (n=374) from Corsica (n=219) and from other European countries (n=155) were analyzed for their volatile fraction composition. Peak features significantly varying in their intensity among the two groups were subjected to post-targeting, with 26 of them tentatively identified. Advanced chemometrics (artificial neural networks – ANN, linear discriminant analysis – LDA, soft independent modeling

of class analogies – SIMCA, and support vector machines – SVM) were applied to discriminate Corsican honeys from others. The model constructed by SVM presented the best results, with a sensitivity of 93.2% and a specificity of 87.2%.

The challenging problem of climate change and its impact on food was tackled by Morimoto et al.[27] who studied the *Pu-erh* tea volatilome from samples harvested in the Yunnan province (China) at different elevations during the pre- and post-monsoon seasons. The study combined the *UT* fingerprinting by template matching on the total detectable volatilome with spectral deconvolution and subtraction for effective analytes targeting. In particular, by peak-regions features matching, 2D patterns were mapped and untargeted features re-aligned across all chromatograms. First, this approach, conducted with tools available in the GC Image software (GC Image LLC, Lincoln, NE, USA), resulted in the tentative identification of 107 of 300 detected compounds by spectral similarity and ${}^{I}I^{T}$ congruence. Second, spectral deconvolution, by Ion Analytics software (Gerstel, Mülheim an der Ruhr, Germany), was conducted on the raw signals to detect known compounds, followed by (MS) subtraction of their spectra from the total ion current chromatogram to reveal untargeted compounds, as described earlier. **Figure 4.2.6** illustrates the process adopted to validate the strategy that combines untargeted/targeted fingerprinting by peak-region features and template matching to sequential spectral deconvolution and subtraction.

Unknowns' spectra were then back matched with in-house and spectral libraries to allow further putative identifications. Of the 300 features, 34% were positively identified with reference standards, 42% were tentatively identified by spectral matching and ${}^{I}I^{T}$ congruence, and the remaining 24% were recoded as unknowns. The authors concluded their comparative evaluation by suggesting an integrated strategy that combines the effectiveness, in terms of computational time, of peak-region features fingerprinting to comprehensively map complex 2D patterns followed by spectral deconvolution and subtraction to highlight unknowns' locations and more efficiently directing post-targeting.

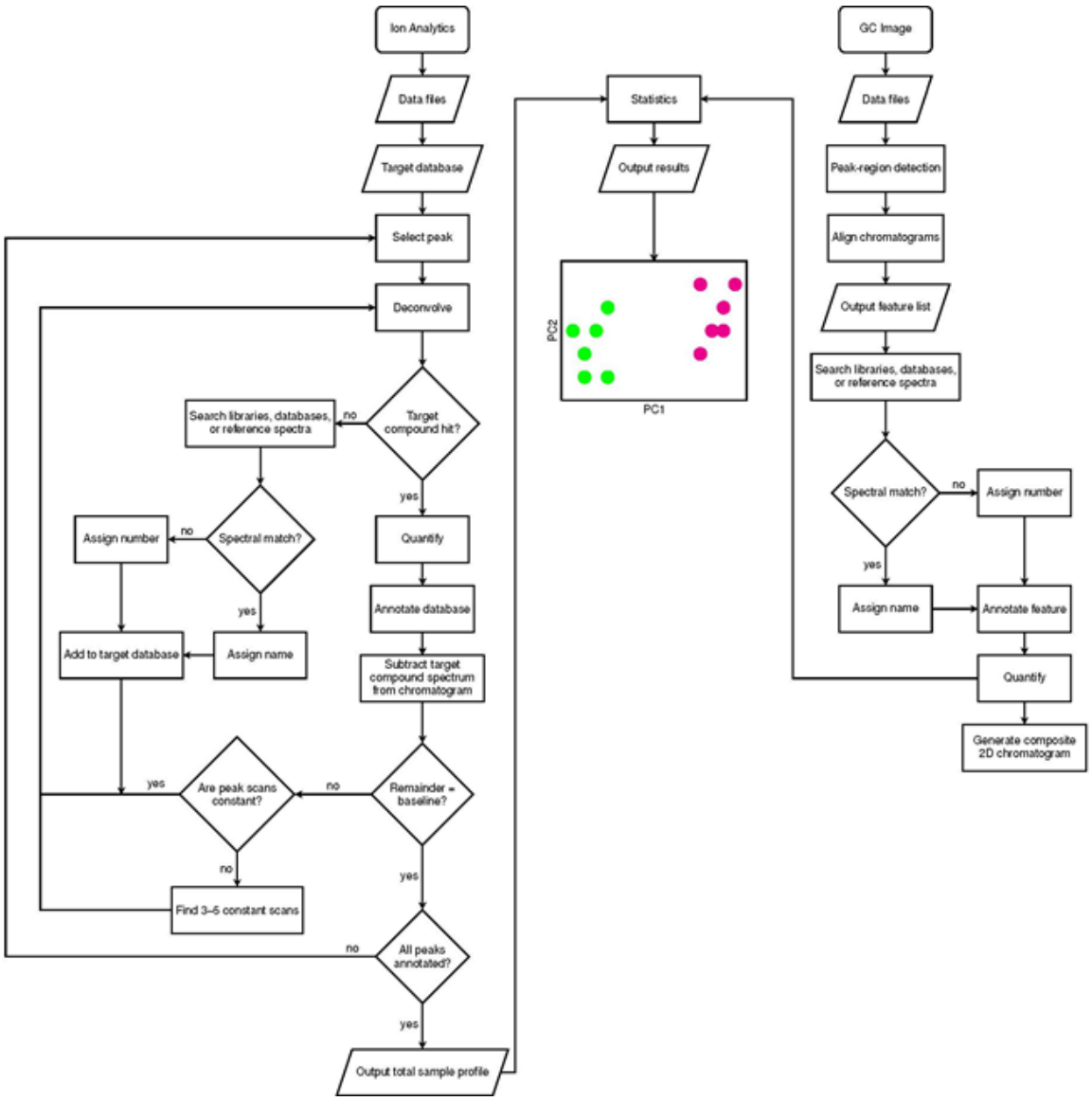

Figure 4.2.6. Flow-charts illustrating the step-by-step procedures applied to cross-validate untargeted/targeted fingerprinting of tea leaves volatile metabolome by combining peak-region features pattern recognition (by GC Image software, GC Image LLC, Lincoln, NE, USA) and deconvolution followed by MS subtraction (by Ion Analytics, Gerstel, Mullheim an der Ruhr, Germany. From Morimoto et al.[27]

Untargeted investigations were applied to support de-risking strategies on cocoa (*Theobroma cacao*). Magagna et al.[80] adopted peak-region features fingerprinting to characterize cocoa volatilome by analyzing samples from seven geographical origins (Mexico, Java, Sao Tomè, Columbia, Venezuela, Ecuador, and Trinidad) and different processing stages (raw, roasted, steamed, and nibs). The 2D peaks patterns were mapped by combining both untargeted and targeted features by the UT fingerprinting workflow. The resulting data set, accounting for 595 reliable peak regions including 130 known analytes, was adopted for classification and discrimination purposes, to explore origin, process, and sensory profile signature. In particular, raw and roasted cocoa were grouped in 3 clusters depending on their origin, and a strong accordance between targeted and untargeted fingerprinting results was revealed. The authors argued that, for classification and origin

discrimination purposes, full untargeted fingerprinting could be preferred, being effective, efficient, and less time-consuming than targeted analysis.

Humston et al.[11,12] studied cocoa volatilome to define robust markers of moisture damage. The authors selected cocoa beans of different geographical origin and with a different extent of moisture damage (usually associated with mold formation). Untargeted investigation by datapoint features with PARAFAC, mathematically resolved analytes of interest from background noise and by interferences, providing useful quantitative information. Twenty-nine informative peak features resulted from the Fisher ratio algorithm on non-moldy *vs.* moldy cocoa samples. They were successfully identified and quantified along 8 different time points. Prediction algorithms were also designed to determine whether moisture damage occurred before any visible sign of mold.

4.2.5.2 Fatty acids methyl esters – FAMEs

Fatty acids (FAs) profiling and fingerprinting provide an interesting example to show the power of GC×GC-MS, especially for its peculiar characteristic of generating ordered patterns for chemically correlated compounds. FAs from plant and animal origins are generally characterized by homologue series with even carbon number within the C_4-C_{24} range. Analysis by gas chromatography requires transformation of FAs into their methyl ester derivatives, namely, fatty acids methyl esters (FAMEs). For odd carbon number FAs and/or branched derivatives, commercial standards are not readily available, a crucial impediment for a targeted approach. Despite these shortcomings, Tranchida and co-workers[21,22] described the role post-processing played in targeting potential biomarkers and/or their disambiguation.

The signature of FAMEs from a refined hazelnuts oil was investigated [21] with a polar × apolar column combination and thermal modulation. The enlarged area of the C15-C24 elution region is shown in **Figure 4.2.7A**. The authors performed an untargeted fingerprinting, based on peak features, to reveal compositional differences within samples while structure-related patterns were further explored for post-targeting. FAMEs are aligned along diagonal retention lines showing increasing relative retention based on carbon number (along the ^{1}D) and presence and geometry of double bonds (along the ^{2}D). Mathematical functions helped in locating unknowns while assigning putative identity in absence of reference compounds [21,22]. Thanks to the enhanced sensitivity, provided by band compression in space of the thermal modulation, a series of rather unexpected odd FAMEs appeared on the 2D space plane; they were: $C_{19:0}$, $C_{21:0}$, $C_{23:0}$, $C_{25:0}$, $C_{19:1}$, $C_{17:2}$, $C_{17:3}$, etc.).

Delmonte et al.[81] went further in generating retention patterns for a FAMEs fraction prepared from human colon adenocarcinoma cells. By on-line sample hydrogenation, obtained at the conjunction between the two chromatographic dimensions of the GC×GC-FID system, FAMEs derivatives were chemically modified providing highly informative retention patterns. Hydrogenation of the analytes was accomplished using the carrier gas, H_2, by passing it through a fused silica capillary coated with Pd installed immediately before the modulation capillary. The resulting elution pattern, illustrated in **Figure 4.2.7B**, shows increasing retention along the first dimension for analytes with the same carbon number but with different numbers of double bonds. Parallel elution series, along the ^{2}D, enable effective post-targeting even for very complex mixtures as for menhaden fish oil.

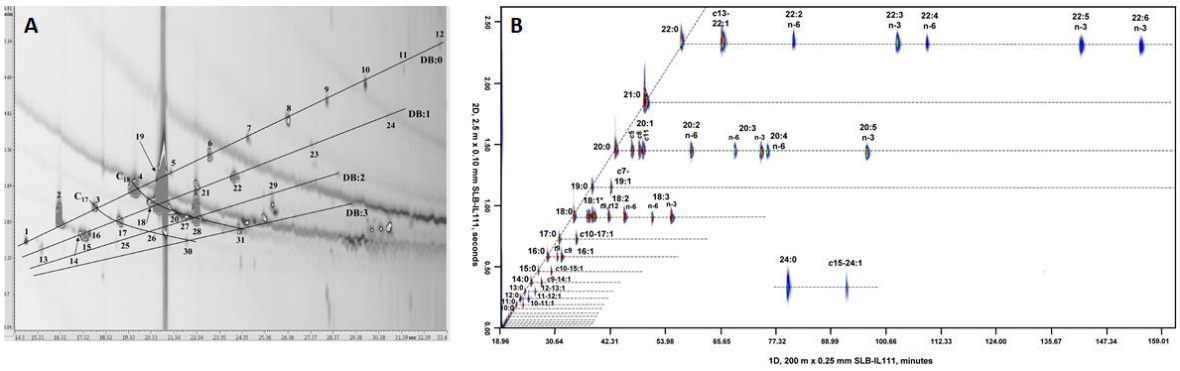

Figure 4.2.7. Contour plot (**4.2.7A**) obtained by GC×GC-FID of FAMEs from refined hazelnut oil. The column set was: ^{1}D SP-2560 (100 m × 0.25 mm ID, 0.20 μm df) and ^{2}D OV-1701 (0.9 m × 0.25 mm ID, 0.25 μm df). The modulator was a longitudinally modulated cryogenic system (LMCS). Oven program: 150 °C to 250 °C (10 min), rate 3 °C/min; P_M 6s. For peaks identification refer to the original paper from Tranchida et al.[21] Double bonds - DB. Contour plot (**4.2.7B**) obtained by GC×GC-FID with on-line hydrogenation of a FAMEs fraction prepared from human colon adenocarcinoma cells. The column set was: ^{1}D SLB-IL111 (200 m × 0.25 mm ID, 0.20 μm df) and ^{2}D SLB-IL111 (2.75m × 0.10 mm ID, 0.08 μm df). Thermal modulation by loop-type modulator. Modulator capillary: 2 m × 0.10 mm ID of deactivated fused silica capillary for the first stage and 0.25 m of the ^{2}D column for the second stage, leaving 2.5 m for the separation process. Oven program: 130 °C (23 min) to 220 °C, rate 1.3 °C/min; P_M 2s. For peaks identifications, refer to the original paper from Del Monte et al.[76] Reprinted with permission from (P. Delmonte, A.R. Fardin-Kia, J.I. Rader, Separation of Fatty Acid Methyl Esters by GC-Online Hydrogenation × GC, Anal. Chem. 85 (2013) 1517–1524). Copyright (2013) American Chemical Society.

4.2.5.3 Food metabolome

Food metabolomics is an emerging area where GC×GC could be of great help.[24] To better understand the interconnection between food chemical composition and variables influencing it (*e.g.*, crop botanical origin, harvesting area, climate impact, post-harvest treatments, storage conditions, and shelf-life), comprehensive untargeted investigations provide larger candidates lists with higher intrinsic informative power. Romo-Pérez et al. [25] investigated primary and specialized metabolites (*i.e.*, formerly defined to as secondary plant metabolites) in onions to understand the effect of cold storage, to prevent early sprouting and rooting, while acquiring knowledge on bulb composition and storability. Onions of different genotypes were selected and analyzed fresh and after 5 months of storage; 120 peak features exhibited meaningful differences as a function of storage. Among them, 56 features were submitted to post-targeting resulting in 43 analytes with a putative identification (spectral matching and $^{I}I^{T}$) and 21 that remained unidentified. In addition, by untargeted peaks distribution, it was possible to clearly discriminate the three studied genotypes in fresh onions while evidencing a confounding role of storage in this discrimination.

Wong et al.[14] focused on *Eucalyptus* spp. leaf oil analyzed by GC×GC coupled to high resolution TOFMS. The detailed untargeted profiling, based on peak features, of specialized metabolites enabled effective chemotaxonomic classification, correctly discriminating *Eucalyptus* species under study while opening further possibilities to explore synergistic relationships between chemical components. More than 400 metabolites were detected; of them, 183 were identified by post-targeting combining mass spectral profile, mass accuracy, and $^{I}I^{T}$. The power of HR-MS was exploited through confident identification of an extended list of metabolites that brings the chemical investigation closer to biologic outcomes.

Peak-region features fingerprinting was successful in deepening the understanding of extreme climate events (drought and temperature) on tea (*Camellia sinensis* L.) primary metabolome.[28] By UT fingerprinting on tea hydrophilic extracts submitted to a standard oximation-silylation protocol, primary metabolites generated complex 2D patterns accounting, on average, for 760 UT peak regions. Of them, 74 analytes were putatively identified, and their biological meaning

validated over available literature. Moreover, the untargeted comparative visualization of 2D patterns supported the fingerprinting process by making easier the location of compositional (absolute and relative) differences between samples. **Figure 4.2.8** shows a comparative visualization, by visual features fingerprinting, between a Yunnan tea harvested at high elevation (**4.2.8A**) and a Fujian tea harvested at low elevation (**4.2.8B**) both during spring season *vs.* a reference sample from Fujian but harvested at high elevation. Pair-wise comparison is rendered with a colorized fuzzy difference. Pixel's saturation (color *vs.* grey tones) indicates the magnitude of the difference between the analyzed and reference images, with grey indicating equal pixel values and bold colors (green or red) indicating large differences. Enlarged areas corresponds to monosaccharides (**a**/**c**) and specialized metabolites belonging to the chlorogenic acid and flavan-3-ol classes (**b**/**d**). Untargeted comparative visualization, when implemented with UT fingerprinting, enables the localization of key primary metabolites and other biomarkers up/down-regulated comparing pre- *vs.* post-monsoon and high- *vs.* low-elevation tea samples.

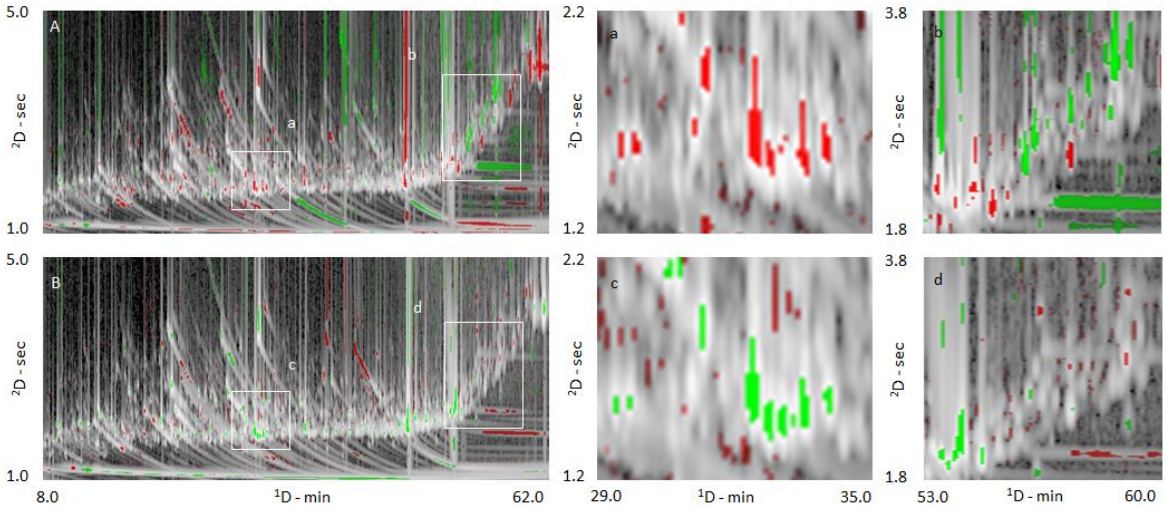

Figure 4.2.8. Comparative visualization, based on visual features, rendered with a colorized fuzzy difference of tea primary metabolites. Pair-wise comparison is between a reference tea sample from Fujian Province (China) harvested at high elevation and (**4.2.8A**) a Yunnan tea from a high elevation and (**4.2.8B**) a Fujian tea from a low elevation. The magnitude of the difference between analyzed and reference images is rendered by pixel saturation, with grey indicating equal pixel values and bold colors indicating large differences. Enlarged areas highlight absolute compositional differences of monosaccharides (**4.2.8 a/c**) and specialized metabolites belonging to chlorogenic acid and flavan-3-ol classes (**4.2.8 b/d**). The column set was configured as follows: ^{1}D DB-5 (30 m × 0.25 mm ID, 0.25 μm df) and a ^{2}D OV-1701 (2 m × 0.1 mm ID, 0.1 μm df). Thermal modulation by loop-type modulator. Modulator capillary: 0.80 m of the ^{2}D column, leaving 1.2 m for the separation process. Oven program: 75 °C (1 min) to 290 °C (15 min), rate 4 °C/min; P_M 5s. From Stilo et al. [28].

New technologies applied to EI-MS based on variable ionization energy systems offer new possibilities. In a study applying tandem ionization TOFMS, the primary metabolome of raw hazelnuts (*Corylus avellana* L.), was explored to better understand compositional differences correlated to cultivar and geographical origin.[23] By UT fingerprinting, based on peak region features, the authors delineated about 140 reliable analytes present in all-but-one sample under study. Of them, 108 analytes were putatively post-targeted. Analytical data were examined to understand the complementary nature of tandem signals (*i.e.*, 70 and 12 eV). Primary metabolites that are also known precursors of key-aroma compounds (*e.g.*, 3-methylbutanal with isoleucine; 2,3-butanedione/2,3-pentanedione with monosaccharides - fructose/glucose derivatives; 2,5-dimethylpyrazine with alanine; 1H-pyrrole, 3-methyl-1H-pyrrole, and 1H-pyrrole-2-carboxaldehyde with ornithine and alanine derivatives) were further investigated to validate linear correlations

between them. When combined, tandem signals provide useful cross-validation of results, support post-targeting by improving the relative abundance of structure-related high molecular weight fragments in the spectra, and extend the dynamic range of the method.

4.2.6 Concluding remarks

Untargeted analysis of GC×GC-MS data is undoubtedly the most profitable strategy for exploring the food metabolome, its variations due to changing inputs (*e.g.*, climate variability), and its interaction with living organisms (*e.g.*, humans, animals, microbes). MDA platforms provide multiple and orthogonal information about each sample's composition and when mass spectrometry is used at the data processing level, the comparative process becomes specific and highly informative, enabling effective post-targeting. Features, if consistently annotated and recorded together with all available analytical metadata, can be treated as "unknown – knowns" and their disambiguation/identification, when libraries and reference compounds are not available, can be postponed (*e.g.*, in *ex-post* analysis) until needed (*e.g.*, until they become important sensory contributors or health and environmental makers).

Although MDA platforms are rapidly evolving to produce robust, stable analytical systems that provide increasing multidimensional, high-density data, such systems result in increasing data analysis challenges due to the need univocally record and annotate detected features as well as re-aligning them across multiple samples and over many years. To date, only a few software tools produce automatic, intuitive 2D chromatograms processing with the possibility for supervised investigation (*e.g.*, scripting) and personalized workflows. If we can conclude that analytical chemistry tools are mature enough to tackle food samples' compositional complexity, we must affirm that analytical strategies and data processing alone cannot provide all answers. Comprehensive sampling, including all relevant variables modulating the phenomenon under study, robust experiment design and interaction between experts in different disciplines are fundamental to achieve higher level information. In this review, we found that the increased separation space offered by GC×GC coupled with MS feature alignment can detect/identify "unknown – knowns" while providing the means to obtain the total, detectable metabolome in agricultural crops and food. We expect, in the near future, a wider spread of untargeted strategies in food-omics; we know such investigations will be successful and convince the analytical chemistry community of the real potential of these tools.

References

(1) Peterson, R. T. Chemical Biology and the Limits of Reductionism. *Nat Chem Biol* **2008**, *4* (11), 635–638.

(2) Ulaszewska, M. M.; Weinert, C. H.; Trimigno, A.; Portmann, R.; Andres Lacueva, C.; Badertscher, R.; Brennan, L.; Brunius, C.; Bub, A.; Capozzi, F.; Cialiè Rosso, M.; Cordero, C. E.; Daniel, H.; Durand, S.; Egert, B.; Ferrario, P. G.; Feskens, E. J. M.; Franceschi, P.; Garcia-Aloy, M.; Giacomoni, F.; Giesbertz, P.; González-Domínguez, R.; Hanhineva, K.; Hemeryck, L. Y.; Kopka, J.; Kulling, S. E.; Llorach, R.; Manach, C.; Mattivi, F.; Migné, C.; Münger, L. H.; Ott, B.; Picone, G.; Pimentel, G.; Pujos-Guillot, E.; Riccadonna, S.; Rist, M. J.; Rombouts, C.; Rubert, J.; Skurk, T.; Sri Harsha, P. S. C.; Van Meulebroek, L.; Vanhaecke, L.; Vázquez-Fresno, R.; Wishart, D.; Vergères, G. Nutrimetabolomics: An Integrative Action for Metabolomic Analyses in Human Nutritional Studies. *Mol. Nutr. Food Res.* **2019**, *63* (1). https://doi.org/10.1002/mnfr.201800384.

(3) Miguel, H.; Carolina, S.; Virginia, G.; Elena, I.; Alejandro, C. Foodomics: MS-based Strategies in Modern Food Science and Nutrition. *Mass Spectrom. Rev.* **2011**, *31* (1), 49–69. https://doi.org/10.1002/mas.20335.

(4) Dunkel, A.; Steinhaus, M.; Kotthoff, M.; Nowak, B.; Krautwurst, D.; Schieberle, P.; Hofmann, T. Nature's Chemical Signatures in Human Olfaction: A Foodborne Perspective for Future Biotechnology. *Angew. Chemie - Int. Ed.* **2014**, *53* (28), 7124–7143. https://doi.org/10.1002/anie.201309508.

(5) Charve, J.; Chen, C.; Hegeman, A. D.; Reineccius, G. A. Evaluation of Instrumental Methods for the Untargeted Analysis of Chemical Stimuli of Orange Juice Flavour. *Flavour Fragr. J.* **2011**, *26* (6), 429–440. https://doi.org/10.1002/ffj.2078.

(6) Wishart, D. S. Metabolomics: Applications to Food Science and Nutrition Research. *Trends in Food Science and Technology*. 2008, pp 482–493. https://doi.org/10.1016/j.tifs.2008.03.003.

(7) Song, X.; Jing, S.; Zhu, L.; Ma, C.; Song, T.; Wu, J.; Zhao, Q.; Zheng, F.; Zhao, M.; Chen, F. Untargeted and Targeted Metabolomics Strategy for the Classification of Strong Aroma-Type Baijiu (Liquor) According to Geographical Origin Using Comprehensive Two-Dimensional Gas Chromatography-Time-of-Flight Mass Spectrometry. *Food Chem.* **2020**, *314*. https://doi.org/10.1016/j.foodchem.2019.126098.

(8) Schrimpe-Rutledge, A. C.; Codreanu, S. G.; Sherrod, S. D.; McLean, J. A. Untargeted Metabolomics Strategies Challenges and Emerging Directions. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (12), 1897–1905. https://doi.org/10.1007/s13361-016-1469-y.

(9) Reichenbach, S. E.; Tian, X.; Cordero, C.; Tao, Q. Features for Non-Targeted Cross-Sample Analysis with Comprehensive Two-Dimensional Chromatography. *J. Chromatogr. A* **2012**, *1226*, 140–148. https://doi.org/10.1016/j.chroma.2011.07.046.

(10) Tranchida, P. Q.; Donato, P.; Cacciola, F.; Beccaria, M.; Dugo, P.; Mondello, L. Potential of Comprehensive Chromatography in Food Analysis. *TrAC - Trends Anal. Chem.* **2013**, *52*, 186–205. https://doi.org/10.1016/j.trac.2013.07.008.

(11) Humston, E. M.; Zhang, Y.; Brabeck, G. F.; McShea, A.; Synovec, R. E. Development of a GCxGC-TOFMS Method Using SPME to Determine Volatile Compounds in Cacao Beans. *J. Sep. Sci.* **2009**, *32* (13), 2289–2295. https://doi.org/10.1002/jssc.200900143.

(12) Humston, E. M.; Knowles, J. D.; McShea, A.; Synovec, R. E. Quantitative Assessment of

Moisture Damage for Cacao Bean Quality Using Two-Dimensional Gas Chromatography Combined with Time-of-Flight Mass Spectrometry and Chemometrics. *Journal of Chromatography A*. 2010, pp 1963–1970. https://doi.org/10.1016/j.chroma.2010.01.069.

(13) Chin, S.-T.; Eyres, G. T.; Marriott, P. J. Application of Integrated Comprehensive/Multidimensional Gas Chromatography with Mass Spectrometry and Olfactometry for Aroma Analysis in Wine and Coffee. *Food Chem.* **2015**, *185*, 355–361. https://doi.org/10.1016/j.foodchem.2015.04.003.

(14) Wong, Y. F.; Perlmutter, P.; Marriott, P. J. Untargeted Metabolic Profiling of Eucalyptus Spp. Leaf Oils Using Comprehensive Two-Dimensional Gas Chromatography with High Resolution Mass Spectrometry: Expanding the Metabolic Coverage. *Metabolomics* **2017**, *13* (5), 1–17. https://doi.org/10.1007/s11306-017-1173-3.

(15) Cordero, C.; Liberto, E.; Bicchi, C.; Rubiolo, P.; Schieberle, P.; Reichenbach, S. E.; Tao, Q. Profiling Food Volatiles by Comprehensive Two-Dimensional Ga Schromatography Coupled with Mass Spectrometry: Advanced Fingerprinting Approaches for Comparative Analysis of the Volatile Fraction of Roasted Hazelnuts (Corylus Avellana L.) from Different Ori. *J. Chromatogr. A* **2010**, *1217* (37), 5848–5858. https://doi.org/10.1016/j.chroma.2010.07.006.

(16) Stilo, F.; Liberto, E.; Reichenbach, S. E.; Tao, Q.; Bicchi, C.; Cordero, C. Untargeted and Targeted Fingerprinting of Extra Virgin Olive Oil Volatiles by Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry: Challenges in Long-Term Studies. *J. Agric. Food Chem.* **2019**, *67* (18), 5289–5302. https://doi.org/10.1021/acs.jafc.9b01661.

(17) Ros, A. D.; Masuero, D.; Riccadonna, S.; Bubola, K. B.; Mulinacci, N.; Mattivi, F.; Lukić, I.; Vrhovsek, U. Complementary Untargeted and Targeted Metabolomics for Differentiation of Extra Virgin Olive Oils of Different Origin of Purchase Based on Volatile and Phenolic Composition and Sensory Quality. *Molecules* **2019**, *24* (16). https://doi.org/10.3390/molecules24162896.

(18) Vaz-Freire, L. T.; da Silva, M. D. R. G.; Freitas, A. M. C. Comprehensive Two-Dimensional Gas Chromatography for Fingerprint Pattern Recognition in Olive Oils Produced by Two Different Techniques in Portuguese Olive Varieties Galega Vulgar, Cobrançosa e Carrasquenha. *Anal. Chim. Acta* **2009**, *633* (2), 263–270. https://doi.org/10.1016/j.aca.2008.11.057.

(19) Cajka, T.; Riddellova, K.; Klimankova, E.; Cerna, M.; Pudil, F.; Hajslova, J. Traceability of Olive Oil Based on Volatiles Pattern and Multivariate Analysis. *Food Chem.* **2010**, *121* (1), 282–289. https://doi.org/10.1016/j.foodchem.2009.12.011.

(20) Stanimirova, I.; Üstün, B.; Cajka, T.; Riddelova, K.; Hajslova, J.; Buydens, L. M. C.; Walczak, B. Tracing the Geographical Origin of Honeys Based on Volatile Compounds Profiles Assessment Using Pattern Recognition Techniques. *Food Chem.* **2010**, *118* (1), 171–176. https://doi.org/10.1016/j.foodchem.2009.04.079.

(21) Tranchida, P. Q.; Giannino, A.; Mondello, M.; Sciarrone, D.; Dugo, P.; Dugo, G.; Mondello, L. Elucidation of Fatty Acid Profiles in Vegetable Oils Exploiting Group-Type Patterning and Enhanced Sensitivity of Comprehensive Two-Dimensional Gas Chromatography. *J. Sep. Sci.* **2008**, *31* (10), 1797–1802. https://doi.org/10.1002/jssc.200800002.

(22) Tranchida, P. Q.; Donato, P.; Dugo, G.; Mondello, L.; Dugo, P. Comprehensive

Chromatographic Methods for the Analysis of Lipids. *TrAC - Trends Anal. Chem.* **2007**, *26* (3), 191–205. https://doi.org/10.1016/j.trac.2007.01.006.

(23) Rosso, M. C.; Mazzucotelli, M.; Bicchi, C.; Charron, M.; Manini, F.; Menta, R.; Fontana, M.; Reichenbach, S. E.; Cordero, C. Adding Extra-Dimensions to Hazelnuts Primary Metabolome Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography Combined with Time-of-Flight Mass Spectrometry Featuring Tandem Ionization: Insights on the Aroma Potential. *J. Chromatogr. A* **2020**, *1614*. https://doi.org/10.1016/j.chroma.2019.460739.

(24) Weinert, C. H.; Egert, B.; Kulling, S. E. On the Applicability of Comprehensive Two-Dimensional Gas Chromatography Combined with a Fast-Scanning Quadrupole Mass Spectrometer for Untargeted Large-Scale Metabolomics. *J. Chromatogr. A* **2015**, *1405*, 156–167. https://doi.org/10.1016/j.chroma.2015.04.011.

(25) Romo-Pérez, M. L.; Weinert, C. H.; Häußler, M.; Egert, B.; Frechen, M. A.; Trierweiler, B.; Kulling, S. E.; Zörb, C. Metabolite Profiling of Onion Landraces and the Cold Storage Effect. *Plant Physiol. Biochem.* **2020**, *146*, 428–437. https://doi.org/10.1016/j.plaphy.2019.11.007.

(26) Schmarr, H.-G.; Bernhardt, J. Profiling Analysis of Volatile Compounds from Fruits Using Comprehensive Two-Dimensional Gas Chromatography and Image Processing Techniques. *J. Chromatogr. A* **2010**, *1217* (4), 565–574. https://doi.org/10.1016/j.chroma.2009.11.063.

(27) Morimoto, J.; Rosso, M. C.; Kfoury, N.; Bicchi, C.; Cordero, C.; Robbat, A. Untargeted/Targeted 2D Gas Chromatography/Mass Spectrometry Detection of the Total Volatile Tea Metabolome. *Molecules* **2019**, *24* (20), 1–14. https://doi.org/10.3390/molecules24203757.

(28) Stilo, F.; Tredici, G.; Bicchi, C.; Robbat, A.; Morimoto, J.; Cordero, C. Climate and Processing Effects on Tea (Camellia Sinensis L. Kuntze) Metabolome: Accurate Profiling and Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography/Time-of-Flight Mass Spectrometry. *Molecules* **2020**, *25* (10), 2447. https://doi.org/10.3390/molecules25102447.

(29) Plumb, R. S.; Johnson, K. A.; Rainville, P.; Smith, B. W.; Wilson, I. D.; Castro-Perez, J. M.; Nicholson, J. K. UPLC/MSE; a New Approach for Generating Molecular Fragment Information for Biomarker Structure Elucidation. *Rapid Commun. Mass Spectrom.* **2006**, *20* (13), 1989–1994. https://doi.org/10.1002/rcm.2550.

(30) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-Independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. & Cell. Proteomics* **2012**, *11* (6), O111.016717. https://doi.org/10.1074/mcp.O111.016717.

(31) Venkatramani, C. J.; Xu, J.; Phillips, J. B. Separation Orthogonality in Temperature-Programmed Comprehensive Two-Dimensional Gas Chromatography. *Anal. Chem.* **1996**, *68* (9), 1486–1492. https://doi.org/10.1021/ac951048b.

(32) Giddings, J. C. Sample Dimensionality: A Predictor of Order-Disorder in Component Peak Distribution in Multidimensional Separation. *J. Chromatogr. A* **1995**, *703* (1–2), 3–15. https://doi.org/10.1016/0021-9673(95)00249-M.

(33) Biedermann, M.; Munoz, C.; Grob, K. Update of On-Line Coupled Liquid Chromatography

– Gas Chromatography for the Analysis of Mineral Oil Hydrocarbons in Foods and Cosmetics. *J. Chromatogr. A* **2017**, *1521*, 140–149. https://doi.org/10.1016/j.chroma.2017.09.028.

(34) Biedermann, M.; Grob, K. Comprehensive Two-Dimensional Gas Chromatography for Characterizing Mineral Oils in Foods and Distinguishing Them from Synthetic Hydrocarbons. *J. Chromatogr. A* **2015**, *1375*, 146–153. https://doi.org/10.1016/j.chroma.2014.11.064.

(35) Yan, D. D.; Wong, Y. F.; Tedone, L.; Shellie, R. A.; Marriott, P. J.; Whittock, S. P.; Koutoulis, A. Chemotyping of New Hop (Humulus Lupulus L.) Genotypes Using Comprehensive Two-Dimensional Gas Chromatography with Quadrupole Accurate Mass Time-of-Flight Mass Spectrometry. *J. Chromatogr. A* **2017**. https://doi.org/10.1016/j.chroma.2017.08.020.

(36) Cordero, C.; Cagliero, C.; Liberto, E.; Nicolotti, L.; Rubiolo, P.; Sgorbini, B.; Bicchi, C. High Concentration Capacity Sample Preparation Techniques to Improve the Informative Potential of Two-Dimensional Comprehensive Gas Chromatography-Mass Spectrometry: Application to Sensomics. *J. Chromatogr. A* **2013**, *1318*, 1–11. https://doi.org/10.1016/j.chroma.2013.09.065.

(37) Magagna, F.; Cordero, C.; Cagliero, C.; Liberto, E.; Rubiolo, P.; Sgorbini, B.; Bicchi, C. Black Tea Volatiles Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography – Mass Spectrometry Combined with High Concentration Capacity Sample Preparation Techniques: Toward a Fully Automated Sensomic Assessment. *Food Chem.* **2017**, *225*, 276–287. https://doi.org/10.1016/j.foodchem.2017.01.003.

(38) Hummel, J.; Selbig, J.; Walther, D.; Kopka, J. The Golm Metabolome Database: A Database for GC-MS Based Metabolite Profiling BT - Metabolomics: A Powerful Tool in Systems Biology; Nielsen, J., Jewett, M. C., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2007; pp 75–95. https://doi.org/10.1007/4735_2007_0229.

(39) Scientific, N.; Databases, T. NIST/EPA/NIH Mass Spectral Library with Search Program: (Data Version: NIST 08, Software Version 2.0f). National Institute of Standards and Technology (NIST): Gaithersburg MD 2005.

(40) McLafferty, F. W. Wiley Registry of Mass Spectral Data, 12th Edition. Wiley 2020.

(41) Adams, R. P. *IDENTIFICATION OF ESSENTIAL OIL COMPONENTS BY GAS CHROMATOGRAPHY/ MASS SPECTROMETRY, 5th Ed.*; Texensis Publishing: Gruver, TX, USA, 2017.

(42) Markes International. Select-EV: The next Generation of Ion Source Technology. *Technical Note*. 2016.

(43) Cordero, C.; Guglielmetti, A.; Bicchi, C.; Liberto, E.; Baroux, L.; Merle, P.; Tao, Q.; Reichenbach, S. E. Comprehensive Two-Dimensional Gas Chromatography Coupled with Time of Flight Mass Spectrometry Featuring Tandem Ionization: Challenges and Opportunities for Accurate Fingerprinting Studies. *J. Chromatogr. A* **2019**, *1597*, 132–141. https://doi.org/10.1016/j.chroma.2019.03.025.

(44) Welthagen, W.; Mitschke, S.; Muhlberger, F.; Zimmermann, R. One-Dimensional and Comprehensive Two-Dimensional Gas Chromatography Coupled to Soft Photo Ionization Time-of-Flight Mass Spectrometry : A Two- and Three-Dimensional Separation Approach. *J. Chromatogr. A* **2007**, *1150*, 54–61. https://doi.org/10.1016/j.chroma.2007.03.033.

(45) Beens, J.; Janssen, H. G.; Adahchour, M.; Brinkman, U. A. T. Flow Regime at Ambient Outlet Pressure and Its Influence in Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2005**, *1086* (1–2), 141–150. https://doi.org/10.1016/j.chroma.2005.05.086.

(46) Cordero, C.; Bicchi, C.; Joulain, D.; Rubiolo, P. Identification, Quantitation and Method Validation for the Analysis of Suspected Allergens in Fragrances by Comprehensive Two-Dimensional Gas Chromatography Coupled with Quadrupole Mass Spectrometry and with Flame Ionization Detection. *J. Chromatogr. A* **2007**, *1150* (1–2), 37–49. https://doi.org/10.1016/j.chroma.2006.08.079.

(47) Purcaro, G.; Tranchida, P. Q.; Ragonese, C.; Conte, L.; Dugo, P.; Dugo, G.; Mondello, L. Evaluation of a Rapid-Scanning Quadrupole Mass Spectrometer in an Apolar × Ionic-Liquid Comprehensive Two-Dimensional Gas Chromatography System. *Anal. Chem.* **2010**, *82* (20), 8583–8590. https://doi.org/10.1021/ac101678r.

(48) Reichenbach, S. E.; Tao, Q.; Cordero, C.; Bicchi, C. A Data-Challenge Case Study of Analyte Detection and Identification with Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry (GC×GC-MS). *Separations* **2019**, *6* (3), 38. https://doi.org/10.3390/separations6030038.

(49) Johnson, K. J.; Synovec, R. E. Pattern Recognition of Jet Fuels: Comprehensive GC × GC with ANOVA-Based Feature Selection and Principal Component Analysis. In *Chemometrics and Intelligent Laboratory Systems*; 2002; Vol. 60, pp 225–237. https://doi.org/10.1016/S0169-7439(01)00198-8.

(50) Mohler, R. E.; Dombek, K. M.; Hoggard, J. C.; Young, E. T.; Synovec, R. E. Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Analysis of Metabolites in Fermenting and Respiring Yeast Cells. *Anal. Chem.* **2006**, *78* (8), 2700–2709. https://doi.org/10.1021/ac052106o.

(51) Pierce, K. M.; Hope, J. L.; Hoggard, J. C.; Synovec, R. E. A Principal Component Analysis Based Method to Discover Chemical Differences in Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry (GC × GC-TOFMS) Separations of Metabolites in Plant Samples. *Talanta* **2006**, *70* (4), 797–804. https://doi.org/10.1016/j.talanta.2006.01.038.

(52) Pierce, K. M.; Hoggard, J. C.; Hope, J. L.; Rainey, P. M.; Hoofnagle, A. N.; Jack, R. M.; Wright, B. W.; Synovec, R. E. Fisher Ratio Method Applied to Third-Order Separation Data to Identify Significant Chemical Components of Metabolite Extracts. *Anal. Chem.* **2006**, *78* (14), 5068–5075. https://doi.org/10.1021/ac0602625.

(53) Pierce, K. M.; Parsons, B. A.; Synovec, R. E. Pixel-Level Data Analysis Methods for Comprehensive Two-Dimensional Chromatography. In *Data Handling in Science and Technology*; Elsevier B.V., 2015; Vol. 29, pp 427–463. https://doi.org/10.1016/B978-0-444-63527-3.00010-2.

(54) Synovec, R. E. Synovec Lab - Gas Chromatography, Liquid Chromatography, and Mass Spectrometry, with Multi-Dimensional Data Analysis.

(55) Marney, L. C.; Christopher Siegler, W.; Parsons, B. A.; Hoggard, J. C.; Wright, B. W.; Synovec, R. E. Tile-Based Fisher-Ratio Software for Improved Feature Selection Analysis of Comprehensive Two-Dimensional Gas Chromatography-Time-of-Flight Mass Spectrometry Data. *Talanta*. 2013, pp 887–895.

https://doi.org/10.1016/j.talanta.2013.06.038.

(56) UOP. *UOP 965-10, Total Cycloparaffins and Total Aromatics in Synthetic Paraffinic Kerosene Fuels by Comprehensive Two-Dimensional GC with Flame Ionization Detection*; Des Plaines, IL, 2010.

(57) UOP. *UOP990-11, Organic Analysis of Distillate by Comprehensive Two-Dimensional Gas Chromatography with Flame Ionization Detection*; Des Plaines, IL, 2011.

(58) Pierce, K. M.; Kehimkar, B.; Marney, L. C.; Hoggard, J. C.; Synovec, R. E. Review of Chemometric Analysis Techniques for Comprehensive Two Dimensional Separations Data. *J. Chromatogr. A* **2012**, *1255*, 3–11. https://doi.org/10.1016/j.chroma.2012.05.050.

(59) Seeley, J. V.; Seeley, S. K. Multidimensional Gas Chromatography: Fundamental Advances and New Applications. *Anal. Chem.* **2013**, *85* (2), 557–578. https://doi.org/10.1021/ac303195u.

(60) Chin, S.-T.; Marriott, P. J. Multidimensional Gas Chromatography beyond Simple Volatiles Separation. *Chem. Commun.* **2014**, *50*, 8819–8833.

(61) Tranchida, P. Q.; Purcaro, G.; Maimone, M.; Mondello, L. Impact of Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry on Food Analysis. *J. Sep. Sci.* **2016**, *39* (1), 149–161. https://doi.org/10.1002/jssc.201500379.

(62) Reaser, B. C.; Watson, Nathanial E. Prebihalo, S. E.; Pinkerton, D. K.; Skogerboe, K. J.; Synovec, R. E. Management and Interpretation of Capillary Chromatographymass Spectrometry Data. In *Hyphenations of Capillary Chromatography with Mass Spectrometry*; Tranchida, P., Mondello, L., Eds.; Elsevier: Amsterdam, 2019; pp 449–480.

(63) Berrier, K. L.; Prebihalo, S. E.; Synovec, R. E. *Advanced Data Handling in Comprehensive Two-Dimensional Gas Chromatography*; Elsevier, 2020; Vol. 12. https://doi.org/10.1016/B978-0-12-813745-1.00007-6.

(64) Cordero, C.; Liberto, E.; Bicchi, C.; Rubiolo, P.; Reichenbach, S. E.; Tian, X.; Tao, Q. Targeted and Non-Targeted Approaches for Complex Natural Sample Profiling by GC×GC-QMS. *J. Chromatogr. Sci.* **2010**, *48* (4), 251–261. https://doi.org/10.1093/chromsci/48.4.251.

(65) Reichenbach, S. E.; Tian, X.; Tao, Q.; Stoll, D. R.; Carr, P. W. Comprehensive Feature Analysis for Sample Classification with Comprehensive Two-Dimensional LC. *J. Sep. Sci.* **2010**, *33* (10), 1365–1374. https://doi.org/10.1002/jssc.200900859.

(66) Reichenbach, S. E.; Tian, X.; Tao, Q.; Ledford, E. B.; Wu, Z.; Fiehn, O. Informatics for Cross-Sample Analysis with Comprehensive Two-Dimensional Gas Chromatography and High-Resolution Mass Spectrometry (GCxGC-HRMS). *Talanta* **2011**, *83* (4), 1279–1288. https://doi.org/10.1016/j.talanta.2010.09.057.

(67) Schmarr, H. G.; Bernhardt, J.; Fischer, U.; Stephan, A.; Müller, P.; Durner, D. Two-Dimensional Gas Chromatographic Profiling as a Tool for a Rapid Screening of the Changes in Volatile Composition Occurring Due to Microoxygenation of Red Wines. *Anal. Chim. Acta* **2010**. https://doi.org/10.1016/j.aca.2010.05.002.

(68) Luhn, S.; Berth, M.; Hecker, M.; Bernhard, J. Using Standard Positions and Image Fusion to Create Proteome Maps from Collections of Two-Dimensional Gel Electrophoresis Images. In *Proteomics*; 2003. https://doi.org/10.1002/pmic.200300433.

(69) Sichilongo, K.; Banda, D. GC-MS Determination of Targeted Pesticides in Environmental

Samples from the Kafue Flats of Zambia. *Bull. Environ. Contam. Toxicol.* **2013**, *91* (5), 510–516. https://doi.org/10.1007/s00128-013-1087-3.

(70) Furtula, V.; Derksen, G.; Colodey, A. Application of Automated Mass Spectrometry Deconvolution and Identification Software for Pesticide Analysis in Surface Waters. *J. Environ. Sci. Health. B.* **2006**, *41* (8), 1259—1271. https://doi.org/10.1080/03601230600962211.

(71) Robbat, A.; Hoffmann, A.; MacNamara, K.; Huang, Y. Quantitative Identification of Pesticides as Target Compounds and Unknowns by Spectral Deconvolution of Gas Chromatographic/Mass Spectrometric Data. *J. AOAC Int.* **2008**, *91* (6), 1467–1477. https://doi.org/10.1093/jaoac/91.6.1467.

(72) Cochran, J. W. Fast Gas Chromatography-Time-of-Flight Mass Spectrometry of Polychlorinated Biphenyls and Other Environmental Contaminants. *J. Chromatogr. Sci.* **2002**, *40* (5), 254–268. https://doi.org/10.1093/chromsci/40.5.254.

(73) Antle, P. M.; Zeigler, C. D.; Livitz, D. G.; Robbat, A. Two-Dimensional Gas Chromatography/Mass Spectrometry, Physical Property Modeling and Automated Production of Component Maps to Assess the Weathering of Pollutants. *J. Chromatogr. A* **2014**, *1364*, 223–233. https://doi.org/10.1016/j.chroma.2014.08.033.

(74) Antle, P. M.; Zeigler, C. D.; Gankin, Y.; Robbat, A. New Spectral Deconvolution Algorithms for the Analysis of Polycyclic Aromatic Hydrocarbons and Sulfur Heterocycles by Comprehensive Two-Dimensional Gas Chromatography-Quadrupole Mass Spectrometery. *Anal. Chem.* **2013**, *85* (21), 10369–10376. https://doi.org/10.1021/ac402336j.

(75) Robbat, A.; Kfoury, N.; Baydakov, E.; Gankin, Y. Optimizing Targeted/Untargeted Metabolomics by Automating Gas Chromatography/Mass Spectrometry Workflows. *J. Chromatogr. A* **2017**, *1505*, 96–105. https://doi.org/10.1016/j.chroma.2017.05.017.

(76) Magagna, F.; Valverde-Som, L.; Ruíz-Samblás, C.; Cuadros-Rodríguez, L.; Reichenbach, S. E.; Bicchi, C.; Cordero, C. Combined Untargeted and Targeted Fingerprinting with Comprehensive Two-Dimensional Chromatography for Volatiles and Ripening Indicators in Olive Oil. *Anal. Chim. Acta* **2016**, *936*, 245–258. https://doi.org/10.1016/j.aca.2016.07.005.

(77) Ros, A. Da; Masuero, D.; Riccadonna, S.; Bubola, K. B.; Mulinacci, N.; Mattivi, F.; Lukić, I.; Vrhovsek, U. Complementary Untargeted and Targeted Metabolomics for Differentiation of Extra Virgin Olive Oils of Different Origin of Purchase Based on Volatile and Phenolic Composition and Sensory Quality. *Molecules* **2019**, *24* (16), 1–17. https://doi.org/10.3390/molecules24162896.

(78) Aith Barbará, J.; Primieri Nicolli, K.; Souza-Silva, É. A.; Camarão Telles Biasoto, A.; Welke, J. E.; Alcaraz Zini, C. Volatile Profile and Aroma Potential of Tropical Syrah Wines Elaborated in Different Maturation and Maceration Times Using Comprehensive Two-Dimensional Gas Chromatography and Olfactometry. *Food Chem.* **2020**, *308*, 125552. https://doi.org/10.1016/j.foodchem.2019.125552.

(79) Chin, S. T.; Marriott, P. J. Review of the Role and Methodology of High Resolution Approaches in Aroma Analysis. *Anal. Chim. Acta* **2015**, *854*, 1–12. https://doi.org/10.1016/j.aca.2014.06.029.

(80) Magagna, F.; Guglielmetti, A.; Liberto, E.; Reichenbach, S. E.; Allegrucci, E.; Gobino, G.; Bicchi, C.; Cordero, C. Comprehensive Chemical Fingerprinting of High-Quality Cocoa at

Early Stages of Processing: Effectiveness of Combined Untargeted and Targeted Approaches for Classification and Discrimination. *J. Agric. Food Chem.* **2017**, *65* (30), 6329–6341. https://doi.org/10.1021/acs.jafc.7b02167.

(81) Delmonte, P.; Fardin-Kia, A. R.; Rader, J. I. Separation of Fatty Acid Methyl Esters by GC-Online Hydrogenation × GC. *Anal. Chem.* **2013**, *85* (3), 1517–1524. https://doi.org/10.1021/ac302707z.

4.3 Comprehensive two-dimensional gas chromatography as a boosting technology in food-omics investigations

Federico Stilo[1], Carlo Bicchi[1], Stephen E. Reichenbach[2,3], Chiara Cordero[1]*

[1] Dipartimento di Scienza e Tecnologia del Farmaco, Università degli Studi di Torino, Via Pietro Giuria 9, I-10125 Torino, Italy

[2] Computer Science and Engineering Department, University of Nebraska – Lincoln, 104 E Avery Hall, Lincoln, NE 68588-0115, USA

[3] GC Image, PO Box 57403, Lincoln, NE 68505-7403, USA

* Corresponding author:

Prof. Chiara Cordero - Dipartimento di Scienza e Tecnologia del Farmaco, Università di Torino, Via Pietro Giuria 9, I-10125 Torino, Italy – e-mail: chiara.cordero@unito.it; phone: +39 011 6707172

Manuscript received: 10 January 2021

Manuscript revised: 11 February 2021

Manuscript accepted: 11 February 2021

Accepted manuscript online: 14 February 2021

Version of Record online: 19 March 2021

Available online: 28 November 2020

DOI: https://doi.org/10.1002/jssc.202100017

JSS – Journal of Separation Science, January 2021, 1-20

4.3.1 Abstract

This review focuses on the role that GC×GC can play within the investigation workflows of food-omics and related disciplines and sub-disciplines, including food metabolomics, nutrimetabolomics, sensomics, and food safety. After a short introductory survey, discussing the intriguing context of system biology and integrationist approaches of investigation, the concepts of analytical dimensions and the key-characteristics of GC×GC are introduced. Through a selection of relevant examples, the boosting role of GC×GC within food-omics is described, providing to the reader evidence of how comprehensive multidimensional separations-based platforms have introduced new concepts and tools in the analytical measurement of complex biological samples.

Key words

comprehensive two-dimensional gas chromatography, sensomics, nutrimetabolomics, food metabolomics, food safety

4.3.2 Introduction

Comprehensive two dimensional gas chromatography (Comprehensive 2DGC or GC×GC) is a highly effective realization of the definition of a multidimensional (MD) technique introduced by Giddings in 1987:[1] "*a separation system [is] multi-dimensional when i) all components of a sample are subjected to two or more separation steps, in which their displacements are dependent on different factors, and when ii) two or more components are substantially separated in any single step, they should remain separated until the end of the separation process."* This definition clearly stated the concept of multidimensionality, restricting the then-contemporary techniques that conformed with it to two-dimensional thin layer (TLC) and paper chromatography, widely used by biochemists. It took only four years for John Phillips with Zaiyou Liu *to materialize* this theoretical concept for gas chromatography (GC) and to develop a fit-for-purpose instrument.[2] The synergic achievements of the concepts developed by Giddings and Phillips have been the basis of other MD chromatographic techniques not only involving GC but also liquid chromatography (LC) and supercritical fluid chromatography (SFC) such as LC×LC, SFC×SFC, and SFC×LC.

As for all innovative techniques, GC×GC had to overcome three phases before achieving a full recognition. The introduction and optimization of its theory and operation covered a time span of about 10 years (1991-2000) during which the petrochemical industry played a fundamental role for GC×GC development because of its continuous need to characterize high complexity sample.[3,4] Next, popularization of the technique covered another time span of about 10 years (2001-2012), during which GC×GC was opened to widely differing disciplines.[5–7] This development also coincided with the routine introduction of fingerprinting approaches in the biological studies in which analysis, or better a comprehensive "view" of high complexity mixtures in a single step, became fundamental for applications in the omics disciplines and in particular to metabolomics (see below).[8] These years also coincided with a considerable evolution of the hardware, standardization of instrumentation, optimization of GC×GC coupling with MS, and development of specially dedicated software.[5,6,9,10] Finally, during the current period, GC×GC has moved out of the academic laboratories to enter routine activity even as analytical innovation influences the evolution of the strategies of analysis (*e.g.*, the adoption "*dilute-and-shoot*" methods).

In this context, research studies adopting GC×GC as primary analytical tool, should: (a) design a suitable configuration of the analytical platform, by combining all relevant analytical dimensions; (b) define the data processing approach that maximizes the effective extraction of relevant information from raw data; and (c) apply suitable data mining to interpret results and achieve higher levels of information about the phenomenon under study. The elements of this rational approach are visualized in **Figure 4.3.1** together with their interrelation with food omics investigative goals. A detailed discussion of individual key elements is outside the scope of this contribution; readers interested in more recent advancements on data processing and data mining for "omics" areas might refer to recent reviews.[9,11–14]

GC×GC is nowadays a self-consistent technique fully independent but complementary to GC, covering important fields of applications that mono-dimensional GC (1D-GC) cannot achieve with a single analysis, such as the analysis of highly complex mixtures with more than 250 components, the practical limit today achievable with 1D GC.

After a (short) schematic introductory survey discussing the principles of system biology and the concept of analytical dimensions, this overview focuses on the role that GC×GC can play with the different disciplines and subdisciplines related to the food field (food-omics) including food metabolomics, nutrimetabolomics, sensomics, and food safety.

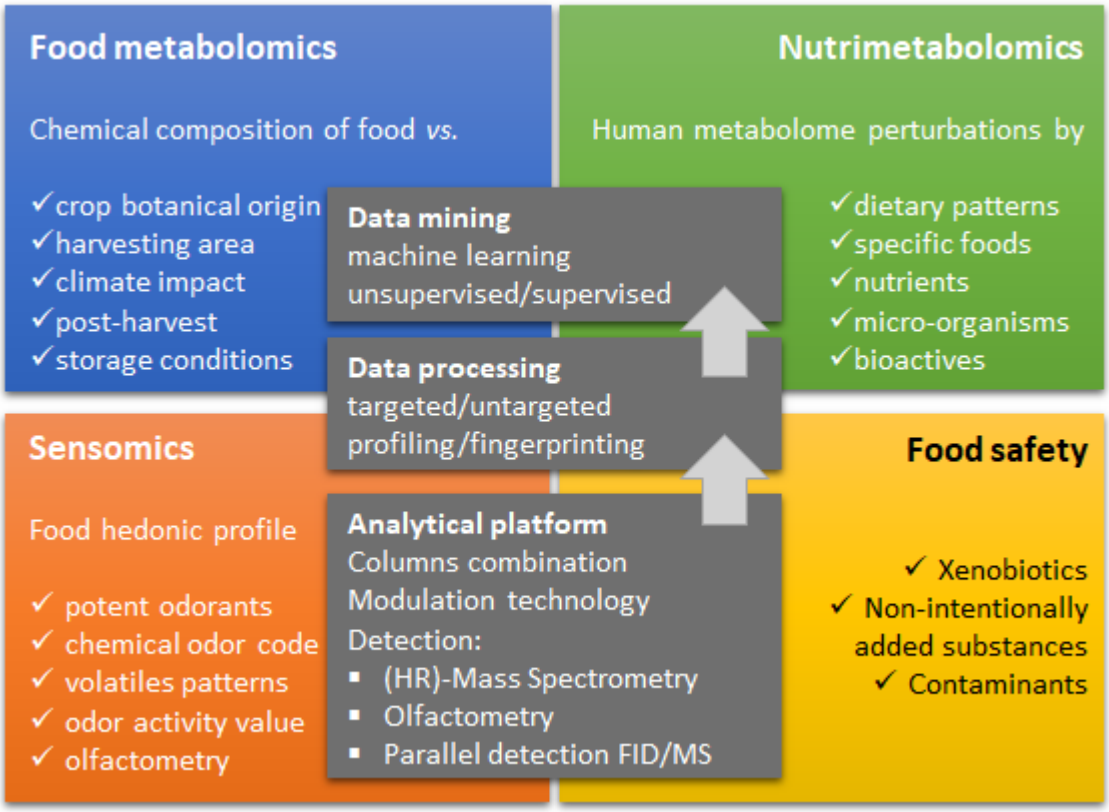

Figure 4.3.1. Main investigation goals of food omics (i.e., food metabolomics, sensomics, nutrimetabolomics, and food safety) correlated with the steps of a rational approach adopting GC×GC as primary analytical tool: analytical platform configuration, data processing and data mining.

4.3.2.1 From reductionism to interactionism: the role of analytical chemistry

The complexity of biological systems and the multi-level interactions behind biological phenomena have motivated the development and evolution of system biology, a discipline that tries to model the complexity by unraveling higher-order network structures and relationships between conditions, taking into account the effect of molecular patterns on multiple targets.[15] From the opposite direction, at least from a conceptual viewpoint, chemical biology, growing in parallel to system biology, offers well-consolidated strategies and approaches aimed at pairing individual compounds and targets. A "one compound/one target" mindset also referred to as reductionist approach biology.[15]

Food chemical investigations moved from reductionism to the integrationist mindset of system biology[15] by following, from one side, the ever increasing demand for high-quality food with high nutritional value and bio-active beneficial components, and, on the other side, the evolution of analytical chemistry measurements concepts and tools. It was in fact foreseen in 1991 by R. Wilson in an editorial for Analytical Chemistry[16] that:: "*A persistent research frontier is the analytical chemistry of the mixtures of chemical substances generated by living organisms. The problems of separation, molecular identification, and quantification of these mixtures are enormous. They are the ultimate molecular mishmash. The challenges they offer have demanded, and produced, many new concepts in chemical measurement science*".

For chromatographers, 1991 marked a turning point with the first paper introducing the concept and the tools to realize comprehensive two-dimensional separation in gas phase.[2] Liu and

Phillips and presented the separation of a standard mixture obtained by coupling in-series two capillary columns, a first dimension (^{1}D) polyethylene glycol (PEG) column [21 m × 0.25 mm id × 0.25 um d_f] and a second dimension (^{2}D) apolar dimethyl polysiloxane column [1 m × 0.10 mm id × 0.10 um d_f] interfaced with a thermal modulator. The separation was achieved in 2.5 min with a modulation period (P_M) of 2 s and a clear ordered pattern was obtained with a relative retention over the separation space patterned by the analytes' relative volatility and polarity.

Since then, although with an "induction period" of several years due to skepticisms of several expert chromatographers who were not fully convinced of the advantages and potential of GC×GC, the technique entered in several application fields as an alternative to 1D-GC or as a replacement of it.[3–5,17–23] The possibilities offered by improved separation, resulting in higher separation power, efficiency, and resolution accompanied by band-compression in space, which improves methods sensitivity, make GC×GC the tool of choice to explore complex samples chemical dimensions,[24] matching integrationist needs for informative patterns of triggering molecules.

However, to achieve a suitable informative level regarding sample composition, multiple analytical dimensions must be combined appropriately. The next section introduces the concept of multidimensional analytical (MDA) platforms and illustrates common GC×GC configurations and their informative dimensions.

4.3.2.2 Analytical dimensions

MDA platforms combine orthogonal principles of measurement through the hyphenation of physicochemical discrimination by chromatography, electromigration, size exclusion, etc. and spectroscopic techniques (*e.g.*, mass spectrometry (MS), nuclear magnetic resonance (NMR), infrared (IR), ultraviolet (UV), etc.). By collecting multiplex analytical information, the process has the potential of discriminating, identifying, and quantifying sample constituents, or homologous classes, even when sample dimensionality is high and the presence of interferents might challenge the measurement process. Moreover, the higher the "orthogonality degree" of the information collected, the larger the opportunities to comprehensively approach food.[13,15,18,25]

Although the integrationist approach looks for patterns of chemicals rather than for single components/biomarkers, the analytical process should provide consistent information on analytes' identities and amounts – relative or absolute. To achieve these goals, MS operating at unit-mass resolution or by exact mass assignments (with high-resolution MS - HRMS) is fundamental. Molecular features and functionality can be derived by fragmentation patterns obtained through electron ionization (EI-MS) either at standard 70 eV or at lower energies,[26] by molecular ions or adducts produced with chemical ionization (CI-MS); or by multiple fragmentations from tandem MS (MS/MS or MS^n). However, due to the great compositional complexity of food fractions that includes hundreds of isomer/isobaric components, univocal identification and robust quantitation cannot be achieved if the measurement is not accompanied by adequate physicochemical discrimination, provided by chromatography (1D or GC×GC). Moreover, when food sensory properties are of interest for the investigation, if not the primary aim, odor quality assignment through olfactometry becomes an essential dimension of the analytical system. The section devoted to sensomics provides a deeper insight on the role of olfactometry, on-line or off-line, combined with GC×GC platforms.

The separation power of GC×GC is undoubtedly a key characteristic that leads to the production of highly detailed 2D chromatographic fingerprints with rationally ordered retention structures that open new and interesting perspectives for data processing, *e.g.*, group-type

analyses,[27–29] while increasing confident identification of unknowns.[30] For example, the appropriate combination of column dimensions, stationary phase chemistries, and modulation dynamics in rationalizing the retention logic of alkyl substituted pyrazines from roasted coffee, was discussed by Ryan *et al.*[31] and by Mondello *et al.*[32] This was one of the first demonstrations of the advantages provided by GC×GC in group-type investigations of food. The authors adopted a polar × apolar column combination (^{1}D PEG × ^{2}D 5% phenyl - 95% dimethylpolysiloxane) and achieved univocal identification by combining linear retention indices (I^T) in the ^{1}D, MS fragmentation patterns by EI and time-of-flight MS (TOF MS) and 2D retention logic. **Figure 4.3.2** illustrates the ordered retention pattern of twenty alkyl pyrazines from a roasted Robusta coffee sampled by headspace solid-phase microextraction (HS-SPME) followed by GC×GC-FID with a longitudinally modulated cryogenic system (MLCS) operating with liquid CO_2 as cryogen. Pyrazines follow a rational order along the ^{1}D by the molecular weight and alkylation degree, whereas along the ^{2}D, the relative retention is more strongly influenced by alkylation.

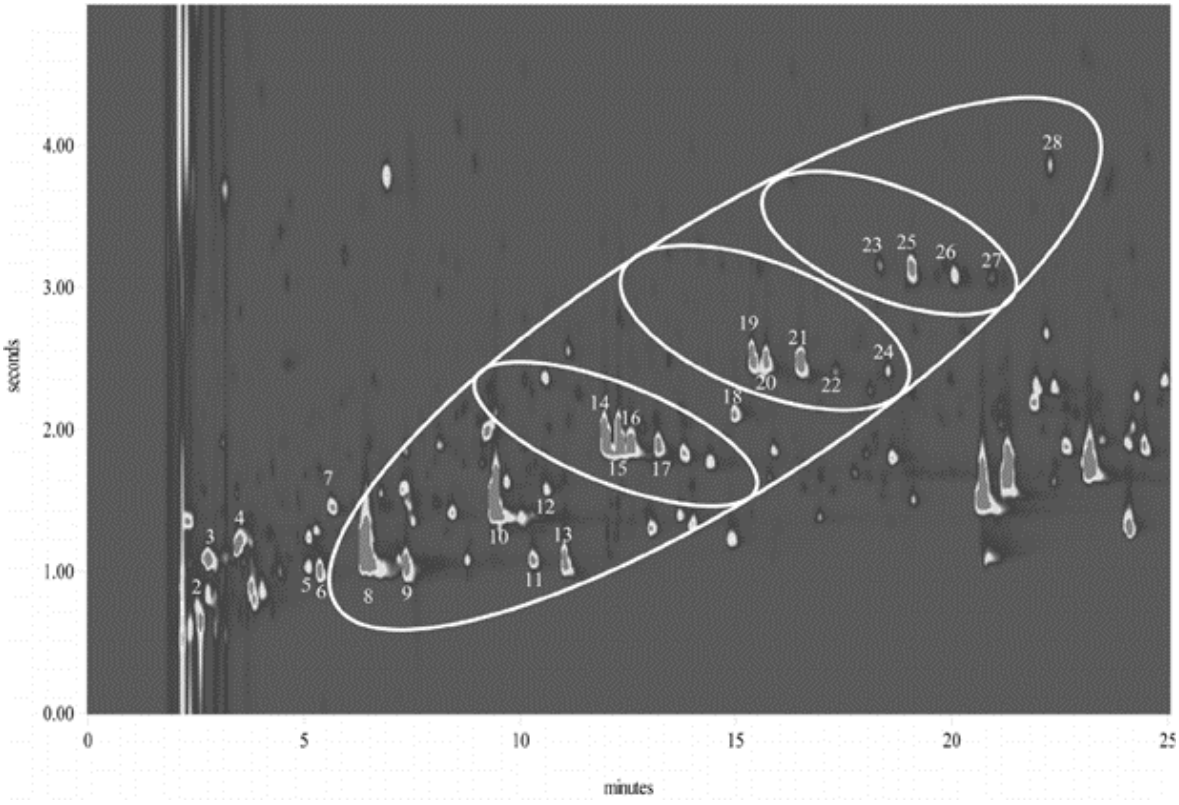

Figure 4.3.2. Ordered retention pattern of twenty alkyl pyrazines from roasted Robusta coffee volatiles sampled by headspace solid-phase microextraction (HS-SPME). Compounds assignment are: (9) for pyrazine, (10) for 2-methylpyrazine, (14) for 2,5-dimethylpyrazine, (15) for 2,6-dimethylpyrazine, (16) for 2-ethylpyrazine, (17) for 2,3-dimethylpyrazine, (19) for 2-ethyl-6-methylpyrazine, (20) for 2-ethyl-5-methylpyrazine, (21) for 2,3,5-trimethylpyrazine, (22) for 2-ethyl-3-methylpyrazine, (23) for 2,6-diethylpyrazine, 25 2-ethyl-3,5-dimethylpyrazine, (26) for 2,3-diethylpyrazine, (27) for 2-ethyl-3,6-dimethylpyrazine, (28) for 3,5-diethyl-2-methylpyrazine. The analyses were performed by a column set consisting of a ^{1}D Supelcowax-10 (30 m × 0.25 mm ID, 0.25 mm) and a ^{2}D SPB-5 (0.1 × 0.1 mm ID, 0.1 mm). Thermal modulation was by longitudinally modulated cryogenic system (LMCS) operating with liquid CO_2 as cryogenics. Modulator capillary: 1 m × 0.10 mm ID of deactivated fused silica capillary. Oven program: 60°C (5 min) to 230°C (5 min), at 1.5°C/min, and to 280°C (2 min) at 50°C/min, P_M 5s. From Mondello *et al.*[32].

Many fermented primary materials or processes for food have clear signatures of homologues deriving from common biosynthetic pathways (*e.g.*, short-chain fatty acids – SCFAs and primary

alcohols in cocoa fermentation),[33–36] or whose formation derives from common reaction frameworks (*e.g.*, Maillard reaction in roasted foods).[37–39] SCFAs, aliphatic alcohols, linear saturated and unsaturated aldehydes, esters, furans, and pyrroles are, among the others, compound classes with rational retention logic that is of great help in identification and disambiguation.

Besides ordered retention logic, the role of ^{1}D I^T is fundamental for disambiguating isomers/isobaric compounds with similar EI-MS patterns. The most popular MS databases, such as those from NIST[40] and Wiley,[41] incorporate experimental and estimated retention indices on most common stationary phases, making possible by MS search software the application of retention-index constraints. This option was pioneered by Mondello and co-workers[42,43] who developed one of the first interactive-databases for flavor and fragrance applications. More recently, Reichenbach *et al.*[44] developed an algorithm for I^T computation that does not require calibration by standard *n*-alkanes or fatty acids methyl esters (FAMEs). By this approach, putative identifications obtained by automatic and comprehensive, untargeted MS search on sample's constituents, generate from their retention times the locking points to compute piecewise linear or logarithmic interpolation of I^T, *e.g.*, the Kovats (with isothermal analysis) or Van den Dool Indices (*i.e.*, linear retention indices). **Figure 4.3.3A** plots, by piecewise linear interpolation function, ^1t_R *vs.* Library I^T for 21 analytes of a terpene standard mixture. I^T are those incorporated into database entry for the top ranked analyte resulted by direct match factor (DMF) similarity. **Figure 4.3.3B** shows the reliable model for preliminary I^T calibration without standardization, obtained by eliminating 3 entries (IDs #8, #18, and #21 [44]) for which analyst supervision revealed incorporated I^T inconsistencies.

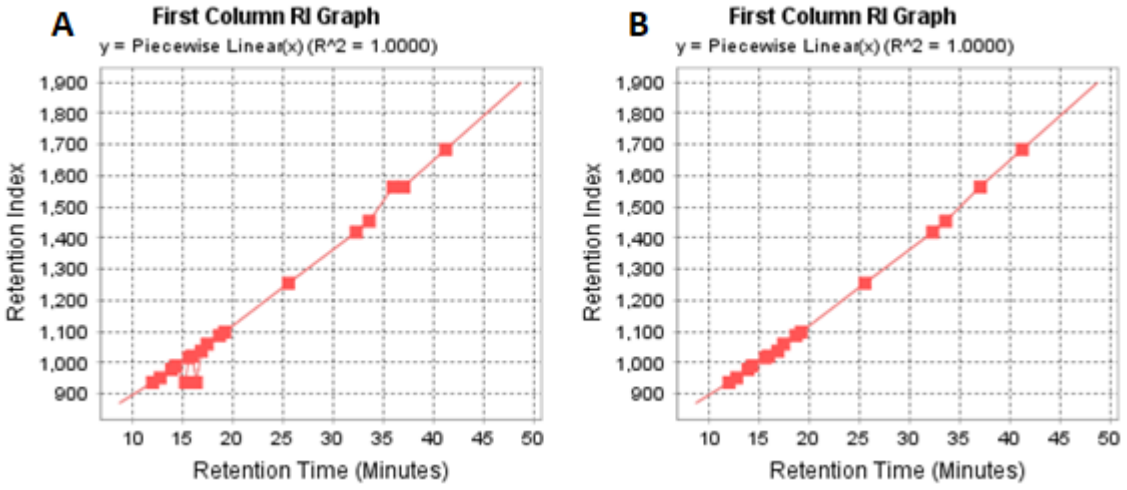

Figure 4.3.3. I^T calibration based on fitting the calibration model to the observed relationship between ^1t_R and the top ranked analyte resulted by direct match factor (DMF) similarity. (**4.3.3A**) 18 of Top 21 analytes (3 #1 hits lacked library I^T); (**4.3.3B**) 15 of Top 21 analytes consistent with piecewise linear I^T calibration model. Adapted from Reichenbach et al.[44]

When the high compositional complexity of processed food prevents the clear identification of component signatures, on both retention patterns and MS spectra, dedicated data processing tools could be of help. If homologues patterns are mutually overlapped, the selective extraction of diagnostic MS signatures from the third analytical dimension by scripting functions[27,28,45,46] enable effective isolation of individual components and classes accompanied by 2D visualization. **Figure 4.3.4** shows pseudocolor images of whole dried milk volatiles extracted by dynamic headspace (D-HS) by means of polydimethylsiloxane (PDMS) trap (**4.3.4A** and **4.3.4B**) or by headspace sorptive extraction (HSSE) on a PDMS sorptive phase. **Figure 4.3.4A** reports the TIC trace obtained by a

polar × medium-polar column combination; **Figure 4.3.4B** shows the resulting image after applying a CLIC™ Expression (GC-Image, LLC Lincoln NE, USA) targeted to linear saturated aldehydes MS spectral signature (prioritizing 57, 82, 95 *m/z*); and **Figure 4.3.4C** shows selected lactones (prioritizing 55, 71, 99 *m/z*).

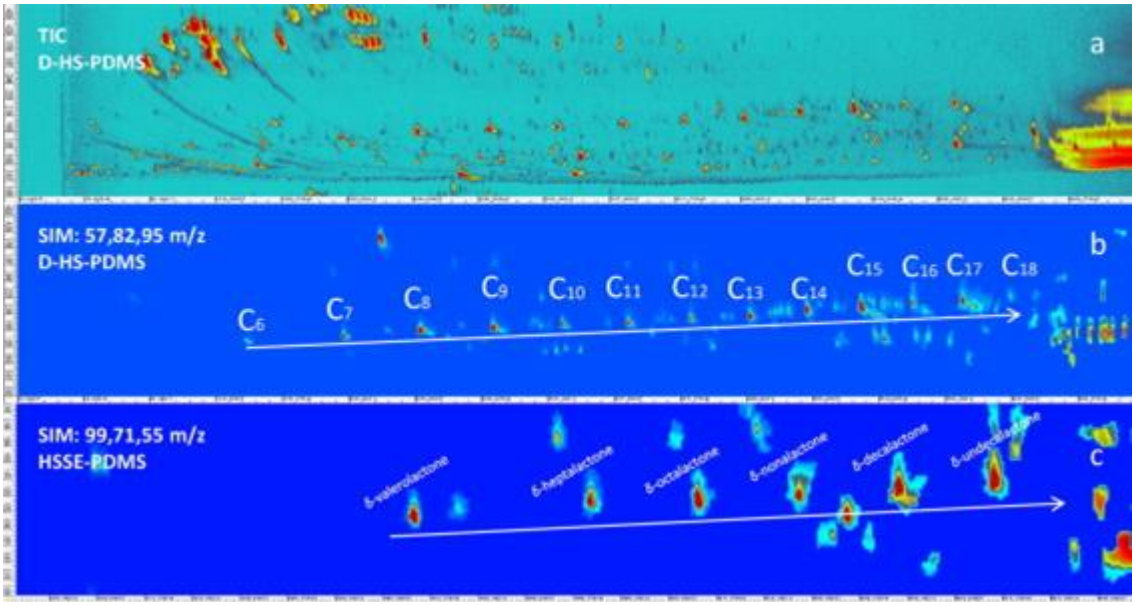

Figure 4.3.4. Pseudocolor 2D images of a whole dried milk sample extracted by dynamic headspace (D-HS) or by headspace sorptive extraction (HSSE). (**4.3.4A**) The TIC trace of the sample headspace (D-HS-PDMS), (**4.3.4B**) the SIM trace of linear saturated aldehydes (57,82,95 *m/z*) and (**4.3.4C**) SIM trace for lactones (55,71,99 *m/z*) recovered by HSSE-PDMS. SIM images were obtained by scripting with CLIC™ Expression (GC Image, LLC Lincoln NE, USA). The column set was by a ^{1}D SolGel-Wax (30 m × 0.25 mm ID, 0.25 mm) and a ^{2}D OV1701 (1 × 0.1 mm ID, 0.1 mm). The system was equipped with a two-stage KT 2004 loop thermal modulator (Zoex Corporation, Houston, TX) cooled with liquid nitrogen controlled by Optimode™ V.2 (SRA Instruments, Cernusco sul Naviglio, MI, Italy). P_M 5s and hot-jet pulse 0.25 s. Modulator capillary: 1 m × 0.10 mm ID of deactivated fused silica capillary. Oven program: 40 °C (1 min) to 170 °C at 2 °C/min and to 250 °C at 20 °C/min (5 min). From Nicolotti *et al.* [46]

MS is a fundamental dimension of MDA for food investigations; besides its confirmatory role, it can improve specificity in the cross-comparative analysis of large sets of samples while also enabling confident untargeted investigations.[13] Worthy of mentioning are MS platforms that operate at variable energy EI[47–49] (viz., Tandem Ionization™), "soft" ionization by laser photo-ionization,[50] or those including chemical ionization (CI).[51] By multiplexing spectra generation with time-switching between two energies,[48] and thanks to a specific design of the ion source, efficient ionization with minimal sensitivity loss is achieved even at low energy (*e.g.*, 10-20 eV)[49,52–56]. Successful applications in food include volatiles and potent odorants in cocoa [49] and in extra-virgin olive oil,[57] and primary metabolites silyl derivatives from hazelnuts.[47]

TOF analyzers are the most popular MS detection platforms in combination with GC×GC; however, just a few of them are capable of achieving high mass resolution, high mass accuracy, in combination with fast scan speeds. Only a few examples of HR-TOF MS or food applications have been demonstrated, but their potential is clear for confident identification and database construction for complex fractions whose components are not fully covered by commercial databases. Tranchida and co-workers dedicated efforts in this direction by studying the unsaponifiable fraction of milk fat from different mammals (cow, buffalo, ewe, and goat),[58,59] and more recently unsaponifiable fraction of vegetable oils.[60] The benefits provided by HR-TOF MS

were mainly related to the sterols class for which both group-type retention ordering and HR-spectra helped in their reliable identification.

On the other hand, more common single quadrupole instruments achieve frequencies and sensitivities suitable to be successfully used for MS detection in combination with GC×GC, although with minor limitations in the mass range extension to afford recording a diagnostic number of data points per peak.

4.3.2.3 Omics fields related to food: open questions to be answered

Commonly, omics refers to the process of "collective characterization and quantification of pools of biological molecules that translate into the structure, function, and dynamics of an organism or organisms".[61] The first omics discipline was genomics, aimed at studying the entire genomes, and was conceptually opposed to genetics, that interrogates individual variants or single genes. To date, a number of different omics have been defined, covering many different fields of genetics, biology, and chemistry. Interestingly, most of the omics related to food attempt to correlate chemical patterns/composition with biological properties of food such as nutritional density and its impact on humans, sensory properties, biological activity, responses to external stimuli such as climate changes on crops, processing practices on semi-finished and finished products, and effects of bacterial metabolism on food composition (e.g., fermented food).

With a comprehensive and reliable definition of the food chemical code, a better understanding on different phenomena is achieved: (a) key food quality markers in primary materials and finished products; (b) bioactive components in food products, extracts, and/or biological fluids; (c) biomarkers in dietary interventions; and (d) human response indicators correlated to metabolic alterations induced by food groups and patterns.

Within the universal framework of foodomics,[62] defined as "a discipline that studies the food and nutrition domains through the application of advanced omics technologies to improve consumer's well-being, health, and confidence", many other omics related to food have been delineated. They focus on more specific interrelations while sharing common investigational strategies and approaches.

Of them, food metabolomics aims at improving the understanding of the relationships between food chemical composition and external variables/stimuli (e.g., crop botanical origin, harvesting area, climate impact, post-harvest treatments, storage conditions, and shelf-life). It represents an emerging area in which GC×GC is demonstrating its central role[63] and boosting effect, providing larger candidates lists with higher intrinsic informative power.

On the other hand, and tightly connected to food metabolomics, the integration of nutritional science with metabolomics delineates the framework of nutrimetabolomics. Within its boundaries, nutrimetabolomics investigates "perturbations of the human metabolome by specific diets, foods, nutrients, micro-organisms or bioactive compounds"[64] while supporting deep mechanistic understanding of the impact of diet and dietary patterns on human health. Its interrelation with the food metabolome, considered as "the part of the human metabolome directly derived from the digestion and biotransformation of foods and their constituents",[65] refers to the presence of both: (a) nutrients that trigger metabolic response in humans, and (b) non-nutrients that might interact with receptors' patterns and metabolic cycles with effects on human metabolism.

When food hedonic profiling is the object of the investigation, sensomics and its well-established workflow, attempts to "molecularize flavor entities of nature" by defining the unique and peculiar odor and taste code capable of evoking food recognition while driving consumers to

pleasant culinary experiences.[66] Sensomics and correlated disciplines such as flavoromics[67] provide solid foundations for objective and rational qualification of food on the basis of chemoreceptive events occurring in our nose and oral cavity.

Food safety and authenticity, although to date have not been developed with an "omics" related framework, have a central role in food analysis. In particular, food safety is a compulsory requirement/pre-requisite for all food products intended for human consumption. For the purpose of this review paper, food safety is included in the discussion due to the potentials of GC×GC in safety assessments and its information power in suggesting supportive actions to improve the overall safety of the food chain. Some interesting examples in the context of food authentication and food identitation are provided in referenced literature.[68–70]

4.3.2.4 Investigation approaches

Well established analytical approaches within metabolomics are those defined as *profiling* and *fingerprinting*.[71,72] They both inform about compositional differences between samples although with a different extent. Profiling is conducted by optimizing the whole analytical process toward known samples' components, along the reductionist track. Detailed information on analytes' retention, mass spectrum, detector response, etc. is collected to allow qualitative and/or quantitative comparisons. If profiling is conducted toward *a priori* defined analytes (*i.e.*, known components),[73] it can be a *targeted profiling*. However, if the analytical process generates distinctive features for all individual components, enabling a truly comprehensive evaluation on sample's constituents, the approach can be defined as *untargeted profiling*.[74,75]

Conceptually different is *fingerprinting*, as it was defined by Fiehn[72] for metabolomics. As a high-throughput process for effective cross-comparative analysis of samples, it does not necessarily achieve the one-feature/one-component resolution of profiling, while it keeps the potential of informing about the comprehensive set of features ideally corresponding to chemical constituents of a sample. In the case of profiling, the level of information (qualitative and quantitative) achievable is function of the analytical dimensions of the system. For example, some spectroscopic techniques (*e.g.*, nuclear magnetic resonance – NMR, near infrared spectroscopy - NIR) can achieve the highest information level, although signals from trace and ultra-trace components could be lost.

Thanks to the rapid and straightforward evolution of MDA platforms, more stable and informative systems are now readily available, offering opportunities to extend and improve the concept of fingerprinting.[27] 2D peak patterns generated by GC×GC, or by LC×LC, can conceptually be compared to human fingertips whose *minutiae features* (*e.g.*, ridge endings and bifurcations) are exploited by fingerprint recognition technology for individuals' identification. In chromatographic fingerprinting, *minutiae features* within 2D peaks patterns, are detected, aligned, and compared across many samples. Depending on system dimensionality, 2D chromatographic features can provide multifaceted information, also referred to as metadata, including retention times in the two chromatographic dimensions (^1t_R, ^2t_R), detector responses, MS spectral signatures and/or additional data related to olfactory quality, parallel detection information, etc. Chromatographic fingerprinting has the potential to give access to the higher levels of information encrypted in MDA data sets. Its fundamental concepts and tools have been recently reviewed.[27]

The next section presents selected research, within food-omics and related fields, that demonstrates the boosting role of GC×GC in developing a more comprehensive understanding of complex phenomena, more accurate and informative results, or simply a new perspective of investigation of complex samples.

4.3.3 Straightforward applications of GC×GC in the food area

4.3.3.1 Food-metabolomics

Food metabolomics by GC×GC can be directed to volatile organic compounds (VOCs), a fraction sometimes referred to as food *volatilome*.[76] By accurate profiling and fingerprinting of VOCs, deeper insights can be achieved on food quality and spoilage, on the effect of climate changes on plant phenotype expression, or to understand the effect of pedoclimatic variations and geographical origin.[35,36,38,77–83] The lipid fraction, by fatty acids methyl esters (FAMEs) profiling and fingerprinting, can be mined for its diagnostic role in terms of quality and origin.[84,85] For food of vegetal origin, plants' primary metabolites (*i.e.*, non-volatile small molecules such as amino acids, sugars, low molecular weight acids, amines, etc.) inform about primary materials' quality [47], harvesting practices, storage conditions and processing impacts,[63,86,87] and abiotic and biotic stress factors effects on crops.[88,89]

Mack *et al.*[90] investigated postharvest ripening impact on kiwifruit. The authors applied untargeted metabolomics to evaluate metabolic changes in kiwifruits across six stages of postharvest ripening plus two non-marketable stages that were connoted by a dramatic water loss. Sugars, sugar-related substances, and organic acids, together with other known and unknown components, showed meaningful variations along post-harvest ripening stages. Of them, sugars followed an incremental trend while organic acids predominantly decreased. In addition, unexpected changes in the concentration of some known and unknown metabolites were observed. One of the most interesting aspects of this study was the multiplatform approach that combined GC×GC-qMS, GC-MS, and NMR on the same samples. Results were thus cross-validated between platforms indicating GC×GC–qMS was the most sensitive platform (of those in the study), allowing the detection of the major part of the kiwifruit primary metabolites. The most concentrated analytes, *e.g.,* main sugars and some organic acids, were better profiled by targeted GC-MS. NMR was least sensitive compared to GC×GC-MS and GC-MS, although for some analytes it showed better reproducibility. The metabolome coverage was quite satisfactory, with about 160 analytes reproducibly measured by GC×GC-qMS and GC-MS. About 45% of all metabolites were identified using an in-house spectral database, while for unknowns it was possible to hypothesize the chemical class, based on the MS fragmentation patterns and relative retention over the 2D space.

By their untargeted multi-platform approach, the authors were able to examine comprehensively the evolution of kiwifruit metabolome during ripening and over-ripening. By the complementary nature of the three platforms, a wider dynamic range of concentrations was covered. The most relevant variations were those of sugars and related substances, which increased during ripening, and of organic acids, which decreased until over-ripening. Other metabolites, such as methanol and xylose, had a maximum in the middle-ripe stages whereas galacturonic acid was detected only in overripe fruits. The study provides a solid foundation for the adoption of GC×GC-qMS as an exploratory analytical platform for untargeted food metabolomics.

When the interest focuses on extreme climate events (drought and temperature effects) and their impact on high-economic value crops, untargeted food metabolomics by GC×GC-TOF MS has demonstrated to be a valid complement to other techniques targeted to non-volatile specialized plant metabolites.[91,92] Stilo *et al.* investigated tea (*Camellia sinensis* L. Kuntze) primary metabolome[89] to better understand plant metabolism adaptation to drought and heavy rains or to daily temperature variations due to elevation. By adapting an extraction/derivatization protocol, well established in biological fluids metabolomics,[93] an extensive coverage of tea primary metabolites was obtained. Of the 780 peak features detected and monitored over more than 40 samples, 74 metabolites were putatively identified based on their relative retention (2D pattern logic and $^{1}D\ I^{T}$) and MS spectral

similarity. Most of the detectable features showed meaningful variations according to season (spring *vs.* summer) and/or to harvest altitude. The known metabolites coverage included 15 amino acids, 21 organic acids, 7 polyalcohols, 14 sugars, and a few specialized metabolites belonging to the phenols class (flavan-3-ols and phenolic acids) and alkaloids (caffeine).

Results revealed that during spring, alanine, aspartic acid, glycine, threonine, valine, phenylalanine, phosphoric acid, xylonic acid, and xylitol were up regulated by the tea plant. When the investigation was extended across samples belonging to two harvest years, fructose, glucose, maltose, and arabinose had consistent up-regulation in spring teas. The impact of altitude, which induces daily temperatures differences, was seen in an up-regulation of alanine, isoleucine, tyrosine, catechin, gallic acid, glycolic acid, malic acid, and ribonic acid in low-elevation teas.

The study, complementing previous work on unprocessed tea focused on volatiles, on specialized metabolites and metals,[88,91,92] extended the existing knowledge on the complex interrelation between tea plants and climate. The authors stated: "…*this work's objective was to identify key primary metabolites and other biomarkers that can contribute to our understanding of the complex relationships and feedback loops that occur between human and natural systems, specifically, strategies to counter climate effects on tea plants.*" The higher level of information achievable by the extended metabolite coverage of GC×GC-TOF MS metabolomics, including known and unknown features consistently tracked across many samples, opens new perspectives in food research. The rich set of metadata (retention times in two dimensions, MS fragmentation patterns, absolute and relative responses, etc.) linked to all untargeted features, can be *ex post* explored for specific discrimination potential with compound identification elucidated by EI-MS spectra, $^{1}D\ I^{T}$, and pattern retention logic.[13]

Fatty acids (FAs) accurate profiling and fingerprinting is another research subject where GC×GC-MS has introduced a new perspective in the investigation. Although lipid fraction chemical dimensionality is not very high, the concurrent presence of several FAs isomers, showing very similar retention behavior, requires high separation power and selectivity for an accurate profiling and deep investigation. GC×GC combines both analytical features (*i.e.*, separation power and differential selectivity on the two chromatographic dimensions) and adds the unique opportunity of generating 2D patterns whose retention logic might help in the identification and disambiguation of components not available as reference standards (*e.g.*, odd carbon number and branched FAs). Tranchida and Mondello dedicated efforts in this application field by profiling and fingerprinting FAs from plants, algae, and animal origin.[84,85] FAs homologue series within the C_4-C_{24} range can be transformed into methyl ester derivatives (fatty acids methyl esters - FAMEs) and then characterized by different column combinations. **Figure 4.3.5A** shows the 2D pattern of human plasma FAMEs analyzed by an apolar × polar column combination. Analytes are aligned along diagonal bands based on carbon number (CN) within C14 and C24 (zones indicated by black arrows) and grouped while forming 2D sub-patterns ordered based on double-bond (DB) number. In the example of **Figure 4.3.5A**, seven DB bands are marked and aligned along the ^{2}D. As additional retention criteria, the relative position of DBs, form diagonal retention bands. The ω-derivatives of $C_{20:5\omega3}$, $C_{20:4\omega3}$, and $C_{20:3\omega3}$ are illustrated in the enlarged area of the chromatogram of **Figure 4.3.5B**.

In the example of **Figure 4.3.5A-B**, analyte identifications were conducted by confirming the presence of 36 FAMEs though reference standard compounds while the identification of the residual 29 components, for which standards were not available, was based on highly ordered nature of the 2D pattern.[94]

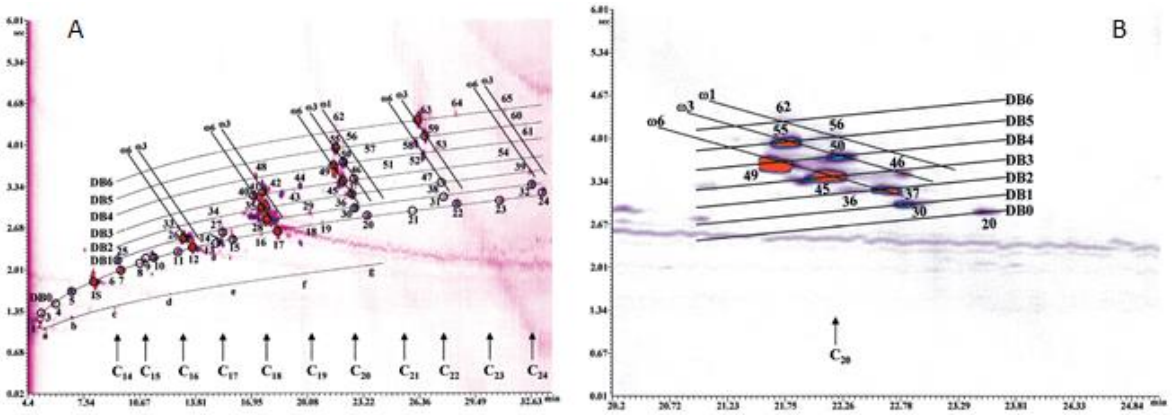

Figure 4.3.5. (**4.3.5A**) 2D pattern of human plasma FAMEs within the C4-C24 range. Analytes are aligned along seven diagonal bands, based on carbon number (CN) within C14 and C24 (zones indicated by lack arrows), and grouped while forming 2D sub-patterns ordered based on double bonds (DB) number. (**4.3.5B**) Enlarged area showing the C20 FAMEs group. The analyses were performed by a column set consisting of a 1D Equity-1 (30 m × 0.25 mm ID, 0.25 mm) and a 2D Supelcowax-10 (0.95 × 0.1 mm ID, 0.1 mm). The GC system was equipped with an LMCS Everest longitudinally modulated cryogenic system (LMCS; Chromatography Concepts, Doncaster, Australia), with a mechanical stepper motor drive for movement of the cryotrap operating with liquid CO2 as cryogenics. Oven program: 180°C to 280°C (10 min), at 3°C/min, PM 6s. Adapted from Tranchida et al.[94]

4.3.3.2 Nutrimetabolomics

When knowledge on food composition is matched with nutritional science, the context of nutrimetabolomics appears clear and highlights a very promising track toward advanced nutritional care, dietary treatments, and personalized nutrition [64].

In this investigation area, the major challenges are related to the great chemical dimensionality of biological fluids (*i.e.*, urine, plasma and saliva) and the difficulty of obtaining highly-ordered patterns for chemically related compounds. Moreover, the large concentration difference between analytes and analyte classes is an additional issue that impacts system contamination, method precision, and long-term reproducibility.

Therefore, most of the efforts in nutrimetabolomics by GC×GC-MS are directed to achieve close-to-optimal separation power in both analytical dimensions accompanied by adequate system/column loadability and appropriate injection to enable full separation/resolution of detectable components within the actual dynamic range of concentration and, consequently, highly reliable quantitation – absolute or relative. Recently, an interdisciplinary group of researchers, belonging to the FoodBAll Consortium (BioNH call of the Joint Programming Initiative "A Healthy Diet for a Healthy Life" (JPI-HDHL),[64] comprehensively revised the nutrimetabolomics strategies while also proposing good practices and standardized protocols to achieve high-quality results.

In this highly challenging context, GC×GC has been successfully applied by Mack *et al.*[95] to reveal coffee intake robust biomarkers. The authors investigated the complex human urine volatilome from a subset of 24 hours urine samples belonging to the cross-sectional study Karlsruhe Metabolomics and Nutrition (KarMeN), performed at the Max Rubner-Institut in Karlsruhe, Germany, between 2011 and 2013. The urine samples were submitted to HS-SPME followed by GC×GC-qMS and combined untargeted/targeted (UT) fingerprinting based on pattern recognition algorithms. Overall, 138 volatiles were reliably detected across all samples with a few of them showing strong-positive correlation to coffee intake. Six most informative volatiles were highlighted: 2-methyl-furan, guaiacol, 2-/3-methyl-butanoic acid, 2-vinylfuran, and 3,4-dimethyl-

2,5-furandione; with the last as the most promising robust biomarker. Of great interest for the nutrimetabolomics context was the evaluation of the most suitable normalization methodology to account for unequal concentration of urine samples that leads to interesting observations applied to the specific context of urine volatilome. The untargeted/targeted investigation workflow is illustrated in **Figure 4.3.6**, including the 2D raw data pre-processing step including batch correction by quality standards, normalization based on different parameters (creatinine content, osmolarity, urine volume, mass spectral total useful signal - MSTUS and probabilistic quotient normalization - PQN), followed by data mining to reveal potential biomarkers through combined evaluation of correlation and statistical meaningfulness. Further insights on the normalization methods and their effect on variation coefficients (CVs%) are provided in the reference.[95]

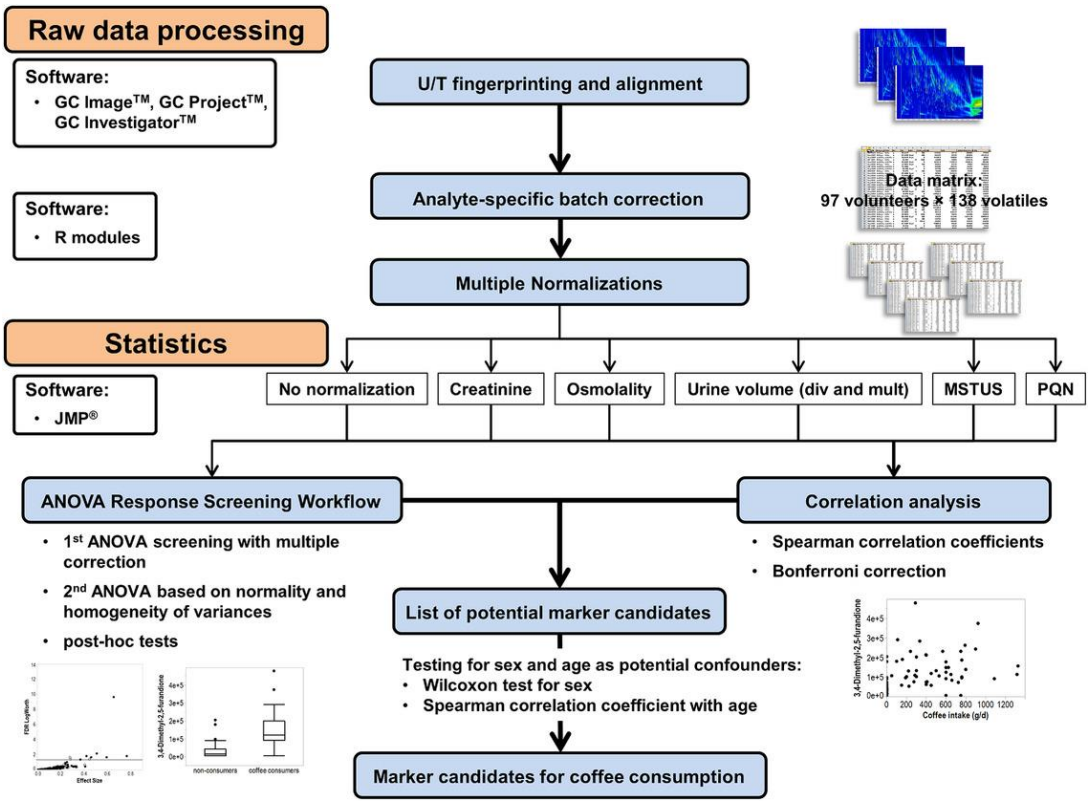

Figure 4.3.6. Untargeted/targeted investigation workflow, including 2D raw data pre-processing step (*i.e.*, batch correction by quality standards and normalization based on different parameters: creatinine content, osmolarity, urine volume, mass spectral total useful signal - MSTUS - and probabilistic quotient normalization - PQN) followed by data mining to reveal potential biomarkers through combined evaluation of correlation and statistical meaningfulness. From Mack et al.[95]

Another interesting study focused on the metabolic transformations of dietary polyphenols by colonic microbiota.[96] The authors compared *in vitro* microbial phenolic metabolite signatures of selected foods and beverages with those from urine of healthy subjects undergoing dietary intervention (8-week clinical trial) with the same foodstuffs. GC×GC-TOF MS was applied in targeted mode for known polyphenolic patterns. Green tea, the most polyphenol-rich food included in the intervention, and its main specialized metabolites (namely phenolic acids) showed patterns of derivatives with modifications on the carbon backbone and extensive oxidation (mainly oxydrilation). Of them 3,4-dihydroxybenzoic acid; 3,4-dihydroxyphenylacetic acid; 3,4-dihydroxyphenylpropionic acid; 3-hydroxybenzoic acid; 3-hydroxyphenylacetic acid; 3-hydroxyphenylpropionic acid; 4-coumaric acid; 4-metylcatechol; 4-hydroxybenzoic acid; benzoic

acid; gallic acid; ferulic acid; sinapic acid; enterodiol; and vanillic acid were those reporting meaningful, positive correlation between urine excretion and *in-vitro* colon model. Moreover, considerably higher amounts of hippuric acid, 3-hydroxybenzoic acid, and ferulic acid were detected in urine of healthy subjects than in the colon model while the hepatic conversion model complemented the residual amounts of phenol-derived metabolites. The study contributed to a better understanding of the bioavailability of diet polyphenols and elucidated their interconnections to urinary metabolites signatures.

Untargeted investigation on urinary metabolites signatures helped in a deeper understanding of the impact of solid vs. liquid fructose formulations in animal models.[56,97] Bressanello *et al.*[56] designed a system combining a 1D column with parallel dual secondary column-dual detection by MS and FID (GC×2GC-MS/FID) to capture changes in metabolic fingerprints from mice urines. Samples were from mice fed with a normal (control diet) or fructose-enriched diet provided in aqueous solution or in solid form. The intervention was followed for 12 weeks and metabolites signatures evolution tracked for single individuals and for sub-population/group.

UT fingerprinting enabled coherent clustering of mice according to dietary manipulation; metabolite fingerprints related to high doses of liquid fructose showed meaningful variations on fructose, glucose, citric, pyruvic, malic, malonic, gluconic, cis-aconitic, succinic and 2-keto glutaric acids, glycine acyl derivatives (N-carboxyglycine, N-butyrylglycine, N-isovaleroylglycine, N-phenylacetylglycine), and hippuric acid concentrations. Post-targeting on additional, informative features highlighted further metabolites not a priori pre-targeted: N-acetyl glucosamine, N-acetyl glutamine, malonyl glycine, methyl malonyl glycine, and glutaric acid, while quantitative GC-MS confirmed the role of several potential biomarkers through accurate quantitative results.

Visual features fingerprinting, based on comparative visualization of urine signatures at different time points of the diet intervention, tracked individual variations while showing the potential role of data processing in supporting personalized strategies. **Figures 4.3.7A** and **4.3.7B** show pseudocolor images of mouse #40 urine at 6 (**4.3.7A**) and 12 weeks (**4.3.7B**) of diet intervention with liquid fructose supplementation. Red graphics locate untargeted/targeted features delineated by the UT fingerprinting workflow. Green graphics instead correspond to column bleeding and/or interferences and were excluded by the computation. **Figure 4.3.7C** shows the comparative visualization rendered with a colorization that emphasized relative differences between patterns. The image is obtained by visual features comparison between the 12 weeks (*reference* pattern) vs. 6 weeks (*analyzed* pattern) urine signatures. Yellow areas over the 2D space in **Figure 4.3.7C** correspond to targeted analytes whose response differences are reported in the list.

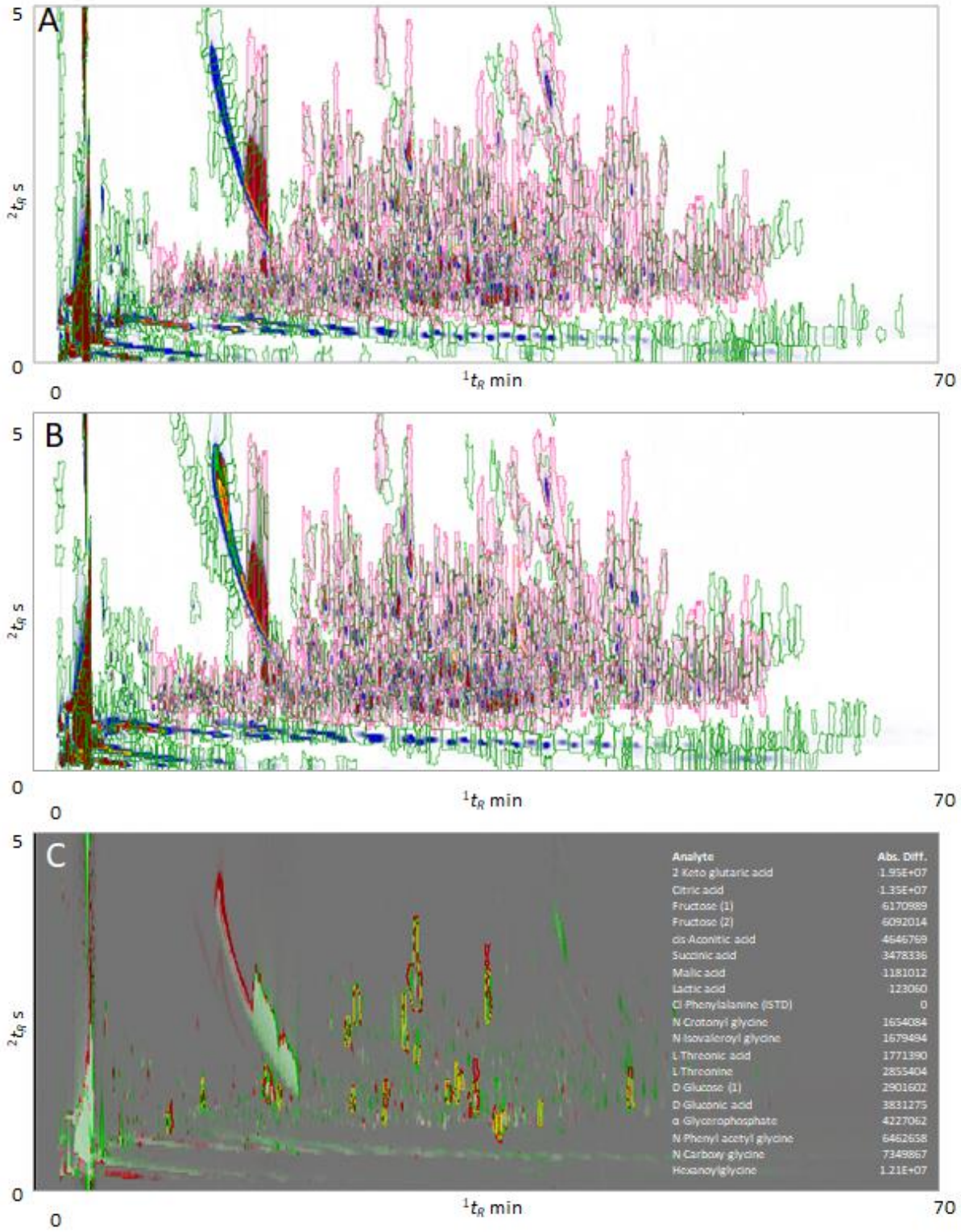

Figure 4.3.7. Visual features fingerprinting (MS signal full scan acquisition) based on comparative visualization of urine samples' signature at different time points of the diet intervention, with red graphics locating untargeted/targeted features, while green graphics correspond to column bleeding and/or interferences and were excluded by the computation. (**4.3.7A**) pseudocolor images of mouse #40 urines at 6 weeks of diet intervention with liquid fructose supplementation – analyzed pattern; (**4.3.7B**) pseudocolor images of mouse #40 urines at 12 weeks of diet intervention with liquid fructose supplementation – reference pattern; (**4.3.7C**) comparative visualization rendered with a colorization that emphasized relative differences between patterns, yellow areas over the

2D space correspond to targeted analytes whose response differences are reported in the list. The column set was by a 1D SE52 (30 m × 0.25 mm ID, 0.25 mm) and a 2D OV1701 (1.4 × 0.1 mm ID, 0.1 mm). The system was equipped with a two-stage KT 2004 loop thermal modulator (Zoex Corporation, Houston, TX) cooled with liquid nitrogen controlled by OptimodeTM V.2 (SRA Instruments, Cernusco sul Naviglio, MI, Italy). PM 5s and hot-jet pulse 0.35 s. Modulator capillary: 0.6 m × 0.10 mm ID of the 2D column. Oven program: 60 °C (1 min) to 300 °C (10 min) at 4 °C/min. From Bressanello et al.[56]

4.3.3.3 Sensomics and Flavoromics

Sensomics aims[98] "*to map the combinatorial code of aroma and taste-active key molecules, which are sensed by human chemosensory receptors and are then integrated by the brain*…." Aroma alone contributes about 80% of the hedonic profile of food [66] and it is modulated by volatile components generally present at trace- and ultra-trace levels (mg/kg or μg/kg) and characterized by low solubility in water media, low polarity, and molecular weights below 300 *amu*. Odorants, interacting with multiple Odorant Receptors (ORs) in the olfactory epithelium, produce a complex pattern of signals (*i.e.,* the Receptor Code).[66,99–101] The integration of these signal patterns by the nervous system produces olfactory perceptions.

In this context, a fundamental contribution to the encryption of food odor code, the so-called *aroma blueprint,*[66,102] is provided by the accurate and comprehensive chemical characterization of all potential ligands and modulator components constituting the Chemical Odor Code. Well-established approaches combine multiple, discrete steps in a systematic workflow. They include volatiles isolation/extraction/concentration; chromatographic separation accompanied by olfactometric screening to identify odor active analytes; accurate quantitation; and validation of aroma contributions by recombination and omission tests.[66,98,103]

GC×GC platforms can implement most of the sensomics tools. Marriott and co-workers made substantial advancements in developing *multi-multidimensional* platforms.[104,105] Instrumental solutions include the possibility of switching between GC×GC and targeted multidimensional gas chromatography system (*i.e.*, switchable GC×GC/targeted MDGC)[106] to achieve full chromatographic resolution in critical regions of the separation space where potent odorants elute. The further, orthogonal dimension of olfactometry was added in a system designed by Chin *et al.*[107] to study aroma-active compounds in coffee and wine. The platform, combining GC-olfactometry (GC-O) and GC×GC with TOF MS/ flame ionization detector (FID)/ flame photometric detector (FPD) in sulfur mode, was used for screening odor-active compounds, separating them by bulk volatiles with GC×GC and identifying with TOF MS and FPD in combination.

Besides the *multi-multidimensional* GC×GC platforms, which have a central role in flavor research, conventional systems are mature enough to become fundamental pillars for sensomics and flavoromics. Nicolotti *et al.*[108] developed a fast, sensomics-based expert system – SEBES, to predict key-aroma compounds of a given food without applying all discrete steps of the conventional workflow [66,98,103]. The authors developed an analytical method for isolation and accurate quantitation of about 100 odorants out of the 226 known key-food odorants (KFOs) listed by Dunkel *et al.*.[66] By a dedicated data-processing step, implementing additional data on odor thresholds, odor activity values (OAV; ratio of concentration to odor threshold) were calculated and sample's combinatorial odor code encrypted. The role of GC×GC was fundamental, as observed by authors, to extend the method dynamic range enabling accurate quantification of analytes present at very different concentrations in real-world samples. Moreover, GC×GC separation efficiency and resolution, enabled a direct injection of crude extracts without further sample preparation steps. Validation of the SEBES system was by comparison with classical sensomics; accuracy was very high with a maximum quantification error of ±20%. As a conclusive remark, the authors stated that "*it was successfully shown that it is possible to characterize key food odorants*

with one single analytical platform and without using the human olfactory system, that is, by artificial intelligence smelling".

Many other applications have emphasized the fundamental role of GC×GC in flavor research. Of interest are those that directly analyze volatiles fractions by headspace (HS) approaches[33,34,38,46,55,57,79,102,109,110]. One of the emerging application areas is extra-virgin olive (EVO) oil volatiles. Within this complex fraction some potent odorants are responsible for positive attributes and of aroma defects.[55,80–82]

Stilo *et al.* [57] explored the complex EVO oil volatilome analysing samples obtained from P*icual* olives harvested at different ripening stages. The combination of HS-SPME to GC×GC-TOF MS featuring hard and soft ionization in tandem (*i.e.*, tandem ionization) enabled the identification of about 130 analytes, over more than 450 detectable volatiles. For most of the targeted analytes with a known role in the definition of the aroma blueprint of EVO oils,[102,111,112] HS linearity was achieved,[113,114] enabling accurate quantitation in a fully automated analysis. Moreover, the adoption of tandem ionization TOF MS extended the dynamic range of the method and the linear range of response, improving method capabilities and informative potential.

4.3.3.4 Food safety

Mineral oil hydrocarbons (MOHs) are complex mixtures of aliphatic hydrocarbon isomers and derivatives (linear -alkanes, branched – isoalkanes and cyclic compounds – cycloalkanes or naphthenes), also referred to as mineral oil saturated hydrocarbons (MOSHs) and of aromatic hydrocarbons (MOAHs). The latter includes aromatic moieties, with one or more benzene rings, and extensive alkyl-substitution. Certain MOAHs fractional components are known for their mutagenicity, toxicity, and tumor promoting effects.

MOSH and MOAH occurrence in food primary materials, semi-finished, and finished products might be related to exogenous contamination sources, whose origin identification would be of great help to implement safety policies for the entire food-chain. MOH fractions quantitative and qualitative assessment represents a great challenge for analytical chemistry; the presence of many foodborne interfering compounds makes mandatory the adoption of suitable confirmatory methods.[115–117]

The state-of-the-art approaches for MOSH/MOAH quantification in food are based on established methodologies that adopt liquid chromatography (LC) off-line or on-line coupled to GC-FID,[118] but the presence of hundreds-to-thousands of isomers producing an unresolved complex mixture (UCM) profile in GC and the lack of a detection dimension with suitable specificity does not guarantee accurate quantitation even by LC-GC.[119,120] Although (HR)MS may be a fundamental dimension to implement, alone it is not sufficient to isolate unique and distinct signals from exogenous MOSHs and MOAHs as required by EU Commission Decision 657/2002.[121]

In this scenario, the potentials of GC×GC were immediately clear. Biedermann and Grob proposed MOHs chromatographic fingerprinting based on medium polar × apolar column combinations capable of discriminating compound classes by pattern retention logic based on relative polarity/volatility. At the same time, through pattern recognition, the fingerprinting survey enables the identification of contamination sources.[122] **Figure 4.3.8** compares chromatographic profiles by reference LC-GC-FID method (left side) corresponding to MOSH and MOAH fractions with the corresponding GC×GC-TOF MS pseudocolor images from Asian rice suspected for MOH contamination.

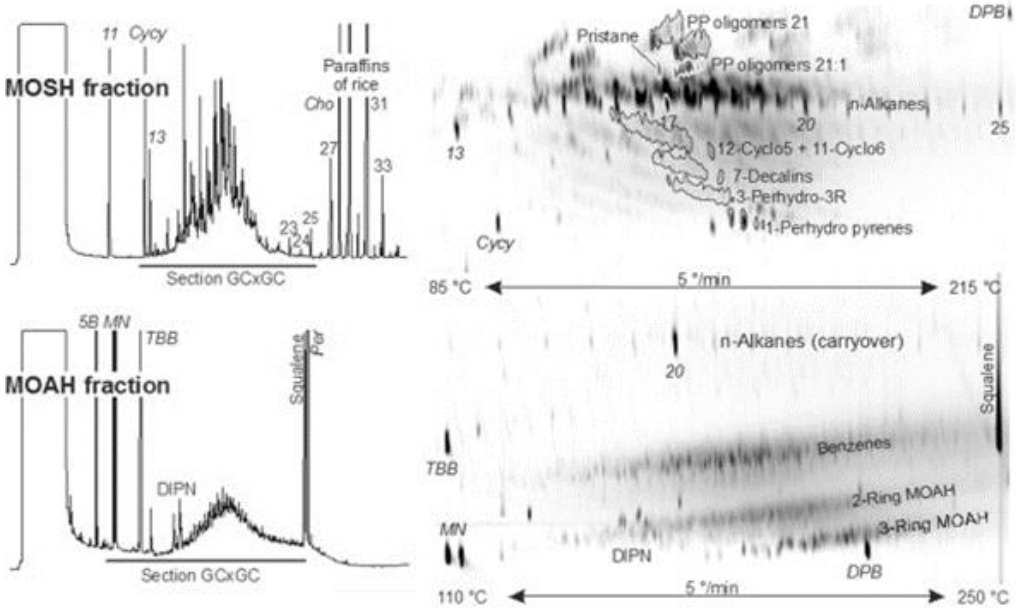

Figure 4.3.8. LC-GC-FID chromatograms (left side) of MOSH and MOAH fractions and the corresponding GC×GC-TOF MS contour plots (right side) from Asian rice samples. Compounds' assignment are (11) for n-undecane, (Cycy) for cyclohexyl cyclohexane, (Cho) for cholestane, (13) for n-tridecane, (20) for n-eicosane (DPB) for diphenoxy benzene, (5B) for n-pentyl benzene, (MN) for 1- and 2-methyl naphthalene, (TBB) for tri tert. butyl benzene, (Per) for perylene. The column set was: 1D column OV-17 (15 m × 0.25 mm × 0.15 μm) and 2D column PS-255 (2.5 m × 0.15 mm × 0.055 μm). Thermal modulation by loop-type modulator. Oven program: 70 °C (3 min) to 310 °C, at 5.0 °C/min; PM 6s. Acronyms are explicated in the text. From Biedermann and Grob.[122]

The fingerprinting potential of GC×GC is illustrated through the ordered pattern of *n*-alkanes from *n*-C13 (internal standard) to *n*-C25 present in uncontaminated rice. Branched paraffines, with longer ^{2}D retention, follow the elution line of *n*-alkanes, while oligomeric polyolefins (POSH) (C21 oligomers) likely originating from polypropylene (PP) packaging, form a distinct cluster. Between *n*-alkanes and POSH, monounsaturated PP oligomers are detected. Slanted elution bands with lower retention in the 2D correspond to naphthenes and cyclic hydrocarbons (*e.g.*, dodecyl cyclopentanes, 12-Cyclo5, and undecyl cyclohexanes, 11-Cyclo6).

The lower plots of **Figure 4.3.8** illustrate the MOAH fraction 1D and 2D profiles. LC-GC produces an unresolved hump while by GC×GC structured elution patterns are more evident. Above all, the group of diisopropyl naphthalenes (DIPN) suggests a possible source of contamination by recycled paperboard.

GC×GC-TOF MS has a crucial role in MOSH-MOAH accurate quantification with opportunities for fingerprinting and identification of contamination sources. The boosting effect of the technique in this area of investigation could deliver truly multidimensional solutions that combine on-line LC pre-fractionation to GC×GC with parallel detection by FID/MS.[123] By this approach, open to untargeted investigations, the concurrent tracing of non-intentionally added substances (NIAS) can be run.[117]

In a more future perspective, GC×GC shows a great potential within the food safety context for monitoring chlorinated paraffins (CPs), synthetic chlorinated *n*-alkanes, widely used in industrial environments as lubricants, paints, coating agents, sealants, etc. According to carbon-chain length, CPs' products are grouped into short chain chlorinated paraffins (SCCPs; C10-C13), medium chain chlorinated paraffins (MCCPs; C14-C17), and long chain chlorinated paraffins (LCCPs; C18-C30).[124] SCCPs and MCCPs recently received attention from the European Food Safety Authority

(EFSA) because of their potential toxicity for humans and persistence in the environment and bioaccumulation.[125]

GC×GC coupled to low-resolution TOF MS was successfully applied to accurate quantitation of SSCPs even in absence of reference calibration standards for individual components. The authors[125] proceeded by establishing a calibration between total response factor (RF) and chlorine (Cl) content based on peak volumes of quantifier ions for 24 congener groups. Results, from CP products and urban air sampling, indicated variations largely among CP materials while seasonal fluctuations on urban air were mostly influenced by average daily temperatures. The highest SCCP concentration was detected in summer compared to winter, with lighter congeners (C_{10} group and Cl_6 group) dominating the gas phase and heavier compounds (C_{13} group and Cl_7 - Cl_8) more distributed in particle phase.

Interesting developments in the direction of accurate assessment accompanied by the identification of the contamination source by informative fingerprinting based on GC×GC- TOF MS could fill the gap evidenced by the EFSA report,[125] which stated: "…only limited data on the occurrence of SCCPs and MCCPs in some fish species were submitted to EFSA. …the Panel noted that dietary exposure will be higher due to the contribution of CPs from other foods." A systematic mapping of the potential environmental sources of contamination crossing the food chain might help and provide solid data to complete risk assessment and make safer our food.

4.3.4 Concluding remarks

Food-omics investigations require a truly comprehensive approach to capture compositional complexity of samples and to establish robust correlations with external stimuli and complex biological phenomena. Comprehensive two-dimensional chromatography has the potential to tackle compositional challenges (chemical dimensionality and dynamic range of concentrations) and provide consistent bases for hypothesis generation. Moreover, dedicated data processing further enriches the toolbox, increasing the investigation potentials while embracing the modern concepts of individual/personalized investigations.

Undoubtedly, GC×GC-MS is a key-analytical platform for food-omics investigations and its widespread could boost research while opening new and concrete perspectives for more comprehensive understanding of food quality, nutritional value, and health benefits.

The industrial request for straightforward solutions to practical problems does not prevent the adoption of MDA platforms and omics concepts even in the industrial research framework. The larger the breadth of an investigation, the better is the understanding of the effects of processing, storage, fermentation, and biotransformation on the overall quality and safety of a food product.

The full potential of GC×GC was probably not clear at its introduction, but its widespread adoption in different areas and the infusion of strategies and concepts from other disciplines, have definitely highlighted its central role of missing technique: "*from a technique that did not exist… to a technique that was missing*".[126]

References

(1) Giddings, J. C. Concepts and Comparisons in Multidimensional Separation. *J. High Resolut. Chromatogr.* **1987**, *10* (5), 319–323. https://doi.org/10.1002/jhrc.1240100517.

(2) Liu, Z.; Phillips, J. B. Comprehensive Two-Dimensional Gas Chromatography Using an on-Column Thermal Modulator Interface. *J. Chromatogr. Sci.* **1991**, *29* (5), 227–231. https://doi.org/10.1093/chromsci/29.6.227.

(3) Adahchour, M.; Beens, J.; Vreuls, R. J. J.; Brinkman, U. A. T. Recent Developments in Comprehensive Two-Dimensional Gas Chromatography (GC × GC). I. Introduction and Instrumental Set-Up. *TrAC - Trends Anal. Chem.* **2006**, *25* (5), 438–454. https://doi.org/10.1016/j.trac.2006.03.002.

(4) Adahchour, M.; Beens, J.; Vreuls, R. J. J.; Brinkman, U. A. T. Recent Developments in Comprehensive Two-Dimensional Gas Chromatography (GC x GC). II. Modulation and Detection. *TrAC - Trends Anal. Chem.* **2006**, *25* (6), 540–553. https://doi.org/10.1016/j.trac.2006.04.004.

(5) Seeley, J. V.; Seeley, S. K. Multidimensional Gas Chromatography: Fundamental Advances and New Applications. *Anal. Chem.* **2013**, *85* (2), 557–578. https://doi.org/10.1021/ac303195u.

(6) Prebihalo, S. E.; Berrier, K. L.; Freye, C. E.; Bahaghighat, H. D.; Moore, N. R.; Pinkerton, D. K.; Synovec, R. E. Multidimensional Gas Chromatography: Advances in Instrumentation, Chemometrics, and Applications. *Anal. Chem.* **2018**, *90* (1), 505–532. https://doi.org/10.1021/acs.analchem.7b04226.

(7) Tranchida, P. Q.; Purcaro, G.; Dugo, P.; Mondello, L.; Purcaro, G. Modulators for Comprehensive Two-Dimensional Gas Chromatography. *TrAC Trends Anal. Chem.* **2011**, *30* (9), 1437–1461. https://doi.org/10.1016/j.trac.2011.06.010.

(8) Fiehn, O. Metabolomics - The Link between Genotypes and Phenotypes. *Plant Mol. Biol.* **2002**, *48* (1–2), 155–171. https://doi.org/10.1023/A:1013713905833.

(9) Pierce, K. M.; Kehimkar, B.; Marney, L. C.; Hoggard, J. C.; Synovec, R. E. Review of Chemometric Analysis Techniques for Comprehensive Two Dimensional Separations Data. *J. Chromatogr. A* **2012**, *1255*, 3–11. https://doi.org/10.1016/j.chroma.2012.05.050.

(10) Pierce, K. M.; Hoggard, J. C.; Mohler, R. E.; Synovec, R. E. Recent Advancements in Comprehensive Two-Dimensional Separations with Chemometrics. *J. Chromatogr. A* **2008**, *1184* (1–2), 341–352. https://doi.org/10.1016/j.chroma.2007.07.059.

(11) Pollo, B. J.; Teixeira, C. A.; Belinato, J. R.; Furlan, M. F.; Cunha, I. C. de M.; Vaz, C. R.; Volpato, G. V.; Augusto, F. Chemometrics, Comprehensive Two-Dimensional Gas Chromatography and "Omics" Sciences: Basic Tools and Recent Applications. *TrAC Trends Anal. Chem.* **2021**, *134*, 116111. https://doi.org/https://doi.org/10.1016/j.trac.2020.116111.

(12) Stilo, F.; Bicchi, C.; Jimenez-carvelo, A. M.; Cuadros-rodriguez, L.; Reichenbach, S. E.; Cordero, C. Trends in Analytical Chemistry Chromatographic Fi Ngerprinting by Comprehensive Two-Dimensional Chromatography : Fundamentals and Tools. *Trends Anal. Chem.* **2021**, *134*, 116133. https://doi.org/10.1016/j.trac.2020.116133.

(13) Stilo, F.; Bicchi, C.; Robbat, A.; Reichenbach, S. E.; Cordero, C. Untargeted Approaches in Food-Omics: The Potential of Comprehensive Two-Dimensional Gas Chromatography/Mass Spectrometry. *TrAC Trends Anal. Chem.* **2021**, *135*, 116162. https://doi.org/10.1016/j.trac.2020.116162.

(14) Bos, T. S.; Knol, W. C.; Molenaar, S. R. A.; Niezen, L. E.; Schoenmakers, P. J.; Somsen, G. W.; Pirok, B. W. J. Recent Applications of Chemometrics in One- and Two-Dimensional Chromatography. *Journal of Separation Science*. 2020, pp 1678–1727. https://doi.org/10.1002/jssc.202000011.

(15) Peterson, R. T. Chemical Biology and the Limits of Reductionism. *Nat Chem Biol* **2008**, *4* (11), 635–638.

(16) Wilson, R. R. Analytical Chemistry of Biological Systems. *Anal. Chem.* **1991**, *63* (7), 379 A. https://doi.org/10.1021/ac00007a600.

(17) Adahchour, M.; Beens, J.; Vreuls, R. J. J.; Brinkman, U. A. T. Recent Developments in Comprehensive Two-Dimensional Gas Chromatography (GC x GC). III. Applications for Petrochemicals and Organohalogens. *TrAC - Trends Anal. Chem.* **2006**, *25* (7), 726–741. https://doi.org/10.1016/j.trac.2006.03.005.

(18) Tranchida, P. Q.; Donato, P.; Cacciola, F.; Beccaria, M.; Dugo, P.; Mondello, L. Potential of Comprehensive Chromatography in Food Analysis. *TrAC - Trends Anal. Chem.* **2013**, *52*, 186–205. https://doi.org/10.1016/j.trac.2013.07.008.

(19) Marriott, P.; Shellie, R. Principles and Applications of Comprehensive Two-Dimensional Gas Chromatography. *TrAC - Trends in Analytical Chemistry*. 2002, pp 573–583. https://doi.org/10.1016/S0165-9936(02)00814-2.

(20) Adahchour, M.; Beens, J.; Vreuls, R. J. J.; Brinkman, U. A. T. Recent Developments in Comprehensive Two-Dimensional Gas Chromatography (GC x GC). IV. Further Applications, Conclusions and Perspectives. *TrAC - Trends Anal. Chem.* **2006**, *25* (8), 821–840. https://doi.org/10.1016/j.trac.2006.03.003.

(21) Cordero, C.; Kiefl, J.; Schieberle, P.; Reichenbach, S. E.; Bicchi, C. Comprehensive Two-Dimensional Gas Chromatography and Food Sensory Properties: Potential and Challenges. *Anal. Bioanal. Chem.* **2015**, *407* (1), 169–191. https://doi.org/10.1007/s00216-014-8248-z.

(22) Cordero, C.; Kiefl, J.; Reichenbach, S. E.; Bicchi, C. Characterization of Odorant Patterns by Comprehensive Two-Dimensional Gas Chromatography: A Challenge in Omic Studies. *TrAC Trends Anal. Chem.* **2019**, *113*, 364–378. https://doi.org/10.1016/j.trac.2018.06.005.

(23) Gruber, B.; Weggler, B. A.; Jaramillo, R.; Murrell, K. A.; Piotrowski, P. K.; Dorman, F. L. Comprehensive Two-Dimensional Gas Chromatography in Forensic Science: A Critical Review of Recent Trends. *TrAC - Trends Anal. Chem.* **2018**, *105*, 292–301. https://doi.org/10.1016/j.trac.2018.05.017.

(24) Giddings, J. C. Sample Dimensionality: A Predictor of Order-Disorder in Component Peak Distribution in Multidimensional Separation. *J. Chromatogr. A* **1995**, *703* (1–2), 3–15. https://doi.org/10.1016/0021-9673(95)00249-M.

(25) Tranchida, P. Q.; Purcaro, G.; Maimone, M.; Mondello, L. Impact of Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry on Food Analysis. *J. Sep. Sci.* **2016**, *39* (1), 149–161. https://doi.org/10.1002/jssc.201500379.

(26) Cordero, C.; Schmarr, H.-G.; Reichenbach, S. E.; Bicchi, C. Current Developments in

Analyzing Food Volatiles by Multidimensional Gas Chromatographic Techniques. *J. Agric. Food Chem.* **2018**, *66* (10), 2226–2236. https://doi.org/10.1021/acs.jafc.6b04997.

(27) Stilo, F.; Bicchi, C.; Jimenez-Carvelo, A. M.; Cuadros-Rodriguez, L.; Reichenbach, S. E.; Cordero, C. Chromatographic Fingerprinting by Comprehensive Two-Dimensional Chromatography: Fundamentals and Tools. *TrAC Trends Anal. Chem.* **2021**, *134*, 116133. https://doi.org/10.1016/j.trac.2020.116133.

(28) Jennerwein, M. K.; Eschner, M.; Gröger, T.; Wilharm, T.; Zimmermann, R. Complete Group-Type Quantification of Petroleum Middle Distillates Based on Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry (GCxGC-TOFMS) and Visual Basic Scripting. *Energy and Fuels* **2014**, *28* (9), 5670–5681. https://doi.org/10.1021/ef501247h.

(29) Vendeuvre, C.; Ruiz-Guerrero, R.; Bertoncini, F.; Duval, L.; Thiébaut, D. Comprehensive Two-Dimensional Gas Chromatography for Detailed Characterisation of Petroleum Products. *Oil Gas Sci. Technol.* **2007**, *62* (1), 43–55. https://doi.org/10.2516/ogst:2007004.

(30) Yan, D. D.; Wong, Y. F.; Tedone, L.; Shellie, R. A.; Marriott, P. J.; Whittock, S. P.; Koutoulis, A. Chemotyping of New Hop (Humulus Lupulus L.) Genotypes Using Comprehensive Two-Dimensional Gas Chromatography with Quadrupole Accurate Mass Time-of-Flight Mass Spectrometry. *J. Chromatogr. A* **2017**. https://doi.org/10.1016/j.chroma.2017.08.020.

(31) Ryan, D.; Shellie, R.; Tranchida, P.; Casilli, A.; Mondello, L.; Marriott, P. Analysis of Roasted Coffee Bean Volatiles by Using Comprehensive Two-Dimensional Gas Chromatography-Time-of-Flight Mass Spectrometry. *J. Chromatogr. A* **2004**, *1054* (1–2), 57–65. https://doi.org/10.1016/j.chroma.2004.08.057.

(32) Mondello, L.; Casilli, A.; Tranchida, P. Q.; Dugo, P.; Costa, R.; Festa, S.; Dugo, G. Comprehensive Multidimensional GC for the Characterization of Roasted Coffee Beans. *J. Sep. Sci.* **2004**, *27* (5–6), 442–450. https://doi.org/10.1002/jssc.200301662.

(33) Magagna, F.; Guglielmetti, A.; Liberto, E.; Reichenbach, S. E.; Allegrucci, E.; Gobino, G.; Bicchi, C.; Cordero, C. Comprehensive Chemical Fingerprinting of High-Quality Cocoa at Early Stages of Processing: Effectiveness of Combined Untargeted and Targeted Approaches for Classification and Discrimination. *J. Agric. Food Chem.* **2017**, *65* (30), 6329–6341. https://doi.org/10.1021/acs.jafc.7b02167.

(34) Magagna, F.; Liberto, E.; Reichenbach, S. E.; Tao, Q.; Carretta, A.; Cobelli, L.; Giardina, M.; Bicchi, C.; Cordero, C. Advanced Fingerprinting of High-Quality Cocoa: Challenges in Transferring Methods from Thermal to Differential-Flow Modulated Comprehensive Two Dimensional Gas Chromatography. *J. Chromatogr. A* **2018**, *1536*, 122–136. https://doi.org/10.1016/j.chroma.2017.07.014.

(35) Humston, E. M.; Zhang, Y.; Brabeck, G. F.; McShea, A.; Synovec, R. E. Development of a GCxGC-TOFMS Method Using SPME to Determine Volatile Compounds in Cacao Beans. *J. Sep. Sci.* **2009**, *32* (13), 2289–2295. https://doi.org/10.1002/jssc.200900143.

(36) Humston, E. M.; Knowles, J. D.; McShea, A.; Synovec, R. E. Quantitative Assessment of Moisture Damage for Cacao Bean Quality Using Two-Dimensional Gas Chromatography Combined with Time-of-Flight Mass Spectrometry and Chemometrics. *J. Chromatogr. A* **2010**, *1217* (12), 1963–1970. https://doi.org/10.1016/j.chroma.2010.01.069.

(37) Kiefl, J.; Pollner, G.; Schieberle, P. Sensomics Analysis of Key Hazelnut Odorants (Corylus Avellana L. 'Tonda Gentile') Using Comprehensive Two-Dimensional Gas Chromatography

in Combination with Time-of-Flight Mass Spectrometry (GC×GC-TOF-MS). *J. Agric. Food Chem.* **2013**, *61* (22), 5226–5235. https://doi.org/10.1021/jf400807w.

(38) Cordero, C.; Liberto, E.; Bicchi, C.; Rubiolo, P.; Schieberle, P.; Reichenbach, S. E.; Tao, Q. Profiling Food Volatiles by Comprehensive Two-Dimensional Ga Schromatography Coupled with Mass Spectrometry: Advanced Fingerprinting Approaches for Comparative Analysis of the Volatile Fraction of Roasted Hazelnuts (Corylus Avellana L.) from Different Ori. *J. Chromatogr. A* **2010**, *1217* (37), 5848–5858. https://doi.org/10.1016/j.chroma.2010.07.006.

(39) Cordero, C.; Bicchi, C.; Rubiolo, P. Group-Type and Fingerprint Analysis of Roasted Food Matrices (Coffee and Hazelnut Samples) by Comprehensive Two-Dimensional Gas Chromatography. *J. Agric. Food Chem.* **2008**, *56* (17), 7655–7666. https://doi.org/10.1021/jf801001z.

(40) Scientific, N.; Databases, T. NIST/EPA/NIH Mass Spectral Library with Search Program: (Data Version: NIST 08, Software Version 2.0f). National Institute of Standards and Technology (NIST): Gaithersburg MD 2005.

(41) McLafferty, F. W. Wiley Registry of Mass Spectral Data, 12th Edition. Wiley 2020.

(42) Liberto, E.; Cagliero, C.; Sgorbini, B.; Bicchi, C.; Sciarrone, D.; Zellner, B. D. A.; Mondello, L.; Rubiolo, P. Enantiomer Identification in the Flavour and Fragrance Fields by "Interactive" Combination of Linear Retention Indices from Enantioselective Gas Chromatography and Mass Spectrometry. *J. Chromatogr. A* **2008**, *1195* (1–2), 117–126. https://doi.org/10.1016/j.chroma.2008.04.045.

(43) d'Acampora Zellner, B.; Bicchi, C.; Dugo, P.; Rubiolo, P.; Dugo, G.; Mondello, L. Linear Retention Indices in Gas Chromatographic Analysis: A Review. *Flavour and Fragrance Journal*. September 2008, pp 297–314. https://doi.org/10.1002/ffj.1887.

(44) Reichenbach, S. E.; Tao, Q.; Cordero, C.; Bicchi, C. A Data-Challenge Case Study of Analyte Detection and Identification with Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry (GC×GC-MS). *Separations* **2019**, *6* (3), 38. https://doi.org/10.3390/separations6030038.

(45) Magagna, F.; Cordero, C.; Cagliero, C.; Liberto, E.; Rubiolo, P.; Sgorbini, B.; Bicchi, C. Black Tea Volatiles Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography – Mass Spectrometry Combined with High Concentration Capacity Sample Preparation Techniques: Toward a Fully Automated Sensomic Assessment. *Food Chem.* **2017**, *225*, 276–287. https://doi.org/10.1016/j.foodchem.2017.01.003.

(46) Nicolotti, L.; Cordero, C.; Cagliero, C.; Liberto, E.; Sgorbini, B.; Rubiolo, P.; Bicchi, C. Quantitative Fingerprinting by Headspace-Two-Dimensional Comprehensive Gas Chromatography-Mass Spectrometry of Solid Matrices: Some Challenging Aspects of the Exhaustive Assessment of Food Volatiles. *Anal. Chim. Acta* **2013**, *798*, 115–125. https://doi.org/10.1016/j.aca.2013.08.052.

(47) Rosso, M. C.; Mazzucotelli, M.; Bicchi, C.; Charron, M.; Manini, F.; Menta, R.; Fontana, M.; Reichenbach, S. E.; Cordero, C. Adding Extra-Dimensions to Hazelnuts Primary Metabolome Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography Combined with Time-of-Flight Mass Spectrometry Featuring Tandem Ionization: Insights on the Aroma Potential. *J. Chromatogr. A* **2020**, *1614*. https://doi.org/10.1016/j.chroma.2019.460739.

(48) Markes International. Select-EV: The next Generation of Ion Source Technology. *Technical Note*. 2016.

(49) Cordero, C.; Guglielmetti, A.; Bicchi, C.; Liberto, E.; Baroux, L.; Merle, P.; Tao, Q.; Reichenbach, S. E. Comprehensive Two-Dimensional Gas Chromatography Coupled with Time of Flight Mass Spectrometry Featuring Tandem Ionization: Challenges and Opportunities for Accurate Fingerprinting Studies. *J. Chromatogr. A* **2019**, *1597*, 132–141. https://doi.org/10.1016/j.chroma.2019.03.025.

(50) Welthagen, W.; Mitschke, S.; Muhlberger, F.; Zimmermann, R. One-Dimensional and Comprehensive Two-Dimensional Gas Chromatography Coupled to Soft Photo Ionization Time-of-Flight Mass Spectrometry : A Two- and Three-Dimensional Separation Approach. *J. Chromatogr. A* **2007**, *1150*, 54–61. https://doi.org/10.1016/j.chroma.2007.03.033.

(51) Warren, C. R. Use of Chemical Ionization for GC-MS Metabolite Profiling. *Metabolomics* **2013**, *9* (SUPPL.1), 110–120. https://doi.org/10.1007/s11306-011-0346-8.

(52) Alam, M. S.; Stark, C.; Harrison, R. M. Using Variable Ionization Energy Time-of-Flight Mass Spectrometry with Comprehensive GC??GC to Identify Isomeric Species. *Anal. Chem.* **2016**, *88* (8), 4211–4220. https://doi.org/10.1021/acs.analchem.5b03122.

(53) Dubois, L. M.; Perrault, K. A.; Stefanuto, P. H.; Koschinski, S.; Edwards, M.; McGregor, L.; Focant, J. F. Thermal Desorption Comprehensive Two-Dimensional Gas Chromatography Coupled to Variable-Energy Electron Ionization Time-of-Flight Mass Spectrometry for Monitoring Subtle Changes in Volatile Organic Compound Profiles of Human Blood. *J. Chromatogr. A* **2017**, *1501*, 117–127. https://doi.org/10.1016/j.chroma.2017.04.026.

(54) Freye, C. E.; Moore, N. R.; Synovec, R. E. Enhancing the Chemical Selectivity in Discovery-Based Analysis with Tandem Ionization Time-of-Flight Mass Spectrometry Detection for Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2018**, *1537*, 99–108. https://doi.org/10.1016/j.chroma.2018.01.008.

(55) Magagna, F.; Valverde-Som, L.; Ruíz-Samblás, C.; Cuadros-Rodríguez, L.; Reichenbach, S. E.; Bicchi, C.; Cordero, C. Combined Untargeted and Targeted Fingerprinting with Comprehensive Two-Dimensional Chromatography for Volatiles and Ripening Indicators in Olive Oil. *Anal. Chim. Acta* **2016**, *936*, 245–258. https://doi.org/10.1016/j.aca.2016.07.005.

(56) Bressanello, D.; Liberto, E.; Collino, M.; Chiazza, F.; Mastrocola, R.; Reichenbach, S. E.; Bicchi, C.; Cordero, C. Combined Untargeted and Targeted Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography: Revealing Fructose-Induced Changes in Mice Urinary Metabolic Signatures. *Anal. Bioanal. Chem.* **2018**, *410* (11), 2723–2737. https://doi.org/10.1007/s00216-018-0950-9.

(57) Stilo, F.; Liberto, E.; Reichenbach, S. E.; Tao, Q.; Bicchi, C.; Cordero, C. Exploring the Extra-Virgin Olive Oil Volatilome by Adding Extra Dimensions to Comprehensive Two-Dimensional Gas Chromatography and Time of Flight Mass Spectrometry Featuring Tandem Ionization: Validation of Ripening Markers in Headspace Linearity Conditio. *J. AOAC Int.* **2020**. https://doi.org/10.1093/jaoacint/qsaa095.

(58) Tranchida, P. Q.; Salivo, S.; Bonaccorsi, I.; Rotondo, A.; Dugo, P.; Mondello, L. Analysis of the Unsaponifiable Fraction of Lipids Belonging to Various Milk-Types by Using Comprehensive Two-Dimensional Gas Chromatography with Dual Mass Spectrometry/Flame Ionization Detection and with the Support of High Resolution Time-

of-Flight Mass . *J. Chromatogr. A* **2013**, *1313*, 194–201. https://doi.org/10.1016/j.chroma.2013.07.089.

(59) Tranchida, P. Q.; Salivo, S.; Franchina, F. A.; Bonaccorsi, I.; Dugo, P.; Mondello, L. Qualitative and Quantitative Analysis of the Unsaponifiable Fraction of Vegetable Oils by Using Comprehensive 2D GC with Dual MS/FID Detection. *Anal. Bioanal. Chem.* **2013**, *405* (13), 4655–4663. https://doi.org/10.1007/s00216-013-6704-9.

(60) Aloisi, I.; Zoccali, M.; Dugo, P.; Tranchida, P. Q.; Mondello, L. Fingerprinting of the Unsaponifiable Fraction of Vegetable Oils by Using Cryogenically-Modulated Comprehensive Two-Dimensional Gas Chromatography-High Resolution Time-of-Flight Mass Spectrometry. *Food Anal. Methods* **2020**, *13* (7), 1523–1529. https://doi.org/10.1007/s12161-020-01773-9.

(61) Comprehensive Analytical Chemistry. In *Fundamentals of Advanced Omics Technologies: From Genes to Metabolites*; Simó, C., Cifuentes, A., García-Cañas, V. B. T.-C. A. C., Eds.; Elsevier, 2014; Vol. 63, p ii. https://doi.org/https://doi.org/10.1016/B978-0-444-62651-6.09993-1.

(62) Capozzi, F.; Bordoni, A. Foodomics: A New Comprehensive Approach to Food and Nutrition. *Genes Nutr.* **2013**, *8* (1), 1–4. https://doi.org/10.1007/s12263-012-0310-x.

(63) Weinert, C. H.; Egert, B.; Kulling, S. E. On the Applicability of Comprehensive Two-Dimensional Gas Chromatography Combined with a Fast-Scanning Quadrupole Mass Spectrometer for Untargeted Large-Scale Metabolomics. *J. Chromatogr. A* **2015**, *1405*, 156–167. https://doi.org/10.1016/j.chroma.2015.04.011.

(64) Ulaszewska, M. M.; Weinert, C. H.; Trimigno, A.; Portmann, R.; Andres Lacueva, C.; Badertscher, R.; Brennan, L.; Brunius, C.; Bub, A.; Capozzi, F.; Cialiè Rosso, M.; Cordero, C. E.; Daniel, H.; Durand, S.; Egert, B.; Ferrario, P. G.; Feskens, E. J. M.; Franceschi, P.; Garcia-Aloy, M.; Giacomoni, F.; Giesbertz, P.; González-Domínguez, R.; Hanhineva, K.; Hemeryck, L. Y.; Kopka, J.; Kulling, S. E.; Llorach, R.; Manach, C.; Mattivi, F.; Migné, C.; Münger, L. H.; Ott, B.; Picone, G.; Pimentel, G.; Pujos-Guillot, E.; Riccadonna, S.; Rist, M. J.; Rombouts, C.; Rubert, J.; Skurk, T.; Sri Harsha, P. S. C.; Van Meulebroek, L.; Vanhaecke, L.; Vázquez-Fresno, R.; Wishart, D.; Vergères, G. Nutrimetabolomics: An Integrative Action for Metabolomic Analyses in Human Nutritional Studies. *Mol. Nutr. Food Res.* **2019**, *63* (1), 1–38. https://doi.org/10.1002/mnfr.201800384.

(65) Scalbert, A.; Brennan, L.; Manach, C.; Andres-Lacueva, C.; Dragsted, L. O.; Draper, J.; Rappaport, S. M.; Van Der Hooft, J. J. J.; Wishart, D. S. The Food Metabolome: A Window over Dietary Exposure. *Am. J. Clin. Nutr.* **2014**, *99* (6), 1286–1308. https://doi.org/10.3945/ajcn.113.076133.

(66) Dunkel, A.; Steinhaus, M.; Kotthoff, M.; Nowak, B.; Krautwurst, D.; Schieberle, P.; Hofmann, T. Nature's Chemical Signatures in Human Olfaction: A Foodborne Perspective for Future Biotechnology. *Angew. Chemie - Int. Ed.* **2014**, *53* (28), 7124–7143. https://doi.org/10.1002/anie.201309508.

(67) Charve, J.; Chen, C.; Hegeman, A. D.; Reineccius, G. A. Evaluation of Instrumental Methods for the Untargeted Analysis of Chemical Stimuli of Orange Juice Flavour. *Flavour Fragr. J.* **2011**, *26* (6), 429–440. https://doi.org/10.1002/ffj.2078.

(68) Purcaro, G.; Barp, L.; Beccaria, M.; Conte, L. S. Fingerprinting of Vegetable Oil Minor Components by Multidimensional Comprehensive Gas Chromatography with Dual Detection. *Anal. Bioanal. Chem.* **2015**, *407* (1), 309–319. https://doi.org/10.1007/s00216-

014-8140-x.

(69) Purcaro, G.; Barp, L.; Beccaria, M.; Conte, L. S. Characterisation of Minor Components in Vegetable Oil by Comprehensive Gas Chromatography with Dual Detection. *Food Chem.* **2016**, *212*, 730–738. https://doi.org/10.1016/j.foodchem.2016.06.048.

(70) Cuadros-Rodríguez, L.; Ruiz-Samblás, C.; Valverde-Som, L.; Pérez-Castaño, E.; González-Casado, A. Chromatographic Fingerprinting: An Innovative Approach for Food "identitation" and Food Authentication - A Tutorial. *Anal. Chim. Acta* **2016**, *909*, 9–23. https://doi.org/10.1016/j.aca.2015.12.042.

(71) Horning, E. C.; Horning, M. G. Human Metabolic Profiles Obtained by Gc and Gc/Ms. *J. Chromatogr. Sci.* **1971**, *9* (3), 129–140. https://doi.org/10.1093/chromsci/9.3.129.

(72) Fiehn, O. Combining Genomics, Metabolome Analysis, and Biochemical Modelling to Understand Metabolic Networks. *Comp. Funct. Genomics* **2001**, *2* (3), 155–168. https://doi.org/10.1002/cfg.82.

(73) Beale, D. J.; Pinu, F. R.; Kouremenos, K. A.; Poojary, M. M.; Narayana, V. K.; Boughton, B. A.; Kanojia, K.; Dayalan, S.; Jones, O. A. H.; Dias, D. A. *Review of Recent Developments in GC–MS Approaches to Metabolomics-Based Research*; Springer US, 2018; Vol. 14. https://doi.org/10.1007/s11306-018-1449-2.

(74) Wolfender, J. L.; Marti, G.; Thomas, A.; Bertrand, S. Current Approaches and Challenges for the Metabolite Profiling of Complex Natural Extracts. *J. Chromatogr. A* **2015**, *1382*, 136–164. https://doi.org/10.1016/j.chroma.2014.10.091.

(75) Alonso, A.; Marsal, S.; Julià, A. Analytical Methods in Untargeted Metabolomics: State of the Art in 2015 . *Frontiers in Bioengineering and Biotechnology* . 2015, p 23.

(76) Bicchi, C.; Maffei, M. *The Plant Volatilome: Methods of Analysis. in High-Throughput Phenotyping in Plants. Methods in Molecular Biology (Methods and Protocols)*; Normanly, J., Ed.; Humana Press: Totowa, NJ, 2012.

(77) Chin, S.-T.; Eyres, G. T.; Marriott, P. J. Application of Integrated Comprehensive/Multidimensional Gas Chromatography with Mass Spectrometry and Olfactometry for Aroma Analysis in Wine and Coffee. *Food Chem.* **2015**, *185*, 355–361. https://doi.org/10.1016/j.foodchem.2015.04.003.

(78) Wong, Y. F.; Perlmutter, P.; Marriott, P. J. Untargeted Metabolic Profiling of Eucalyptus Spp. Leaf Oils Using Comprehensive Two-Dimensional Gas Chromatography with High Resolution Mass Spectrometry: Expanding the Metabolic Coverage. *Metabolomics* **2017**, *13* (5), 1–17. https://doi.org/10.1007/s11306-017-1173-3.

(79) Stilo, F.; Liberto, E.; Reichenbach, S. E.; Tao, Q.; Bicchi, C.; Cordero, C. Untargeted and Targeted Fingerprinting of Extra Virgin Olive Oil Volatiles by Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry: Challenges in Long-Term Studies. *J. Agric. Food Chem.* **2019**, *67* (18), 5289–5302. https://doi.org/10.1021/acs.jafc.9b01661.

(80) Ros, A. Da; Masuero, D.; Riccadonna, S.; Bubola, K. B.; Mulinacci, N.; Mattivi, F.; Lukić, I.; Vrhovsek, U. Complementary Untargeted and Targeted Metabolomics for Differentiation of Extra Virgin Olive Oils of Different Origin of Purchase Based on Volatile and Phenolic Composition and Sensory Quality. *Molecules* **2019**, *24* (16), 1–17. https://doi.org/10.3390/molecules24162896.

(81) Vaz-Freire, L. T.; da Silva, M. D. R. G.; Freitas, A. M. C. Comprehensive Two-Dimensional Gas Chromatography for Fingerprint Pattern Recognition in Olive Oils Produced by Two Different Techniques in Portuguese Olive Varieties Galega Vulgar, Cobrançosa e Carrasquenha. *Anal. Chim. Acta* **2009**, *633* (2), 263–270. https://doi.org/10.1016/j.aca.2008.11.057.

(82) Cajka, T.; Riddellova, K.; Klimankova, E.; Cerna, M.; Pudil, F.; Hajslova, J. Traceability of Olive Oil Based on Volatiles Pattern and Multivariate Analysis. *Food Chem.* **2010**, *121* (1), 282–289. https://doi.org/10.1016/j.foodchem.2009.12.011.

(83) Stanimirova, I.; Üstün, B.; Cajka, T.; Riddelova, K.; Hajslova, J.; Buydens, L. M. C.; Walczak, B. Tracing the Geographical Origin of Honeys Based on Volatile Compounds Profiles Assessment Using Pattern Recognition Techniques. *Food Chem.* **2010**, *118* (1), 171–176. https://doi.org/10.1016/j.foodchem.2009.04.079.

(84) Tranchida, P. Q.; Giannino, A.; Mondello, M.; Sciarrone, D.; Dugo, P.; Dugo, G.; Mondello, L. Elucidation of Fatty Acid Profiles in Vegetable Oils Exploiting Group-Type Patterning and Enhanced Sensitivity of Comprehensive Two-Dimensional Gas Chromatography. *J. Sep. Sci.* **2008**, *31* (10), 1797–1802. https://doi.org/10.1002/jssc.200800002.

(85) Tranchida, P. Q.; Donato, P.; Dugo, G.; Mondello, L.; Dugo, P. Comprehensive Chromatographic Methods for the Analysis of Lipids. *TrAC - Trends Anal. Chem.* **2007**, *26* (3), 191–205. https://doi.org/10.1016/j.trac.2007.01.006.

(86) Romo-Pérez, M. L.; Weinert, C. H.; Häußler, M.; Egert, B.; Frechen, M. A.; Trierweiler, B.; Kulling, S. E.; Zörb, C. Metabolite Profiling of Onion Landraces and the Cold Storage Effect. *Plant Physiol. Biochem.* **2020**, *146*, 428–437. https://doi.org/10.1016/j.plaphy.2019.11.007.

(87) Schmarr, H.-G.; Bernhardt, J. Profiling Analysis of Volatile Compounds from Fruits Using Comprehensive Two-Dimensional Gas Chromatography and Image Processing Techniques. *J. Chromatogr. A* **2010**, *1217* (4), 565–574. https://doi.org/10.1016/j.chroma.2009.11.063.

(88) Morimoto, J.; Rosso, M. C.; Kfoury, N.; Bicchi, C.; Cordero, C.; Robbat, A.; Cialiè Rosso, M.; Kfoury, N.; Bicchi, C.; Cordero, C.; Robbat, A. Untargeted/Targeted 2D Gas Chromatography/Mass Spectrometry Detection of the Total Volatile Tea Metabolome. *Molecules* **2019**, *24* (20), 1–14. https://doi.org/10.3390/molecules24203757.

(89) Stilo, F.; Tredici, G.; Bicchi, C.; Robbat, A.; Morimoto, J.; Cordero, C. Climate and Processing Effects on Tea (Camellia Sinensis L. Kuntze) Metabolome: Accurate Profiling and Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography/Time-of-Flight Mass Spectrometry. *Molecules* **2020**, *25* (10), 2447. https://doi.org/10.3390/molecules25102447.

(90) Mack, C.; Wefers, D.; Schuster, P.; Weinert, C. H.; Egert, B.; Bliedung, S.; Trierweiler, B.; Muhle-Goll, C.; Bunzel, M.; Luy, B.; Kulling, S. E. Untargeted Multi-Platform Analysis of the Metabolome and the Non-Starch Polysaccharides of Kiwifruit during Postharvest Ripening. *Postharvest Biol. Technol.* **2017**, *125*, 65–76. https://doi.org/10.1016/j.postharvbio.2016.10.011.

(91) Kfoury, N.; Morimoto, J.; Kern, A.; Scott, E. R.; Orians, C. M.; Ahmed, S.; Griffin, T.; Cash, S. B.; Stepp, J. R.; Xue, D.; Long, C.; Robbat, A. Striking Changes in Tea Metabolites Due to Elevational Effects. *Food Chem.* **2018**, *264*, 334–341. https://doi.org/10.1016/j.foodchem.2018.05.040.

(92) Kfoury, N.; Scott, E. R.; Orians, C. M.; Ahmed, S.; Cash, S. B.; Griffin, T.; Matyas, C.; Stepp, J. R.; Han, W.; Xue, D.; Long, C.; Robbat, A. Plant-Climate Interaction Effects: Changes in the Relative Distribution and Concentration of the Volatile Tea Leaf Metabolome in 2014–2016. *Front. Plant Sci.* **2019**, *10* (November), 1–10. https://doi.org/10.3389/fpls.2019.01518.

(93) Chan, E. C. Y.; Pasikanti, K. K.; Nicholson, J. K. Global Urinary Metabolic Profiling Procedures Using Gas Chromatography-Mass Spectrometry. *Nat. Protoc.* **2011**, *6* (10), 1483–1499. https://doi.org/10.1038/nprot.2011.375.

(94) Quinto Tranchida, P.; Costa, R.; Donato, P.; Sciarrone, D.; Ragonese, C.; Dugo, P.; Dugo, G.; Mondello, L. Acquisition of Deeper Knowledge on the Human Plasma Fatty Acid Profile Exploiting Comprehensive 2-D GC. *J. Sep. Sci.* **2008**, *31* (19), 3347–3351. https://doi.org/10.1002/jssc.200800289.

(95) Mack, C. I.; Egert, B.; Liberto, E.; Weinert, C. H.; Bub, A.; Hoffmann, I.; Bicchi, C.; Kulling, S. E.; Cordero, C. Robust Markers of Coffee Consumption Identified Among the Volatile Organic Compounds in Human Urine. *Molecular Nutrition and Food Research*. 2019. https://doi.org/10.1002/mnfr.201801060.

(96) Vetrani, C.; Rivellese, A. A.; Annuzzi, G.; Adiels, M.; Borén, J.; Mattila, I.; Orešič, M.; Aura, A.-M. Metabolic Transformations of Dietary Polyphenols: Comparison between in Vitro Colonic and Hepatic Models and in Vivo Urinary Metabolites. *J. Nutr. Biochem.* **2016**, *33*, 111–118. https://doi.org/https://doi.org/10.1016/j.jnutbio.2016.03.007.

(97) Mastrocola, R.; AS, C.; Chiazza, F.; Ferrocino, I.; Bitonto, V.; Nigro, D.; Querio, G.; Rantsiou, K.; JC, C.; Aragno, M.; Liberto, E.; Masini, E.; Cordero, C.; Cocolin, L.; M., C. Toxicological Impact of High Fructose Intake on Gut Microbiota and Liver/Intestine Integrity: Any Differences between Solid and Liquid Formulations? *Diabetologia* **2017**, *60*, S245–S245.

(98) Schieberle, P.; Hofmann, T. Mapping the Combinatorial Code of Food Flavors by Means of Molecular Sensory Science Approach. In *Food Flavors: Chemical, Sensory and Technological Properties*; 2011; pp 413–438. https://doi.org/10.1201/b11187-22.

(99) Firestein, S. How the Olfactory System Makes Sense of Scent. *Nature* **2001**, *413* (6852), 211–218.

(100) Breer, H.; Fleischer, J.; Strotmann, J. The Sense of Smell: Multiple Olfactory Subsystems. *Cell. Mol. Life Sci.* **2006**, *63* (13), 1465–1475. https://doi.org/10.1007/s00018-006-6108-5.

(101) Audouze, K.; Tromelin, A.; Le Bon, A. M.; Belloir, C.; Petersen, R. K.; Kristiansen, K.; Brunak, S.; Taboureau, O. Identification of Odorant-Receptor Interactions by Global Mapping of the Human Odorome. *PLoS One* **2014**, *9* (4). https://doi.org/10.1371/journal.pone.0093037.

(102) Purcaro, G.; Cordero, C.; Liberto, E.; Bicchi, C.; Conte, L. S. Toward a Definition of Blueprint of Virgin Olive Oil by Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2014**, *1334*, 101–111. https://doi.org/10.1016/j.chroma.2014.01.067.

(103) Grosch, W. Evaluation of the Key Odorants of Foods by Dilution Experiments, Aroma Models and Omission. *Chem. Senses* **2001**, *26* (5), 533–545. https://doi.org/10.1093/chemse/26.5.533.

(104) Kulsing, C.; Nolvachai, Y.; Marriott, P. J. Concepts, Selectivity Options and Experimental Design Approaches in Multidimensional and Comprehensive Two-Dimensional Gas

Chromatography. *TrAC - Trends Anal. Chem.* **2020**, *130*, 115995. https://doi.org/10.1016/j.trac.2020.115995.

(105) Chin, S. T.; Eyres, G. T.; Marriott, P. J. System Design for Integrated Comprehensive and Multidimensional Gas Chromatography with Mass Spectrometry and Olfactometry. *Anal. Chem.* **2012**, *84* (21), 9154–9162. https://doi.org/10.1021/ac301847y.

(106) Maikhunthod, B.; Morrison, P. D.; Small, D. M.; Marriott, P. J. Development of a Switchable Multidimensional/Comprehensive Two-Dimensional Gas Chromatographic Analytical System. *J. Chromatogr. A* **2010**, *1217* (9), 1522–1529. https://doi.org/10.1016/j.chroma.2009.12.078.

(107) Chin, S. T.; Eyres, G. T.; Marriott, P. J. Identification of Potent Odourants in Wine and Brewed Coffee Using Gas Chromatography-Olfactometry and Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2011**, *1218* (42), 7487–7498. https://doi.org/10.1016/j.chroma.2011.06.039.

(108) Nicolotti, L.; Mall, V.; Schieberle, P. Characterization of Key Aroma Compounds in a Commercial Rum and an Australian Red Wine by Means of a New Sensomics-Based Expert System (SEBES)—An Approach To Use Artificial Intelligence in Determining Food Odor Codes. *J. Agric. Food Chem.* **2019**, *67* (14), 4011–4022. https://doi.org/10.1021/acs.jafc.9b00708.

(109) Dugo, G.; Franchina, F. A.; Scandinaro, M. R.; Bonaccorsi, I.; Cicero, N.; Tranchida, P. Q.; Mondello, L. Elucidation of the Volatile Composition of Marsala Wines by Using Comprehensive Two-Dimensional Gas Chromatography. *Food Chem.* **2014**, *142*, 262–268. https://doi.org/10.1016/j.foodchem.2013.07.061.

(110) Stilo, F.; Liberto, E.; Spigolon, N.; Genova, G.; Rosso, G.; Fontana, M.; Reichenbach, S. E.; Bicchi, C.; Cordero, C. An Effective Chromatographic Fingerprinting Workflow Based on Comprehensive Two-Dimensional Gas Chromatography - Mass Spectrometry to Establish Volatiles Patterns Discriminative of Spoiled Hazelnuts (Corylus Avellana L.). *Food Chem.* **2020**, *340* (September 2020), 128135. https://doi.org/10.1016/j.foodchem.2020.128135.

(111) Morales, M. T.; Luna, G.; Aparicio, R. Comparative Study of Virgin Olive Oil Sensory Defects. *Food Chem.* **2005**, *91* (2), 293–301. https://doi.org/10.1016/j.foodchem.2004.06.011.

(112) Oliver-Pozo, C.; Trypidis, D.; Aparicio, R.; Garciá-González, D. L.; Aparicio-Ruiz, R. Implementing Dynamic Headspace with SPME Sampling of Virgin Olive Oil Volatiles: Optimization, Quality Analytical Study, and Performance Testing. *J. Agric. Food Chem.* **2019**, *67* (7), 2086–2097. https://doi.org/10.1021/acs.jafc.9b00477.

(113) Kolb, B.; Ettre, L. S. *Static Headspace – Gas Chromatography : Theory and Practice Second Edition*, Second.; Wiley-VCH: New York, 2006.

(114) Stilo, F.; Cordero, C.; Sgorbini, B.; Bicchi, C.; Liberto, E. Highly Informative Fingerprinting of Extra-Virgin Olive Oil Volatiles: The Role of High Concentration-Capacity Sampling in Combination with Comprehensive Two-Dimensional Gas Chromatography. *Separations* **2019**, *6* (3), 34. https://doi.org/10.3390/separations6030034.

(115) Biedermann, M.; Munoz, C.; Grob, K. Epoxidation for the Analysis of the Mineral Oil Aromatic Hydrocarbons in Food. An Update. *J. Chromatogr. A* **2020**, *1624*. https://doi.org/10.1016/j.chroma.2020.461236.

(116) Zoccali, M.; Barp, L.; Beccaria, M.; Sciarrone, D.; Purcaro, G.; Mondello, L. Improvement of Mineral Oil Saturated and Aromatic Hydrocarbons Determination in Edible Oil by Liquid-Liquid-Gas Chromatography with Dual Detection. *J. Sep. Sci.* **2016**, *39* (3), 623–631. https://doi.org/10.1002/jssc.201501247.

(117) Biedermann, M.; Grob, K. Advantages of Comprehensive Two-Dimensional Gas Chromatography for Comprehensive Analysis of Potential Migrants from Food Contact Materials. *Anal. Chim. Acta* **2019**, *1057*, 11–17. https://doi.org/10.1016/j.aca.2018.10.046.

(118) Weber, S.; Schrag, K.; Mildau, G.; Kuballa, T.; Walch, S. G.; Lachenmeier, D. W. Analytical Methods for the Determination of Mineral Oil Saturated Hydrocarbons (MOSH) and Mineral Oil Aromatic Hydrocarbons (MOAH)—A Short Review. *Anal. Chem. Insights* **2018**, *13*, 117739011877775. https://doi.org/10.1177/1177390118777757.

(119) Biedermann, M.; Grob, K. On-Line Coupled High Performance Liquid Chromatography–Gas Chromatography for the Analysis of Contamination by Mineral Oil. Part 2: Migration from Paperboard into Dry Foods: Interpretation of Chromatograms. *J. Chromatogr. A* **2012**, *1255*, 76–99. https://doi.org/10.1016/j.chroma.2012.05.096.

(120) Biedermann, M.; Munoz, C.; Grob, K. Update of On-Line Coupled Liquid Chromatography – Gas Chromatography for the Analysis of Mineral Oil Hydrocarbons in Foods and Cosmetics. *J. Chromatogr. A* **2017**, *1521*, 140–149. https://doi.org/10.1016/j.chroma.2017.09.028.

(121) Commission, E. Commission Decision 2002/657/EC Implementing Council Directive 96/23/EC Concerning the Performance of Analytical Methods and the Interpretation of Results. *Off. J. Eur. Union* **2002**, *L221* (23 May 1996), 8–36.

(122) Biedermann, M.; Grob, K. Comprehensive Two-Dimensional Gas Chromatography for Characterizing Mineral Oils in Foods and Distinguishing Them from Synthetic Hydrocarbons. *J. Chromatogr. A* **2015**, *1375*, 146–153. https://doi.org/10.1016/j.chroma.2014.11.064.

(123) Pantò, S.; Collard, M.; Purcaro, G. Comprehensive Gas Chromatography Coupled to Simultaneous Dual Detection (TOF-MS/FID) as a Confirmatory Method for MOSH and MOAH Determination in Food. *Curr. Trends Mass Spectrom.* **2020**, No. July 2020, 15–20.

(124) Zou, Y.; Niu, S.; Dong, L.; Hamada, N.; Hashi, Y.; Yang, W.; Xu, P.; Arakawa, K.; Nagata, J. Determination of Short-Chain Chlorinated Paraffins Using Comprehensive Two-Dimensional Gas Chromatography Coupled with Low Resolution Mass Spectrometry. *J. Chromatogr. A* **2018**, *1581–1582*, 135–143. https://doi.org/10.1016/j.chroma.2018.11.004.

(125) Schrenk, D.; Bignami, M.; Bodin, L.; Chipman, J. K.; del Mazo, J.; Grasl-Kraupp, B.; Hogstrand, C.; Hoogenboom, L.; Leblanc, J. C.; Nebbia, C. S.; Ntzani, E.; Petersen, A.; Sand, S.; Schwerdtle, T.; Vleminckx, C.; Wallace, H.; Brüschweiler, B.; Leonards, P.; Rose, M.; Binaglia, M.; Horváth, Z.; Ramos Bordajandi, L.; Nielsen, E. Risk Assessment of Chlorinated Paraffins in Feed and Food. *EFSA J.* **2020**, *18* (3), 1–220. https://doi.org/10.2903/j.efsa.2020.5991.

(126) Cordero, C.; Bicchi, C. Third GCxGC-MS School - Sic et Simpliciter - Mass Spectrometry Group Italian Chemical Society. Turin, Italy.

Chapter 5

Published research papers

5.1 Highly informative fingerprinting of extra-virgin olive oil volatiles: The role of high concentration-capacity sampling in combination with comprehensive two-dimensional gas chromatography

Federico Stilo[1], Chiara Cordero[1]*, Barbara Sgorbini[1], Carlo Bicchi[1] and Erica Liberto[1]

[1]Dipartimento di Scienza e Tecnologia del Farmaco, Università di Torino, Via Pietro Giuria 9, I-10125 Turin, Italy

*****Corresponding author:

Dr. Chiara Cordero - Dipartimento di Scienza e Tecnologia del Farmaco, Università di Torino, Via Pietro Giuria 9, I-10125 Torino, Italy – e-mail: chiara.cordero@unito.it; phone: +39 011 6707172;

Received: June 25, 2019

Revised: July 10, 2019

Accepted: July 11, 2019

Published: July 15, 2019

DOI: 10.3390/separations6030034

Separations 2019, 6(3), 34

5.1.1 Abstract

The study explores the complex volatile fraction of extra-virgin olive oil by combining high concentration-capacity headspace approaches with comprehensive two-dimensional gas chromatography, which is coupled with time of flight mass spectrometry. The static headspace techniques in this study are: (a) Solid-phase microextraction, with multi-polymer coating (SPME-Divinylbenzene/Carboxen/Polydimethylsiloxane), which is taken as the reference technique; (b) headspace sorptive extraction (HSSE) with either a single-material coating (polydimethylsiloxane—PDMS) or a dual-phase coating that combines PDMS/Carbopack and PDMS/EG (ethyleneglycol); (c) monolithic material sorptive extraction (MMSE), using octa-decyl silica combined with graphite carbon (ODS/CB); and dynamic headspace (d) with either PDMS foam, operating in partition mode, or Tenax TA™, operating in adsorption mode. The coverage of both targeted and untargeted 2D-peak-region features, which corresponds to detectable analytes, was examined, while concentration factors (CF) for a selection of informative analytes, including key-odorants and off-odors, and homolog-series relative ratios were calculated, and the information capacity was discussed. The results highlighted the differences in concentration capacities, which were mainly caused by polymer-accumulation characteristics (sorptive/adsorptive materials) and its amount. The relative concentration capacity for homologues and potent odorants was also discussed, while headspace linearity and the relative distribution of analytes, as a function of different sampling amounts, was examined. This last point is of particular interest in quantitative studies where accurate data is needed to derive consistent conclusions.

Key words

comprehensive two-dimensional gas chromatography-time of flight mass spectrometry; extra virgin olive oil; high concentration-capacity sampling; headspace solid-phase microextraction; dynamic headspace

5.1.2 Introduction

Comprehensive two-dimensional gas chromatography (GC × GC) is a multidimensional separation technique that enables the in-depth chemical characterization of the complex food volatilome[1]. It combines, in a single analytical platform, two separation dimensions with mass spectrometry, *i.e.*, an orthogonal measurement principle that is fundamental for analyte identification and quantitation, and automated sample preparation. Such configured platforms deliver highly efficient profiling (detailed investigation of single molecular entities) with the intrinsic fingerprinting potential, and can provide accurate and informative cross-comparative analyses [1].

The chemical characterization of the olive-oil volatilome is a challenging, although fundamental, task that is part of the quality assessment process. The composition of the volatile fraction, also referred to as the chemical signature, is an informative and diagnostic tool for oil quality characterization and sensory qualification [2–5]. Only a few of the considerable number of detectable volatiles are responsible for the positive and negative attributes that delineate olive oil sensory profiles. In fact, olive oil is, to date, the only food product whose sensory attributes are officially regulated by EU legislation, and standardized sensory assessment protocols [6,7], in the form of smelling and tasting experiments, are run by constantly updated and trained panelists. Virgin olive oil is classified into three categories, extra-virgin (EV), virgin (V), and lampante oil, according to the presence/absence and the intensity of coded defects (*i.e.*, fusty/muddy sediment, musty/humid/earthy, winey/vinegary, rancid) and the perception of the "fruity" taste.

Improved separation power and detection sensitivity are needed to efficiently extract information on the presence of potent odorants, sometimes at trace and ultra-trace concentration levels. These features, if accompanied by a structured logic of elution for chemically correlated compounds, can provide highly confident and accurate chemical characterizations, while offering new perspectives to the important problems of quality and authenticity assessment [8].

GC × GC has been adopted to characterize the olive-oil volatilome in studies that aim to define the volatile signatures of olives that differed in terms of variety, origin, and process technology [9–11] using both targeted and targeted/untargeted analyte distributions. A significant step ahead was made by Purcaro et al. [2], who explored the 2D-patterns of volatiles, after headspace solid-phase microextraction (HS-SPME) sampling, to delineate the coded defects in the chemical signature of the oil (*i.e.*, blueprint). Olive ripening and its impact on volatiles distribution and oil quality has been studied by Magagna et al. [12], who also introduced a systematic strategy for efficient untargeted and targeted investigations, which was based on pattern recognition by template matching; the strategy was defined as combined untargeted/targeted (UT) fingerprinting and has been recently extended to several other applications in the fields of food [13–15], and nutrimetabolomics [16].

All of the above-referenced studies have exploited HS-SPME as a sampling strategy; it combines the advantages of gas-phase extraction approaches with the possibility of achieving suitable enrichment factors that match method-sensitivity requirements [17–22]. However, as the zeroth dimension of an analytical process [23], HS-SPME, and more generally any sampling procedure, may impact on method information potential by discriminating analytes in function of one of their specific characteristics (polarity, volatility, etc.). If the focus of the investigation is potent odorants, the ideal sampling system should comply for: (a) Appropriate/tunable extraction selectivity; (b) high extraction efficiency toward ultra-trace analytes with high odor potency; (c) mild interaction mechanisms (sorption/partition is preferable) that limit the formation of artifacts that may be induced during the thermo-desorption of volatiles at high temperatures; and (d) the full integration of all operation steps in the analytical system [19,20,24].

In this context, it would be interesting to compare the effectiveness of a number of gas-phase extraction procedures, more specifically headspace-sampling approaches, in delineating informative volatile patterns in extra-virgin olive oil. In fact, conventional SPME can show limited extraction capability for ultra-trace odorant analysis and/or suffer from headspace saturation [25], towards major HS components. Interesting solutions to enhance the sensitivity of the HS-SPME method have been presented by Chin et al. [26], who proposed that cumulative multiple HS-SPME samplings be used in combination with a number of different fiber coatings, and followed by successive GC injections, which are delayed over time to achieve odor detection limits for GC-olfactometry (GC-O) screenings of wine aroma. More recently, Oliver-Pozo et al. [25] have developed a dynamic headspace (DHS) sampling system that enables volatiles to be accumulated in SPME fibers, while also providing higher enrichment factors than the static headspace (SHS) approach and better aldehyde and alcohol recovery. The above-mentioned methods unfortunately have some limitations, such as automation difficulties and, in the case of GC-O screenings, the fact that replicate analyses and dilution experiments are not possible.

In this scenario, a systematic investigation of the different and complementary HS sampling methods, combined with high-resolution fingerprinting by GC × GC coupled to time-of-flight mass spectrometry (TOF MS), would be of great interest, especially when the fingerprinting includes key-odorants that are responsible for the positive and negative sensory attributes of olive oil.

In this study, a selection of HS approaches has been used to study the complex volatilome of a commercial EV olive oil. They include enriched SHS with: (a) SPME with a multi-polymer coating (divinylbenzene/carboxen/polydimethylsiloxane—DVB/CAR/PDMS), taken as the reference technique; (b) headspace sorptive extraction (HSSE) either with a single-material coating (PDMS) or dual-phase coating that combines PDMS/Carbopack [27] and PDMS/EG [28] (ethyleneglycol); (c) monolithic material sorptive extraction (MMSE) by octa-decyl silica combined with graphite carbon (ODS/CB); and (d) D-HS, with either PDMS foam, operating in partition mode, or Tenax TA™, operating in adsorption mode.

The coverage of both targeted and untargeted peak-region features that correspond to detectable analytes has been examined, while concentration factors (CF) have been calculated for a selection of informative analytes, including key-odorants and off-odors. Homolog-series relative ratios have also been calculated and information capacity has been discussed.

5.1.3 Materials and methods

5.1.3.1 Reference Compounds and Samples

The pure reference compounds for the confirmation of the identity of potent odorants, *n*-alkanes (*n*-C9 to *n*-C25) for linear retention index (I^T) determination, and the reference compounds for internal standardization, α- and β-thujone, for SPME (see below) were obtained from Sigma-Aldrich (Milan, Italy). Pure dibutyl phthalate was used for internal standard (IS) working-solution preparation (0.1 g/L) and was purchased from Merck (Milan, Italy).

A commercial sample of extra virgin olive oil was selected from those collected as part of the Italian "Violin" Project (valorization of Italian olive products through innovative analytical tools—AGER Fondazioni in rete per la Ricerca Agroalimentare). In particular, the olive oil used was an EV olive oil with a protected geographical indication (PGI) quality label from Azienda Agricola Mori Concetta, PGI Toscano, olives Mariolo cultivar (San Casciano in Val di Pesa, Firenze, Italy).

A reference oil from International Olive Council (IOC) for the fusty/muddy defect was kindly supplied by Prof. Lanfranco Conte from the University of Udine.

5.1.3.2 Headspace Solid Phase Microextraction

Automated HS-SPME was performed using an MPS-2 multipurpose sampler (Gerstel, Mülheim a/d Ruhr, Germany) installed on the GC × GC-TOF MS system. SPME fibers were obtained from Supelco (Bellefonte, PA, USA) and consisted of divinylbenzene/carboxen/polydimethylsiloxane—DVB/PDMS/CAR d_f 50/30 μm–2 cm. Fibers were conditioned before use, as recommended by the manufacturer. Sampling conditions and thermal desorption parameters are summarized in **Table 5.1.1**.

Table 5.1.1. Sampling devices and conditions adopted in the study.

Acronym	Sampling Approach	Sample Weight/Volume	Temperature and Time	Other
SPME-TRIF	HS-SPME—DVB/CAR/PDMS	1.500 g oil Sampling vial: 20 mL	Temperature: 40 °C Sampling time: 60 min	Constant stirring Desorption time: 5 (min) S/SL injector: 250 °C Split ratio 1:10
HSSE-TW1	HSSE—Twister™ PDMS 1 cm	1.500 g oil Sampling vial: 20 mL	Temperature: 40 °C Sampling time: 60 min	TDU conditions: from 30 °C to 27.0°C (5 min) at 60 °C/min; Flow mode: Splitless Transfer line: 270 °C. CIS-4 PTV injector temp: −50 °C Coolant: Liquid CO_2; Injection temp program: From −50 °C to 270 °C (10 min) at 12 °C/s. Inlet operated in split mode: Split ratio 1:10.
HSSE-TW2	HSSE—Twister™ PDMS 2 cm			
HSSE-PDMS/CPB	HSSE—Twister™ PDMS—Carbopack B™			
HSSE-PDMS/EG	HSSE—Twister™ PDMS—Ethylene glycol EG			
MMSE-ODS	MMSE-ODS			
MMSE-ODS/GC	MMSE ODS—Graphite carbon			
DHS-TENAX	D-HS TENAX TA™	1.500 g oil Sampling vial: 20 mL	Incubation: 40 °C Sampling: room temperature Carrier: nitrogen Sampling flow: 10 mL/min Sampling time: 20 min	
DHS-PDMS	D-HS PDMS (foam)			

5.1.3.3 Headspace Sorptive Extraction

HSSE sampling was performed using commercial Twister™ devices. 100% PDMS d_f 500 μm 1 cm and 2 cm long twisters, as well as EG/Silicone (PDMS/EG copolymer) twisters were supplied by Gerstel (Mülheim a/d Ruhr, Germany). PDMS-Carbopack B™ d_f 500 μm–2 cm Dual Phase (DP) twisters were obtained from the Research Institute for Chromatography—RIC (Kortrijk, Belgium). Sampling was carried out in a thermostatic bath with constant stirring; HSSE twisters were suspended in the vapor phase with a stainless-steel wire, and volatiles were thus transferred to GC × GC-TOF MS by a MPS-2 multipurpose sampler (Gerstel, Mülheim a/d Ruhr, Germany) equipped with a Thermo Desorption Unit (TDU) and a CIS-4 PTV injector (Gerstel, Mülheim a/d Ruhr, Germany). Sampling conditions and thermal desorption parameters are reported in **Table 5.1.1**.

5.1.3.4 Monolithic Material Sorptive Extraction

MMSE sampling was performed using commercial devices, named MonoTrap™ (GL Sciences, Tokyo, Japan), in the form of monolithic rods consisting of a combination of octa-decyl silica and graphite carbon (ODS/GC). Sampling was carried out in a thermostatic bath with constant stirring; MonoTraps were suspended in the vapor phase with the stainless-steel wire supplied by the manufacturer, and volatiles were thus transferred to GC × GC-TOF MS by an MPS-2 multipurpose sampler (Gerstel, Mülheim a/d Ruhr, Germany) equipped with a Thermo Desorption Unit (TDU) and a CIS-4 PTV injector (Gerstel, Mülheim a/d Ruhr, Germany). Sampling conditions and thermal desorption parameters are reported in **Table 5.1.1**.

5.1.3.5 Dynamic Headspace Sampling

Dynamic headspace sampling was performed using traps assembled in the authors' laboratory. They consisted of (a) 50 mg (±2) of Tenax TA™—60/80 meshes from Supelco (Bellefonte, PA, USA) and (b) 100% PDMS foams (15 mm length—30 mg ± 2) supplied by Gerstel (Mülheim a/d Ruhr, Germany). Packing materials were assembled on inert, single taper, glass liners for the TDU unit.

During sampling, traps were gas-tight connected to the outlet of a 20 mL sampling vial kept at 40 °C, and analytes were stripped with nitrogen at 10 mL/min for 20 min (200 mL of total volume). Traps were maintained at room temperature during sampling to increase extraction efficiency. Sampling conditions and thermal desorption parameters are given in **Table 5.1.1**.

5.1.3.6 GC×GC-MS Instrument Set-up and Analytical Conditions

GC × GC analyses were performed on an Agilent 7890B GC unit coupled with a Bench TOF-Select™ system (Markes International, Llantrisant, UK) featuring Tandem Ionization™. The ion source and transfer line were set at 270 °C. The MS optimization option was set to operate in single ionization mode with a mass range between 40 and 300 m/z; the data-acquisition frequency was 100 Hz; filament voltage was set at 1.60 V. Electron ionization 70 eV.

The system was equipped with a two-stage KT 2004 loop thermal modulator (Zoex Corporation, Houston, TX, USA) cooled with liquid nitrogen, and controlled by Optimode™ V.2 (SRA Instruments, Cernusco sul Naviglio, MI, Italy). The hot-jet pulse time was set at 250 ms, the modulation period (P_M) was 4 s, and cold-jet total flow was progressively reduced as a linear function, from 40% of the mass flow controller (MFC), at initial conditions, to 8% at the end of the run.

The column set was configured as follows: ^{1}D SolGel-Wax column (100% polyethylene glycol; 30 m × 0.25 mm d_c, 0.25 µm d_f) from SGE Analytical Science (Ringwood, Australia) coupled with a ^{2}D OV1701 column (86% polydimethylsiloxane, 7% phenyl, 7% cyanopropyl; 2 m × 0.1 mm d_c, 0.10 µm d_f), from J&W (Agilent, Little Falls, DE, USA). The two columns were connected in series by a µ-union (SGE Analytical Science) and the first meter of the capillary was wrapped in the modulator slit acting as modulator capillary (*i.e.*, the loop capillary). Columns were placed in the same oven and no temperature offset was applied to the two dimensions. The carrier gas was helium at a constant flow of 1.3 mL/min. The oven temperature program was from 40 °C (2 min) to 240 °C at 3.5 °C/min (10 min).

The *n*-alkanes liquid sample solution for I^Ts determination was analyzed under the following conditions: Split/splitless injector in split mode, split ratio 1:50, injector temperature 250 °C, and injection volume 1 μL.

5.1.3.7 Analytes Identification

Analytes were identified on the basis of their linear retention indices (I^T) and MS electron impact (MS-EI) spectra that were either compared to those of authentic standards (where available) or tentatively identified through their EI-MS fragmentation patterns and I^T. The list of targeted analytes is reported in **Table 5.1.2** together with their retention times (^1t_R, ^2t_R), I^T, odor qualities, and odor thresholds, as reported in reference literature, and their correlation with coded defects [2].

Table 5.1.2. List of targeted analytes together with their retention times in the two dimensions (^1t_R and ^1t_R), linear retention times—I^T, their known role in defining attributes (defects of qualities), odor quality, odor threshold—OT (mg/Kg), and reference literature for data on sensory features. Sensory defect and quality acronyms: Fusty—F; Vinegary—V; Rancid—R; Mold—M; Morchia—Mo; and Fruity—Fr.

Compound	^1t_R (min)	^2t_R (s)	I^T	Attributes	Odor Quality	OT (mg/kg)	Ref
Heptane	4.34	1.09	750		Alkane		
Octane	5.59	1.89	800	F/V/R	Alkane	0.94	[2]
1-Octene	6.09	1.68	820	M	-	0.08	
Ethyl acetate	6.75	1.35	850	F/V	Pineapple	0.94	[2]
Butanal	7.00	1.04	857	F/M	Pungent, green	0.018	
Ethanol	7.67	1.14	883	V	Alcohol	30	[2]
Pentanal	7.75	1.35	892		-		
Nonane	7.82	2.34	895		Alkane		
3,4-Diethyl-1,5-hexadiene (RS+SR)	8.50	2.36	917		-		
3,4-Diethyl-1,5-hexadiene (meso)	8.66	2.40	923		-		
3-Methylbutanal	8.75	2.61	927	F/V	Malty	0.0054	[3]
3-Pentanone	8.84	1.47	930	V	Ether	70	[2]
(*Z*)-3-Ethyl-1,5-octadiene	9.92	2.61	973		-		
1-Penten-3-one	10.17	1.47	983	M	Mustard	0.00073	[3]
(*E*)-3-Ethyl-1,5-octadiene	10.42	2.61	993		-		
Ethyl butanoate	10.59	1.77	1000	F	Sweet, fruity	0.03	[2]
(*E*)-2-Butenal	10.75	1.38	1010		Green, fruit		
Butyl acetate	12.00	1.73	1046	F	Green, fruity, pungent, sweet	0.3	[2]
Hexanal	12.25	1.77	1054	F/Mo/V/R	Green apple, grassy	0.08	[2]
(*E*,*Z*)-3,7-Decadiene	12.25	2.74	1054		-		
(*E*,*E*)-3,7-Decadiene	12.58	2.74	1065		-		
(*Z*)-Pent-2-enal	14.00	1.51	1108	Mo	Strawberry, fruit, tomato, green, pleasant		
(*E*)-Pent-2-enal	14.08	1.52	1110	V	Green, apple, tomato, pungent	0.3	[2]
Ethyl benzene	14.25	1.78	1115	Fr	Strong		
1-Penten-3-ol	14.70	0.20	1129				
1-Butanol	15.42	1.26	1142	V/M	Winey	0.15	[3]
2-Heptanone	16.42	1.89	1161	V	Sweet, fruity	0.3	
Heptanal	16.50	1.89	1169	R	Oily, fatty, woody	0.5	
Limonene	16.91	2.15	1181		Citrus, mint		
1-Pentanol	17.17	1.45	1190	F/M/V	Fruity	3	[2]

(*Z*)-2-Hexenal	17.54	1.61	1198	Fr	Green leaves, cut grass	0.003	[4]
(*E*)-2-Hexenal	18.00	1.64	1208	Mo/V/F/R	Bitter almond, green	0.42	[3]
3-Methylbutan-1-ol	18.35	0.94	1215	F/M/Mo	Whiskey, malt, burnt	0.1	[2]
Ethyl hexanoate	18.36	1.94	1216	F	Apple peel, fruit		
1-Hexanol	19.00	1.71	1231	Fr	Fruity, banana, soft	0.4	[2]
Styrene	19.25	1.35	1237		Balsamic, gasoline		
Hexyl acetate	20.33	2.02	1263	Fr	Green, fruity, sweet	1.04	[4]
2-Octanone	20.83	2.06	1274	V	Mold, green	0.51	
Octanal	21.08	2.02	1280	Mo/R	Fatty, sharp	0.32	[2]
1-Octen-3-one	21.58	1.89	1292	Mo	Mushroom, mold	0.01	
(Z)-2-Penten-1-ol	21.75	1.22	1296		Butter, pungent		
(*E*)-4,8-Dimethyl-1,3,7-nonatriene	21.92	2.36	1300		-		
(*Z*)-3-Hexen-1-ol acetate	22.08	1.68	1304		Green, banana		
(*E*)-2-Penten-1-ol	22.09	1.02	1304		Butter, pungent	1.04	[4]
1-Heptanol	22.31	1.89	1309		Herb		
(*Z*)-2-Heptenal	22.58	1.77	1315	R	Oxidized, tallowy	0.042	[2]
(*E*)-2-Heptenal	22.67	1.81	1317	Mo/R	-	0.005	[2]
Ethyl pentanoate	23.08	2.27	1327	M	-	0.0015	
6-Methylhept-5-en-2-one	23.17	1.85	1329	Mo/F/R	Pungent, green	1	[2]
(*Z*)-3-Hexen-1-ol	24.08	1.30	1350	F/R/V	Green	1.5	[2]
(*E*)-3-Hexen-1-ol	24.92	1.35	1369	V/F	Green	6	[4]
1-Octanol	25.50	1.96	1383	Mo	Moss, nut, mushroom	0.1	
Nonanal	25.75	2.19	1388	R	Fatty, waxy, pungent	0.15	[2]
(*E*,*Z*)-2,4-Hexadienal	25.76	1.46	1388		Green		
(*E*)-2-Hexen-1-ol	25.92	1.26	1392	V	Green grass, leaves	5	[2]
(*E*,*E*)-2,4-Hexadienal	26.00	1.48	1395		-		
(*Z*)-2-Octenal	27.15	1.86	1420		Green leaf, walnut		
(*E*)-2-Octenal	27.25	1.89	1424	R	Green, nut, fat	0.004	[2]
Ethyl octanoate	27.50	2.36	1429	V	Fruit, fat	10	
1-Octen-3-ol	27.83	1.47	1437	Mo	Mold, earthy	0.05	
Acetic acid	28.50	0.97	1453	F/V/R	Sour, vinegary	0.5	[2]
(*Z*,*E*)-2,4-Heptadienal	28.58	1.60	1455	R/Mo/F	Fatty, rancid	0.36	[2]
(*E*,*Z*)-2,4-Heptadienal	28.66	3.24	1457	R/Mo/F	Fatty, rancid	10	[2]
(*E*,*E*)-2,4-Heptadienal	28.75	1.73	1459	R/Mo/F	Fatty, rancid	4	[2]
1-Nonanol	30.02	2.02	1487		Fresh, clean, floreal	0.28	[3]
Copaene	30.16	2.99	1492		Wood, spice		
Decanal	30.25	2.23	1494	R	Penetrating, sweet, waxy	0.65	[2]
(*E*)-Octa-3,5-dien-2-one	30.91	1.73	1507	V/Mo	Geranium-like	0.0005	[4]
(*Z*)-2-Nonenal	31.43	1.94	1521	R	Green, fatty	0.0045	[3]
(*E*)-2-Nonenal	31.75	1.98	1530	R	Paper-like, fatty	0.9	[3]
Propanoic acid	32.24	0.78	1541		Pungent, acidic		
(*E*)-6-Methylhepta-3,5-dien-2-one	33.83	1.64	1568	V/Mo	-	0.38	[2]
Undecanal	34.49	2.02	1597		Waxy, aldehydic, soapy		
Methyl benzoate	34.91	1.43	1608		Phenolic, prune, lettuce		
Butanoic acid	35.75	1.01	1630		-		
(*E*,*E*)-2,4-Nonadienal	35.82	1.63	1632	R	Watermelon	2.5	[2]
Ethyl decanoate	35.91	2.48	1634	V	Grape	10	[2]
(*E*)-2-Decenal	36.08	2.02	1638	R	Painty, fishy, fatty	0.01	[2]
1-Decanol	36.20	2.06	1642		Fatty, waxy, floral, orange		

Methyl butanoate	37.41	1.05	1672	F	Ether, fruit, sweet	0.06	
γ-Hexalactone	37.91	1.56	1684		Herbal, coconut, sweet		
Dodecanal	38.65	2.12	1704		Soapy, waxy, citrus		
3,4-Dimethyl-2,5-Furandione	39.16	1.35	1717		-		
α-Farnesene	40.15	2.23	1744		Wood, sweet		
Pentanoic acid	40.16	1.05	1751		Sweet, acidic, sharp		
(*E*,*Z*)-2,4-Decadienal	40.58	1.73	1756	R	Deep-fried	0.01	[2]
Phenylethyl alcohol	40.80	0.26	1763				
γ-Heptalactone	41.66	1.60	1783		Sweet, coconut, nutty		
1-Undecanol	41.93	2.10	1792		Fresh, waxy, rose, soapy		
(*E*,*E*)-2,4-Decadienal	42.25	1.64	1800	R	Deep-fried	0.18	[2]
Tridecanal	42.57	2.24	1809	R	Flower, sweet, must, clean		
Geranylacetone	43.90	1.85	1846		Magnolia, green		
Butyl benzoate	43.99	1.64	1848		Balsamic, mild, fruity		
Hexanoic acid	44.16	1.01	1853		Sweet, sour, fatty	0.7	[3]
γ-Octalactone	45.66	1.77	1891		Sweet, coconut, creamy		
Tetradecanal	46.31	2.36	1914	R	Fatty, lactonic, coconut, woody		
1-Dodecanol	47.57	2.14	1951		Soapy, waxy, clean		
Heptanoic acid	47.99	1.01	1963		Waxy, cheesy, fruity	0.1	[2]
γ-Nonalactone	49.49	1.85	2007		Fatty, coconut		
Pentadecanal	49.89	2.48	2020	R	Fresh, waxy		
Octanoic acid	51.58	1.01	2072		Rancid, soapy, cheesy	3	[2]
γ-Decalactone	53.16	1.98	2109		Fruity, fresh, peach		
Hexadecanal	53.31	2.58	2126	R	Cardboard		
1-Tridecanol	54.23	2.17	2155		Musty		
Nonanoic acid	54.91	1.05	2176		Fatty, waxy, cheesy		
Methyl palmitate	55.90	2.44	2208		Oily, waxy, fatty, orris		
Ethyl palmitate	57.06	2.61	2247		Waxy, fruity, creamy		
Decanoic acid	58.24	1.05	2286		Soapy, waxy, fruity		
Palmitic acid	59.56	2.69	2332		Waxy, creamy, fatty, soapy		
Heptadecanal	59.72	2.74	2338	R	-		
1-Tetradecanol	60.40	2.21	2361		Coconut		
Butyl palmitate	62.23	3.11	2438				

5.1.3.8 Method-Performance Parameters

A simple validation protocol was designed to establish method performance in terms of precision for quantitative descriptors (*i.e.*, 2D peak volumes measured on analytes target ion—*Ti*). This protocol included experiments on HS-SPME with DVB/CAR/PDMS (SPME-TRIF), HSSE with 100% PDMS Twister™ (HSSE-TW1), and DHS sampling with PDMS foams (DHS-PDMS). Precision data (intra and inter-week precision on retention times and 2D peak volumes on selected odorants *Ti*), were evaluated by replicating analyses (six replicates) over a period of three weeks. Results are reported as Supplementary Material— **Supplementary Table 5.1.1** and are expressed as percentage relative standard deviation (RSD%).

5.1.3.9 Raw Data Acquisition and GC×GC Data Handling

Data were acquired using TOF-DS software (Markes International, Llantrisant, UK) and processed using GC Image ver 2.8 (GC Image, LLC, Lincoln, NE, USA). Statistical analyses were performed using XLStat (Addinsoft, Paris, France).

5.1.4 Results and Discussion

This study aimed to evaluate how sample preparation can impact upon the high-resolution fingerprinting of the olive-oil volatilome when GC × GC-TOF MS is exploited for both its untargeted and targeted investigation potential. The oil volatilome has been chosen as the model here because of its chemical complexity, also referred to as chemical dimensionality [29], and the high informational density it brings to oil-quality characterization and sensory evaluation.

The performance of the different sample preparation approaches is evaluated by considering untargeted and targeted peak-region features using the UT fingerprinting strategy, while a focus on some odor-active compounds is also discussed in view of their relevant roles in delineating olive-oil aroma.

The next paragraph will introduce the olive-oil volatilome by illustrating 2D-peak patterns as they result from a polar × medium-polarity column combination.

5.1.4.1 Extra-Virgin Olive Oil Complex Volatilome by GC×GC Fingerprinting

The chemical complexity of the olive-oil volatilome can be effectively described by the concept defined by Giddings, known as chemical dimensionality [29], which was introduced to describe the degree of order/disorder that can be achieved in multidimensional separations. Volatiles in olive oil are generated from multiple chemical reactions, mainly promoted by endogenous or exogenous enzymes, that occur in olive primary metabolites during fruit ripening and, later, in post-harvest and processing stages. In addition, storage and shelf-life may add additional complexity, resulting in thousands of volatiles that belong to different chemical classes and differ in their polarity, volatility and concentration.

High resolution separations and orthogonal detection by mass spectrometry are fundamental for the accurate fingerprinting of volatiles. In addition, the possibility of obtaining structured separation patterns for chemically correlated compounds is of great help; analyte identification can be confirmed by observing analyte relative elution, while for unknowns, information about their relative polarity and volatility can be reliably hypothesized because of the multiple retention mechanisms used by the technique. **Figure 5.1.1A** shows the 2D pattern of the PGI Toscano extra virgin olive oil (EVOO), which was taken as a reference sample for the study. The number of detectable 2D-peaks, over a signal-to-noise ratio (SNR) threshold of 50, is about 1500, and reliable identification was possible for 114 of them by matching $^{1}D\ I^{T}$ and MS spectra with those collected in commercial and in-house databases [30,31].

Of the most important classes of informative volatiles, compounds formed from lipoperoxide cleavage, also referred to as the lipoxygenase (LOX) signature (**Figure 5.1.1B**), are fundamental for the definition of fresh-green and fruity notes, which are considered positive attributes. Of the C6 unsaturated alcohols and aldehydes, hexanal was connoted by green apple and grassy notes, (Z)-3-hexenal had green and grassy odor, (E)-2-hexenal was described as bitter almond and green, (Z)-3-hexenol and (E)-3-hexenol had both green notes, (E)-2-hexenol evokes green grass and leaf odors,

while (E,Z)-2,4-hexadienal had a green odor. All of these compounds were formed from linoleic and linolenic acid oxidative cleavage, as promoted by lipoxygenase (LOX) and hydroperoxide lyase (HPL) pathways [32].

Figure 5.1.1C illustrates the homologous series of linear saturated and unsaturated aldehydes, together with a few ketones that were most likely formed by non-enzymatic hydroperoxide cleavage. This last group of analytes generally provides information on shelf-life evolution [33]; increasing concentrations of potent odorants of this class bring rancid and fatty notes. Heptanal, octanal, and nonanal, although possessing different odor potencies, had fatty and waxy notes, decanal and undecanal were described as waxy and fatty, while the series of (E)-2-unsaturated aldehydes (*i.e.*, from (E)-2-heptenal to (E)-2-Decenal) had odors that evoked apple and green leaf, up to fatty and tallowy notes for the higher homologues.

The enlarged area of **Figure 5.1.1D** shows the retention region of a group of branched unsaturated hydrocarbons, eluting later in the ^{2}D. They were identified by Angerosa et al. [34] in olives at early stages of ripening. They were 3,4-diethyl-1,5-hexadiene (RS + SR), 3,4-diethyl-1,5-hexadiene (meso), (*5Z*) and (*5E*)-3-Ethyl-1,5-octadiene, (*E*,*Z*)- and (*E*,*E*)-3,7-decadiene and (*E*)-4,8-Dimethyl- 1,3,7-nonatriene [12,34].

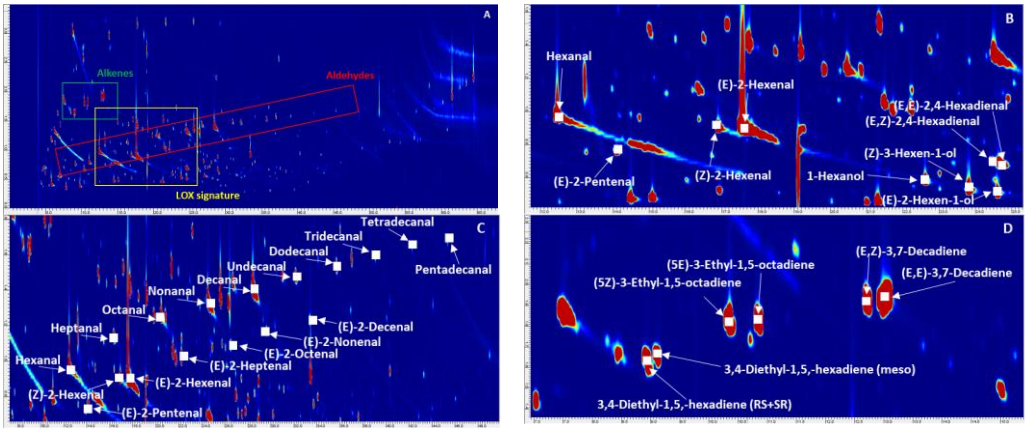

Figure 5.1.1. 2D pattern of the PGI Toscano EVOO (**5.1.1A**), together with some informative patterns of volatiles. The lipoxygenase (LOX) signature is shown in panel (**5.1.1B**), linear saturated and unsaturated aldehydes are illustrated in panel (**5.1.1C**), while panel (**5.1.1D**) shows the enlarged area of branched unsaturated hydrocarbons correlated to olive fruit freshness [34].

This chemical complexity can also be explored by simple datapoint feature fingerprinting [35]; this pointwise approach, also explored by Vaz-Freire et al. [9] in a study focused on olive oils from different cultivars, enables point-by-point, or pixel-by-pixel, chromatogram comparisons to be performed. In a GC × GC-TOF MS chromatogram, every datapoint corresponded to a detector event, *i.e.*, a single MS spectrum. Features located at the same retention times in a pair of chromatograms were implicitly matched using this approach. **Figure 5.1.2** shows the 2D-patterns of the analyzed EV olive oil (**Figure 5.1.2A**), and of a reference oil from IOC for the fusty/muddy defect (**Figure 5.1.2B**). The comparative visualization was rendered as the colorized fuzzy ratio in **Figure 5.1.2C**, and analyte relative abundance in the two samples was highlighted by color-coding (green, red, and light-grey). In this specific pair-wise comparison, performed on the normalized total ion current (TIC) response to the smooth concentration effect, several compounds were present in a higher relative ratio in the analyzed sample (*e.g.*, fusty/muddy oil). They were 2,3-butanediol, 2-butenal, 3-methyl-1-butanol acetate, 3,4-dimethyl-2-hexanone, 2-heptanone, heptanal, 6-methyl-5-hepten-2-one, nonanal, propanoic acid, butanoic acid,

3-methyl butanoic acid, and pentanoic acid. On the other hand, red colored datapoints correspond to compounds that were more abundant in the reference sample (*e.g.*, EV oil). They were 2-methyl-1-butanol, (*E*)-2-hexenal, (*Z*)-3-hexenol, hexanol, (*E*)-2-hexenol, acetic acid, and dodecanol.

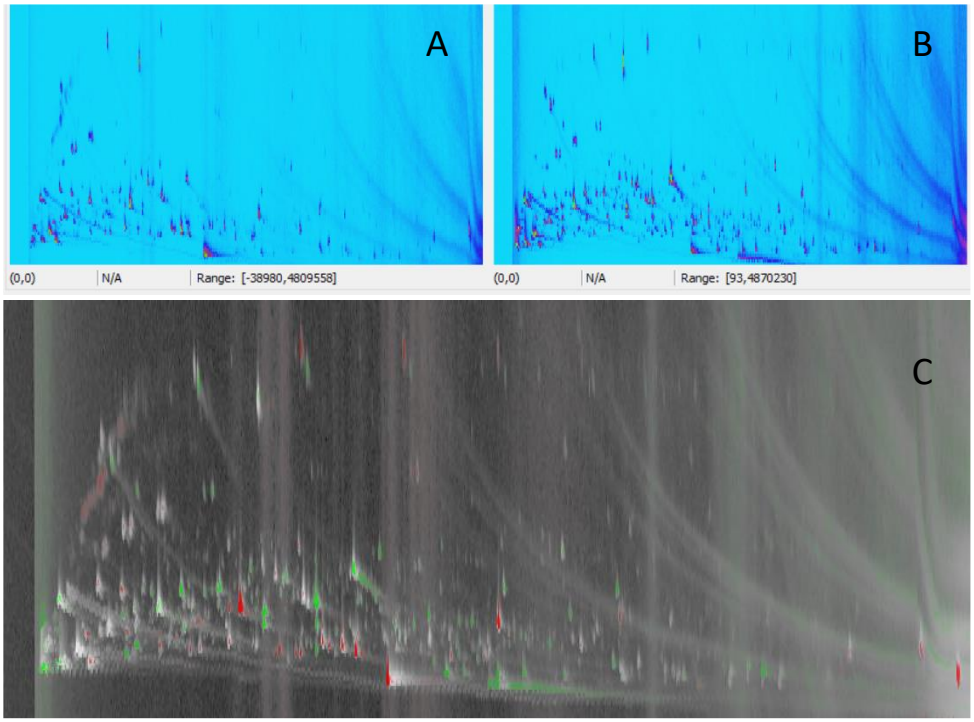

Figure 5.1.2. 2D pattern of the PGI Toscano EVO oil (**5.1.2A**) and that of a reference oil from IOC for the fusty/muddy defect (**5.1.2B**). Comparative visualization is rendered in (**5.1.2C**) as colorized fuzzy ratio and analyte relative abundance in the two samples is highlighted by color-coding (green, red, and light-grey). For details see text.

This last approach is a clear example of how high-resolution bi-dimensional separation can effectively compare sample patterns and give prompt results on compositional differences. The same approach, performed on 1D-GC profiles, would fail for minor components or for those affected by co-elution issues, although it would be effective for more highly abundant peaks/components.

The next paragraph will discuss the results of the differential information provided by the explored sampling approaches, which are based on untargeted peak-region distribution.

5.1.4.2 Sampling Information Potentials Based on Untargeted Data

Smart-template-concept based pattern recognition [36], was used to study sampling effectiveness and 2D-pattern information potential. The template corresponds to the pattern of 2D-peaks and/or their corresponding graphic objects created over the 2D-peak contour of a reference image(s) (single or composite image) [16]. This template is then used to recognize similar peak patterns in an analyzed image(s) [37]. Template objects (2D-peak and/or graphic) carry various metadata such as retention times, I^T, mass spectrum, compound name, compound group, informative ions and their relative ratios, additional constraint functions, and qualifier functions. Typical constraint functions are those that limit positive correspondence to analytes that show an MS-fragmentation pattern similarity above a fixed threshold, while qualifier functions may provide information about quality indicators, as calculated via scripts that

are developed ad hoc. These functions enable highly specific cross-comparison of data, providing 2D-peak re-alignment across samples with high consistency.

Peak-region features, introduced by Reichenbach and co-workers [38,39], are template objects that give considerable assistance in compensating for temporal inconsistencies, detector fluctuations and highly variable sample compositions. They provide greater robustness than peak-features methods, while offering all the advantages of one-feature-to-one-analyte selectivity. The combination of both, 2D-peaks and peak-regions was adopted for combined untargeted and targeted fingerprinting—UT fingerprinting strategy [5,12,13,16,40].

In UT fingerprinting, a group of reliable peaks, which positively match across all or most chromatograms in a set [41], is established and then used to re-align chromatograms [42], before their combination into a single-composite chromatogram. The composite chromatogram corresponds to the sum of the re-aligned datapoint responses in the 2D retention time plane. It can therefore be treated as a regular 2D-chromatogram for peak detection and metadata extraction. The sub-set of reliable 2D-peaks and all the peak-regions extracted by peak outlines in the composite chromatogram are collected in a feature template, or consensus template, which covers the chemical dimensionality of the whole sample-set, and that is capable of capturing chemical variability with high specificity. The subset of known compounds can be completed from all detected analytes by filling their metadata fields (compound name, ion ratios, I^T); this subset—reported in **Table 5.1.2**—can be separately processed for the interpretation of results.

A schematic of the UT fingerprinting process is illustrated in Supplementary Material—**Supplementary Figure 5.1.1** together with some details on targeted and untargeted 2D peaks and peak-regions.

In this specific application, UT fingerprinting is extremely useful since it enables a consistent re-alignment of detected features (UT peaks) when different sampling approaches are applied. In this context, the cross-comparative analysis aims at revealing 2D-pattern differences brought by the extraction techniques rather than those related to the different composition of a selection of samples.

The distribution of about 1500 untargeted and targeted 2D-peak-regions is illustrated as a heat-map in **Figure 5.1.3**. Analyte responses (absolute 2D-peak volume) were normalized using the Z-score (*i.e.*, mean subtraction and normalization to the standard-deviation) and clustered (hierarchical clustering—HC) based on Spearman rank correlation. For each sampling approach, three analytical replicates were computed.

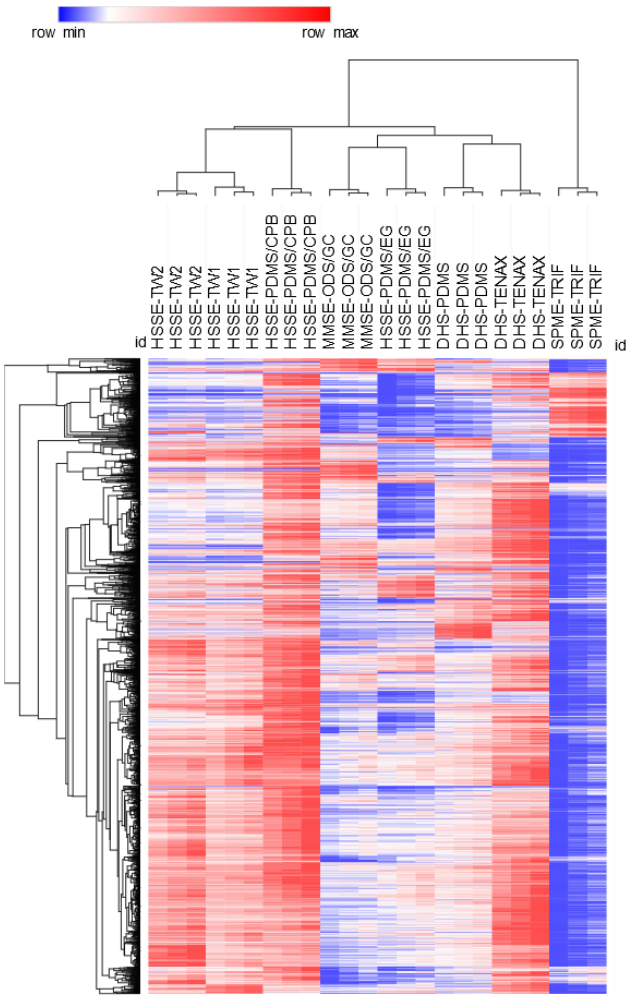

Figure 5.1.3. Heat-map showing the analyte-response (absolute 2D-peak volume) distribution resulting from the different sampling approaches. Data were normalized by Z-score (*i.e.*, mean subtraction and normalization to the standard deviation) and clustered (hierarchical clustering—HC) based on Spearman rank correlation. For each sampling approach, three analytical replicates were computed.

The tested techniques with higher amounts of polymer gave better results in terms of concentration capacity, as expected; the approach with the highest TIC response, calculated over all UT peaks, was HSSE-PDMS/CPB (1.56×10^7), as indicated by the predominance of red colored spots on the heat-map. This was followed by DHS-TENAX (1.17×10^7) and then by DHS-PDMS (8.89×10^6). As expected, SPME-TRIF was the approach with the lowest overall TIC response. However, its coverage for key-analytes is quite good, as will be illustrated in the next section.

These following preliminary considerations can be confirmed upon observing the HC results (**Figure 5.1.3**): SPME-TRIF clusters independently of the other approaches; HSSE with PDMS twisters (HSSE-TW1 and HSSE-TW2) and the combination of PDMS and Carbopack B (HSSE-PDMS/CPB) were all clustered together with the sub-cluster of the two PDMS devices, which differed in the amount of extraction polymer. Interestingly, both DHS approaches were closely clustered, as were MMSE-ODS/GC and HSSE-PDMS/EG, which, however, showed limited accumulation capacity.

The next section will discuss sampling performance towards a selection of targeted analytes of interest for olive-oil sensory profiles.

5.1.4.3 Focus on Informative Targeted Analyte Signatures

Of the group of analytes that were targeted, the class of alcohols provides information on LOX activity in ripened fruits, and on enzymatic reactions promoted by bacteria and molds. Their concentration factors (CFs), obtained from **Equation 5.1.1**, are calculated over SPME-TRIF, which is taken as the reference technique:

Equation 5.1.1 $$CF_i = \frac{A_i\ dev\ x}{A_1 SPME\ TRIF},$$

where A_i is the 2D-chromatographic area of the *i*-analyte obtained by applying a certain sampling device/approach (*dev x*) and A_i *SPME TRIF* is the analyte *i* chromatographic area resulting from the HS-SPME approach, which is taken as the reference technique.

Results were visualized in the heat-map of **Figure 5.1.4A**; alcohols were, in general, better recovered by MMSE-ODS/GC and DHS-PDMS sampling, with the latter showing a mean CF of 120 and a median of 3.36. The alcohols that were recovered most by DHS-PDMS sampling were 2-methyl-1-propanol (CF 1548) and 1-tetradecanol (CF 557). Indeed, DHS shows lower CF values for the most volatile members of the linear series (*i.e.*, ethanol, pentanol, hexanol, heptanol, octanol, and nonanol), most probably because of their breakthrough (see below for further comments). The two C6 unsaturated alcohols ((*E*) and (*Z*)-3-hexenol), with a high information potential being fundamental for their green note contribution to the overall flavor, were better recovered by all devices, except SPME-TRIF. It is worth noting that their relative abundance in the sample was so high that their detection by HS sampling was not generally limiting. Phenylethyl alcohol, the aromatic member of this chemical class, showed the opposite tendency, being better enriched by SHS with HSSE-PDMS and dual-phase twisters HSSE-PDMS/CBP.

Another interesting chemical class is that of unsaturated aldehydes; they are formed by the β-scission of unsaturated fatty-acid hydroperoxides, and their odor thresholds are generally lower than those of the saturated homologs. They have been described in the volatile fraction of defected oils [2] (rancid, moldy and fusty) and contribute with fatty and rancid notes. **Figure 5.1.4B** shows CF values within a sub-set of techniques for the most relevant members of this series. In this case, the HSSE approach gave higher average CFs than the other techniques. For (*E*)-2-octenal and (*E*)-2-nonenal, MMSE-ODS/GC reported CFs of 10 and 13, showing good selectivity, compared to SPME-TRIF, for these two analytes, which were connoted by very low OTs, *i.e.*, 0.004 and 0.9 mg/kg.

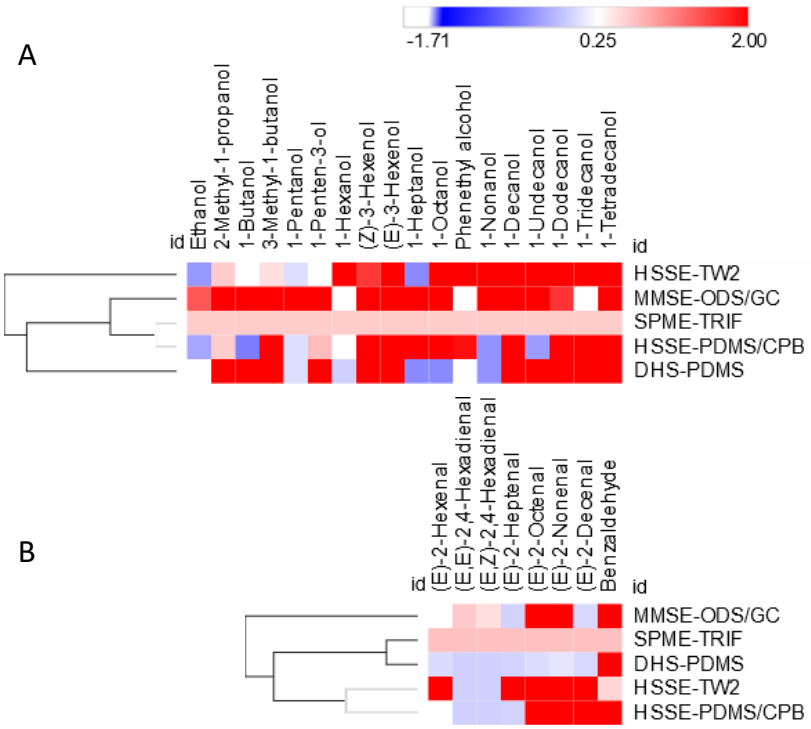

Figure 5.1.4. Heat-map illustrating concentration factor (CF) values for the alcohol series (**5.1.4A**) and unsaturated aldehydes (**5.1.4B**) relative to the selected devices/approaches, with solid-phase microextraction (SPME)-TRIF being taken as the reference.

The results could also be examined by observing the relative concentration capacity of each device/approach toward selected analytes and by taking the most effective one as a reference. In a few words, a unit value was assigned to the most effective approach towards each single analyte and the ratios between analyte responses were calculated in a 0–1 range for the other sampling devices. Sampling selectivity was emphasized with this parameter, if calculated for homologs, therefore it is a parameter to be considered in method optimization. Results for saturated aldehydes, alcohols, and short-chain fatty acids are shown in histograms in **Figures 5.1.5A–C**.

Saturated aldehydes, the series from C6 to C17, showed an interesting trend. The static headspace approaches (SPME-TRIF, MMSE-ODS/GC, and HSSE), better recover the most volatile members of the series (from C6 up to C12), provide higher amount of the accumulating polymer, and more uniform relative analyte recovery, although they do so to different extents. For example, HSSE-TW2 achieves unit values for C6–C9 and C12, and is one of the best performing devices for these analytes. For higher homologs in the series, C12–C17, the highest relative concentration capacity was shown by DHS-PDMS. Interestingly, opposite trends are observed with SPME-TRIF, which discriminates this series in favor of the most volatile species, and DHS-PDMS, which better enriches the less volatile members (C13–C17).

For the alcohols' series (**Figure 5.1.5B**), the most effective approach to enrich the higher homologs, from C9 to C14, was HSSE-TW2. Complementary behavior is shown by MMSE-ODS/GC and DHS-PDMS; they effectively enrich alcohols from C2 to C8. SPME-TRIF represents this chemical class well and displays less discrimination than observed for aldehydes.

Fatty acids (**Figure 5.1.5C**), within the C2 to C10 range, were well characterized by HSSE-TW2, which maximized the extraction for those with lower volatility. On the other hand, volatility discrimination is evident in the SPME-TRIF profile.

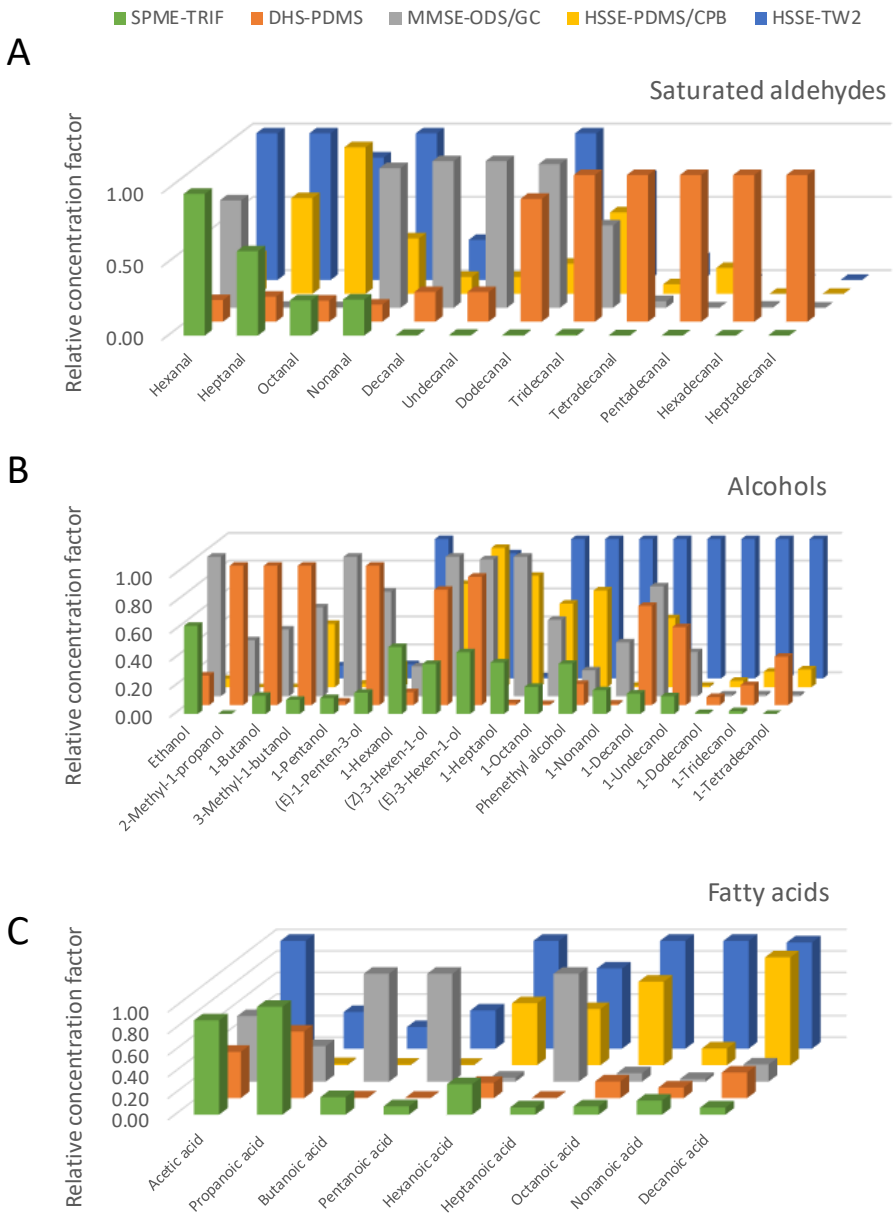

Figure 5.1.5. Relative concentration capacity of each device/approach towards selected analyte classes (saturated aldehydes—(**5.1.5A**), alcohols—(**5.1.5B**), and fatty acids—(**5.1.5C**)) with the most effective approach being taken as the reference.

These experimental results demonstrate how high concentration-capacity (HCC) HS can provide useful information on olive-oil volatile-fraction compositions and also enable trace and ultra-trace analytes to be quantitatively recovered. However, as demonstrated by observing CF trends in absolute and relative terms, this information is partial and can only be adopted for cross-sample analyses. Any conclusion about the quantitative distribution of volatiles in the sample would be erroneous if not obtained via accurate quantitation methods [43,44].

To quantify volatiles from a complex matrix, standard addition (SA) or multiple headspace extraction (MHE) should be adopted. These methods require headspace linearity conditions [45], for all targeted analytes, meaning that no saturation effects should compromise result accuracy. The next paragraph will focus on the evaluation of headspace linearity for a selection of potent odorants after SPME-TRIF sampling in the selected time–temperature conditions. Considerations about internal standard (IS) quantitation will also be discussed as this approach is widely, although often erroneously, adopted in several applications.

5.1.4.4 Headspace Linearity and Its Impact on Analyte Relative Distribution

Volatile fraction profiling [46], can be run for effective cross-sample comparison, as in the case of a selection of olive oils that differ in origin, extraction technology, and storage time, so analytes and/or informative markers can be compared using chromatographic quantitative indicators that are based on peak areas (raw areas, percentage area), and peak-volume percentage in the case of GC × GC (raw volume, percentage volume), or IS normalization (normalized area, normalized volume). These methods, which are based on relative/normalized responses, have been accepted by the scientific community for several application fields [47], despite being inaccurate and misleading if treated as absolute concentration indicators.

They do not take into consideration several concurring effects, such as the effect exerted by the condensed phase (*i.e.*, matrix effect) on the release of an analyte into the HS or the displacement and multiple-equilibria that occur with adsorption polymers (carboxen—CAR and divinylbenzene—DVB).

Procedures that compensate, or model, the matrix effect are those known as quantitation approaches, and these are based on either external or internal calibration with authentic standards or stable isotopologues of the target analytes.

Those suitable for liquid samples are: (a) External calibration in matrix-matched blank samples; (b) standard addition (SA) by spiking the sample with known incremental amounts of analyte(s); (c) stable isotope dilution assays (SIDA), which are a specific application of SA; and (d) multiple headspace extraction (MHE).

HS linearity must be accomplished for an accurate quantitation of targeted analytes, whichever approach is applied. [45] This condition is verified when, under pre-determined sampling conditions (*e.g.*, temperature, time, and phase-ratio), the condensed phase (liquid or solid sample) releases a minimal analyte amount into the HS without saturation and, in addition, a linear function can be established between the analyte concentration in the sample (C_0) and its concentration in the gas phase (C_g). Linearity is easily achievable for trace and sub-trace components, but becomes challenging in multi-analyte quantitation with complex volatile fractions. The linear range depends on analyte partition (K_{hs}) and activity coefficients. It generally ranges between 0.1% and 1% of actual concentration in the sample and can be tuned by modifying the sampling extraction phase (adsorption/sorption mechanisms), the amount of extraction polymers, the sampling temperature and time, as well as by modifying the phase ratio between HS (V_h) and the condensed phase volume (V_s).

In this study, HS linearity by SPME-TRIF sampling, which was taken as the reference technique, was explored, for the analytes listed in **Table 5.1.2**, by analyzing different amounts of EV olive oil, between 1.500 and 0.100 g, chosen on the basis of previous studies [3,48–51]. Results for normalized peak volumes and percentage responses are reported in Supplementary Material - **Supplementary Table 5.1.2**, while they are summarized in **Table 5.1.3** for a selection of potent

odorants. Responses, obtained from three analytical replicates randomly distributed within a uniform sampling batch, were normalized over the IS (*i.e.*, α-tujone) and refer to 1.500 g of sample (the amount adopted for sampling-device screening) or to 0.100 g. At 0.100 g, at least 50% of the analytes failed to reach HS saturation for the applied sampling conditions (20 mL sampling vial, 40 °C, 60 min).

Normalized volume ratios, calculated between 0.1 and 1.5 g of sampling amount, differ widely within the group of analytes considered. A histogram of **Figure 5.1.6** summarizes the results. In particular, an average value of 1 (highlighted by a red line) was obtained for (*Z*)-2-penten-1-ol, hexanal, (*E*)-2-hexenal, ethyl acetate, and 2-pentanol. On the other hand, responses were about 10- and 2.6-fold higher for 1-octanol and (*E*)-2-decenal, respectively, with 0.1 g of sampling amount. In this last case, displacement effects that occur on the CAR-DVB material reasonably influence the extraction. On the other hand, response ratios for (*Z*)-3-hexen-1-ol, (*E*)-2-penten-1-ol, (*E*)-2-hexen-1-ol, (*E*)-2-octenal, and (*E*,*E*)-2,4-hexadienal were lower at 0.1 g than at 1.5 g. The results clearly highlight the complexity of the multiple-equilibria that coexist in HS-SPME-TRIF sampling and suggest that deeper investigations into HS linearity conditions were needed to validate the hypothesis of HS saturation at the very least. SHS experiments are required to investigate the coexisting effects, such as displacement and phase ratio. Multiple headspace extraction (MHE) was therefore applied to confirm these findings.

The MHE approach is a dynamic, stepwise gas extraction carried out on the sample headspace; it was introduced for S-HS applications and adapted to HS-SPME, resulting in a technique referred to as MHS-SPME [45,52–56].

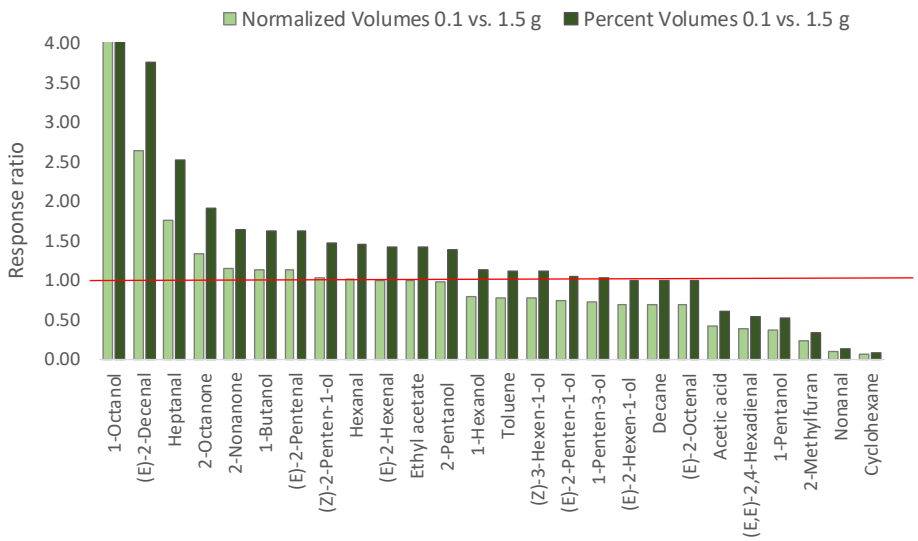

Figure 5.1.6. Normalized response ratios (normalized volumes and % volumes) calculated between 0.1 and 1.5 g for selected analytes.

It consists of three main steps:

Step 1. Exhaustive analyte extraction from calibration solutions in a range that matches real-sample concentrations.

Step 2. Exhaustive analyte extraction from selected samples that show comparable matrix effects in order to define HS linearity boundaries.

Step 3. Use of MHE on samples of interest.

Steps 1 and 2 define the function according to the cumulative instrumental response that is the results of consecutive extractions from the same sample/calibration solution. After four to six consecutive extractions, in HS linearity [45,54], exhaustiveness is accomplished and the decrease in the analyte response (chromatographic peak area) should be exponential.

By summing the instrumental response (A) from each HS extraction, the total response that is virtually generated by the analyte amount in the sample (A_T) can be estimated - **Equation 5.1.2**:

Equation 5.1.2 $$A_T = \sum_{i=1}^{=\infty} A_i = A_1 \frac{1}{(1- e^{-q})} = \frac{A_1}{(1- \beta)},$$

where A_T is the total estimated response (chromatographic area), A_1 is the response after the first extraction, and q is a constant associated with the response exponential decay (β) over consecutive extractions. The term q is obtained from the natural logarithm of the analyte response as a function of the number of extractions, and a linear regression equation (**Equation 5.1.3**) can be calculated:

Equation 5.1.3 $$\ln A_i = a\,(i-1) + b,$$

where i is the number of extraction steps, b is the intercept on the y axis, and a is the slope. β (e^{-q}) is analyte and matrix dependent and can be adopted to confirm, or refute, HS linearity.

In this study, β values and decay functions (all $R^2 \geq 0.995$) were estimated over four successive extractions on sample headspace from 1.500, 1.000, 0.500, and 0.100 g of EV olive oil. Results are reported in **Table 5.1.3** for the 0.100 g sampling. It is worth noting that HS saturation occurred in the range between 0.500–1.500 g with relative β values of ~1 for most of the analytes examined.

Table 5.1.3. Normalized and % volumes obtained by sampling 0.100 and 1.500 g of reference EV oil, as calculated after the first extraction. % Error refers to the relative error in response indicators as calculated between the two sampling amounts, and by taking the lowest as reference. For MHS- SPME-TRIF experiments, slope (β) and decay function formulae are reported.

	Sampling Amount 0.1 g		**Sampling Amount 1.5 g**			**MHS-SPME-TRIF (0.1 g)**	
	Norm. Volume	**% Volume**	**Norm. Volume**	**% Volume**	**% Error**	***β***	**Decay Function**
1-Octanol	1.10	9.42×10^{-3}	0.11	6.75×10^{-4}	−89.75	0.67	y = −0.40x + 14.1
(*E*)-2-Decenal	0.92	7.82×10^{-3}	0.35	2.07×10^{-3}	−62.06	0.59	y = −0.54x + 14.7
Heptanal	1.36	1.16×10^{-2}	0.77	4.62×10^{-3}	−43.25	0.47	y = −0.76x + 14.6
2-Octanone	0.20	1.71×10^{-3}	0.15	8.95×10^{-4}	−25.33	0.76	y = −0.28x + 12.5
2-Nonanone	0.17	1.46×10^{-3}	0.15	8.91×10^{-4}	−12.96	0.62	y = −0.48x + 12.7
1-Butanol	3.38	2.88×10^{-2}	2.97	1.77×10^{-2}	−12.05	0.37	y = −1.00x + 16.1
(*E*)-2-Pentenal	25.42	2.17×10^{-1}	22.45	1.34×10^{-1}	−11.65	0.28	y = −1.28x + 17.5
(*Z*)-2-Penten-1-ol	9.06	7.73×10^{-2}	8.81	5.26×10^{-2}	−2.74	0.39	y = −0.94x + 17.1
Hexanal	366.32	3.13×10^{0}	360.14	2.15×10^{0}	−1.69	0.48	y = −0.74x + 20.6
(*E*)-2-Hexenal	0.15	1.24×10^{-3}	0.15	8.68×10^{-4}	0.21	0.67	y = −0.74x + 23.6
Ethyl acetate	0.61	5.23×10^{-3}	0.62	3.68×10^{-3}	0.67	0.63	y = −0.46x + 13.4
2-Pentanol	0.19	1.61×10^{-3}	0.19	1.15×10^{-3}	2.84	0.69	y = −0.37x + 12.7
1-Hexanol	304.90	2.60×10^{0}	384.66	2.30×10^{0}	26.16	0.75	y = −0.28x + 19.8
Toluene	1.14	9.72×10^{-3}	1.46	8.71×10^{-3}	28.23	0.44	y = −0.81x + 14.6
(*Z*)-3-Hexen-1-ol	71.87	6.13×10^{-1}	92.98	5.55×10^{-1}	29.37	0.74	y = −0.30x + 18.5
(*E*)-2-Penten-1-ol	220.87	1.88×10^{0}	301.14	1.80×10^{0}	36.34	0.28	y = −0.82x + 20.3
1-Penten-3-ol	191.90	1.64×10^{0}	267.17	1.59×10^{0}	39.23	0.44	y = −0.81x + 20.0
(*E*)-2-Hexen-1-ol	171.70	1.46×10^{0}	247.77	1.48×10^{0}	44.30	0.73	y = −0.32x + 19.3
Decane	0.34	2.90×10^{-3}	0.49	2.93×10^{-3}	44.54	0.73	y = −0.32x + 13.0
(*E*)-2-Octenal	0.41	3.48×10^{-3}	0.59	3.52×10^{-3}	44.89	0.44	y = −0.40x + 13.1
Acetic acid	15.99	1.36×10^{-1}	37.64	2.25×10^{-1}	135.36	0.45	y = −0.81x + 17.9
(*E*,*E*)-2,4-Hexadienal	19.17	1.64×10^{-1}	50.81	3.03×10^{-1}	165.05	0.48	y = −0.73x + 17.9
1-Pentanol	3.54	3.02×10^{-2}	9.89	5.90×10^{-2}	179.08	0.51	y = −0.68x + 15.8
2-Methylfuran	0.25	2.17×10^{-3}	1.11	6.64×10^{-3}	338.20	0.57	y = −0.56x + 13.1
Nonanal	0.69	5.88×10^{-3}	7.88	4.70×10^{-2}	1043.06	0.58	y = −0.55x + 15.6
Cyclohexane	15.81	1.35×10^{-1}	273.40	1.63×10^{0}	1629.52	0.34	y = −1.07x + 18.0

Based on these data, any conjecture that is made as to the amount of analytes in a sample derived from HS-SPME sampling without having applied suitable quantitative strategies (*e.g.*, MHE or SA) would be inconsistent and may lead to erroneous results. In practice, if internal standardization is used to estimate analyte concentrations in the sample using HS-SPME-TRIF sampling on 1.500 g, as reported in some research papers that deal with EV olive oil volatilome, relative errors would range between −90% (for 1-octanol) to 1630% (for cyclohexane). These errors are just an underestimation of the actual ones, since additional sources of error, such as detection or response factors and chromatographic extra-column effects, are not computed. Only external calibration [43] or suitable flame ionization detector (FID) response-factor estimation [43,44],

accompanied by MHS-SPME or SA, can lead to the accurate quantitation of multiple analytes in samples headspace.

Alternatively, sorption-based materials (PDMS or polyethylene glycol—PEG) with relatively high amounts of extraction polymer were able to overcome most of the limitations highlighted in the examples discussed herein. New commercial devices that benefit from the advantages of SPME, in terms of automation and instrument integration, such as SPME arrows [57,58] and Hisorb™ solutions (Markes International) deserve consideration.

5.1.5 Conclusions

This study showed how high concentration capacity headspace sampling could successfully be integrated in a GC × GC-TOF MS platform for highly informative fingerprinting of the complex EV olive oil volatilome. The influence of different variables on extraction effectiveness (CFs and relative CFs) was shown focusing the attention on potent odorants and/or on key-markers known to be correlated with oil sensory defects. Among the others, SPME-TRIF confirmed its good quali-quantitative coverage of the different chemical dimensions present in the EV oil volatilome. However, to derive consistent and accurate quantitative considerations, headspace linearity should be accomplished at the sampling stage. When saturation occurs in the sample headspace, analytes displacements and distribution on the extraction polymers may change giving unrealistic results in quantitative terms.

Several studies erroneously apply HS-SPME followed by IS "apparent" quantitation by working outside the boundaries of headspace linearity and/or by using adsorption polymers (CAR and DVB above all) that may be affected by displacement and competition phenomena between analytes. In addition, such apparent quantitation approaches do not take into account the actual analytes distribution constants (K_{HS} and K_D) and detector response factors that, in the case of MS detection, may vary greatly analyte by analyte.

5.1.6 Supplementary Material

Supplementary material at the Google Drive's link:
https://drive.google.com/drive/folders/1dw3d3BvjJIrAmzTWl_qhkf1GFTmHyIBF?usp=sharing

References

(1) Cordero, C.; Kiefl, J.; Reichenbach, S. E.; Bicchi, C. Characterization of Odorant Patterns by Comprehensive Two-Dimensional Gas Chromatography: A Challenge in Omic Studies. *TrAC Trends Anal. Chem.* **2019**, *113*, 364–378. https://doi.org/10.1016/j.trac.2018.06.005.

(2) Purcaro, G.; Cordero, C.; Liberto, E.; Bicchi, C.; Conte, L. S. Toward a Definition of Blueprint of Virgin Olive Oil by Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2014**, *1334*, 101–111. https://doi.org/10.1016/j.chroma.2014.01.067.

(3) Romero, I.; García-González, D. L.; Aparicio-Ruiz, R.; Morales, M. T. Validation of SPME-GCMS Method for the Analysis of Virgin Olive Oil Volatiles Responsible for Sensory Defects. *Talanta* **2015**, *134*, 394–401. https://doi.org/10.1016/j.talanta.2014.11.032.

(4) Aparicio, R.; Morales, M. T.; Aparicio-Ruiz, R.; Tena, N.; García-González, D. L. Authenticity of Olive Oil: Mapping and Comparing Official Methods and Promising Alternatives. *Food Res. Int.* **2013**, *54* (2), 2025–2038. https://doi.org/10.1016/j.foodres.2013.07.039.

(5) Stilo, F.; Liberto, E.; Reichenbach, S. E.; Tao, Q.; Bicchi, C.; Cordero, C. Untargeted and Targeted Fingerprinting of Extra Virgin Olive Oil Volatiles by Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry: Challenges in Long-Term Studies. *J. Agric. Food Chem.* **2019**, *67* (18), 5289–5302. https://doi.org/10.1021/acs.jafc.9b01661.

(6) Commission of the European Communities. Commission Regulation (Eec) N° 2568/91. *Official Journal of the European Communities*. 1991, pp 1–83.

(7) International Oil Council. COI/T.20/DOC.15/Rev.10 SENSORY ANALYSIS OF OLIVE OIL - METHOD FOR THE ORGANOLEPTIC ASSESSMENT OF VIRGIN OLIVE OIL. 2018, p COI/T.20/DOC.15/Rev.10.

(8) Stilo, F.; Liberto, E.; Bicchi, C.; Reichenbach, S. E.; Cordero, C. GC×GC–TOF-MS and Comprehensive Fingerprinting of Volatiles in Food: Capturing the Signature of Quality. *LCGC Eur.* **2019**, *32* (5), 234–242.

(9) Vaz-Freire, L. T.; da Silva, M. D. R. G.; Freitas, A. M. C. Comprehensive Two-Dimensional Gas Chromatography for Fingerprint Pattern Recognition in Olive Oils Produced by Two Different Techniques in Portuguese Olive Varieties Galega Vulgar, Cobrançosa e Carrasquenha. *Anal. Chim. Acta* **2009**, *633* (2), 263–270. https://doi.org/10.1016/j.aca.2008.11.057.

(10) Cajka, T.; Riddellova, K.; Klimankova, E.; Cerna, M.; Pudil, F.; Hajslova, J. Traceability of Olive Oil Based on Volatiles Pattern and Multivariate Analysis. *Food Chem.* **2010**, *121* (1), 282–289. https://doi.org/10.1016/j.foodchem.2009.12.011.

(11) Lukić, I.; Carlin, S.; Horvat, I.; Vrhovsek, U. Combined Targeted and Untargeted Profiling of Volatile Aroma Compounds with Comprehensive Two-Dimensional Gas Chromatography for Differentiation of Virgin Olive Oils According to Variety and Geographical Origin. *Food Chem.* **2019**, *270* (March 2018), 403–414. https://doi.org/10.1016/j.foodchem.2018.07.133.

(12) Magagna, F.; Valverde-Som, L.; Ruíz-Samblás, C.; Cuadros-Rodríguez, L.; Reichenbach, S. E.; Bicchi, C.; Cordero, C. Combined Untargeted and Targeted Fingerprinting with Comprehensive Two-Dimensional Chromatography for Volatiles and Ripening Indicators

in Olive Oil. *Anal. Chim. Acta* **2016**, *936*, 245–258. https://doi.org/10.1016/j.aca.2016.07.005.

(13) Magagna, F.; Guglielmetti, A.; Liberto, E.; Reichenbach, S. E.; Allegrucci, E.; Gobino, G.; Bicchi, C.; Cordero, C. Comprehensive Chemical Fingerprinting of High-Quality Cocoa at Early Stages of Processing: Effectiveness of Combined Untargeted and Targeted Approaches for Classification and Discrimination. *J. Agric. Food Chem.* **2017**, *65* (30), 6329–6341. https://doi.org/10.1021/acs.jafc.7b02167.

(14) Magagna, F.; Liberto, E.; Reichenbach, S. E.; Tao, Q.; Carretta, A.; Cobelli, L.; Giardina, M.; Bicchi, C.; Cordero, C. Advanced Fingerprinting of High-Quality Cocoa: Challenges in Transferring Methods from Thermal to Differential-Flow Modulated Comprehensive Two Dimensional Gas Chromatography. *J. Chromatogr. A* **2018**, *1536*, 122–136. https://doi.org/10.1016/j.chroma.2017.07.014.

(15) Reichenbach, S. E.; Zini, C. A.; Nicolli, K. P.; Welke, J. E.; Cordero, C.; Tao, Q. Benchmarking Machine Learning Methods for Comprehensive Chemical Fingerprinting and Pattern Recognition. *J. Chromatogr. A* **2019**, *1595*, 158–167. https://doi.org/10.1016/j.chroma.2019.02.027.

(16) Bressanello, D.; Liberto, E.; Collino, M.; Chiazza, F.; Mastrocola, R.; Reichenbach, S. E.; Bicchi, C.; Cordero, C. Combined Untargeted and Targeted Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography: Revealing Fructose-Induced Changes in Mice Urinary Metabolic Signatures. *Anal. Bioanal. Chem.* **2018**, *410* (11), 2723–2737. https://doi.org/10.1007/s00216-018-0950-9.

(17) Bicchi, C.; Cordero, C.; Rubiolo, P. A Survey on High-Concentration-Capability Headspace Sampling Techniques in the Analysis of Flavors and Fragrances. *J. Chromatogr. Sci.* **2004**, *42* (8).

(18) Bicchi, C.; Cordero, C.; Liberto, E.; Rubiolo, P.; Sgorbini, B. Automated Headspace Solid-Phase Dynamic Extraction to Analyse the Volatile Fraction of Food Matrices. *J. Chromatogr. A* **2004**, *1024* (1–2), 217–226. https://doi.org/10.1016/j.chroma.2003.10.009.

(19) Risticevic, S.; Vuckovic, D.; Lord, H. L.; Pawliszyn, J. 2.21 – Solid-Phase Microextraction. In *Comprehensive Sampling and Sample Preparation*; 2012; pp 419–460. https://doi.org/10.1016/B978-0-12-381373-2.00055-7.

(20) Lord, H. L.; Pfannkoch, E. A. Sample Preparation Automation for GC Injection. In *Comprehensive Sampling and Sample Preparation*; Elsevier, 2012; Vol. 2, pp 597–612. https://doi.org/10.1016/B978-0-12-381373-2.00061-2.

(21) Ross, C. F. *2.02 – Headspace Analysis*; Elsevier, 2012; Vol. 2. https://doi.org/10.1016/B978-0-12-381373-2.10036-5.

(22) Bicchi, C.; Cordero, C.; Liberto, E.; Sgorbini, B.; Rubiolo, P. Headspace Sampling of the Volatile Fraction of Vegetable Matrices. *J. Chromatogr. A* **2008**, *1184* (1–2), 220–233. https://doi.org/10.1016/j.chroma.2007.06.019.

(23) Cordero, C.; Schmarr, H.-G.; Reichenbach, S. E.; Bicchi, C. Current Developments in Analyzing Food Volatiles by Multidimensional Gas Chromatographic Techniques. *J. Agric. Food Chem.* **2018**, *66* (10), 2226–2236. https://doi.org/10.1021/acs.jafc.6b04997.

(24) Cordero, C.; Kiefl, J.; Schieberle, P.; Reichenbach, S. E.; Bicchi, C. Comprehensive Two-Dimensional Gas Chromatography and Food Sensory Properties: Potential and Challenges.

Anal. Bioanal. Chem. **2015**, *407* (1), 169–191. https://doi.org/10.1007/s00216-014-8248-z.

(25) Oliver-Pozo, C.; Trypidis, D.; Aparicio, R.; Garciá-González, D. L.; Aparicio-Ruiz, R. Implementing Dynamic Headspace with SPME Sampling of Virgin Olive Oil Volatiles: Optimization, Quality Analytical Study, and Performance Testing. *J. Agric. Food Chem.* **2019**, *67* (7), 2086–2097. https://doi.org/10.1021/acs.jafc.9b00477.

(26) Chin, S. T.; Eyres, G. T.; Marriott, P. J. Cumulative Solid Phase Microextraction Sampling for Gas Chromatography-Olfactometry of Shiraz Wine. *J. Chromatogr. A* **2012**, *1255*, 221–227. https://doi.org/10.1016/j.chroma.2012.03.084.

(27) Bicchi, C.; Cordero, C.; Liberto, E.; Sgorbini, B.; David, F.; Sandra, P.; Rubiolo, P. Influence of Polydimethylsiloxane Outer Coating and Packing Material on Analyte Recovery in Dual-Phase Headspace Sorptive Extraction. *J. Chromatogr. A* **2007**, *1164* (1–2), 33–39. https://doi.org/10.1016/j.chroma.2007.07.026.

(28) Bicchi, C.; Cordero, C.; Liberto, E.; Rubiolo, P.; Sgorbini, B.; David, F.; Sandra, P. Dual-Phase Twisters: A New Approach to Headspace Sorptive Extraction and Stir Bar Sorptive Extraction. *J. Chromatogr. A* **2005**, *1094* (1–2), 9–16. https://doi.org/10.1016/j.chroma.2005.07.099.

(29) Giddings, J. C. Sample Dimensionality: A Predictor of Order-Disorder in Component Peak Distribution in Multidimensional Separation. *J. Chromatogr. A* **1995**, *703* (1–2), 3–15. https://doi.org/10.1016/0021-9673(95)00249-M.

(30) NIST/EPA/NIH Mass Spectral Library with Search Program Data Version: NIST V17.

(31) Adams, R. P. *Identification of Essential Oil Components by Gas Chromatography—Mass Spectroscopy*; Allured Publishing: New York, 1995.

(32) Feussner, I.; Wasternack, C. The Lipoxygenase Pathway. *Annu. Rev. Plant Biol.* **2002**, *53* (1), 275–297. https://doi.org/10.1146/annurev.arplant.53.100301.135248.

(33) Berlitz, H. D.; Grosch, W.; Schieberle, P. *Food Chemistry*; Springer, 2009.

(34) Angerosa, F.; Camera, L.; D'Alessandro, N.; Mellerio, G. Characterization of Seven New Hydrocarbon Compounds Present in the Aroma of Virgin Olive Oils. *J. Agric. Food Chem.* **1998**, *46* (2), 648–653. https://doi.org/10.1021/jf970352y.

(35) Reichenbach, S. E.; Tian, X.; Cordero, C.; Tao, Q. Features for Non-Targeted Cross-Sample Analysis with Comprehensive Two-Dimensional Chromatography. *J. Chromatogr. A* **2012**, *1226*, 140–148. https://doi.org/10.1016/j.chroma.2011.07.046.

(36) Reichenbach, S. E.; Carr, P. W.; Stoll, D. R.; Tao, Q. Smart Templates for Peak Pattern Matching with Comprehensive Two-Dimensional Liquid Chromatography. *J. Chromatogr. A* **2009**, *1216* (16), 3458–3466. https://doi.org/10.1016/j.chroma.2008.09.058.

(37) GC Image™. *GC Image GCxGC Edition Users' Guide*; 2017.

(38) Cordero, C.; Liberto, E.; Bicchi, C.; Rubiolo, P.; Reichenbach, S. E.; Tian, X.; Tao, Q. Targeted and Non-Targeted Approaches for Complex Natural Sample Profiling by GCxGC-QMS. *J. Chromatogr. Sci.* **2010**, *48* (4), 251–261.

(39) Reichenbach, S. E.; Tian, X.; Tao, Q.; Stoll, D. R.; Carr, P. W. Comprehensive Feature Analysis for Sample Classification with Comprehensive Two-Dimensional LC. *J. Sep. Sci.* **2010**, *33* (10), 1365–1374. https://doi.org/10.1002/jssc.200900859.

(40) Magagna, F.; Liberto, E.; Reichenbach, S. E.; Tao, Q.; Carretta, A.; Cobelli, L.; Giardina, M.; Bicchi, C.; Cordero, C. Advanced Fingerprinting of High-Quality Cocoa: Challenges in Transferring Methods from Thermal to Differential-Flow Modulated Comprehensive Two Dimensional Gas Chromatography. *J. Chromatogr. A* **2017**. https://doi.org/10.1016/j.chroma.2017.07.014.

(41) Reichenbach, S. E.; Tian, X.; Boateng, A. A.; Mullen, C. A.; Cordero, C.; Tao, Q. Reliable Peak Selection for Multisample Analysis with Comprehensive Two-Dimensional Chromatography. *Anal. Chem.* **2013**, *85* (10), 4974–4981. https://doi.org/10.1021/ac303773v.

(42) Rempe, D. W.; Reichenbach, S. E.; Tao, Q.; Cordero, C.; Rathbun, W. E.; Zini, C. A. Effectiveness of Global, Low-Degree Polynomial Transformations for GCxGC Data Alignment. *Anal. Chem.* **2016**, *88* (20), 10028–10035. https://doi.org/10.1021/acs.analchem.6b02254.

(43) Sgorbini, B.; Cagliero, C.; Liberto, E.; Rubiolo, P.; Bicchi, C.; Cordero, C. Strategies for Accurate Quantitation of Volatiles from Foods and Plant-Origin Materials: A Challenging Task. *J. Agric. Food Chem.* **2019**, acs.jafc.8b06601. https://doi.org/10.1021/acs.jafc.8b06601.

(44) Cordero, C.; Guglielmetti, A.; Sgorbini, B.; Bicchi, C.; Allegrucci, E.; Gobino, G.; Baroux, L.; Merle, P. Odorants Quantitation in High-Quality Cocoa by Multiple Headspace Solid Phase Micro-Extraction: Adoption of FID-Predicted Response Factors to Extend Method Capabilities and Information Potential. *Anal. Chim. Acta* **2019**, *1052*, 190–201. https://doi.org/10.1016/j.aca.2018.11.043.

(45) Kolb, B.; Ettre, L. S. *Static Headspace-Gas Chromatography : Theory and Practice*; Wiley-VCH: New York, 2006.

(46) Cordero, C.; Liberto, E.; Bicchi, C.; Rubiolo, P.; Schieberle, P.; Reichenbach, S. E.; Tao, Q. Profiling Food Volatiles by Comprehensive Two-Dimensional Gas Chromatography Coupled with Mass Spectrometry: Advanced Fingerprinting Approaches for Comparative Analysis of the Volatile Fraction of Roasted Hazelnuts (Corylus Avellana L.) from Different Ori. *J. Chromatogr. A* **2010**, *1217* (37).

(47) Brevard, H.; Cantergiani, E.; Cachet, T.; Chaintreau, A.; Demyttenaere, J.; French, L.; Gassenmeier, K.; Joulain, D.; Koenig, T.; Leijs, H.; Liddle, P.; Loesing, G.; Marchant, M.; Saito, K.; Scanlan, F.; Schippa, C.; Scotti, A.; Sekiya, F.; Sherlock, A. Guidelines for the Quantitative Gas Chromatography of Volatile Flavouring Substances, from the Working Group on Methods of Analysis of the International Organization of the Flavor Industry (IOFI). *Flavour Fragr. J.* **2011**, *26* (5), 297–299. https://doi.org/10.1002/ffj.2061.

(48) Vichi, S.; Pizzale, L.; Conte, L. S.; Buxaderas, S.; Lopez-Tamames, E. Solid Phase Microextraction in the Analysis of Virgin Olive Oil Volatile Fraction: Characterization of Virgin Oils from Two Distinct Geographical Areas of Northern Italy. *J. Agric. Food Chem.* **2003**, 6577.

(49) Cavalli, J. F.; Fernandez, X.; Lizzani-Cuvelier, L.; Loiseau, A. M. Comparison of Static Headspace, Headspace Solid Phase Microextraction, Headspace Sorptive Extraction, and Direct Thermal Desorption Techniques on Chemical Composition of French Olive Oils. *J. Agric. Food Chem.* **2003**, *51* (26), 7709–7716. https://doi.org/10.1021/jf034834n.

(50) Morales, M. T.; Luna, G.; Aparicio, R. Comparative Study of Virgin Olive Oil Sensory Defects. *Food Chem.* **2005**, *91* (2), 293–301.

https://doi.org/10.1016/j.foodchem.2004.06.011.

(51) Nigri, S.; Oumeddour, R.; Fernandez, X. Analysis of Some Algerian Virgin Olive Oils by Headspace Solid Phase Micro-Extraction Coupled to Gas Chromatography/Mass Spectrometry. *Riv. Ital. delle Sostanze Grasse* **2012**, *89* (1), 54–61. https://doi.org/10.1016/j.jhin.2016.11.012.

(52) Costa, R.; Albergamo, A.; Bua, G. D.; Saija, E.; Dugo, G. Determination of Flavor Constituents in Particular Types of Flour and Derived Pasta by Heart-Cutting Multidimensional Gas Chromatography Coupled with Mass Spectrometry and Multiple Headspace Solid-Phase Microextraction. *LWT - Food Sci. Technol.* **2017**, *86*, 99–107. https://doi.org/10.1016/j.lwt.2017.07.047.

(53) Cagliero, C.; Bicchi, C.; Cordero, C.; Rubiolo, P.; Sgorbini, B.; Liberto, E. Fast Headspace-Enantioselective GC-Mass Spectrometric-Multivariate Statistical Method for Routine Authentication of Flavoured Fruit Foods. *Food Chem.* **2012**, *132* (2), 1071–1079. https://doi.org/10.1016/j.foodchem.2011.10.106.

(54) Nicolotti, L.; Cordero, C.; Cagliero, C.; Liberto, E.; Sgorbini, B.; Rubiolo, P.; Bicchi, C. Quantitative Fingerprinting by Headspace-Two-Dimensional Comprehensive Gas Chromatography-Mass Spectrometry of Solid Matrices: Some Challenging Aspects of the Exhaustive Assessment of Food Volatiles. *Anal. Chim. Acta* **2013**, *798*, 115–125. https://doi.org/10.1016/j.aca.2013.08.052.

(55) Sgorbini, B.; Bicchi, C.; Cagliero, C.; Cordero, C.; Liberto, E.; Rubiolo, P. Herbs and Spices: Characterization and Quantitation of Biologically-Active Markers for Routine Quality Control by Multiple Headspace Solid-Phase Microextraction Combined with Separative or Non-Separative Analysis. *J. Chromatogr. A* **2015**, *1376*, 9–17. https://doi.org/10.1016/j.chroma.2014.12.007.

(56) Griglione, A.; Liberto, E.; Cordero, C.; Bressanello, D.; Cagliero, C.; Rubiolo, P.; Bicchi, C.; Sgorbini, B. High-Quality Italian Rice Cultivars: Chemical Indices of Ageing and Aroma Quality. *Food Chem.* **2015**, *172*, 305–313. https://doi.org/10.1016/j.foodchem.2014.09.082.

(57) Kremser, A.; Jochmann, M. A.; Schmidt, T. C. PAL SPME Arrow - Evaluation of a Novel Solid-Phase Microextraction Device for Freely Dissolved PAHs in Water. *Anal. Bioanal. Chem.* **2016**, *408* (3), 943–952. https://doi.org/10.1007/s00216-015-9187-z.

(58) Helin, A.; R??nkk??, T.; Parshintsev, J.; Hartonen, K.; Schilling, B.; L??ubli, T.; Riekkola, M. L. Solid Phase Microextraction Arrow for the Sampling of Volatile Amines in Wastewater and Atmosphere. *J. Chromatogr. A* **2015**, *1426*, 56–63. https://doi.org/10.1016/j.chroma.2015.11.061.

5.2 Chromatographic fingerprinting by template matching for data collected by comprehensive two-dimensional gas chromatography

Federico Stilo[1], Chiara Cordero[1]*, Carlo Bicchi[1], Daniela Peroni[2], Qingping Tao[3] and Stephen E Reichenbach[3,4*]

[1]Dipartimento di Scienza e Tecnologia del Farmaco, Università degli Studi di Torino, Turin, Italy

[2]SRA Instruments, Cernusco sul Naviglio, Milan, Italy

[3]GC Image LLC, Lincoln, NE, USA

[4]Computer Science and Engineering Department, University of Nebraska, Lincoln, NE, USA

*Co-Corresponding authors:

Prof. Chiara Cordero - Dipartimento di Scienza e Tecnologia del Farmaco, Università di Torino, Via Pietro Giuria 9, I-10125 Torino, Italy – e-mail: chiara.cordero@unito.it;

Prof. Stephen E Reichenbach - Computer Science and Engineering Department, University of Nebraska, Lincoln, 260 Avery Hall, Lincoln, NE, 68588-0115, USA – e-mail: reich@cse.unl.edu

Published: September 2, 2020

DOI: 10.3791/61529

J. of Visual Experiment. 2020, 163, 1-20

5.2.1 Abstract

Data processing and evaluation are critical steps of comprehensive two-dimensional gas chromatography, particularly when coupled to mass spectrometry. The rich information encrypted in data may be highly valuable but difficult to access efficiently. Data density and complexity can lead to long elaboration times and require laborious, analyst-dependent procedures. Effective yet accessible data processing tools, therefore, are key to enabling the spread and acceptance of this advanced multidimensional technique in laboratories for the daily use. The data analysis protocol presented in this work uses chromatographic fingerprinting and template matching to achieve the goal of highly automated deconstruction of complex two-dimensional chromatograms into individual chemical features for advanced recognition of informative patterns within individual chromatograms and across sets of chromatograms. The protocol delivers high consistency and reliability with little intervention. At the same time, analyst supervision is possible in a variety of settings and constraint functions that can be customized to provide flexibility and capacity to adapt to different needs and goals. Template matching is shown here to be a powerful approach to explore extra virgin olive oil volatilome. Cross-alignment of peaks is performed not only for known targets, but also for untargeted compounds, which significantly increases the characterization power for a wide range of applications. Examples are presented to evidence the performance for the classification and comparison of chromatographic patterns from sample sets analyzed under similar conditions.

Key words

comprehensive two-dimensional gas chromatography, 2D patterns of analytes, template matching, combined untargeted and targeted fingerprinting, 2D chromatographic misalignment, template transform, cross-comparative analysis, foodomics

5.2.2 Introduction

Comprehensive two-dimensional gas chromatography combined with the time-of-flight mass spectrometric detection (GC×GC-TOF MS) is nowadays the most informative analytical approach for the chemical characterization of complex samples.[1–5] In GC×GC columns are serially connected and interfaced by a modulator, *e.g.,* a thermal or valve-based focusing interface, that traps eluting components from the first dimension (^{1}D) column before their re-injection into the second-dimension (^{2}D) column. This operation is done within a fixed modulation time-period (P_M), generally ranging between 0.5 to 8 s. By thermal modulation, the process includes cryo-trapping and focusing of the eluting band with some benefits for the overall separation power.

Although GC×GC is a two-dimensional separation technique, the process produces sequential data values. The detector analog-to-digital (A/D) converter collects the chromatographic signal output at a certain frequency then the digitalized data, together with related metadata (information about the data), are stored in a file with specific proprietary formats. GC×GC systems employ an A/D converter to map the intensity of the chromatographic signal to a digital number (DN) as a function of time in the two analytical dimensions. Single-channel detectors (e.g., flame ionization detector - FID, electron capture detector - ECD, sulphur chemiluminescence detector SCD, etc.) produce single values per sampling time, whereas multichannel detectors (e.g., mass spectrometric detector MS) produce multiple values (typically, over a spectral range) per sampling time along the analytical run.

To visualize 2D data, elaboration starts with rasterization of single modulation period (or cycle) data values as a column of pixels (picture elements corresponding to detector events). Along the ordinate (Y-axis, bottom-to-top) is visualized the ^{2}D separation time. Pixel columns are sequentially processed so that the abscissa (X-axis, left-to-right) reports ^{1}D separation time. This ordering presents the 2D data in a right-handed Cartesian coordinate system, with the ^{1}D retention ordinal as the first index into the array.

Data processing of 2D chromatograms gives access to a higher level of information than raw data, enabling 2D peak detection, peak identification, extraction of response data for quantitative analysis, and cross-comparative analysis.

The 2D peak patterns can be treated as sample's unique fingerprint and detected compounds as minutiae features for effective cross-comparative analysis. This approach, known as template-based fingerprinting,[6,7] was inspired by biometric fingerprinting.[6] Automatic biometric fingerprint verification systems, in fact, rely on unique fingertip characteristics, *e.g.,* ridge bifurcations and endings, localized and extracted from inked impressions or detailed images. These characteristics, named minutiae features, are then cross-matched with available stored templates.[8,9]

As mentioned above, every GC×GC separation pattern is composed of 2D peaks rationally distributed over a two-dimensional plane. Each peak corresponds to a single analyte, has its informative potential, and can be treated as a single feature for comparative pattern analysis.

In this contribution, we present an effective approach for chemical fingerprinting by GC×GC-TOF MS featuring tandem ionization. The goal is to catalog features comprehensively and quantitatively from a set of chromatograms.

Compared to existing commercial software or in-house routines[10,11] that employ a peak-features approach, template-based fingerprinting is characterized by high specificity, efficiency, and limited computational time. In addition, it has an intrinsic flexibility that enables the cross-alignment of minutia features (*i.e.*, 2D peaks) between severely misaligned chromatograms as those acquired by different instrumentation or in long-time frame studies.[12–14]

The basic operations of the proposed method are described briefly to guide the reader to a good understanding of the 2D pattern complexity and information power. Then, by exploring the instrument output data matrix, chemical identification is performed and known targeted analytes located over the two-dimensional space. The template of targeted peaks is then built and applied to a series of chromatograms acquired within the same analytical batch. Metadata related to retention times, spectral signatures, and responses (absolute and relative) are extracted from re-aligned patterns of targeted peaks and adopted to reveal compositional differences in the sample set.

As an additional, unique step of the processing, combined untargeted and targeted (UT) fingerprinting is also performed on pre-targeted chromatograms to extend the fingerprinting potential to both known and unknown analytes. The process produces a UT template for a truly comprehensive comparative analysis that can be largely automated.

As a final step, the method performs the cross-alignment of features in two parallel detector signals produced with high and low electron ionization energies (70 and 12 eV).

The protocol is quite flexible in supporting analyses of a single chromatogram or a set of chromatograms and with variable chromatography and/or multiple detectors. Here, the protocol is demonstrated with a commercially available GC×GC Software suite (see **Table of materials**) combined to a MS library and search software (see **Table of Materials**). Some of the necessary tools are available in other software and similar tools could be implemented independently from descriptions in the literature by Reichenbach and co-workers.[15–19] Raw data for the demonstration derive from a research study on extra-virgin olive (EVO) oil from Italy conducted in authors' laboratory. In particular, the volatile fraction (*i.e.,* volatilome) of Italian EVO oils is sampled by headspace solid phase microextraction (HS-SPME) and analyzed by GC×GC-TOF MS to capture diagnostic fingerprints for quality and sensory qualification of samples. Details on samples, sampling conditions and analytical set-up are provided in the **Table of Materials.** Contact the authors to obtain data for demonstration and testing of the protocol.

Steps 1-6 describe preprocessing of the chromatograms. Steps 7-9 describe processing and analysis of individual chromatograms. Steps 10-12 describe template creation and matching, which are the basis for cross-sample analysis. Steps 13-16 describe applying the protocol across a set of chromatograms, with Steps 14-16 for UT analysis.

5.2.3 Protocol

1. Import raw data

NOTE: This creates a two-dimensional raster array for visualization and processing.

1.1 Launch the Image software.

1.2 Select **File | Import**; navigate to and choose the raw data file "VIOLIN 101.lsc"; then click "**Open**". The chromatogram opens in this software.

In the Import dialog, set the Modulation Period (P_M) to "**3.5**" s; then click "**OK**".

NOTE: Some acquisition software may not record the modulation period.

Select **File | Save Image As…**; navigate to the desired folder; enter the name "**Oil 1 RAW.gci**"; then click "**SAVE**".

NOTE: This file is included in the supplemental archive, which can be opened for Step 2.

2. Shift modulation phase

NOTE: This puts all peaks in each modulation cycle into the same image column, including those peaks that wrap around the end of the modulation period into the void time of the next modulation period.[20]

2.1 Select Processing | Shift Phase.

2.2 In the Shift Phase dialog, set the Shift Amount to "**-0.8**" s; then click "**OK**".

3. Correct baseline[21]

3.1 Select **Graphic | Draw Rectangle**.

3.2 Click-and-drag to draw a rectangle in the image where no peaks are detected.

3.3 Select **Tools | Visualize Data**; note the mean and standard deviation of the detector signal, here 21,850 ± 1,455 SD unit-less digital number (DN); then, close the tool.

3.4 Select **Processing | Correct Baseline**.

4. Colorize the chromatographic image using a value map and color map[20]

4.1 Select **View | Colorize**.

4.2 In the Colorize dialog, select the **Import/Export** tab; choose the "**#AAAA**" custom color map; then click "**Import**".

4.3 On the Value Mapping controls, set the value range to the minimum and maximum values; then click "**OK**".

5. Detect 2D peaks (*i.e.,* blobs) for analytes[18]

5.1 Select **Processing | Detect Blobs** with the default settings; then, observe that some peaks are split and there are spurious detections.

5.2 Select **Configure | Settings | Blob Detection**; then set Smoothing to 0.1 ^{1}D samples and 2.0 ^{2}D modulations and set Minimum Volume (*i.e.*, threshold for the summed values) to 1.00 E6; then click "**OK**".

5.3 Select **Processing | Detect Blobs** with the new settings; then, observe the improvements.

6. Filter 2D peaks

NOTE: This is done to automatically remove meaningless detections, *e.g.*, due to column bleeds along the ^{1}D and strikes or tailings along the ^{2}D.

6.1 Select **Processing | Interactive Blob Detection**.

6.2 Note the blob detection settings; then, click "**Detect**".

6.3 In the Advanced Filter builder, click "**Add**"; then, in the New Constraint dialog, select "**Retention II**"; then, click "**OK**".

6.4 With the Constraint sliders, set the minimum and maximum ^{2}D retention times for the filter so as to reduce the number of false peaks without losing true peaks.

6.5 Click "**Apply**"; then click "**Yes**" to save to the detection settings with the new filter.

NOTE: More advanced tools may be required to deal with particular detection problems, *e.g.,* ion-peak detection or deconvolution for co-elutions.[22]

7. Calibrate linear retention indices

NOTE: Perform this step [23] (I^T) for the specific retention times across the set of retention index (RI) standards (typically *n*-alkanes).

7.1 Select **Configure | RI Table | Retention Index (Col I)**.

7.2 On the RI Table Configuration dialog, click "**Import**"; then, select the RI calibration file (in CSV format with name, retention time, and retention index).

7.3 Select **File | Save Image As…**; navigate to the desired folder; enter the name "Oil 1 LRI CALIBRATED.gci"; then click "SAVE".

NOTE: This file is included in the supplemental archive, which can be opened for Step 8.

8. Search for peak spectra in the NIST17 MS Library[24]

8.1 Select **Configure | Settings | Search Library**.

8.2 In the Search Library dialog, set Type of Spectrum to "**Peak MS**", Intensity Threshold to "100", NIST Search Type to "**Simple (Similarity)**", NIST RI Column Type to "**Standard Polar**", and NIST RI Tolerance to "**10**"; then, click "**OK**". NIST MS Search offers many other settings that here are set to the defaults.

8.3 Select **Processing | Search Library for All Blobs.**

9. Review and correct analyte identifications

9.1 On the tool palette, set the cursor mode to **Blob | Select Blobs**.

9.2 In the Image view, right-click on the desired peak.

9.3 On the Blob **Properties** dialog, inspect blob properties; then, click "**Hit List**".

9.4 Inspect the hit list; then, if the identification is incorrect, select the checkmark beside the correct identification.

9.5 In the Blob Properties dialog, enter the Group Name to designate chemical class and any other desired metadata; then, click "**OK**".

9.6 Select **File | Save Image As…**; navigate to the desired folder; enter the name "Oil 1 COLORIZED for Template construction.gci"; then click "SAVE".

NOTE: This file is included in the supplemental archive, which can be opened for Step 10.

10. Create a template with targeted peaks[15]

10.1 In the Image view (still in Select Blobs mode from Step 9.1), select desired peaks with a click on the first peak and **CTL + click** on additional peaks.

10.2 On the tool palette, click the "**Add to Template**" button.

10.3 When the template is complete, select **File | Save Template**, specify the folder and file name; then, click "**Save**".

10.4 Select **File | Close Image**.

NOTE: At this point, these instructions continue with the template created in this fashion by the authors to include the desired target peaks, available in the supplemental archive as "Targeted tamplate.bt".

11. Match and apply the template

Note: Matching recognizes the template pattern in the detected peaks a new chromatogram. Applying the matching sets identifications and other metadata in the new chromatogram from the template.

11.1 Select **File | Open Image**; navigate to and select the "**Oil 2 COLORIZED.gci**" chromatogram file (which already is preprocessed); then, click "**Open**".

11.2 On the tool palette, set the cursor mode to **Template | Select Objects**.

11.3 Select **Template | Load Template**.

11.4 In the Load Template dialog, click "**Browse**"; navigate to and select the targeted peaks template "**Targeted template.bt**"; then, click "Open".

11.5 In the Load Template dialog, click "**Load**", then "**Dismiss**".

11.6 In the Image view, right-click on a template peak; then, inspect its object properties, including the qCLIC and reference MS.

11.7 Select **Template | Interactive Match Template**.

11.8 In the Interactive Match interface, click "**Match All**"; then, review the matching results both in the table and in the image, in which each template peak is marked with unfilled circles and, if a match is made, there is a link to a filled circle for the detected peak.

11.9 Edit the matches as desired; when satisfied, click "**Apply**" to transfer metadata from the template to the chromatogram.

NOTE: Matching constraints, such as the qCLIC, help match the correct pattern among the detected peaks of the new chromatogram. Constraint parameters include the type of MS signature used as template reference (*peak MS* or *blob MS*) and the threshold values for spectral similarity (DMF and Reverse Match Factor, RMF). Here, parameters are set based on previous studies[13,14] to limit false negative matches: *peak MS* and DMF and RMF similarity threshold 700.

12. Transform the template for substantially different chromatography

NOTE: This step is not necessary unless chromatographic conditions vary substantially causing the template to be misaligned with a new chromatogram, such as can be the case over long-term studies or after a new column is installed. In such cases, the template can be geometrically transformed in the chromatographic retention time plane to better fit the new chromatogram.[12,13] In this example, the peak patterns of the template and chromatogram are similar, but differ in the retention time geometry, such as would be seen for different chromatographic conditions.

12.1 Repeat Steps 11.2-11.5, except navigate to, select, and load "**Targeted template 2.bt**".

12.2 Select **Template | Interactive Match Template**; then, click "**Edit Transform**".

12.3 In the Transform Template interface, vary the ^{1}D and ^{2}D scales, translations, and shears to better align the template with the detected peaks; then, click "**Transform Template**".

12.4 With the transformed template, click "**Edit Match**"; then, repeat Steps 11.8 -11.9.

13. Perform targeted analysis across a set of chromatograms

Note: The targeted template is matched to each of a set of chromatograms to establish correspondences between targeted analytes, then consistent cross-sample features are extracted for pattern recognition.

13.1 Perform preprocessing (Steps 1-6) and template matching (Steps 11.1-11.9) for all chromatograms in the set (*i.e.*, targeted chromatograms of oils). (Alternatively, automate this step with Project software or similar software, not described here.)

13.2 Launch the Investigator software.

13.3 Select **File | Load Images…**; then, select and open the image files saved at Step 13.1.

13.4 In the Load Options dialog, on the Features tab, select "As is"; on the Attributes tab, select specific attributes: retention times, volume, percent response, and volume ratio; and, on the Class Assignment tab, assign the chromatograms to classes "Tuscany" or "Sicily".

13.5 Click "OK" to process the chromatograms, which, after processing is complete, can be opened and examined if desired.

13.6 Click on the Compounds tab to review metric values and statistics for specific analytes across chromatograms, then:

13.6.1 Click on the Attributes tab to review values and statistics for specific metrics across chromatograms.

13.6.2 Click on the Summary tab to review the summary statistics for both compounds and features. If the chromatograms are from different classes, as in this case oils produced from olives harvested in two different Regions of Italy, then the Summary tab lists Fisher ratio statistics (F and FDR) which provide insights into features for discriminating between classes.

13.6.3 View various charts on all tabs and, if desired, perform Principal Component Analysis (PCA) on the Attributes tab.

14. Construct an *U*ntargeted and *T*argeted (*UT*) template

14.1 Launch the Investigator software.

14.2 Select **File | Load Images…**; then, select and open the image files saved at Step

14.3 In the Load Options dialog, check "Use configuration" and browse to and select "OIL CONFIGURATION.cfg"; on the Feature tab, select "Auto generate a feature template", specify an Output Folder and output File Name, set "Preferred Reliable Peaks" to "Most Relaxed and Composite Image Normalization to "Total Image Volume"; then, click "OK" to begin processing, which may require from minutes to hours depending on the number and complexity of the chromatograms to produce a template composited from the detected UT peaks in all chromatograms, here named "UT template 70 relaxed.bt".

14.4 In the Image software, select **File | Open Image…**; in the Open dialog, navigate to and select "OIL 2 COLORIZED.gci", then click "Open".

14.5 On the tool palette, set the cursor mode to **Template | Select Objects**; then, select **Template | Load Template**.

14.6 In the Load Template dialog, click "Browse"; navigate to and select the UT template "UT template 70 relaxed.bt"; then, click "Load", then "Dismiss". This is a hybrid template of targeted peaks (generated at Step 10) and untargeted reliable peaks (generated at Step 14.3).

14.7 As in Step 11.6, in the Image view, right-click on a template peak; then, inspect its object properties, including the qCLIC and reference MS, and, if desired, edit the template object.

15. Perform UT analysis across a set of chromatograms

NOTE: This is the same as Step 13, but with the UT template instead of the target template.

15.1 Perform preprocessing (Steps 1-6).

15.2 In the Investigator software, select **File -> Load Images…**; then, select and open the image files saved at Step 13.1.

15.3 In the Load Options dialog, on the Features tab, select "Apply an existing feature template"; click "Browse"; then, navigate to and select the UT template "UT template 70 relaxed.bt". As in Step 13.4, on the Attributes tab, select specific attributes: retention times, volume, percent response, and volume ratio; and, on the Class Assignment tab, assign the chromatograms to classes Tuscany or Sicily. Finally, click "OK" to analyze the chromatograms.

15.4 Examine the analysis as in Step 13.6.

15.5 Select **File | Save Analysis**. Here, the analysis has been saved as "Feature Jove su 70 eV.gca".

16. Modify the UT template for parallel MS analysis

Note: The analysis was performed with both 70 eV and 12 eV (*i.e.*, high and low) electron ionization energies.[25,26]

16.1 Perform preprocessing (Steps 1-6) on each 12 eV chromatogram.

16.2 Open one of the 12 eV chromatograms*, e.g.,* "Oil 2 12 eV.gci" and load the UT template "UT template 70 relaxed.bt" as described in Steps 11.1-11.6.

16.3 If necessary, adjust the template to fit the detected 12 eV peaks as described in Step 12. Here, there is no significant misalignment because the tandem signals are multiplexed. However, it should be noted that because the different ionization settings produce different fragmentations, it is necessary to relax constraints for the qCLIC constraints on DMF and RMF spectral similarity (not demonstrated here).

16.4 Select **File | Save Template**; specify the folder and file name, e.g., "UT template 12.bt"; then, click "Save".

16.5 Apply the 12 eV UT template to the 12 eV chromatograms as described in Step 15 for the 70 eV UT template and the 70 eV chromatograms.

16.6 Examine the cross-chromatographic analysis of the 12 eV data, similarly to Steps 13.6 and 15.4.

16.7 Select **File | Save Analysis**. Here, the analysis has been saved as "Feature Jove su 12 eV.gca".

5.2.4 Results

GC×GC-TOF MS patterns of high-quality extra-virgin olive oil volatilome exhibit about 500 2D peaks above a signal-to-noise ratio (SNR) threshold of 100. Such a threshold was defined by previous investigations on food volatiles (Cordero, Guglielmetti, et al., 2019; Stilo et al., 2019) as the minimum relative signal over threshold to obtain reliable spectra for cross-comparative analysis. Components are distributed over the chromatographic space according to their relative retention in the two chromatographic dimensions, and specifically based on their volatility/polarity in the 1D and volatility in the 2D. Here, column combination is polar × semi-polar (i.e., Carbowax 20M × OV1701).

The 2D pattern shows a high degree of order. Relative retention patterns for homologous series and classes are shown in **Figure 5.2.1A** with annotations (graphics for groups and bubbles for peaks) for linear saturated hydrocarbons (black), unsaturated hydrocarbons (yellow), linear saturated aldehydes (blue), mono-unsaturated aldehydes (red), polyunsaturated aldehydes (salmon), primary alcohols (green), and short-chain fatty acids (cyano).

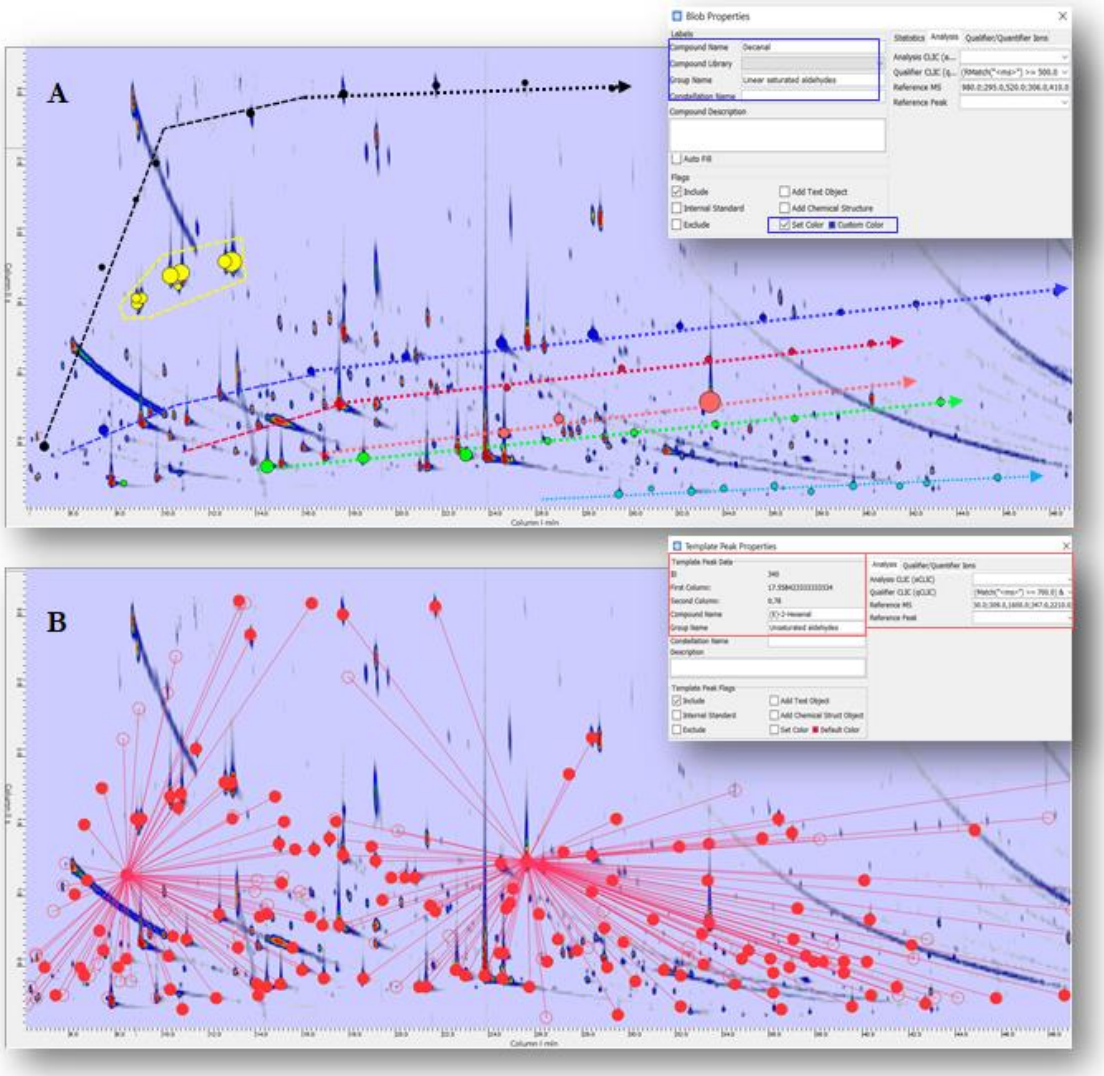

Figure 5.2.1. Bidimensional contour plot and targeted template. (**5.2.1A**) Contour plot of the volatile fraction of an extra virgin olive oil from Tuscany. Ordered patterns of homolog series and classes are highlighted with different colors and lines: linear saturated hydrocarbons - black line and 2D contours, unsaturated hydrocarbons - yellow, linear saturated aldehydes - blue, mono-unsaturated aldehydes - red, polyunsaturated aldehydes -salmon, primary alcohols - green and short-chain fatty acids – cyano. (**5.2.1B**) Overimposed targeted template of known analytes (red colored circles) with connection lines linking Internal Standards (ISs). Panels show 2D peak/blob properties metadata (Decanal) or Template peak properties.

Detected 2D peaks then can be identified by comparing the average MS spectrum extracted from the entire 2D peak (*blob* spectrum) or from the largest spectrum (*apex* spectrum). **Figure 5.2.2** illustrates the output of the *apex* spectrum search for the "blob 5" and returns a high similarity match (first 10 hits) for (*E*)-2-hexenal. Databases explored are those pre-selected by the analyst at Step 8 of the method.

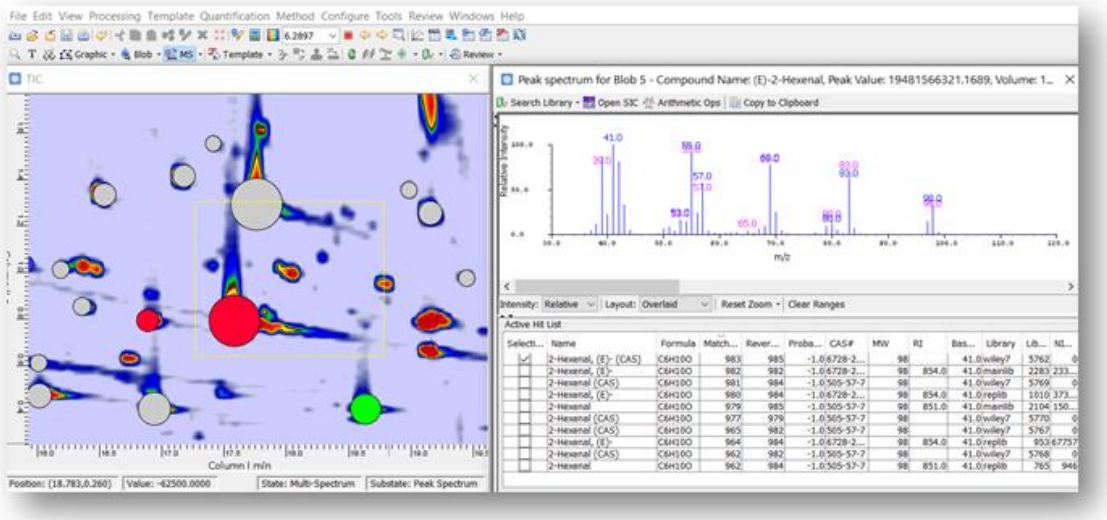

Figure 5.2.2. Apex MS search. Output of the apex MS search for the "blob 5". List of the database entries with the highest similarity match and related metadata available from the library.

The identification is validated by active retention indexing. The experimental I^T value was calculated for the 2D peaks, so that at this stage the library search prioritizes results with coherent values of tabulated I^T. Tolerance windows can be customized based on analyst experience, reliability of reference database values according to stationary phase and analytical conditions applied. New tools for smart calibration of linear retention indices without experimental calibration with n-alkanes, have been recently developed and discussed in a study by Reichenbach *et al.*[22]

The collection of identified 2D peaks, *i.e.*, targeted peaks, can be adopted to build a template of targeted peaks to promptly establish reliable correspondences between the same compound across all sample chromatograms. The collection of targeted template peaks is visualized in **Figure 5.2.1B**. Red circles correspond to the 196 targeted compounds, including two Internal Standards (IS) linked to template peaks with connection lines. IS are used for response normalization and connection lines help to visualize which of the included IS will be adopted to normalize each 2D peak/blob response.

In **Figure 5.2.1B**, filled circles indicate positive matches between template peak and the actual pattern while empty circles are for template peaks for which the correspondence was not verified. False negative matches can be limited by appropriate selection of threshold parameters, reference spectra and constraint functions.[13,14,18,19] For complex patterns with multiple coelutions, ion peak detection functions, based on spectral deconvolution, are advisable and could be a valid option [19]. Template peak metadata are shown in the enlarged panel of **Figure 5.2.2** for (*E*)-2-hexenal.

The specificity of template matching relies on the possibility to apply constraint functions that limit positive correspondence to those candidate peaks that, falling within the search window of the algorithm, have MS spectral similarity above a certain threshold. In this case, at Step 11, similarity thresholds[24] were set at 700 according to previous experiments aimed at defining optimal parameters limiting false negative matches.[14] Highlighted areas of the template peak properties in **Figure 5.2.2** show the information about the reference MS spectrum string and the qCLIC constraint function (*i.e.*, (Match("<ms>") >= 700.0) & (RMatch("<ms>") >= 700.0)).

By applying the template to all chromatograms of a set, one could encounter challenging situations as in the case of partial misalignment of patterns. This can be due to oven temperature

inconsistencies, carrier gas flow/pressure instabilities, or because of a manual intervention on the system as in the case of column substitution or modulator loop-capillary replacement.[14,27] **Figure 5.2.3** shows a situation of a partial misalignment between the targeted template and the actual chromatogram. For minimal misalignments, an interactive template transforms (**Figure 5.2.3**, control panel) can reposition of template peaks for a better fit. Once repositioned, the template can be matched to establish correspondences. In the example, the template (**Figure 5.2.3**, Step 12) peaks correctly match with the actual 2D pattern. In case of severe misalignments, not discussed here, the repetition of match-transform-update actions can iteratively adapt the template peaks position to the actual peak pattern.[12–14]

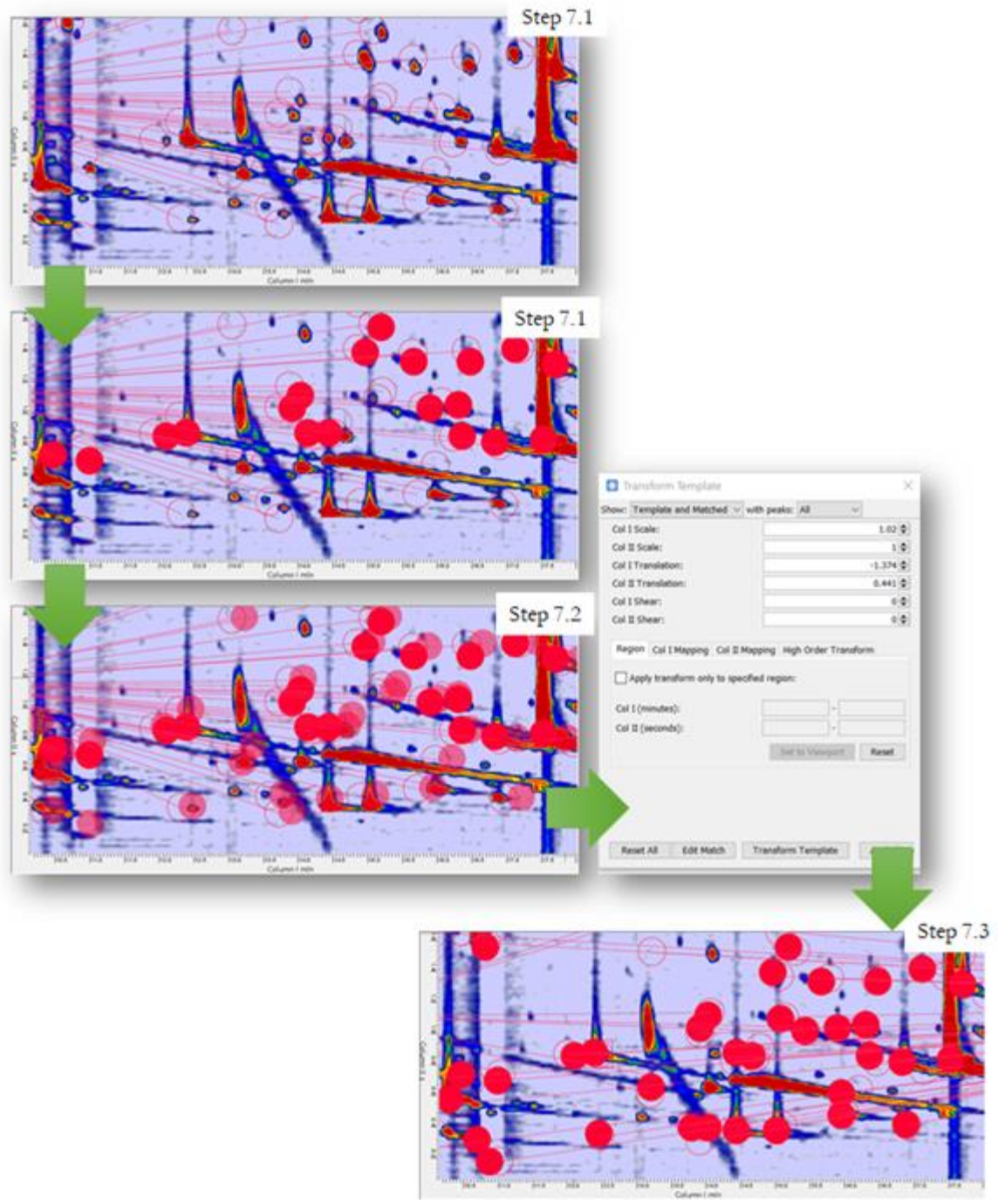

Figure 5.2.2. Template realignment. Workflow illustrating the steps that allow re-alignment of the template by transformation.

Here, the targeted peaks, *i.e.*, known analytes, provide about 40% of the chromatographic result (196 targeted peaks of about 500 detectable peaks on average). The other 60% of compounds, together with the information they bring, are not taken into consideration in targeted analysis. To make the investigation truly comprehensive, consistent cross-alignment of untargeted 2D peaks also should be established. The first application where template matching was extended to all detectable analytes dealt with the complex volatilome of roasted coffee.[7] This process is automated in Investigator, here shown at Steps 14-15.

In this process, pre-targeted images belonging to the sample set under study (20 samples) are used to define reliable peaks by cross matching of all image patterns.[28] Subsequently, a composite chromatogram is built from which to identify UT reliable peaks and peak-regions (*i.e.*, 2D peaks footprint) in the so-called feature template.[17]

For analyses acquired at 70 eV, the process determined 144 reliable peaks with relaxed reliability,[28] 76 of which belong to the targeted peaks list. Based on these 144 reliable peaks, the process aligns all chromatograms consistently with the average retention times of the reliable peaks and then combines them to create a composite chromatogram. **Figure 5.2.4** shows a list of all samples labelled according to the production Region of the oil (left) and the list of reliable peaks/blob volumes in each sample (right).

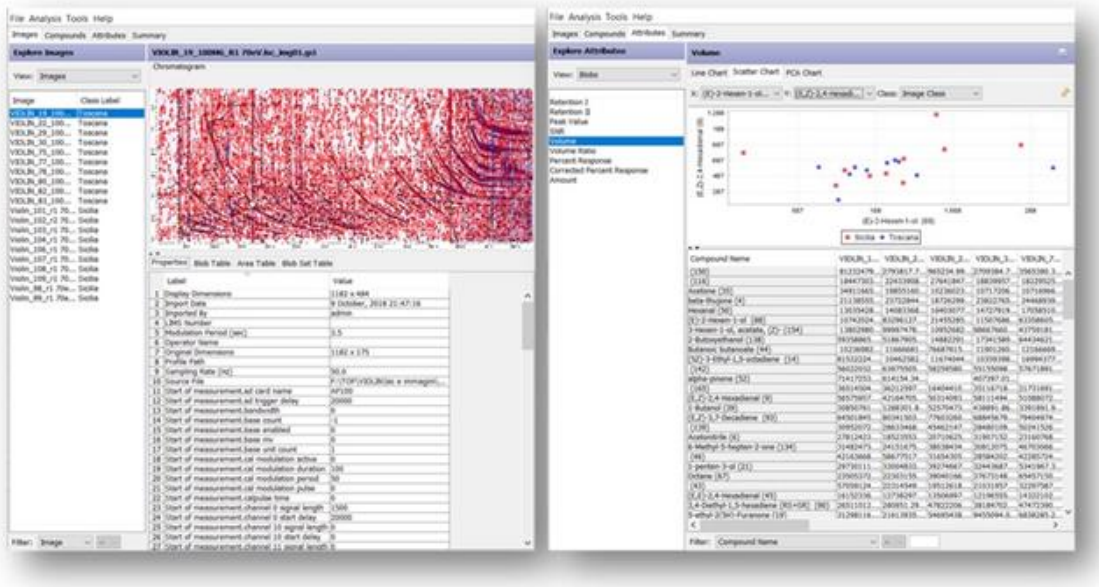

Figure 5.2.3. GC Investigator interface. Investigator panel with all selected images labelled according to the production Region of the oil (left) and the list of reliable peaks/blob volumes in each sample (right).

The untargeted feature template is composed of 2D peaks from analytes detected in the composite chromatogram, shown in **Figure 5.2.5A**, that are matched by the reliable-peaks template (n=168 – red circles for targeted peaks and green circles for untargeted peaks). The mass spectra of the composite peaks, as well as their retention times, are recorded in the feature template as shown for (*Z*)-3-hexenol acetate in the enlarged area. Peak-regions are shown in **Figure 5.2.5B** as red colored graphics; they are instead defined by the outlines of all 2D peaks detected in the composite chromatogram (n=3578).

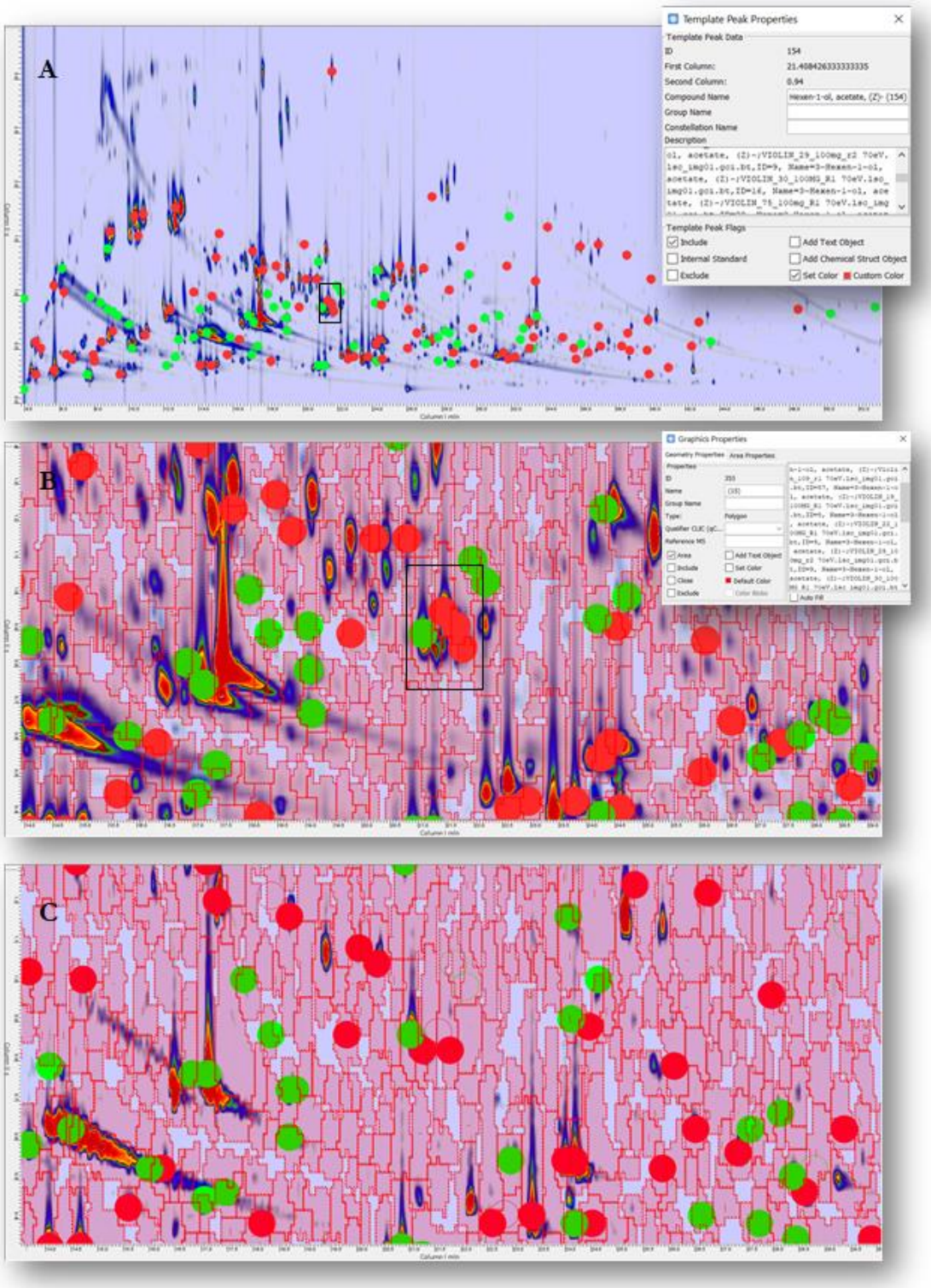

Figure 4.2.5. Targeted and UT template. (**5.2.5A**) Reliable peaks as resulting from the automated processing at Step 11; red circles correspond to known analytes while green circles are unknowns. In the superimposed panel, template object properties are shown for the (Z)-3-hexenal. (**5.2.5B**) Enlarged area that shows the UT peaks (red and green circles) and peak-regions (red graphics) of the UT template matched on a sample oil acquired at 70 eV ionization energy. (**5.3.5C**) UT template matched on a sample oil acquired at 12 eV ionization energy.

When unsupervised pattern recognition by Principal Component Analysis is applied to targeted peaks distribution within the 20 analyzed samples, Sicilian and Tuscany oils cluster separately suggesting that pedo-climatic conditions and terroir impact the relative prevalence of volatiles. Results are shown in **Figure 5.2.6A** and the PCA results from the reliable peaks distribution are shown in **Figure 5.2.6B**. The two approaches cross-validate that oils from different geographical areas have different, while coherent, chemical signatures whether targeted or untargeted compounds, or both, are mapped.

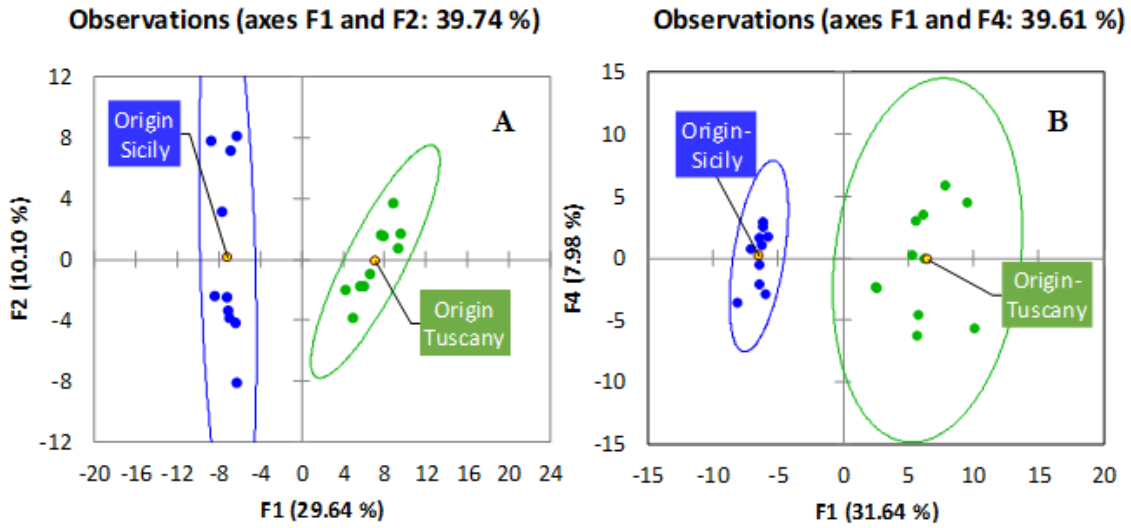

Figure 5.2.6. They show the natural conformation of samples (oils from Tuscany and Sicily) as they result by (**5.2.6A**) targeted peaks distribution or (**5.2.6B**) UT peaks distribution.

Finally, the software enables prompt and effective re-alignment of patterns across parallel detection channels. In this application, the re-alignment is proposed for tandem ionization signals. The ion source of the MS multiplexes between two ionization energies (*i.e.*, 70 and 12 eV) at an acquisition frequency of 50 Hz per channel.[29] The two resulting chromatographic patterns are closely aligned while spectral data (*i.e.*, spectral signatures and responses) and they bring complementary information with different dynamic ranges.[25,26] The aligned patterns allow extracting features (2D peaks and peak-regions) with univocal IDs (*i.e.*, chemical names for targeted peaks and unique numbering # for untargeted peaks and peak-regions).

Template matching allows effective cross-alignment. In this situation, there is not much misalignment, but MS constraints must be relaxed to allow matches for UT peaks. On the other hand, featured UT peak-regions, that have no MS constraints, are promptly matched without any false negative matches. **Figure 5.2.5C** shows an enlarged area of a 12 eV chromatogram where the feature template built from 70 eV data is matched. Reliable UT peaks are positively matched because of the lowered qCLIC constraints (*e.g.*, DMF threshold at 600). To note, at 12 eV, there are fewer detected peaks due to the limited fragmentation induced by low ionization energy.

5.2.5 Conclusions

Visualization of GC×GC-TOF MS data is a fundamental step for an appropriate understanding of the results achieved by comprehensive two-dimensional separations. Image plots with customized colorization allow analysts to appreciate detector response differences and so the differential distribution of sample components. This visual approach changes completely the analyst

perspective on chromatograms interpretation and elaboration. This first step, once understood and confidently used by chromatographers, opens a new perspective in further processing.

Another fundamental aspect of data processing is the accessibility to the full data matrix (*i.e.*, MS spectral data and responses) for all sample points, each of which corresponds to a single detector event. In this respect 2D peaks integration, so that the collection of detector events corresponding to a single analyte represent a critical step. In the current protocol. 2D peaks detection is based on the watershed algorithm[18] with some adaptations included to improve detection sensitivity in case of partial co-eluting compounds. To make this process more specific, deconvolution has to be done and more sophisticated procedures adopted. This is possible by performing a ion peak detection for MS data; the algorithm processes the data array and isolates the response from single analytes based on spectral profiles.[19,30]

An important yet critical step of the protocol, and of any GC×GC-MS data interpretation process, relates to analytes identification. This procedure, proposed at steps 8 and 9, in absence of a confirmatory analysis with authentic standards, has to be carefully conducted by the analyst. Automated actions are available in any commercial software; they include MS spectral signature similarity evaluation against collected reference spectra (i.e., spectral libraries) and evaluation of characteristic ratios among qualifier/quantifier ions. However, additional confirmatory criteria are needed to disambiguate identification of isomers. The protocol proposes the adoption of linear retention indexes to prioritize the list of candidates; the limit here relates to the availability of retention data and its consistency.

The main characteristic that makes this approach unique is template matching.[12,13,15,28] Template matching enables 2D pattern recognition in a very effective, specific and intuitive way. It can be set, in terms of sensitivity and specificity, by applying customized threshold values and/or constraint functions while the analyst can supervise the procedure by actively interacting with transform function parameters. The peculiarity of this process relies on the possibility to cross-align targeted and untargeted peaks information between samples of a uniform batch but also between samples acquired with the same nominal conditions despite medium-to-severe misalignment. Advantages of this operation relate to the possibility to preserve all targeted analytes identifications, which is a time-consuming task for the analyst, and all metadata saved for targeted and untargeted peaks in previous elaboration sessions.

Template matching is also very effective in terms of computational time; low-resolution MS data files consists of about 1-2- Gb of packed data while high-resolution MS analyses may reach 10-15 Gb per single analytical run. Template matching does not process the full data matrix every time but, at first, performs retention time alignment between chromatograms using template peaks then, processes candidate peaks within the search window for their similarity match with reference in the template. In case of severe misalignment, the most challenging situation, global second-order polynomial transforms performed better than local methods while reducing computational time.[13]

For the GC×GC technique to spread widely beyond academia and research laboratories, data processing tools have to facilitate basic operations for visualization and chromatograms inspection; identification of analytes should offer the possibility to adopt standardized algorithms and procedures (*e.g.,* NIST search algorithm and I^T calibration); and cross-comparative analysis should be intuitive, effective and supported by interactive tools. The proposed approach addresses these needs while offering advanced options and tools to deal with complex situations such as, for example, analytes co-elution, multiple analytes calibration, group-type analysis, and parallel detection alignment.

The referenced literature well covers many possible scenarios where GC×GC and, more generally, comprehensive two-dimensional chromatography, offer unique solutions and reliable results that cannot be achieved by 1D-chromatography in single run analysis.[5,31,32] Although GC×GC is a most powerful tool that increases separation capacity and sensitivity, there are always limitations to separation power, sensitivity, and other systemic capacities. As those systemic limits are approached, data analysis becomes progressively more difficult. Therefore, research and development must continue to improve the analytical tools at our disposal.

References

(1) Tranchida, P. Q.; Donato, P.; Cacciola, F.; Beccaria, M.; Dugo, P.; Mondello, L. Potential of Comprehensive Chromatography in Food Analysis. *TrAC - Trends Anal. Chem.* **2013**, *52*, 186–205.

(2) Cordero, C.; Kiefl, J.; Reichenbach, S. E.; Bicchi, C. Characterization of Odorant Patterns by Comprehensive Two-Dimensional Gas Chromatography: A Challenge in Omic Studies. *TrAC - Trends in Analytical Chemistry*. 2019, pp 364–378.

(3) Cordero, C.; Kiefl, J.; Schieberle, P.; Reichenbach, S. E.; Bicchi, C. Comprehensive Two-Dimensional Gas Chromatography and Food Sensory Properties: Potential and Challenges. *Analytical and Bioanalytical Chemistry*. Springer Verlag 2015, pp 169–191.

(4) Adahchour, M.; Beens, J.; Vreuls, R. J. J.; Brinkman, U. A. T. Recent Developments in Comprehensive Two-Dimensional Gas Chromatography (GC × GC). I. Introduction and Instrumental Set-Up. *TrAC - Trends Anal. Chem.* **2006**, *25* (5), 438–454.

(5) Prebihalo, S. E.; Berrier, K. L.; Freye, C. E.; Bahaghighat, H. D.; Moore, N. R.; Pinkerton, D. K.; Synovec, R. E. Multidimensional Gas Chromatography: Advances in Instrumentation, Chemometrics, and Applications. *Anal. Chem.* **2018**, *90* (1), 505–532.

(6) Cordero, C.; Liberto, E.; Bicchi, C.; Rubiolo, P.; Schieberle, P.; Reichenbach, S. E.; Tao, Q. Profiling Food Volatiles by Comprehensive Two-Dimensional Gas Chromatography Coupled with Mass Spectrometry: Advanced Fingerprinting Approaches for Comparative Analysis of the Volatile Fraction of Roasted Hazelnuts (Corylus Avellana L.) from Different Ori. *J. Chromatogr. A* **2010**, *1217* (37).

(7) Cordero, C.; Liberto, E.; Bicchi, C.; Rubiolo, P.; Reichenbach, S. E.; Tian, X.; Tao, Q. Targeted and Non-Targeted Approaches for Complex Natural Sample Profiling by GC×GC-QMS. *J. Chromatogr. Sci.* **2010**, *48* (4).

(8) Maio, D.; Maltoni, D. Direct Gray-Scale Minutiae Detection in Fingerprints. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**.

(9) Jain, A. K.; Hong, L.; Pankanti, S.; Bolle, R. An Identity-Authentication System Using Fingerprints. *Proc. IEEE* **1997**, *85* (9), 1365–1388.

(10) Parsons, B. A.; Marney, L. C.; Siegler, W. C.; Hoggard, J. C.; Wright, B. W.; Synovec, R. E. Tile-Based Fisher Ratio Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry (GC × GC-TOFMS) Data Using a Null Distribution Approach. *Anal. Chem.* **2015**, *87* (7), 3812–3819.

(11) Pierce, K. M.; Kehimkar, B.; Marney, L. C.; Hoggard, J. C.; Synovec, R. E. Review of Chemometric Analysis Techniques for Comprehensive Two Dimensional Separations Data. *J. Chromatogr. A* **2012**, *1255*, 3–11.

(12) Reichenbach, S. E.; Rempe, D. W.; Tao, Q.; Bressanello, D.; Liberto, E.; Bicchi, C.; Balducci, S.; Cordero, C. Alignment for Comprehensive Two-Dimensional Gas Chromatography with Dual Secondary Columns and Detectors. *Anal. Chem.* **2015**, *87* (19), 10056–10063.

(13) Rempe, D. W.; Reichenbach, S. E.; Tao, Q.; Cordero, C.; Rathbun, W. E.; Zini, C. A. Effectiveness of Global, Low-Degree Polynomial Transformations for GCxGC Data Alignment. *Anal. Chem.* **2016**, *88* (20), 10028–10035.

(14) Stilo, F.; Liberto, E.; Reichenbach, S. E.; Tao, Q.; Bicchi, C.; Cordero, C. Untargeted and Targeted Fingerprinting of Extra Virgin Olive Oil Volatiles by Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry: Challenges in Long-Term Studies. *J. Agric. Food Chem.* **2019**, *67* (18), 5289–5302.

(15) Reichenbach, S. E.; Carr, P. W.; Stoll, D. R.; Tao, Q. Smart Templates for Peak Pattern Matching with Comprehensive Two-Dimensional Liquid Chromatography. *J. Chromatogr. A* **2009**, *1216* (16), 3458–3466.

(16) Reichenbach, S. E.; Tian, X.; Tao, Q.; Ledford, E. B.; Wu, Z.; Fiehn, O. Informatics for Cross-Sample Analysis with Comprehensive Two-Dimensional Gas Chromatography and High-Resolution Mass Spectrometry (GCxGC-HRMS). *Talanta* **2011**, *83* (4), 1279–1288.

(17) Reichenbach, S. E.; Tian, X.; Cordero, C.; Tao, Q. Features for Non-Targeted Cross-Sample Analysis with Comprehensive Two-Dimensional Chromatography. *J. Chromatogr. A* **2012**, *1226*, 140–148.

(18) Latha, I.; Reichenbach, S. E.; Tao, Q. Comparative Analysis of Peak-Detection Techniques for Comprehensive Two-Dimensional Chromatography. *J. Chromatogr. A* **2011**, *1218* (38), 6792–6798.

(19) Reichenbach, S. E.; Tao, Q.; Cordero, C.; Bicchi, C. A Data-Challenge Case Study of Analyte Detection and Identification with Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry (GC×GC-MS). *Separations* **2019**, *6* (3).

(20) Reichenbach, S. E. *Chapter 4 Data Acquisition, Visualization, and Analysis*; 2009; Vol. 55.

(21) Reichenbach, S. E.; Ni, M.; Zhang, D.; Ledford Jr., E. B. Image Background Removal in Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2003**, *985* (1–2).

(22) Reichenbach, S. E.; Tao, Q.; Cordero, C.; Bicchi, C. A Data-Challenge Case Study of Analyte Detection and Identification with Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry (GC×GC-MS). *Separations* **2019**, *6* (3), 38.

(23) van Den Dool, H.; Dec. Kratz, P. A Generalization of the Retention Index System Including Linear Temperature Programmed Gas—liquid Partition Chromatography. *J. Chromatogr. A* **1963**, *11*, 463–471.

(24) NIST Standard Reference Database 1A. NIST/EPA/NIH Mass Spectral Library with Search Program. National Institute of Standards and Technology (NIST): Gaithersburg MD 2017.

(25) Cialiè Rosso, M.; Mazzucotelli, M.; Bicchi, C.; Charron, M.; Manini, F.; Menta, R.; Fontana, M.; Reichenbach, S. E.; Cordero, C. Adding Extra-Dimensions to Hazelnuts Primary Metabolome Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography Combined with Time-of-Flight Mass Spectrometry Featuring Tandem Ionization: Insights on the Aroma Potential. *J. Chromatogr. A* **2020**, *1614* (460739), 1–11.

(26) Cordero, C.; Guglielmetti, A.; Bicchi, C.; Liberto, E.; Baroux, L.; Merle, P.; Tao, Q.; Reichenbach, S. E. Comprehensive Two-Dimensional Gas Chromatography Coupled with Time of Flight Mass Spectrometry Featuring Tandem Ionization: Challenges and Opportunities for Accurate Fingerprinting Studies. *J. Chromatogr. A* **2019**, *1597*, 132–141.

(27) Ni, M.; Reichenbach, S. E.; Visvanathan, A.; TerMaat, J.; Ledford, E. B. Peak Pattern Variations Related to Comprehensive Two-Dimensional Gas Chromatography Acquisition. In *Journal of Chromatography A*; 2005; Vol. 1086, pp 165–170.

(28) Reichenbach, S. E.; Tian, X.; Boateng, A. A.; Mullen, C. A.; Cordero, C.; Tao, Q. Reliable Peak Selection for Multisample Analysis with Comprehensive Two-Dimensional Chromatography. *Anal. Chem.* **2013**, *85* (10), 4974–4981.

(29) Markes International. Select-EV: The next Generation of Ion Source Technology. *Technical Note*. 2016.

(30) Tao;, Q.; Reichenbach;, S. E.; Heble;, C.; Wu, Z. New Investigator Tools for Finding Unique and Common Components in Multiple Samples with Comprehensive Two-Dimensional Chromatography. *Chromatogr. Today* **2018**, No. February-March.

(31) Seeley, J. V.; Seeley, S. K. Multidimensional Gas Chromatography: Fundamental Advances and New Applications. *Anal. Chem.* **2013**, *85* (2), 557–578.

(32) Tranchida, P. Q.; Aloisi, I.; Giocastro, B.; Mondello, L. Current State of Comprehensive Two-Dimensional Gas Chromatography-Mass Spectrometry with Focus on Processes of Ionization. *TrAC - Trends Anal. Chem.* **2018**, *105*, 360–366.

5.3 Untargeted and Targeted Fingerprinting of Extra Virgin Olive Oil Volatiles by Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry: Challenges in Long-Term Studies

Federico Stilo[1], Erica Liberto[1], Stephen E. Reichenbach[2,3], Qingping Tao[3], Carlo Bicchi[1] and Chiara Cordero[1*]

[1]Dipartimento di Scienza e Tecnologia del Farmaco, Università degli Studi di Torino, Turin, Italy

[2]Computer Science and Engineering Department, University of Nebraska, Lincoln, NE, USA

[3]GC Image LLC, Lincoln, NE, USA

*Corresponding author:

Prof. Chiara Cordero - Dipartimento di Scienza e Tecnologia del Farmaco, Università di Torino, Via Pietro Giuria 9, I-10125 Torino, Italy – e-mail: chiara.cordero@unito.it; phone: +39 011 6707172;

Received: March 14, 2019

Revised: April 17, 2019

Accepted: April 17, 2019

Published: April 17, 2019

DOI: 10.1021/acs.jafc.9b01661

J. Agric. Food Chem. 2019, 67, 5289−5302

5.3.1 Abstract

Comprehensive two-dimensional gas chromatography coupled with mass spectrometric detection (GC×GC-MS) offers an information-rich basis for effective chemical fingerprinting of food. However, GC×GC-MS yields 2D-peak patterns (*i.e.*, sample 2D fingerprints) whose consistency may be affected by variables related to either the analytical platform or to the experimental parameters adopted for the analysis.

This study focuses on the complex volatile fraction of extra-virgin olive oil and addresses 2D-peak patterns variations, including MS signal fluctuations, as they may occur in long-term studies where pedo-climatic, harvest year or shelf-life changes are studied. 2D-pattern misalignments are forced by changing chromatographic settings and MS acquisition. All procedural steps, preceding pattern recognition by template matching, are analyzed and a rational workflow defined to accurately re-align patterns and analytes metadata.

Signal-to-noise ratio (SNR) detection threshold, reference spectra extraction, and similarity match factor threshold are critical to avoid false-negative matches. Distance thresholds and polynomial transform parameters are key for effective template matching. In targeted analysis (supervised workflow) with optimized parameters, method accuracy reaches 92.5% (*i.e.*, % of true-positive matches) while for combined untargeted and targeted (*UT)* fingerprinting (unsupervised workflow), accuracy reaches 97.9 %. Response normalization also is examined, evidencing good performance of multiple internal standard normalization that effectively compensates for discriminations occurring during injection of highly volatile compounds. The resulting workflow is simple, effective, and time efficient.

Key words

Comprehensive two-dimensional gas chromatography coupled to time-of-flight mass spectrometry; extra-virgin olive oil volatiles; template matching; combined untargeted and targeted (UT) fingerprinting; data alignment in long-term studies

5.3.2 Introduction

One of the most important features of *-omics* approaches applied to food is the multiplicity of encrypted information that they can provide.[1] Because of the complexity of the food field, reliable results often can be obtained only with projects spread over years. Clear examples are monitoring chemical key characteristics of food crops whose cultivation has been moved in non-native countries or continents or submitted to the effect of climate changing. The possibility of comparing and/or merging data obtained over wide time-ranges, not necessarily resulting from rigorously standardized conditions (or improved analytical instrumentation), is therefore mandatory to achieve reliable and useful results. Comprehensive two-dimensional gas chromatography (GC×GC) is one of the most informative separation techniques for chemical characterization of complex fractions of volatiles from food. [2–4] It enables highly effective fingerprinting [5] and, when combined with mass spectrometric detection (GC×GC-MS), it has the intrinsic potential to provide a detailed profiling, giving access to higher level information encrypted in complex patterns of volatiles, for example: sample origin, technological signature, and aroma.[3,6–9]

Each analytical run produces dense and multidimensional data, so that elaboration and interpretation of chemical information is challenging. Moreover, 2D-peak patterns representing the sample 2D fingerprint, are defined by a series of variables also related to the analytical platform and to the experimental parameters adopted for the analysis. The choice of flow modulation instead of thermal/cryogenic modulation, MS detection by fast scanning quadrupoles *vs.* time-of-flight MS, low-resolution MS *vs.* high-resolution MS as well as GC×GC stationary phase combination, columns lengths and diameters, carrier gas linear-velocities, modulation period (P_M) and oven temperature programming greatly impact on 2D-patterns signature and informational density.[10]

Although most of these parameters, once fixed after method development and optimization, are kept constant (*e.g.*, column set-up, carrier gas flows, and modulation parameters) or can be standardized as the MS tuning and optimization, some others represent a source of random variability that must be considered when fingerprinting and pattern recognition studies extend over time and/or across different platforms.

For mono-dimensional (1D) GC-MS applications, possible strategies for chromatographic alignment and data normalization are: (*a*) linear retention indexing (van Den Dool and Kratz or Kovats indices) or retention time locking methods based on pressure/flow adjustments (*i.e.*, retention time locking) [11–13] to accurately locate target analytes along the analytical run; (*b*) chromatographic realignment; [14–16] (*c*) internal standardization for response normalization by single or multiple Internal Standards (IS) addition; and (*d*) external standard normalization by adopting single or multiple External Standards (ES). These strategies are effective and routinely adopted in peak-features based applications[5]. However, for GC×GC, these 1D-GC strategies may be ineffective especially for retention inconsistencies that result from two, almost independent, separation steps. On the other hand, the peculiar nature of 2D-peak patterns offers the possibility of exploiting pattern recognition algorithms for fast and effective fingerprinting. So, strategies for pattern alignment and normalization are needed.

Pattern recognition approaches based on peak-region features,[5] implemented with the smart template concept, [17] use different transform functions to facilitate recognition of 2D peak patterns based on retention time coordinates and establish correspondences between 2D-peaks, or 2D-peak-regions, from a *reference* pattern to those in an *analyzed* pattern even in presence of retention times shifts[17–19] and/or when severe misalignments occur because of different modulation regimes.[20,21]

Pattern correspondences are at the basis of the re-alignment of untargeted/targeted 2D-peaks or 2D-peak-regions across a samples-set to enable fingerprinting investigations.[3] Furthermore, the

specificity and reliability of pattern matching can be improved by including constraint functions, operating on the third dimension of the data, *i.e.*, the MS signature (*i.e.*, EI-MS fragmentation pattern). Typical functions are those that limit positive correspondences to 2D-peaks with spectral similarity match above a certain threshold or, more simply, for 2D-peaks that comply for specific m/z relative ratios between informative fragments of the spectrum.

In a scenario that food chemical fingerprinting must be extended over long time-frames, as for example to cover different harvest years or shelf-life modifications, strategies and tools for data re-alignment and normalization are necessary[22,23] together with more rational strategies and intuitive operative protocols/workflows to guide analysts over the present limits deriving from analytical data misalignment.

This study addresses, for the first time, 2D-peak patterns variations occurring in long-term studies that might impact the effectiveness of combined untargeted and targeted fingerprinting (*UT* fingerprinting).

The complex 2D patterns of volatiles from Extra Virgin Olive oil of different quality are here studied as model application. The ccomprehensive profiling and fingerprinting of olive oil volatiles can be, in fact, strategically extended over wide-time frames, *e.g.*, over harvest years or across the shelf-life of the product, to build databases for authenticity and valorisation of products. A concrete challenge faced through the collection of profiling data from Italian extra-virgin olive oils within the *Violin* project.[24]

Despite the great potential of GC×GC in exploiting the chemical dimensionality of olive oil volatile fractions, just a few studies are available in this field and none of them address challenges posed by long time-frame studies. Vaz Freire *et al.*[25] adopted an image-features approach[26] to investigate characteristic distributions of volatiles. An open-source image analysis software (Image J, National Institutes of Health) was used to extract detector response information from 2D regions over the separation space. Image-features with a high discrimination potential were selected by Principal Component Analysis (PCA) and targeted profiling was then combined to locate known analytes within most informative 2D regions.

Studies aimed at defining geographical origin indicators or cultivar markers include those by Cajka et al.[27] who adopted GC×GC-TOF-MS to identify 44 compounds able to discriminate extra-virgin olive oils based on their different geographical origin and production year, and by Lukić et al. [28] who applied a peak-features approach to reveal compositional differences between oils obtained by different olive cultivars and geographical areas. They considered, as potentially informative, both untargeted and targeted analytes as they were extracted from the raw data-set on the basis of relative retention and spectral features. Magagna et al. [8] first developed an integrated strategy for *UT* fingerprinting based on template matching, to define olive ripening indicators, while Purcaro et al.[9] combined targeted and untargeted analysis to delineate chemical blueprints of olive oil aroma defects.

The olive oil volatiles 2D-patterns considered were obtained in a one-year study during which misalignments and inconsistencies were introduced by varying column lengths and restrictions, modulation period (P_M), and operating with the time-of-flight (TOF) MS with differently optimized parameters. The processing steps, preceding template matching, are analyzed to define a rational workflow enabling consistent pattern recognition and to provide solid foundation for data processing procedures to be adopted in such challenging scenarios.

5.3.3 Materials and methods

5.3.3.1 Reference compounds and solvents

Pure reference standards of α- and β-thujone and methyl-2-octynoate used as Internal Standards (ISs), *n*-alkane standards (n-C7 to n-C25) used for linear retention index (I^T) calibration and pure reference compounds for targeted analytes' identity confirmation were supplied by Merk (Sigma-Aldrich srl Italy, Milan, Italy). Cyclohexane (HPLC grade) for n-alkane standards and pure dibutyl phthalate for ISTDs working solutions were from Merk.

5.3.3.2 Olive Oil samples

Extra Virgin Olive oils (EVO oils), supplied by the University of Granada (Spain), Prof. Luis Cuadros-Rodríguez, were obtained from olives of the *Picual* cultivar, harvested in the regions of Granada *Altipiano* named *Baza* and *Benamaurel*, and grown under differing production and irrigation practices. [8] Samples were obtained by mixing olives from five different trees from the same plot in duplicate batches (A and B). Olives for oil production were collected at four different ripening stages: November 10-12, 2014; November 24-28, 2014; December 16-17, 2014; and January 12-15, 2015, and classified by oil quality (Extra Virgin - EVOO, Virgin-VOO or *Lampante-LOO*). Samples acronyms and characteristics are summarized in **Table 5.3.1**. Oil qualifications were by a certified laboratory (ISO 17025:2018) [8] and according to Commission Regulation (EEC) No 2568/91 of 11 July 1991 and IOC Standard COI/T.15/NC No 3/Rev. 12. Some quality indices are reported in **Table 5.3.1**, including the sensory panel test results.

Table 5.3.1. List of samples together with acronym, harvest region, harvest stage, quality parameters according to COMMISSION REGULATION (EEC) No 2568/91 of 11 July 1991 and IOC Standard COI/T.15/NC No 3/Rev. 12, sensory evaluation results (Md: median of defects – Mf: median of fruity notes) and commercial classification.

Sample Acronym	Region	Harvest stage	Acidity (%)	Peroxide index (mEq O_2/kg)	K_{232}	K_{270}	ΔK	Md	Mf	Classification
Baza-1-A	Baza	November 10-12	0.2	5	1.84	0.2	0	0	5	EVOO
Baza-2-A	Baza	November 24-28	0.2	3	1.6	0.2	0	0	4.1	EVOO
Baza-3-A	Baza	December 16-17	0.2	5	1.17	0.2	0	> 0.00	1.3	VOO
Baza-4-A	Baza	January 12-15	0.4	11	1.11	0.1	0	> 0.00	0	LOO
Baza-1-B	Baza	November 10-12	0.2	4	1.92	0.2	0	0	5.2	EVOO
Baza-2-B	Baza	November 24-28	0.1	3	1.65	0.2	0	0	3.8	EVOO
Baza-3-B	Baza	December 16-17	0.2	6	1.28	0.1	0	> 0.00	1.7	VOO
Baza-4-B	Baza	January 12-15	0.4	13	1.12	0.1	0	> 0.00	0	LOO
Bena-1-A	Benamaurel	November 10-12	0.2	5	1.61	0.2	0	0	4.4	EVOO
Bena-2-A	Benamaurel	November 24-28	0.2	4	1.53	0.2	0	0	4.3	EVOO
Bena-3-A	Benamaurel	December 16-17	0.2	8	1.19	0.1	0	0	3.1	EVOO
Bena-4-A	Benamaurel	January 12-15	0.4	19	1.05	0.1	0	> 0.00	0	LOO
Bena-1-B	Benamaurel	November 10-12	0.1	4	1.64	0.4	0	0	4.2	EVOO
Bena-2-B	Benamaurel	November 24-28	0.2	3	1.48	0.2	0	0	4.3	EVOO
Bena-3-B	Benamaurel	December 16-17	0.2	6	1.51	0.1	0	0	2.9	EVOO
Bena-4-B	Benamaurel	January 12-15	0.2	14	1.05	0.1	0	> 0.00	0	LOO

5.3.3.3 Headspace solid phase microextraction sampling devices and conditions

Volatiles were sampled from samples by headspace (HS) solid phase microextraction (SPME). DVB/CAR/PDMS d_f 50/30 μm 2 cm length fiber (Supelco, Belle-fonte, PA, USA) was chosen based on previous studies [9] and conditioned before use as recommended by the manufacturer. The ISs were pre-loaded onto the SPME device, [29,30] before sampling, in a 20 mL headspace vial containing a 5.0 μL of α/β-thujone and methyl-2-octynoate at 100 mg L^{-1} in dibutyl phthalate standard solution. ISs were equilibrated at 40°C and pre-loaded by exposing the SPME device for 5 min.

Sampling was carried out on 0.100 ± 0.005 g of oil samples precisely weighed in 20 mL headspace vials. The very low amount of sample was chosen to comply with HS linearity conditions for most of the key-analytes responsible of samples discrimination. [8] Sampling was at 40°C for 60 min. After extraction, the SPME device was automatically transferred to the Split/Splitless (S/SL) injection port of the GC×GC system kept at 260°C for 5 min. Each sample was analyzed in triplicate.

5.3.3.4 Instrument set-up and analysis conditions

GC×GC analyses were performed on an Agilent 7890 GC chromatograph (Agilent Technologies, Wilmington DE, USA) coupled to a Markes BenchTOF-Select™ mass spectrometer featuring Tandem ionization™ (Markes International, Llantrisant, UK). The GC transfer line was at 270°C. TOF MS tuning parameters were set for single ionization at 70 eV and for tandem ionization at 70 and 12 eV; the scan range was set at 35-350 m/z with a spectra acquisition frequency of 100 Hz for single eV and 50 Hz/channel for tandem ionization.

The system was equipped with a two-stage KT 2004 loop type thermal modulator (Zoex Corporation, Houston, TX) cooled with liquid nitrogen controlled by Optimode v2.0 (SRA Intruments, Cernusco sul Naviglio, Milan, Italy). Modulation periods (P_M) and hot jet pulse times are detailed in **Table 5.3.2**, along with other parameters. A Mass Flow Controller (MFC) reduced the cold-jet stream from 45% to 8% of the total flow with a linear function along the run duration. A fused silica capillary loop (1.0 m × 0.1 mm id) was used in the modulator slit.

Table 5.3.2. Set-up 1 and 2 columns characteristics, settings and operative pressures.

	Set-up 1	***Set-up 2***
^{1}D Columns	^{1}D: SolGelWax™ (30 m, 0.25mm d_c, 0.25 μm d_f) Batch N° 1238274C06	^{1}D: SolGelWax™ (30 m, 0.25mm d_c, 0.25 μm d_f) Batch N° 1315621E03
^{1}D Carrier gas settings	He carrier @ 1.3 mL/min - constant flow conditions Average velocity ($^1\bar{u}$): 15.3 cm/s Initial head-pressure (relative) 234 kPa Outlet pressure (absolute) 285 kPa Hold-up 3.27 min	He carrier @ 1.3 mL/min - constant flow conditions Average velocity ($^1\bar{u}$): 12.8 cm/s Initial head-pressure (relative) 290 kPa Outlet pressure (absolute) 349 kPa Hold-up 3.89 min
^{2}D Columns	^{2}D: OV1701 Mega (1.0 m, 0.10 mm d_c, 0.10 μm d_f) Loop-capillary: deactivated fused silica (1.0 m, 0.10 mm d_c,) Restriction toward MS: none	^{2}D: OV1701 Mega (1.0 m, 0.10 mm d_c, 0.10 μm d_f) Loop-capillary: deactivated fused silica (1.0 m, 0.10 mm d_c,) Restriction toward MS: deactivated fused silica (1 m, 0.10 mm d_c,)
^{2}D Carrier gas settings	He carrier @ 1.3 mL/min - constant flow conditions Average velocity ($^2\bar{u}$): 157 cm/s Mid-point pressure (relative) 184 kPa Hold-up 1.28 s	He carrier @ 1.3 mL/min - constant flow conditions Average velocity ($^2\bar{u}$): 128 cm/s Mid-point pressure (relative) 248 kPa Hold-up 2.35 s
Modulation	P_M: 3.5 s - Hot-Jet pulse time: 250 ms	P_M: 4s - Hot-Jet pulse time: 250 ms

The column set was configured as follows: ^{1}D SolGel-Wax column (100% polyethylene glycol; 30 m × 0.25 mm d_c × 0.25 μm d_f) from Trajan Analytical Science (Ringwood, Australia) coupled with a ^{2}D OV1701 column (86% polydimethylsiloxane, 7% phenyl, 7% cyanopropyl; 1 m × 0.1 mm d_c × 0.10 μm d_f) from Mega (Legnano, Milan, Italy). A capillary restriction towards the MS was used to generate a differential pressure-drop influencing actual carrier gas linear velocities along the column in *Set-up 2* (1 m × 0.1 mm d_c deactivated silica). The GC Split/Splitless (S/SL) injector port was set at 260°C and operated in split mode with a split ratio 1:20.

The carrier gas was helium at a constant nominal flow of 1.3 mL/min. The oven temperature programming was set as follows: 40°C (2 min) to 240°C (10 min) at 3.5°C/min. Carrier gas average linear velocities ($^{1}\bar{u}$ and $^{2}\bar{u}$), pressure settings, and hold-up times are reported in **Table 5.3.2** and were obtained by basic calculations with a reference temperature of 60°C.

For I^T determination, 1μL of the n-alkanes sample solution was injected with a split ratio 1:50.

Data were acquired by TOF-DS™ (Markes International, Llantrisant, UK) and processed by GC Image ver. 2.8 (GC Image, LLC Lincoln, NE, USA).

5.3.3.5 UT fingerprinting workflow

The distribution of detectable analytes over the 2D chromatographic space in a GC×GC separation is at the core of pattern recognition based on the *smart template* concept. [17] The template is a pattern of 2D-peaks and/or graphic objects (geometrical objects delineating 2D-peak contour) built over a *reference* image(s) (single or composite image) [31] and then used to recognize similar patterns of 2D-peaks in an *analyzed* image(s).[32] Each template object (2D-peak and/or graphic) can carry various metadata including: compound chemical name, retention times, I^T, mass spectrum, informative ions and their relative ratios, constraint functions to limit peak correspondences above certain thresholds, and qualifier functions examining metadata information by script functions.

Peak-region features[33] are of great help in the presence of temporal inconsistencies and detector fluctuations. Peak-regions attempt to define one chromatographic region around each analyte peak, thereby achieving the one-feature-to-one-analyte selectivity but with greater robustness than can be achieved with single 2D-peak detection.[26] 2D-peaks and peak-regions are features adopted in the combined targeted and untargeted fingerprinting (*UT* fingerprinting) strategy.[8,20,31,34]

UT fingerprinting establishes a group of *reliable* peaks, positively matched across *all* or *most-of* chromatograms in set,[35] and then uses them to align chromatograms[18] for their combination into a single, composite chromatogram. From the composite chromatogram, 2D-peaks (*i.e.*, the combination of the re-aligned responses in the 2D retention time plane) are detected and their outlines are recorded to define peak-region objects. The set of reliable 2D-peaks and peak-regions objects are collected in the *feature* template, or *consensus* template, covering the whole sample-set variability and capable of cross-corresponding chemical feature patterns among samples. Within all detected analytes, the subset of targeted compounds can be highlighted by completing their metadata fields (compound name, ion ratios, I^T) and computed together with untargeted features during the data processing.

A schematic of the UT fingerprinting process is illustrated in the Supplementary Material – **Supplementary Figure 5.3.1** together with some details on targeted and untargeted 2D peaks and peak-regions.

5.3.4 Results and discussion

This section discusses all steps, supervised by the analyst, aiming at selecting key parameters to generate a targeted template or a *reliable* or *consensus* template covering the entire volatile fraction, and its chemical dimensionality,[6] of extra-virgin olive oil. In particular, some critical parameters are examined for their impact on template matching efficacy: (*a*) 2D-peaks response thresholds (Signal-to-Noise ratio - SNR and Volume-to-Noise ratio – VNR); (*b*) MS reference signature to be used for spectral constraints; (*c*) MS similarity thresholds; and (*d*) the template matching transform and settings to compensate for 2D-peak patterns variability to achieve an effective, reliable and comprehensive chemical fingerprinting.

The next paragraphs illustrate the strategy adopted to generate misaligned patterns and the subsequent workflow designed to re-align templates. Results are critically discussed in term of accuracy (*i.e.*, true positive matches) and data inter-batch transferability (*i.e.*, response normalization).

5.3.4.1 Pattern misalignment challenges

Chromatographic pattern distortions and misalignments were induced, generating worst-case scenarios, by changing the following parameters: (*a*) columns were from different commercial batches; (*b*) a post-column restriction was added generating a pressure drop between column inlet and outlet, influencing carrier gas linear velocities in both analytical dimensions; and (*c*) the P_M was set at 4 or 3.5 s generating an absolute ^{2}D retention misalignment. All the other parameters, carrier gas nominal flow, oven temperature programming and injection conditions were kept constant. Analytical conditions for the two resulting set-ups (*i.e.*, *Set-up 1* and *Set-up 2*) are summarized in **Table 5.3.2**. **Figure 5.3.1** shows the colorized plot of an oil sample obtained from olives at an early ripening stage (*Baza-1-A*) analyzed by the two set-ups (**Figure 5.3.1A** for *Set-up 1* and **Figure 5.3.1B** for *Set-up 2*). Pattern differences are visible and are related to the different chromatographic efficiencies (peak-width – ^1W and ^2W – **Table 5.3.2** data) that impact resolution and to the absolute retentions that affect system orthogonality.[36–38].

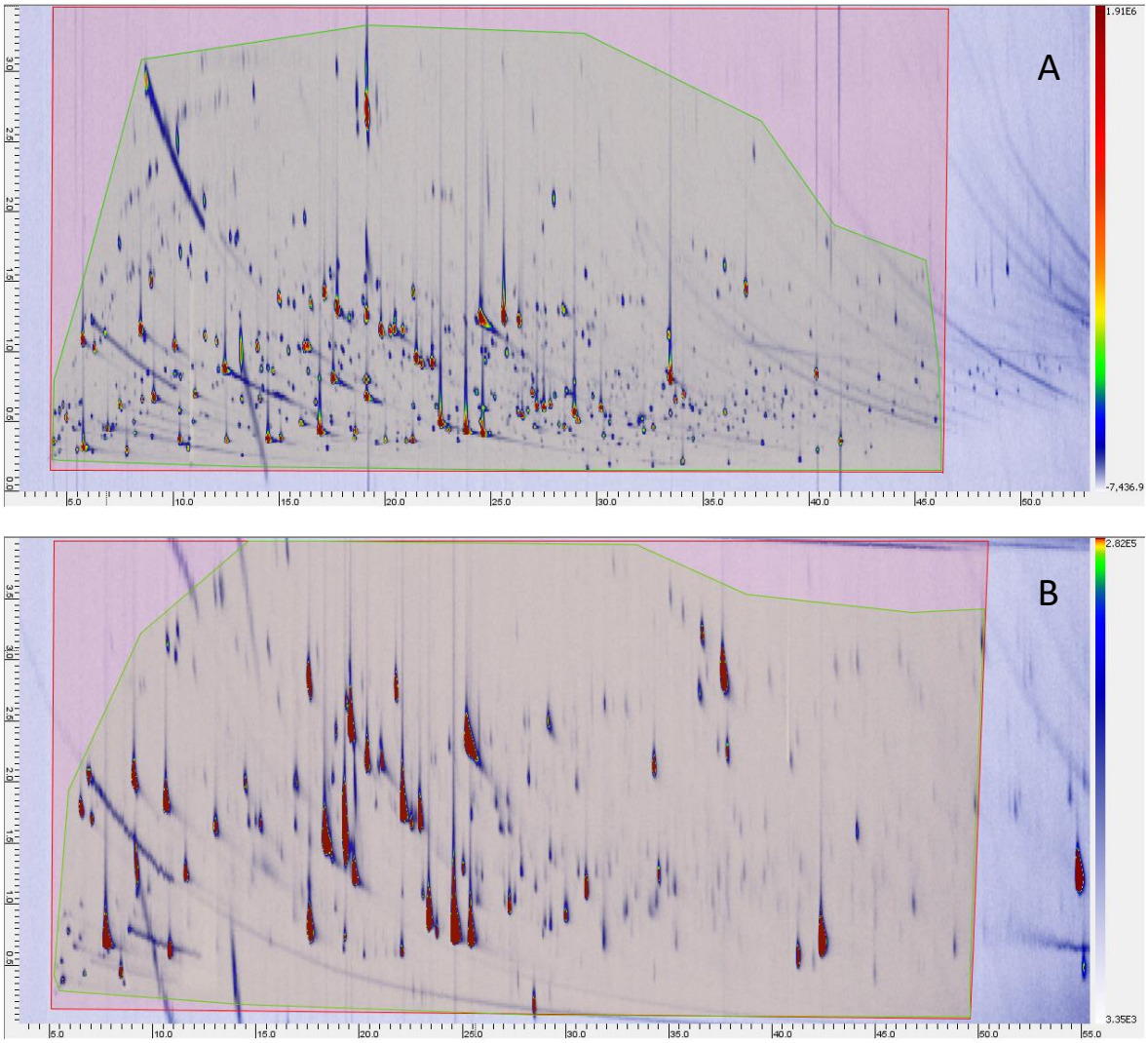

Figure 5.3.1. Colorized plot of the Baza-1-A sample analyzed with Set-up 1 (**5.3.1A**) and Set-up 2 (**5.3.1B**). Red colored areas indicate the available separation space while green areas include all targeted peaks elution area. For details, see the text.

The global misalignment between peaks patterns is visualized in **Figure 5.3.2** and evaluated by calculating analyte relative retention in both separation dimensions against reference peaks.[21] *(Z)-3-hexen-1-ol* acetate, which elutes in the middle area of the chromatographic plane, was arbitrarily chosen as reference/centroid, while *phenol,* the last-eluting marker, was used to normalize each analyte's relative position *i*.

The ^{1}D relative retention (^{1}D RR) is calculated by **Equation 5.3.1**:

Equation. 5.3.1 $$^{1}\text{D RR} = (^{1}t_{Ri} - {}^{1}t_{R(Z)\text{-3-Hexen-1-ol acetate}}) \; / \; {}^{1}t_{R\text{phenol}}$$

where $^{1}t_{Ri}$ corresponds to the first-dimension retention time expressed in minutes for the targeted peak *i*, *(Z)-3-Hexen-1-ol acetate* is the reference peak, and *phenol* is the last eluting peak. The ^{2}D relative retention (^{2}D RR) is calculated through **Equation 5.3.2**:

Equation. 5.3.2 $$^{2}\text{D RR} = (^{2}t_{Ri} - {}^{2}t_{R(Z)\text{-3-Hexen-1-ol acetate}}) \; / \; P_M$$

where $^{2}t_{Ri}$ corresponds to the second dimension retention time expressed in minutes for the targeted peak *i*, *(Z)-3-hexen-1-ol acetate* is the reference peak, and P_M the modulation time. [21]

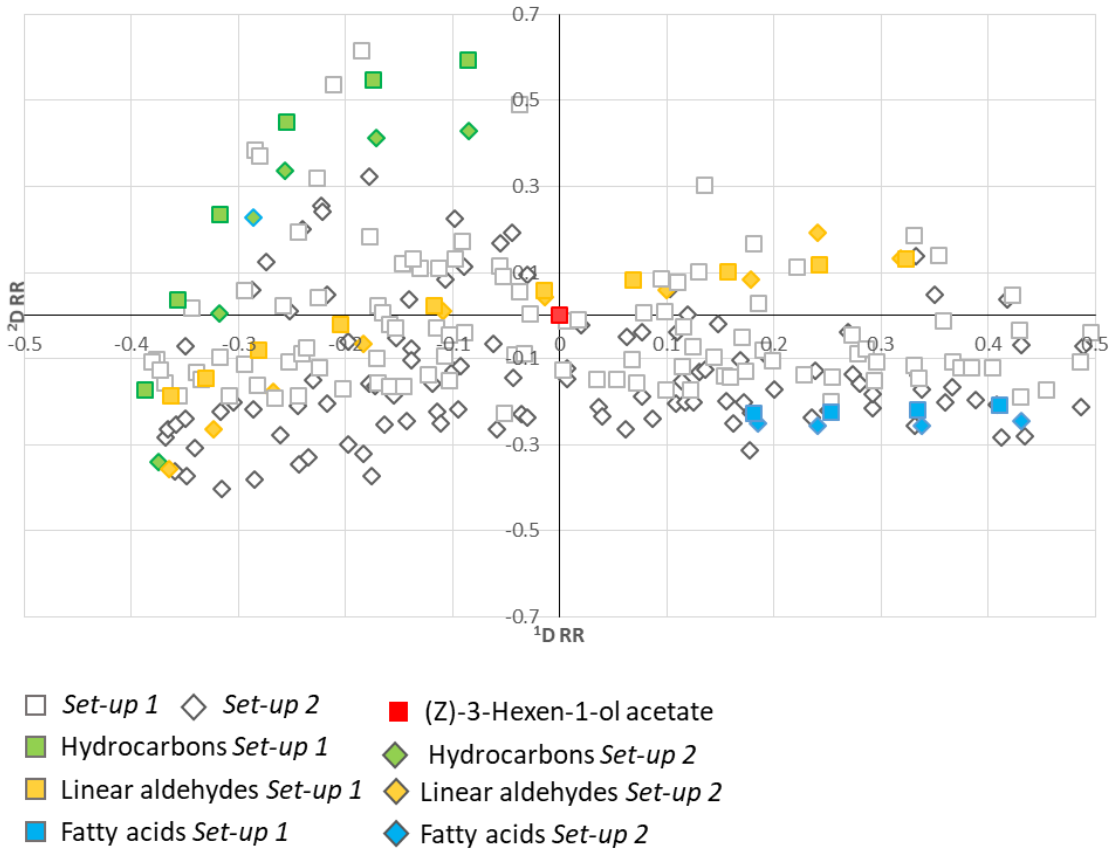

Figure 5.3.2. Dispersion graph resulting from the relative position of targeted peak analytes (white indicators) from the two set-ups in the normalized retention times space: homologous series n-alkanes - green indicators, linear saturated aldehydes - orange indicator

As visualized in **Figure 5.3.2**, there is a dramatic impact on the ^{2}D absolute and relative retention. This effect is due primarily to the different P_M applied (4 vs. 3.5 seconds) and to the actual carrier gas linear velocities ($^{1}\bar{u}$ and $^{2}\bar{u}$) and operative pressures (initial head-pressure and mid-point pressure). Analytes falling in the third quadrant show an higher ^{2}D k in *Set-up 1,* whereas this effect decreases as a function of increasing ^{1}D k (retention).

Interestingly, the two patterns, although misaligned on the normalized retention times space, keep coherent the group-type separation for homologous series. Normal alkanes (*n*-alkanes), shown with green indicators, mostly in the second quadrant; linear aldehydes, shown with orange indicators, spanning mostly across the first and third quadrants; and short chain fatty acids, shown with cyan indicators, appearing in the forth quadrant, all are rationalized over the 2D space.

The next step of the study addresses detector response variations and examines threshold parameters for 2D-peaks descriptors to adopt for consistent template matching.

5.3.4.2 Supervised workflow for reliable targeted template construction

MS detector response fluctuations due to tuning, optimization, and/or other factors directly impact on absolute response and background noise intensity. Such performance issues also may affect template matching effectiveness and analyte identity confirmation, as a consequence of variable peak detection and the varying quality and reliability of 2D-peak spectra adopted as reference for matching. In this study, TOF MS was set differently: in *Set-up 1,* it multiplexed between

high and low ionization energies (Tandem Ionization ™ - 70 and 12 eV) at 50 Hz acquisition rate for each channel, whereas in *Set-up 2,* TOF MS operated in single electron energy acquisition mode (70 eV) at 100 Hz. Therefore, MS was tuned differently [39] and output signals exhibited different absolute responses (total ion current) and background noise intensities.

The signals resulting from the same sample (*Baza*-1-A), whose patterns are illustrated in **Figure 5.3.1A** and **5.3.1B**, have the following characteristics:

- Background noise sampled in the middle of the chromatogram within a 50x50 acquisition-point window reported an average absolute total intensity of 75,500 counts (RSD% ^{1}D 0.6 and ^{2}D 5.98) for *Set-up 1* and 167,000 counts (RSD% ^{1}D 0.5 and ^{2}D 1.57) for *Set-up 2*. Supplementary data visually illustrates performance evaluation operations (Supplementary material – **Supplementary Figure 5.3.2**).
- After background subtraction[40], the average intensity was 38,000 counts (RSD% ^{1}D 1.19 and ^{2}D 8.9) for *Set-up 1* and 88,000 counts (RSD% ^{1}D 0.9 and ^{2}D 2.49) for *Set-up 2*.
- The number of detected 2D-peaks above SNR 15 were 770 for *Set-up 1* and 500 for *Set-up 2*.
- Within the detected 2D-peaks, SNR values ranged between 15-13,000 in *Set-up 1* and between 15-3,000 in *Set-up 2*.
- Volume-to-noise ratio (VNR) values ranged between 100-14,100 in *Set-up 1* and 100-6,700 in *Set-up 2*.

Experimental results indicate that MS, operating with a single ionization energy at 70 eV, results in greater absolute and relative background noise (*e.g.*, 1.9 times) compared to the tandem ionization settings. Interestingly, the noise fluctuations are greater along the ^{1}D (RSD% values) where column bleeding increases as a function of temperature programming. Background noise subtraction has almost the same effect, in terms of noise suppression, and in both cases, signal intensity is halved compared to the initial values.

With respect to peak detection, *Set-up 1* exhibited better chromatographic efficiency (**Table 5.3.3** $^{1}W_{0.1}$ and $^{2}W_{0.1}$) and resulted in a larger number of detected peaks over SNR≥15, with a wide range of variation, *i.e.*, 15-13,000 SNR, whereas in *Set-up 2* maximum SNR achieved only a value of 3,000. On the other hand, VNR, which corresponds to the ratio of analyte 2D-peak volume to the standard error (SE ($\sigma/\sqrt{n}$)), is not so influenced by peak-width as SNR. It informs about the dynamic range of the MS response[41] and with *Set-up 1* this appears 10 times greater (up to 14,100 vs. up to 6,700 VNR) compared to *Set-up 2*. Although with *Set-up 1* the number of detected peaks over a SNR of 15 was higher, a greater volume standard deviation (σ) was computed.

Table 5.3.3. List of all targeted analytes together with their experimental I^T, identification criterion (a) authentic reference compound or (b) $I^T \pm 20$ and spectral similarity direct match ≥ 850, elution order (#Rank) in the two set-up; retention times (^1t_R and ^2t_R) and relative standard deviation (RSD%) calculated over all analyzed samples, peak-width at 10% of peak height estimated on each dimension ($^1W_{0.1}$ (min) and $^2W_{0.1}$ (sec)).

				Set-up 1							***Set-up 2***					
Compounds	**Exp I^T**	**Identification**	**# Rank**	**^1t_R (min)**	**RSD %**	**^2t_R (sec)**	**RSD %**	**$^1W_{0.1}$ (min)**	**$^2W_{0.1}$ (sec)**	***# Rank***	**^1t_R (min)**	**RSD %**	**^2t_R (sec)**	**RSD %**	**$^1W_{0.1}$ (min)**	**$^2W_{0.1}$ (sec)**
Hexane	778	a	1	4.36	0.77	0.33	3.53	0.12	0.05	1	4.80	8.45	1.06	5.80	0.40	0.15
2,2,4-Trimethylpentane	789	b	2	4.63	0.73	0.56	0.00	0.10	0.06	2	5.44	3.53	0.83	3.07	0.36	0.13
Heptane	795	a	3	4.78	0.00	0.58	0.00	0.10	0.05	3	5.53	3.61	0.60	5.38	0.38	0.16
Cyclohexane	802	b	4	4.96	0.00	0.49	2.34	0.08	0.04	6	5.88	3.49	0.96	7.99	0.29	0.11
1,3-Pentadiene	813	b	5	5.25	0.00	0.42	0.00	0.37	0.17	4	5.60	0.00	0.54	4.26	0.53	0.17
Propanal	819	a	6	5.44	0.62	0.28	0.00	0.25	0.12	5	5.82	0.00	0.52	7.62	0.40	0.15
Octane	830	a	7	5.70	0.59	1.06	0.00	0.10	0.09	8	6.28	1.18	1.68	1.04	0.33	0.15
Acetone	831	a	8	5.74	0.59	0.29	4.03	0.25	0.11	10	6.67	0.00	0.74	8.48	0.53	0.37
1-Octene	850	b	9	6.26	0.54	1.01	1.15	0.12	0.08	7	6.26	0.62	1.01	3.78	0.40	0.23
Tetrahydrofuran	858	b	10	6.46	0.52	0.48	0.00	0.23	0.16	9	6.33	1.82	0.48	13.50	0.36	0.18
Butanal	867	a	11	6.67	0.51	0.42	0.00	0.23	0.13	11	7.47	2.36	0.91	8.41	0.42	0.23
Ethyl acetate	873	a	12	6.86	0.49	0.43	2.71	0.18	0.05	13	7.74	0.01	1.08	0.05	0.33	0.16
2,3-Dimethylheptane	893	b	13	7.43	0.45	1.76	1.14	0.23	0.17	12	7.73	0.86	1.99	11.30	0.38	0.13
2-Methylbutanal	895	a	14	7.44	0.01	0.60	0.21	0.22	0.07	15	8.27	8.41	1.16	8.24	0.43	0.24
Ethanol	913	a	15	7.78	0.43	0.26	0.00	0.19	0.09	14	7.78	2.47	0.38	3.53	0.42	0.29
1-Methoxyhexane	932	b	16	8.42	0.40	1.15	1.01	0.14	0.09	16	9.09	0.43	2.21	2.66	0.29	0.31
2-Ethylfurane	932	b	17	8.42	0.40	0.54	0.00	0.29	0.07	17	9.11	0.74	1.10	3.64	0.38	0.17
3,4-Diethyl-1,5-hexadiene (RS+SR)	942	a	18	8.63	0.31	1.61	1.90	0.21	0.08	18	9.15	2.76	2.74	13.90	0.66	0.16
2,3-Butanedione	955	a	19	8.94	0.38	0.37	3.09	0.37	0.05	19	9.18	3.71	0.45	11.80	0.42	0.13
Pentanal	956	a	20	9.03	0.00	0.65	0.02	0.30	0.23	22	10.00	0.04	1.26	0.05	0.37	0.16
3,4-Diethyl-1,5-hexadiene (meso)	960	a	21	9.17	0.16	1.65	1.87	0.23	0.10	20	9.63	2.81	2.78	15.05	0.54	0.14
3-Methylnonane	960	b	22	9.20	0.37	2.33	0.50	0.31	0.27	21	9.65	0.69	2.48	12.00	0.51	0.19
Acetonitrile	980	b	23	9.69	0.00	0.27	0.04	0.15	0.10	23	10.27	0.03	0.86	0.29	0.30	0.32
(Z)-1-Methoxy-3-hexene	991	b	24	10.01	0.34	1.03	1.12	0.14	0.09	25	10.66	0.36	1.99	4.07	0.24	0.24
Decane	997	a	25	10.18	0.01	2.51	0.03	0.18	0.08	24	10.48	0.06	3.31	0.09	0.32	0.15
(5Z)-3-Ethyl-1,5-Octadiene	999	a	26	10.21	0.17	1.81	1.45	0.24	0.13	26	10.89	2.15	3.20	13.70	0.87	0.20
1-Penten-3-one	1000	a	27	10.25	0.33	0.56	0.00	0.21	0.06	27	10.96	0.70	1.11	3.74	0.42	0.23
(5E)-3-Ethyl-1,5-Octadiene	1010	a	28	10.63	0.11	1.74	1.55	0.24	0.15	29	11.06	2.19	3.12	12.08	0.72	0.18
Propan-1-ol	1011	a	29	10.65	0.00	0.29	0.04	0.19	0.08	31	11.43	0.02	0.65	0.11	0.34	0.18
α-Pinene	1011	a	30	10.66	0.00	1.62	0.01	0.20	0.06	30	11.18	0.01	2.77	0.17	0.28	0.16

(E)-2-Butenal	1017	a	31	10.95	1.11	0.50	4.00	0.21	0.15	28	11.05	1.51	0.59	6.55	0.40	0.15
Toluene	1020	a	32	10.99	0.31	0.68	2.94	0.18	0.07	32	11.61	0.67	1.37	4.68	0.33	0.17
1-Decene	1032	b	33	11.41	0.00	2.06	0.01	0.17	0.05	33	11.98	0.38	2.98	0.53	0.43	0.20
2,3-Pentanedione	1033	a	34	11.50	0.00	0.52	0.00	0.21	0.06	35	12.19	0.32	1.16	0.27	0.36	0.23
4-Methyldecane	1051	b	35	12.04	0.56	2.83	0.41	0.29	0.31	34	12.02	0.56	2.93	2.44	0.53	0.38
(E,Z)-3,7-Decadiene	1054	a	36	12.15	0.14	1.99	1.23	0.20	0.09	36	12.91	2.41	3.47	12.90	0.58	0.35
Hexanal	1059	a	37	12.39	0.27	0.86	0.00	0.21	0.07	37	13.09	1.28	1.72	6.23	0.42	0.19
Isobutyl alcohol	1061	a	38	12.46	0.27	0.34	0.00	0.16	0.04	38	13.10	0.88	0.78	6.51	0.29	0.17
(E,E)-3,7-Decadiene	1077	a	39	12.91	0.15	2.03	1.15	0.19	0.11	39	13.22	2.33	3.53	14.36	0.46	0.33
2,4,6-Trimethyldecane	1090	b	40	13.26	0.25	3.11	0.37	0.29	0.33	43	13.99	1.93	3.26	8.52	0.44	0.35
β-Pinene	1093	a	41	13.59	0.00	1.58	0.00	0.17	0.05	41	13.77	0.01	1.70	0.22	0.25	0.12
Undecane	1097	a	42	13.75	0.00	2.85	0.01	0.19	0.09	48	14.30	0.01	3.61	0.09	0.32	0.16
3-Penten-2-one	1100	a	43	13.88	0.00	0.60	0.00	0.27	0.14	40	13.73	1.70	0.69	6.19	0.42	0.22
1-Methoxy-2-propanol	1101	a	44	13.90	0.01	0.39	0.03	0.38	0.09	44	14.10	0.02	0.48	0.74	0.41	0.22
3-Methylbutyl acetate	1101	a	45	13.90	0.24	1.02	0.00	0.14	0.10	42	13.95	3.01	1.34	9.21	0.69	0.18
1-Methoxy-1-propanol	1103	b	46	13.98	0.24	0.39	2.94	0.39	0.16	45	14.12	1.84	0.66	10.90	0.36	0.32
(E)-2-Pentenal	1004	a	47	14.04	0.03	0.64	0.31	0.28	0.12	47	14.27	0.80	0.88	0.62	0.24	0.27
Ethylbenzene	1106	b	48	14.12	0.00	0.86	0.00	0.16	0.07	46	14.24	1.51	1.31	7.54	0.42	0.30
1,4-Dimethylbenzene	1115	b	49	14.41	0.00	0.87	1.32	0.16	0.17	49	14.59	0.53	0.95	10.60	0.47	0.25
1-Butanol	1115	a	50	14.41	0.00	0.36	0.00	0.21	0.06	50	14.98	1.81	1.22	8.66	0.36	0.23
1,3-Dimethylbenzene	1121	b	51	14.64	0.00	0.84	0.00	0.14	0.06	51	15.09	0.44	1.75	6.31	0.64	0.35
(Z)-3-Hexenal	1024	a	52	14.77	0.01	0.73	0.06	0.29	0.13	52	15.42	0.75	1.21	3.67	0.32	0.29
Butyl 2-methylpropanoate	1128	b	53	14.94	0.00	1.36	0.02	0.12	0.08	54	15.62	0.01	2.12	0.05	0.33	0.15
1-Penten-3-ol	1130	a	54	15.01	0.22	0.36	0.00	0.16	0.05	53	15.53	0.86	0.99	12.50	0.36	0.24
2-Methylpropyl butyrate	1141	a	55	15.40	0.00	1.33	0.87	0.31	0.21	56	15.74	1.30	1.55	8.44	0.38	0.15
ß-Myrcene	1147	a	56	15.65	0.22	1.33	0.87	0.14	0.09	55	15.68	1.07	1.59	6.74	0.34	0.32
Heptanal	1162	a	57	16.28	0.00	1.02	0.00	0.19	0.08	59	17.07	2.64	2.00	6.10	0.44	0.18
1,3-Xylene	1163	b	58	16.33	0.00	0.83	1.39	0.18	0.11	57	16.61	0.97	1.34	4.40	0.44	0.29
2-Ehylhexanal	1166	a	59	16.45	0.00	1.33	0.87	0.14	0.09	60	17.17	3.19	2.3	13.60	0.51	0.29
(Z)-2-Hexenal	1172	a	60	16.68	0.00	0.71	1.88	0.16	0.07	58	16.98	2.24	0.97	14.10	0.47	0.32
3-Methyl-1-butanol	1177	b	61	16.86	0.00	0.79	1.47	0.16	0.07	61	17.50	3.19	1.45	13.60	0.51	0.29
2-Methyl-1-butanol	1177	b	62	16.86	0.00	0.41	0.03	0.16	0.08	63	17.69	0.01	1.10	0.25	0.38	0.25
Limonene	1182	a	63	17.09	0.00	1.41	0.82	0.16	0.09	62	17.57	0.44	2.87	19.40	0.36	0.26
Eucalyptol	1190	a	64	17.37	0.00	1.54	0.01	0.32	0.15	65	17.94	0.01	2.43	0.12	0.37	0.24
(E)-2-Hexenal	1193	a	65	17.50	0.00	0.81	1.43	0.31	0.12	64	17.80	2.43	1.50	3.66	0.31	0.19
Dodecane	1197	a	66	17.68	0.00	3.02	0.01	0.19	0.09	66	18.18	0.01	3.68	0.08	0.33	0.16
Terpinene	1229	a	67	18.92	0.18	1.35	0.85	0.12	0.09	67	19.20	0.34	1.71	3.07	0.20	0.30
(E)-ß-Ocimene	1234	a	68	19.08	0.00	1.27	1.82	0.16	0.11	69	19.49	0.35	2.64	5.06	0.27	0.21
1-Pentanol	1236	a	69	19.15	0.18	0.11	6.74	0.33	0.17	68	19.33	2.60	0.92	7.57	0.69	0.31
Styrene	1244	b	70	19.57	0.01	0.62	0.01	0.31	0.22	70	19.98	0.01	1.39	0.10	0.58	0.43
1-Dodecene	1249	b	71	19.74	0.17	2.67	0.87	0.27	0.32	71	19.98	1.75	2.74	12.80	0.33	0.30
Hexyl acetate	1250	a	72	19.78	0.00	1.13	1.02	0.18	0.09	74	20.59	0.97	2.35	15.40	0.42	0.22

3-Hydroxy-2-butanone	1256	a	73	19.93	0.17	0.63	1.84	0.37	0.21	73	20.58	1.16	1.02	4.55	0.36	0.21
1,2,3-Trimethylbenzene	1261	b	74	20.18	0.00	0.96	0.00	0.27	0.09	72	20.35	2.01	1.06	3.73	0.38	0.27
Octanal	1266	a	75	20.36	0.00	1.11	0.25	0.20	0.08	75	20.91	0.47	1.65	1.13	0.39	0.33
2-Ethyl-2-hexenal	1278	a	76	20.77	0.00	1.14	0.02	0.17	0.09	76	21.34	0.02	2.13	0.08	0.48	0.32
(Z)-3-Hexenyl acetate	1300	a	77	21.43	0.16	0.94	0.00	0.14	0.09	77	21.95	0.47	1.98	1.01	0.27	0.33
N,N-Dimethylformamide	1302	b	78	21.53	0.00	0.50	0.00	0.21	0.15	79	22.28	1.95	1.47	13.30	0.36	0.37
(Z)-2-Heptenal	1306	a	79	21.66	0.16	0.90	0.00	0.25	0.08	78	22.24	1.55	1.38	5.59	0.29	0.29
6-Methyl-5-hepten-2-one	1319	a	80	22.17	0.00	0.91	1.26	0.18	0.08	80	22.86	0.45	1.87	1.22	0.20	0.29
1-Hexanol	1328	a	81	22.89	0.03	0.36	0.44	0.31	0.16	81	23.35	0.15	1.03	1.51	0.40	0.26
(Z)-3-Hexen-1-ol	1346	a	82	23.74	0.00	0.42	0.00	0.21	0.11	82	23.73	0.85	1.03	6.28	0.22	0.24
(E,E)-2,4-Hexadienal	1369	a	83	24.38	0.00	0.58	0.00	0.18	0.05	85	25.39	0.26	1.22	2.84	0.33	0.49
Nonanal	1370	a	84	24.44	0.00	1.24	1.61	0.29	0.10	84	24.77	1.48	1.77	14.10	0.49	0.31
(E)-2-Hexen-1-ol	1373	a	85	24.56	0.00	0.40	0.00	0.21	0.09	83	24.71	2.20	0.91	3.37	0.31	0.31
(E)-3-Octen-2-one	1382	a	86	24.85	0.00	0.96	0.01	0.28	0.10	86	25.41	0.01	1.82	0.03	0.50	0.33
α-Thujone (ISTD)	**1402**	**a**	**87**	**25.55**	**0.00**	**1.24**	**0.02**	**0.25**	**0.16**	**88**	**26.47**	**0.02**	**2.20**	**0.13**	**0.31**	**0.31**
(E)-2-Octenal	1415	a	88	25.73	0.00	0.98	2.04	0.27	0.14	87	25.83	1.13	1.01	11.90	0.80	0.24
1-Octen-3-ol	1417	a	89	25.78	0.00	0.34	0.00	0.18	0.08	90	26.83	0.01	1.15	0.09	0.20	0.29
β-Thujone (ISTD)	1424	a	90	26.26	0.00	1.21	0.01	0.15	0.09	89	26.59	0.01	2.21	0.10	0.21	0.20
1-Heptanol	1429	a	91	26.43	0.00	0.53	2.19	0.18	0.05	93	27.19	0.28	1.17	2.65	0.27	0.30
1-Ethenyl-4-ethylbenzene	1430	b	92	26.48	0.00	0.86	0.00	0.37	0.13	91	26.87	0.72	1.81	1.68	0.51	0.25
Furfural	1434	a	93	26.77	0.00	0.34	0.00	0.17	0.06	95	27.54	0.01	1.17	0.22	0.32	0.30
(E,E)-2,4-Heptadienal	1437	a	94	26.89	0.00	0.69	1.67	0.14	0.06	96	27.80	1.06	1.44	0.00	0.62	0.25
5-Methyl-2-(1-methylethyl)-cyclohexanone	1446	b	95	27.13	0.00	1.31	0.88	0.21	0.13	92	27.05	0.28	1.37	2.41	0.33	0.25
Butyl-2-ethylhexanoate	1453	a	96	27.42	0.00	1.89	1.22	0.39	0.29	94	27.33	0.88	1.97	7.99	0.40	0.19
2-Ethyl-1-hexanol	1460	a	97	27.72	0.00	0.60	0.00	0.21	0.11	97	28.06	0.02	1.46	0.23	0.34	0.20
(Z)-Hepten-4-ol	1469	a	98	28.18	0.00	0.46	0.02	0.19	0.08	99	28.93	0.01	1.17	0.23	0.39	0.15
Decanal	1474	a	99	28.35	0.00	1.30	1.54	0.25	0.09	98	28.60	1.50	1.89	14.30	0.33	0.25
(E)-2-Hepten-1-ol	1476	a	100	28.41	0.00	0.44	0.02	0.41	0.16	100	29.20	0.00	0.97	0.03	0.55	0.33
3,5-Octadien-2-one	1485	a	101	28.88	0.00	0.77	0.01	0.21	0.12	101	29.52	0.02	1.56	0.10	0.60	0.31
Benzaldehyde	1494	a	102	29.03	0.12	0.49	2.34	0.18	0.04	104	29.95	0.47	1.06	15.40	0.44	0.22
6-Undecanone	1505	a	103	29.38	0.00	1.53	0.01	0.38	0.14	105	29.96	0.01	2.30	0.05	0.44	0.25
Propanoic acid	1506	a	104	29.40	0.01	0.14	0.06	0.35	0.09	106	30.29	0.02	0.96	0.12	0.47	0.23
(E)-2-Nonenal	1509	a	105	29.58	0.00	1.05	1.10	0.41	0.17	102	29.66	1.42	1.16	16.5	0.36	0.15
Linaloool	1514	a	106	29.75	0.00	0.67	1.73	0.18	0.06	103	29.90	0.26	0.72	2.90	0.38	0.12
1-Octanol	1525	a	107	30.16	0.00	0.57	2.01	0.14	0.05	107	30.69	0.25	1.57	2.40	0.16	0.22
Nonyl acetate	1567	b	108	31.15	0.00	1.33	0.01	0.25	0.16	108	30.96	0.00	1.29	0.06	0.40	0.28
5-Methyl-2(5H)-furanone	1589	b	109	31.46	0.11	0.46	0.00	0.54	0.07	109	32.49	0.24	1.02	3.08	0.62	0.25
Undecanal	1598	a	110	32.08	0.00	1.37	1.69	0.53	0.28	111	32.76	0.12	2.72	0.71	0.78	0.50
Butanoic acid	1610	a	111	32.57	0.10	0.16	0.00	0.39	0.09	110	32.64	0.42	1.45	7.74	0.47	0.23
Butyrolactone	1611	a	112	32.61	0.00	0.44	0.00	0.35	0.09	112	32.78	0.12	0.94	3.34	0.24	0.27

1-Nonanol	1628	a	113	33.66	0.00	0.63	1.82	0.16	0.05	113	34.22	0.11	1.43	2.20	0.20	0.23
Ethyl benzoate	1637	a	114	34.03	0.10	0.67	1.71	0.12	0.06	114	34.53	0.84	1.35	10.90	0.40	0.28
5-Ethyl-2(5H)-furanone	1643	b	115	34.28	0.20	0.41	10.2	0.25	0.16	115	35.04	0.61	1.24	8.38	0.36	0.33
(Z)-3-Nonen-1-ol	1647	a	116	34.44	0.00	0.57	0.02	0.21	0.07	116	35.04	0.01	1.11	0.26	0.44	0.30
Dodecanal	1684	a	117	35.64	0.00	1.42	1.41	0.51	0.21	117	36.25	0.38	2.49	8.17	0.47	0.36
α-Muurolene	1701	a	118	35.99	0.00	1.61	1.89	0.16	0.13	119	36.86	0.28	2.52	12.80	0.40	0.30
3,4-Dimethyl-2,5-furandione	1701	b	119	35.99	0.00	0.54	0.02	0.25	0.11	120	36.96	0.01	1.41	0.12	0.55	0.28
Pentanoic acid	1710	a	120	36.16	0.00	0.17	0.21	0.32	0.20	118	36.80	0.01	0.94	0.14	0.60	0.34
1,4-Cyclohex-2-enedione	1710	b	121	36.16	0.00	0.43	0.03	0.33	0.19	121	37.13	0.00	1.28	0.21	0.49	0.22
α-Farnesene	1725	a	122	36.98	0.00	1.43	0.02	0.15	0.06	122	37.13	0.01	2.16	0.11	0.28	0.15
(E,E)-2,4-Decadienal	1732	a	123	37.20	0.09	0.91	2.55	0.37	0.20	123	37.64	0.10	0.95	3.20	0.64	0.35
1-Decanol	1737	a	124	37.37	0.04	0.69	0.72	0.15	0.07	124	37.77	0.22	1.61	0.98	0.50	0.13
Methyl salicylate	1743	b	125	37.57	0.00	0.57	2.04	0.25	0.11	125	38.11	1.57	1.16	9.73	0.56	0.41
2-(2-Butoxyethoxy)-ethanol	1754	a	126	37.92	0.00	0.52	0.00	0.25	0.16	126	38.39	0.70	1.31	5.33	0.31	0.35
δ-Pentalactone	1767	b	127	38.33	0.00	0.52	0.00	0.39	0.16	127	39.36	0.00	1.18	2.74	0.60	0.32
3-Phenyl-2-propenal	1794	b	128	39.20	0.00	0.52	0.00	0.21	0.07	128	40.26	0.10	1.14	3.51	0.40	0.29
Hexanoic acid	1805	a	129	39.51	0.09	0.20	0.00	0.31	0.22	129	40.47	0.10	0.83	7.78	0.58	0.43
(Z)-6,10-Dimethyl-5,9-undecadien-2-one	1821	b	130	40.02	0.00	1.13	2.05	0.35	0.20	130	40.70	0.09	2.12	0.99	0.31	0.29
1-Undecanol	1824	a	131	40.11	0.09	0.75	0.86	0.21	0.11	131	41.19	0.19	1.64	4.80	0.29	0.16
Butyl benzoate	1830	b	132	40.31	0.00	0.82	0.02	0.41	0.27	132	41.27	0.00	1.69	0.14	0.59	0.30
Benzyl alcohol	1832	a	133	40.37	0.00	0.28	0.00	0.28	0.14	133	41.29	0.00	0.99	0.26	0.46	0.27
Heptanoic acid	1845	a	134	40.91	0.18	0.23	1.63	0.33	0.15	134	41.38	0.88	0.98	7.93	0.76	0.27
Phenylethyl alcohol	1869	a	135	41.44	0.16	0.34	0.00	0.29	0.11	135	41.42	0.16	0.83	12.00	0.56	0.34
4-Phenyl-3-buten-2-one	1914	b	136	42.82	0.00	0.56	0.00	0.40	0.14	136	43.77	0.09	1.12	1.61	0.64	0.25
1-Dodecanol	1928	a	137	43.23	0.00	0.81	1.42	0.23	0.13	137	43.89	0.32	1.68	8.43	0.38	0.31
Phenol	1956	a	138	44.04	0.00	1.27	3.97	0.43	0.34	138	44.84	1.51	1.66	18.3	0.53	0.41
Average			-	-	*0.11*	-	*0.87*	*0.24*	*0.12*	-	-	*0.97*	-	*5.39*	*0.42*	*0.25*
Min			-	-	*0.00*	-	*0.00*	*0.08*	*0.04*	-	-	*0.00*	-	*0.00*	*0.16*	*0.11*
Max			-	-	*1.11*	-	*10.20*	*0.54*	*0.34*	-	-	*8.45*	-	*19.40*	*1.04*	*0.50*

Based on the differences observed in the absolute response, spectral quality fluctuations were expected especially for low-intensity or threshold peaks. Therefore, the next step was to define threshold parameters for template construction with the objective of achieving 100% true-positive matches (accuracy) in presence of random variability over the application of the method in a short term and with the same method set-up. Therefore, these tests were performed between analytical replicates of the same oil sample to define benchmark values and then validated over a new sample acquired in the same conditions. Tests were done on the three analytical replicates of *Baza-4-A* acquired by *Set-up 1*; validation was on the three replicates of *Bena-4-A* acquired by the same set-up.

Threshold values for candidate peaks populating a template were set for SNR and NIST Similarity match factor (direct match factor - DMF); as reference, spectral signatures were tested for the average spectrum (named *blob spectrum*) and the highest modulation spectrum (named *peak spectrum*). SNR values were varied step-wise in a range between 10 and 100, covering 2D peaks with a percent response between 0.01 and 0.04, while the DMF threshold was set at 800 or 700.[42] Templates were built with ten 2D-peaks with SNR within the fixed range and homogeneously distributed over the 2D space. Experimental results, expressed as % of true-positive matches, are reported in **Table 5.3.4**.

Table 5.3.4. template matching results based on Set-up 1 templates including 2D peaks with SNR ranging from 10 to 100. Similarity DMF threshold applied are 800 or 700 while reference spectrum is blob (average 2D-peak spectrum) or peak (highest modulation spectrum). The upper part of the table refers to benchmark values obtained by applying Set-up 1 templates over replicated analyses of the same sample (*i.e.*, Baza-4-A); lower part of the table refers about results of Set-up 1 templates over replicated analyses of another sample (*i.e.*, Bena-4-A).

***Baza-4-A* Set-up *1* – three replicates**

			Similarity threshold 800 – *Blob* MS		Similarity threshold 700 – *Peak* MS	
SNR	**% Response**	**Peaks n°**	**% Matching**	**Matched peaks n°**	**% Matching**	**Matched peaks n°**
10 ± 2	0.01	Column bleeding or interferences				
30 ± 2	0.02	10	10.00 %	1	40.00 %	4
50 ± 2	0.02	10	10.00 %	1	90.00 %	9
70 ± 2	0.03	10	40.00 %	4	100.00 %	10
90 ± 2	0.04	10	40.00 %	4	100.00 %	10
100 ± 2	0.04	10	70.00 %	7	100.00 %	10

Baza 4-A* Set-up *1* over *Bena-4-A* Set-up *1

			Similarity threshold 800 – *Blob* MS		Similarity threshold 700 – *Peak* MS	
SNR	**% Response**	**Peaks n°**	**% Matching**	**Matched peaks n°**	**% Matching**	**Matched peaks n°**
10 ± 2	0.01	Column bleeding or interferences				
30 ± 2	0.02	10	10.00 %	1	10.00 %	1
50 ± 2	0.02	10	10.00 %	1	20.00 %	1
70 ± 2	0.03	8	25.00 %	2	87.50 %	7
90 ± 2	0.04	5	40.00 %	2	100.00 %	5
100 ± 2	0.04	7	42.86 %	3	100.00 %	7

2D Peaks with SNR values below 50 are connoted by inconsistent MS spectral signatures resulting in false-negative matches even when DMF threshold is lowered from 800 to 700. For these peaks, neither the *blob spectrum* nor the *peak spectrum* are sufficiently reliable to carry consistent information for identity confirmation. For 2D-peaks with a SNR within 50-100 in the reference chromatogram, the rate of positive matches increases from 10% to 70% when MS constraints are kept at 800 DMF and *blob spectrum* considered. The rate of true-positive matches reaches 100% with the combination of lower DMF threshold at 700 and *peak spectrum* taken as reference MS. Note, in these cases no false positives were revealed, meaning that correspondences were just established between peaks generated by the same chemical entity.

Results suggest that a SNR cut-off should be defined, based on 2D data particulars, to limit inconsistencies at targeted identity confirmation level. The validation of these settings was by

applying the templates resulting from the reference sample *Baza-4-A* to *Bena-4-A* replicates on the same column set-up. Results, reported in **Table 5.3.4**, confirmed the need for a SNR threshold of at least 50 with better performances at DMF threshold of 700 with a reference template spectrum collected from the highest modulation (*peak spectrum*). In this case, some template peaks were unmatched (true-negative matches) because they were not detected in the analyzed sample (below method Limit of Detection LOD).

Rules for template peaks thresholds and reference spectra were then applied to build a reference *targeted template* of known analytes. This fully supervised approach aimed at characterizing the distribution of marker compounds known for their role in defining olive oil sensory quality or to refer about olives ripening status[8,9]. It was also the complement of the *UT* fingerprinting process and included targeted peaks listed in **Table 5.3.3**. Analytes identifications were by combining ^{1}D retention data (I^{T}) with MS fragmentation pattern similarity above 900 DMF adopting commercial [42,43] and in-house databases or, when possible, by authentic standards.

The template of 126 2D-peaks was built by inspecting samples patterns obtained with *Set-up 1*; reference peaks inclusion was limited to those analytes showing a SNR≥100, the reference spectra was from the highest modulation (*e.g.*, *peak spectrum*), and the MS constraint was set at 700 DMF and 700 Reverse Match Factor (RMF).[42]

Results for targeted template matching are summarized in **Table 5.3.5**. The average matching within *Set-up 1* samples achieved 97% (122 over 126 peaks). Further comments will follow for template matching for *Set-up 2* patterns.

Table 5.3.5. template matching results for targeted template (supervised workflow) from Set-up 1 applied to Set-up 2 samples and for feature template (unsupervised and automatic workflow) built over Set-up 1 samples and applied on Set-up 2 samples

	Targeted template (126 peaks)				***Feature* template (257 reliable peaks)**			
	Set-up 1		***Set-up 2***		***Set-up 1***		***Set-up 2***	
Samples	**%**	**Peaks n°**	**%**	**Peaks n°**	**%**	**Peaks n°**	**%**	**Peaks n°**
Baza-1	94.44	119	91.27	115	100.00	257	97.28	250
Baza-2	97.62	123	92.06	116	100.00	257	98.44	253
Baza-3	98.41	124	95.24	120	100.00	257	98.83	254
Baza-4	100.00	126	96.03	121	99.22	255	96.86	247
Bena-1	93.65	118	88.89	112	100.00	257	98.44	253
Bena-2	96.03	121	91.27	115	100.00	257	96.50	248
Bena-3	97.62	123	92.06	116	98.83	254	99.22	256
Bena-4	98.41	124	92.86	117	100.00	257	98.05	252
Average	*97.02*	*122*	*92.46*	*116*	*99.75*	*256*	*97.95*	*252*
RSD%	*2.22*	*2.22*	*2.47*	*2.47*	*0.45*	*0.46*	*0.99*	*1.22*

5.3.4.3 Template transformation

Once template construction was established with simple rules for confident identification and effective matching, the next step was the selection of matching algorithm (transform) and related parameters to effectively transform the template to match the peak pattern showing severe misalignment. To approach this challenge, global polynomial transformations were tested, as it was successful in complex re-alignment problems such as those posed by method translation from thermal to differential flow modulated platforms. [18,21] Global, low-degree transformation functions (second-degree or third-degree polynomials) are successful when a sufficient number of alignment points, at least 10 for affine and 30 for second-degree polynomial, are available to guide the pattern re-alignment.[18]

The strategy here applied included the re-alignment of the targeted template built over *Set-up 1* analyses to chromatograms from *Set-up 2*. The first step included the adjustment of the *distance*

threshold parameters for the ^{1}D and ^{2}D, which correspond to the horizontal and vertical distance threshold that limits the after-transformation distance between template and candidate 2D-peaks. These distances are expressed in inter-sample distances (*i.e.*, pixel dimensions). To compensate for the greater ^{2}D misalignment the 2D distance threshold was step-wise incremented from a minimum value of 5 up to 25. On the ^{1}D, where the misalignment was minimal, a threshold of 10 was sufficient to avoid false-negative matches.

At the same time, affine and second-degree polynomial matching transforms were tested for performance.[18] These functions are available in the current version of GC Image (*i.e.*, v2.8). The reader interested in a deeper insight on transform functions and their effectiveness to re-align complex patterns can refer to previous research by Reichenbach and co-workers.[17,18,35]

The *Set-up 1* targeted template was applied to the first (arbitrarily chosen) sample pattern (*i.e.*, *Baza-4-A*) obtained with *Set-up 2*. At first, the number of true-positive matches was higher for the second-degree transform, so the iterative process of *match-and-transform* was continued. Iterating the process of matching and template transform allows the template to be adapted to the actual pattern while increasing, step-by-step, the number of matched peaks up to the maximum number that corresponds to all targeted analytes actually detected/confirmed in the analyzed pattern. In practice, this operation increments the number of alignments points available step-by-step, thus improving the quality of the global template transformation at each step.

Experimental results for the application of the targeted template adapted to *Set-up 1* over the patterns of *Set-up 2* resulted in a 65 positive matches over 126 template peaks (51% - 5.4% RSD); after transformation by taking these 65 alignment points, the successive matching step achieved 95 positive matches with a 75% of matched peaks; then, the next step matched 110 peaks, 87%. The maximum number of matched peaks, shown in **Table 5.3.5**, was 121 (96%) and was achieved after one additional matching step. In practice, for a full and effective re-alignment of the targeted template, a variable number of iterations between 3 and 5 was applied.

All such results are listed in **Table 5.3.5**; benchmark values for maximum template matching performance are those corresponding to the application of the targeted template to *Set-up 1* samples (first column). On average matching performance was better for *Set-up 1* patterns (97% of true-positive matches), although the loss of accuracy on *Set-up 2* pattern was just 5% (92.46% *vs.* 97.02%). These results present a solid foundation for the application of this experimented strategy to a fully unsupervised approach as that for the reliable template construction.

The next section illustrates the process of *feature template* construction over samples patterns from *Set-up 1* and its successive alignment over *Set-up 2* patterns. Accuracy results are discussed as % of true-positive matches. [44]

5.3.4.4 UT fingerprinting: feature template construction

The feature template was built over a subset of *Set-up 1* chromatograms with the first analytical replicate of all analyzed samples. The 2D chromatograms were pre-processed for rasterization, background subtraction, and 2D peaks detection above a SNR threshold of 100. Data processing was then conducted within a component of the GC Image software suite (Image Investigator™) using the previously validated settings:

- SNR≥100 as threshold value for template peaks;
- *peak spectrum* as MS reference to upload in the template;
- DMF similarity threshold at 700;
- matching transform parameters with ^{1}D distance threshold of 10 and ^{2}D of 25.

Additional settings, specific for this process, included an option for reliable peak inclusion that was set as *most relaxed*: with this setting the algorithm considers as *reliable peaks* all 2D-peaks that

match across at least half of the chromatograms. Reliable peaks are fundamental as they are used as alignment points for the transform function when the *feature template* is used to cross-align a large set of chromatograms including those obtained with a different set-up.

The final *feature template* accounted for 257 reliable peaks and 1500 peak-regions. The reliable template built over *Set-up 1* chromatograms is shown in Supplementary material – **Supplementary Figure 5.3.3**. Matching constraints for MS spectrum similarity were applied on 2D-peak features and results are summarized in **Table 5.3.5**. The average % of matching for *Set-up 1* chromatograms was 99.75 (±0.45 RSD%); when the feature template is transformed to match for *Set-up 2* patterns, the % of matching is slightly lower 97.95 (±0.99 RSD%) but evidences the high accuracy of the process.

Once re-aligned almost all chemical features are detected in all samples patterns, proceeding in a sort of data fusion, the final step aimed at defining the best 2D-peak response descriptor for cross-sample analysis.

5.3.4.5 Response normalization and samples clustering

As evidenced by signal intensity evaluation and by 2D-peaks statistics (SNR and VNR distribution), pattern cross-alignment is not sufficient to compensate for random variations across measurement sessions and impacting detector response. Response normalization is mandatory to allow consistent cross-comparison of data set. The removal of unwanted intensity variation (*i.e.*, normalization) is referred to as signal drift correction, removal of batch effect, scaling, and matrix effects removal, and can be approached differently as function of the study objectives.[23] Normalization is, in fact, a fundamental step because it may affect the outcomes of a study; the meaningfulness of differentially abundant analytes may vary depending on the normalization method.[23,40]

In this study, we tested three simple approaches within those generally adopted in volatiles profiling studies.[3,45] The first included multiple IS normalization with α and β thujone and methyl octynoate that were pre-loaded into the SPME device before headspace sampling of olive oils.[30] As an alternative method, the analyte % response (calculated on the 2D-Volume) was considered. It was obtained by: (a) normalizing analyte 2D-Volume over all detected 2D-peaks above fixed thresholds or (b) normalizing analyte 2D-Volume over all *UT* peaks included in the feature template. In this last approach 2D-peaks from column bleeding and from interferents were excluded.

Results are illustrated, for a selection of informative analytes covering different volatilities, polarities and amounts, in **Figure 5.3.3**.

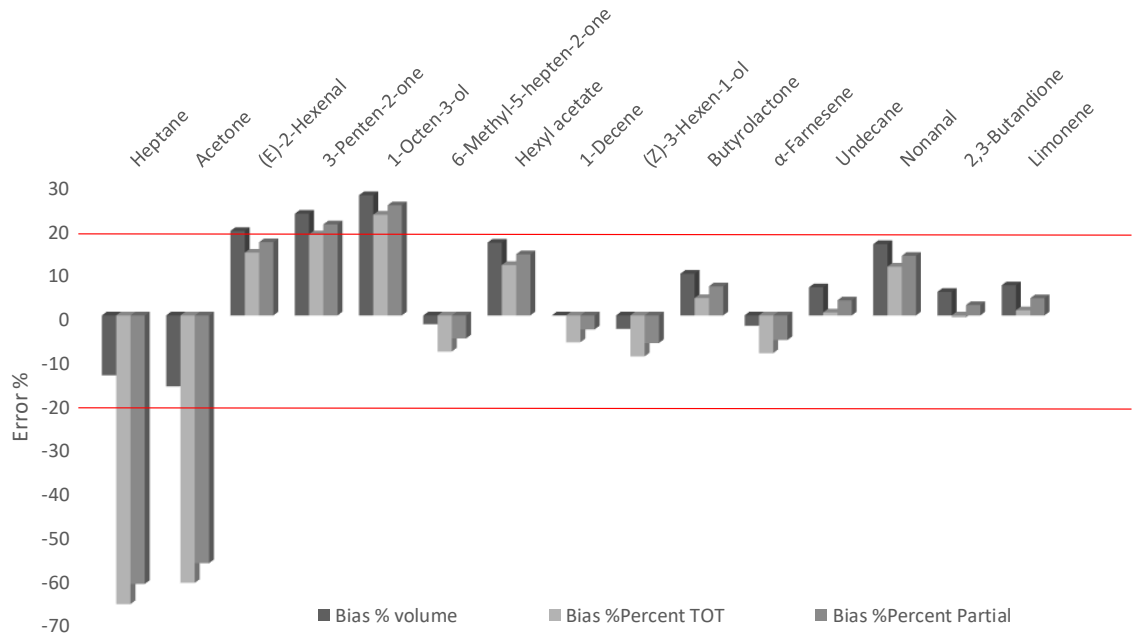

Figure 5.3.3. Error %, according to Equation 3, calculated between 2D peaks response (Normalized 2D Volume over IS, % response over all 2D peaks detected or over UT peaks) taking Set-up 2 as reference for a selection of informative analytes. Red dotted lines indicate boundaries for acceptance at ±20%.

The bias is computed as Error %, according to **Equation 5.3.3**, and between 2D peaks response (Normalized 2D Volume over IS, % response over all 2D peaks detected or over UT peaks) taking *Set-up 2* as reference.

Equation 5.3.3. $\text{Error \%} = ((Response_i\ Set\text{-}up\ 2 - Response_i\ Set\text{-}up\ 1)/\ Response_i\ Set\text{-}up\ 2) * 100$

Normalization by IS(s) performs, on average, more effectively than those on % response (Average Error % 11.2 vs. 16.2); it better compensates for response fluctuations derived by S/SL injection discriminations here impacting on highly volatiles (heptane and acetone) and due to the different operative head-pressures applied to the two set-up. For highly volatile analytes, the Error % computed for IS normalization drops below 16 while for % response it reaches 50-60%. On the other hand, % response indicators well compensate for detection variability on less abundant analytes as 1-octen-3 ol (0.01%), 2-butanone (0.01%), 3-penten-2-one (0.02%), 1-octanol (0.04%), and octanal (0.06%). Note, response variations also are influenced by chromatographic performance; 2D-peaks showing long tails (carbonyl derivatives or unsaturated alcohols) or distorted by overloading phenomena may be split into multiple 2D-peaks. In these cases, supervision is needed to merge all 2D-peaks belonging to the same analytes in a single one.

Although ISs normalization gave better results for analytes heavily discriminated, it requires a dedicated sample preparation with ISs pre-loading before sampling that may impact on the global analysis time. In addition, a careful selection of standards is necessary by focusing on compounds non-present in the samples under study while covering the suitable range of volatility and polarity. In this scenario, % response normalization is attractive being simpler and less time consuming although it does not rule out the use of quality control procedures as for example external standardization (ES) or multiple quality control samples analysis.

As a proof of concept, a PCA was run on normalized responses obtained from samples analyzed by the two set-ups. The data matrix consisted of 126 chemical variables corresponding to the targeted analytes listed in **Table 5.3.2**, and 16 samples corresponding to 2 plots × 4 ripening stages × 2 set-up. Technical replicates were averaged; therefore, the final data matrix was 126 × 16 dimensioned. Results are shown in the score plot of **Figure 5.3.4A**; observations (samples) are rationally distributed over the Cartesian plane according to the ripening stage of olives (visible along the F1 from right to left) and in accordance with oil quality (*i.e.* extra-virgin, virgin or *lampante*). Results are in agreement with those from the original study by Magagna et al.[8] confirming the informative role of several analytes related to positive aroma attributes (*i.e.*, (*E*)-2-Hexen-1-ol, (*Z*)-3-Hexen-1-ol and (*E,E*)-2,4-Hexadienal) with a relevant role in early ripening stages and 6-Methyl-5-hepten-2-one with higher impact on later stages were oils were classified as *lampante* or virgin.

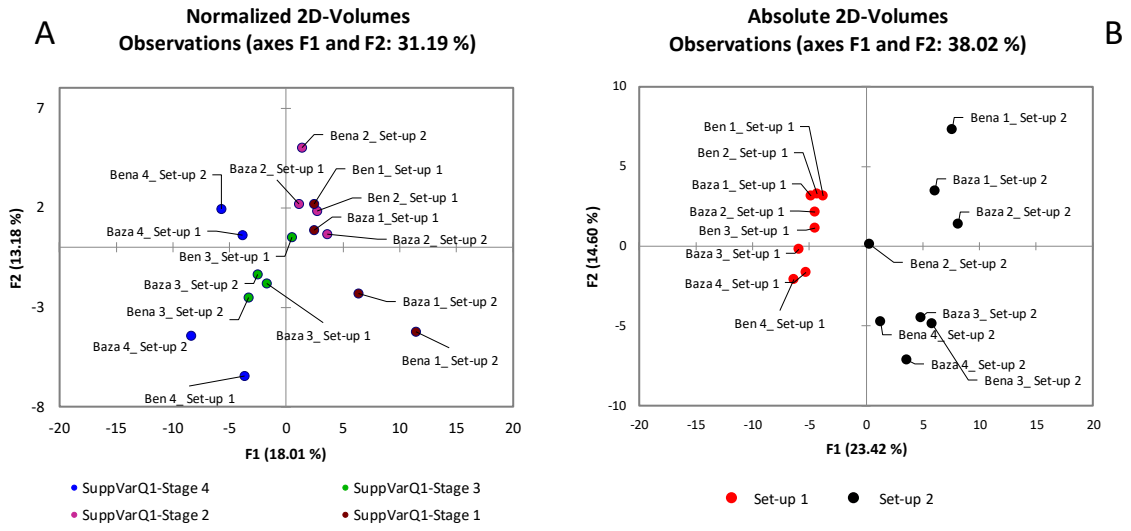

Figure 5.3.4. PCA score plots based on normalized responses for 126 targeted analytes (**Table 5.3.2**) resulting from samples (Baza and Benamaurel plots) analyzed with set-up 1 and set-up 2 (**5.3.4A**), colors correspond to different ripening stages of olives (1-4). (**5.3.4B**). Scores plot in based on analytes absolute responses without normalization while confidence ellipses clearly highlight the samples clusters influenced by batch effect (set-up 1 red indicators and set-up 2 black indicators.

Measurements from the two set-ups are homogeneously distributed over the Cartesian plane confirming the effectiveness of response normalization on the "batch effect". The latter is evident in **Figure 5.3.4B** where the PCA was conducted on analytes 2D volumes without normalization. Here measurements from the two set-ups are independently clustered, as also indicated by confidence ellipses, and well discriminated along the F1. On the other side, as a conformation on the consistency of the measurements obtained by the two set-up, along F2 for both clusters, the ripening stage is coherently distributed from the upper-side of the graph to bottom.

5.3.5 Supplementary Material

Supplementary material at the Google Drive's link:
https://drive.google.com/drive/folders/1dw3d3BvjJIrAmzTWl_qhkf1GFTmHyIBF?usp=sharing

References

(1) Miguel, H.; Carolina, S.; Virginia, G.; Elena, I.; Alejandro, C. Foodomics: MS-based Strategies in Modern Food Science and Nutrition. *Mass Spectrom. Rev.* **2011**, *31* (1), 49–69. https://doi.org/10.1002/mas.20335.

(2) Tranchida, P. Q.; Donato, P.; Cacciola, F.; Beccaria, M.; Dugo, P.; Mondello, L. Potential of Comprehensive Chromatography in Food Analysis. *TrAC Trends Anal. Chem.* **2013**, *52*, 186–205. https://doi.org/10.1016/j.trac.2013.07.008.

(3) Cordero, C.; Kiefl, J.; Reichenbach, S. E.; Bicchi, C. Characterization of Odorant Patterns by Comprehensive Two-Dimensional Gas Chromatography: A Challenge in Omic Studies. *TrAC Trends Anal. Chem.* **2019**, *113*, 364–378. https://doi.org/10.1016/j.trac.2018.06.005.

(4) Cordero, C.; Kiefl, J.; Schieberle, P.; Reichenbach, S. E.; Bicchi, C. Comprehensive Two-Dimensional Gas Chromatography and Food Sensory Properties: Potential and Challenges. *Anal. Bioanal. Chem.* **2015**, *407* (1), 169–191. https://doi.org/10.1007/s00216-014-8248-z.

(5) Reichenbach, S. E.; Tian, X.; Tao, Q.; Ledford, E. B.; Wu, Z.; Fiehn, O. Informatics for Cross-Sample Analysis with Comprehensive Two-Dimensional Gas Chromatography and High-Resolution Mass Spectrometry (GCxGC-HRMS). *Talanta* **2011**, *83* (4), 1279–1288. https://doi.org/10.1016/j.talanta.2010.09.057.

(6) Cordero, C.; Liberto, E.; Bicchi, C.; Rubiolo, P.; Schieberle, P.; Reichenbach, S. E.; Tao, Q. Profiling Food Volatiles by Comprehensive Two-Dimensional Ga Schromatography Coupled with Mass Spectrometry: Advanced Fingerprinting Approaches for Comparative Analysis of the Volatile Fraction of Roasted Hazelnuts (Corylus Avellana L.) from Different Ori. *J. Chromatogr. A* **2010**, *1217* (37), 5848–5858. https://doi.org/10.1016/j.chroma.2010.07.006.

(7) Kiefl, J.; Schieberle, P. Evaluation of Process Parameters Governing the Aroma Generation in Three Hazelnut Cultivars (Corylus Avellana L.) by Correlating Quantitative Key Odorant Profiling with Sensory Evaluation. *J. Agric. Food Chem.* **2013**, *61* (22), 5236–5244. https://doi.org/10.1021/jf4008086.

(8) Magagna, F.; Valverde-Som, L.; Ruíz-Samblás, C.; Cuadros-Rodríguez, L.; Reichenbach, S. E.; Bicchi, C.; Cordero, C. Combined Untargeted and Targeted Fingerprinting with Comprehensive Two-Dimensional Chromatography for Volatiles and Ripening Indicators in Olive Oil. *Anal. Chim. Acta* **2016**, *936*, 245–258. https://doi.org/10.1016/j.aca.2016.07.005.

(9) Purcaro, G.; Cordero, C.; Liberto, E.; Bicchi, C.; Conte, L. S. Toward a Definition of Blueprint of Virgin Olive Oil by Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2014**, *1334*, 101–111. https://doi.org/10.1016/j.chroma.2014.01.067.

(10) Mondello, L.; C. Lewis, A.; D. Bartle, K. *Multidimensional Chromatography*; Wiley-VCH: New York, 2001. https://doi.org/10.1002/0470845775.

(11) Etxebarria, N.; Zuloaga, O.; Olivares, M.; Bartolomé, L. J.; Navarro, P. Retention-Time Locked Methods in Gas Chromatography. *J. Chromatogr. A* **2009**, *1216* (10), 1624–1629. https://doi.org/10.1016/j.chroma.2008.12.038.

(12) Klee, M. S.; Quimby, B. D.; Blumberg, L. M. No Title. US Patent 5,987,959 A, 1999.

(13) Blumberg, L. M. No Title. US Patent 6,634,211 B1, 2003.

(14) Jiang, W.; Zhang, Z. M.; Yun, Y.; Zhan, D. J.; Zheng, Y. B.; Liang, Y. Z.; Yang, Z. Y.; Yu, L. Comparisons of Five Algorithms for Chromatogram Alignment. *Chromatographia* **2013**, *76* (17–18), 1067–1078. https://doi.org/10.1007/s10337-013-2513-8.

(15) Niu, W.; Knight, E.; Xia, Q.; McGarvey, B. D. Comparative Evaluation of Eight Software Programs for Alignment of Gas Chromatography-Mass Spectrometry Chromatograms in Metabolomics Experiments. *J. Chromatogr. A* **2014**, *1374*, 199–206. https://doi.org/10.1016/j.chroma.2014.11.005.

(16) Stanimirova, I.; Daszykowski, M. Exploratory Analysis of Metabolomic Data. *Compr. Anal. Chem.* **2018**, *82*, 227–264. https://doi.org/10.1016/bs.coac.2018.08.005.

(17) Reichenbach, S. E.; Carr, P. W.; Stoll, D. R.; Tao, Q. Smart Templates for Peak Pattern Matching with Comprehensive Two-Dimensional Liquid Chromatography. *J. Chromatogr. A* **2009**, *1216* (16), 3458–3466. https://doi.org/10.1016/j.chroma.2008.09.058.

(18) Rempe, D. W.; Reichenbach, S. E.; Tao, Q.; Cordero, C.; Rathbun, W. E.; Zini, C. A. Effectiveness of Global, Low-Degree Polynomial Transformations for GCxGC Data Alignment. *Anal. Chem.* **2016**, *88* (20), 10028–10035. https://doi.org/10.1021/acs.analchem.6b02254.

(19) Reichenbach, S. E.; Rempe, D. W.; Tao, Q.; Bressanello, D.; Liberto, E.; Bicchi, C.; Balducci, S.; Cordero, C. Alignment for Comprehensive Two-Dimensional Gas Chromatography with Dual Secondary Columns and Detectors. *Anal. Chem.* **2015**, *87* (19), 10056–10063. https://doi.org/10.1021/acs.analchem.5b02718.

(20) Magagna, F.; Liberto, E.; Reichenbach, S. E.; Tao, Q.; Carretta, A.; Cobelli, L.; Giardina, M.; Bicchi, C.; Cordero, C. Advanced Fingerprinting of High-Quality Cocoa: Challenges in Transferring Methods from Thermal to Differential-Flow Modulated Comprehensive Two Dimensional Gas Chromatography. *J. Chromatogr. A* **2018**, *1535*, 122–136. https://doi.org/10.1016/j.chroma.2017.07.014.

(21) Cordero, C.; Rubiolo, P.; Reichenbach, S. E.; Carretta, A.; Cobelli, L.; Giardina, M.; Bicchi, C. Method Translation and Full Metadata Transfer from Thermal to Differential Flow Modulated Comprehensive Two Dimensional Gas Chromatography: Profiling of Suspected Fragrance Allergens. *J. Chromatogr. A* **2017**, *1480*, 70–82. https://doi.org/10.1016/j.chroma.2016.12.011.

(22) Mommers, J.; Ritzen, E.; Dutriez, T.; van der Wal, S. A Procedure for Comprehensive Two-Dimensional Gas Chromatography Retention Time Locked Dual Detection. *J. Chromatogr. A* **2016**, *1461*, 153–160. https://doi.org/10.1016/j.chroma.2016.07.052.

(23) De Livera, A. M.; Sysi-Aho, M.; Jacob, L.; Gagnon-Bartsch, J. A.; Castillo, S.; Simpson, J. A.; Speed, T. P. Statistical Methods for Handling Unwanted Variation in Metabolomics Data. *Anal. Chem.* **2015**, *87* (7), 3606–3615. https://doi.org/10.1021/ac502439y.

(24) Progetto Ager. *Violin Project*; 2016.

(25) Vaz-Freire, L. T.; da Silva, M. D. R. G.; Freitas, A. M. C. Comprehensive Two-Dimensional Gas Chromatography for Fingerprint Pattern Recognition in Olive Oils Produced by Two Different Techniques in Portuguese Olive Varieties Galega Vulgar, Cobrançosa e Carrasquenha. *Anal. Chim. Acta* **2009**, *633* (2), 263–270. https://doi.org/10.1016/j.aca.2008.11.057.

(26) Reichenbach, S. E.; Tian, X.; Cordero, C.; Tao, Q. Features for Non-Targeted Cross-Sample

Analysis with Comprehensive Two-Dimensional Chromatography. *J. Chromatogr. A* **2012**, *1226*, 140–148. https://doi.org/10.1016/j.chroma.2011.07.046.

(27) Cajka, T.; Riddellova, K.; Klimankova, E.; Cerna, M.; Pudil, F.; Hajslova, J. Traceability of Olive Oil Based on Volatiles Pattern and Multivariate Analysis. *Food Chem.* **2010**, *121* (1), 282–289. https://doi.org/10.1016/j.foodchem.2009.12.011.

(28) Lukić, I.; Carlin, S.; Horvat, I.; Vrhovsek, U. Combined Targeted and Untargeted Profiling of Volatile Aroma Compounds with Comprehensive Two-Dimensional Gas Chromatography for Differentiation of Virgin Olive Oils According to Variety and Geographical Origin. *Food Chem.* **2019**, *270* (March 2018), 403–414. https://doi.org/10.1016/j.foodchem.2018.07.133.

(29) Setkova, L.; Risticevic, S.; Linton, C. M.; Ouyang, G.; Bragg, L. M.; Pawliszyn, J. Solid-Phase Microextraction-Gas Chromatography-Time-of-Flight Mass Spectrometry Utilized for the Evaluation of the New-Generation Super Elastic Fiber Assemblies. *Anal. Chim. Acta* **2007**, *581* (2), 221–231. https://doi.org/10.1016/j.aca.2006.08.022.

(30) Wang, Y.; O'Reilly, J.; Chen, Y.; Pawliszyn, J. Equilibrium In-Fibre Standardisation Technique for Solid-Phase Microextraction. *J. Chromatogr. A* **2005**, *1072* (1), 13–17. https://doi.org/10.1016/j.chroma.2004.12.084.

(31) Bressanello, D.; Liberto, E.; Collino, M.; Chiazza, F.; Mastrocola, R.; Reichenbach, S. E.; Bicchi, C.; Cordero, C. Combined Untargeted and Targeted Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography: Revealing Fructose-Induced Changes in Mice Urinary Metabolic Signatures. *Anal. Bioanal. Chem.* **2018**, *410* (11), 2723–2737. https://doi.org/10.1007/s00216-018-0950-9.

(32) GC Image™. *GC Image GCxGC Edition Users' Guide*; 2017.

(33) Cordero, C.; Liberto, E.; Bicchi, C.; Rubiolo, P.; Reichenbach, S. E.; Tian, X.; Tao, Q. Targeted and Non-Targeted Approaches for Complex Natural Sample Profiling by GCxGC-QMS. *J. Chromatogr. Sci.* **2010**, *48* (4), 251–261.

(34) Magagna, F.; Guglielmetti, A.; Liberto, E.; Reichenbach, S. E.; Allegrucci, E.; Gobino, G.; Bicchi, C.; Cordero, C. Comprehensive Chemical Fingerprinting of High-Quality Cocoa at Early Stages of Processing: Effectiveness of Combined Untargeted and Targeted Approaches for Classification and Discrimination. *J. Agric. Food Chem.* **2017**, *65* (30), 6329–6341. https://doi.org/10.1021/acs.jafc.7b02167.

(35) Reichenbach, S. E.; Tian, X.; Boateng, A. A.; Mullen, C. A.; Cordero, C.; Tao, Q. Reliable Peak Selection for Multisample Analysis with Comprehensive Two-Dimensional Chromatography. *Anal. Chem.* **2013**, *85* (10), 4974–4981. https://doi.org/10.1021/ac303773v.

(36) Cordero, C.; Rubiolo, P.; Cobelli, L.; Stani, G.; Miliazza, A.; Giardina, M.; Firor, R.; Bicchi, C. Potential of the Reversed-Inject Differential Flow Modulator for Comprehensive Two-Dimensional Gas Chromatography in the Quantitative Profiling and Fingerprinting of Essential Oils of Different Complexity. *J. Chromatogr. A* **2015**, *1417*, 79–95. https://doi.org/10.1016/j.chroma.2015.09.027.

(37) Cordero, C.; Rubiolo, P.; Sgorbini, B.; Galli, M.; Bicchi, C. Comprehensive Two-Dimensional Gas Chromatography in the Analysis of Volatile Samples of Natural Origin: A Multidisciplinary Approach to Evaluate the Influence of Second Dimension Column Coated with Mixed Stationary Phases on System Orthogonality. *J. Chromatogr. A* **2006**, *1132* (1–2),

268–279. https://doi.org/10.1016/j.chroma.2006.07.067.

(38) Venkatramani, C. J.; Xu, J.; Phillips, J. B. Separation Orthogonality in Temperature-Programmed Comprehensive Two-Dimensional Gas Chromatography. *Anal. Chem.* **1996**, *68* (9), 1486–1492. https://doi.org/10.1021/ac951048b.

(39) Markes International. Select-EV: The next Generation of Ion Source Technology. *Technical Note*. 2016.

(40) Reichenbach, S. E.; Ni, M.; Zhang, D.; Ledford, E. B. Image Background Removal in Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2003**, *985* (1–2), 47–56. https://doi.org/10.1016/S0021-9673(02)01498-X.

(41) Cordero, C.; Guglielmetti, A.; Bicchi, C.; Liberto, E.; Baroux, L.; Merle, P.; Tao, Q.; Reichenbach, S. E. Comprehensive Two-Dimensional Gas Chromatography Coupled with Time of Flight Mass Spectrometry Featuring Tandem Ionization: Challenges and Opportunities for Accurate Fingerprinting Studies. *J. Chromatogr. A* **2019**, *1597*, 132–141. https://doi.org/10.1016/j.chroma.2019.03.025.

(42) NIST/EPA/NIH Mass Spectral Library with Search Program Data Version: NIST V17.

(43) Adams, R. P. *Identification of Essential Oil Components by Gas Chromatography—Mass Spectroscopy*; Allured Publishing: New York, 1995.

(44) Kiefl, J.; Cordero, C.; Nicolotti, L.; Schieberle, P.; Reichenbach, S. E.; Bicchi, C. Performance Evaluation of Non-Targeted Peak-Based Cross-Sample Analysis for Comprehensive Two-Dimensional Gas Chromatography-Mass Spectrometry Data and Application to Processed Hazelnut Profiling. *J. Chromatogr. A* **2012**, *1243*, 81–90. https://doi.org/10.1016/j.chroma.2012.04.048.

(45) Sgorbini, B.; Cagliero, C.; Liberto, E.; Rubiolo, P.; Bicchi, C.; Cordero, C. Strategies for Accurate Quantitation of Volatiles from Foods and Plant-Origin Materials: A Challenging Task. *J. Agric. Food Chem.* **2019**, acs.jafc.8b06601. https://doi.org/10.1021/acs.jafc.8b06601.

5.4 A step forward in the equivalence between thermal and differential-flow modulated comprehensive two-dimensional gas chromatography methods

Federico Stilo[1], Elena Gabetti[1], Carlo Bicchi[1], Andrea Carretta[2], Daniela Peroni[2], Stephen E. Reichenbach[3,4], Chiara Cordero[1*], James Mc Curry[5]

[1] Università degli Studi di Torino Turin - Italy E-M@il: chiara.cordero@unito.it

[2] SRA Intruments SpA, Cernusco sul Naviglio, Milan, Italy

[3] University of Nebraska-Lincoln, NE USA

[4] GC Image LLC, Lincoln, NE USA

[5] Agilent Technologies, Gas Phase Separations Division, Wilmington DE, USA

* Corresponding author:

Prof. Dr. Chiara Cordero - Dipartimento di Scienza e Tecnologia del Farmaco, Università di Torino, Via Pietro Giuria 9, I-10125 Torino, Italy – e-mail: chiara.cordero@unito.it ; phone: +39 011 6702197

Received: May 22, 2020

Revised: July 6, 2020

Accepted: July 7, 2020

Published: July 8, 2020

DOI: 1 0.1016/j.chroma.2020.461396

Journal of Chromatography A 2020, 1627, 461396

5.4.1 Abstract

Comprehensive two-dimensional gas chromatography (GC×GC) based on flow-modulation (FM) is gaining increasing attention as an alternative to thermal modulation (TM), the recognized GC×GC benchmark, thanks to its lower operational cost and rugged performance. An accessible, rational procedure to perform method translation between the two platforms would be highly valuable to facilitate compatibility and consequently extend the flexibility and applicability of GC×GC. To enable an effective transfer, the methodology needs to ensure preservation of the elution pattern, separation power, and sensitivity.

Here, a loop-type thermal modulation system with dual detection (TM-GC×GC-MSD/FID) used for the targeted analysis of allergens in fragrances is selected as reference method. Initially, six different columns configurations are systematically evaluated for the flow-modulated counterpart. The set up providing the most consistent chromatographic separation (20 m x 0.18 mm d_c x 0.18 µm d_f + 1.8 m x 0.18 mm d_c x 0.18 µm d_f) is further evaluated to assess its overall performance in terms of sensitivity, linearity, accuracy, and pattern reliability. The experimental results convincingly show that the method translation procedure is effective and allows successful transfer of the target template metadata. Additionally, the FM-GC×GC-MSD/FID system is suitable for challenging applications such as the quantitative profiling of complex fragrance materials.

Key words

Two-dimensional comprehensive gas chromatography-mass spectrometry and flame ionization detection; reverse-inject differential flow modulation; suspected fragrance allergens; method translation; method limit of detection; repeatability and precision

5.4.2 Introduction

Comprehensive two-dimensional gas chromatography GC×GC coupled with mass spectrometry (MS) is a powerful technique for detailed profiling and effective fingerprinting of medium-to-high complexity samples. Thermal modulators implementing cryogenic cooling are widely used and, to date, considered the "gold standard" for GC×GC. The effective in-space band-focusing induced by this modulator results in a peak capacity gain (G_n) that is close to the achievable theoretical limit.[1] At the same time, the signal-to-noise ratio greatly increases resulting in a sensitivity gain of one order of magnitude compared to a conventional 1D-GC analysis. Despite these advantages, thermal modulators have some drawbacks related to hardware and operational costs limiting their widespread adoption in quality control and high-throughput screening.

Differential-flow modulators (FMs), such as those based on the Seeley *et al.* design,[2,3] are an interesting alternative to thermal modulators (TMs). Configurations can have an adjustable volume/length accumulation loop, as those proposed by Tranchida *et al.*[4–6] Large accumulation loop volumes limit the overloading, extends the re-injection period and provides multi-stage dynamics with some benefits on separation power and peak symmetry. The first commercial FM used fixed volume accumulation loop devices obtained with Capillary Flow Technology (CFT) microfluidic plates. They implement both the forward fill/flush (FFF) injection dynamics described by Seeley *et al.* [7] and the reverse fill/flush (RFF) dynamics connoted by a more efficient band re-injection, improved 2D peak widths and symmetry, and effective handling of collection-channel overloading.[8–12] Commercial RFF modulators are available from Agilent Technologies[13] and by Sep-Solve with the FM named Insight™.[14] More recently, Seeley *et al.*[15] proposed the multi-mode modulator (MMM). This device, as it is engineered in the commercial platforms by LECO (Flux™), enables the adoption of conventional column combinations and carrier gas operational flow in both separation dimensions but is characterized by a low duty cycle.

The growing interest in robust and cryogen-free modulators certainly is driven by the possibility they offer to describe in depth the chemical dimensionality of samples[16] with a relative ease of use and low operational costs. FM gives access to peculiar features of GC×GC separations such as group-type characterization, accurate profiling, and advanced fingerprinting based on 2D separation patterns.[17–21] However, FM dynamics are connoted by a limited flexibility in terms of operative flows in the two separation dimensions which in turn require a careful selection of column dimensions/characteristics to fully exploit the separation potential.

If the price to pay is mainly related to the actual separation power of the system, absolute method sensitivity is another important issue to consider; to date this method characteristic lacks dedicated research especially in the perspective of application transfer between TM to FM platforms. This study fills this gap by systematically examining six different column combinations, almost equivalent to a reference TM system, for their chromatographic performances and method's figures of merit. Method translation principles [22–25] are here applied for a rational and effective translation of the reference methodology developed for a loop-type TM system to the six FM tested column configurations, instead of a trial-and-error approach to set chromatographic parameters. By this rational approach, the first dimension (^{1}D) elution order and resolution of the original TM method are preserved and resulting 2D peak patterns are coherent between mutually translated methods. At the same time, in view of routine applications, analysis speed is also evaluated.

The best performing configuration is then examined for method performance parameters in terms of linearity over a 3 order of magnitude in analytes concentrations, sensitivity and quantitation accuracy. As a challenging application, raw fragrance materials of medium complexity are considered and a selection of targeted analytes referred to as "established contact allergens in

humans" by the EU Scientific Committee on Consumer Safety[26] are subjected to quantitative profiling. They included 60 analytes (single compounds or mixtures of isomers) covering a wide range of polarity and volatility.

5.4.3 Materials and methods

5.4.3.1 Raw materials, pure reference compounds and solvents

Pure standards of *n*-alkanes (from *n*-C9 to *n*-C25) for Linear Retention Indices (I^T) calibration were from Merck (Milan, Italy). Pure standards (or isomers mixtures) of tested analytes listed in **Table 5.4.1** were purchased from Merck (Milan, Italy) or kindly provided by Firmenich SA (Geneva, Switzerland). Solvents (cyclohexane and dichloromethane) were all HPLC-grade from Merck (Milan, Italy). Pure standards of 1,4-dibromobenzene and 4,4'- dibromobiphenyl used as Internal Standards (ISTDs) were from Merck (Milan, Italy).

Commercial raw fragrance materials for accuracy assessment were kindly provided by Farotti srl (Rimini, Italy). Test sample #1 (TS1) consisted of a *citrus-like* fragrance while test sample #2 was a *flowery-like* fragrance (TS2).

Table 5.4.1. List of analytes included in the MMix together with FM-GC×GC-MS/FID precision data on 1D and 2D retention times (^1t_R and ^2t_R), 2D peak absolute volumes and normalized volumes on FID and MS channels; linearity (R^2) and accuracy (relative error %) at two spiking levels. "$" refers to accuracy data calculated on the TIC-MS signal instead of FID.

	Precision - Repeatability										**Linearity**		**Accuracy – *Relative error %***			
	^{1}D Retention (^1t_R)			***^{2}D Retention (^2t_R)***			***FID –Responses %RSD***		***MS TIC – Responses %RSD***				***Citrus-like TS1***		***Flowery-like TS2***	
Compound Name	**min**	**St dev**	**% RSD**	**sec**	**St dev**	**% RSD**	**Vol.**	**Norm. Vol.**	**Vol.**	**Norm. Vol.**	***R^2* FID**	***R^2* MS**	**+1 mg/L**	**+10 mg/L**	**+1 mg/L**	**+10 mg/L**
1,4-Dibromobenzene	15.17	0.08	0.50	1.35	0.03	1.87	3.69	0.00	2.55	0.00	-	-	-	-	-	-
4,4'-Dibromobiphenyl	33.55	0.00	0.00	2.00	0.00	0.25	3.57	0.00	1.68	0.00	-	-	-	-	-	-
Benzaldehyde	8.70	0.00	0.00	1.21	0.01	0.63	0.45	1.63	2.07	4.39	0.998	0.994	-7.21	-2.27	-15.07	-12.29
α-Pinene	8.80	0.00	0.00	0.38	0.00	0.00	1.21	0.03	2.95	9.66	0.998	0.989	-5.06	-9.36	-0.22	-6.16
β-Pinene	9.85	0.00	0.00	0.51	0.00	0.57	2.02	0.84	0.90	0.07	0.996	0.999	-5.80	3.16	-0.11	-2.75
Benzyl alcohol	10.63	0.03	0.27	1.34	0.02	1.14	0.24	1.42	1.33	1.41	0.998	0.998	15.26	-6.18	-8.44	2.25
α-Terpinene	10.83	0.03	0.27	0.52	0.02	3.87	1.10	0.08	1.40	4.39	0.998	0.997	-21.40	-2.51	12.23	-2.78
Salicylaldehyde	10.88	0.03	0.26	1.34	0.01	0.65	0.70	1.88	3.32	2.60	0.998	0.998	19.74	22.04	-1.86	5.47
Limonene	11.15	0.00	0.00	0.51	0.01	2.03	1.81	0.63	2.59	2.06	0.996	0.990	-6.44	-7.62	-0.87	16.91
Terpinolene	12.75	0.00	0.00	0.64	0.01	1.20	14.58	13.41	2.11	2.91	0.993	0.998	-6.79	-0.10	-7.41	-13.09
Linalool	12.83	0.03	0.22	0.65	0.01	1.55	8.66	9.84	1.81	0.80	0.996	0.996	9.43	-15.61	1.62	-9.05
Camphor	13.95	0.00	0.00	1.13	0.01	0.89	3.24	2.06	0.72	1.10	0.998	0.997	6.39	2.05	-2.60	2.74
Menthol	14.90	0.00	0.00	0.72	0.01	1.44	2.73	1.55	0.69	3.33	0.995	0.996	-8.86	-1.21	-21.80	-0.30
Folione	15.17	0.03	0.19	1.11	0.01	0.45	2.61	1.43	5.63	5.58	0.994	0.997	-9.91	5.22	-6.26	-6.95
Methyl salicylate	15.23	0.03	0.19	1.28	0.01	0.98	2.51	1.33	4.29	4.42	0.997	0.998	-11.74	-0.67	-0.64$	-13.38$
α-Terpineol	15.33	0.03	0.19	0.88	0.01	1.19	3.13	1.95	8.65	1.31	0.997	0.997	-5.16	0.50	-3.82	-13.03
Citronellol	16.27	0.03	0.18	0.72	0.01	1.44	2.10	0.92	3.28	0.33	0.998	0.993	-2.79	-2.75	-8.83	-4.81
Neral	16.47	0.03	0.18	0.97	0.01	0.79	2.63	1.45	3.16	3.96	0.998	0.997	-1.36$	6.28$	-1.59	-5.68
Carvone	16.50	0.00	0.00	1.20	0.01	0.84	1.22	0.04	4.90	6.15	0.998	0.996	-8.43	0.36	0.60	-0.88
Cinnamaldehyde	16.90	0.05	0.30	1.76	0.01	0.72	3.78	2.60	4.32	6.27	0.998	0.997	-4.59	15.87	-6.79	8.98
Geraniol	16.93	0.03	0.17	0.84	0.01	1.19	0.67	0.51	2.82	3.90	0.998	0.995	17.64	7.77	9.97	8.58
Linalyl acetate	17.10	0.00	0.00	0.66	0.02	2.30	2.66	3.84	4.91	1.61	0.998	0.998	17.12	-6.75	-7.48	-1.31
Geranial	17.18	0.03	0.17	0.98	0.01	0.78	2.14	0.96	0.55	0.42	0.998	0.998	-16.86	11.12	-0.58$	-5.43$
Anise alcohol	17.30	0.05	0.29	1.81	0.02	1.12	1.24	2.42	4.67	8.44	0.998	0.995	-9.85	1.08	-13.22	-10.00
Hydroxycitronellal	17.47	0.03	0.17	1.03	0.01	1.28	2.02	0.84	0.51	0.86	0.997	0.997	-1.03	-3.02	-20.49	-3.19

Anethole trans	17.72	0.03	0.16	1.22	0.01	0.71	4.70	3.52	4.78	1.31	0.998	0.998	-8.64	-0.18	-10.81	-6.72
Cinnamyl alcohol	17.97	0.03	0.16	1.67	0.02	0.91	0.73	0.45	1.34	5.23	0.997	0.998	4.06	13.18	-15.09	8.44
DMBCA	18.65	0.00	0.00	1.07	0.01	1.18	1.03	0.15	3.73	1.81	0.999	0.998	5.08	1.92	-20.04	-6.82
Eugenol	19.47	0.03	0.15	1.34	0.01	0.43	3.28	2.10	1.68	4.40	0.998	0.998	0.03	5.80	-11.21	-10.68
Vanillin	20.05	0.05	0.25	2.15	0.03	1.40	11.32	10.15	11.65	1.31	0.997	0.996	15.05	16.41	12.39	4.13
δ-Damascone	20.22	0.03	0.14	0.94	0.00	0.31	11.86	10.69	5.32	4.68	0.978	0.997	0.31	-8.10	-7.76	0.42
Geranyl acetate	20.23	0.03	0.14	0.87	0.03	3.04	7.85	9.02	4.19	0.19	0.996	0.997	-3.15	13.68	16.80	-6.11
β-Damascenone	20.38	0.03	0.14	1.09	0.01	0.80	4.10	2.92	7.46	8.32	0.997	0.997	-4.00	2.41	-14.15	-4.19
α-Damascone	20.65	0.00	0.00	1.01	0.01	1.25	3.70	2.52	0.22	0.47	0.991	0.992	-6.62	-2.16	-11.71	6.89
Coumarin	21.00	0.05	0.24	2.59	0.03	1.16	2.49	1.31	3.93	3.14	0.997	0.994	-9.81	9.51	-12.50	-17.43
Majantol	21.02	0.03	0.14	1.19	0.01	1.11	4.08	2.91	6.62	1.70	0.998	0.995	-4.25	6.74	-7.89	4.30
β-Damascone	21.17	0.03	0.14	1.03	0.01	0.56	3.41	2.23	1.14	5.97	0.998	0.991	-6.38	-5.46	-14.72	-15.45
Isoeugenol (E)	21.75	0.00	0.00	1.43	0.01	0.73	3.17	1.99	3.03	2.75	0.998	0.998	-19.77	9.51	-14.49	3.96
β-Caryophyllene	21.87	0.03	0.13	0.70	0.01	1.43	3.54	2.36	4.12	7.26	0.993	0.996	-6.30	-1.75	-4.24	-1.37
Ebanol (Z isomer)	22.10	0.00	0.00	0.72	0.01	1.40	3.82	2.64	3.99	1.37	0.998	0.998	4.60	3.64	-13.60	-9.55
Ebanol (E isomer)	22.30	0.00	0.00	0.73	0.01	1.73	3.46	2.28	3.19	0.06	0.997	0.998	-2.12	3.69	-16.47	-8.40
Isomethylionone alpha	22.92	0.03	0.13	0.87	0.01	0.58	4.63	3.45	2.25	0.80	0.996	0.997	-4.56	-1.99	-5.61	-1.88
Eugenyl acetate	23.35	0.05	0.21	1.63	0.01	0.31	4.13	2.95	1.13	3.74	0.997	0.999	-9.11	10.74	-7.44	15.63
Lilial	23.67	0.03	0.12	1.19	0.00	0.24	4.18	3.00	2.51	4.43	0.997	0.996	4.33	-0.18	-5.39	-0.46
Propylidene phtalide	24.57	0.03	0.12	1.79	0.01	0.32	2.71	1.53	0.81	2.08	0.998	0.998	-27.26	9.69	-3.70	10.25
Amyl salicylate	24.93	0.03	0.12	1.03	0.01	0.84	0.28	0.90	6.05	2.63	0.995	0.994	-5.19	1.27	-14.88	-8.85
Isoeugenyl acetate	25.38	0.03	0.11	1.65	0.01	0.46	4.20	3.02	5.88	3.67	0.998	0.992	-9.74	13.06	-6.90	-2.71
Amylcinnamaldehyde alpha	26.47	0.03	0.11	1.24	0.02	1.68	0.21	1.39	5.26	5.58	0.984	0.993	-3.74	22.29	3.02	4.89
Lyral (minor isomer)	26.48	0.03	0.11	1.34	0.01	0.65	11.57	10.40	3.63	5.56	0.996	0.999	3.32	-22.34	-9.54	-4.87
Lyral (major isomer)	26.63	0.03	0.11	1.36	0.01	0.85	3.33	2.16	1.08	0.23	0.997	0.996	-11.96[$]	14.81[$]	-12.89	5.80
ISO E Super (major isomers)	27.23	0.03	0.11	0.97	0.01	1.07	3.99	2.81	2.33	1.08	0.998	0.996	-12.23	12.79	-10.56	10.33
Amylcinnamyl alcohol alpha	27.28	0.03	0.11	1.24	0.01	1.01	4.08	2.90	3.02	7.75	0.997	0.990	-5.37	10.41	-17.93	1.20
α-Santalol	27.42	0.03	0.11	0.98	0.01	0.51	2.87	1.69	2.28	5.97	0.996	0.994	-4.36	5.60	-23.17	-4.82
Farnesol	28.22	0.03	0.10	0.86	0.00	0.34	0.47	0.71	4.35	4.58	0.997	0.994	-2.05	9.18	5.50	-14.12
β-Santalol	28.25	0.00	0.00	1.07	0.01	1.17	3.50	2.32	6.66	2.70	0.997	0.998	-3.83	15.37	-12.40	-17.09
α-Hexylcinnamaldehyde	28.63	0.03	0.10	1.17	0.01	0.89	4.83	3.65	6.24	0.05	0.998	0.997	1.03	8.23	-9.73	5.06
Benzyl benzoate	28.78	0.03	0.10	1.88	0.01	0.53	5.25	4.07	7.76	3.91	0.997	0.997	-11.89	8.91	-19.23	-1.93
Acetylcedrene	29.48	0.03	0.10	1.08	0.01	0.71	3.93	2.75	0.09	0.60	0.997	0.999	-3.30	7.53	-11.61	6.86
Benzyl salicylate	31.00	0.05	0.16	1.76	0.01	0.59	1.87	3.04	1.32	0.40	0.997	0.997	2.85	12.59	2.01	10.91
Galaxolide (major isomers)	31.50	0.05	0.16	1.22	0.01	0.47	3.38	2.20	1.53	1.61	0.997	0.997	-4.21	10.25	-8.28	13.56
Hexadecanolactone	32.73	0.03	0.09	1.08	0.01	0.71	3.31	2.13	1.64	1.31	0.993	0.985	8.05	12.62	-10.55	2.86
Benzyl cinnamate	35.03	0.03	0.08	1.92	0.01	0.30	3.09	1.91	0.12	5.82	0.997	0.998	-5.41	19.90	-4.13	19.67
Sclareol	37.82	0.03	0.08	1.23	0.00	0.00	1.87	3.05	3.58	11.29	0.998	0.986	0.15	24.33	-8.06	15.02
Average	/	/	0.12	/	/	0.98	3.51	2.71	3.30	3.20	0.996	0.996	-3.07	4.58	-7.44	-0.69

5.4.3.2 Reference solutions and calibration mixtures

Standard Stock Solutions (SS) of reference analytes were prepared at a concentration of 10 mg/mL in dichloromethane or cyclohexane and stored at -18°C. The Model Mixture (MMix) stock solution was prepared by mixing suitable amounts of SS at a final concentration of 200 mg/L in cyclohexane. Fresh calibration solutions were prepared every week by diluting suitable amounts of MMix in cyclohexane. Calibration levels covered were: 0.1-0.2-0.5-1-5-10-20-50-100 mg/L. ISTDs were at a final concentration of 50 mg/L. Standard reference solutions for purity evaluation (by 1D-GC-FID) were prepared from SS at a nominal concentration of 100 mg/L in cyclohexane.

Raw fragrances TS1 and TS2 were diluted 20% (w/v) immediately before analysis in cyclohexane. For accuracy evaluation, spiked samples were prepared by adding suitable volumes of MMix up to +10 mg/L and +1 mg/L concentration levels. ISTDs were added to all analyzed samples at 50 mg/L.

5.4.3.3 GC×GC with reverse-inject differential flow modulation: instrument set-up

GC×GC analyses with reverse-inject differential flow modulation were run with a GC-MS system consisting of an Agilent 7890A GC unit provided with a 4513A auto injector sampler (Agilent, Little Falls, DE, USA) and coupled to an Agilent 5977B HES (High Efficiency Source) fast quadrupole MS detector (Agilent, Little Falls, DE, USA) operating in EI mode at 70 eV and a fast FID detector. The GC transfer line was set at 280°C. The MS was tuned using the *HES Tune* option. The scan range was set to m/z 40-240 with a scanning rate of 12,500 amu/s to obtain a spectrum generation frequency of 28 Hz. The flame ionization detector (FID) conditions were base temperature 280°C, H_2 flow 40 mL/min, air flow 350 mL/min, make-up (N_2) 20 mL/min, and sampling frequency 200 Hz.

The system was equipped with a reverse-inject FM (Supplementary Material – **Supplementary Figure 5.4.1**) consisting of a CFT plate connected to a three-way solenoid valve that receives a controlled supply of carrier gas (helium) from an auxiliary electronic pressure control module (EPC). The CFT plate schematic and modulation dynamics description is provided in **Supplementary Figure 5.4.1**.

5.4.3.4 GC×GC with thermal modulation: instrument set-up

The TM GC×GC system consisted of an Agilent 7890B GC unit with a 4513A auto injector sampler (Agilent, Little Falls, DE, USA) coupled with a Bench TOF-Select™ time of flight mass spectrometer (Markes International, Llantrisant, UK). Electron ionization was set at 70 eV. The ion source and transfer line were set at 290°C. The MS optimization option was set to operate in Single Ionization with a mass range between 35 and 550 m/z; data acquisition frequency was 100 Hz; filament voltage was set at 1.60 V. For parallel detection, the FID was set with a base temperature of 280°C, H_2 flow 40 mL/min, air flow 350 mL/min, make-up (N_2) 20 mL/min, and sampling frequency 200 Hz.

The system was equipped with a two-stage KT 2004 loop thermal modulator (Zoex Corporation, Houston, TX) cooled with liquid nitrogen controlled by Optimode™ V.2 (SRA Instruments, Cernusco sul Naviglio, MI, Italy). The hot jet pulse time was set at 250 ms, modulation period was 5 s, and cold-jet total flow was progressively reduced with a linear function from 35% of Mass Flow Controller (MFC) at initial conditions to 5% at the end of the run.

Injections of the Calibration mixtures (CAL), as well as those for I^T_S determination, were carried out with a 4513A auto injector under the following conditions: injection mode: split, split ratio: 1/20 for CAL and 1/50 for *n*-alkanes, injection volume 2 μL, temperature 270°C.

5.4.3.5 Column set, connections and auxiliary control modules

The reference method (*i.e.*, TM-GC×GC) columns configuration and those tested with FM-GC×GC to achieve comparable method performance are summarized in **Table 5.4.2**. Pressure settings (S/SL injector and Auxiliary EPC), carrier gas (helium) volumetric flows in the two dimensions, linear velocities across capillaries, and oven temperature programs also are reported. Calculations were by reference equations and/or by a validated pneumatic model designed for the CFT plate.[27]

Connections between the second dimension (^{2}D) column and deactivated silica capillaries toward MS and FID for parallel detection were by a three-way un-purged splitter (G3181B, Agilent, Little Falls, DE, USA) while columns connection in the TM-GC×GC was by deactivated ultimate unions (G3182-61580 Agilent, Little Falls, DE, USA). ^{1}D columns DB-1, ^{2}D columns OV17 and deactivated capillaries were from Agilent - J&W (Little Falls, DE, USA).

5.5.3.6 Performance parameters: reference equations

To evaluate the performances of the tested FM configurations compared to the reference method, several chromatographic performance parameters were considered. The reference parameters and equations are described here.

Re-injection pulse width ($^2\sigma_i$) directly affects the actual $^2\sigma_t$ with an additive effect on ^{2}D peak-broadening due to the chromatographic process ($^2\sigma_c$). Re-injection pulses were defined as peak standard deviation ($^2\sigma_i$) and estimated on un-retained solvent peaks from FID channel (200 Hz sampling frequency).[1]

The net separation measure ($S_{GC\times GC}$)[28] describes system's separation ability under the experimental conditions applied. $S_{GC\times GC}$ extends the concept of separation measure (S) to GC×GC separations [29] and refers to the product of S in each chromatographic dimension:

Equation 5.4.1 $$S_{GC\times GC} = S_1 \times S_2$$

where S_1 and S_2 are calculated, for ^{1}D and ^{2}D respectively, using the reference equation:

Equation 5.4.2 $$S = {\Delta t}/{\sigma_{av}}$$

where Δt is the arbitrary time interval between two peaks *a* and *b*, $\Delta t = t_b - t_a$, and σ_{av} is the peak- standard deviation (σ) of *a* and *b*:

Equation 5.4.3 $$\sigma_{av} = \frac{(\sigma_a + \sigma_b)}{2}$$

In this study the time interval was that between the first (*i.e.*, benzaldehyde) and the last (*i.e.*, sclareol) eluting peaks of the MMix for the ^{1}D and the P_M for the ^{2}D.

Pattern coherence was evaluated by relative retention (RR) in the two chromatographic dimensions [11] and taking as reference centroid methyl salicylate and sclareol as last eluting peak. In the ^{2}D the relative retention is normalized to P_M. Here follows reference equations:

Equation 5.4.4 $$^1\text{D } RR = (^1\text{D Rt}_i - {}^1\text{D Rt}_{\text{methyl salicylate}})/{}^1\text{D Rt}_{\text{sclareol}}$$

Equation 5.4.5 $^2D\ RR = (^2D\ Rt_i - {}^2D\ Rt_{methyl\ salicylate})/P_M$

Within method performance parameters, linearity in the calibration range (0.1 – 100 mg/L) was evaluated with the determination coefficient of the linear model (R^2) while limits of detection (x_{LOD}) for MS and FID channels were calculated according to EU guidelines[30] as:

Equation 5.4.6 $$x_{LOD} = 3.9\frac{s_{y,b}}{b}$$

where $s_{y,b}$ is the standard deviation of the blank signal and b is the slope of the calibration curve within the lower calibration levels (*i.e.*, 0.1-1 mg/L).

Precision was estimated over a one-week validation protocol as repeatability[31] and expressed as percent relative standard deviation (% RSD). It was calculated on retention times in the two dimensions (^1t_R and ^2t_R) at all calibration points (n=8) and for all analytical replicates (n= 2). Repeatability on absolute and normalized 2D volumes were calculated for the analytical replicates in the middle of the calibration range at 1 and 10 mg/L.

Accuracy was estimated initially at two spiking levels (*i.e.*, 1 and 10 mg/L in the final sample) and for two commercial fragrances of medium complexity. Bias was expressed as *relative error* % according to the following equation:

Equation 5.4.7 $$Relative\ error\ \% = \frac{(x_m - x_{exp})}{x_{exp}} \times 100$$

where x_m is the estimated amount and x_{exp} is the expected amount after spiking. Accuracy was established analysing TS1 and TS2 samples and spiked ones in triplicate. The *relative error %* reported in **Table 5.4.1** are those resulting from FID signals except for analytes affected by co-elutions and reported in the table with the symbol "$".

5.4.3.7 Data acquisition and 2D data processing

Data were acquired by TOF-DS software (Markes International, Llantrisant, UK) in the reference TM-GC×GC method and Enhanced MassHunter (Agilent Technologies, Little Falls, DE, USA) in the translated FM-GC×GC methods. 2D data were processed by GC Image® GC×GC Edition Software, Release 2.9 (GC Image, LLC Lincoln NE, USA).

5.4.4 Results and Discussion

5.4.4.1 Background for the present study and reference method

In previous studies, we successfully applied the principles of method translation from a reference method, implemented with a loop-type thermal modulator GC×GC-MS/FID platform, to a reverse-inject differential flow modulated GC×2GC-MS/FID platform.[11,12] The configuration tested in the FM-GC×GC consisted of a ^{1}D with reduced internal diameter and length, compared to the reference set up, and two-parallel ^{2}D columns each one directed to a different detector (MS and FID). The column combination included a polyethylene glycol (PEG) stationary phase in the ^{1}D and 86% polydimethylsiloxane, 7% phenyl, 7% cyanopropyl in the ^{2}D. Based on the models developed by Blumberg and Klee,[22,32] translatable parameters were set to preserve ^{1}D peak elution order, ^{1}D peak capacity, and chromatographic resolution. Temperature programming was therefore modified according to the estimated speed gain and corresponding to the ratio between column void times (t_{Mref} and t_{Mtr}). The operation was supported by the method translation software and available as free application on the web.[33]

Results were satisfactory and included a reduction of a factor of 2 for the total analysis time (t_A) of the translated method (32.67 min instead of 65.53 min of the original method) and the preservation of the elution order and of the relative retention in the two chromatographic dimensions (*i.e.*, pattern coherence). Pattern coherence, between mutually translatable methods, enabled effective transfer of metadata from the reference methodology by template matching algorithms.[11,12,34]

More recently, Aloisi et al.[35] explored the possibility of defining an equivalent standard column set between TM and FM GC×GC. Their strategy was driven by the choice of two equivalent column sets, in terms of separation power, in consideration of the flow restrictions posed by the two systems. Their set up included a TM platform with a ^{1}D 30 m x 0.25 mm d_c x 0.25 µm d_f and a ^{2}D of 1.5 m x 0.25 mm d_c x 0.25 µm d_f with a delay loop of 1.5 m x 0.18 mm d_c while the FM was with a ^{1}D 20 m x 0.18 d_c x 0.18 µm d_f and a ^{2}D of 5 m x 0.32 d_c x 0.25 µm d_f. The two set-ups provided almost equivalent separation power (referred to the efficiency expressed as number of theoretical plates *N*) and analytes relative retention in the two dimensions. However, the FM method had lower sensitivity because of the compensation for the lower re-injection efficiency of the FM system. The authors report that "… *A sample amount 4 x higher was introduced onto the ^{1}D column in the FM analysis, to compensate for the higher sensitivity of CM (i.e., cryo-modulator)*".[35]

Although the sensitivity drop observed with the FM-GC×2GC-MS/FID set up in translated conditions [12] was less drastic, the FM method did not match TM performances. In the mentioned study [12], cocoa volatiles fingerprinting covered 75 of the 130 targeted peaks (58%) and 450 of the 595 (76%) reliable peak-regions compared to the reference TM procedure.

In this study, to make a step forward in the direction of matching, at the same time, separation power and sensitivity, the RFF FM modulator is tested in its full flexibility by combining three different ^{1}D columns with two ^{2}Ds for a total of 6 configurations. The application context is that of the routine quantification of established volatile allergens in fragrances and the reference method that proposed by Belhassen *et al.* [36]. The system included a loop-type TM, with liquid nitrogen, and a parallel dual-secondary column/dual parallel detection configuration (*i.e.*, TM-GCx2GC-MS/FID). The linearity ranges examined were between 2-100 mg/kg for MS and 100-10,000 mg/kg for FID. Accuracy was good and quantitation bias was below 20% of error for the majority of the analytes (85%).[36]

For this study, the reference TM method adopted for benchmarking FM configurations implied a longer ^{1}D column, compared to that of Belhassen *et al.,*[36] (*e.g.*, 60 m × 0.25 mm d_c x 0.25 µm d_f) and a single ^{2}D of wider diameter (*e.g.,* 1.8 m × 0.18 mm d_c x 0.18 µm d_f). Parallel detection by time-of-flight mass spectrometry (TOF MS) and FID was obtained by post-column splitting with a passive-tee junction and a flow ratio of about 30:70 (MS/FID) in order to balance the relative sensitivity of the two channels. **Table 5.4.2** reports in detail the reference method column configuration, helium carrier flows and linear velocities as they were estimated by reference equations, oven programming and total analysis time (t_A), modulation parameters and operative pressure at the inlet (p_i) at the midpoint between the two dimensions (p_{mid}) and at the tee-union. The TM-GC×GC-TOF MS/FID method was tested for its linearity within 0.1 to 100 mg/L; calibration levels below 1 mg/L were explored because of the industrial needs of a quantitation method able to monitor regulated substances even below the conformity limits. The platform including TOF MS and FID (30:70) enabled to cover this requirement for both channels. In addition, the larger ^{2}D column d_c compensates for the limited loadability of 0.1 mm d_c columns while helping in situations where highly abundant components may overload it to the detriment of both ^{2}D separations and TOF MS ionization efficiency.

Based on the reference method, the six different FM combinations are detailed in **Table 5.4.2** (Set-up #1a and b; #2a and b; #3a and b). The rationale for their design was based on limitations due to the modulation dynamics, which requires low carrier flow in the ^{1}D and high flows in the ^{2}D. ^{1}D columns tested were therefore 0.10 mm and 0.18 mm d_c with variable phase ratios to enable higher loadability (*e.g.*, 10 m x 0.10 mm x 0.1 or 0.4 μm d_f). ^{2}D columns were set to afford adequate loadability and efficiency to match with the benchmark peak-capacity. Chromatographic performance parameters were at first examined to evaluate the best configuration. The next section reports experimental results on chromatographic performance in a critical perspective.

Table 5.4.2. Reference and translated methods settings, including columns characteristics, initial head-pressure (pi), helium volumetric flows, and hold-up times on the basis of reference equations. Oven temperature programming and total analysis time (t_A) are also reported. Operative conditions include also modulation parameters.

	^{1}D	^{2}D	Connections and capillaries	Oven programming	Modulation parameters
Reference method TM-GC×GC-TOFMS/FID	DB1 60 m × 0.25 mm d_c x 0.25 µm d_f He @ 2.0 mL/min - constant flow Initial head-pressure (p_i relative) 254.7 kPa Outlet pressure (p_{mid} absolute) 163.7 kPa Hold-up 3.52 min - Outlet velocity 46.97 cm/s	OV17 1.8 m × 0.18 mm d_c x 0.18 µm d_f He @ 2.0 mL/min - constant flow Initial head-pressure (p_{mid} absolute) 163.7 kPa Hold-up 1.8 sec - Outlet velocity 106.6 cm/s	Loop-capillary deactivated silica: 1.0 m, 0.10 mm d_c MS/FID split ratio 70:30 to MS: 0.7 m, 0.10 mm d_c to FID: 1.1 m, 0.18 mm d_c	60°C(1') to 280°C (10') @ 4°/min t_A= 48.35 min	P_M = 5s hot-jet pulse: 250 ms
Translated methods FM-GC×GC-MS/FID					
Set-up #1a	DB1 10 m × 0.10 mm d_c x 0.10 µm d_f He @ 0.27 mL/min - constant flow Initial head-pressure (p_i relative) 305.27 kPa Outlet pressure (p_{aux} absolute) 278 kPa Hold-up 0.89 min - Outlet velocity 18.80 cm/s	OV17 1.8 m × 0.18 mm d_c x 0.18 µm d_f He @ 8.0 mL/min - constant flow Initial head-pressure (p_{aux} relative) 177 kPa Hold-up 0.6 sec - Outlet velocity 292.57 cm/s	MS/FID split ratio 70:30 to MS: 0.5 m, 0.10 mm d_c to FID: 1.1 m, 0.18 mm d_c bleeding capillary: 6.37 m, 0.10 mm d_c	60°C(0.25') to 280°C (2.52') @ 15.89°/min t_A= 12.37 min; t_A % reduction: 25.6%	P_M = 2s pulse time: 150 ms
Set-up #1b	DB1 10 m × 0.10 mm d_c x 0.10 µm d_f He @ 0.27 mL/min - constant flow Initial head-pressure (p_i relative) 288.54 kPa Outlet pressure (p_{aux} absolute) 253 kPa Hold-up 0.83 min - Outlet velocity 19.95 cm/s	OV17 2.5 m × 0.25 mm d_c x 0.25 µm d_f He @ 11.0 mL/min - constant flow Initial head-pressure (p_{aux} relative) 151.94 kPa Hold-up 2.58 sec - Outlet velocity 225.65 cm/s	MS/FID split ratio 70:30 to MS: 0.5 m, 0.10 mm d_c to FID: 1.1 m, 0.18 mm d_c bleeding capillary: 5.11 m, 0.10 mm d_c	60°C(0.24') to 280°C (2.37') @ 16.87°/min t_A= 11.67 min; t_A % reduction: 24.1%	P_M = 4s pulse time: 150 ms
Set-up #2a	DB1 10 m × 0.10 mm d_c x 0.40 µm d_f He @ 0.27 mL/min - constant flow Initial head-pressure (p_i relative) 307.9 kPa Outlet pressure (p_{aux} absolute) 278 kPa Hold-up 0.88 min - Outlet velocity 18.95 cm/s	OV17 1.8 m × 0.18 mm d_c x 0.18 µm d_f He @ 8.0 mL/min - constant flow Initial head-pressure (p_{aux} relative) 177 kPa Hold-up 0.6 sec - Outlet velocity 292.57 cm/s	MS/FID split ratio 70:30 to MS: 0.5 m, 0.10 mm d_c to FID: 1.1 m, 0.18 mm d_c bleeding capillary: 6.37 m, 0.10 mm d_c	60°C(1.01') to 280°C (10.08') @ 3.97°/min t_A= 48.80 min; t_A % increase: 100.9%	P_M = 3s pulse time: 150 ms
Set-up #2b	DB1 10 m × 0.10 mm d_c x 0.40 µm d_f He @ 0.27 mL/min - constant flow Initial head-pressure (p_i relative) 291.29 kPa Outlet pressure (p_{aux} absolute) 253 kPa Hold-up 0.83 min - Outlet velocity 20.01 cm/s	OV17 2.5 m × 0.25 mm d_c x 0.25 µm d_f He @ 11.0 mL/min - constant flow Initial head-pressure (p_{aux} relative) 151.94 kPa Hold-up 2.58 sec - Outlet velocity 225.65 cm/s	MS/FID split ratio 70:30 to MS: 0.5 m, 0.10 mm d_c to FID: 1.1 m, 0.18 mm d_c bleeding capillary: 5.11 m, 0.10 mm d_c	60°C(0.95') to 280°C (9.5') @ 4.21°/min t_A= 46.05 min; t_A % reduction: 95.2%	P_M = 4.5s pulse time: 150 ms
Set-up #3a	DB1 20 m × 0.18 mm d_c x 0.18 µm d_f He @ 0.5 mL/min - constant flow Initial head-pressure (p_i relative) 227.73 kPa Outlet pressure (p_{aux} absolute) 278 kPa Hold-up 2.72 min - Outlet velocity 13.34 cm/s	OV17 1.8 m × 0.18 mm d_c x 0.18 µm d_f He @ 8.0 mL/min - constant flow Initial head-pressure (p_{aux} relative) 177 kPa Hold-up 0.6 sec - Outlet velocity 292.57 cm/s	MS/FID split ratio 70:30 to MS: 0.5 m, 0.10 mm d_c to FID: 1.1 m, 0.18 mm d_c bleeding capillary: 6.06 m, 0.10 mm d_c	60°C(0.77') to 280°C (7.74') @ 5.17°/min t_A= 37.85 min; t_A % reduction: 78.3%	P_M = 3s pulse time: 150 ms
Set-up #3b	DB1 20 m × 0.18 mm d_c x 0.18 µm d_f He @ 0.5 mL/min - constant flow Initial head-pressure (p_i relative) 207.12 kPa	OV17 2.5 m × 0.25 mm d_c x 0.25 µm d_f He @ 11.0 mL/min - constant flow Initial head-pressure (p_{aux} relative) 151.94 kPa	MS/FID split ratio 70:30 to MS: 0.5 m, 0.10 mm d_c to FID: 1.1 m, 0.18 mm d_c	60°C(0.72') to 280°C (7.16') @ 5.58°/min	P_M = 4.5s pulse time: 150 ms

	Outlet pressure (p_{aux} absolute) 253 kPa Hold-up 2.52 min - Outlet velocity 13.21 cm/s	Hold-up 2.58 sec - Outlet velocity 225.65 cm/s	bleeding capillary: 2.76 m, 0.10 mm d_c	t_A= 34.95 min; t_A % reduction: 72.3%	

5.4.4.2 Chromatographic performances of FM-GC×GC-MS/FID in translated conditions

The workflow to translate chromatographic parameters is visualized in the Supplementary Material – **Supplementary Figure 5.4.2**. In practice, estimated operative pressures at the inlet (p_i) and outlet (p_{out} or p_{mid}) of the ^{1}D column in the reference TM method are input in the calculator and used by the model to translate conditions for the FM method. For the FM configuration, the p_i and p_{out} or p_{aux} are set based on the *a priori* fixed flow conditions in both dimensions. The model calculates the oven temperature programming for the FM method by normalizing it according to the system void time (t_M).

Each modulation period (P_M) was defined after a scouting run with each configuration and evaluating the ^{1}D baseline peak-width (w_b) to obtain a comparable modulation ratio (M_R) for all methods [37].

Results are visualized as pseudocolorized images in **Figure 5.4.1** for the MMix at 10 mg/L from the FID signal. The accordance between relative retention in both chromatographic dimensions (*i.e.*, pattern coherence) was evaluated through peaks relative retention against a centroid (methyl salicylate) and the last eluting peak (sclareol) for the ^{1}D and against the P_M for the ^{2}D (**Equation 5.4.4** and **Equation 5.4.5**). Results are visualized in Supplementary Material - **Supplementary Figure 5.4.3** for all configurations. The reference method, visualized in **Figure 5.4.1A**, was characterized by an average re-injection pulse of 20 ms against an average value of 40 ms for the FM-GC×GC (**Figure 5.4.2A**). The average $^1\sigma$ for reference peaks (first and last eluting) was 2.34 s with a resulting S_1 of 934 (**Figure 5.4.2C** and **5.4.2D**). Supplementary Material - **Supplementary Table 5.4.1** reports $^1\sigma$ and $^2\sigma$ for all targeted peaks. On the other hand, the best performing set for FM, considering only ^{1}D separation efficiency by $^1\sigma$, was Set up #1a the one combining ^{1}D 10 m x 0.1 mm d_c x 0.1 µm d_f with a ^{2}D of 0.18 mm d_c that showed an average $^1\sigma$ of 1.85 s. However, the S_1 value of this combination was only 304 (**Figure 5.4.2D**) due to the lower capacity factors (k) expressed by this set-up.

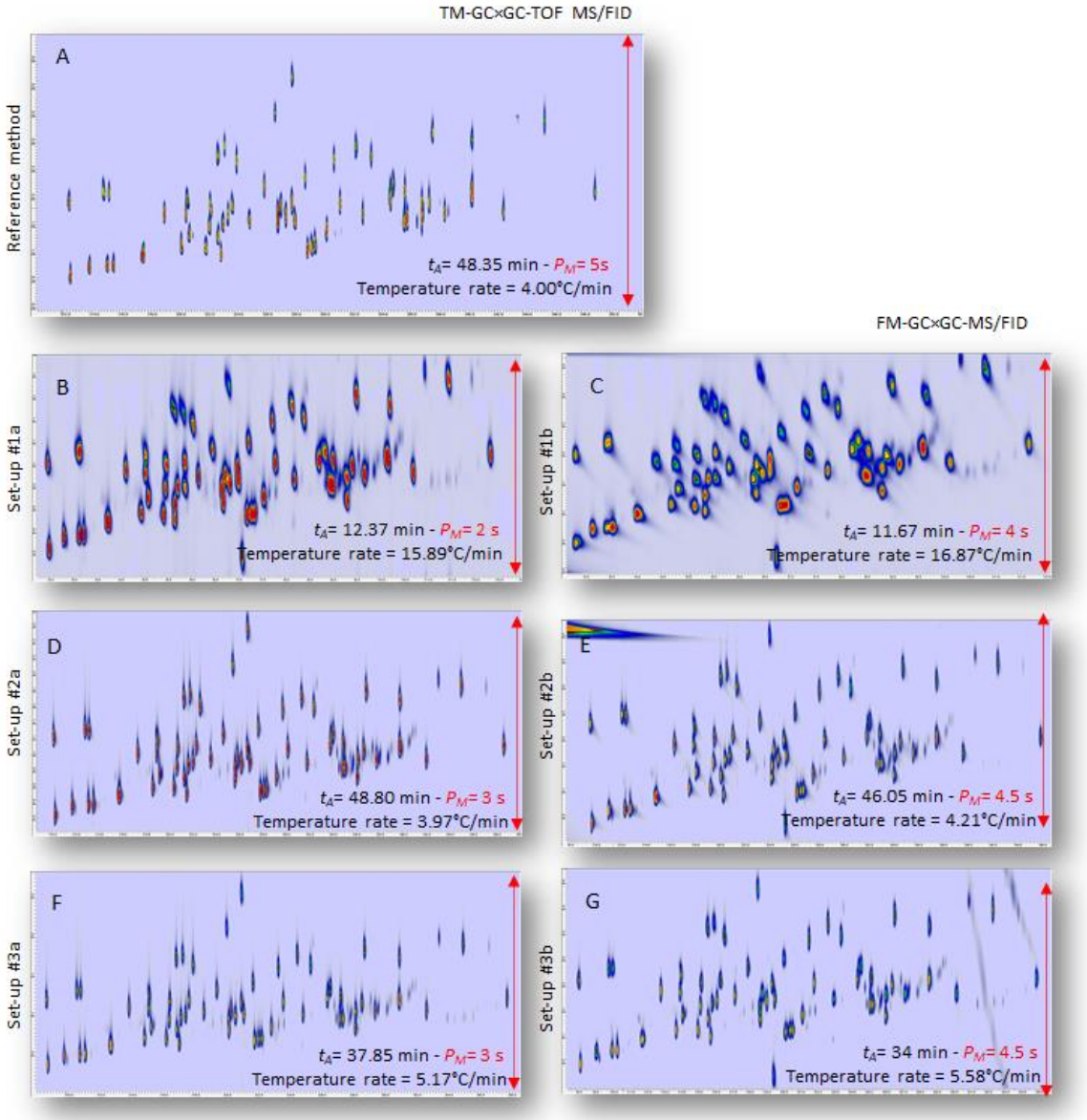

Figure 5.4.1. MMix calibration solution at 10 mg/L analyzed with the different configurations. (**5.4.1A**) reference TM-GC×GC-TOFMS/FID; (**5.4.1B**) – FM Set-up #1a [DB1 10 m × 0.10 mm dc x 0.10 µm df + OV17 1.8 m × 0.18 mm dc x 0.18 µm df]; (**5.4.1C**) – FM Set-up #1b [DB1 10 m × 0.10 mm dc x 0.10 µm df + OV17 2.5 m × 0.25 mm dc x 0.25 µm df]; (**5.4.1D**) – FM Set-up #2a [DB1 10 m × 0.10 mm dc x 0.40 µm df + OV17 1.8 m × 0.18 mm dc x 0.18 µm df]; (**5.4.1E**) – FM Set-up #2b [DB1 10 m × 0.10 mm dc x 0.40 µm df + OV17 2.5 m × 0.25 mm dc x 0.25 µm df]; (**5.4.1F**) FM Set-up #3a [DB1 20 m × 0.18 mm dc x 0.18 µm df + OV17 1.8 m × 0.18 mm dc x 0.18 µm df]; (**5.4.1G**) – FM Set-up #3b [DB1 20 m × 0.18 mm dc x 0.18 µm df + OV17 2.5 m × 0.25 mm dc x 0.25 µm df].

Conversely, average $^2\sigma$ were almost comparable for all FM methods (average value of 0.11 s in **Figure 5.4.1E**) and, in turn, even better than those estimated for the reference method (*i.e.*, $^2\sigma$ 0.17 s). For this reason, all FM systems had comparable separation power in the ^{2}D (**Figure 5.4.1F**) with S_2 values ranging between 23 for the Set up #1a [10 m x 0.1 mm d_c x 0.1 µm d_f + 1.8 m x 0.18 mm d_c x 0.18 µm d_f] and 42 for the Set up #3b [20 m x 0.18 mm d_c x 0.18 µm d_f + 2.5 m x 0.25 mm d_c x 0.25 µm d_f].

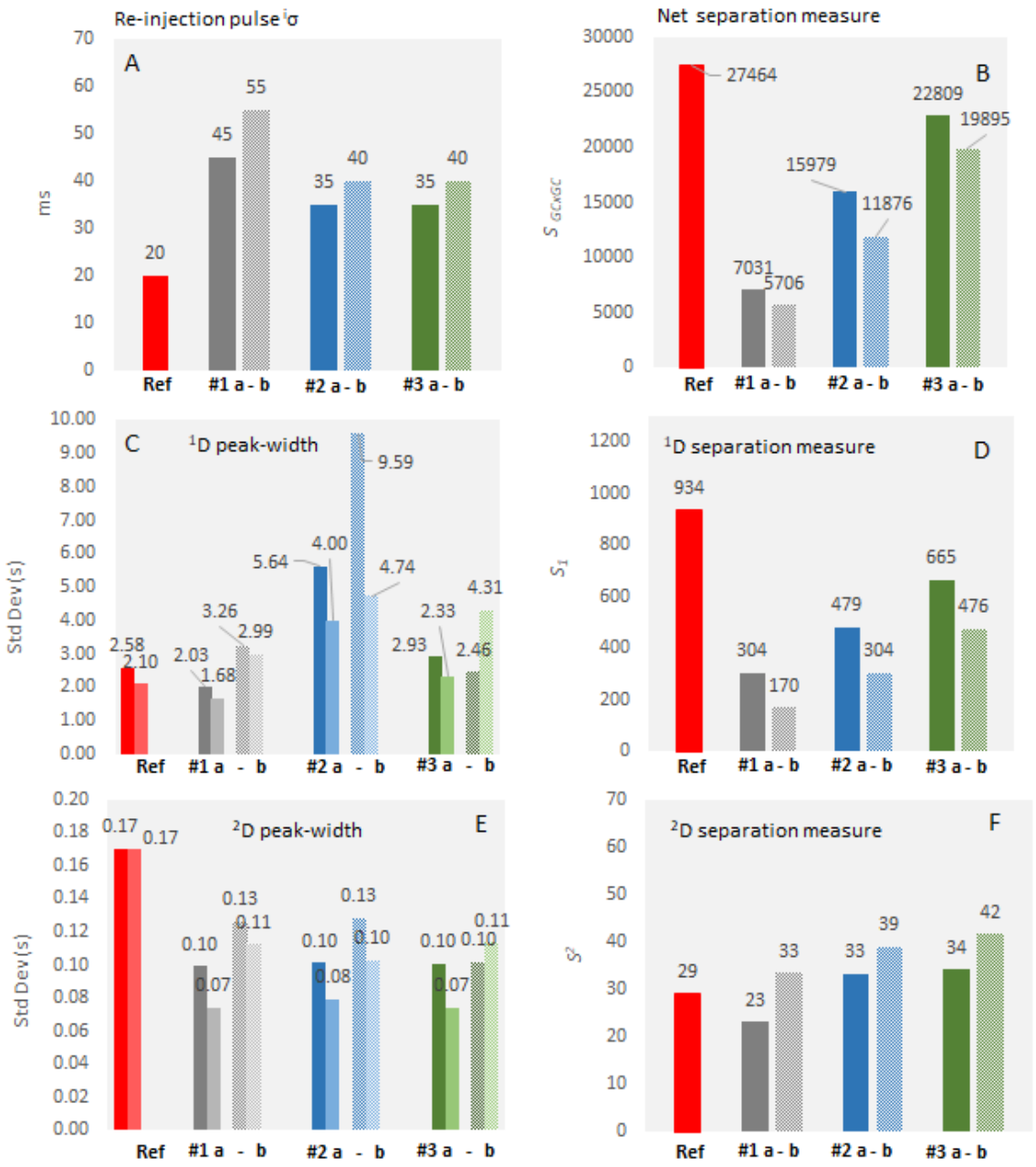

Figure 5.4.2. separation performances for the reference TM-GC×GC-TOFMS/FID method (red bars) compared to translated FM- GC×GC-MS/FID set-up (#1a and b – grey; #2a and b blue; #3a and b green). Performances refer to (**5.4.2A**) re-injection pulse width (iσ); (**5.4.2B**) net separation measure (SGC×GC); (**5.4.2C**) 1D peak-width expressed as standard deviation (1σ); (**5.4.2D**) 1D separation measure (S1); (**5.4.2E**) 2D peak-width expressed as standard deviation (2σ); (**5.4.2F**) 2D separation measure (S2).

The best performing approach in terms of separation power (*i.e.*, $S_{GC\times GC}$) was Set up #3a, [20 m x 0.18 mm d_c x 0.18 µm d_f + 1.8 m x 0.18 mm d_c x 0.18 µm d_f], with an $S_{GC\times GC}$ value of 22809 against 27464 (~ 83%) of the TM method and a shorter analysis time (~ 78%).

Based on these premises, the FM-GC×GC-MS/FID set up #3a, consisting of [DB1 of 20 m x 0.18 mm d_c x 0.18 µm d_f + OV17 of 1.8 m x 0.18 mm d_c x 0.18 µm d_f], was selected to proceed with

the evaluation of performance parameters in a one-week validation protocol. The next section reports experimental results on linearity, limit of detection (LOD), pattern reliability, and accuracy.

5.4.4.3 Method performance parameters of the translated FM-GC×GC-MS/FID

5.4.4.3.1 Linearity

The calibration ranges explored for reference and translated methods were designed to span 3 to 4 orders of magnitude of concentrations, as it is in general with natural and synthetic fragrance materials.[36,38] Calibration at low levels, between 0.1 and 1 mg/L, was explored to cover trace amounts for analytes of concern.

The reference method confirmed its good linearity at the MS channel (TIC signal) within both: (a) the 0.1-20 mg/L range with a median R^2 of 0.9983 – mean 0.9980 (min 0.9954 / max 0.9998); and (b) in the full range 0.1-100 mg/L with median R^2 of 0.9942 – mean 0.9902 (min 0.9522 / max 0.9999). Benzaldehyde exhibited the worst performance with R^2 0.9522. Linearity at the FID was satisfactory; the median value for R^2 was 0.9963 – mean 0.9959 (min 0.9949 / max 0.9987) within 0.1-100 mg/L although better performances were registered in the higher calibration range (10-100 mg/L) with R^2 median value of 0.9996 – mean 0.9995 (min 0.9984 / max 0.9999). The reference method means (red cross mark) and median (red line) are reported in red in the scatter plot of **Figure 5.4.3A** showing linearity results. R^2 values calculated on linear regression models for the translated method are reported in **Table 5.4.1**.

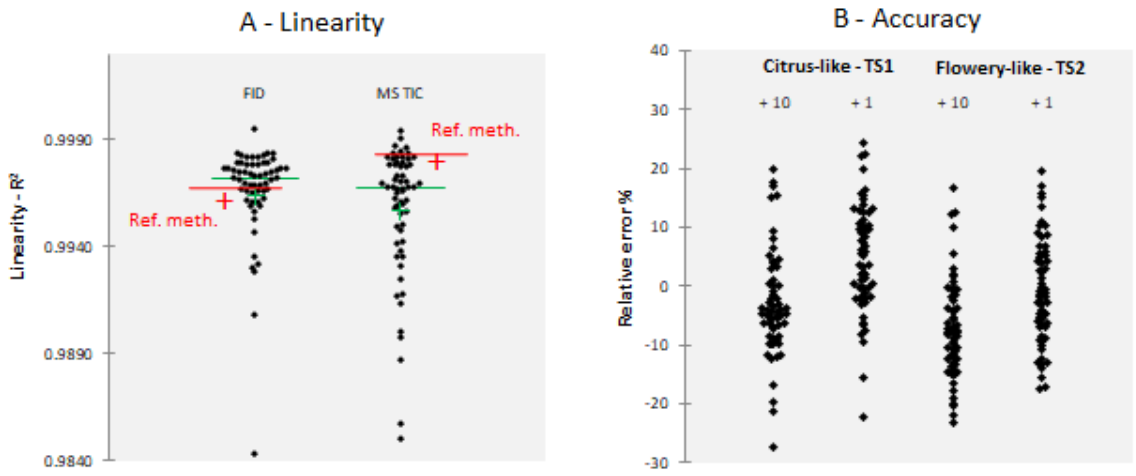

Figure 5.4.3. (**5.4.3A**) scatter diagram referring of linearity of calibration models (coefficient of determination R2) obtained with the FM-GC×GC-MS/FID method and set-up #3a; red marks report means and median of the reference methodology. (**5.4.3B**) shows accuracy results for the two tested raw materials spiked at 1 and 10 mg/L level. Accuracy is reported as relative error % - see section 2.6 for details.

The translated candidate method exhibited very good linearity; it has to be considered that the FM-GC×GC includes a single quadrupole MS with high efficiency source (HES) and actual sampling frequency was lower (*i.e.*, 28 Hz) compared to the TOF MS operating at 100 Hz. Despite these configuration differences, the qMS data was highly satisfactory, with results visualized in the scatter diagrams of **Figure 5.4.3A**. Median R^2 value for the MS TIC signal was 0.9967 – mean 0.9957 (median – green line / mean – green mark) with a min of 0.9850 for hexadecanolactone and a max of 0.9994 for eugenyl acetate in the range 0.1-20 mg/L. Conversely, FID in the full range

(0.1-100 mg/L), had median value for R^2 of 0.9972 – mean 0.9964 (median – green line / mean – green mark) with a min of 0.9784 for damascenone delta and a max of 0.9995 for dimethylbenzylcarbinyl acetate (DMBCA).

Results indicate that, in terms of linearity within the examined ranges, FM-GC×GC-MS/FID has performances comparable to the reference method. The FID channel has indeed better linear models although, as it will be discussed in the next section, absolute sensitivity for this channel is slightly lower.

5.4.4.3.2 Limit of detection

Absolute sensitivity was estimated according to the EU guidelines for food and feed,[30] generally more restrictive than those for other fields of application (**Equation 5.4.6**). Results are reported as histograms in **Figure 5.4.4** for FID (**Figure 5.4.4A**) and MS (**Figure 5.4.4B**). On average, the MS detection channel had higher sensitivity: the mean LOD value of the reference method was 7.25 µg/L, with a maximum value for sclareol (*i.e.*, 29.5 µg/L) and a minimum for limonene (*i.e.*, 2.93 µg/L). The translated method followed exactly the same trend, with slightly higher LODs (+ 1.4%). To note: the two platforms were equipped with different MS systems, consequently this data should be read in light of linearity performances. However, if one considers the FM-GC×GC as a suitable system for routine controls, reliability in the established conditions are not affected by the slower acquisition frequency of the qMS.

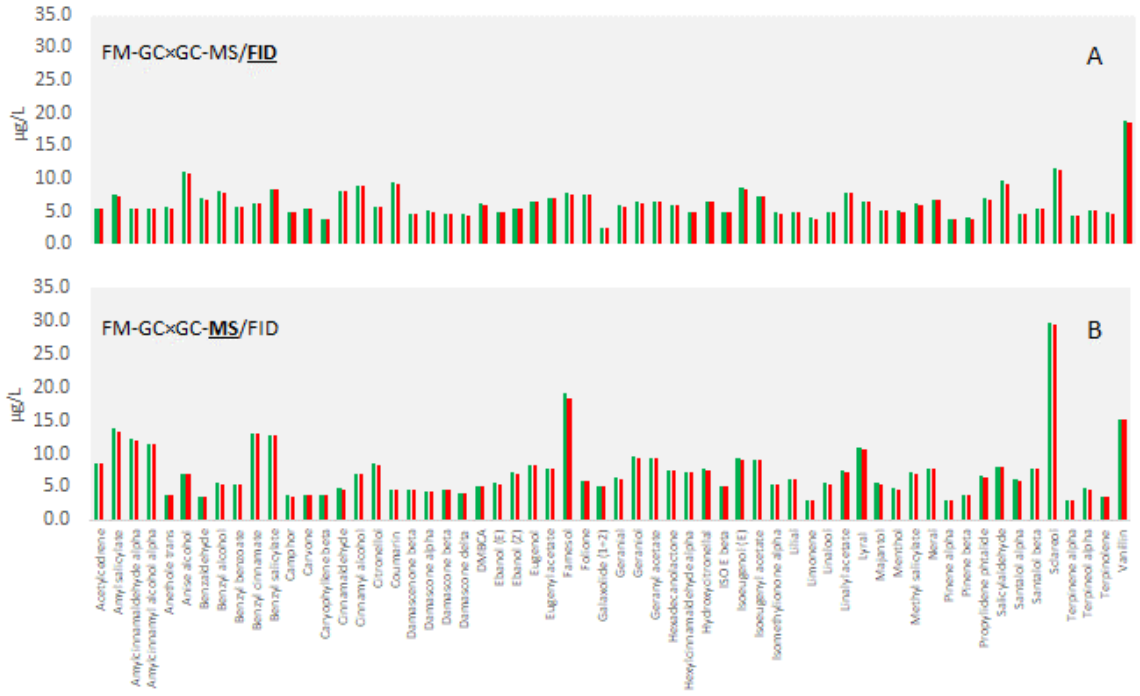

Figure 5.4.4. Histograms showing LOD values (µg/L) estimated for the TM- GC×GC-TOFMS/FID (reference – red bars) and FM-GC×GC-MS/FID method and set-up #3a (green bars) on FID signal (**5.4.4A**) and TIC MS signal (**5.4.4B**).

In accordance with MS results, the FID channel sensitivity with the FM-GC×GC was perfectly comparable to that of the TM-GC×GC, revealing an average LOD of 6.36 µg/L *vs.* 6.25 µg/L of the TM (+ 1%).

LODs also inform about the relative sensitivity of the two detectors (*i.e.,* MS and FID) and, at the same time, confirm that TIC MS exceeds FID of a factor of 2.3. Of course, by selecting diagnostic ion traces, MS can be even more sensitive and, at the same time, more flexible enabling to overcome co-elution issues.

The next section examines pattern reliability, through retention times precision and responses stability.

5.4.4.3.3 Repeatability: retention times and responses

Retention time stability is a fundamental characteristic for GC×GC separations, since a primary criterion for analytes identification is their position in the 2D pattern. **Table 5.4.1** reports precision data, expressed as RSD % on ${}^{1}t_R$ and ${}^{2}t_R$ calculated over 8 calibration points and 2 analytical replicates each (n=16 runs). Along the ^{1}D, absolute retention times for the FM method were highly similar showing a RSD % of 0.12. Slightly higher values were obtained for ^{2}D retention, with an RSD% of 0.98. To note, retention of analytes with ^{2}D tailing and/or distortion effects is less precise; linalool resulted in a 1.55 RSD % while vanillin and α-amyl cinnamaldehyde had RSD% of 1.40 and 1.68, respectively.

Responses were indeed highly stable; for the FID channel, absolute 2D volumes registered an average precision of 3.51 % (RSD) while normalized values (over respective ISTDs) were on average 2.71%. The TIC MS signal was comparable with RSD% of 3.30 and 3.20 for absolute and normalized responses, respectively.

The next section briefly presents accuracy data on medium complexity fragrances spiked at 1 and 10 mg/L levels.

5.4.4.3.4 Accuracy: medium complexity fragrance mixtures

Accuracy was preliminarily assessed for targeted analytes spiked in commercial raw fragrances at 1 and 10 mg/L concentration levels. Bias was expressed as *relative error %* (**Equation 5.4.7**) and calculated on the FID signal. In case of co-elutions, the TIC-MS data were adopted and indicated in Table with the symbol "$". Results are reported in **Table 5.4.1** and visualized as scatter plots in **Figure 5.4.3B**. Supplementary Material - **Supplementary Figure 5.4.4** shows pseudocolorized chromatographic images of raw commercial fragrances spiked at 1 and 10 mg/L together with targeted peaks template (coloured circles) and connection lines for ISTDs.

The *relative error* at the higher spiking level (*i.e.*, + 10 mg/L) was lower for *flowery-like* TS2 sample with a median of 6.16 %, calculated on absolute values, compared to the 8.82% at the lower level (*i.e.*, + 1 mg/L). For the *citrus-like* TS1 sample, median values were 7.53% (+ 10 mg/L) and 5.8% (+ 1 mg/L). Minimum and maximum error values were always below ±30%. Results are in line with those validated for the same analytes in the reference method,[36] and indicate that the translated FM-GC×GC-MS/FID method is a good candidate for a routine quantification of targeted analytes in medium complexity fragrances.

5.4.5 Conclusions

This study evidences the flexibility of RFF FM-GC×GC while suggesting a rational approach to translate chromatographic conditions by keeping coherent separation patterns and avoiding chromatographic distortions (overloading of the accumulation loop, generation of asymmetrical 2D peaks etc.). Moreover, the method translation enables the operator to obtain a separation power in line with a reference methodology with TM-GC×GC and, thanks to a rational procedure, to exploit the flexibility by acting on column characteristics that have direct impact on re-injection efficiency and analysis time.

The best FM configuration, when tested for performances of interest in the context of quantitative profiling, demonstrated linearity, sensitivity, and accuracy comparable to the TM counterpart. However, the need for higher flows to the ^{2}D of a FM system, at least to achieve adequate separation power, slightly limits system performances resulting either in an equivalent separation power at the cost of sensitivity[33] or in a sensitivity and quantitation consistency at the cost of ~20 % separation power.

5.4.6 Supplementary Material

Supplementary material at the Google Drive's link:
https://drive.google.com/drive/folders/1dw3d3BvjJIrAmzTWl_qhkf1GFTmHyIBF?usp=sharing

References

(1) Klee, M. S.; Cochran, J.; Merrick, M.; Blumberg, L. M. Evaluation of Conditions of Comprehensive Two-Dimensional Gas Chromatography That Yield a near-Theoretical Maximum in Peak Capacity Gain. *J. Chromatogr. A* **2015**, *1383*, 151–159. https://doi.org/10.1016/j.chroma.2015.01.031.

(2) Seeley, J. V.; Micyus, N. J.; Bandurski, S. V.; Seeley, S. K.; McCurry, J. D. Microfluidic Deans Switch for Comprehensive Two-Dimensional Gas Chromatography. *Anal. Chem.* **2007**, *79* (5), 1840–1847. https://doi.org/10.1021/ac061881g.

(3) Seeley, J. V.; Seeley, S. K. Multidimensional Gas Chromatography: Fundamental Advances and New Applications. *Anal. Chem.* **2013**, *85* (2), 557–578. https://doi.org/10.1021/ac303195u.

(4) Tranchida, P. Q.; Franchina, F. A.; Dugo, P.; Mondello, L. Flow-Modulation Low-Pressure Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2014**, *1372*, 236–244. https://doi.org/10.1016/j.chroma.2014.10.097.

(5) Tranchida, P. Q.; Franchina, F. A.; Dugo, P.; Mondello, L. Use of Greatly-Reduced Gas Flows in Flow-Modulated Comprehensive Two-Dimensional Gas Chromatography-Mass Spectrometry. *J. Chromatogr. A* **2014**, *1359*, 271–276. https://doi.org/10.1016/j.chroma.2014.07.054.

(6) Tranchida, P. Q.; Maimone, M.; Franchina, F. A.; Bjerk, T. R.; Zini, C. A.; Purcaro, G.; Mondello, L. Four-Stage (Low-)Flow Modulation Comprehensive Gas Chromatography□ quadrupole Mass Spectrometry for the Determination of Recently-Highlighted Cosmetic Allergens. *J. Chromatogr. A* **2016**, *1439*, 144–151. https://doi.org/10.1016/j.chroma.2015.12.002.

(7) Bueno, P. A.; Seeley, J. V. Flow-Switching Device for Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2004**, *1027* (1–2), 3–10. https://doi.org/10.1016/j.chroma.2003.10.033.

(8) Griffith, J. F.; Winniford, W. L.; Sun, K.; Edam, R.; Luong, J. C. A Reversed-Flow Differential Flow Modulator for Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2012**, *1226*, 116–123. https://doi.org/10.1016/j.chroma.2011.11.036.

(9) Duhamel, C.; Cardinael, P.; Peulon-Agasse, V.; Firor, R.; Pascaud, L.; Semard-Jousset, G.; Giusti, P.; Livadaris, V. Comparison of Cryogenic and Differential Flow (Forward and Reverse Fill/Flush) Modulators and Applications to the Analysis of Heavy Petroleum Cuts by High-Temperature Comprehensive Gas Chromatography. *J. Chromatogr. A* **2015**, *1387*, 95–103. https://doi.org/10.1016/j.chroma.2015.01.095.

(10) Cordero, C.; Rubiolo, P.; Cobelli, L.; Stani, G.; Miliazza, A.; Giardina, M.; Firor, R.; Bicchi, C. Potential of the Reversed-Inject Differential Flow Modulator for Comprehensive Two-Dimensional Gas Chromatography in the Quantitative Profiling and Fingerprinting of Essential Oils of Different Complexity. *J. Chromatogr. A* **2015**, *1417*, 79–95. https://doi.org/10.1016/j.chroma.2015.09.027.

(11) Cordero, C.; Rubiolo, P.; Reichenbach, S. E.; Carretta, A.; Cobelli, L.; Giardina, M.; Bicchi, C. Method Translation and Full Metadata Transfer from Thermal to Differential Flow Modulated Comprehensive Two Dimensional Gas Chromatography: Profiling of Suspected Fragrance Allergens. *J. Chromatogr. A* **2017**, *1480*, 70–82.

https://doi.org/10.1016/j.chroma.2016.12.011.

(12) Magagna, F.; Liberto, E.; Reichenbach, S. E.; Tao, Q.; Carretta, A.; Cobelli, L.; Giardina, M.; Bicchi, C.; Cordero, C. Advanced Fingerprinting of High-Quality Cocoa: Challenges in Transferring Methods from Thermal to Differential-Flow Modulated Comprehensive Two Dimensional Gas Chromatography. *J. Chromatogr. A* **2018**, *1536*, 122–136. https://doi.org/10.1016/j.chroma.2017.07.014.

(13) Agilent Reverse-Inject modulator.

(14) SepSolve Flow Modulator.

(15) Seeley, J. V.; Schimmel, N. E.; Seeley, S. K. The Multi-Mode Modulator: A Versatile Fluidic Device for Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2018**, *1536*, 6–15. https://doi.org/10.1016/j.chroma.2017.06.030.

(16) Giddings, J. C. Sample Dimensionality: A Predictor of Order-Disorder in Component Peak Distribution in Multidimensional Separation. *J. Chromatogr. A* **1995**, *703* (1–2), 3–15. https://doi.org/10.1016/0021-9673(95)00249-M.

(17) Reichenbach, S. E.; Tian, X.; Cordero, C.; Tao, Q. Features for Non-Targeted Cross-Sample Analysis with Comprehensive Two-Dimensional Chromatography. *J. Chromatogr. A* **2012**, *1226*, 140–148. https://doi.org/10.1016/j.chroma.2011.07.046.

(18) Cordero, C.; Liberto, E.; Bicchi, C.; Rubiolo, P.; Reichenbach, S. E.; Tian, X.; Tao, Q. Targeted and Non-Targeted Approaches for Complex Natural Sample Profiling by GC×GC-QMS. *J. Chromatogr. Sci.* **2010**, *48* (4), 251–261. https://doi.org/10.1093/chromsci/48.4.251.

(19) Kiefl, J.; Cordero, C.; Nicolotti, L.; Schieberle, P.; Reichenbach, S. E.; Bicchi, C. Performance Evaluation of Non-Targeted Peak-Based Cross-Sample Analysis for Comprehensive Two-Dimensional Gas Chromatography-Mass Spectrometry Data and Application to Processed Hazelnut Profiling. *J. Chromatogr. A* **2012**, *1243*, 81–90. https://doi.org/10.1016/j.chroma.2012.04.048.

(20) Purcaro, G.; Cordero, C.; Liberto, E.; Bicchi, C.; Conte, L. S. Toward a Definition of Blueprint of Virgin Olive Oil by Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2014**, *1334*, 101–111. https://doi.org/10.1016/j.chroma.2014.01.067.

(21) Magagna, F.; Valverde-Som, L.; Ruíz-Samblás, C.; Cuadros-Rodríguez, L.; Reichenbach, S. E.; Bicchi, C.; Cordero, C. Combined Untargeted and Targeted Fingerprinting with Comprehensive Two-Dimensional Chromatography for Volatiles and Ripening Indicators in Olive Oil. *Anal. Chim. Acta* **2016**, *936*, 245–258. https://doi.org/10.1016/j.aca.2016.07.005.

(22) Klee, M. S.; Blumberg, L. M. Theoretical and Practical Aspects of Fast Gas Chromatography and Method Translation. *J. Chromatogr. Sci.* **2002**, *40* (5), 234–247. https://doi.org/10.1093/chromsci/40.5.234.

(23) Blumberg, L. M. Theory of Fast Capillary Gas Chromatography. Part 1: Column Efficiency. *J. High Resolut. Chromatogr.* **1997**, *20* (11), 597–604. https://doi.org/10.1002/jhrc.1240201106.

(24) Blumberg, L. M. Theory of Fast Capillary Gas Chromatography Part. 2: Speed of Analysis. *J. High Resolut. Chromatogr.* **1997**, *20* (12), 679–687. https://doi.org/10.1002/jhrc.1240201212.

(25) Blumberg, L. M. Theory of Fast Capillary Gas Chromatography - Part 3: Column Performance vs. Gas Flow Rate. *HRC J. High Resolut. Chromatogr.* **1999**, *22* (7), 403–413. https://doi.org/10.1002/(SICI)1521-4168(19990701)22:7<403::AID-JHRC403>3.0.CO;2-R.

(26) SCCP (Scientific Committee on Consumer Products). Opinion on Fragrance Allergens in Cosmetic Products. *Eur. Comm.* **2011**, No. January, 1–136. https://doi.org/10.2773/ISBN.

(27) Giardina, M.; McCurry, J. D.; Cardinael, P.; Semard-Jousset, G.; Cordero, C.; Bicchi, C. Development and Validation of a Pneumatic Model for the Reversed-Flow Differential Flow Modulator for Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2018**, *1577*, 72–81. https://doi.org/10.1016/j.chroma.2018.09.022.

(28) Blumberg, L. M.; Klee, M. S. Metrics of Separation in Chromatography. *J. Chromatogr. A* **2001**, *933* (1–2), 1–11. https://doi.org/10.1016/S0021-9673(01)01256-0.

(29) Blumberg, L. M. Comprehensive Two-Dimensional Gas Chromatography: Metrics, Potentials, Limits. In *Journal of Chromatography A*; 2003; Vol. 985, pp 29–38. https://doi.org/10.1016/S0021-9673(02)01416-4.

(30) Wenzl, T.; Haedrich, J.; Schaechtele, A.; Robouch, P.; Stroka, J. *Guidance Document on the Estimation of LOD and LOQ for Measurements in the Field of Contaminants in Feed and Food. EUR 28099 EN*; 2016. https://doi.org/10.2787/8931.

(31) Eurachem. *Eurachem Guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics.*; 2014. https://doi.org/978-91-87461-59-0.

(32) Blumberg, L. M.; Klee, M. S. Optimal Heating Rate in Gas Chromatography. *J. Microcolumn Sep.* **2000**, *12* (September), 508–514. https://doi.org/10.1002/1520-667X(2000)12:9<508::AID-MCS5>3.0.CO;2-Y.

(33) Agilent Method Translation software.

(34) Rempe, D. W.; Reichenbach, S. E.; Tao, Q.; Cordero, C.; Rathbun, W. E.; Zini, C. A. Effectiveness of Global, Low-Degree Polynomial Transformations for GCxGC Data Alignment. *Anal. Chem.* **2016**, *88* (20), 10028–10035. https://doi.org/10.1021/acs.analchem.6b02254.

(35) Aloisi, I.; Schena, T.; Giocastro, B.; Zoccali, M.; Tranchida, P. Q.; Caramão, E. B.; Mondello, L. Towards the Determination of an Equivalent Standard Column Set between Cryogenic and Flow-Modulated Comprehensive Two-Dimensional Gas Chromatography. *Anal. Chim. Acta* **2020**, *1105*, 231–236. https://doi.org/10.1016/j.aca.2020.01.040.

(36) Belhassen, E.; Bressanello, D.; Merle, P.; Raynaud, E.; Bicchi, C.; Chaintreau, A.; Cordero, C. Routine Quantification of 54 Allergens in Fragrances Using Comprehensive Two-Dimensional Gas Chromatography-Quadrupole Mass Spectrometry with Dual Parallel Secondary Columns. Part I: Method Development. *Flavour Fragr. J.* **2018**, *33* (1), 63–74. https://doi.org/10.1002/ffj.3416.

(37) Khummueng, W.; Harynuk, J.; Marriott, P. J. Modulation Ratio in Comprehensive Two-Dimensional Gas Chromatography. *Anal. Chem.* **2006**, *78* (13), 4578–4587. https://doi.org/10.1021/ac052270b.

(38) Chaintreau, A.; Cicchetti, E.; David, N.; Earls, A.; Gimeno, P.; Grimaud, B.; Joulain, D.; Kupfermann, N.; Kuropka, G.; Saltron, F.; Schippa, C. Collaborative Validation of the Quantification Method for Suspected Allergens and Test of an Automated Data Treatment.

J. Chromatogr. A **2011**, *1218* (43), 7869–7877. https://doi.org/10.1016/j.chroma.2011.08.072.

5.5 Exploring the extra-virgin olive oil volatilome by adding extra dimensions to comprehensive two-dimensional gas chromatography and time of flight mass spectrometry featuring tandem ionization: validation of ripening markers in headspace linearity conditions

Federico Stilo[1], Erica Liberto[1], Stephen E. Reichenbach[2,3], Qingping Tao[3], Carlo Bicchi[1] and Chiara Cordero[1*]

[1]Dipartimento di Scienza e Tecnologia del Farmaco, Università degli Studi di Torino, Turin, Italy
[2]Computer Science and Engineering Department, University of Nebraska – Lincoln, NE, USA
[3]GC Image LLC, Lincoln, NE, USA

*Corresponding author:
Prof. Chiara E. Cordero - Dipartimento di Scienza e Tecnologia del Farmaco, Università di Torino, Via Pietro Giuria 9, I-10125 Torino, Italy – e-mail: chiara.cordero@unito.it; phone: +39 011 6707172;

Received: March 25, 2020

Revised: June 29, 2020

Accepted: June 29, 2020

Published: August 05, 2020

DOI: 10.1093/jaoacint/qsaa095

Journal of AOAC INTERNATIONAL 2020

5.5.1 Abstract

Comprehensive two-dimensional gas chromatography combined with time-of-flight mass spectrometry (GC×GC-TOF MS) is the most informative analytical approach for chemical characterization of the complex food volatilome. Key analytical features include separation power and resolution enhancement; improved sensitivity; and structured separation patterns from chemically correlated analytes. In this study, we explore the complex extra-virgin olive oil volatilome by combining headspace (HS) solid-phase microextraction, applied under HS linearity conditions to GC×GC-TOF MS featuring hard and soft ionization in tandem. Multiple analytical dimensions are combined in a single run and evaluated in terms of chemical dimensionality; method absolute and relative sensitivity; identification reliability provided by spectral signatures acquired at 70 and 12 eV; and dynamic and linear range of response provided by soft ionization.

Method effectiveness is validated on a sample set of oils from *Picual* olives at different ripening stages. Ripening markers (3,4-diethyl-1,5-hexadiene (*RS/SR*), 3,4-diethyl-1,5-hexadiene (meso), *(5Z)*-3-ethyl-1,5-octadiene, *(5E)*-3-ethyl-1,5-octadiene, *(E,Z)*-3,7-decadiene and *(E,E)*-3,7-decadiene, *(Z)*-2-hexenal, *(Z)*-3-hexenal and *(Z)*-3-hexenal, *(E)*-2-pentenal, *(Z)*-2-pentenal, 1-pentanol, 1-penten-3-ol, 3-pentanone, and 1-penten-3-one) and quality indexes (*(Z)*-3-Hexenal/Nonanal, *(Z)*-3-Hexenal/Octane, *(E)*-2-Pentenal/Nonanal, and *(E)*-2-Pentenal/Octane) are confirmed for their validity in HS linearity conditions.

For the complex olive oil volatilome, the proposed approach offers concrete advantages for the validation of the informative role of existing analytes while suggesting new potential markers to be studied in larger sample sets. The accurate fingerprinting of volatiles by HS-SPME operating in HS linearity conditions followed by GC×GC-TOF MS featuring tandem ionization gives the opportunity to improve the quality of analytical data and reliability of results.

Key words

comprehensive two-dimensional gas chromatography, time-of-flight mass spectrometry featuring tandem ionization; extra-virgin olive oil volatilome, ripening markers, headspace linearity

5.5.2 Introduction

Comprehensive two-dimensional gas chromatography combined with time-of-flight mass spectrometric detection (GC×GC-TOF MS) is nowadays the most informative analytical approach for chemical characterization of the complex food volatilome.[1–3] By this technique, the detailed profiling of known analytes (*i.e.*, targeted compounds) is extended from two- to five-fold compared to mono-dimensional (1D) GC-MS,[2,4–6] while a new concept of fingerprinting can be exploited by investigating bi-dimensional (2D) separation patterns with dedicated algorithms.[7,8] Key analytical features of GC×GC include: (a) separation power and resolution enhancement, provided by the appropriate combination of two separation dimensions with orthogonal principles of discrimination,[9,10] (b) improved sensitivity, due to the effective band focusing-in-space produced by thermal modulators; and (c) structured separation patterns from chemically related groups of analytes, that are useful for identity confirmation and structural elucidation. All these characteristics make GC×GC-TOF MS the platform of choice to obtain the highest level of information encrypted in the complex volatiles fractions of foods, while supporting reliable and robust results within the challenging scenario of food quality assessment. Origin traceability, technological impact, and aroma quality are areas of active research for GC×GC food applications.[2,4,5,11–14]

Although GC×GC-TOF MS combined to advanced fingerprinting data processing can be considered a well-established technique, several challenging aspects have to be considered when volatiles must be accurately profiled, and their relative or absolute amounts adopted to define specific sample characteristics. The main challenge relates to sampling representativeness in terms of both qualitative and quantitative distributions of analytes and their actual information about sample composition.

Headspace sampling (HS) should be designed carefully in agreement with research aims with quantitative results validated through accepted protocols and/or guidelines. Very often, internal normalization procedures replace accurate quantitation by standard addition (SA), external calibration (ESTD), or stable isotope dilution (SI), and may provide inaccurate and misleading results.[15–17]

In this study, we explore the complex extra-virgin olive oil (EVOO) volatilome by combining HS-SPME, applied under HS linearity conditions,[15,18] to GC×GC-TOF MS featuring hard and soft ionization in tandem (*i.e.*, tandem ionization). The combination of multiple analytical dimensions in a single run is examined in terms of: (a) chemical dimensionality,[19] (b) method absolute and relative sensitivity achieved by the combination of tandem signals acquired at two different ionization energies (*e.g.*, 70 and 12 eV); (c) identification reliability provided by complementary spectral signatures. Then, method effectiveness is validated on a sample set consisting of EVOOs from *Picual* olives at different ripening stages. Ripening markers[20] and quality indexes[13,21–24] are tested for their validity in HS linearity conditions.

5.5.3 Materials and methods

5.5.3.1 Reference compounds and chemicals

Pure standards of *n*-alkanes (from n-C6 to n-C25) for system evaluation and Linear Retention Index (I^T) calibration, internal standards (ISs) α-thujone, β-thujone and methyl-2-octynoate for response normalization, and pure reference compounds for targeted analytes' identity confirmation were supplied by Merck (Milan, Italy). Cyclohexane (HPLC grade) for *n*-alkanes standard solution (at 100 mg/L) and pure dibutyl phthalate for ISs working solutions (at 100 mg/L) were also from Merck (Milan, Italy).

5.5.3.2 Olive oil samples

Extra Virgin Olive Oil (EVOO) samples were supplied from the University of Granada (Spain) by Prof. Luis Cuadros-Rodríguez. They were obtained from olives of the *Picual* cultivar harvested in the Granada Altipiano region, in a plot named *Baza*. Each sample was available in duplicate and obtained by mixing olives from five different trees to have homogeneous and representative samples. Olives for oil production were harvested at four different ripening stages: November 10−12, 2014; November 24−28, 2014; December 16−17, 2014; and January 12−15, 2015, and classified by oil quality (Extra Virgin - EVOO; Virgin - VOO; or *Lampante* - LOO). Samples' acronyms and characteristics are summarized in **Table 5.5.1**. Oil qualifications were by a certified laboratory (ISO 17025:2018) and according to Commission Regulation (EEC) No. 1604/2019 of September 27, 2019 and IOC Standard COI/T.15/NC No 3/Rev. 12. Some quality indices are reported in **Table 5.5.1**, including the sensory panel test results.

Table 5.5.1. List of analyzed samples. They were produced under controlled conditions in a plot named Baza (Altipiano de Granada, Spain). Samples are listed with acronyms, harvest period, quality parameters according to Commission Implementing Regulation (EEC) No 1604/2019 of September 27, 2019, sensory evaluation results (Md: median of defects – Mf: median of fruity notes), and commercial classification.

Sample Acronym	Harvest period	Acidity (%)	Peroxide index (mEq O_2/kg)	K_{232}	K_{270}	ΔK	Md	Mf	Classification
Baza-1-A	November 10-12	0.20	5	1.84	0.20	0.00	0.00	5.0	EVOO
Baza-2-A	November 24-28	0.20	3	1.60	0.20	0.00	0.00	4.1	EVOO
Baza-3-A	December 16-17	0.20	5	1.17	0.20	0.00	> 0.00	1.3	VOO
Baza-4-A	January 12-15	0.40	11	1.11	0.10	0.00	> 0.00	0.0	LOO
Baza-1-B	November 10-12	0.20	4	1.92	0.20	0.00	0.00	5.2	EVOO
Baza-2-B	November 24-28	0.10	3	1.65	0.20	0.00	0.00	3.8	EVOO
Baza-3-B	December 16-17	0.20	6	1.28	0.10	0.00	> 0.00	1.7	VOO
Baza-4-B	January 12-15	0.40	13	1.12	0.10	0.00	> 0.00	0.0	LOO

5.5.3.3 Headspace solid phase micro extraction sampling devices and conditions

Volatiles were sampled by automated headspace solid-phase microextraction (HS-SPME) with a SPR Autosampler for GC (SepSolve-Analytical). Different commercial fiber coatings were tested in order to evaluate their performance and effectiveness in describing the olive oil volatilome. In particular, the tested coatings were: (*a*) Divinylbenzene/Carboxen/Polydimethylsiloxane (DVB/CAR/PDMS) 50/30 µm (stationary phase/film thickness) - 1 cm length; (*b*) DVB/PDMS 65 µm - 1 cm length; (*c*) Polyacrylate (PA) 85 µm - 1 cm length; and (*d*) PDMS 75 µm - 1 cm length. All fibers/devices were from Merck (Bellefonte, PA, USA).

The ISs for system performance evaluation and responses normalization were preloaded onto the SPME device according to the procedure proposed by Wang *et al.*[25,26] by exposing the extraction phase to the headspace of a 5.0 µL of α/β-thujone and methyl 2-octynoate solution at 100 mg L^{-1} in dibutyl phthalate placed in a 20 mL vial for 5 minutes at 40 °C. Sampling was carried out on 0.100 ± 0.005 g of olive oil precisely weighed in 20 mL headspace vials, matching for headspace linearity conditions [15] and kept at 40 °C for 60 min under constant stirring. After sampling, the SPME device was automatically transferred to the split/splitless injection port of the GC×GC

system kept at 270°C; desorption time was set at 5 min. Each sample was analyzed in triplicate and precision data on retention times and targeted 2D peaks normalized volumes expressed as % Relative Standard Deviation (% RSD). Average % RSD for ^1t_R was 2.2, while for ^2t_R it was 0.4. 2D Peak volumes mean %RSD was different as a function of the fiber considered. In decreasing order, % RSD on 2D Peak volumes were 12.1% for PA, 10.7% for PDMS, 6.2% for DVB/PDMS and 4.9 % for DVB/CAR/PDMS. Based on these precision data, successive elaborations considered the average value deriving from the three analytical replicates of each sample.

5.5.3.4 Comprehensive two-dimensional gas chromatography: system configuration and parameters

GC×GC-MS analyses were run with a system configuration consisting of an Agilent 7890B GC unit (Agilent Technologies, Wilmington DE, USA) coupled with a BenchTOF-Select™ time-of-flight mass spectrometer featuring tandem ionization (Markes International, Llantrisant, UK). The GC transfer line was set at 270°C as was the ion source. TOF MS acquisition was between 40-350 m/z at 50 Hz per channel and multiplexing ionization energy was between 70 and 12 eV (*i.e.*, tandem ionization).

The column set was configured as follows: ^{1}D SolGel-Wax column (100% polyethylene glycol, 30 m × 0.25 mm d_c, 0.25 μm d_f) from Trajan Scientific and Medical (Ringwood, Australia) coupled with a ^{2}D OV1701 column (86% polydimethylsiloxane, 7% phenyl, 7% cyanopropyl, 1 m × 0.1 mm d_c, 0.10 μm d_f) from Mega (Legnano, Milan, Italy). Fused silica capillary loop dimensions were 1.0 m length and 0.1 mm d_c. Fiber thermal desorption into the GC injector port was under the following conditions: split/splitless injector in split mode, split ratio 1:20, injector temperature 270°C, and thermal desorption 5 minutes. Carrier gas was helium at a constant flow of 1.3 mL/min. The oven temperature program was from 40°C (1 min) to 200°C at 3°C/min then to 250°C at 10°C/min (5 min). No secondary oven was adopted in the system set-up.

The *n*-alkanes liquid sample solution for I^T determination was analyzed under the following conditions: split/splitless injector in split mode, split ratio 1:50, injector temperature 270°C, and injection volume 1μL.

The system was equipped with a two-stage KT 2004 loop type thermal modulator (Zoex Corporation, Houston, TX) cooled with liquid nitrogen and controlled by Optimode v2.0 (SRA Instruments, Cernusco sul Naviglio, Milan, Italy). Modulation period (P_M) was set at 4s, with a hot jet pulse time of 250 ms. A Mass Flow Controller (MFC) reduced the cold-jet stream from 45% to 8% of the total flow with a linear function along the run duration.

5.5.3.5 Data acquisition and elaboration software

Raw data acquisition was by TOF-DS software (Markes International, Llantrisant, UK). 2D data processing was by GC Image GC×GC Software, ver. 2.9 (GC Image, LLC, Lincoln NE, USA). Data analysis and chemometrics were by XLSTAT (Addinsoft, New York, USA).

5.5.4 Results and Discussion

Despite the great potential of GC×GC in exploring the chemical complexity of olive oil volatilome, just a few studies are available in this field and none of them address the challenging aspect of HS linearity to obtain reliable and representative fingerprints of volatiles. Vaz Freire *et al.*[27] investigated the volatiles patterns of Portuguese olive varieties *Galega Vulgar*, *Cobrançosa* e

Carrasquenha by combining HS-SPME with a multi-coating fiber (*e.g.*, 2 cm 50/30 μm DVB/Carboxen/PDMS) with GC×GC-TOF MS and image-features data analysis [7]. The developed method included the HS-SPME sampling of 12 g of olive oil in a 20 mL HS vial for 30' at 40°C. These conditions exceed HS analyte linearity[15] but provide a good coverage of samples chemical dimensions. Cajka et al.[28] developed a GC×GC-TOF-MS profiling approach able to define geographical origin indicators and cultivar markers. They considered 914 samples over three production seasons and obtained from olives harvested in Liguria in northwest Italy (n = 210) and other regions of Italy, Spain, France, Greece, Cyprus, and Turkey (n = 704). Sampling was carried out by a DVB/Carboxen/PDMS 2 cm 50/30 μm fiber on 2.0 g of olive oil placed in a 10 mL HS vial. Sampling temperature was 40°C for 15' extraction time. More recently Lukić et al.[29] surveyed different olive cultivars and geographical areas by sampling 3.00 g of oil in 20 mL HS vials at 50°C for 40' and Da Ros et al.[4] profiled volatiles from Italian EVOOs from 2.0 g of sample in 20 mL HS vial at 40°C for 30'. In both cases, targeted analyte normalized responses over the internal standard were reported as concentrations (mg/kg) relative to the IS. Volatiles sampling outside HS linearity conditions brings to inaccurate quantitation by IS normalization that ranges between -90 % for highly volatiles to + 1600 % for medium-to-low volatiles.[15]

In this scenario, the possibility of capturing in great detail the chemical complexity of EVOOs while avoiding HS saturation effects is of great interest, especially if response data can be adopted as marker or indicator of sample qualities. However, due to the very wide dynamic range of concentrations of EVOOs volatiles that span over 3 orders of magnitude,[6,30] profiling methods face several challenges and can benefit greatly from GC×GC TOF MS key-features.

Based on previous studies on cocoa volatilome[31] and on hazelnuts primary metabolites signatures [32], tandem ionization is used here to test its capability to extend method dynamic range through relative and absolute sensitivity expressed by hard (70 eV) and soft (12 eV) ionization, when acquired in tandem across a single analytical run.[33,34] Moreover, tandem signal data are examined for their complementary nature to provide reliable identity confirmation, to add additional specificity to the profiling/fingerprinting method and to cross-validate the role of ripening markers and quality indicators.

Following a logic path, the next sections introduce the complexity of EVOO volatilome as it results from multidimensional analysis and discuss the complementary information provided by HS-SPME carried out with different fibers and additional features deriving from tandem signal data. The validation of ripening markers and quality indexes complete the discussion while providing evidence of the concrete advantages of a multi-dimensional investigation.

5.5.4.1 Informative chemical patterns in the EVOO volatilome

EVOO volatiles fractions are complex mixtures of compounds belonging to different chemical classes and covering a wide range of volatility, as illustrated by the list of 136 targeted compounds in **Table 5.5.**2. Analytes are listed according to their elution order in the polar × medium polar column configuration and are reported together with their retention times in both chromatographic dimensions (^1t_R - min and ^2t_R - sec), experimental I^T, odor threshold (OT) values,[22] and sensory descriptors. Volatiles were reliably identified by MS spectral similarity (NIST Identity algorithm),[35] fixing a threshold value of 850 for direct match factor (DMF) and 875 for the reverse match factor (RMF), and linear retention index (I^T) coherence ($I^T \pm 15$ units). For those analytes where a pure reference compound was available, identity was confirmed by direct injection of reference standard solutions.

Table 5.5.2. List of the 136 targeted compounds together with their retention times in the chromatographic dimensions (^1t_R min, ^2t_R sec). experimental I^T, odor threshold (µg/kg), odor descriptor, and relative concentration factor calculated for the different fibers (Divinylbenzene/Carboxen/Polydimethylsiloxane - DVB/CAR/PDMS; DVB/PDMS; Polyacrylate – PA; PDMS) from TIC signal. Analytes are grouped according to chemical classes.

Chemical class	Compound Name	^1t_R (min)	^2t_R (sec)	Exp I^T	Odor threshold (µg/kg)	Odor descriptor	PA	DVB-PDMS	PDMS	DVB-CAR-PDMS
Saturated aldehydes	Hexanal	12.87	1.66	1220	75	Tallowy, leaf-like	20	80	23	100
	Heptanal	16.73	1.98	1317	-	Green, aldehydic	40	2	59	100
	Octanal	20.93	2.20	1419	320	Fatty, sharp	38	2	59	100
	Nonanal	25.07	2.36	1518	150	Fatty, waxy, pungent	15	100	24	100
	Decanal	29.07	2.52	1617	650	Penetrating, sweet, waxy	45	100	88	65
	Benzaldehyde	29.93	0.88	1639	-	Almond, burnt sugar	25	66	18	100
	Undecanal	32.87	2.62	1715	-	Sweet, fatty, floral-citrus	70	100	92	82
	Dodecanal	36.47	2.74	1815	-	Soapy, fatty, aldehydic	16	32	100	50
	4-Ethylbenzaldehyde	37.33	1.22	1839	-	Almond, bitter, almond	0	100	13	93
	Tridecanal	39.87	2.84	1911	-	Fresh, clean, soapy, citrus	64	95	95	100
	Cinnamaldehyde	40.27	0.96	1923	-	Sweet, spice, cinnamon	0	100	0	68
	Tetradecanal	43.07	2.96	2005	-	Fatty, waxy, amber, incense	23	43	100	50
	Pentadecanal	46.32	3.07	2111	-	Fresh, waxy	73	87	93	100
	Hexadecanal	49.51	3.18	2209	-	-	12	22	100	13
Average Saturated aldehydes							**31**	**66**	**62**	**80**
Unsaturated aldehydes	(Z)-2-Pentenal	13.73	1.22	1243	-	-	18	42	21	100
	(E)-2-Pentenal	14.60	1.24	1265	300	Pungent, apple-like	13	29	16	100
	(Z)-3-Hexenal	15.27	1.34	1282	3	Leaf-like	8	68	10	100
	(Z)-2-Hexenal	17.40	1.54	1335	-	Fruity	18	67	23	100
	(E)-2-Hexenal	18.13	1.52	1352	420	Bitter almond, green	8	65	13	100
	(E)-2-Heptenal	22.40	1.68	1454	5	Fatty, almond-like	43	89	66	100
	(E,Z)-2,4-Hexadienal	25.27	1.06	1523	-	Green	0	8	0	100

	(E,E)-2,4-Hexadienal	25.53	1.04	1529	-	Green	0	10	0	100
	(E)-2-Octenal	26.53	1.84	1554	4	Fresh, fatty, waxy	34	76	49	100
	(E,E)-2,4-Heptadienal	27.80	1.28	1585	3620	Fatty, green, oily	0	100	10	99
	(E)-2-Nonenal	30.47	2.00	1653	900	Fatty, green, cucumber	32	61	39	100
	(E)-2-Decenal	34.27	2.10	1754	10	Waxy, fatty, earthy	12	44	15	100
	(E)-2-Undecenal	37.80	2.22	1852	-	Fresh, fruity, orange	16	26	23	100
	(E)-2-Dodecenal	40.67	2.38	1877	-	Citrus, waxy, green	49	58	62	100
	(E)-2-Tetradecenal	47.13	2.63	2069	-	-	31	41	72	100
Average Unsaturated aldehydes							**19**	**53**	**29**	**100**
Alcohols	Ethanol	8.53	0.34	1091	30000	Alcoholic, ethereal	14	37	26	100
	2-Methyl-1-propanol	12.93	0.64	1222	480	-	52	75	42	100
	1-Butanol	14.93	0.70	1274	150	Winey	27	67	39	100
	1-Penten-3-ol	15.53	0.72	1289	-	Pungent	8	76	11	100
	3-Methyl-1-butanol	17.40	0.80	1335	100	-	15	16	19	100
	1-Pentanol	19.13	0.76	1376	470	Sweet, pungent	20	57	28	100
	(Z)-2-Penten-1-ol	21.60	0.60	1435	-	Green, plastic, rubber	9	45	7	100
	(E)-2-Penten-1-ol	21.87	0.62	1441	-	Mushroom	9	100	7	45
	2-Heptanol	21.93	1.30	1442	10	Fruity, green, earthy	27	64	40	100
	1-Hexanol	23.27	0.86	1474	400	Fruity, banana, soft	9	92	2	100
	(Z)-3-Hexen-1-ol	24.47	0.76	1503	1500	Leaf-like	19	14	19	100
	(Z)-2-Hexen-1-ol	25.27	0.72	1523	5000	Green grass, leaves	0	100	17	30
	(E)-2-Hexen-1-ol	25.60	0.72	1531	-	Green grass, leaves	22	0	0	100
	1-Heptanol	27.20	0.98	1570	10	Herb	44	88	54	100
	6-Methyl-5-hepten-2-ol	27.47	1.06	1577	2000	Green, sweet	17	82	100	72
	1-(2-Methoxy-1-methylethoxy)-2-propanol	28.00	1.10	1590	-	-	0	16	11	100
	2-Ethyl-1-hexanol	28.47	1.16	1601	-	-	31	100	14	15

	1-Octanol	30.93	1.10	1665	100	Moss, nut, mushroom	20	100	20	90
	1-Nonanol	34.47	1.20	1759	280	Fresh, clean, floral	29	100	33	78
	1-Decanol	37.87	1.32	1854	-	Fatty, waxy	62	100	69	90
	1-Undecanol	41.07	1.46	1946	-	Fresh, waxy, rose, soapy	69	88	98	100
	Benzyl alcohol	41.27	0.50	1952	-	Sweet, fruity	34	100	16	78
	Phenylethyl Alcohol	42.33	0.64	1984	-	Sweet, floral, fresh	21	72	12	100
	1,4-Butanediol	42.40	0.34	1986	-	-	63	96	100	93
	1-Dodecanol	44.13	1.58	2038	-	Earthy, soapy, waxy, fatty	54	100	84	96
	Phenol	44.93	0.34	2063	-	-	73	41	72	100
	1-Tridecanol	47.00	1.70	2128	-	Musty	67	87	100	90
	2-Phenoxyethanol	48.80	0.58	2186	-	-	27	46	39	100
	1-Tetradecanol	49.80	1.82	2219	-	Fruity, waxy, orris, coconut	71	89	100	89
Average Alcohols							**31**	**71**	**41**	**89**
Fatty acids	Acetic acid	27.27	0.20	1572	500	Sour, vinegary	16	60	35	100
	Hexanoic acid	40.53	0.44	1931	700	Goat-like, sweaty	30	89	30	100
	Heptanoic acid	43.73	0.50	2026	100	-	32	100	38	76
	Octanoic acid	46.80	0.60	2121	3000	Fatty, waxy, rancid, oily	22	100	24	57
	Nonanoic acid	49.67	0.68	2215	-	Sweaty, waxy	82	92	100	92
Average Fatty acids							**42**	**81**	**56**	**88**
Ketones	Acetone	6.67	0.42	1026	-	Pungent	12	36	39	100
	3-Pentanone	9.67	1.12	1127	-	Ethereal, acetone	4	38	8	100
	1-Penten-3-one	10.87	0.92	1163	50	Pungent, spicy	9	53	17	100
	2-Heptanone	16.67	2.00	1317	300	Sweet, fruity	29	73	41	92
	2-Octanone	20.80	2.16	1415	510	Mould, green	11	98	16	100
	3-Hydroxy-2-butanone	20.67	0.60	1412	-	Buttery	16	67	23	100
	6-Methyl-5-hepten-2-one	22.80	1.72	1463	1000	Pungent, green	12	95	21	100

	2-Nonanone	24.93	2.32	1514	-	Fresh, sweet, green	12	12	32	100
	2-Decanone	28.87	2.48	1612	-	Orange, floral, peach	39	37	57	100
	2-Undecanone	32.44	2.54	1713	-	Fruity, waxy	54	70	77	100
	Acetophenone	34.33	1.00	1756	-	Sweet pungent almond	57	84	24	100
	2-Dodecanone	36.20	2.70	1808	-	Fruity, floral, orange	51	66	77	100
	2-Cyclohexene-1,4-dione	37.13	0.78	1833	-	-	14	30	18	100
	2-Tridecanone	39.60	2.82	1903	-	Fatty, waxy, coconut	46	100	58	50
	(Z)-6,10-Dimethyl-5,9-Undecadien-2-one	40.93	2.16	1942	-	-	42	100	76	95
	2-Tetradecanone	42.80	2.94	1997	-	-	65	93	93	100
	4-Phenyl-3-buten-2-one	43.87	1.06	2030	-	Sweet, spice, cinnamon	11	100	12	67
	5,6-Decanedione	49.80	1.12	2219	-	-	43	51	60	100
Average Ketones							**37**	**70**	**50**	**93**
Lactones	Butyrolactone	33.60	0.78	1736	-	Creamy, oily, caramel	87	100	25	71
	δ-Valerolactone	37.10	0.98	1786	-	Herbal, tonka, sweet	71	100	43	56
	ɣ-Caprolactone	37.51	0.92	1822	-	-	13	100	9	59
	ε-Hexalactone	41.80	1.06	1968	-	-	71	89	100	96
	ɣ-Nonalactone	46.07	1.46	2097	-	Coconut, creamy, waxy	30	100	28	71
	γ-Decalactone	49.27	1.56	2202	-	Fresh, oily, waxy, peach	65	100	51	90
Average Lactones							**48**	**92**	**50**	**79**
Esters	Ethyl acetate	7.67	0.68	1058	940	Fruity, sweet, green	9	100	19	42
	Butyl acetate	12.47	1.60	1210	300	Pear	22	55	26	100
	Hexyl acetate	20.20	2.22	1401	1040	Fruity	15	100	26	99
	(Z)-3-Hexenyl acetate	21.93	1.80	1442	1040	Sweet	9	83	18	100
	Methyl benzoate	33.33	1.06	1729	-	-	11	100	6	83
	2-Ethyl-3-hydroxyhexyl 2-methylpropanoate	38.47	1.08	1871	-	-	21	100	19	93
	2-Phenethyl acetate	39.60	1.20	1903	-	Floral, rose, sweet	0	0	0	100

	Methyl 2-oxohexanoate	43.87	0.56	2030	-	-	15	100	6	60
	Methyl anisate	46.80	0.96	2121	-	Herbal, anise,sweet	14	13	13	100
Average Esters							**22**	**79**	**22**	**85**
Hydrocarbons	(Z)-2-Pentene	5.27	0.32	945	-	-	0	100	8	74
	n-Hexane	5.33	0.48	949	-	-	0	100	17	84
	1,4-Pentadiene	5.60	0.38	964	-	-	0	25	6	100
	n-Heptane	5.67	0.92	968	-	Sweet, ethereal	0	100	23	41
	n-Octane	6.47	1.72	1014	940	Solvent, unpleasant	6	34	27	100
	1-Octene	7.00	1.62	1031	-	Gasoline	17	61	55	100
	3,4-Diethyl-1,5-hexadiene (RS/SR)	9.33	2.74	1117	-	-	0	72	6	100
	3,4-Diethyl-1,5-hexadiene (meso)	9.47	2.78	1121	-	-	0	55	7	100
	(5Z)-3-Ethyl-1,5-octadiene	10.60	3.20	1155	-	-	2	89	6	100
	(5E)-3-Ethyl-1,5-octadiene	11.07	3.12	1169	-	-	3	97	7	100
	(E,Z)-3,7-Decadiene	12.80	3.66	1218	-	-	1	88	4	100
	(E,E)-3,7-Decadiene	13.07	3.68	1225	-	-	2	100	4	72
	Tridecene	19.13	1.76	1376	-	-	0	100	23	87
	(E)-4,8-Dimethyl-1,3,7-nonatriene	21.67	2.80	1436	-	-	0	100	7	93
	(Z)-1-Ethenyl-4-ethyl-benzene	26.73	1.64	1559	-	-	0	95	5	100
	(E)-1-Ethenyl-4-ethyl-benzene	27.13	1.66	1568	-	-	0	100	0	73
	1,3-Diethenylbenzene	31.40	1.26	1677	-	-	0	63	0	100
	1,4-Diethenylbenzene	32.07	1.26	1694	-	-	0	57	4	100
	n-Heptadecane	35.60	0.72	1790	-	-	0	57	36	100
Average Hydrocarbons							**6**	**79**	**13**	**92**
Terpenes	δ-3-carene	15.47	3.18	1287	-	-	31	38	48	100
	Limonene	17.40	2.88	1335	-	Citrus, mint	0	100	8	90
	(Z)-β-ocimene	18.73	2.48	1366	-	Floral, citrus	0	100	8	90

	(E)-β-ocimene	19.40	2.50	1382	-	-	0	100	5	81
	Linalool	30.53	1.26	1655	-	Citrus	11	100	19	64
	Pinocarvone	31.73	2.09	1740	-	Minty	11	23	100	35
	Eremophilene	36.60	3.20	1818	-	-	17	100	50	86
	(E,E)-α-Farnesene	37.60	2.88	1847	-	-	31	100	39	84
	Limonene dioxide	47.13	1.24	2132	-	Menthol	87	94	100	96
Average Terpenes							**17**	**78**	**33**	**86**
Others	1-Methoxy hexane	9.00	1.96	1107	-	-	8	68	20	100
	(Z)-1-Methoxy-3-hexene	10.53	1.82	1153	-	-	6	55	13	100
	Benzyl methyl ether	24.93	1.30	1514	-	-	11	10	11	100
	Furfural	27.73	0.86	1593	-	Bready, brown, sweet	7	17	100	28
	Isovanillin, TBDMS derivative	28.87	0.76	1612	-	-	11	100	27	43
	5-Methyl-2-furaldehyde	29.01	0.88	1582	-	Caramellic, brown	0	0	100	9
	Ethylene glycol	29.65	1.23	1621	-	-	100	0	1	1
	1,2-Dimethoxypropanol	29.80	1.00	1636	-	-	0	14	7	100
	2-Furanmethanol	34.27	0.36	1754	-	Alcoholic, musty, sweet	51	62	52	100
	5-Ethyl-2(5H)-furanone	37.87	0.88	1854	-	-	9	97	0	100
	Dimethyl sulfone	42.00	0.46	1974	-	Sulforous, burnt	18	100	20	41
	Diethylene glycol	44.07	0.36	2036	-	Sweet	100	33	32	33
Average others							**25**	**63**	**37**	**69**

The EVOO volatilome expresses a high chemical dimensionality, a parameter defined by Giddings to describe the degree of order or disorder in multidimensional separations.[19] In this perspective, ordered elution patterns for chemically correlated compounds are a peculiar characteristic of the technique. Analytes are distributed according to their polarity/volatility along the ^{1}D, while the selectivity of the ^{2}D enables differential retention along the orthogonal axis of separation discriminating analytes based on their apolar moieties. Primary alcohols and short chain fatty acids elute earlier in the ^{2}D, as illustrated for the *Baza-2-A* sample in **Figure 5.5.1A** and **5.5.1D**, while linear saturated/unsaturated aldehydes and carbonyl compounds (*i.e.*, hydroperoxides cleavage products) are more retained along the ^{2}D axis (**Figure 5.5.1D**). This group of analytes is informative about shelf-life evolution and its abundance is generally correlated with the presence of rancid and fatty notes [21,36]. Saturated and unsaturated hydrocarbons are chemical groups with the highest retention in the ^{2}D (**Figure 5.5.1A** and **5.5.1B**). In particular, 3,4-diethyl-1,5-hexadiene (*RS* or *SR*), 3,4-diethyl-1,5-hexadiene (meso), *(5Z)*-3-ethyl-1,5-octadiene*, (5E)*-3-ethyl-1,5-octadiene, *(E,Z)*-3,7-decadiene, and *(E,E)*-3,7-decadiene are diagnostic markers of early ripening stages [13]. Finally, lipoxygenases (LOXs) derived compounds (C5-C6 alcohols, aldehydes, and ketones – **Figure 5.5.1C**) generated from linoleic and linolenic acids oxidative cleavage are correlated with the perception of positive attributes (*i.e.*, green and fruity notes) and are well resolved over the 2D space.

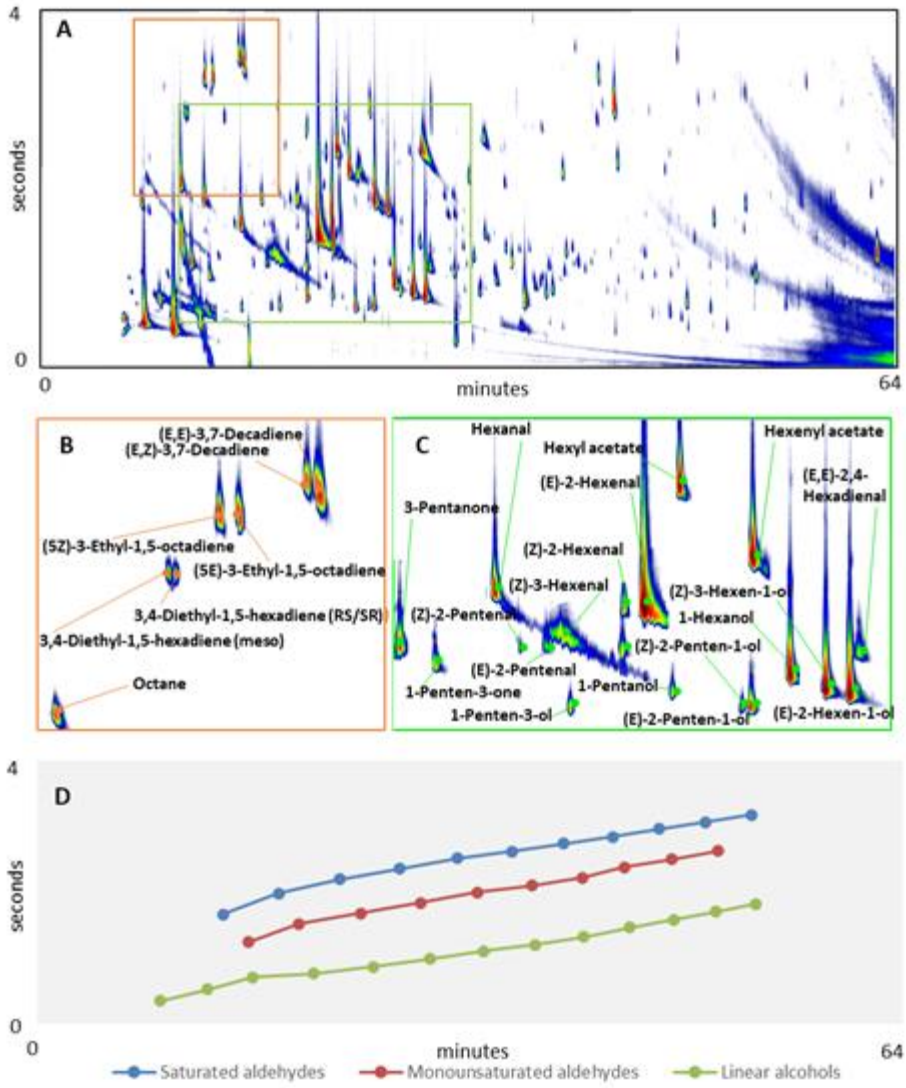

Figure 5.5.1. (**5.5.1A**) Contour plot of the volatile fraction of Baza-1-A sample (0.100 g – 20 mL vial – 40°C 60 min sampling) sampled by HS-SPME with a DVB/CAR/PDMS fiber. Enlarged areas highlight retention patterns for (**5.5.1B**) alkenes and (**5.5.1C**) LOX derived aldehydes and alcohols. (**5.5.1 D**) The ordered elution for saturated and mono-unsaturated aldehydes and primary alcohols are evidenced by a dispersion graph.

Ordered patterns are not only a clear advantage for identification purposes and/or for a more accurate extraction of response data, but also an analytical opportunity since they open new perspectives for dedicated processing approaches defined as *chromatographic fingerprinting.* [5,8,11] An example in this direction is given by visual feature fingerprinting applied here to reveal the differential information provided by HS-SPME with different fiber coatings (described in the next section).

5.5.4.2 Exploring the information capacity of HS-SPME

Sample preparation by HS-SPME deserves a separate discussion within the analytical dimensions exploited in this study. It can be considered as the zeroth dimension of the system[37] as it applies the initial discrimination of analytes as a function of specific characteristics (polarity, volatility, etc.). It should match with the sample's chemical dimensions[19] to give access to the multiple levels of information encrypted in analyte patterns. Headspace sampling approaches

implementing a concentration step are the elective route to profitably explore complex volatilome patterns [15]. In particular, HS-SPME, the most popular high concentration capacity (HCC) technique [38–44], has been widely adopted in this field of applications. Here, we explored the differential effect of extraction coating selectivity and efficiency by testing commercially available SPME fibers with single or multiple phases operating with different extraction mechanisms (partition/absorption or adsorption). This survey is not *per se* a novelty, but it should be a recommended practice when a highly complex matrix is studied.[28] In this application, headspace linearity[18] was also checked for the most informative analytes[15] to verify if under pre-determined temperature/time/phase-ratio conditions, a minimal analyte amount is released from the condensed phase avoiding saturation, thus establishing a linear correspondence between the analyte concentration in the sample (C_0) and its concentration in the gas phase (C_g).[18]

Four different fibers were tested: PDMS (apolar, partition/absorption), PA (medium polar, partition/absorption), DVB/PDMS (apolar/medium polar, absorption/adsorption), DVB/CAR/PDMS (apolar/medium polar, absorption/adsorption). The relative concentration factors were calculated to describe sampling efficiency. In practice, the absolute chromatographic response (2D Peak Volume from Total Ion Current – TIC) of each analyte was divided by that obtained from the most effective fiber. Results were re-scaled in a 0-100 range where 0 corresponds to "not detected" and "100" to the best performing coating. Results are visualized by spider diagram in **Figure 5.5.2,** where analytes are grouped according to their chemical class; individual values are reported in **Table 5.5.2**.

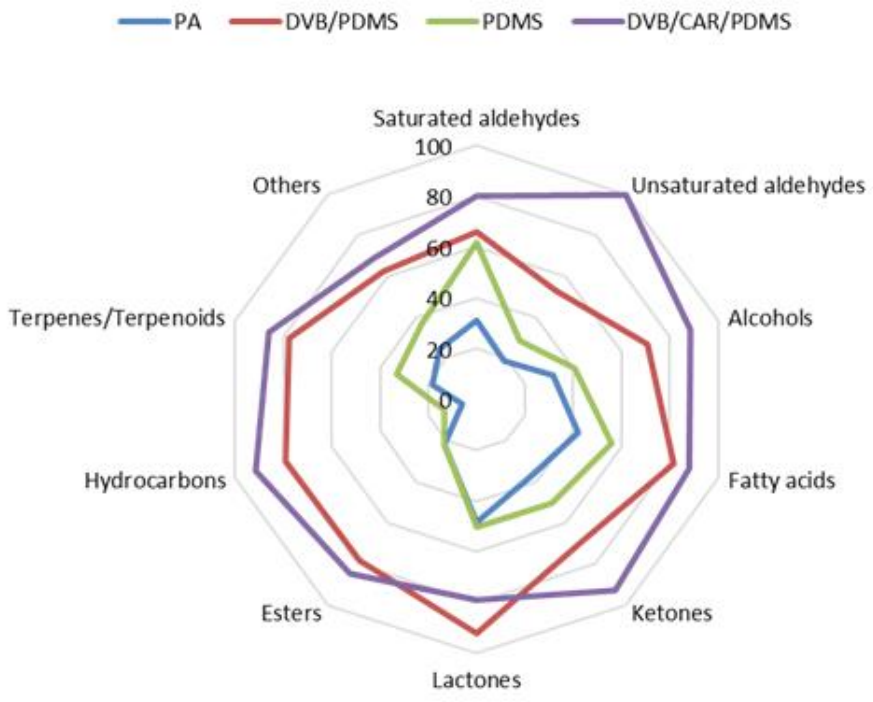

Figure 5.5.2. Spider diagram for tested sampling fibers (Divinylbenzene/Carboxen/Polydimethylsiloxane - DVB/CAR/PDMS; DVB/PDMS; Polyacrylate – PA; PDMS) showing the average relative concentration factor calculated for selected analytes belonging to different chemical classes.

As expected, the DVB/CAR/PDMS fiber, visualized with a violet color, was best in terms of efficiency (sampling capacity) and representativeness. Its wide-spectrum selectivity, given by the combination of polarities and extraction mechanisms (*i.e.,* adsorption and sorption), well covers the

EVOO volatilome dimensionality. In particular, DVB/CAR/PDMS is very effective toward alcohols and carbonyls, including most of key-aroma compounds resulting from the LOXs pathway and hydroperoxides cleavage. On the other hand, PDMS/DVB fiber (red color in **Figure 5.5.2**) showed good performances especially for lactones (*e.g.*, butyrolactone, δ-valerolactone, and ɣ-nonalactone), connoted by *creamy*, *caramel*, *waxy*, and *fresh* notes. Then PDMS fiber (green color in **Figure 5.5.2**) had good affinity - higher than 60 - only for saturated aldehydes, in particular C12-C16, and long-chain alcohols (*e.g.*, 6-methyl-5-hepten-2-one, 1-tridecanol, 1-tetradecanol), while PA fiber (blue color in **Figure 5.5.2**) was the worst in terms of extraction efficiency.

The differential attitudes of tested sampling fibers can be shown with visual features fingerprinting.[7] This strategy is a pixel based comparative visualization between image pairs where the information related to targeted analytes is preserved and easily retrieved during result inspection. GC×GC raw data are represented as an *a [m, n]* matrix, where *a* is the *analyzed* chromatogram with indexed pixels by ^{1}D retention time, *m*, and ^{2}D retention time, *n*. Each pixel corresponds to the detector data (*i.e.*, a fragmentation spectrum) while its colorization reveals differential detector response.[45,46] **Figure 5.5.3** shows, through a *fuzzy difference* colorization, the relative response differences between a *Baza-2-A* oil sampled by DVB/CAR/PDMS (*reference* image) and the resulting pattern for the same sample with PDMS fiber (*analyzed* image). Green coloured pixels/regions indicate a relative intensity higher for DVB/CAR/PDMS while those represented in red indicate a relative intensity higher for PDMS. White-grey coloured pixels correspond to analytes with the same relative intensity in the two chromatograms. Red bands, enclosed in white rectangles, at higher ^{1}D elution time correspond to ^{2}D column bleeding.

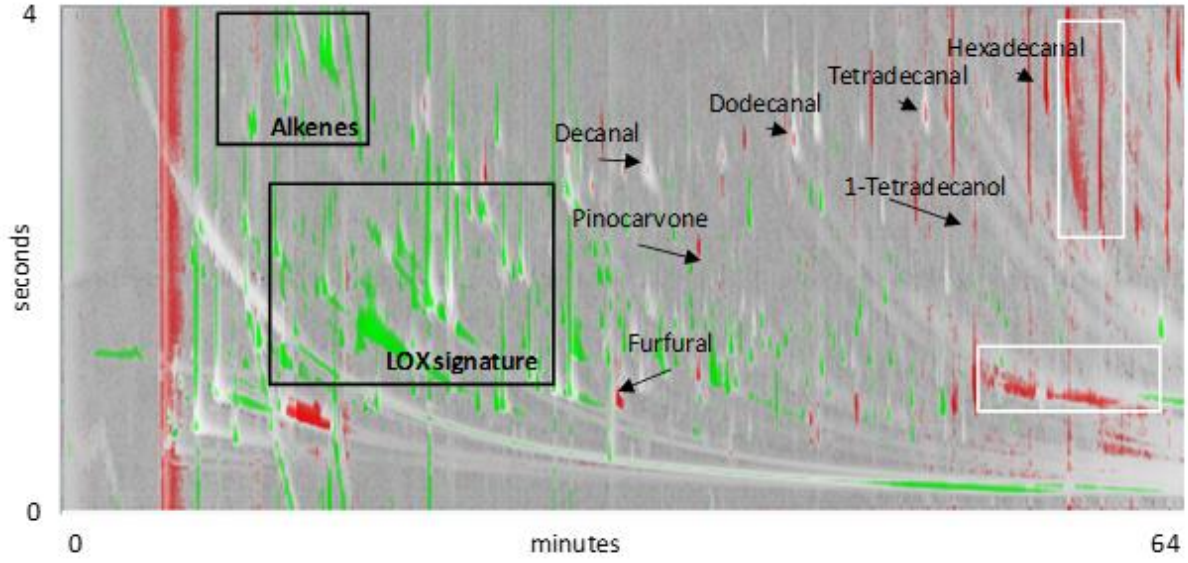

Figure 5.5.3. Visual features fingerprinting represented by colorized fuzzy ratio rendering. The analyzed chromatogram Baza-2-A oil is sampled by PDMS fiber while the reference chromatogram is by DVB/CAR/PDMS fiber. Green coloured pixels correspond to detector relative response values higher for DVB/CAR/PDMS while those in red indicate relative intensities higher for the PDMS. White-grey pixels correspond to response values with the same relative intensity in the two chromatograms. White squares highlight 2D column bleeding.

The predominance of green colour indicates an overall higher extraction performance for the DVB/CAR/PDMS fiber, in particular for some informative regions such as LOX signatures and alkenes, confirming data provided by **Figure 5.5.2** and **Table 5.5.2**. However, a higher abundance provided by PDMS extraction is shown by red spots corresponding to compounds having a polar function and, except for furfural, a relatively high MW (range 150-220 *amu*): pinocarvone, 1-

tetradecanol, dodecanal, tetradecanal, and hexadecanal. These compounds, as long chain aldehydes and alcohols, characterize sensory defects associated to oxidation during olive oil storage.[22]

5.5.4.3 Complementary information provided by tandem ionization TOF MS

The possibility to vary electron ionization energy during the analytical run by multiplexing between two values is a peculiar characteristic of the analytical platform adopted in this study. The complementary nature of tandem signals, acquired at 70 and 12 eV, in terms of spectral signatures differences and relative sensitivity, have been demonstrated in other application fields.[31–33,47] Here, possible benefits provided by this additional analytical dimension are considered for the validation of analytical indices of ripening and EVOO quality.

The next paragraphs show experimental results and discuss tandem signals absolute and relative sensitivity toward a selection of informative targeted compounds and their information potential in terms of fragmentation pattern similarity/difference.

a) Tandem signals complementarity: absolute and relative sensitivity.

To qualify soft ionization signals as an additional dimension of the system, absolute and relative sensitivity were considered and compared to standard, 70 eV signals. The number of detected peaks over a fixed threshold was calculated as an indicator of absolute sensitivity. **Table 5.5.3** indicates, for all samples considered in the study, the number of detected 2D peaks with signal-to-noise ratio (SNR) of 100 or more. This value (*e.g.*, 100 SNR) resulted from a previous study aimed at defining thresholds above which reliable spectra for cross-comparative analysis data can be extracted.[14]

Table 5.5.3. Number of detected 2D peaks above a signal-to-noise (SNR) of 100 from 70 eV and 12 eV ionization channels. SNR range indicates the minimum (*e.g.*, fixed threshold) and maximum value recorded from analyses.

	n° 2D peaks SNR > 100	SNR range (threshold – max)
Baza-1-A/B		
70 eV	332	100 - 6120
12 eV	187	100 - 12150
Baza-2-A/B		
70 eV	211	100 – 3950
12 eV	163	100 - 10440
Baza-3-A/B		
70 eV	272	100 - 6550
12 eV	178	100 - 13270
Baza-4-A/B		
70 eV	288	100 - 5032
12 eV	145	100 - 11420

On average, the number of detected peaks at 12 eV only reaches the 62% (RSD = 18.8%) of those detected at 70 eV, confirming that a higher ionization energy fragmentation generates a larger number of fragments and higher analyte responses. On the other hand, surprisingly, the SNR range is always greater at 12 eV (2 to 2.5-fold) indicating that at lower ionization energy the dynamic range

of responses is wider, with possible benefits for method dynamic range and linearity (see results at section *Validation of ripening indicators in headspace linearity conditions*).

SNRs were recorded based on 2D peak apexes for a selection of informative analytes to evaluate the relative sensitivity; the results are shown in **Table 5.5.4**. Values are from pre-processed chromatograms after back-ground noise subtraction. This operation was applied to all chromatograms with a well-established algorithm [48] implemented in the processing platform that calculates the average spectral noise across the 2D time space. Results in **Table 5.5.4** refer of SNRs at 12 and 70 eV (average of three analytical replicates and two biological replicates) for selected compounds detected in *Baza-2-A/B* samples.

For the most abundant volatiles in EVOO samples, *e.g.*, (*E*)-2-hexenal and (*Z*)-3-hexen-1-ol, at 12 eV SNRs are enhanced reaching 2.6 and 2.1 times the 70 eV value. Other analytes for which the 12 eV resulted in a higher relative sensitivity compared to 70 eV (with SNR ratios) are: ethyl acetate (2.8), 1-penten-3-ol (1.5), 1-penten-3-one (1.8), 3-pentanone (1.69), (*E*)-2-hexen-1-ol (2.2), (*E,E*)-α-farnesene (1.3), (5*E*)-3-ethyl-1,5-octadiene (1.8), (5*Z*)-3-ethyl-1,5-octadiene (1.6), and (*E,E*)-3,7-decadiene (1.5).

It has to be stressed that although 12 eV signal shows a higher SNR for some analytes, it does not necessarily turn out in a higher method sensitivity on this detection channel. The absolute response, that indicates the actual number of fragments generated in the ion source, was indeed lower at 12 eV compared to that at 70 eV as it is also confirmed by the number of detected 2D peaks above SNR 100 and reported in **Table 5.5.3**.

Table 5.5.4. Signal-to-noise (SNR) values estimated for a selection of informative analytes in the Baza-2-A/B samples (mean values); direct match factor (DMF) and reverse match factor (RMF) values of the same analytes calculated by the NIST Identity Search algorithm between spectra collected at 12 eV vs. NIST database entries, 70 eV vs. 12 eV, and 12 eV vs. NIST. Data on base peak (BP) m/z values at 70 and 12 eV also are reported.

		Relative sensitivity		Spectral similarity – NIST identity algorithm							
		SNRs		70 eV *vs.* NIST database		70 eV *vs.* 12 eV		12 eV *vs.* NIST database		Base peak (*m/z*)	
	Compounds	70 eV	12 eV	DMF	RFM	DMF	RFM	DMF	RFM	70 eV	12 eV
LOX signature	Ethyl acetate	2464	6836	866	885	724	754	592	620	43	61
	1-Pentanol	265	105	865	886	781	796	698	715	42	70
	1-Penten-3-ol	673	983	870	888	797	798	688	698	57	57
	1-Penten-3-one	640	1170	893	908	763	784	636	661	55	55
	3-Pentanone	943	1551	934	938	775	777	600	621	57	86
	1-Hexanol	1683	1503	914	925	793	798	654	678	56	56
	(Z)-3-Hexenyl acetate	1375	1441	901	917	781	787	656	669	67	82
	Hexanal	985	642	885	895	750	762	602	618	41	56
	Hexyl acetate	631	233	924	929	773	792	607	625	43	56
	(E)-2-Hexenal	3950	10440	880	895	795	822	707	737	55	69
	(E)-2-Hexen-1-ol	3396	7558	871	890	785	801	706	715	57	82
	(Z)-3-Hexen-1-ol	3947	8460	930	953	805	829	662	692	41	82
	(E,E)-2,4-Hexadienal	483	512	873	876	786	781	676	691	81	81
Terpenes	Limonene	586	474	927	955	771	775	591	606	68	136
	Linalool	140	94	921	923	789	804	634	664	71	84
	(E,E)-α-Farnesene	1596	2013	973	978	798	809	607	636	93	123
Aldeh	Heptanal	43	<10	833	939	752	797	651	659	43	70
	Octanal	302	79	877	960	790	833	684	690	43	84

	(E)-2-Heptenal	99	38	917	927	784	815	649	679	43	82
	(E)-2-Octenal	66	13	884	893	781	767	642	649	41	57
	(E)-2-Nonenal	85	35	889	900	775	796	654	664	43	70
	(E)-2-Decenal	443	137	937	960	821	845	700	722	43	70
	Nonanal	1068	430	886	908	750	781	616	628	57	98
	Decanal	221	49	858	877	748	767	606	625	43	82
	Dodecanal	329	94	932	955	792	805	662	667	43	82
Alkenes	3,4-Diethyl-1,5-hexadiene (meso)	489	454	888	913	791	821	695	720	69	69
	3,4-Diethyl-1,5-hexadiene (RS/SR)	420	318	893	920	772	813	648	675	69	69
	(5E)-3-Ethyl-1,5-octadiene	1412	2482	908	929	776	791	644	660	69	109
	(5Z)-3-Ethyl-1,5-octadiene	1167	1876	909	929	791	787	648	656	69	109
	(E,E)-3,7-Decadiene	1516	2275	851	869	774	785	666	670	69	95
	(E,Z)-3,7-Decadiene	1022	1092	857	877	766	806	679	700	43	109
	Average value			**897**	**917**	**778**	**796**	**650**	**668**		
	Standard deviation			**29.9**	**29.0**	**19.4**	**21.3**	**35.1**	**34.9**		
	RSD%			**3.3**	**3.2**	**2.5**	**2.7**	**5.4**	**5.2**		

b) Tandem signals complementarity: fragmentation pattern similarity

Tandem signals also provide complementary information from spectral signatures. This aspect has direct impact on method specificity while opening to an additional criterion for identity confirmation and cross-validation of quantitative results. Specificity that is very useful in those applications where co-elution occur and diagnostic ions are not present on 70 eV MS signature.[49]

DMF and RMF were calculated by extracting the MS fragmentation pattern from 2D peak apexes of selected informative analytes to evaluate spectral similarity/dissimilarity. **Table 5.5.4** reports the results with spectral similarity of informative analytes calculated on experimental spectra recorded at 70 eV vs. NIST reference, 70 eV vs. 12 eV, and 12 eV vs. NIST reference. The first comparison, i.e., 70 eV vs. NIST was done to evaluate the actual spectral consistency at 70 eV during tandem ionization operations. This spectral quality represents the benchmark for accurate identification or analytes and/or targets identity confirmation against spectral databases.

On average, selected peaks showed an average DMF value of 897 (RSD = 3.3%) and RMF of 917 (RSD = 3.2%) when spectra are compared to those collected in a commercial database (i.e., NIST). This result supports the reliability of analyte identification with tandem ionization. On the other hand, data from 70 eV vs. 12 eV are of help to evaluate the dissimilarity of hard and soft ionization spectra, and therefore to support the use of this instrumental option for volatilome investigations. In this case, DMF was on average 778 (RSD = 2.5%) with a minimum of 724 for ethyl acetate and a maximum value of 821 for (E)-2-decenal. RMF was on average 796 (RSD = 2.7%) with a minimum of 754 for ethyl acetate and a maximum value of 845 for (E)-2-decenal. However, comparing 12 eV spectra vs. NIST database, the dissimilarity is even lower with average DMF of 650 (RSD = 5.4%) and RMF of 668 (RSD = 5.2%).

Figure 5.5.4 shows the elution region of informative alkenes [13] from an EVOO sample obtained from olives at the earliest ripening stage (Baza-1-A) and recorded at 70 eV (**Figure 5.5.4A**) and at 12 eV (**Figure 5.5.4B**). The fragmentation patterns, in head-to-tail layout, illustrate spectral differences for (5E)-3-ethyl-1,5-octadiene (**Figure 5.5.4C**) and (*E,E*)-3,7-decadiene (**Figure 5.5.4D**). Spectra at 70 eV are quite similar (DMF=916 and RMF=921) and their reliable identification needs retention index confirmation. However, at 12 eV, the two spectra show lower similarity (DMF=837 and RMF=845) and different base peaks (BPs): 109 *m/z* for (5E)-3-ethyl-1,5-octadiene and 95 m/z for (*E,E*)-3,7-decadiene. In addition, the molecular ion at 138 *m/z* is enhanced for both analytes at 12 eV compared to 70 eV.

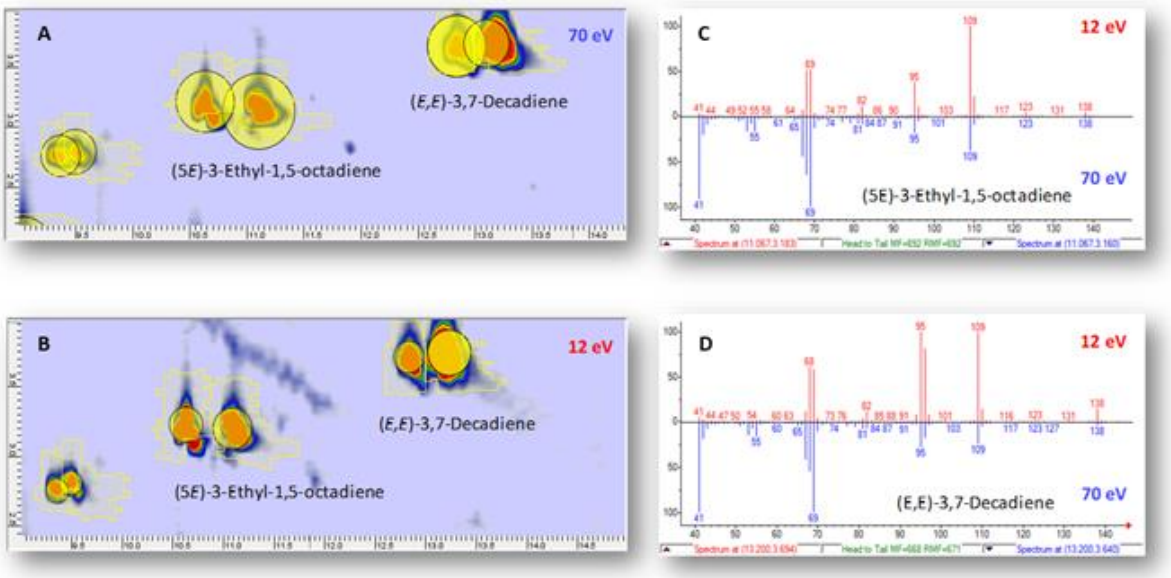

Figure 5.5.4. Contour plot corresponding to the retention region of informative alkenes acquired in tandem at 70 eV (**5.5.4A**) and 12 eV (**5.5.4B**). Head-to-tail spectra compare fragmentation patterns obtained at 70 and 12 eV for (5E)-3-ethyl-1,5-octadiene (**5.5.4C**) and (*E,E*)-3,7-decadiene (**5.5.4D**).5

The impact on spectral signatures and on BP distribution is detailed in **Table 5.5.**4. At 12 eV, BP values are generally higher than those at 70 eV, two exceptions are 1-penten-3-one and 3,4-diethyl-1,5-hexadiene. This characteristic is of great interest because it offers the possibility to add a confirmation point for reliable analyte identification and enables cross-validation of profiling data. This latter possibility is further discussed in the next section.

5.5.4.4 Validation of ripening indicators in headspace linearity conditions

The availability of samples deriving from olives harvested in different ripening stages and resulting in a different quality, as illustrated in **Table 5.5.1**, offered the opportunity (a) to investigate the evolution of oil volatilome, (b) to observe the presence of known markers[13,20] and their trends as a function of oil quality, and (c) to confirm their reliability and robustness when headspace linearity is verified.

Existing literature highlights the role of several compounds as ripening indicators. In particular, 3,4-diethyl-1,5-hexadiene (*RS* or *SR*), 3,4-diethyl-1,5-hexadiene (meso), *(5Z)*-3-ethyl-1,5-octadiene, *(5E)*-3-ethyl-1,5-octadiene, *(E,Z)*-3,7-decadiene, and *(E,E)*-3,7-decadiene are correlated to earlier stages of olive harvesting [24] independently from olive cultivar and geographical origin. On the other hand, both C6 aldehydes (*i.e.*, *(Z)*-2-hexenal, *(E)*-3-hexenal, and *(Z)*-3-hexenal), except for hexanal, and C5 aldehydes, alcohols, and ketones (*i.e.*, *(E)*-2-pentenal, *(Z)*-2-pentenal, 1-pentanol, 1-penten-3-ol, 3-pentanone, and 1-penten-3-one) show a decreasing trend during ripening[13,20] in *Picual*, *Chétoui* and *Arbequina* cultivars. These compounds belong to the LOX pathway, contributing to pleasant aroma and positively correlated with bitterness and pungency of olive oils; their decreasing concentrations contribute to the downgrading of the analyzed olive oil from EVOO (*Baza-1* and *Baza-2*) to VOO (*Baza-3*) and finally LO (*Baza-4*).

Conversely, other compounds have opposite trends, *i.e.*, they increase during ripening and over-ripening as well as during olive oil storage. They are *(E)*-2-heptenal, octanal, and nonanal that are formed from fatty acids hydroperoxides,[20] *(E,E)*-α-farnesene, 6-methyl-5-hepten-2-one, and octane

that characterize later stages of harvesting.[13,20] With the exception of 6-methyl-5-hepten-2-one, these analytes confirm their informative role within the sample set under study and by linear HS sampling.

Some ripening indexes of *Picual* cultivar defined by Magagna *et al.* [13] can be validated in linearity conditions. They include: *(Z)*-3-Hexenal/Nonanal, *(Z)*-3-Hexenal/Octane, *(E)*-2-Pentenal/Nonanal, and *(E)*-2-Pentenal/Octane. The histograms of **Figure 5.5.5** show the trend of ripening indexes as they result from tandem signals at 70 eV (blue bars) and at 12 eV (orange bars). Data are normalized in a 0-100 range to allow a proper comparison between the two ionization channels; trends were estimated by fitting with linear, exponential, or polynomial functions to delineate indexes evolution along harvest stage; and the accuracy of fittings is assessed by the determination coefficient (R^2). Interestingly, values of *(Z)*-3-Hexenal/Nonanal and *(Z)*-3-Hexenal/Octane ratios are better fit with different types of trend-lines: linear and polynomial in the first case, linear and exponential in the second case; for *(E)*-2-Pentenal/Nonanal and *(E)*-2-Pentenal/Octane, both 70 eV and 12 eV data fit better to polynomial trend-lines. In all cases, the determination coefficient is higher for trend-line generated on 12 eV data. The different trend lines are an additional proof of the complementary nature of tandem signals and of their actual differences in dynamic and linearity range of responses.

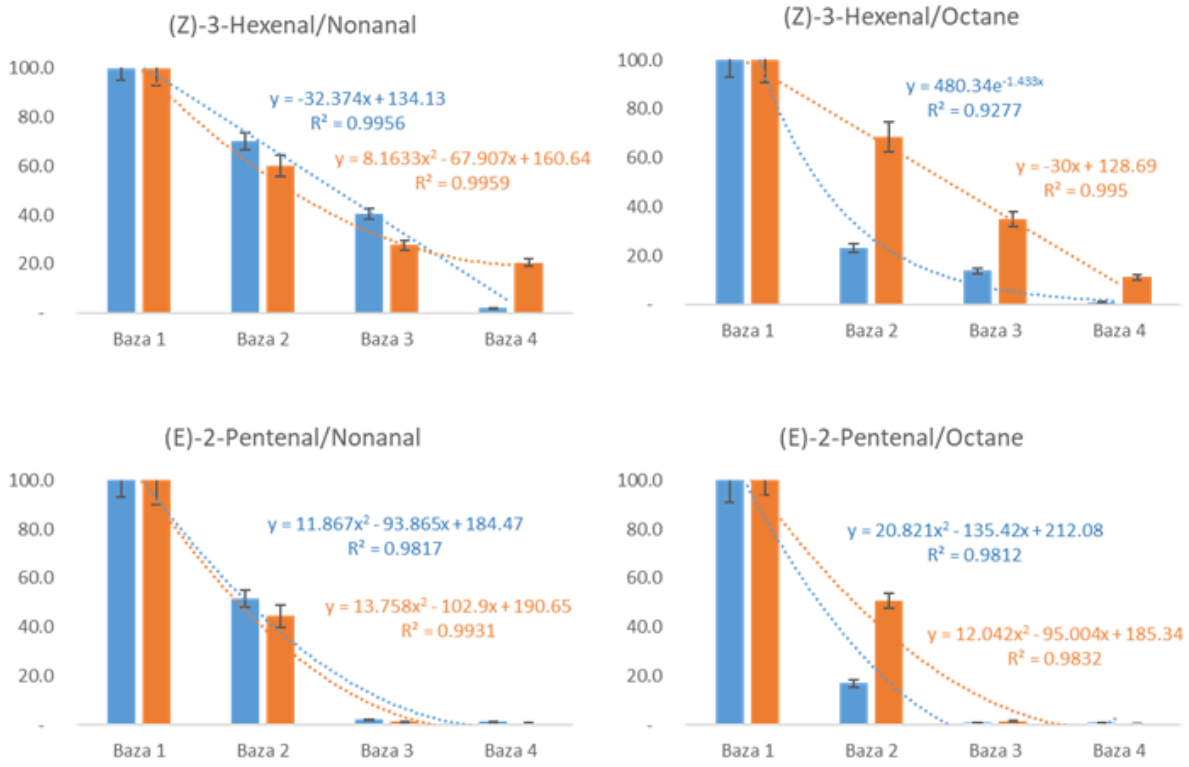

Figure 5.5.5. Histograms illustrating the response ratios between informative analytes (*e.g.*, ripening indexes)[13] calculated on averaged responses at 70 eV (blue bars) and 12 eV (orange bars) from oils obtained from olives at different ripening stages. Trend lines are accompanied by model functions and determination coefficients (R^2).

Eight new potential markers of ripening could be also highlighted. For some of them, linear HS conditions enable exploration of their actual dynamic range that is compressed in HS saturation. Trends are illustrated in histograms of **Figure 5.5.6.** Hachicha Hbaieb *et al.* [20] reported, for *Chétoui* and *Arbequina* olive cultivars, an increase for all C5 LOX-compounds and not significant alterations in C6 LOX-alcohols, but here thanks to the wider dynamic range due to HS linearity *(Z)*-3-hexen-

1-ol clearly decreased in oils produced with overripe olives. In particular, the decrease of *(Z)*-3-hexen-1-ol matches with the trend of other analytes whose formation is well correlated with the LOX pathway, *e.g.*, C5 and C6 aldehydes, alcohols, and ketones. Moreover, new informative analytes distinctive of earlier harvesting stages can be identified: methyl benzoate, 2-phenoxyethanol, and *(E)*-4,8-dimethyl-1,3,7-nonatriene.

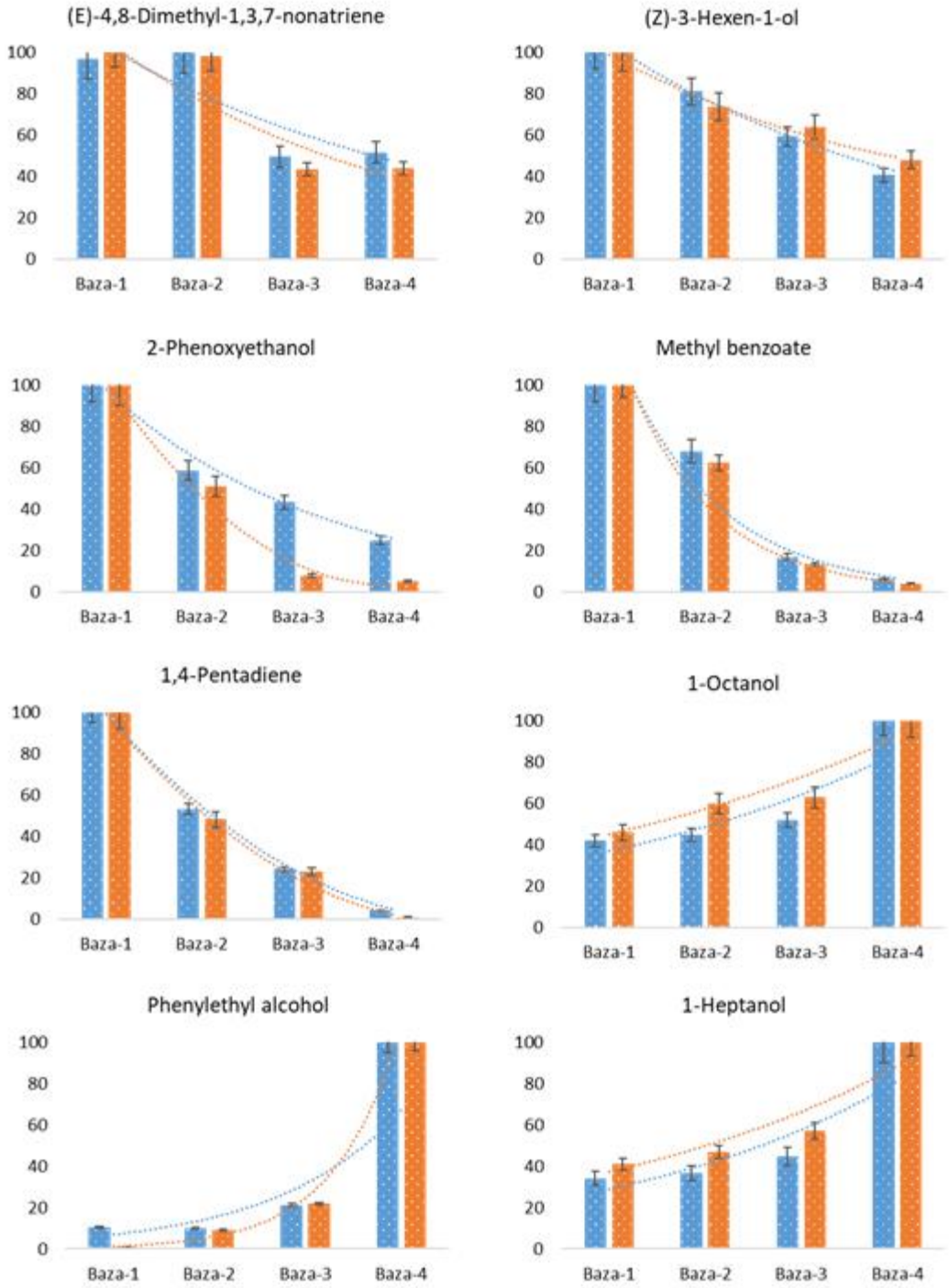

Figure 5.5.6. Histograms illustrating response trends for informative analytes (*e.g.*, ripening markers 20,21) calculated on averaged responses at 70 eV (blue bars) and 12 eV (orange bars) from oils obtained from olives at different ripening stages.

On the other hand, overripe olives yield oils with low sensory quality and higher relative amounts of phenylethyl alcohol, 1-heptanol, and 1-octanol. Note, 1-heptanol and 1-octanol are compounds characterized by low OTs (respectively 10 and 100 μg/kg) and correlated to musty, moss, and mushroom notes.

Very interestingly, HS linearity, although impacting on method absolute sensitivity, opens to new potential markers of *Picual* olives ripening and oils sensory quality extending the method informative potential while suggesting that additional dimensions at the MS detection level could be beneficial in terms of results cross-validation and robustness. The proposed approach could be effectively applied in an untargeted fashion (*i.e.*, combined untargeted and targeted UT fingerprinting) to exploit the full informative potential and to a larger set of samples to understand the general meaning of specific analytes over different cultivars and geographical areas.

5.5.5 Conclusions

The accurate fingerprinting of volatiles by HS-SPME operating in HS linearity conditions followed by GC×GC-TOF MS featuring tandem ionization gives the opportunity to improve the quality of analytical data and reliability of results. Highly resolved patterns from complex volatiles fraction can be effectively investigated since analytes identification is supported by: (a) stable spectral signatures at 70 eV that can be corroborated by complementary information provided by soft ionization at 12 eV (or lower energies) where diagnostic ions with higher m/z ratio are emphasized; (b) ordered retention patterns for chemically correlated compounds that inform about unknowns functionality; (c) high sensitivity provided by the in-space band compression of thermal modulation, a characteristic that is fundamental when HS linearity is matched at sampling level; and (d) wider dynamic and linear range of response provided by soft ionization that extends method capabilities when analytes span a wide concentration range.

For the complex olive oil volatilome, the proposed approach offers concrete advantages for the validation of the informative role of existing analytes while suggesting new potential markers to be studied in larger sample sets.

References

(1) Tranchida, P. Q.; Donato, P.; Cacciola, F.; Beccaria, M.; Dugo, P.; Mondello, L. Potential of Comprehensive Chromatography in Food Analysis. *TrAC Trends Anal. Chem.* **2013**, *52*, 186–205. https://doi.org/10.1016/j.trac.2013.07.008.

(2) Cordero, C.; Kiefl, J.; Reichenbach, S. E.; Bicchi, C. Characterization of Odorant Patterns by Comprehensive Two-Dimensional Gas Chromatography: A Challenge in Omic Studies. *TrAC Trends Anal. Chem.* **2019**, *113*, 364–378. https://doi.org/10.1016/j.trac.2018.06.005.

(3) Cordero, C.; Kiefl, J.; Schieberle, P.; Reichenbach, S. E.; Bicchi, C. Comprehensive Two-Dimensional Gas Chromatography and Food Sensory Properties: Potential and Challenges. *Anal. Bioanal. Chem.* **2015**, *407* (1), 169–191. https://doi.org/10.1007/s00216-014-8248-z.

(4) Ros, A. D.; Masuero, D.; Riccadonna, S.; Bubola, K. B.; Mulinacci, N.; Mattivi, F.; Lukić, I.; Vrhovsek, U. Complementary Untargeted and Targeted Metabolomics for Differentiation of Extra Virgin Olive Oils of Different Origin of Purchase Based on Volatile and Phenolic Composition and Sensory Quality. *Molecules* **2019**, *24* (16). https://doi.org/10.3390/molecules24162896.

(5) Purcaro, G.; Cordero, C.; Liberto, E.; Bicchi, C.; Conte, L. S. Toward a Definition of Blueprint of Virgin Olive Oil by Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2014**, *1334*, 101–111. https://doi.org/10.1016/j.chroma.2014.01.067.

(6) Vichi, S.; Pizzale, L.; Conte, L. S.; Buxaderas, S.; Lopez-Tamames, E. Solid Phase Microextraction in the Analysis of Virgin Olive Oil Volatile Fraction: Characterization of Virgin Oils from Two Distinct Geographical Areas of Northern Italy. *J. Agric. Food Chem.* **2003**, 6577.

(7) Reichenbach, S. E.; Tian, X.; Cordero, C.; Tao, Q. Features for Non-Targeted Cross-Sample Analysis with Comprehensive Two-Dimensional Chromatography. *J. Chromatogr. A* **2012**, *1226*, 140–148. https://doi.org/10.1016/j.chroma.2011.07.046.

(8) Reichenbach, S. E.; Tian, X.; Tao, Q.; Ledford, E. B.; Wu, Z.; Fiehn, O. Informatics for Cross-Sample Analysis with Comprehensive Two-Dimensional Gas Chromatography and High-Resolution Mass Spectrometry (GCxGC-HRMS). *Talanta* **2011**, *83* (4), 1279–1288. https://doi.org/10.1016/j.talanta.2010.09.057.

(9) Cordero, C.; Rubiolo, P.; Sgorbini, B.; Galli, M.; Bicchi, C. Comprehensive Two-Dimensional Gas Chromatography in the Analysis of Volatile Samples of Natural Origin: A Multidisciplinary Approach to Evaluate the Influence of Second Dimension Column Coated with Mixed Stationary Phases on System Orthogonality. *J. Chromatogr. A* **2006**, *1132* (1–2), 268–279. https://doi.org/10.1016/j.chroma.2006.07.067.

(10) Venkatramani, C. J.; Xu, J.; Phillips, J. B. Separation Orthogonality in Temperature-Programmed Comprehensive Two-Dimensional Gas Chromatography. *Anal. Chem.* **1996**, *68* (9), 1486–1492. https://doi.org/10.1021/ac951048b.

(11) Cordero, C.; Liberto, E.; Bicchi, C.; Rubiolo, P.; Schieberle, P.; Reichenbach, S. E.; Tao, Q.

Profiling Food Volatiles by Comprehensive Two-Dimensional Ga Schromatography Coupled with Mass Spectrometry: Advanced Fingerprinting Approaches for Comparative Analysis of the Volatile Fraction of Roasted Hazelnuts (Corylus Avellana L.) from Different Ori. *J. Chromatogr. A* **2010**, *1217* (37), 5848–5858. https://doi.org/10.1016/j.chroma.2010.07.006.

(12) Kiefl, J.; Schieberle, P. Evaluation of Process Parameters Governing the Aroma Generation in Three Hazelnut Cultivars (Corylus Avellana L.) by Correlating Quantitative Key Odorant Profiling with Sensory Evaluation. *J. Agric. Food Chem.* **2013**, *61* (22), 5236–5244. https://doi.org/10.1021/jf4008086.

(13) Magagna, F.; Valverde-Som, L.; Ruíz-Samblás, C.; Cuadros-Rodríguez, L.; Reichenbach, S. E.; Bicchi, C.; Cordero, C. Combined Untargeted and Targeted Fingerprinting with Comprehensive Two-Dimensional Chromatography for Volatiles and Ripening Indicators in Olive Oil. *Anal. Chim. Acta* **2016**, *936*, 245–258. https://doi.org/10.1016/j.aca.2016.07.005.

(14) Stilo, F.; Liberto, E.; Reichenbach, S. E.; Tao, Q.; Bicchi, C.; Cordero, C. Untargeted and Targeted Fingerprinting of Extra Virgin Olive Oil Volatiles by Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry: Challenges in Long-Term Studies. *J. Agric. Food Chem.* **2019**, *67* (18), 5289–5302. https://doi.org/10.1021/acs.jafc.9b01661.

(15) Stilo, F.; Cordero, C.; Sgorbini, B.; Bicchi, C.; Liberto, E. Highly Informative Fingerprinting of Extra-Virgin Olive Oil Volatiles: The Role of High Concentration-Capacity Sampling in Combination with Comprehensive Two-Dimensional Gas Chromatography. *Separations* **2019**, *6* (3), 34. https://doi.org/10.3390/separations6030034.

(16) Cordero, C.; Guglielmetti, A.; Sgorbini, B.; Bicchi, C.; Allegrucci, E.; Gobino, G.; Baroux, L.; Merle, P. Odorants Quantitation in High-Quality Cocoa by Multiple Headspace Solid Phase Micro-Extraction: Adoption of FID-Predicted Response Factors to Extend Method Capabilities and Information Potential. *Anal. Chim. Acta* **2019**, *1052*, 190–201. https://doi.org/10.1016/j.aca.2018.11.043.

(17) Sgorbini, B.; Cagliero, C.; Liberto, E.; Rubiolo, P.; Bicchi, C.; Cordero, C. Strategies for Accurate Quantitation of Volatiles from Foods and Plant-Origin Materials: A Challenging Task. *J. Agric. Food Chem.* **2019**, acs.jafc.8b06601. https://doi.org/10.1021/acs.jafc.8b06601.

(18) Kolb, B.; Ettre, L. S. *Static Headspace – Gas Chromatography : Theory and Practice Second Edition*, Second.; Wiley-VCH: New York, 2006.

(19) Giddings, J. C. Sample Dimensionality: A Predictor of Order-Disorder in Component Peak Distribution in Multidimensional Separation. *J. Chromatogr. A* **1995**, *703* (1–2), 3–15. https://doi.org/10.1016/0021-9673(95)00249-M.

(20) Hachicha Hbaieb, R.; Kotti, F.; Gargouri, M.; Msallem, M.; Vichi, S. Ripening and Storage Conditions of Chétoui and Arbequina Olives: Part I. Effect on Olive Oils Volatiles Profile. *Food Chem.* **2016**, *203*, 548–558. https://doi.org/10.1016/j.foodchem.2016.01.089.

(21) Morales, M. T.; Luna, G.; Aparicio, R. Comparative Study of Virgin Olive Oil Sensory Defects. *Food Chem.* **2005**, *91* (2), 293–301. https://doi.org/10.1016/j.foodchem.2004.06.011.

(22) Kalua, C. M.; Allen, M. S.; Bedgood, D. R.; Bishop, A. G.; Prenzler, P. D.; Robards, K. Olive Oil Volatile Compounds, Flavour Development and Quality: A Critical Review. *Food Chem.* **2007**, *100* (1), 273–286. https://doi.org/10.1016/j.foodchem.2005.09.059.

(23) Angerosa, F.; Servili, M.; Selvaggini, R.; Taticchi, A.; Esposto, S.; Montedoro, G. Volatile Compounds in Virgin Olive Oil: Occurrence and Their Relationship with the Quality. *J. Chromatogr. A* **2004**, *1054* (1–2), 17–31. https://doi.org/10.1016/j.chroma.2004.07.093.

(24) Angerosa, F.; Camera, L.; D'Alessandro, N.; Mellerio, G. Characterization of Seven New Hydrocarbon Compounds Present in the Aroma of Virgin Olive Oils. *J. Agric. Food Chem.* **1998**, *46* (2), 648–653. https://doi.org/10.1021/jf970352y.

(25) Wang, Y.; O'Reilly, J.; Chen, Y.; Pawliszyn, J. Equilibrium In-Fibre Standardisation Technique for Solid-Phase Microextraction. *J. Chromatogr. A* **2005**, *1072* (1), 13–17. https://doi.org/10.1016/j.chroma.2004.12.084.

(26) Setkova, L.; Risticevic, S.; Linton, C. M.; Ouyang, G.; Bragg, L. M.; Pawliszyn, J. Solid-Phase Microextraction-Gas Chromatography-Time-of-Flight Mass Spectrometry Utilized for the Evaluation of the New-Generation Super Elastic Fiber Assemblies. *Anal. Chim. Acta* **2007**, *581* (2), 221–231. https://doi.org/10.1016/j.aca.2006.08.022.

(27) Vaz-Freire, L. T.; da Silva, M. D. R. G.; Freitas, A. M. C. Comprehensive Two-Dimensional Gas Chromatography for Fingerprint Pattern Recognition in Olive Oils Produced by Two Different Techniques in Portuguese Olive Varieties Galega Vulgar, Cobrançosa e Carrasquenha. *Anal. Chim. Acta* **2009**, *633* (2), 263–270. https://doi.org/10.1016/j.aca.2008.11.057.

(28) Cajka, T.; Riddellova, K.; Klimankova, E.; Cerna, M.; Pudil, F.; Hajslova, J. Traceability of Olive Oil Based on Volatiles Pattern and Multivariate Analysis. *Food Chem.* **2010**, *121* (1), 282–289. https://doi.org/10.1016/j.foodchem.2009.12.011.

(29) Lukić, I.; Carlin, S.; Horvat, I.; Vrhovsek, U. Combined Targeted and Untargeted Profiling of Volatile Aroma Compounds with Comprehensive Two-Dimensional Gas Chromatography for Differentiation of Virgin Olive Oils According to Variety and Geographical Origin. *Food Chem.* **2019**, *270* (March 2018), 403–414. https://doi.org/10.1016/j.foodchem.2018.07.133.

(30) Oliver-Pozo, C.; Trypidis, D.; Aparicio, R.; Garciá-González, D. L.; Aparicio-Ruiz, R. Implementing Dynamic Headspace with SPME Sampling of Virgin Olive Oil Volatiles: Optimization, Quality Analytical Study, and Performance Testing. *J. Agric. Food Chem.* **2019**, *67* (7), 2086–2097. https://doi.org/10.1021/acs.jafc.9b00477.

(31) Cordero, C.; Guglielmetti, A.; Bicchi, C.; Liberto, E.; Baroux, L.; Merle, P.; Tao, Q.; Reichenbach, S. E. Comprehensive Two-Dimensional Gas Chromatography Coupled with Time of Flight Mass Spectrometry Featuring Tandem Ionization: Challenges and

Opportunities for Accurate Fingerprinting Studies. *J. Chromatogr. A* **2019**, *1597*, 132–141. https://doi.org/10.1016/j.chroma.2019.03.025.

(32) Rosso, M. C.; Mazzucotelli, M.; Bicchi, C.; Charron, M.; Manini, F.; Menta, R.; Fontana, M.; Reichenbach, S. E.; Cordero, C. Adding Extra-Dimensions to Hazelnuts Primary Metabolome Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography Combined with Time-of-Flight Mass Spectrometry Featuring Tandem Ionization: Insights on the Aroma Potential. *J. Chromatogr. A* **2020**, *1614*. https://doi.org/10.1016/j.chroma.2019.460739.

(33) Dubois, L. M.; Perrault, K. A.; Stefanuto, P. H.; Koschinski, S.; Edwards, M.; McGregor, L.; Focant, J. F. Thermal Desorption Comprehensive Two-Dimensional Gas Chromatography Coupled to Variable-Energy Electron Ionization Time-of-Flight Mass Spectrometry for Monitoring Subtle Changes in Volatile Organic Compound Profiles of Human Blood. *J. Chromatogr. A* **2017**, *1501*, 117–127. https://doi.org/10.1016/j.chroma.2017.04.026.

(34) Markes International. Select-EV: The next Generation of Ion Source Technology. *Technical Note*. 2016.

(35) NIST/EPA/NIH Mass Spectral Library with Search Program Data Version: NIST V17.

(36) Stilo, F.; Liberto, E.; Bicchi, C.; Reichenbach, S. E.; Cordero, C. GC×GC–TOF-MS and Comprehensive Fingerprinting of Volatiles in Food: Capturing the Signature of Quality. *LCGC Eur.* **2019**, *32* (5), 234–242.

(37) Cordero, C.; Schmarr, H.-G.; Reichenbach, S. E.; Bicchi, C. Current Developments in Analyzing Food Volatiles by Multidimensional Gas Chromatographic Techniques. *J. Agric. Food Chem.* **2018**, *66* (10), 2226–2236. https://doi.org/10.1021/acs.jafc.6b04997.

(38) Risticevic, S.; Vuckovic, D.; Lord, H. L.; Pawliszyn, J. 2.21 – Solid-Phase Microextraction. In *Comprehensive Sampling and Sample Preparation*; 2012; pp 419–460. https://doi.org/10.1016/B978-0-12-381373-2.00055-7.

(39) Lord, H. L.; Pfannkoch, E. A. Sample Preparation Automation for GC Injection. In *Comprehensive Sampling and Sample Preparation*; Elsevier, 2012; Vol. 2, pp 597–612. https://doi.org/10.1016/B978-0-12-381373-2.00061-2.

(40) Ross, C. F. *2.02 – Headspace Analysis*; Elsevier, 2012; Vol. 2. https://doi.org/10.1016/B978-0-12-381373-2.10036-5.

(41) Bicchi, C.; Cordero, C.; Liberto, E.; Sgorbini, B.; Rubiolo, P. Headspace Sampling of the Volatile Fraction of Vegetable Matrices. *J. Chromatogr. A* **2008**, *1184* (1–2), 220–233. https://doi.org/10.1016/j.chroma.2007.06.019.

(42) Cavalli, J. F.; Fernandez, X.; Lizzani-Cuvelier, L.; Loiseau, A. M. Comparison of Static Headspace, Headspace Solid Phase Microextraction, Headspace Sorptive Extraction, and Direct Thermal Desorption Techniques on Chemical Composition of French Olive Oils. *J. Agric. Food Chem.* **2003**, *51* (26), 7709–7716. https://doi.org/10.1021/jf034834n.

(43) Liberto, E.; Bicchi, C.; Cagliero, C.; Cordero, C.; Rubiolo, P.; Sgorbini, B. Chapter 1: Headspace Sampling: An “Evergreen” Method in Constant Evolution to Characterize Food Flavors through Their Volatile Fraction. *Food Chemistry, Function and Analysis*.

2020, pp 3–37. https://doi.org/10.1039/9781788015752-00001.

(44) Bicchi, C.; Cordero, C.; Rubiolo, P. A Survey on High-Concentration-Capability Headspace Sampling Techniques in the Analysis of Flavors and Fragrances. *J. Chromatogr. Sci.* **2004**, *42* (8), 402–409. https://doi.org/10.1093/chromsci/42.8.402.

(45) Cordero, C.; Bicchi, C.; Rubiolo, P. Group-Type and Fingerprint Analysis of Roasted Food Matrices (Coffee and Hazelnut Samples) by Comprehensive Two-Dimensional Gas Chromatography. *J. Agric. Food Chem.* **2008**, *56* (17), 7655–7666. https://doi.org/10.1021/jf801001z.

(46) Hollingsworth, B. V.; Reichenbach, S. E.; Tao, Q.; Visvanathan, A. Comparative Visualization for Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2006**, *1105* (1–2), 51–58. https://doi.org/10.1016/j.chroma.2005.11.074.

(47) Freye, C. E.; Moore, N. R.; Synovec, R. E. Enhancing the Chemical Selectivity in Discovery-Based Analysis with Tandem Ionization Time-of-Flight Mass Spectrometry Detection for Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2018**, *1537*, 99–108. https://doi.org/10.1016/j.chroma.2018.01.008.

(48) Reichenbach, S. E.; Ni, M.; Zhang, D.; Ledford, E. B. Image Background Removal in Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2003**, *985* (1–2), 47–56. https://doi.org/10.1016/S0021-9673(02)01498-X.

(49) Cordero, C.; Gabetti, E.; Bicchi, C.; Merle, P.; Belhassen, E. GC×GC with Parallel Detection by FID and TOF MS Featuring Tandem Ionization: Extra-Dimensions for Great Flexibility in Fragrance Allergens Profiling. In *43rd International Symposium on Capillary Chromatography & the 16th GCxGC Symposium.*; Armstrong, D. W., Schug, K. A., Eds.; Fort Worth, Texas, USA, 2019; p 1.

5.6 Chromatographic fingerprinting enables effective discrimination and *identitation* of high-quality Italian extra-virgin olive oils

Federico Stilo[1], Ana M. Jiménez-Carvelo[2,*], Erica Liberto[1], Carlo Bicchi[1], Stephen E. Reichenbach[3,4], Luis Cuadros-Rodríguez[2], and Chiara Cordero[1,*]

[1] Dipartimento di Scienza e Tecnologia del Farmaco, Università degli Studi di Torino, Turin, Italy

[2] Department of Analytical Chemistry, Faculty of Science, University of Granada, Av. Fuentenueva S/N, E-18071, Granada, Spain

[3] University of Nebraska-Lincoln, NE USA

[4] GC Image LLC, Lincoln, NE USA

* Corresponding authors:

Prof. Chiara Cordero – Dipartimento di Scienza e Tecnologia del Farmaco, Università di Torino; Via Pietro Giuria 9, I-10125 Torino, Italy e-mail: chiara.cordero@unito.it; phone: +39 011 6707172

Dr. Ana M. Jiménez Carvelo - Department of Analytical Chemistry, Faculty of Sciences, University of Granada, Av. Fuentenueva S/N, E-18071, Granada, Spain e-mail: amariajc@ugr.es; phone: +34 958240797

Submitted to Journal of Agricultural and Food Chemistry: May 21, 2021

5.6.1 Abstract

The challenging process of high-quality food authentication takes advantages by highly informative chromatographic fingerprinting and its *identitation* potential. In this study, the unique chemical traits of the complex volatile fraction of extra-virgin olive oils from Italian production, are captured by comprehensive two-dimensional gas chromatography coupled to time-of-flight mass spectrometry and explored by pattern recognition algorithms. The consistent re-alignment of untargeted and targeted features over 73 samples, including oils obtained by different olives cultivar (n=24), harvest years (n=3) and processing technologies, provides solid foundation for samples identification and discrimination based on production Region (n=6).

Through a dedicated multivariate statistics workflow, *identitation* is achieved by two-level PLS regression, that highlights Region diagnostic patterns accounting between 58 and 82 of untargeted and targeted compounds, while samples classification is by sequential application of SIMCA models, one for each production Region. Samples are correctly classified in five of the six single-class models and quality parameters (*i.e.*, sensitivity, specificity, precision, efficiency, and area under the receiver operating characteristic curve (AUC)) are equal to 1.00.

Key words

Comprehensive two-dimensional gas chromatography; extra virgin olive oil; combined untargeted and targeted (UT) fingerprinting; *identitation* and authentication; geographical origin.

5.6.2 Introduction

Olive oil (OO) is one of the pillars of the Mediterranean diet and represents the main source of fats in the countries of the Mediterranean basin.[1] In particular, *extra-virgin* olive oil (EVOO) is recognized as the most valuable product among the edible oils,[2] it is extracted from fresh olive fruits (*Olea europeae L.*) by mechanical or physical technologies that preserve the composition of the lipid fraction, while limiting autoxidation reactions and alterations of its native quality.[3]

The reason of the increasing demand for olive oil of high-quality, *i.e.*, EVOO (Commission of the European Communities, 1991; "EU Food Qual. Labels," 2021; IOC, 2015), relates not only on its nutritional and healthy values, due to the presence of antioxidants (*i.e.*, tocopherols and phenolic compounds) and high oleic acid content, but also on its peculiar sensory characteristics [2] strongly related to olives cultivar, pedoclimatic conditions of the harvest region, olives ripeness and extraction technology ("EU Food Qual. Labels," 2021).

In this context, any analytical methodology capable of delineating chemical patterns informative of the different functional variables influencing the composition of EVOO is useful and has the potential to support the valorization of high-quality products, facilitate sensory quality evaluation/screenings at the basis of commercial classification as well as to counteract fraudulent practices.[7] In this latter context, the accurate fingerprinting of the unsaponifiable fraction and of minor components by comprehensive two-dimensional gas chromatography coupled to mass spectrometry (GC×GC-MS) and/or with parallel flame ionization detection (MS/FID), was successful in identifying admixtures of OO with other fats and/or establish the product freshness/shelf-life.[8]

Active research in the development of GC fingerprinting methodologies includes also investigations on EVOO volatiles. The mono-dimensional (1D)-GC fingerprinting accompanied by accurate profiling were recently applied to validate the role of sesquiterpene hydrocarbons as geographical origin markers[7] in EVOOs from different cultivars and production areas. In their study, Quintanilla Casas *et al.*[7] confirmed the superior discrimination power of the total volatiles fingerprint (100% correct classification) obtained by GC-MS raw data processing followed by suitable supervised chemometrics, compared to the targeted profiling of selected sesquiterpenoids which correctness ranged between 46 to 100% of correct classification, as a function of the production country.

Moreover, by the volatiles fingerprinting based on 1D-GC, it was also possible to support the commercial classification of OO based on sensory panel evaluation.[9] EVOOs are in fact characterized by peculiar yet essential aroma qualities such as *green*, *grassy* and *fruity* notes, whose perception is at the basis of commercial classification based on EU regulations. Sensory quality in fact, together with compositional/chemical standards to be complied,[5] and absence of off-flavors, guide the OOs classification in EVOO (median of the positive attributes ≥ 0), *virgin* olive oil (VOO) and *lampante* olive oil (LOO) in presence of sensory defects (*rancid*, *fusty/muddy*, *musty*, and *winey/vinegary)*, even at lower levels (*e.g.*, median of the defects $\neq 0$).[5]

In this study we make a step forward in the direction of validating a powerful and highly-flexible chromatographic fingerprinting workflow with superior *identitation*[10] and classification effectiveness compared to existing tools. The improved separation capacity of GC×GC, the analytes retention logic over the separation space, and the comprehensive capture of component's features generated by TOF MS detection, make the resulting 2D fingerprints as sample's unique traits for effective and reliable authentication. Moreover, the specificity of the third information dimension of the system (*i.e.*, EI-MS fragmentation patterns), gives access to a higher informative level as any confirmatory analytical technique.

Compared to existing studies adopting GC×GC as profiling and/or fingerprinting technique,[11–13] the combined information deriving from untargeted and targeted features is here explored in the challenging scenario of Italian high-quality EVOOs production connoted by an impressive heritage of olives genetic varieties, with about 540 different registered cultivars,[14] and 46 Protected Designation of Origin (PDO) products from different geographical locations *(i.e.*, Regions) over the entire territory.

The challenge posed by the complexity and high chemical dimensionality of EVOOs volatiles is tackled by a dedicated workflow, named combined untargeted and targeted fingerprinting *(UT fingerprinting)*,[15] where the information from known and unknown components patterns are accurately tracked across many samples and their *identitation*, discrimination, and classification power examined in great detail with a focus on Regional characters. Furthermore, the synergy between profiling and fingerprinting is also examined by observing the distribution of key-aroma compounds and potent odorants strongly correlated to positive and/or negative odor qualities.[16–18]

5.6.3 Materials and methods

5.6.3.1 Reference compounds and solvent

Pure reference standards of α- and β-thujone and methyl-2-octynoate used as internal standards (ISs), *n*-alkanes (from *n*-C7 to *n*-C25) used for linear retention index (I^T) calibration, and pure reference compounds for identity confirmation were supplied by Merck (Milan, Italy). Cyclohexane (HPLC grade) for *n*-alkanes dilution and pure dibutyl phthalate used to prepare ISs working solutions were also from Merck (Milan, Italy).

5.6.3.2 Extra virgin olive oil samples

Extra virgin olive oils (EVOOs) were supplied within the VIOLIN project[19] selection. They were obtained by olives of different cultivars harvested between 2016-2018 over the Italian territory; samples were all certified as EVOOs by accredited laboratories (ISO 17025:2018) and by the official sensory panel test. Details on the sample-set under study, counting seventy-three samples, are provided in **Supplementary Table 5.6.1** together with harvest/production Regions (*i.e.*, Umbria n=7, Garda lake n=10, Lazio n=11, Puglia n=12, Sicilia n=13, and Toscana n=20). **Supplementary Figure 5.6.1**, shows geographical locations of selected EVOOs production sites.

5.6.3.3 Headspace solid phase microextraction devices and sampling conditions

Volatiles were sampled by headspace (HS) solid phase microextraction (SPME). A divinylbenzene/carboxen/polydimethyl siloxane (DVB/CAR/PDMS) d_f 50/30 μm 2 cm length fiber (Supelco, Bellefonte, PA, USA) was chosen based on its sampling effectiveness on EVOOs volatiles and previous research.[16,20–22] SPME fibers were conditioned before use as recommended by the manufacturer.

The ISs were preloaded onto the SPME device by sampling 5.0 μL of α/β-thujone and methyl 2-octynoate ISs solution (100 mg L^{-1}) placed in a 20 mL headspace vial. ISs pre-loading is by exposing the SPME device to the HS kept at 40 °C for 5 min.

Sampling was carried out on 0.100 ± 0.005 g of oil samples, precisely weighed in 20 mL headspace vials, at 40 °C for 60 min under constant stirring. The amount of sample was chosen matching for HS linearity conditions for most of the characteristic analytes of EVOOs volatile fraction.[20,23] After extraction, the SPME device was automatically transferred to the split/splitless injection port of the GC×GC system kept at 250 C and thermal desorption was for 5 min.

5.6.3.4 GC×GC-TOF MS: instrument set-up and conditions

GC×GC analyses were performed on an Agilent 7890B GC unit (Agilent Technologies, Wilmington DE, USA) coupled to a Markes BenchTOF-Select™ mass spectrometer featuring Tandem Ionization™ (Markes International, Llantrisant, UK). The GC transfer line was set at 270 °C. TOF MS tuning parameters were set for single ionization at 70 eV and the scan range was set at 40-350 m/z with a spectra acquisition frequency of 100 Hz. The system was equipped with a two-stage KT 2004 loop-type thermal modulator (Zoex Corporation, Houston, TX) cooled with liquid nitrogen and controlled by Optimode v2.0 (SRA Intruments, Cernusco sul Naviglio, Milan, Italy). Modulation period (P_M) and hot jet pulse times were set, respectively, at 3.5 s and 300 ms, with a cold jet stream at the mass flow controller (MFC) from 40% to 8% of the total flow along the run duration. No secondary oven was adopted in the GC×GC set-up.

5.6.3.5 GC×GC columns and settings

The column set was configured as follows: ^{1}D DB-HeavyWax™ column (100% polyethylene glycol; 30 m × 0.25 mm d_c × 0.25 μm d_f) from Agilent J&W (Wilmington, DE, USA) coupled with a ^{2}D OV1701 column (86% polydimethylsiloxane, 7% phenyl, 7% cyanopropyl; 1 m × 0.1 mm d_c × 0.10 μm d_f) from Agilent Technologies (Wilmington, DE, USA). A fused silica capillary loop (1.0 m × 0.1 mm d_c) was used in the modulator slit.

The GC split/splitless injector port was kept at 250 °C and operated in split mode with a split ratio 1:20. The carrier gas was helium at a constant nominal flow of 1.3 mL min^{-1}. The oven temperature programming was set as follows: from 40 °C (2 min) to 240 °C (10 min) at 3.5 °C min^{-1}.

The *n*-alkanes solution for I^T determination was analyzed under the following conditions: split/splitless injector in split mode, split ratio 1:50, injector temperature 250 °C, and injection volume 1 μL.

5.6.3.6 Combined Untargeted and Targeted (UT) fingerprinting workflow

The data processing workflow was designed to comprehensively capture the chemical signature of volatiles form EVOO samples by computing both peak and peak-region features from untargeted (unknowns) and targeted components located over the 2D space. The approach, named *UT* fingerprinting, was designed on EVOO volatile patterns and further adapted to compositional peculiarities of samples in other fields.[24] In this study, the targeting (*i.e.*, identification) of analytes

was done as last step of the process after chromatograms re-alignment over *reliable* peaks from untargeted components/features.

The generation of untargeted features (*i.e.*, peaks and peak-regions) and their re-alignment across all samples chromatograms is by *template matching* [25] and actively uses metadata, collected for 2D peaks and peak-regions (*i.e.*, retention times, MS spectrum, and detector response) above a signal-to-noise (S/N) threshold value of 100,[22] to establish correspondences across 2D patterns. Re-alignment specificity, is by active constraints on MS similarity [*i.e.*, threshold value 750 for direct match factor DMF and reverse match factor RMF according to NIST MS Search algorithm, ver. 2.0 (National Institute of Standards and Technology, Gaithersburg, MD, USA)] between template (*reference*) and candidate (*analyzed*) "peak spectra".[22,26,27]

The chromatographic fingerprinting was performed automatically by GC Image Investigator™ V2.9 (GC Image LLC, Lincoln NE, USA) on a random selection of samples chromatograms (n=25) acquired across a timeframe of two-weeks. It aligned the 25 chromatograms through *reliable* peaks for registration and generated a *composite* chromatogram over which peak-region features were delineated and extracted to form a *feature* template for further processing. Reliable peaks in this study were those that positively matched across *all-but-one* of the selected 25 chromatograms (*i.e.*, most constrained condition option).

The resulting feature template includes untargeted (*reliable*) peaks and peak-regions comprehensively capturing the chemical composition of samples. **Figure 5.6.1A** shows the pseudocolor image of a Sicilian EVOO (#S1) overlaid with 591 peak-regions (red graphics) and 159 targeted peaks (green circles). Targeting of informative compounds, including EVOO key-aroma compounds, ripening indicators and potent odorants responsible of coded defects,[28] was performed at the end of the re-alignment process over the entire set of chromatograms (n=73).

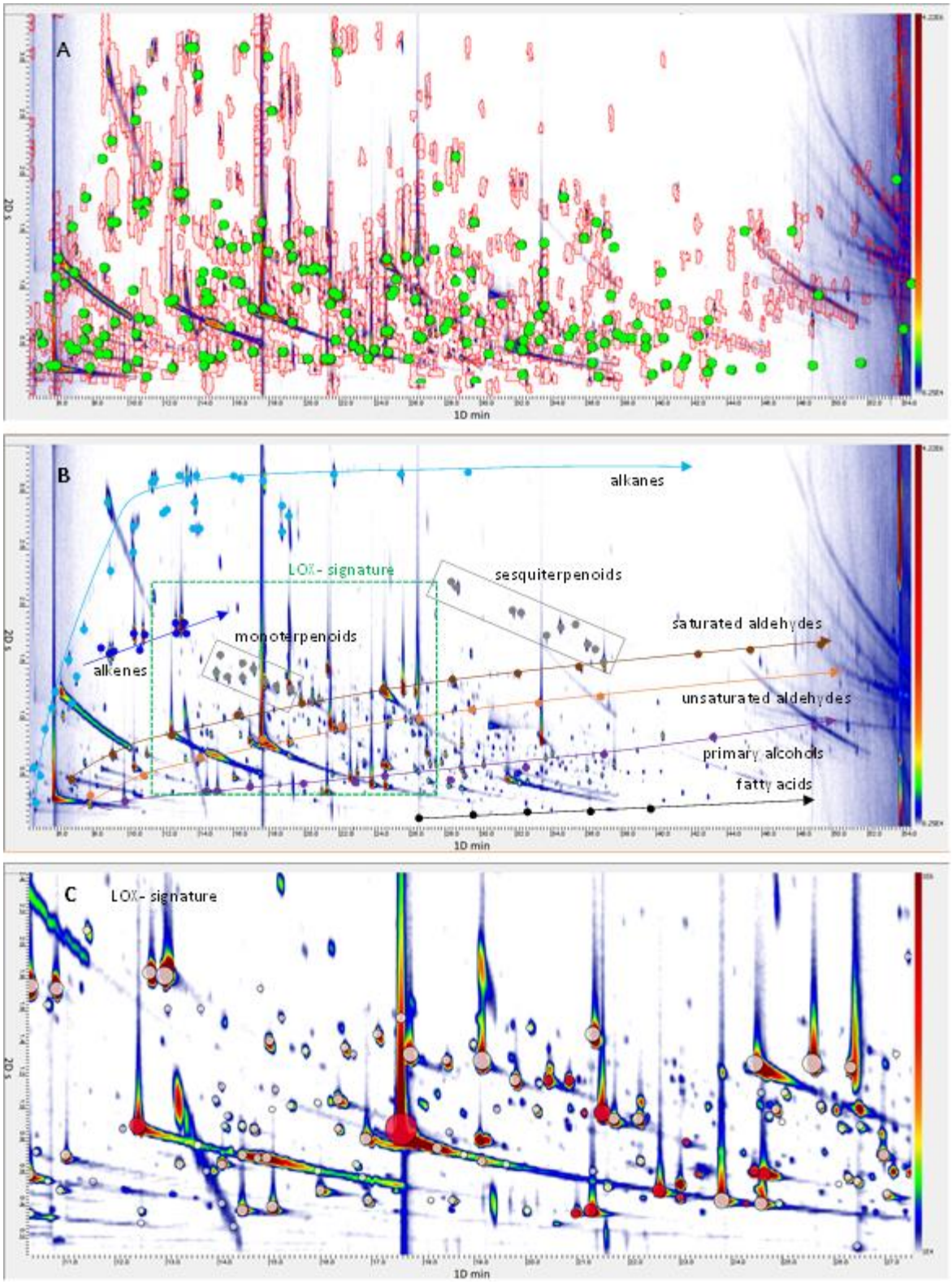

Figure 5.6.1. Pseudocolor image (**5.6.1A**) of a Sicilian EVOO (Sicilia origin – ID#S1) volatile fraction comprehensively mapped through untargeted and targeted (UT) peak-regions (red graphics); identified/targeted analytes (*i.e.,* targeted compounds) are highlighted by green circles. (**5.6.1B**) Patterns of analytes, following a retention logic based on the relative retention exerted by the polar × semipolar column combination adopted (alkanes - cyano, alkenes - blue, saturated - brown and unsaturated - orange aldehydes, alcohols - purple, terpenoids - grey and fatty acids - black) are highlighted. (**5.6.1C**) Shows the enlarged area of LOX derivatives (red circles).

Table 5.6.1. List of 159 target peaks, together with retention times in the two analytical dimensions (^1t_R, ^2t_R), % relative standard deviation (%RSD) calculated over six analytical replicates over two-weeks and referred to retention times and 2D peak volumes; experimental ^{1}D I^T and tabulated I^T values; identification criteria (a) reference compound confirmation or (b) spectral direct match similarity and $I^T \pm 20$. Odor descriptors as reported in reference literature.[16,18,34,35,42]

Chemical class	Compound Name	^1t_R (min)	^1t_R %RSD	^2t_R (s)	^2t_R %RSD	2D-Peak Volumes %RSD	Exp I^T	Lit I^T	Identification	Odor descriptors
Alcohols	2-Propanol	7.64	0.39	0.26	1.33	5.37	907	912	a	*Alcoholic, musty*
	2-Methyl-1-propanol	12.48	0.76	0.34	0.79	14.20	1064	1081	a	-
	1-Butanol	14.41	0.87	0.36	2.75	13.89	1116	1124	a	*Winey*
	1-Penten-3-ol	15.05	0.73	0.36	1.71	6.94	1133	1139	a	*Pungent, apple-like*
	3-Methyl-1-butanol	16.92	0.79	0.40	0.88	6.94	1180	1184	b	-
	1-Pentanol	18.08	0.75	0.48	2.91	20.83	1210	1216	a	*Sweet, pungent*
	(Z)-2-Penten-1-ol	20.94	0.50	0.34	0.92	15.22	1282	1289	a	*Green, plastic, rubber*
	(E)-2-Penten-1-ol	21.29	0.72	0.34	4.50	10.55	1291	1296	a	*Mushroom*
	(Z)-3-Hexen-1-ol	22.58	0.99	0.46	3.54	4.05	1324	1344	a	*Green, leafy*
	1-Hexanol	22.98	0.69	0.50	1.37	4.72	1335	1338	a	*Fruity, banana, soft*
	(Z)-2-Hexen-1-ol	23.74	0.74	0.42	3.78	6.08	1355	1375	a	*Leaf-like*
	(E)-2-Hexen-1-ol	24.21	0.00	0.70	3.85	4.66	1367	1379	a	*Green grass, leaves*
	1-Butoxy-2-propanol	24.56	0.29	0.40	3.60	2.77	1376	1363	b	-
	1-Heptanol	26.43	0.88	0.52	2.60	14.77	1425	1423	a	*Herbal*
	1-Octen-3-ol	26.72	0.91	0.56	0.64	2.47	1433	1437	a	*Mushroom, earthy*
	2-Ethyl-1-hexanol	27.77	0.20	0.60	2.12	3.32	1462	1470	a	-
	1-Octanol	29.75	0.23	0.66	0.86	13.60	1517	1518	a	*Moss, nut, mushroom*
	1-Nonanol	33.02	0.55	0.66	1.59	14.41	1610	1616	a	*Fresh, clean, floral*
	4-Butoxy-1-butanol	34.83	0.04	0.54	1.21	2.64	1664	1668	b	-
	1-Decanol	36.98	0.87	0.70	3.16	4.15	1730	1738	a	*Fatty, waxy*
	2-(2-butoxyethoxy)-Ethanol	37.92	0.94	0.52	0.93	19.81	1759	1364	b	-
	Benzyl alcohol	40.37	0.87	0.28	3.79	14.78	1839	1846	a	*Sweet, fruity*
	1,4-Butanediol	41.30	0.83	0.28	0.71	5.05	1867	1861	b	-
	Phenylethyl alcohol	41.48	0.66	0.34	0.10	2.45	1875	1877	a	*Sweet, floral, fresh*
	1-Dodecanol	42.88	0.08	0.68	0.77	9.55	1921	1924	a	*Earthy, soapy, fatty, waxy*
	Phenol	43.93	0.77	1.60	3.59	12.40	1956	1957	a	-
	1-Tetradecanol	47.56	0.48	0.65	2.55	11.68	2118	2137	b	*Musty*
	2-Phenoxyethanol	47.83	0.29	1.58	4.54	15.91	2130	2145	b	-

	1-Hexadecanol	54.25	0.34	1.06	3.01	4.02	2345	2356	b	*Waxy, clean*
Esters	Ethyl acetate	6.88	0.35	0.42	3.94	13.34	866	875	a	*Fruity, sweet, green*
	2,2-Dimethylpropanoate	7.70	0.58	0.72	2.18	16.23	909	913	a	-
	Butyl acetate	12.13	0.73	0.84	3.12	14.29	1055	1064	a	*Pear*
	Isoamyl acetate	13.94	0.91	1.02	0.85	18.09	1104	1109	a	*Sweet, fruity*
	2-Methylpropyl butanoate	14.82	0.01	1.08	1.39	7.43	1127	1139	b	*Sweet, fruity, pineapple*
	Butyl isobutyrate	14.99	0.28	1.36	1.11	15.72	1131	1140	b	*Sweet, fruity, apple*
	Butanoic butanoate	17.73	0.43	1.28	4.96	19.18	1201	1212	a	-
	3-Hydroxy-2-butanone	20.01	0.84	0.34	2.91	12.18	1259	1270	a	*Sweet, buttery, creamy*
	Hexyl acetate	20.42	0.81	1.14	3.31	19.31	1269	1275	a	*Fruity*
	(Z)-3-Hexenyl acetate	21.47	0.49	0.92	3.33	19.79	1296	1310	a	*Sweet*
	Butyl 2-ethylhexanoate	27.42	0.54	1.88	4.55	6.49	1452	1459	a	-
	Methyl benzoate	30.67	0.59	0.58	3.50	10.67	1546	1560	a	-
	Ethyl benzoate	32.55	0.40	0.72	4.92	3.38	1597	1612	a	*Sweet, fruity*
	Methyl salicylate	37.57	0.00	0.56	2.37	3.99	1748	1755	b	*Sweet, minty*
	Butyl benzoate	40.31	0.79	0.82	1.30	20.29	1837	1846	a	*Balsamic, amber*
Lactones	4-Hydroxy-2-hexenoic acid lactone	31.50	0.35	0.46	3.27	15.41	1567	/	b	*Fruity, minty*
	Butyrolactone	32.20	0.72	0.42	2.15	20.63	1587	1601	a	*Creamy, oily, caramel*
	β-Angelica lactone	34.24	0.15	0.42	1.45	15.98	1647	1664	b	-
	δ-Pentalactone	35.12	0.59	0.52	0.56	3.22	1672	1684	*a*	*Herbal, sweet*
	λ-Hexalactone	38.09	0.35	0.60	3.51	5.84	1765	/	a	*Tonka, creamy*
Fatty acids	Acetic acid	26.43	0.13	0.12	3.94	15.78	1425	1427	a	*Sour, vinegary*
	Propanoic acid	29.52	0.23	0.14	3.99	18.99	1510	1516	a	*Acidic, pungent*
	Butanoic acid	31.97	0.35	0.62	3.17	20.39	1580	1581	a	*Cheesy, sharp*
	Pentanoic acid	36.05	0.96	0.30	0.24	5.53	1700	1704	a	*Cheesy, acidic*
	Hexanoic acid	39.55	0.15	0.20	0.13	13.26	1812	1817	a	*Goat-like, sweaty*
	Heptanoic acid	42.64	0.23	0.26	4.87	14.40	1914	1920	a	-
	Octanoic acid	45.73	0.04	0.26	4.27	5.55	2053	2058	a	*Fatty, waxy, rancid, oily*
	Nonanoic acid	48.59	0.02	0.42	1.11	10.84	2181	2180	a	*Sweaty, waxy*
	Decanoic acid	51.63	0.54	0.30	4.35	16.35	2239	2240	a	*Soapy, rancid*
Hydrocarbons	*n*-Hexane	3.91	0.40	0.34	1.50	10.88	600	/	a	-
	Cyclopentane	4.20	0.95	0.22	2.12	3.19	631	/	b	*Petroleum*
	2,4-Dimethylhexane	4.38	0.39	0.54	0.92	10.70	649	/	b	-

	1,4-Pentadiene	4.61	0.79	0.26	2.53	16.93	675	/	b	-
	Cyclohexane	4.96	0.92	0.50	0.31	14.37	700	719	a	-
	n-Heptane	4.96	0.16	0.72	4.22	15.01	700	/	a	*Sweet, ethereal*
	2-Methylheptane	5.31	0.27	0.88	3.15	6.70	754	/	b	-
	n-Octane	5.72	0.99	1.06	2.15	15.96	800	/	a	*Solvent, unpleasant*
	1-Octene	6.24	0.95	1.00	0.84	5.71	822	838	b	*Gasoline*
	2,3-Dimethylheptane	6.48	0.64	1.64	0.57	9.44	839	847	b	-
	2,4-Dimethyl-1-heptene	6.77	0.83	1.12	2.12	18.67	859	878	b	-
	n-Nonane	7.62	0.33	1.72	3.00	4.08	900	/	a	-
	Benzene	8.11	0.37	0.48	1.61	15.84	924	934	a	*Aromatic*
	3,4-Diethyl-1,5-hexadiene (meso)	8.87	0.50	1.54	3.34	6.90	952	966	a	-
	3,4-Diethyl-1,5-hexadiene (RS+SR)	9.10	0.59	1.52	1.50	15.25	961	968	a	-
	n-Decane	10.21	0.97	2.44	1.90	5.35	1000	/	a	-
	(5Z)-3-Ethyl-1,5-octadiene	10.33	0.62	1.70	4.40	18.95	1005	1006	a	-
	(5E)-3-Ethyl-1,5-octadiene	10.79	0.04	1.72	4.27	13.16	1018	1012	a	-
	4-Methyldecane	10.79	0.89	2.72	0.23	10.79	1018	1022	b	-
	Toluene	11.03	0.54	0.68	4.86	13.14	1024	1024	a	*Sweet*
	1-Decene	11.43	0.88	2.04	0.73	20.43	1035	1039	b	-
	(*E,Z*)-3,7-Decadiene	12.66	0.36	1.82	0.49	19.42	1069	1069	a	-
	(*E,E*)-3,7-Decadiene	12.95	0.87	1.80	1.42	6.64	1077	1077	a	-
	n-Undecane	13.77	0.44	2.86	1.99	10.91	1100	/	a	-
	1,3-Dimethylbenzene	14.12	0.69	0.86	0.50	14.17	1109	1122	b	-
	Ethylbenzene	14.47	0.62	0.86	4.45	11.26	1118	1125	b	-
	1-Dodecene	17.56	0.62	2.50	0.89	19.10	1197	1192	b	-
	n-Dodecane	17.68	0.22	3.04	0.89	2.11	1200	/	a	*Alkane*
	Styrene	19.13	0.56	0.66	0.29	8.42	1237	1242	b	*Sweet, balsamic*
	(E)-4,8-Dimethylnona-1,3,7-triene	20.07	0.59	1.46	4.56	8.14	1260	1266	b	-
	1,2,3-Trimethylbenzene	20.18	0.78	0.96	1.11	13.72	1263	1282	b	-
	n-Tridecane	21.64	0.90	3.08	3.31	10.77	1300	/	a	-
	1-Ethenyl-4-ethylbenzene	26.13	0.51	0.86	4.38	13.83	1417	/	b	-
Terpenoids	α-Pinene	10.68	0.88	1.62	1.26	17.60	1015	1017	a	*Herbal, woody*
	β-Pinene	13.07	0.84	1.60	2.12	3.80	1081	1072	a	*Terpenic, pine*
	ß-Myrcene	15.69	0.54	1.32	2.11	15.33	1149	1154	a	*Spicy, woody*

	Limonene	17.09	0.68	1.40	4.56	18.55	1185	1190	a	*Citrus*
	Eucalyptol	17.38	0.59	1.54	1.17	6.53	1192	1195	a	*Herbal, minty*
	Terpinene	18.43	0.81	1.24	2.99	12.14	1219	1221	a	-
	(E)-ß-Ocimene	19.13	0.79	1.24	0.58	11.92	1237	1239	a	-
	α-Copaene	28.35	0.70	2.16	1.24	17.55	1478	1485	a	*Woody, spicy*
	Linalool	29.40	0.78	0.90	1.21	8.86	1507	1507	a	*Citrus*
	α-Muurolene	36.40	0.92	1.54	1.43	17.84	1711	1714	b	*Woody*
	α-Farnesene	36.98	0.16	1.44	1.30	10.91	1730	1740	a	-
Saturated aldehydes	Propanal	5.43	0.05	0.28	1.73	4.23	769	762	a	*Ethereal, musty*
	Butanal	6.71	0.43	0.42	2.09	14.68	856	861	a	*Cocoa, pungent*
	Pentanal	9.10	0.01	0.64	2.06	19.63	961	965	a	*Fermented, winey*
	Hexanal	12.43	0.40	0.86	4.41	14.35	1063	1066	a	*Tallowy, leaf-like*
	Heptanal	15.87	0.55	1.02	1.18	20.24	1154	1161	a	*Green, aldehydic*
	2-Ethylhexanal	16.45	0.74	1.32	0.89	18.25	1169	1187	a	-
	Octanal	19.78	0.19	1.14	2.02	14.25	1253	1268	a	*Fatty, sharp*
	Nonanal	24.44	0.39	1.22	1.24	20.93	1373	1380	a	*Fatty, waxy, pungent*
	Decanal	28.35	0.09	1.30	3.84	5.74	1478	1475	a	*Penetrating, sweet, waxy*
	Undecanal	32.08	0.57	1.34	3.93	12.16	1583	1585	a	*Sweet, fatty, floral-citrus*
	Dodecanal	35.64	0.78	1.40	2.65	14.35	1688	1688	a	*Soapy, fatty, aldehydic*
	Tridecanal	38.79	0.74	1.42	1.18	7.32	1787	1792	b	*Soapy, fresh, clean*
Unsaturated /aromatic aldehydes	(E)-2-Butenal	10.27	1.00	0.90	0.66	14.45	1003	1002a	a	-
	(Z)-2-Pentenal	13.24	0.01	0.64	4.81	13.17	1085	1101	a	-
	(E)-2-Pentenal	14.06	0.96	0.64	2.76	13.05	1107	1111	a	*Pungent, apple-like*
	3-Methyl-2-butenal	16.16	0.67	0.58	0.61	19.81	1161	1164	a	*Fruity, sweet*
	(Z)-2-Hexenal	16.86	0.13	0.78	3.32	14.98	1179	1183	a	*Fruity*
	(E)-2-Hexenal	17.50	0.80	0.82	1.59	2.06	1195	1204	a	*Almond, green*
	(Z)-2-Heptenal	18.32	0.96	0.72	3.02	8.08	1216	1218	a	*Green, oily, sharp*
	2-Ethyl-2-hexenal	20.83	0.45	1.14	1.94	13.83	1279	1285	a	-
	(E)-2-Heptenal	21.93	0.51	1.06	0.69	15.83	1307	1318	a	*Fatty, almond-like*
	(E,Z)-2,4-Hexadienal	24.38	0.29	0.58	4.48	16.10	1371	1373	a	*Green*
	(E,E)-2,4-Hexadienal	24.50	0.38	0.61	2.32	12.38	1375	1376	a	*Green*

	(E)-2-Octenal	24.97	0.29	0.86	1.01	19.78	1386	1391	a	*Fresh, fatty, waxy*
	(E,E)-2,4-Heptadienal	26.89	0.83	0.68	0.96	15.59	1438	1441	a	*Fatty, green, oily*
	Benzaldehyde	29.05	0.16	0.50	3.46	10.58	1497	1499	a	*Almond, burnt sugar*
	(E)-2-Nonenal	29.28	0.76	1.10	4.60	17.09	1503	1509	a	*Fatty, green, cucumber*
	(E)-2-Decenal	33.43	0.44	1.36	3.20	13.67	1622	1625	a	*Waxy, fatty, earthy*
	(E,E)-2,4-Decadienal	37.22	0.74	0.90	1.98	3.81	1737	1740	a	*Fatty, oily*
	2,4-Dimethylbenzaldehyde	38.68	0.21	0.64	4.59	4.09	1783	1789	b	*Almond, bitter*
Ketones	Acetone	5.72	0.08	0.30	4.97	14.82	800	819	a	*Pungent, apple-like*
	2-Butanone	7.12	0.83	0.44	1.48	10.16	880	887	a	*Ethereal, chemical*
	3-Buten-2-one	8.17	0.40	0.40	2.16	13.95	926	931	a	*Sweet*
	2,3-Butanedione	8.75	0.27	0.38	0.41	13.05	948	954	a	*Buttery, sweet*
	1-Penten-3-one	10.27	0.27	0.56	2.38	6.61	1003	1019	a	*Pungent, spicy*
	2,3-Pentanedione	10.97	0.14	0.52	2.12	14.05	1023	1026	a	*Buttery, toasted*
	3-Penten-2-one	13.88	0.63	0.58	2.75	19.70	1103	1106	a	*Fruity, acetone*
	4-Heptanone	14.06	0.47	1.10	0.25	19.18	1107	1118	b	*Fruity, cheesy*
	3-Heptanone	15.11	0.26	1.08	4.68	9.32	1134	1141	a	*Green, fatty*
	2-Heptanone	16.33	0.27	1.02	2.66	5.21	1166	1169	a	*Sweet, fruity*
	2-Octanone	19.22	0.45	1.10	1.97	13.31	1239	1244	a	*Mould, green*
	6-Methyl-5-hepten-2-one	21.70	0.11	0.90	3.33	18.52	1302	1313	a	*Pungent, green*
	(E)-3-Octen-2-one	24.68	0.69	0.92	4.03	12.79	1379	1384	a	-
	5-Methyl-2-(1-methylethyl)cyclohexanone	27.13	0.07	1.30	2.64	11.88	1444	1448	b	-
	3,5-Octadien-2-one	28.88	0.24	0.76	0.14	10.74	1492	1492	a	*Fruity, fatty, mushroom*
	2-Decanone	29.46	0.38	1.54	2.53	11.03	1508	1515	a	*Orange, floral, peach*
Others	Tetrahydrofuran	6.48	0.29	0.46	3.65	16.69	838	845	b	-
	1-Methoxyhexane	8.46	0.86	1.12	0.81	11.63	937	941	b	-
	2-Ethylfurane	8.46	0.03	0.54	3.58	2.11	937	944	a	-
	Acetonitrile	9.74	0.75	0.26	2.91	10.67	985	988	b	-
	Furfural	28.18	0.84	0.40	3.16	4.56	1473	1477	a	*Bready, brown, sweet*
	Dimethyl sulfoxide	30.28	0.13	0.38	0.19	3.85	1532	1549	b	*Fatty, oily*
	1-Chloro dodecane	34.48	0.98	1.76	3.71	20.48	1653	1661	b	-
	3,4-Dimethyl-2,5-furandione	35.41	0.52	0.66	4.77	2.66	1681	1685	b	-
	Acetamide	36.46	0.01	0.18	3.30	9.61	1713	1725	b	-

	Phthalide	53.43	0.66	1.76	3.84	17.87	2312	2323	a	-
	Diethyl phthalate (IS dilution solvent)	53.84	0.07	0.62	0.12	18.26	2329	2332	a	-
	Average %RSD value	**/**	**0.34**	**/**	**3.01**	**11.98**	**/**			**-**

Identifications were confirmed by authentic standards when available in authors' laboratory (criterion "a" in **Table 5.6.1**) or by spectral similarity DMF ≥900 and RMF ≥950 and I^T tolerance ± 20 units (criterion "b" in **Table 5.6.1**). **Table 5.6.1** lists target analytes with ^{1}D and ^{2}D retention times (^1t_R; ^2t_R), precision data (see ***section 5.6.3.7***), experimental (Exp.) and tabulated (Lit.) ^{1}D I^T values, and odor descriptors as reported in reference literature.

The output table collecting 2D peaks and peak-regions aligned across all chromatograms with features-related metadata (^{1}D and ^{2}D retention times, MS spectrum, base peak and molecular ion m/z, and TIC response) were stored and made available for further processing.

Supplementary Table 5.6.2 lists untargeted and targeted peak-region features included in the *UT* template, together with their experimental ^{1}D I^T values, retention times in the two analytical dimensions (^{1}t$_R$, ^{2}t$_R$), % relative standard deviation (% RSD) on retention times across all analyses, and reference MS spectral signature from *peak-apex* spectrum.

5.6.3.7 Method performance parameters

Repeatability was evaluated on analytical descriptors considered fundamental for an accurate chromatographic fingerprinting based on both 2D peak patterns and analytes responses. Therefore, %RSD was calculated on retention times and analytes % response (% normalized 2D volumes over IS) for all targeted compounds and on analytical replicates of the same sample analyzed every two-days over the two-weeks of the study (n = 6). Results are reported in **Table 5.6.1**. Mean %RSD on retention times were, respectively, 0.34% for the ^{1}D (^1t_R) and 3.01% for the ^{2}D (^2t_R). Maximum %RSD on percent response was instead 20.93%, reported for nonanal, while the mean value was 11.98%.

5.6.3.8 Data acquisition and 2D data processing

Data were acquired by TOF-DS software (Markes International, Llantrisant, UK) and processed by GC Image V2.9 suite (GC Image, LLC Lincoln, NE, USA).

The peak-region features data files from each chromatogram were exported in '.xls' format (Microsoft Excel), and then converted to MATLAB format (version R2017b). All the multivariate analysis was performed using PLS_Toolbox 8.6.1 (Eigenvector Research, Manson, WA USA) for MATLAB environment (MathWorks Inc., Massachusetts, USA, R2017b). Principal component analysis (PCA), partial least-squares regression (PLS) and soft independent modelling for class analogy (SIMCA) were applied as exploratory analysis, variable selection and classification method, respectively. In addition, Microsoft Excel spreadsheet was used for similarity analysis.

5.6.4 Results and Discussion

Chromatographic fingerprinting based on comprehensive two-dimensional separations has a great potential for discrimination and identification of samples based on their chemical signatures, a process described as *identitation*.[10] Moreover, it offers further advantages, when mass spectrometry is used at the detection level, providing additional information for analytes putative identification. This step gives access to a higher information level on samples properties and characteristics.[7,17,29]

The strategy adopted to encrypt the hidden information from volatiles patterns of EVOOs harvested in different Italian Regions collects information by *Untargeted* and *Targeted (UT)* features. It is a fingerprinting approach designed to comprehensively map all detectable volatiles from GC×GC-TOF MS analyses.[15] Chromatograms processing was by a validated workflow described in ***Section 5.6.3.6***; the output was a data matrix 73 × 519 dimensioned (*i.e.*, samples × features) with a sub set of 159 identified (targeted) compounds.

Next section, while introducing the chemical dimensionality and the information encrypted on EVOOs volatile fraction, highlights the fundamental role of high-resolution separations and retention pattern logic at the basis of samples *identitation*. Machine learning, based on multivariate statistics and modelling algorithms, will be presented as key-tools to access the higher level of information to identify distinctive Regional markers patterns.

5.6.4.1 The complex and multidimensional EVOO volatilome

EVOO is highly appreciated by consumers because of its unique and characteristic flavor, which reflects the chemical complexity and dimensionality[30] of its volatile fraction, characterized by the presence of many compounds, especially carbonyls (*e.g.*, aldehydes, ketones), esters, alcohols, and hydrocarbons (*e.g.*, linear, aromatic, terpenoids etc.). Odor active compounds, with low odor perception threshold, and volatiles lacking of sensory features (*i.e.*, interferents),[31] concur in the modulation of the "odor code" while triggering aroma perception which objectification by instrumental methods is challenging.[9,32] However, EVOO volatiles encrypt additional information about relevant functional variables including olive cultivars, olive trees harvest region and local pedoclimatic conditions, olive ripeness, technological processes, and storage condition.[1,2,33,34] Methodologies capable of comprehensively and reliably capturing the complexity of OO volatilome, have the potential of being valid and robust fingerprinting tools for discrimination, *identitation*,[10] and valorization of high-quality product. This latter process, might help in supporting European Union food quality labelling ("EU Food Qual. Labels," 2021) while promoting food fraud counteractions in the world market of EVOO.

Figure 5.6.1A shows the pseudocolor image of a Sicilian EVOO (Sicilia origin – #S1) volatile fraction comprehensively mapped through untargeted and targeted (*UT*) peak-regions (red graphics); identified/targeted analytes (*i.e.*, targeted compounds) are highlighted by green circles. Patterns of analytes, following a retention logic based on the relative retention exerted by the polar × semipolar column combination adopted, are highlighted in **Figure 5.6.1B** and **5.6.1C**.

Compounds deriving from linoleic and linolenic acids oxidative cleavage, promoted by lipoxygenase (LOX) and hydroperoxide lyase (HPL) pathways, constitute the LOX signature (green color area in **Figure 5.6.1B** and enlarged area in **Figure 5.6.1C**), which is the most abundant fraction in high-quality EVOOs.[1,12] It is characterized by the presence of C_6 and C_5 compounds, in particular aldehydes, alcohols, ketones, and esters (*e.g.*, hexanal, *(E)*-2-hexenal, 1-penten-3-ol, 1-hexanol, 1-penten-3-one, hexyl acetate etc.), fundamental to define positive attributes as *fruity* and *green*.[13,20]

Saturated and unsaturated aldehydes (respectively in brown and orange) are mainly produced by oxidation of the unsaturated fatty acids.[35] While C_6 and C_5 unsaturated aldehydes from LOX are correlated to positive attributes, the others, with higher molecular weight and low odor threshold (*e.g.*, *(E)*-2-heptenal, *(E)*-2-octenal, *(E)*-2-decenal, heptanal, octanal and nonanal) are indicated in many studies as responsible of the *rancid* off-flavor with unpleasant and penetrating notes.[18,35,36] Alcohols (purple line in **Figure 5.6.1B**), represented by thirty congeners here identified, have a strong retention in the ^{1}D polar column and are well separated by informative carbonyls. Of them the most relevant are 1-octen-3-ol, 1-nonanol, and 1-decanol, because of their decisive role in defining sensory defects eliciting *fatty, rancid, earthy,* and *mushroom-like* notes.[23,34]

Short-chain fatty acids (**Figure 5.6.1B** - black line) derive from the oxidation of the corresponding aldehydes[18,35] with propanoic and butanoic acids as the most odor active, followed by pentanoic and heptanoic acids. Their presence was correlated to the perception of *rancid* and *fusty* defects.[1,35,36]

Hydrocarbons (**Figure 5.6.1B** in cyano) have a negligible contribution in the definition of the EVOO flavor, although some unsaturated derivatives (*i.e.*, 3-ethyl-1,5-octadiene and 4,8-dimethyl-1,3,7-nonatriene) were linked to *green* and *fruity* notes,[12] or to *rancid* and *fishy* aroma.[1] Moreover, a series of C10 alkenes, *i.e.*, 3,4-diethyl-1,5-hexadiene (*RS* or *SR*), 3,4-diethyl-1,5-hexadiene (meso), *(5Z)*-3-ethyl-1,5-octadiene, *(5E)*-3-ethyl-1,5-octadiene, *(E,Z)*-3,7-decadiene, and *(E,E)*-3,7-decadiene, which elution region is highlighted in blue on **Figure 5.6.1B**, are known to be diagnostic markers of early ripening stages of olives (Angerosa, Camera, D'Alessandro, & Mellerio, 1998), while *n*-octane is an indicator of over-ripening.[3,15,23]

The presence and the abundance of terpenes (grey rectangles in **Figure 5.6.1B**) is of particular interest because of their role as indicators of geographical origin[7,13] or of ripening, *e.g.*, α-farnesene.[3] Moreover, they contribute to define positive attributes, as *wood*, *lemon* and *rose-like* odors.[12,38]

Lactones are generally detected in low but variable amounts in EVOO, and their relative concentration is cultivar-specific.[34] Esters as well, closely eluting to lactones, contribute in defining *fruity* notes with C_6 and C_5 derivatives deriving from the LOX pathway that dominates the class.[1,20,34]

5.6.4.2 Multivariate analysis

Firstly, an exploratory unsupervised analysis was carried out applying PCA; data structure was examined to whether geographical origin-related intrinsic groupings of olive oils samples were detectable. Then, six two-level PLS regression models (one for each concerned Italian region) were built in order to obtain the variable importance in prediction (VIP) scores and to select the variables which contribute the most to characterize each EVOO belonging to a particular Italian region against the rest of the samples. From these PLS models the six characteristic volatile patterns, one for of each geographical region, were delineated and a similarity analysis of the characteristic pattern of each region was carried out by applying the nearness index. Finally, six one-input class SIMCA classification models were developed and validated. **Figure 5.6.2** shows the multivariate analysis workflow designed to capture informative and diagnostic patterns capable of correctly classifying/discriminating EVOO production regions.

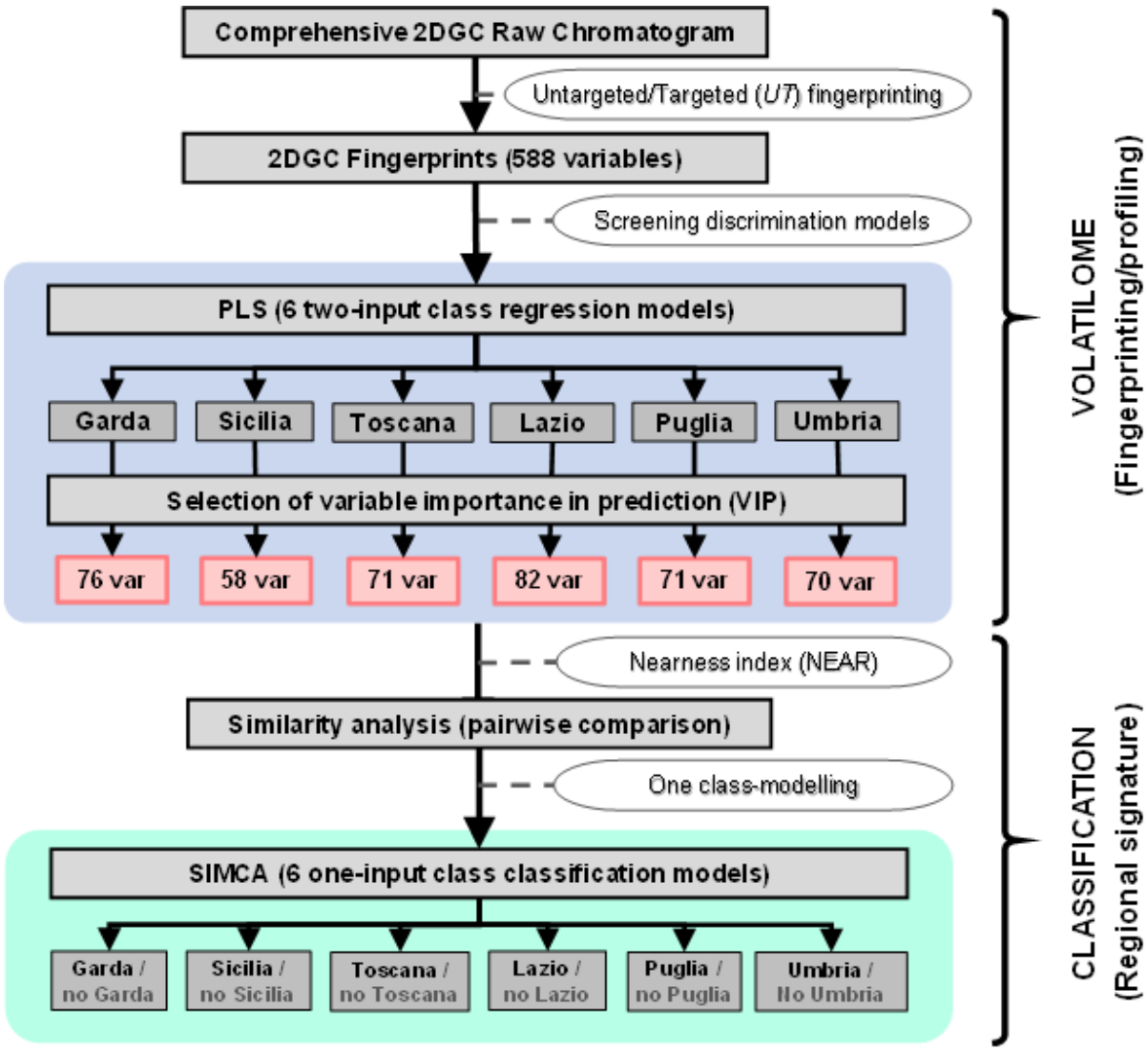

Figure 5.6.2. Workflow including data processing (*i.e.,* fingerprinting and profiling) and machine learning

5.6.4.2.1. Exploratory analysis

PCA was initially performed considering the 591 variables (*i.e.*, peak-region features) per sample (n=73). After inferring from this first PCA model, three variables were removed: phthalide, *(E)*-2-hexenal and toluene because the related loadings were very large in all cases, and they were masking the behavior of the other variables. Finally, a new PCA with 588 variables was developed and all the successive multivariate analysis steps were carried out with these variables.

The new PCA model was built with 12 principal components, which explained a total of variance of the 79.73%. **Figure 5.6.3** displays the score plot on PC1 *vs.* PC2. Some particular grouping trends were observed for the olive oil samples from Sicilia, Lazio and Umbria. In addition, Garda and Puglia were spread over the bottom and top halves of the PC1.

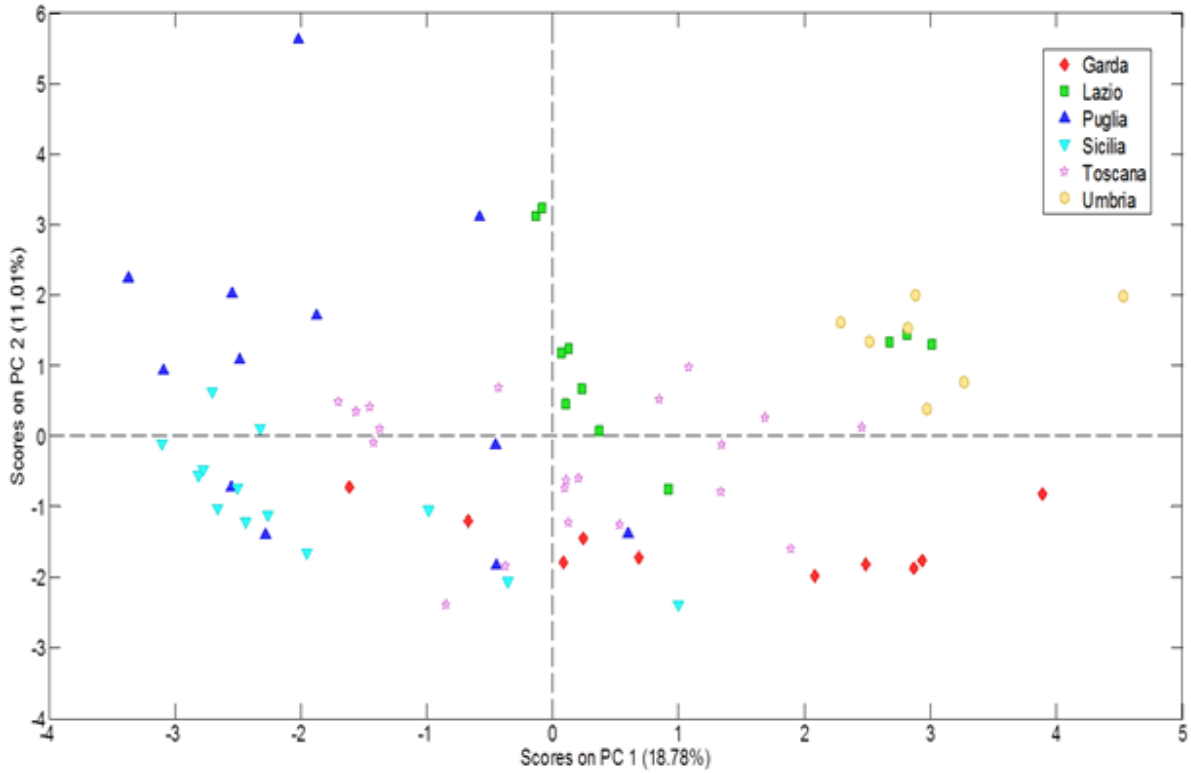

Figure 5.6.3. Principal Component Analysis – PCA score plot on PC1 vs. PC2 accounting for a 79.73 % of the total explained variance. The PCA model is based on the TIC % Response from UT peak-regions comprehensively covering the chromatographic space. EVOOs from different Italian Regions are displayed in different colors.

5.6.4.2.2. Variable selection – Characteristic profile

The variable importance in the projection (VIP) score, that summarize the overall contribution of each variable to the PLS model, was used as variable selection strategy in order to highlight characteristic volatile patterns for each Region. The "greater-than-one-rule" was applied for selecting the VIP scores, and only about 12% from the total number of variables (588) were selected as characteristics. On this way the number of selected variables per region was Garda (76), Sicilia (58), Toscana (71), Lazio (82), Puglia (71) and Umbria (70), counting together a total of 121 variables which were displayed in a bar-chart shown in **Supplementary Figure 5.6.2**. **Supplementary Tables 5.6.3-8** list, for each Italian Region, the specific variables and include both untargeted and targeted components.

5.6.4.2.3. Similarity study

The similarity analysis among the Region characteristic patterns was carried out by calculating similarity indices. Such indices are defined as a number between 0 to 1 which describes the equivalence of two objects characterized by multivariate data; the value 0 indicates maximum difference, and 1 implies maximum similarity. In this study the nearness index (NEAR)[39] was employed; it is calculated by **Equation 5.6.1**:

Equation 5.6.1

$$\mathrm{NEAR}\,(x_C, x_R) = 1 - \frac{\sqrt{\Sigma\,(x_{C_i} - x_{R_i})^2}}{\sqrt{\Sigma\,(x_{C_i} + x_{R_i})^2}}$$

where x_{c_i} and x_{r_i} symbolise each element of the considered and reference characteristic profiles, respectively.

Equation 5.6.1 could be also re-formulated in matrix notation as reported in **Equation 5.6.2**:

Equation 5.6.2

$$\mathrm{NEAR}(\mathbf{X_C}, \mathbf{X_R}) = 1 - \left[\sqrt{\frac{(\mathbf{X_C} - \mathbf{X_R}) \times (\mathbf{X_C} - \mathbf{X_R})^T}{(\mathbf{X_C} + \mathbf{X_R}) \times (\mathbf{X_C} + \mathbf{X_R})^T}} \right]$$

where, correspondingly, $\mathbf{X_{C_i}}$ and $\mathbf{X_{Rr_i}}$ are the considered and reference characteristic profile vectors, respectively (the superscript T is denoting the transposed matrix).

To carry out the similarity study a new reduced tertiary vector consisting of 0, 1 and 2 codes for each Region was built from the regional characteristic profiles; results reported in **Table 5.6.2**. The following rules were applied to establish the aforementioned codes:

- 0: was assigned to those variables not selected as part of the Region characteristic pattern, *e.g.*, the variable 32 corresponding to methyl benzoate was selected only for Puglia profile, and thus this variable was codified with the value 0 for the remaining reduced tertiary vectors.
- 1: was assigned to those variables whose VIP scores ranged from 0 to 1, *e.g.*, the variable 46 corresponding to ethyl benzoate was selected for Lazio, Puglia, and Umbria profiles.
- 2: was assigned to those variables whose VIP scores was greater than 1, *e.g.*, the variable 8 corresponding to α-copaene had a VIP score larger than 1 for Garda, Sicilia, Toscana, Lazio and Puglia profiles and a VIP score between 0 and 1 for the Umbria profile. Thus, in the five pattern (Garda, Sicilia, Toscana, Lazio and Puglia) this variable was codified with the value 2 and in Umbria having the value 1.
-

Table 5.6.3. Reduced tertiary vector from six studied Regions from Italy.

VIP Variables	GARDA	SICILIA	TOSCANA	LAZIO	PUGLIA	UMBRIA
α-Pinene	2	0	1	1	1	2
α-Copaene	2	2	2	2	2	1
Terpinene	1	0	0	0	0	0
Pentanal	1	2	2	1	1	2
Nonanal	1	0	2	1	2	0
n-Octane	0	0	0	1	0	2
n-Hexane	2	0	0	1	0	0
Methyl benzoate	0	0	0	0	1	0
Limonene	2	0	2	2	2	2
Hexyl acetate	0	0	1	2	0	1
Hexanal	1	1	2	1	1	1
Ethyl benzoate	0	0	0	1	1	1
Dodecanal	2	1	1	1	2	0
Diethyl phtalate	0	0	0	2	2	0
Cyclopentane	0	0	0	1	0	1

Cyclohexane	2	1	1	0	1	2
Butyl isobutyrate	0	1	0	1	0	1
Butanoic butanoate	1	0	1	2	2	2
Benzaldehyde	0	0	0	0	1	0
Acetonitrile	0	0	0	1	2	1
6-Methyl-5-hepten-2-one	0	0	0	0	0	1
4-Hydroxy-2-hexenoic acid lactone	2	2	2	2	1	2
3,4-Diethyl-1,5-hexadiene (RS+SR)	0	1	0	1	2	1
3,4-Diethyl-1,5-hexadiene (meso)	1	0	0	1	0	0
2-Ethyl-2-hexenal	0	0	1	2	0	0
2-Ethyl-1-hexanol	0	0	1	0	1	0
1-Penten-3-one	1	0	2	2	1	2
1-Penten-3-ol	2	0	0	1	0	0
1-Methoxyhexane	1	0	1	0	0	0
1-Hexanol	1	2	2	2	2	2
1-Butanol	0	0	0	1	0	0
1,4-Pentadiene	1	0	0	1	0	0
(Z)-3-Hexenyl acetate	1	2	2	1	2	2
(Z)-3-Hexen-1-ol	2	2	2	2	2	2
(Z)-2-Pentenal	0	0	0	0	1	0
(Z)-2-Hexenal	1	1	1	0	1	0
(E,Z)-3,7-Decadiene	2	2	2	2	2	2
(E,Z)-2,4-Hexadienal	0	0	0	1	0	0
(E,E)-3,7-Decadiene	2	2	2	2	2	2
(E,E)-2,4-Hexadienal	1	2	0	0	0	0
(E)-ß-Ocimene	1	0	0	1	0	0
(E)-2-Pentenal	0	1	1	0	0	0
(E)-2-Penten-1-ol	1	1	1	2	2	1
(E)-2-Hexen-1-ol	1	2	2	1	2	2
(E)-2-Heptenal	0	0	1	0	0	0
(5Z)-3-Ethyl-1,5-octadiene	1	1	1	2	2	2
(5E)-3-Ethyl-1,5-octadiene	2	2	2	2	2	2
160	0	0	0	0	1	0
166	0	0	0	0	0	1
175	0	1	1	1	0	1
178	0	0	0	1	0	0

187	1	0	0	0	0	0
188	2	1	1	0	1	0
190	0	1	0	0	0	0
201	0	1	1	2	0	1
205	0	1	0	1	1	1
210	0	0	0	1	1	0
213	1	1	1	1	1	2
215	2	0	1	1	0	0
227	1	2	1	0	0	0
232	0	0	2	1	1	2
238	1	1	2	2	1	2
243	1	2	1	0	0	0
244	0	0	0	1	0	1
245	2	0	0	0	0	2
254	2	1	2	2	2	0
258	1	1	0	0	0	1
262	2	0	1	0	1	0
267	2	1	1	0	2	1
268	0	0	0	0	0	1
274	1	0	2	1	2	1
275	1	1	1	1	1	0
276	1	0	1	0	0	1
278	1	0	0	0	1	0
280	0	0	0	0	0	1
285	1	1	1	1	2	1
294	1	0	0	0	0	0
299	1	2	1	2	1	2
305	1	1	1	2	1	1
314	2	0	1	2	2	2
318	2	2	2	2	2	2
323	2	0	2	2	1	2
331	1	2	1	2	2	2
340	1	2	2	2	2	2
341	0	0	0	0	1	0
349	0	1	2	1	1	2
359	0	0	0	0	1	0
374	2	2	2	1	1	2
375	0	0	0	1	0	0

377	1	2	1	2	2	2
382	2	2	2	2	2	2
386	2	0	1	0	0	0
388	2	0	1	1	0	1
396	0	0	1	0	1	0
400	2	0	0	1	1	0
413	2	2	2	2	2	2
418	2	2	2	2	2	0
422	2	1	1	1	2	1
424	0	1	0	0	0	0
430	1	0	0	0	0	0
445	0	0	0	1	0	0
454	1	2	2	1	2	2
457	0	0	0	1	0	1
458	1	2	2	1	2	0
468	0	0	0	0	1	0
475	2	0	0	0	0	0
480	0	0	0	1	1	0
498	0	1	0	1	1	1
504	1	0	1	1	1	1
505	2	1	2	1	2	2
510	2	0	0	1	0	0
523	1	2	1	1	0	0
529	1	0	1	1	0	1
538	0	0	1	0	0	0
549	1	0	0	0	0	0
550	0	1	0	0	1	0
554	1	2	1	0	0	0
555	2	2	2	1	0	2
561	0	1	0	0	0	0
573	2	1	2	1	2	1
588	1	1	2	1	1	2

Once the reduced tertiary vectors from characteristic patterns were pairwise compared, a similarity matrix was constructed from the found NEAR values, which is shown in the **Figure 5.6.4**.

UMBRIA						
PUGLIA						0.609
LAZIO					0.656	0.662
TOSCANA				0.658	0.667	0.662
SICILIA			0.646	0.558	0.579	0.569
GARDA		0.548	0.649	0.584	0.584	0.556
	GARDA	SICILIA	TOSCANA	LAZIO	PUGLIA	UMBRIA

Figure 5.6.4. Similarity matrix based on NEAR index resulting from pairwise comparison of Regional patterns. Details are provided in Section 5.6.4.2.3.

As can be seen in the similarity matrix, in all cases the NEAR value is significantly less than 1, indicating that the volatiles profile/pattern between the regions is significantly dissimilar. Therefore, it may be used to classify samples according to geographical origin. In addition, there were 5 variables out of the 121 selected, that were present in all characteristic patterns having a code higher than 1. It was therefore decided to remove them from the classification models as their contribution to the discrimination among Regions would not be relevant.

5.6.4.2.4. Classification according to harvest/production Region unique signature

The most conventional way to develop a classification model is based on building a model with two input-classes, the target class and the non-target class, but a valid alternative is performing the same classification method by training with a single input-class, *i.e.*, the target class.[40] Working with one input-class classification has significant advantages in food authentication, the model is trained using the data from representative samples from genuine foods (target class) and no other samples are required. In fact, some authors have stated that it is advisable to develop the models using one-class classifier in the case of food authentication. Indeed, if well-known discriminant method such as partial lest squares-discriminant analysis (PLS-DA) is used and a new sample does not belong to any of such classes, the discriminant analysis is unable to properly define the belonging of the sample to one particular class. Conversely, one-class classifier such as SIMCA, establishes if the acceptance is around the target class delimiting the target samples from other classes.[41]

SIMCA involves building a classification method in which each class of the training set is modeled independently and the assignment of an unknown sample as belonging to a specific class is based on the nearest distance to the corresponding regions established in the space of principal components. Six one input-class SIMCA models were built, one for each Italian Region. Each individual model was developed using the 116 untargeted/targeted features which were selected in at least one of the Regional characteristic patterns. The aim was to generate overall models suitable for application in routine analysis. Otherwise, should it be required to classify a sample of unknown

origin, which characteristic variables would be selected or chosen? In this way, any classification model developed can be applied and it will be possible to assign a class to the sample. **Table 5.6.3** shows the numbers of PCs chosen for each model and the samples used in the training and validation step.

Table 5.6.4. Characteristics of the SIMCA models.

Origin	PCs	% variance	Training set	Validation set*
Garda	7	99.97	10 samples (Garda)	73 samples
Sicilia	8	93.77	13 samples (Sicilia)	73 samples
Toscana	12	94.11	20 samples (Toscana)	73 samples
Lazio	7	99.82	11 samples (Lazio)	73 samples
Puglia	8	94.49	12 samples (Puglia)	73 samples
Umbria	5	97.23	7 samples (Umbria)	73 samples

*** For the validation were employed all the samples analyzed**

Class boundaries were established for each pre-defined target class model on the basis of the values of the Hotelling's T^2 and residual Q statistics. The classification criterion of the samples regarding each Region were defined using a combination of the reduced T^2 and Q statistic values. Thus, for a sample to be considered as belonging to certain target class, both values must be less than 1.0.

Because the number of available samples from each Italian Region was limited, each single-class model training was carried out employing all the samples belonging to the concerned target class. Then, all 73 samples, both those belonging to the target class and those not, were used for validation purpose. All the samples were correctly classified in five of the six single-class models and the quality parameters such as sensitivity, specificity, precision, efficiency (accuracy) and area under the receiver operating characteristic curve (AUC) were equal to 1.00 [40]. The only model that misclassified one of the samples was the Garda model, in which a Garda sample was considered as non-Garda. Thus, in this model the sensitivity, specificity, precision, efficiency (accuracy) and AUC were equal to 0.90, 1.00, 1.00, 0.99 and 0.95 respectively. The classification plots of each model are shown in the supplementary material (**Supplementary Figures 5.6.3-8**).

5.6.4.3 Regional signatures and GC×GC identitation potential

Based on the information shown in **Table 5.6.2**, it is possible to derive some conclusions about peculiar chemical traits specific to certain Regions. For example, compounds #28 (*n*-hexane), #109 (1-penten-3-ol) and #386, #475 and #510 (all unidentified) are characteristic of the Garda region. In the same way, compound #141 ((*E,E*)-2,4-hexadienal) is specific of Sicilia samples, compound #95 (2-ethyl-2-hexenal) of Lazio and compound #27 (*n*-octane) of Umbria. Further assignments could be identified as characteristic of more than one Region, for example, compound #245 (unidentified) is associated with Garda and Umbria. In the same way, following this assignment methodology, and considering the presence/absence of a few volatile compounds previously selected, a classification tree rule could be deduced in order to classify undoubtedly any sample of EVOO from any of the six considered Regions. However, it might be beneficial to have a larger set of representative olive oil samples from each of the Regions for such a classification tree to be sufficiently reliable.

The classification strategy proposed in this study is based on using the whole *UT* fingerprint of volatiles that is established by considering simultaneously all the compounds that have been selected as characteristic of at least one of the Regions concerned. In this way, the one-class SIMCA classification models are applied sequentially to any EVOO sample, regardless of geographic origin, so that the oil is assigned to one of the Regions. This overall classification approach based on the use of *UT* fingerprinting, *i.e.*, *identitation*, overcomes the main drawback for the routine application of single-step multivariate models.

Moreover, the strategy takes full advantage by the high-resolution power of GC×GC that effectively maps all detectable volatile components including: (a) those related to major functional variables (*e.g.,* olive cultivar,[12] olives ripening stage,[15] harvest year and processing technology[11]), here playing a confounding role for Regional classification; and (b) several potent odorants delineating EVOO sensory features. The latter might be masked by co-elution phenomena occurring in 1D-GC[17] while resulting in less effective *identitation* and poorly informative profiling processes.

5.6.5 Supplementary material

Supplementary material at the Google Drive's link:
https://drive.google.com/drive/folders/1dw3d3BvjJIrAmzTWl_qhkf1GFTmHyIBF?usp=sharing

References

(1) Angerosa, F.; Servili, M.; Selvaggini, R.; Taticchi, A.; Esposto, S.; Montedoro, G. Volatile Compounds in Virgin Olive Oil: Occurrence and Their Relationship with the Quality. *Journal of Chromatography A*. October 2004, pp 17–31.

(2) Cecchi, L.; Migliorini, M.; Giambanelli, E.; Rossetti, A.; Cane, A.; Mulinacci, N. New Volatile Molecular Markers of Rancidity in Virgin Olive Oils under Nonaccelerated Oxidative Storage Conditions. *J. Agric. Food Chem.* **2019**, *67* (47), 13150–13163.

(3) Hachicha Hbaieb, R.; Kotti, F.; Gargouri, M.; Msallem, M.; Vichi, S. Ripening and Storage Conditions of Chétoui and Arbequina Olives: Part I. Effect on Olive Oils Volatiles Profile. *Food Chem.* **2016**, *203*, 548–558.

(4) International Olive Council COI/T.15/NC No 3/Rev. 15. *International Trade Standard Applying to Olive Oils and Olive-Pomace Oils International Trade Standard Applying to Olive Oils and Olive-Pomace Oils. COI/T.15/NC No 3/Rev. 15*; 2019.

(5) Commission of the European Communities. Commission Regulation (Eec) Nº 2568/91. *Official Journal of the European Communities*. 1991, pp 1–83.

(6) EU Food Quality Labels https://ec.europa.eu/info/news/new-search-database-geographical-indications-eu-2020-nov-25_en (accessed Feb 3, 2021).

(7) Quintanilla-Casas, B.; Bertin, S.; Leik, K.; Bustamante, J.; Guardiola, F.; Valli, E.; Bendini, A.; Toschi, T. G.; Tres, A.; Vichi, S. Profiling versus Fingerprinting Analysis of Sesquiterpene Hydrocarbons for the Geographical Authentication of Extra Virgin Olive Oils. *Food Chem.* **2020**, *307* (March 2019).

(8) Purcaro, G.; Barp, L.; Beccaria, M.; Conte, L. S. Fingerprinting of Vegetable Oil Minor Components by Multidimensional Comprehensive Gas Chromatography with Dual Detection. *Anal. Bioanal. Chem.* **2015**, *407* (1), 309–319.

(9) Quintanilla-Casas, B.; Marin, M.; Guardiola, F.; García-González, D. L.; Barbieri, S.; Bendini, A.; Toschi, T. G.; Vichi, S.; Tres, A. Supporting the Sensory Panel to Grade Virgin Olive Oils: An in-House-Validated Screening Tool by Volatile Fingerprinting and Chemometrics. *Foods* **2020**, *9* (10), 1–14.

(10) Cuadros-Rodríguez, L.; Ruiz-Samblás, C.; Valverde-Som, L.; Pérez-Castaño, E.; González-Casado, A. Chromatographic Fingerprinting: An Innovative Approach for Food "identitation" and Food Authentication - A Tutorial. *Analytica Chimica Acta*. 2016, pp 9–23.

(11) Vaz-Freire, L. T.; da Silva, M. D. R. G.; Freitas, A. M. C. Comprehensive Two-Dimensional Gas Chromatography for Fingerprint Pattern Recognition in Olive Oils Produced by Two Different Techniques in Portuguese Olive Varieties Galega Vulgar, Cobrançosa e Carrasquenha. *Anal. Chim. Acta* **2009**, *633* (2), 263–270.

(12) Ros, A. Da; Masuero, D.; Riccadonna, S.; Bubola, K. B.; Mulinacci, N.; Mattivi, F.; Lukić, I.; Vrhovsek, U. Complementary Untargeted and Targeted Metabolomics for Differentiation of Extra Virgin Olive Oils of Different Origin of Purchase Based on Volatile and Phenolic Composition and Sensory Quality. *Molecules* **2019**, *24* (16), 1–17.

(13) Lukić, I.; Carlin, S.; Horvat, I.; Vrhovsek, U. Combined Targeted and Untargeted Profiling of Volatile Aroma Compounds with Comprehensive Two-Dimensional Gas

Chromatography for Differentiation of Virgin Olive Oils According to Variety and Geographical Origin. *Food Chem.* **2019**, *270*, 403–414.

(14) Aprea, E.; Gasperi, F.; Betta, E.; Sani, G.; Cantini, C. Variability in Volatile Compounds from Lipoxygenase Pathway in Extra Virgin Olive Oils from Tuscan Olive Germoplasm by Quantitative SPME/GC-MS. *J. Mass Spectrom.* **2018**, *53* (9), 824–832.

(15) Magagna, F.; Valverde-Som, L.; Ruíz-Samblás, C.; Cuadros-Rodríguez, L.; Reichenbach, S. E.; Bicchi, C.; Cordero, C. Combined Untargeted and Targeted Fingerprinting with Comprehensive Two-Dimensional Chromatography for Volatiles and Ripening Indicators in Olive Oil. *Anal. Chim. Acta* **2016**, *936*, 245–258.

(16) Purcaro, G.; Cordero, C.; Liberto, E.; Bicchi, C.; Conte, L. S. Toward a Definition of Blueprint of Virgin Olive Oil by Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2014**, *1334*, 101–111.

(17) Quintanilla-Casas, B.; Bustamante, J.; Guardiola, F.; García-González, D. L.; Barbieri, S.; Bendini, A.; Toschi, T. G.; Vichi, S.; Tres, A. Virgin Olive Oil Volatile Fingerprint and Chemometrics: Towards an Instrumental Screening Tool to Grade the Sensory Quality. *LWT* **2020**, *121*, 108936.

(18) Neugebauer, A.; Granvogl, M.; Schieberle, P. Characterization of the Key Odorants in High-Quality Extra Virgin Olive Oils and Certified Off-Flavor Oils to Elucidate Aroma Compounds Causing a Rancid Off-Flavor. *J. Agric. Food Chem.* **2020**, *68* (21), 5927–5937.

(19) Progetto Ager. Violin project https://olivoeolio.progettoager.it/index.php/olio-e-olivo-per-la-ricerca/itemlist/user/49-violin.

(20) Stilo, F.; Cordero, C.; Sgorbini, B.; Bicchi, C.; Liberto, E. Highly Informative Fingerprinting of Extra-Virgin Olive Oil Volatiles: The Role of High Concentration-Capacity Sampling in Combination with Comprehensive Two-Dimensional Gas Chromatography. *Separations* **2019**, *6* (3), 34.

(21) Vichi, S.; Guadayol, J. M.; Caixach, J.; Lopeztamames, E.; Buxaderas, S.; López-Tamames, E.; Buxaderas, S. Comparative Study of Different Extraction Techniques for the Analysis of Virgin Olive Oil Aroma. *Food Chem.* **2007**, *105* (3), 1171–1178.

(22) Stilo, F.; Liberto, E.; Reichenbach, S. E.; Tao, Q.; Bicchi, C.; Cordero, C. Untargeted and Targeted Fingerprinting of Extra Virgin Olive Oil Volatiles by Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry: Challenges in Long-Term Studies. *J. Agric. Food Chem.* **2019**, *67* (18), 5289–5302.

(23) Stilo, F.; Liberto, E.; Reichenbach, S. E.; Tao, Q.; Bicchi, C.; Cordero, C. Exploring the Extra-Virgin Olive Oil Volatilome by Adding Extra Dimensions to Comprehensive Two-Dimensional Gas Chromatography and Time of Flight Mass Spectrometry Featuring Tandem Ionization: Validation of Ripening Markers in Headspace Linearity Conditio. *J. AOAC Int.* **2020**, No. June, 1–14.

(24) Stilo, F.; Bicchi, C.; Jimenez-Carvelo, A. M.; Cuadros-Rodriguez, L.; Reichenbach, S. E.; Cordero, C. Chromatographic Fingerprinting by Comprehensive Two-Dimensional Chromatography: Fundamentals and Tools. *TrAC - Trends in Analytical Chemistry*. Elsevier Ltd 2021, p 116133.

(25) Reichenbach, S. E.; Carr, P. W.; Stoll, D. R.; Tao, Q. Smart Templates for Peak Pattern Matching with Comprehensive Two-Dimensional Liquid Chromatography. *J. Chromatogr. A* **2009**, *1216* (16), 3458–3466.

(26) Cordero, C.; Guglielmetti, A.; Bicchi, C.; Liberto, E.; Baroux, L.; Merle, P.; Tao, Q.; Reichenbach, S. E. Comprehensive Two-Dimensional Gas Chromatography Coupled with Time of Flight Mass Spectrometry Featuring Tandem Ionization: Challenges and Opportunities for Accurate Fingerprinting Studies. *J. Chromatogr. A* **2019**, *1597*, 132–141.

(27) Kiefl, J.; Cordero, C.; Nicolotti, L.; Schieberle, P.; Reichenbach, S. E.; Bicchi, C. Performance Evaluation of Non-Targeted Peak-Based Cross-Sample Analysis for Comprehensive Two-Dimensional Gas Chromatography-Mass Spectrometry Data and Application to Processed Hazelnut Profiling. *J. Chromatogr. A* **2012**, *1243*, 81–90.

(28) IOC. International Olive Council. *Newsl. No 144 DECEMBER 2019* **2019**, No. December.

(29) Stilo, F.; Bicchi, C.; Robbat, A.; Reichenbach, S. E.; Cordero, C. Untargeted Approaches in Food-Omics: The Potential of Comprehensive Two-Dimensional Gas Chromatography/Mass Spectrometry. *TrAC Trends Anal. Chem.* **2020**, *135*, 116162.

(30) Giddings, J. C. Sample Dimensionality: A Predictor of Order-Disorder in Component Peak Distribution in Multidimensional Separation. *J. Chromatogr. A* **1995**, *703* (1–2), 3–15.

(31) Cordero, C.; Kiefl, J.; Reichenbach, S. E.; Bicchi, C. Characterization of Odorant Patterns by Comprehensive Two-Dimensional Gas Chromatography: A Challenge in Omic Studies. *TrAC - Trends in Analytical Chemistry*. 2019, pp 364–378.

(32) Casadei, E.; Valli, E.; Aparicio-Ruiz, R.; Ortiz-Romero, C.; García-González, D. L.; Vichi, S.; Quintanilla-Casas, B.; Tres, A.; Bendini, A.; Toschi, T. G. Peer Inter-Laboratory Validation Study of a Harmonized SPME-GC-FID Method for the Analysis of Selected Volatile Compounds in Virgin Olive Oils. *Food Control* **2021**, *123* (July 2020).

(33) Purcaro, G.; Nasir, M.; Franchina, F. A.; Rees, C. A.; Aliyeva, M.; Daphtary, N.; Wargo, M. J.; Lundblad, L. K. A.; Hill, J. E. Breath Metabolome of Mice Infected with Pseudomonas Aeruginosa. *Metabolomics* **2019**, *15* (1), 1–10.

(34) Kalua, C. M.; Allen, M. S.; Bedgood, D. R.; Bishop, A. G.; Prenzler, P. D.; Robards, K. Olive Oil Volatile Compounds, Flavour Development and Quality: A Critical Review. *Food Chem.* **2007**, *100* (1), 273–286.

(35) Morales, M. T.; Luna, G.; Aparicio, R. Comparative Study of Virgin Olive Oil Sensory Defects. *Food Chem.* **2005**, *91* (2), 293–301.

(36) Aparicio, R.; Morales, M. T.; Aparicio-Ruiz, R.; Tena, N.; García-González, D. L. Authenticity of Olive Oil: Mapping and Comparing Official Methods and Promising Alternatives. *Food Res. Int.* **2013**, *54* (2), 2025–2038.

(37) Angerosa, F.; Camera, L.; D'Alessandro, N.; Mellerio, G. Characterization of Seven New Hydrocarbon Compounds Present in the Aroma of Virgin Olive Oils. *J. Agric. Food Chem.* **1998**, *46* (2), 648–653.

(38) Cavalli, J. F.; Fernandez, X.; Lizzani-Cuvelier, L.; Loiseau, A. M. Characterization of Volatile Compounds of French and Spanish Virgin Olive Oils by HS-SPME: Identification of Quality-Freshness Markers. *Food Chem.* **2004**, *88* (1), 151–157.

(39) Valverde-Som, L.; Ruiz-Samblás, C.; Rodríguez-García, F. P.; Cuadros-Rodríguez, L. Multivariate Approaches for Stability Control of the Olive Oil Reference Materials for Sensory Analysis - Part I: Framework and Fundamentals. *J. Sci. Food Agric.* **2018**, *98* (11), 4237–4244.

(40) Cuadros-Rodríguez, L.; Pérez-Castaño, E.; Ruiz-Samblás, C. Quality Performance Metrics in Multivariate Classification Methods for Qualitative Analysis. *TrAC - Trends in Analytical Chemistry*. Elsevier B.V. June 2016, pp 612–624.

(41) Rodionova, O. Y.; Titova, A. V.; Pomerantsev, A. L. Discriminant Analysis Is an Inappropriate Method of Authentication. *TrAC Trends Anal. Chem.* **2016**, *78*, 17–22.

(42) Angerosa, F. Influence of Volatile Compounds on Virgin Olive Oil Quality Evaluated by Analytical Approaches and Sensor Panels. *European Journal of Lipid Science and Technology*. 2002, pp 639–660.

5.7 Delineating the extra-virgin olive oil aroma blueprint by multiple headspace solid phase microextraction and differential-flow modulated comprehensive two-dimensional gas chromatography

Federico Stilo[1], Maria del Pilar Segura Borrego[2], Carlo Bicchi[1], Sonia Battaglino[1], Raquel Maria Callejón Fernadez[2], Maria Lourdes Morales[2], Stephen E. Reichenbach[3,4], James Mc Curry[5], Daniela Peroni[6], Chiara Cordero[1]*

Authors' affiliation:

1. University of Turin, Dipartimento di Scienza e Tecnologia del Farmaco, Turin, Italy

2. Área de Nutrición y Bromatología, Dpto. de Nutrición y Bromatología, Toxicología y Medicina Legal, Facultad de Farmacia, Universidad de Sevilla, Sevilla, Spain

3. Computer Science and Engineering Department, University of Nebraska – Lincoln, Lincoln, NE, USA

4. GC Image LLC, Lincoln, Nebraska, USA

5. Agilent Technologies, Gas Phase Separations Division, Wilmington DE, USA

6. SRA Intruments SpA, Cernusco sul Naviglio, Milan, Italy

*Corresponding author:

Dr. Chiara Cordero - Dipartimento di Scienza e Tecnologia del Farmaco, Università di Torino, Via Pietro Giuria 9, I-10125 Torino, Italy – e-mail: chiara.cordero@unito.it; phone: +39 011 6707172

Received: February 27, 2021

Revised: April 26, 2021

Accepted: April 30, 2021

Published: May 19, 2021

DOI: https://doi.org/10.1016/j.chroma.2021.462232

J. Agric. Food Chem. 2021, 462232

5.7.1 Abstract

Comprehensive two-dimensional gas chromatography with parallel mass spectrometry and flame ionization detection (GC×GC-MS/FID) enables effective chromatographic fingerprinting of complex samples by comprehensively mapping untargeted and targeted components. Moreover, the complementary characteristics of MS and FID open the possibility of performing multi-target quantitative profiling with great accuracy. If this synergy is applied to the complex volatile fraction of food, sample preparation is crucial and requires appropriate methodologies capable of providing true quantitative results.

In this study, untargeted/targeted (*UT*) fingerprinting of extra-virgin olive oil volatile fractions is combined with accurate quantitative profiling by multiple headspace solid phase microextraction (MHS-SPME). External calibration on fifteen pre-selected analytes and FID predicted relative response factors (RRFs) enable the accurate quantification of forty-two analytes in total, including key-aroma compounds, potent odorants, and olive oil geographical markers.

Results confirm good performances of comprehensive UT fingerprinting in developing classification models for geographical origin discrimination, while quantification by MHS-SPME provides accurate results and guarantees data referability and results transferability over years. Moreover, by this approach the extent of internal standardization procedure inaccuracy, largely adopted in food volatiles profiling, is measured. Internal standardization refers of an average relative error of 208 % for the fifteen calibrated compounds, with an overestimation of + 538% for (*E*)-2-hexenal, the most abundant yet informative volatile of olive oil, and a -89% and -80% for (*E*)-2-octenal and (*E*)-2-nonenal respectively, analytes with a lower HS distribution constant.

Compared to existing methods based on 1D-GC, the current procedure offers better separation power and chromatographic resolution that greatly improve method specificity and selectivity and results in lower LODs and LOQs, high calibration performances (*i.e.*, R^2 and residual distribution), and wider linear range of responses.

As an *artificial intelligence smelling machine*, the MHS-SPME-GC×GC-MS/FID method is here adopted to delineate extra-virgin olive oil aroma blueprints; an objective tool with great flexibility and reliability that can improve the quality and information power of each analytical run.

Key words

Extra-virgin olive oil volatiles; comprehensive two-dimensional gas chromatography; parallel detection MS/FID; predicted relative response factors; reverse-inject differential-flow modulation; quantitative analysis

5.7.2 Introduction

Comprehensive two-dimensional gas chromatography (GC×GC) coupled to mass spectrometry (MS), by combining physico-chemical discrimination of sample constituents with spectrometric diagnostic signatures, provides qualitative and quantitative information about single analytes and/or groups of analytes and is the basis for in-depth comprehensive investigations with fingerprinting and profiling.[1–3] The improved separation power of GC×GC, compared to one-dimensional (1D) GC, accompanied by logical retention patterns for chemically related compounds and specialized data processing techniques, make GC×GC-MS one of the most suitable platforms for accurate and informative investigations on complex samples. Moreover, GC×GC-MS performance, information dimensions, and flexibility are crucial to achieve reliable and consistent results when the analytical process is used to answer many different questions about functional variables related to a sample's chemical composition.[1,4]

When the fraction under study poses challenges because of the large dynamic range of concentrations covered and consists of analytes with a wide polarity range within a relatively narrow volatility interval, chromatographic resolution and efficiency are fundamental to achieve appropriate method performances. The role of sample preparation, as the zeroth dimension of the system and as a key step of the analytical method, is crucial and its design/set-up requires careful consideration of the analysis' final goal(s). If a targeted profiling of selected components is the goal, sample preparation efforts can be directed to known analytes and performance parameters optimized to achieve high specificity and selectivity, appropriate repeatability, and accuracy for those selected targets. However, if the aim of the investigation is untargeted fingerprinting of all detectable constituents, then sample preparation must be comprehensive and minimally analytes/class-specific, in order to limit discriminations and improve the breadth of the analysis.[1] Moreover, sample preparation should be able to exploit a large dynamic range of analytes' concentrations while providing solid foundations for quantitative cross-comparisons.

In the context of complex volatile fractions of food origin, GC×GC-MS has been demonstrated to be very effective for both untargeted and targeted investigations (e.g., combined untargeted/targeted fingerprinting approaches)[5,6] by combining high-throughput fingerprinting with quantitative profiling[7] in the same analysis. In these applications, the role of headspace (HS) solid phase microextraction (SPME) as the sampling approach, is central, and its main limits, related to the actual volume/amount of extracting/accumulating phase, are fully compensated by the analytical performances of GC×GC.

In this study, we make a step forward in the exploitation of the HS-SPME-GC×GC-MS potentials by designing a procedure capable of performing comprehensive chromatographic fingerprinting of the complex volatile fraction of high-quality extra-virgin olive oil (EVOO) while providing accurate quantitative data on a large list of targeted analytes (*i.e.*, targeted quantitative profiling)[1] with a high informative potential related to samples' sensory quality and qualification. Moreover, the procedure is highly automated, limiting manual operations to a few simple steps, and implemented on a robust, reliable, and commercially available analytical platform in order to be considered suitable for high-throughput screenings and quality control assessments. The differential-flow modulation technology was chosen as a core element of the GC×GC platform. Its stable performance and relative ease of use,[8–13] accompanied by the possibility of rationally translating validated methods from thermal modulated systems, [8,14,15] are key-aspects to design a method complying with minimal performance requirements in EU quality standards for analytical measurements.[16,17]

Mediterranean countries offer ideal conditions for olive tree (*Olea europaea* L.) cultivation[18] and have preserved and valorized olive oil (OO) manufacturing traditions including harvest methodologies and milling technologies.[19,20] High-quality OO, complying with EU Regulations and

International Standards for production, chemical composition, and sensory quality, are labelled as *extra-virgin* (EV)OOs (EEC No 2568/91 and its amendments; IOC/T.20/Doc. No 15/Rev. 10 2018). Within them, due to peculiar characteristics of olive cultivar(s), pedo-climatic conditions of the harvest region, and traditions related to milling technology, the EU recognizes a further additional quality through quality schemes and labels. The Protected Designation of Origin (PDO) products, for example, are those which "*have the strongest links to the place in which they are made*" and "*every part of the production, processing and preparation process must take place in the specific region*".[21]

In this context, the possibility to accurately map high-quality EVOO volatiles (*i.e.,* volatiles fingerprinting) with additional information about the concentrations/amounts of informative components, represents a step forward in the rationalization of the quality concept. To date, EVOO quality is defined by its compliance with physicochemical indices (*e.g.,* free acidity, peroxide index, UV absorbance) and by absence of sensory defects and presence of a perceivable *fruity* attribute (*i.e.*, median M>0). Nevertheless, these standards do not provide elements for valorization and discrimination of products with superior sensory quality or obtained within PDO and Protected Geographical Indication (PGI) recognized protocols. A methodology, capable of generating a sample's fingerprint with identitation potentials[22] accompanied by the accurate quantification of selected chemical markers, might fill this gap while improving the knowledge on EVOO quality signature, its *aroma blueprint,*[23,24] and its unicity across production regions and harvest years.

Within EU quality standards for analytical methodologies, specificity and selectivity are matched by a suitable combination of separation phase chemistries in the two chromatographic dimensions; sensitivity is achieved by careful tuning the columns combination dimensions (first dimension – ^{1}D and second dimension – ^{2}D lengths and diameters) and differential flows; identity confirmation is achieved by MS spectral signature matching constrained by two retention time points ($^{1}t_R$ and $^{2}t_R$); and quantitative accuracy is achieved by external calibration and data cross-validation with two parallel detectors (*i.e.*, MS and flame ionization detector FID). The latter, *i.e.* parallel detection by MS/FID, extends the method's linear range over two-to-three orders of magnitude of actual analytes' concentrations,[25] and opens to the possibility of adopting reliable response factors for quantitative estimations.[10,26]

To achieve accurate quantitative results, the multiple headspace extraction (MHE) approach is combined with the enrichment capacity of SPME with a multi-component fiber (*i.e.*, divinylbenzene/carboxen/polydimethyl siloxane) extensively adopted in EVOO volatiles' profiling.[27,28] The challenging aspect of the MHS-SPME procedure consists in the need of avoiding headspace saturation, the basis of quantification inaccuracy of many methods, while enabling multi-analyte quantification even without external calibration.

This study adds a further, advanced, and extremely flexible tool for quality control and valorization of high-quality EVOOs, acting as a bridge between 1D-GC based methods for fingerprinting [29] and/or quantitative profiling of selected key-markers[27,30] to advanced high-information methods based on GC×GC technology[5,24,31,32]. Methods performances are tested over a set of fifty EVOO samples from Italian top-quality production,[33,34] obtained from different olives cultivars and from three harvest areas. Fingerprinting potentials are explored briefly and then compared to the targeted quantitative profiling information power.

Quantification accuracy is demonstrated over a set of fifteen informative chemicals subjected to external quantification by MHS-SPME and toward an extended set including forty-two volatiles for which predicted FID relative response factors (RRFs) can be applied. Results are critically compared to existing methods based on 1D-GC-FID or 1D-GC-MS, and the advantages derived from improved information potential are discussed in the perspective of an objective qualification of high-quality products, including their sensory features, by reliable and standardized instrumental methods.

5.7.3 Materials and methods

5.7.3.1 Reference standards and solvents

Pure standards of *n*-alkanes (from *n*-C9 to *n*-C25) used for Linear Retention Indices (I^T) calibration, of α/β-thujone and 2-methyloctynoate used as Internal Standards (ISTDs), and solvents (cyclohexane and dibutyl phthalate – 99% of purity) were from Merck (Milan, Italy).

The following key-aroma compounds and potent odorants, selected according to reference literature[35–41] and adopted for external calibration, were from Sigma Aldrich (Milan, Italy): (*E*)-2-pentenal (CAS 1576-87-0), (*E*)-2-penten-1-ol (CAS 1576-96-1), (*Z*)-2-penten-1-ol (CAS 1576-95-0), 1-penten-3-ol (CAS 616-25-1), 1-pentanol (CAS 71-41-0), 2-pentanol (CAS 6032-29-7), ethyl acetate (CAS 141-78-6), (*E,E*)-2,4-hexadienal (CAS 142-83-6), (*E*)-2-hexenal (CAS 6728-26-3), (*Z*)-3-hexen-1-ol (CAS 928-96-1), hexanal (CAS 66-25-1), 1-hexanol (CAS 111-27-3), heptanal (CAS 111-70-6), (*E*)-2-octenal (CAS 2548-87-0), and (*E*)-2-nonenal (CAS 18829-56-6).

5.7.3.2 Reference solutions and calibration mixtures

Standard stock solutions of reference analytes were prepared at a concentration of 10 g/L in cyclohexane and stored at -18°C for one-week. The Reference Working Solution (RWS) was prepared by mixing suitable amounts of standard stock solutions to reach the concentration of 0.250 g/L using dibutyl phthalate as solvent. Calibration Solutions (CS) were prepared by diluting suitable amounts of RWS in dibutyl phthalate to reach the final concentrations: 30, 25, 20, 15, 10, 5, 2, 1, 0.5, 0.2 mg/L. Calibration curves were built by analyzing 5 µL of each CSs by MHS-SPME while matching absolute amounts of 150, 125, 100, 100, 75, 50, 25, 10, 5, 2.5, 1 ng. ISTDs working solution for standard-in fiber pre-loading[42] was prepared at 0.100 g/L in dibutyl phthalate and stored at -18°C in sealed vials. A schematic diagram of the operative procedure is reported in the **Supplementary material**.

5.7.3.3 Extra virgin olive oil samples

EVOO samples were collected within the VIOLIN project (Valorization of Italian OLive products through INnovative analytical tools).[34] They were obtained from olives of different cultivars harvested in 2017 over the Italian territory. Details on the sample-set under study are provided in **Table 5.7.1** together with harvest regions (*i.e.*, *Sicilia*, *Toscana,* and *Garda* lake) shown in Supplementary Material - **Supplementary Figure 5.7.1**.

Table 5.7.1. List of analyzed samples with harvest region/origin, identification codes, olives cultivar, and EU quality labelling and/or additional certifications (i.e., organic cultivation and production).

Origin	Sample ID	Cultivar	Additional qualifications
Garda	G1	*Leccino*	PDO
	G2	*Casaliva, Leccino*	PDO
	G3	*Casaliva, Leccino*	PDO
	G4	*Casaliva, Leccino*	PDO
	G5	*Casaliva, Leccino, Moraiolo, Pendolino*	PDO
	G6	*Casaliva*	PDO
	G7	*Casaliva, Frantoio, Leccino*	
	G8	*Coratina*	
	G9	*Grignano*	
	G10	*Grignano, Favarol, Pendolino, Trepp*	PDO

	G11	Blend	
	G12	*Casaliva*	Organic
	G13	*Casaliva, Frantoio, Leccino*	PDO
	G14	*Casaliva, Frantoio, Leccino*	
	G15	*Casaliva*	
Sicilia	S1	*Nocellara del Belice*	
	S2	*Nocellara del Belice*	Organic
	S3	*Cerasuola e Nocellara del Belice*	Organic
	S4	Blend	PDO
	S5	*Nocellara del Belice*	
	S6	*Tonda Iblea*	
	S7	*Cerasuola, Nocellara del Belice, Biancolilla*	PGI
	S8	*Nocellara del Belice*	PDO
	S9	*Biancolilla*	PDO
	S10	*Nocellara del Belice*	PDO
	S11	*Tonda Iblea*	PDO
	S12	*Nocellara Messinese*	
	S13	*Nocellara Etnea*	
	S14	*Nocellara Etnea*	
	S15	*Nocellara del Belice*	
	S16	*Nocellara del Belice*	
	S17	*Biancolilla*	
	S18	*Cerasuola*	
Toscana	T1	*Frantoio, Moraiolo, Maurino, Picholine*	
	T2	Blend	PGI
	T3	Blend	Organic
	T4	Blend	PDO
	T5	Blend	PGI
	T6	Blend	PGI
	T7	Blend	Organic, PGI
	T8	*Moraiolo, Frantoio, Leccino, Americano*	Organic
	T9	*Frantoio, Moraiolo, Leccino*	PGI
	T10	*Moraiolo*	PGI
	T11	Blend	
	T12	*Olivastra Saggianese*	PDO
	T13	*Olivastra Saggianese*	PDO
	T14	Blend	PDO
	T15	*Correggiolo, Leccino, Frantoio*	PDO
	T16	*Frantoio, Leccino, Moraiolo*	
	T17	*Frantoio, Leccino, Moraiolo, Pendolino*	

5.7.3.4 Multiple headspace solid phase microextraction: devices and conditions

Volatiles from EVOO samples were extracted by HS-SPME with a divinylbenzene/carboxen/polydimethyl siloxane (DVB/CAR/PDMS) fiber (d_f 50/30 μm; 2 cm length) from Supelco (Bellefonte, PA, USA) chosen based on literature on sampling effectiveness for EVOO's informative compounds.[24,32,37,43] The SPME fiber was conditioned before use as recommended by the manufacturer.

The ISTDs for response normalization and quality control were preloaded onto the SPME device[42,44] by sampling 5.0 μL of α/β-thujone and methyl 2-octynoate ISs solution (0.100 g/L) placed in a 20 mL headspace vial. ISTDs pre-loading was performed by exposing the SPME device in the HS at a temperature of 40 °C for 5 min.

Sampling was carried out on 0.100 ± 0.005 g of oil, precisely weighed, in 20 mL headspace vials, kept at 40°C for 60 minutes under constant agitation. The very low amount of sample was chosen to match HS linearity conditions for most of the characteristic analytes of the EVOO volatilome.[37,45] After extraction, the SPME device was automatically transferred to the split/splitless injection port of the GC×GC system, kept at 250 °C, and thermal desorption was for 5 min. Samples were analyzed in three replicates randomly distributed over a two-week time frame.

MHS-SPME from samples and calibration solutions was conducted by applying the above indicated conditions, and the number of consecutive extraction steps was set to five, achieving an almost exhaustive extraction for the analytes under study.[46]

5.7.3.5 GC×GC-MS/FID with reverse-inject differential flow modulation: instrument set-up and conditions

Automated MHS-SPME, as described in **Section 5.7.3.4**, was performed by a multipurpose sampler, model MPS-2 (Gerstel, Mülheim a/d Ruhr, Germany), installed on a GC×GC system equipped with a reverse-inject differential-flow modulator based on capillary flow technology™ (Agilent Technologies, Little Falls, DE, USA). The Agilent 7890B GC unit was coupled to an Agilent 5977B HES (high efficiency source) fast quadrupole MS detector (Agilent Technologies, Little Falls, DE, USA) operating in electron ionization mode at 70 eV. Ion source and transfer-line temperatures were set at 280 °C, and the quadrupole temperature was set at 240°C. The scan range was set between 45 and 240 m/z, achieving an actual data acquisition frequency of 30 Hz. Parallel detection was by a fast FID with base temperature 280 °C; H_2 flow 40 mL/min, air flow 350 mL/min, and sampling frequency 200 Hz.

The column set was configured as: ^{1}D HeavyWax™ column (100% polyethylene glycol - PEG; 20 m × 0.18 mm *dc* × 0.18 μm *df*) coupled with ^{2}D DB17 column (50% phenyl-methylpolysiloxane; 1.8 m × 0.18 mm *dc* × 0.18 μm *df*), both from Agilent Technologies (Wilmington, DE, USA). The microfluidic splitter (G3181B, Agilent, Little Falls, DE, USA). The resulting split ratio was 70:30 FID/MS. The bleeding capillary of deactivated fused silica (6.06 m × 0.1 mm *dc*) installed on the modulator plate was dimensioned according to a previously validated flow calculator.[15]

The GC split/splitless injector port was set at 250 °C and operated in pulsed-split mode (250 kPa overpressure applied to the injection port until 2 min) with a split ratio 1:20. A special design liner for SPME thermal-desorption (Merck) was used to improve analytes transfer into the ^{1}D column and limit band broadening in space. The carrier gas was helium at a nominal flow of 0.4 mL/min along the ^{1}D column and 10 mL/min along the ^{2}D column. The oven temperature program was set as: from 40 °C (2.29 min) to 240°C (11') at 3.06 °C/min and determined by method translation of a reference method previously validated in the authors' laboratory.[10,32] The modulation period (P_M) was set at 3s and pulse time at 150 ms.

The *n*-alkanes liquid sample solution for I^T determination was analyzed under the following conditions: split/splitless injector in split mode, split ratio 1:50, injector temperature 250 °C, and injection volume 1 µL.

5.7.3.6 External standard calibration by MHS-SPME-GC×GC-MS/FID

Calibration curves were built to cover analyte absolute amounts in the analyzed samples within the range 1-150 ng. External standard calibration was conducted on both MS and FID traces. For the MS detection channel, extracted target ions (Ti) were selected for each analyte; m/z values are reported in **Table 5.7.2**. Up to three additional qualifier ions (Q1, Q2, Q3) were also monitored for quality evaluation. For the FID channel, external calibration was based on 2D peak volumes; analytes affected by coelution issues were only quantified by MS. Additional details on the calibration/quantification procedure are discussed in **Section 5.7.3.2**.

Table 5.7.2 lists target analytes subjected to external calibration together with their odor quality, odor threshold (OT), experimental I^T, Ti and Qs m/z, regression functions parameters including calibration range, determination coefficients (R^2), and precision.

Table 5.7.2. List of the fifteen targeted odorants with their odor quality, odor threshold in oil as reported in literature (OT ng/g), experimental I^T, target ion used for quantification (Ti), additional qualifier ions (Q#n), calibration range covered, calibration function for MS and FID channels with the corresponding determination coefficient (R^2), intermediate precision data expressed as % relative standard deviation (RSD %) obtained by replicated quantitative measurements on a representative sample, limit of detection (LOD) and limit of quantification (LOQ).

							MHS-SPME calibration MS				MHS-SPME calibration FID					
Targeted analyte	**Odor quality**	**OT (ng/g)**	**Exp. I^T**	**Ti (m/z)**	**Q#n (m/z)**	**Range (ng)**	**Regression equation**			**Precision**	**Regression equation**			**Precision**	**LOD (ng/g)**	**LOQ (ng/g)**
							m	***q***	***R²***	**RSD%**	***m***	***q***	***R²***	**RSD%**		
Ethyl acetate	Fruity, sweet, winey	940§	865	70	88;61	1-5	0.26	0.81	0.996	9.2	0.10	0.03	1	8.9	0.47	1.57
						5-100	0.40	0.11	0.995		0.81	2.38	0.996			
Hexanal	Green-apple, grass	300$	1072	72	82;56	1-10	0.85	1.00	0.995	9.5	1.34	1.44	0.998	9.8	0.33	1.1
						10-150	0.64	4.52	0.993		0.74	7.23	0.995			
2-Pentanol	Musty, fermented	380£	1095	73	87;55;45	1-100	1.61	0.79	0.997	7	0.98	0.68	0.996	9.7	0.16	0.53
***(E)*-2-Pentenal**	Pungent, apple-like	300§	1121	84	69;55	1-5	1.07	0.11	0.995	3.7	0.34	0.16	0.996	2.9	0.47	1.57
						5-125	1.17	0.74	0.995		1.10	0.22	0.995			
1-Penten-3-ol	Pungent, butter	400§	1151	57	86;71	1-100	2.08	0.41	0.996	9.6	1.19	0.49	0.998	7.3	0.34	1.13
Heptanal	Citrus-like, fatty	500$	1179	96	81;70;55	1-125	0.99	0.95	0.999	1.9	1.18	0.86	0.997	4.9	0.22	0.73
***(E)*-2-Hexenal**	Bitter almond, green, fruity	320$	1213	83	98;69;55	1-5	0.82	0.00	0.995	6.2	0.12	0.01	1	4.3	0.32	1.07
						5-125	0.93	0.16	0.995		1.38	0.20	0.995			
1-Pentanol	Sweet, pungent	470§	1240	70	87;55	1-5	1.35	0.03	0.995	2.4	0.21	0.02	0.999	4.2	0.39	1.3
						5-125	1.16	0.20	0.997		1.25	0.29	0.996			
***(Z)*-2-Penten-1-ol**	Green, almond	250£	1306	68	86;57	1-125	1.60	0.09	0.997	2.6	1.27	0.32	0.996	6.3	0.29	0.97
***(E)*-2-Penten-1-ol**	Mushroom, earthy	250£	1314	68	86;57	1-125	1.36	0.95	0.998	1.7	1.02	0.46	0.995	2.1	0.47	1.57
1-Hexanol	Fruity, banana, soft	400§	1346	84	102;69;56	1-5	1.49	0.15	0.996	7.8	0.24	0.01	1	3.8	0.26	0.87
						5-125	2.05	2.09	0.999		1.51	0.15	0.997			

***(Z)*-(3)-Hexen-1ol**	Banana, fresh, grass	1100£	1379	67	100;82;55	1-125	1.62	0.36	0.995	2.7	1.51	1.30	0.995	3.6	0.32	1.07
***(E,E)*-2,4Hexadienal**	Green	270£	1400	81	96;67;53	1-125	1.68	3.83	0.995	7.4	2.26	1.40	0.996	5.1	0.29	0.97
***(E)*-2-Octenal**	Fatty, nutty	120$	1428	83	97;70;55	1-125	1.00	2.03	0.995	9.1	1.26	1.94	0.996	7.8	0.32	1.07
***(E)*-2-Nonenal**	Fatty, green, soapy	140$	1534	96	111;83;70	1-100	0.27	0.56	0.991	3.5	1.14	2.01	0.998	4	0.07	0.23

5.7.3.7 Data acquisition and 2D data processing

Data were acquired by Enhanced MassHunter (Agilent Technologies, Little Falls, DE, USA) and processed by GC Image® suite, Release2.9 (GC Image, LLC, Lincoln NE, USA). Statistical analysis and chemometrics were by XLSTAT statistical and data analysis solution software (Addinsoft 2020, New York, USA).

5.7.4 Results and Discussion

5.7.4.1 The information potential of olive oil volatiles patterns

EVOO volatile fractions are complex mixtures of compounds belonging to different chemical classes (*e.g.*, aldehydes, ketones, alcohols, esters, lactones, hydrocarbons, terpenes), whose relative distributions reflect concurrent functional variables strongly correlated to the above mentioned quality concepts: cultivar, geographical origin, quality (*i.e.,* EVOO, virgin olive oil – VOO, and lampante olive oil – LOO), olive ripeness, technological processes, and storage conditions.[24] Within this complex fraction, potent odorants are those components that, reaching the olfactory epithelium, dissolve into the mucus and interact with olfactory receptors, triggering olfaction and modulating the final aroma perception.[36] Potent odorants in their natural concentration in food define the odor code or aroma blueprint[23] and are at the basis of a rational sensory coding process, recently defined as "artificial intelligence smelling", of great interest for food industry and market.[47]

In this study, chromatographic fingerprinting was based on the combined untargeted/targeted (UT) fingerprinting process[5,32,45] while the accurate quantification was primarily directed to a selection of impacting odorants defining high-quality EVOO aroma. Contributing to *fresh-green* and *fruity* notes, positive attributes[20] in EVOOs, are C5 and C6 oxygenated compounds (*i.e.*, alcohols and carbonyls). In particular, (*E*)-2-pentenal, (*Z*)-2-penten-1-ol, 1-penten-3-ol, 1-pentanol, (*E,E*)-2,4-hexadienal, (*E*)-2-hexenal, (*Z*)-3-hexen-1-ol, hexanal, and 1-hexanol were targeted and subjected to quantitative profiling by external standard calibration. They belong to the so-called lipoxygenase (LOX) signature[36–40] and are formed by enzymatic oxidation of linolenic and linoleic acids through the LOX pathway.[40]

Compounds responsible for sensorial defects also were included, although they were expected to be at very low concentrations, if not below the method's limit of detection (LOD), due to the quality level and freshness of selected samples. This group included some C_7-C_{10} linear saturated and unsaturated aldehydes deriving from fatty acids hydroperoxides cleavage: heptanal, (*E*)-2-octenal, and (*E*)-2-nonenal. They provide information on the autoxidation process and shelf-life evolution, and their increasing concentration is correlated to the perception of *rancid* and *fatty* notes.[35,37,38]

The targeted compounds list was completed by analytes correlated to well-known sensory defects.[36,38,39] (*E*)-2-penten-1-ol, with *mushroom-like* and *earthy* sensations; 2-pentanol, related to *musty* and *fermented* perception; and ethyl acetate, responsible for the *winey* note.[36,38,39] **Table 5.7.2** lists targeted analytes together with their odor quality, odor threshold (OT) in oil as reported in literature, experimentally determined I^T, and additional information related to the quantification procedure. Details are discussed in the next sections.

Figure 5.7.1 shows pseudocolor images of volatile patterns from an EVOO sample analyzed by a polar × semi-polar column combination (*i.e.*, PEG × OV1701) and resulting from the parallel detection by MS (**Figure 5.7.1A**) and FID (**Figure 5.7.1B**). Enlarged areas of the chromatographic plane highlight the complexity of the patterns, the chromatographic resolution of analyte clusters obtained by combining the two separation dimensions, and the wide range of response variations (perceivable by colorization) spanned by detected volatiles in the two detection channels. The list of targeted analytes, including those that were not subjected to quantitative assessment, is provided

in Supplementary Material - **Supplementary Table 5.7.1**, together with their retention times in the two chromatographic dimensions (^1t_R - min, ^2t_R - sec), experimental I^T, odor descriptor, and OT (ng/g) Supplementary Material - **Supplementary Table 5.7.2** lists untargeted and targeted peak features, identified by unique labelling on both detection channels (*i.e.*, #ID and chemical names), together with retention times in the two chromatographic dimensions (^1t_R - min, ^2t_R - sec), experimental I^T, and raw MS spectra in tabular form.

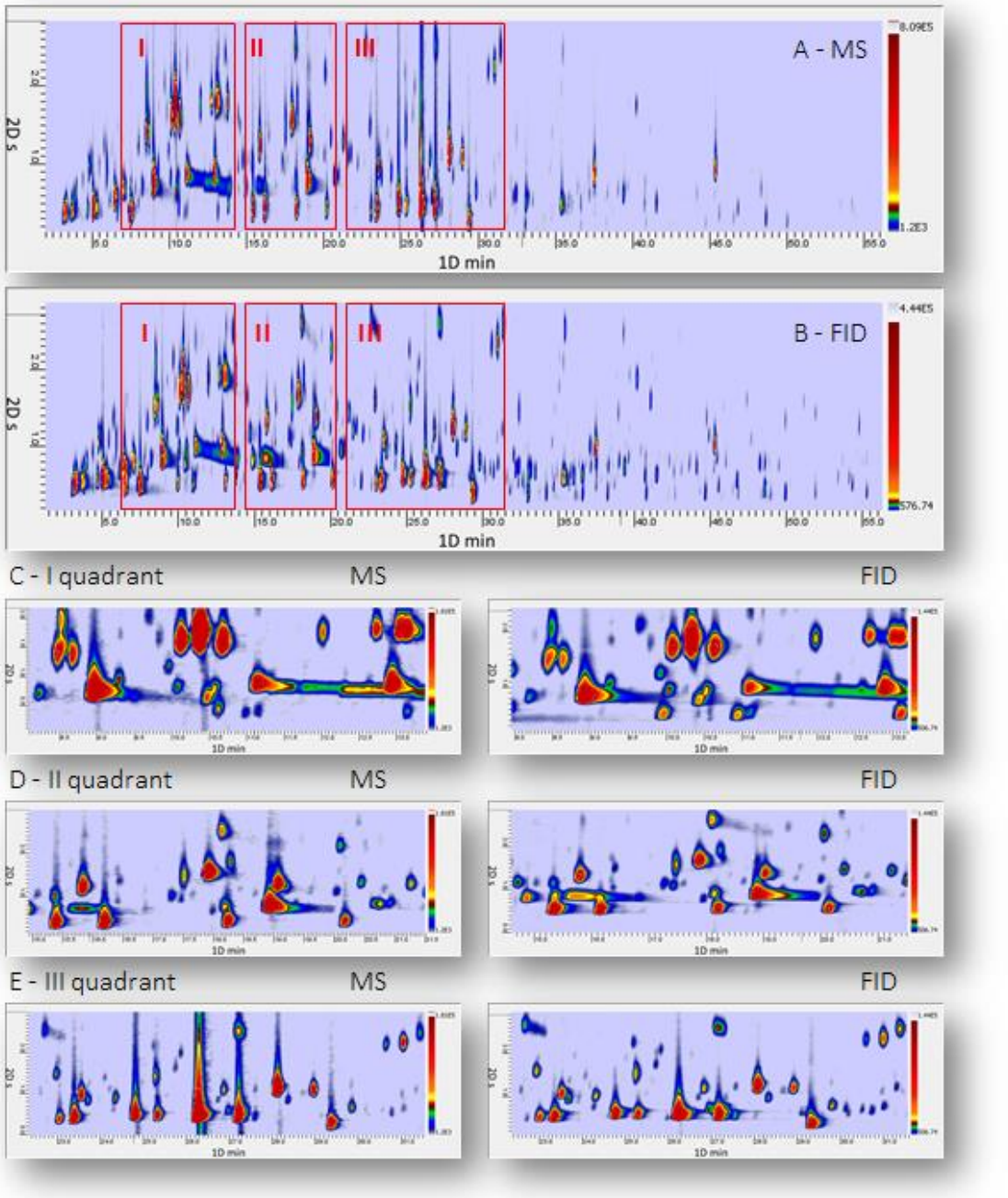

Figure 5.7.1. Pseudocolor images of the volatiles pattern of S#1 EVOO sample from *Sicilia* analyzed by a polar × semi-polar column combination (*i.e.*, PEG × OV1701) and resulting from the parallel detection by MS (**5.7.1A**) and FID (**5.7.1B**). Enlarged areas of the chromatographic plane (red rectangles in quadrant *I* - **5.7.1C**, *II* - **5.7.1D**, and *III* – **5.7.1E**) highlights the complexity of the pattern and the chromatographic resolution of analytes clusters.

5.7.4.2 Accuracy of multiple headspace extraction vs. internal standardization for selected potent odorants

Quantitative analysis is one of the most challenging aspects related with HS sampling, especially when carried out through high concentration capacity techniques based on accumulating polymers/materials, as for SPME. The main issues related to erroneous results are standardization and/or normalization of accumulating polymer(s) performance(s) and accuracy of the selected quantification approach.[48]

Among the most common approaches, those based on internal standard(s) (ISs) require careful consideration about the quality of information that they provide. ISs methods are fast and simple and take into account and compensates for detector response variations and sampling variability when applied to liquid samples/extracts, but cannot be considered accurate [17] if analytes subjected to quantification have different HS partition constants (K_{HS}) and accumulating polymer/material distribution constants (K_D) under the applied sampling conditions. Moreover, when MS is adopted, the differential fragmentation rate and/or relative response of selected *Ti* might add a further quantification error related to the specific response factor of the analytes *vs.* that of ISs.[49]

To overcome inaccuracy issues related to the lack of an appropriate modelling of the actual phenomena occurring in the HS sampling, external calibration is compulsory although in the case of HS methods, the matrix effect would require its implementation in the form of standard addition (SA) procedure.[50] Moreover, SA provides accurate results if HS sampling operates in equilibrium conditions and/or if analytes HS saturation does not occur.[37,45,50,51]

To achieve accurate quantification of multiple volatile targets in a single run, by operating in a wide dynamic range of analytes concentration, adopting non-equilibrium sampling with high efficient yet standard multi-component SPME devices, MHE method is the elective route.[49,52] MHE was introduced by Suzuki *et al.*[53] and McAuliffe[54] for static HS sampling to measure the area representative of the total amount of the target analyte in the investigated matrix. It was further developed and modeled by Kolb and Ettre[52] and then successfully extended to HS-SPME sampling.[48,50,55–58] Principles of operation and fundamental equations are well detailed in reference literature[48,50,55–58] and briefly summarized in the **Supplementary material**.

Table 5.7.2 reports, for the fifteen target odorants, information about calibration ranges (expressed in ng, as absolute amount of analyte in the HS), calibration functions for MS and FID channels accompanied by determination coefficients (R^2), and intermediate precision expressed as the characteristic % relative standard deviation (RSD%) obtained by replicated quantitative measurements of a representative sample over two-weeks. *(E)*-2-pentenal, 1-pentanol, ethyl acetate, *(E)*-2-hexenal, hexanal, and 1-hexanol required a two-step calibration to match for linearity on MS channel. Limit of detection (LOD) and limit of quantitation (LOD) also are reported for the FID channel. They were estimated according to EU protocol for food safety analytical methods performance parameters[59] by extrapolation from calibration curves.

Due to the lack of certified reference materials (CRM), quantification accuracy was defined through recovery % and precision[17]. Recovery was determined by multiple extraction (until exhaustive extraction) of CSs (i.e., blank matrix spiked with known amounts of target analytes) and of selected oil samples; quantification error (i.e., inaccuracy or bias) was then expressed as % relative error (%RE). Moreover, recovery data were validated by internal cross-matching between MS and FID values. Results on accuracy are visualized with histograms in **Figure 5.7.2A** with dark blue and orange bars accompanied by imprecision intervals. Amounts correspond to the average values of replicated analyses (n=3) of the 25 ng CS experiment. Expected values (*i.e.*, 25 ng) are marked with a green line while boundary, the light green bar, highlights a ± 20% error tolerance.[17]

Of relevance are results on quantification error occurring by applying the ISs approach (see Supplementary Material - **Supplementary Equation 5.7.1**) with a relative response factor RFF=1 (light blue and orange bars in **Figure 5.7.2A**).

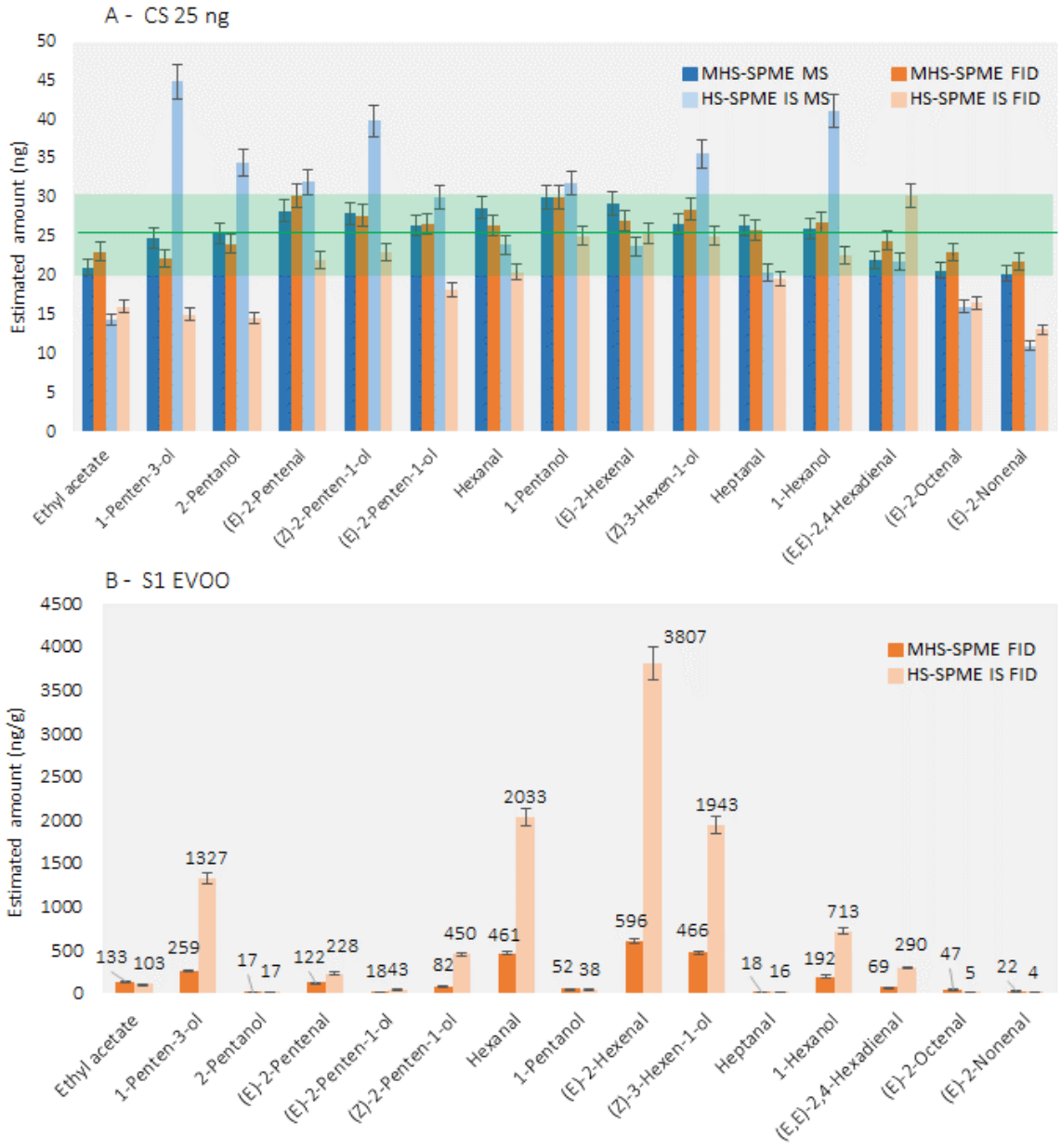

Figure 5.7.2: Histograms illustrating accuracy of MHS-SPME *vs.* HS-SPME with IS normalization on (**5.7.2A**) the 25 ng CS (blue and orange bars). Expected values (*i.e.*, 25 ng) are marked with a green line while boundary highlights a ± 20% error tolerance. In **5.7.2B**, accuracy is shown for a EVOO sample. Results correspond to the average values of replicated analyses (n=3).

The MHS-SPME method results can be considered accurate according to EU quality standards for analytical methods,[17] this is true for both MS and FID channels, respectively dark blue and orange bars in **Figure 5.7.2A**. Almost all analytes were quantified with a %RE lower than 20. For the MS channel, the error is 11.7% with a maximum of 25.8% for (*E*)-2-nonenal, whereas for the FID, it achieves 9.5% with a maximum of 21.1% resulting from the quantification of (*E*)-2-

pentenal. %RE increases up to 80 when the IS method is applied. In particular, for the MS detection channel, the %RE is 35.6, with eleven of fifteen compounds quantified having an error greater than 20% and a maximum value of 79.3% for 1-penten-3-ol. Accuracy was slightly better for the FID channel with an average error of 21.1%, with eight of fifteen compounds quantified with %RE greater than 20 and a maximum of 47.7 for quantification of (*E*)-2-nonenal.

Of interest are quantification results estimated on the test sample (*i.e.*, EVOO S1 from *Sicilia*) and shown in histogram of **Figure 5.7.2B.** Here the %RE between the MHS-SPME approach (taken as reference for comparison) and HS-SPME quantification based on the IS was calculated for the FID channel. The average %RE for the IS approach was 208% with a maximum of +538% for (*E*)-2-hexenal. This analyte is the most abundant in the EVOO volatile fraction and is highly informative, being a key-aroma compound validated by sensomics.[35].Its overestimation might lead to erroneous conclusions about a sample's aroma blueprint. Similar results were obtained for other highly abundant components, such as 1-penten-3-ol, hexanal, (Z)-3-hexen-1-ol, and 1-hexanol, all characterized by steeper decay curves, *i.e.*, lower β values (further details are commented below).

On the other hand, for compounds present in lower amounts and characterized by relatively flat decay curves and higher β values, as in the case of (*E*)-2-octenal and (*E*)-2-nonenal, underestimation was respectively -89% and -80%. Supplementary material - **Supplementary Table 5.7.3** reports accuracy (*i.e.,* recovery and precision) results for analytes subjected to external calibration by MHS-SPME on selected CSs (1-10-25-100 ng) and on test samples S1, S8, T3, T10, G6, and G12, where exhaustive extractions were conducted until LOD.

These results provide proof of evidence on how inaccurate might be quantification that approximates detector relative response factors to unity (*i.e.*, by using a single IS for quantification) and does not consider analytes' specific K_{HS} and K_D under the applied conditions. Moreover, the adoption of non-equilibrium HS sampling – at least not for all targeted analytes – and the multi component characteristics of the SPME device (e.g., combining adsorption and absorption mechanisms) add further variables that make complex and challenging EVOO volatiles quantification.

5.7.4.3 Extending quantification to non-calibrated analytes by adopting FID predicted relative response factors

EVOO volatile fractions are highly complex, and the presence of a large number of informative components requires greater efforts to extend external calibrations to all analytes matching with HS linearity in the defined conditions. However, the possibility to extend the quantification potential of the analytical method to a large set of volatiles while keeping an appropriate accuracy is attractive. If achieved, batch-to-batch data transferability is possible, even for long-time frame studies and for different GC×GC technologies.

With parallel MS and FID detection, additional options for reliable quantification are available. In particular, while MS detection achieves high specificity by selecting specific ion traces in cases of co-elution and/or the presence of interferents,[60] FID provides stable structure-specific response factors in a wider range of concentrations.[51,61]

The adopted platform, combining high-resolution and efficient separation of analytes by GC×GC with parallel detection by MS/FID, opens the possibility of adopting FID predicted RRFs based on combustion enthalpies and molecular structure.[61] The alignment of the separation patterns from MS and FID detection, as illustrated in **Figure 5.7.1**, enables reliable analyte identification using multiple criteria (retention in two dimensions and MS spectral signature) and across the parallel channels, while quantification can be extended over uncalibrated compounds by estimating their FID RRFs with **Equation 5.7.1**.

Equation 5.7.1 $$\mathrm{RRF} = 10^3 * \left(\frac{MW_i}{MW_{IS}}\right) * (-61.3 + 88.8\eta_C + 18.7\eta_H - 41.3\eta_O + 6.4\eta_N + 64.0\eta_S - 20.2\eta_F - 23.5\eta_{Cl} - 10.2\eta_{Br} - 1.75\eta_I + 127\eta_{benz})^{-1}$$

where η_C, η_H, η_O, η_N, η_S, η_F, η_{Cl}, η_{Br}, η_I, and η_{benz} corresponds to the number of carbon, hydrogen, oxygen, nitrogen, sulfur, fluorine, chlorine, bromine, and iodine atoms, and the number of benzene rings; and MW_i and MW_{IS} are the molecular weights of the analyte *i* and the *IS* adopted for the development of the model by de Saint Laumer *et al.*[61]

In this study, quantification was obtained on the peak area of each component normalized versus α-thujone as IS and corrected with its RRF calculated with Eq.1 to align combustion enthalpies normalized versus 1-hexanol here considered as internal reference. The analyte-specific RRF was corrected to the 1-hexanol/methyl octanoate ratio (*i.e.,* $RRF_{i,1\text{-hexanol}} = 0.933/RRF_{i,\ \text{methyl octanoate}}$) to adapt the model to 1-hexanol.

With predicted RRFs, quantitative profiling was extended to additional potent odorants, characterizing both positive attributes and specific sensorial defects, and to phenotypic markers. **Table 5.7.3** reports the RRF values calculated for targeted/calibrated analytes and for additional compounds of interest selected among those listed in Supplementary Material - **Supplementary Table 5.7.1**. They are: 3-pentanone, 1-penten-3-one, α-pinene, camphene, β-pinene, δ-3-carene, limonene, eucalyptol, hexyl acetate, octanal, *(Z)*-3-hexenyl acetate, *(E)*-2-heptenal, 6-methyl-5-hepten-2-one, nonanal, *(E)*-2-hexen-1-ol, acetic acid, *(E,E)*-2,4-heptadienal, α-copaene, benzaldehyde, propanoic acid, 1-octanol, butanoic acid, butyrolactone, γ-hexalactone, propanal, (*Z*)-3-hexenal, and α-farnesene.

Table 5.7.3. Extended list of targeted analytes, reported with their experimental I^T, molecular weight (MW) and molecular formula. Relative Response Factors (RRF) are calculated based on Equation 1 and are normalized over 1-hexanol taken as internal reference. β values (± SD) are calculated on the entire sample set while accuracy data is reported as Relative Error (RE%) on calibration solutions at 25 ng (CS25) by MHS-SPME and between MHS-SPME and HS-SPME with IS normalization on FID signals.

				Predicted Relative Response Factor data						MHS-SPME vs. HS-SPME accuracy		
Targeted analytes	**Exp. I^T**	**MW**	**Formula**	**η_C**	**η_H**	**η_O**	**η_{Benz}**	**RRF**	**β (±SD)**	**Relative Error % - CS25 MHS-SPME FID *vs.* MS**	**Amount µg/kg MHS-SPME**	**Amount µg/kg HS-SPME IS**
Propanal	802	58.1	C_3H_6O	3	6	1	0	1.92	0.54 (±0.07)	-	86	134
Ethyl acetate	865	88.1	$C_4H_8O_2$	4	8	2	0	2.23	0.87 (±0.08)	9	133	103
3-Pentanone	956	86.1	$C_5H_{10}O$	5	10	1	0	1.49	0.52 (±0.05)	-	1775	741
1-Penten-3-one	1007	84.1	C_5H_8O	5	8	1	0	1.56	0.51 (±0.05)	-	50	182
α-Pinene	1026	136.2	$C_{10}H_{16}$	10	16	0	0	1.11	0.82 (±0.07)	-	910	88
Hexanal	1072	100.2	$C_6H_{12}O$	6	12	1	0	1.40	0.61 (±0.03)	8	461	2083
Camphene	1078	136.2	$C_{10}H_{16}$	10	16	0	0	1.11	0.83 (±0.05)	-	12	3
2-Pentanol	1107	88.2	$C_5H_{12}O$	5	12	1	0	1.42	0.77 (±0.11)	6	17	17
***(E)*-2-Pentenal**	1121	84.1	C5H8O	5	8	1	0	1.56	0.53 (±0.02)	6	122	228
β-Pinene	1129	136.2	$C_{10}H_{16}$	10	16	0	0	1.11	0.92 (±0.06)	-	29	12
(Z)-3-Hexenal	1132	98.1	$C_6H_{10}O$	6	10	1	0	1.45	0.74 (±0.03)	-	44	26
δ-3-Carene	1150	136.2	$C_{10}H_{16}$	10	16	0	0	1.11	0.90 (±0.04)	-	2	7
1-Penten-3-ol	1151	86.1	$C_5H_{10}O$	5	10	1	0	1.49	0.62 (±0.02)	11	259	1327
Heptanal	1179	114.2	$C_7H_{14}O$	7	14	1	0	1.34	0.82 (±0.05)	3	18	16
Limonene	1190	136.2	$C_{10}H_{16}$	10	16	0	0	1.11	0.92 (±0.06)	-	125	53
Eucalyptol	1205	154.3	$C_{10}H_{18}O$	10	18	1	0	1.26	0.94 (±0.03)	-	596	3807
***(E)*-2-Hexenal**	1213	98.1	$C_6H_{10}O$	6	10	1	0	1.45	0.63 (±0.02)	8	1981	3814
1-Pentanol	1240	88.2	$C_5H_{12}O$	5	12	1	0	1.42	0.71 (±0.03)	3	52	38
Hexyl acetate	1271	144.2	$C_8H_{16}O_2$	8	16	2	0	1.52	0.88 (±0.08)	-	20	155

Octanal	1285	128.2	$C_8H_{16}O$	8	16	1	0	1.29	0.94 (±0.04)	-	10	17
***(Z)*-2-Penten-1-ol**	1306	86.1	$C_5H_{10}O$	5	10	1	0	1.49	0.61 (±0.02)	1	82	450
***(E)*-2-Penten-1-ol**	1314	86.1	$C_5H_{10}O$	5	10	1	0	1.49	0.65 (±0.03)	1	18	43
***(Z)*-3-Hexenyl acetate**	1317	142.2	$C_8H_{14}O_2$	8	14	2	0	1.57	0.89 (±0.04)	-	119	1245
***(E)*-2-Heptenal**	1321	112.2	$C_7H_{12}O$	7	12	1	0	1.38	0.87 (±0.05)	-	14	39
6-Methyl-5-hepten-2-one	1327	126.2	$C_8H_{14}O$	8	14	1	0	1.33	0.91 (±0.03)	-	24	25
1-Hexanol	1346	102.2	$C_6H_{14}O$	6	14	1	0	1.35	0.79 (±0.02)	3	192	713
***(Z)*-(3)-Hexen-1-ol**	1379	100.2	$C_6H_{12}O$	6	12	1	0	1.40	0.75 (±0.01)	7	466	1943
Nonanal	1393	142.2	$C_9H_{18}O$	9	18	1	0	1.26	0.95 (±0.02)	-	27	58
***(E,E)*-2,4-Hexadienal**	1400	96.1	C_6H_8O	6	8	1	0	1.51	0.81 (±0.12)	10	69	290
***(E)*-2-Hexen-1-ol**	1401	100.2	$C_6H_{12}O$	6	12	1	0	1.40	0.83 (±0.07)	-	824	776
Acetic acid	1409	60.1	$C_2H_4O_2$	2	4	2	0	5.06	0.78 (±0.09)	-	426	140
***(E)*-2-Octenal**	1428	126.2	$C_8H_{14}O$	8	14	1	0	1.33	0.95 (±0.04)	10	47	5
***(E,E)*-2,4-Heptadienal**	1464	110.2	$C_7H_{10}O$	7	10	1	0	1.42	0.86 (±0.05)	-	24	29
α-Copaene	1491	204.4	$C_{15}H_{24}$	15	24	0	0	1.09	0.77 (±0.06)	-	53	119
Benzaldehyde	1524	106.1	C_7H_6O	7	6	1	1	1.28	0.91 (±0.04)	-	13	31
***(E)*-2-Nonenal**	1534	140.2	$C_9H_{16}O$	9	16	1	0	1.29	0.95 (±0.02)	15	22	4
Propanoic acid	1539	74.1	$C_3H_6O_2$	3	6	2	0	2.88	0.86 (±0.05)	-	54	17
1-Octanol	1551	130.2	$C_8H_{18}O$	8	18	1	0	1.26	0.95 (±0.03)	-	7	17
Butanoic acid	1578	88.1	$C_4H_8O_2$	4	8	2	0	2.23	0.93 (±0.05)	-	9	3
Butyrolactone	1613	86.1	$C_4H_6O_2$	4	6	2	0	2.43	0.92 (±0.04)	-	37	27
γ-Hexalactone	1670	114.1	$C_{10}H_{18}O_2$	10	18	2	0	0.96	0.94 (±0.03)	-	12	2
α-Farnesene	1753	204.4	$C_{15}H_{24}$	15	24	0	0	1.09	0.78 (±0.03)	-	7	17

The list includes additional compounds formed along the LOX pathway and correlated to *green* and *fruity* notes: C5 ketones (*i.e.*, 3-pentanone and 1-penten-3-one), C6 esters (*i.e.*, hexyl acetate, (*Z*)-3-hexenal, and *(Z)*-3-hexenyl acetate), and C6 alcohols (*i.e.*, *(E)*-2-hexen-1-ol).[24,38,62–64]

Limonene, α-pinene, β-pinene, camphene, δ-3-carene, eucalyptol, α-copaene, and α-farnesene are terpenoids mainly known to be genetic and/or geographic markers,[65] nevertheless, they are odor active and could be associated with the perception of *herbal*, *pine,* and *citrus-like* notes.[64,66] 6-Methyl-5-hepten-2-one is a ripening indicator with *fruity-like* odor and propanal is present in high-quality EVOOs and described as *fresh, fruity* and *malty*.[35,45,67,68]

Compounds related to sensorial defects include butyrolactone and ɣ-hexalactone, described as contributing to the definition of the undesirable *oily* aroma,[63] acetic acid, propanoic acid, and butanoic acid, associated with *vinegary*, *musty* and *rancid* defects, [38,62] 1-octanol, usually found in oxidized samples[19,39] and with a *mushroom-like* odor; benzaldehyde, whose presence is related to *moldy* and *earthy* defects.[38] Finally octanal, nonanal, *(E)*-2-heptenal, and *(E,E)*-2,4-heptadienal contribute to *rancid* and *fatty* sensations.

Quantification results for the 42 targeted compounds are reported in Supplementary Material - **Supplementary Table 5.7.4**. The information they bring to the sample discrimination is considered in the next section. Italian EVOOs aroma blueprint is defined by considering key-aroma compounds concentration over the odor detection threshold (*i.e.*, odor activity value OAV).

5.7.4.4 Differential information encrypted in volatiles patterns: response vs. quantitative data

Volatiles fingerprints captured by mapping the responses of both untargeted and targeted features and 2D pattern information deriving from selected target volatiles bring differential information. The regional influence on the volatiles fraction, for example, can be examined by supervised chemometrics. Partial least square discriminant analysis (PLS-DA) was adopted to develop classification models based on the production region. Results are greatly illustrative on the information encrypted on global fingerprint and/or on 2D patterns with selected targets.

Figure 5.7.3A shows the score plot resulting from the PLS-DA model based on the global fingerprint information (i.e., *UT* fingerprinting process) and a cross-validation by *leave-n-out* approach (CV=4). The model was developed on the entire sample set (n=50) and resulted in an average sensitivity of 92.9% and a specificity of 97.4%. In particular, higher sensitivity was obtained for *Sicilia* EVOOs (100%) and lower sensitivity for *Toscana* samples (84.2%), with three EVOOs from *Toscana* classified as *Garda* and one *Garda* sample misclassified as *Sicilia*. Results confirm the effectiveness of the comprehensive volatiles fingerprinting for geographical classification of EVOOs.[29,65].By examining the variables importance in the projection (VIP) score, which summarizes the overall contribution of each variable to the PLS-DA model, components with a higher informative role in describing the characteristic regional signatures were selected. Those having VIP value ± SD higher than 1, were 27, of them 11 untargeted and 16 targeted. Within the identified analytes, those with the highest VIP ranking were: (*E*)-2-hexenal, phenylethyl alcohol, ethyl acetate, 1-penten-3-ol, butyl butanoate, 2-buten-1-ol, 2-ethylfuran, 1-pentanol, tridecane, (Z)-2-hexenal, (Z)-3-hexenol, butyrolactone, δ-3-carene, propanal, 3-pentanone, and α-farnesene.

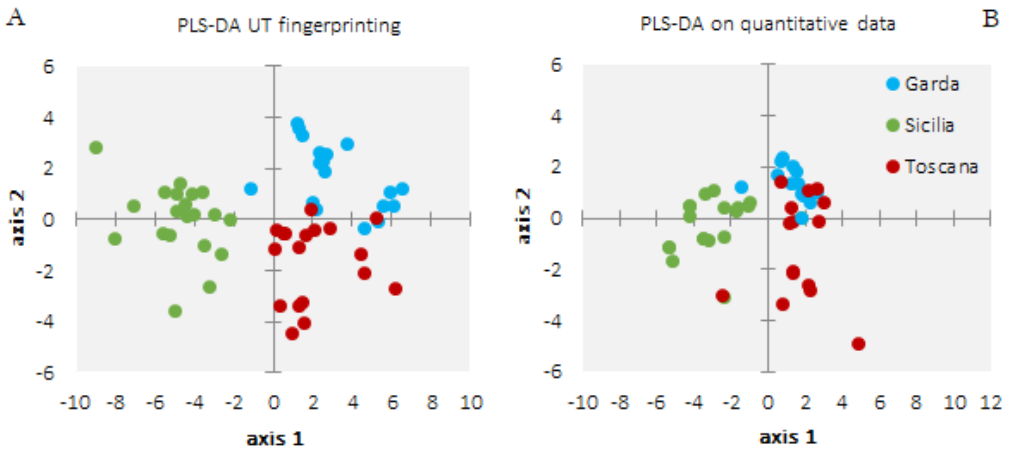

Figure 5.7.3. PLS-DA score plots resulting from the model for regional discrimination of samples and based on the UT peaks responses (**5.7.3A**) or quantitated target compounds from **Table 5.7.3** (**5.7.3B**).

Figure 5.7.3B shows the score plot resulting from the PLS-DA based on the 42 quantified compounds. Results of a full cross-validation (CV=4) performed on the entire sample set, gave an average sensitivity of 73.3% and specificity of 87.1%. Nearly all EVOOs from *Sicilia* were correctly classified (one misclassified as *Toscana*), whereas the number of misclassified samples between *Toscana* and *Garda* regions increased compared to the *UT* fingerprinting model. For targeted profiling data, the list of compounds selected by VIPs scores, in decreasing order of relevance for regional classification, were: (*E*)-2-hexenal, 1-penten-3-ol, 3-pentanone, 1-pentanol, (*Z*)-3-hexen-1-ol, α-copaene, ethyl acetate, and (*Z*)-3-hexenyl acetate. Score plot and performance parameters clearly show that the discrimination capacity of the model based on the quantified compounds was markedly lower. Analytes' selection was in fact driven by their role in defining positive or negative attributes to the overall aroma and on analytical variables not necessarily related to the regional influence on volatiles expression (*i.e.,* absence of HS saturation effect ($\beta < 0.95$) and co-elutions). However, results obtained by accurate MHS-SPME quantification are likely more robust, referable, and transferable in time and between different platforms.

Boxplots in **Figure 5.7.4** show the quantitative distribution of selected EVOO volatiles from the three production regions. Interestingly, discriminant compounds mostly belong to the LOX signature: (*E*)-2-hexenal is significantly higher in EVOOs from *Toscana* and *Garda* compared to *Sicilia*; (*Z*)-3-hexen-1-ol and 3-pentanone are more abundant in *Sicilia* samples, and hexyl acetate shows a higher amount in *Garda* and *Sicilia* compared to *Toscana* samples. The sesquiterpenoid α-copaene is more abundant in EVOOs from *Sicilia* and 1-pentanol is on average more concentrated in *Sicilia* samples.

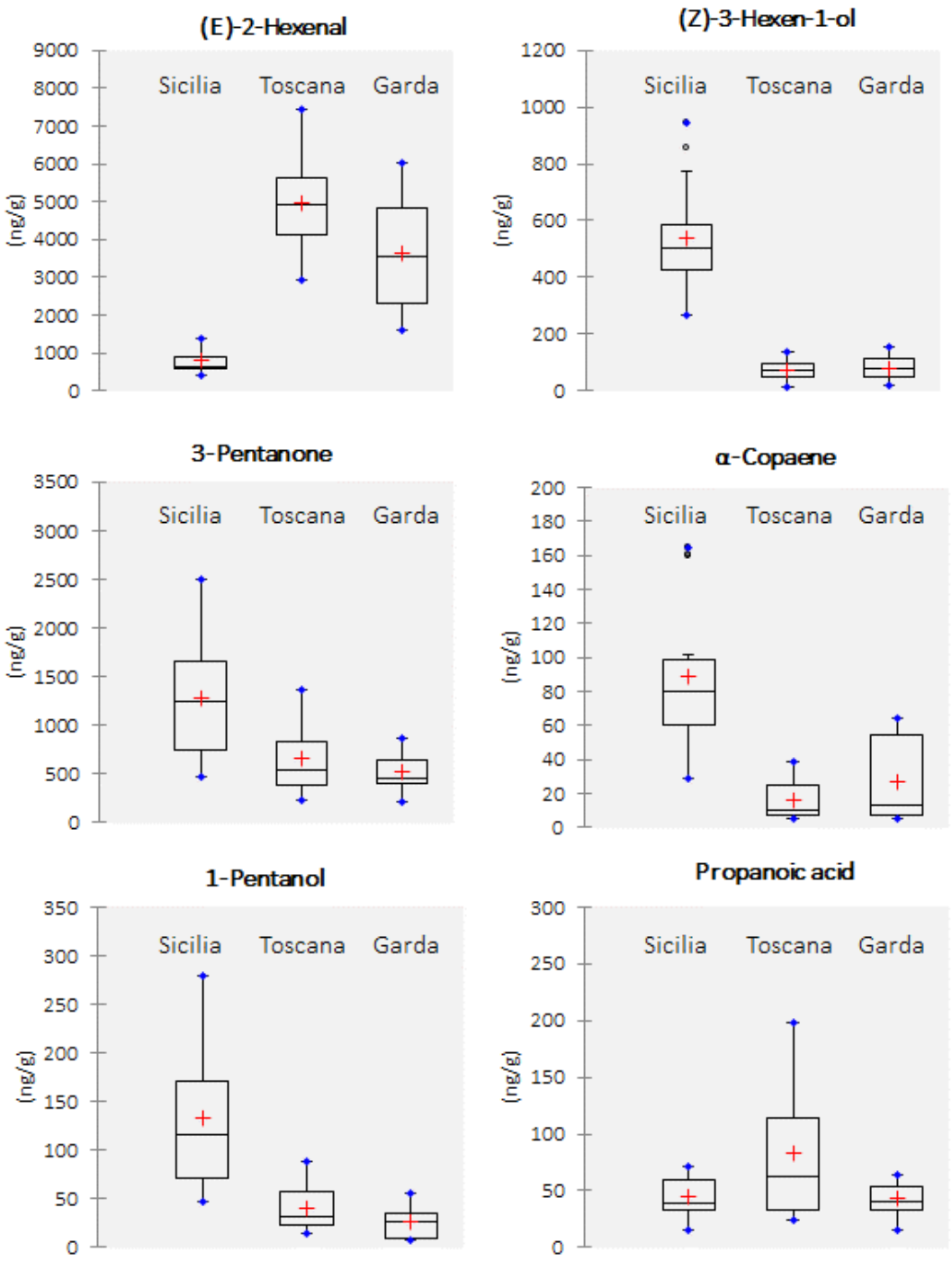

Figure 5.7.4. Boxplots illustrating the median, minimum and maximum and mean quantitative values (red mark) for discriminant targeted compounds.

5.7.4.5 Italian high-quality EVOO aroma blueprint

To develop a realistic picture about the sensory contribution of quantified analytes on the overall EVOO aroma, odor activity values (OAVs) might be useful. By molecular sensory science principles (*i.e.*, sensomics),[23] odorants having an OAV >1 and with positive validation in aroma recombinates, are considered key-aromas.[51] Although this simplification could lead to a misleading

interpretation of the actual role played by odor-active compounds in modulating the aroma perception of a food, the concept has been validated over hundreds of food products.[23] OAV is a robust yet objective tool to identify character odorants and to rank them over the multitude of volatiles characterizing a given product.

The relevance of quantitated analytes in terms of their sensory contribution to the aroma perception of EVOO samples is illustrated in spider-diagrams of **Figure 5.7.5.** Median values of OAVs for all potent odorants, *i.e.*, those reporting an OAV > 1 in at least one sample, are visualized for each production region (**Figure 5.7.5A** *Garda*, **Figure 5.7.5B** *Toscana*, and **Figure 5.7.5C** *Sicilia*) and in decreasing order of OAVs taking the *Garda* as reference. Analytes are accompanied by their odor descriptors. Pictograms clearly show differences in the aroma blueprints of EVOOs from these different regions.

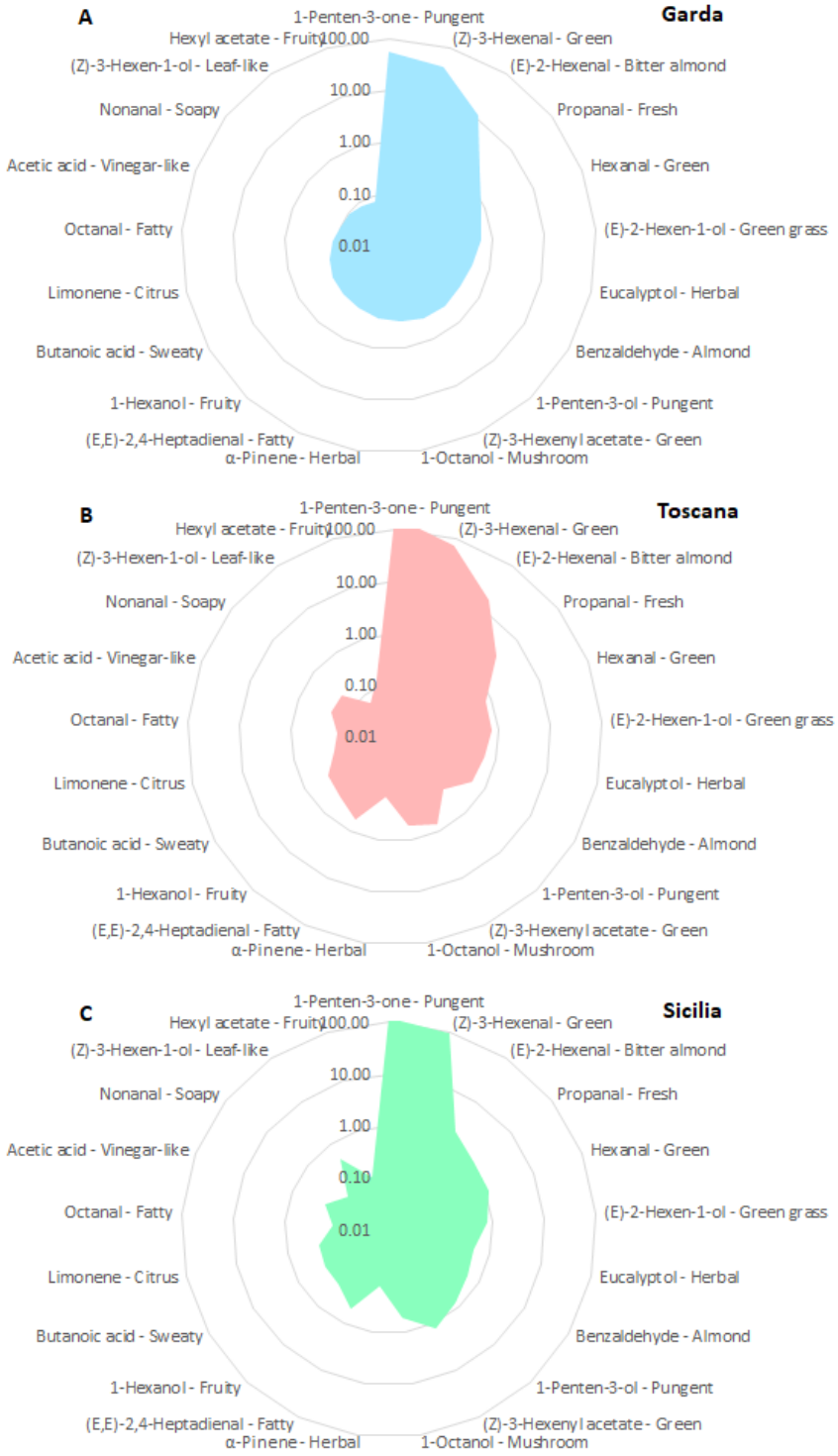

Figure 5.7.5. Spider-diagrams showing the EVVOs aroma blueprint and based on calculated OAVs (reported in logarithmic scale). Key-aroma compounds are reported in descending order taking the Garda lake group as reference. Odor qualities are also indicated.

In all regions, compounds belonging to the LOX signature and related to positive attributes as *pungent, green* and *grassy* notes, dominate the overall aroma of EVOOs, even if in different proportions; *e.g.*, (E)-2-hexenal has higher OAV in *Garda* and *Toscana* than in *Sicilia*, whereas (Z)-3-

hexenyl acetate has a higher impact on *Toscana* and *Sicilia* EVOOs aroma).[40,69] Compounds related to *fruity* notes, *e.g.* (Z)-3-hexen-1-ol, might evoke the perception of *banana-like* and *grassy* qualities, documented in EVOOs from *Sicilia.*[69] 1-Hexanol with *fruity* and *banana-like* perception has higher relevance in *Garda* and *Toscana* samples.

The important role of terpenoids, highlighted by profiling data as regional markers, is here seen in the OAVs for limonene and eucalyptol, likely contributing to the definition of *citrus, herbal, terpenic* qualities. Moreover, several EVOOs from the Garda lake area show α-pinene with an OAV > 1, likely eliciting *herbal* and *woody* notes.

Interestingly, among the 42 quantified analytes, only four are correlated to sensorial defects (*i.e.*, acetic acid, (*E,E*)-2,4-heptadienal, (*E*)-2-nonenal, and nonanal)[24,35,36,39] and only in a few samples is their OAV > 1. This result was expected in that the analyzed EVOOs had sensory classifications with the median of the coded defects equal to zero.

5.7.5 Conclusions

The possibility of performing highly informative chromatographic fingerprinting, extended to both untargeted and targeted volatile components, accompanied by accurate multi-target quantification performed on two parallel detection channels (MS/FID), has been demonstrated in the challenging context of high-quality EVOO valorization and characterization. Compared to existing methods based on 1D-GC-FID or 1D-GC-MS, the current procedure offers higher separation power and chromatographic resolution that improve method specificity and selectivity. Higher chromatographic effectiveness results in lower LODs and LOQs[27] and high quality calibration (*i.e.*, R^2 and residual distribution)[27] provides accurate quantification in a wider range of concentrations. Moreover, parallel detection by MS/FID offers the possibility of cross-validation of quantitative results and the option of adopting FID RRFs to extend quantification to uncalibrated analytes.

The crucial yet fundamental role of an effective, fully automated, and highly repeatable sample preparation approach, *i.e.*, MHS-SPME, is further highlighted even in a "true" quantification approach by evidencing the inaccuracy of semi-quantitative methodologies (*e.g.*, HS-SPME and IS normalization). By exploring the possibility of adopting MHS-SPME-GC×GC-MS/FID as artificial intelligence smelling machine[47] for EVOOs, it has been demonstrated how the great flexibility and reliability of multidimensional analytical techniques implemented with commercially available instrumentation can improve the quality and the information power of each analytical run.

5.7.6 Supplementary material

Supplementary material at the Google Drive's link:

https://drive.google.com/drive/folders/1dw3d3BvjJIrAmzTWl_qhkf1GFTmHyIBF?usp=sharing

References

(1) Stilo, F.; Bicchi, C.; Robbat, A.; Reichenbach, S. E.; Cordero, C. Untargeted Approaches in Food-Omics: The Potential of Comprehensive Two-Dimensional Gas Chromatography/Mass Spectrometry. *TrAC Trends Anal. Chem.* **2021**, *135*, 116162. https://doi.org/10.1016/j.trac.2020.116162.

(2) Stilo, F.; Bicchi, C.; Jimenez-Carvelo, A. M.; Cuadros-Rodriguez, L.; Reichenbach, S. E.; Cordero, C. Chromatographic Fingerprinting by Comprehensive Two-Dimensional Chromatography: Fundamentals and Tools. *TrAC - Trends Anal. Chem.* **2021**, *134*, 116133. https://doi.org/10.1016/j.trac.2020.116133.

(3) Cordero, C.; Kiefl, J.; Reichenbach, S. E.; Bicchi, C. Characterization of Odorant Patterns by Comprehensive Two-Dimensional Gas Chromatography: A Challenge in Omic Studies. *TrAC Trends Anal. Chem.* **2019**, *113*, 364–378. https://doi.org/10.1016/j.trac.2018.06.005.

(4) Tranchida, P. Q.; Purcaro, G.; Maimone, M.; Mondello, L. Impact of Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry on Food Analysis. *J. Sep. Sci.* **2016**, *39* (1), 149–161. https://doi.org/10.1002/jssc.201500379.

(5) Magagna, F.; Valverde-Som, L.; Ruíz-Samblás, C.; Cuadros-Rodríguez, L.; Reichenbach, S. E.; Bicchi, C.; Cordero, C. Combined Untargeted and Targeted Fingerprinting with Comprehensive Two-Dimensional Chromatography for Volatiles and Ripening Indicators in Olive Oil. *Anal. Chim. Acta* **2016**, *936*, 245–258. https://doi.org/10.1016/j.aca.2016.07.005.

(6) Reichenbach, S. E.; Zini, C. A.; Nicolli, K. P.; Welke, J. E.; Cordero, C.; Tao, Q. Benchmarking Machine Learning Methods for Comprehensive Chemical Fingerprinting and Pattern Recognition. *J. Chromatogr. A* **2019**, *1595*, 158–167. https://doi.org/10.1016/j.chroma.2019.02.027.

(7) Nicolotti, L.; Cordero, C.; Cagliero, C.; Liberto, E.; Sgorbini, B.; Rubiolo, P.; Bicchi, C. Quantitative Fingerprinting by Headspace-Two-Dimensional Comprehensive Gas Chromatography-Mass Spectrometry of Solid Matrices: Some Challenging Aspects of the Exhaustive Assessment of Food Volatiles. *Anal. Chim. Acta* **2013**, *798*, 115–125. https://doi.org/10.1016/j.aca.2013.08.052.

(8) Cordero, C.; Rubiolo, P.; Reichenbach, S. E.; Carretta, A.; Cobelli, L.; Giardina, M.; Bicchi, C. Method Translation and Full Metadata Transfer from Thermal to Differential Flow Modulated Comprehensive Two Dimensional Gas Chromatography: Profiling of Suspected Fragrance Allergens. *J. Chromatogr. A* **2017**, *1480*, 70–82. https://doi.org/10.1016/j.chroma.2016.12.011.

(9) Cordero, C.; Rubiolo, P.; Cobelli, L.; Stani, G.; Miliazza, A.; Giardina, M.; Firor, R.; Bicchi, C. Potential of the Reversed-Inject Differential Flow Modulator for Comprehensive Two-Dimensional Gas Chromatography in the Quantitative Profiling and Fingerprinting of Essential Oils of Different Complexity. *J. Chromatogr. A* **2015**, *1417*, 79–95. https://doi.org/10.1016/j.chroma.2015.09.027.

(10) Stilo, F.; Gabetti, E.; Bicchi, C.; Carretta, A.; Peroni, D.; Reichenbach, S. E.; Cordero, C.; McCurry, J. A Step Forward in the Equivalence between Thermal and Differential-Flow Modulated Comprehensive Two-Dimensional Gas Chromatography Methods. *J. Chromatogr. A* **2020**, *1627*, 461396. https://doi.org/10.1016/j.chroma.2020.461396.

(11) Franchina, F. A.; Maimone, M.; Tranchida, P. Q.; Mondello, L. Flow Modulation Comprehensive Two-Dimensional Gas Chromatography–Mass Spectrometry Using ≈4 ML Min−1 Gas Flows. *J. Chromatogr. A* **2016**, *1441*, 134–139. https://doi.org/10.1016/j.chroma.2016.02.041.

(12) Seeley, J. V.; Micyus, N. J.; Bandurski, S. V.; Seeley, S. K.; McCurry, J. D. Microfluidic Deans Switch for Comprehensive Two-Dimensional Gas Chromatography. *Anal. Chem.* **2007**, *79* (5), 1840–1847. https://doi.org/10.1021/ac061881g.

(13) Seeley, J. V.; Micyus, N. J.; McCurry, J. D.; Seeley, S. K. Comprehensive Two-Dimensional Gas Chromatography with a Simple Fluidic Modulator. *American Laboratory*. 2006, pp 24–26.

(14) Klee, M. S.; Blumberg, L. M. Theoretical and Practical Aspects of Fast Gas Chromatography and Method Translation. *J. Chromatogr. Sci.* **2002**, *40* (5), 234–247. https://doi.org/10.1093/chromsci/40.5.234.

(15) Giardina, M.; McCurry, J. D.; Cardinael, P.; Semard-Jousset, G.; Cordero, C.; Bicchi, C. Development and Validation of a Pneumatic Model for the Reversed-Flow Differential Flow Modulator for Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2018**, *1577*, 72–81. https://doi.org/10.1016/j.chroma.2018.09.022.

(16) Eurachem. *Eurachem Guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics.*; 2014. https://doi.org/978-91-87461-59-0.

(17) Commission, E. Commission Decision 2002/657/EC Implementing Council Directive 96/23/EC Concerning the Performance of Analytical Methods and the Interpretation of Results. *Off. J. Eur. Union* **2002**, *L221* (23 May 1996), 8–36.

(18) Grigg, D. Olive Oil, the Mediterranean and the World. *GeoJournal* **2001**, *53* (2), 163–172.

(19) Cecchi, L.; Migliorini, M.; Giambanelli, E.; Rossetti, A.; Cane, A.; Melani, F.; Mulinacci, N. Headspace Solid-Phase Microextraction-Gas Chromatography-Mass Spectrometry Quantification of the Volatile Profile of More than 1200 Virgin Olive Oils for Supporting the Panel Test in Their Classification: Comparison of Different Chemometric Approaches. *J. Agric. Food Chem.* **2019**, *67* (32), 9112–9120. https://doi.org/10.1021/acs.jafc.9b03346.

(20) IOC. International Olive Council. *Newsl. No 144 DECEMBER 2019* **2019**, No. December.

(21) EU Quality Schemes.

(22) Cuadros-Rodríguez, L.; Ruiz-Samblás, C.; Valverde-Som, L.; Pérez-Castaño, E.; González-Casado, A. Chromatographic Fingerprinting: An Innovative Approach for Food "identitation" and Food Authentication - A Tutorial. *Anal. Chim. Acta* **2016**, *909*, 9–23. https://doi.org/10.1016/j.aca.2015.12.042.

(23) Dunkel, A.; Steinhaus, M.; Kotthoff, M.; Nowak, B.; Krautwurst, D.; Schieberle, P.; Hofmann, T. Nature's Chemical Signatures in Human Olfaction: A Foodborne Perspective for Future Biotechnology. *Angew. Chemie - Int. Ed.* **2014**, *53* (28), 7124–7143. https://doi.org/10.1002/anie.201309508.

(24) Purcaro, G.; Cordero, C.; Liberto, E.; Bicchi, C.; Conte, L. S. Toward a Definition of Blueprint of Virgin Olive Oil by Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2014**, *1334*, 101–111. https://doi.org/10.1016/j.chroma.2014.01.067.

(25) Belhassen, E.; Bressanello, D.; Merle, P.; Raynaud, E.; Bicchi, C.; Chaintreau, A.; Cordero, C. Routine Quantification of 54 Allergens in Fragrances Using Comprehensive Two-Dimensional Gas Chromatography-Quadrupole Mass Spectrometry with Dual Parallel Secondary Columns. Part I: Method Development. *Flavour Fragr. J.* **2018**, *33* (1), 63–74. https://doi.org/10.1002/ffj.3416.

(26) Cordero, C.; Guglielmetti, A.; Sgorbini, B.; Bicchi, C.; Allegrucci, E.; Gobino, G.; Baroux, L.; Merle, P. Odorants Quantitation in High-Quality Cocoa by Multiple Headspace Solid Phase Micro-Extraction: Adoption of FID-Predicted Response Factors to Extend Method Capabilities and Information Potential. *Anal. Chim. Acta* **2019**, *1052*. https://doi.org/10.1016/j.aca.2018.11.043.

(27) Casadei, E.; Valli, E.; Aparicio-Ruiz, R.; Ortiz-Romero, C.; García-González, D. L.; Vichi, S.; Quintanilla-Casas, B.; Tres, A.; Bendini, A.; Toschi, T. G. Peer Inter-Laboratory Validation Study of a Harmonized SPME-GC-FID Method for the Analysis of Selected Volatile Compounds in Virgin Olive Oils. *Food Control* **2021**, *123* (July 2020). https://doi.org/10.1016/j.foodcont.2020.107823.

(28) Vichi, S.; Pizzale, L.; Conte, L. S.; Buxaderas, S.; López-Tamames, E. Solid-Phase Microextraction in the Analysis of Virgin Olive Oil Volatile Fraction: Characterization of Virgin Olive Oils from Two Distinct Geographical Areas of Northern Italy. *J. Agric. Food Chem.* **2003**, *51* (22), 6572–6577. https://doi.org/10.1021/jf030269c.

(29) Quintanilla-Casas, B.; Bustamante, J.; Guardiola, F.; García-González, D. L.; Barbieri, S.; Bendini, A.; Toschi, T. G.; Vichi, S.; Tres, A. Virgin Olive Oil Volatile Fingerprint and Chemometrics: Towards an Instrumental Screening Tool to Grade the Sensory Quality. *LWT* **2020**, *121*, 108936. https://doi.org/https://doi.org/10.1016/j.lwt.2019.108936.

(30) Quintanilla-Casas, B.; Marin, M.; Guardiola, F.; García-González, D. L.; Barbieri, S.; Bendini, A.; Toschi, T. G.; Vichi, S.; Tres, A. Supporting the Sensory Panel to Grade Virgin Olive Oils: An in-House-Validated Screening Tool by Volatile Fingerprinting and Chemometrics. *Foods* **2020**, *9* (10), 1–14. https://doi.org/10.3390/foods9101509.

(31) Vaz-Freire, L. T.; da Silva, M. D. R. G.; Freitas, A. M. C. Comprehensive Two-Dimensional Gas Chromatography for Fingerprint Pattern Recognition in Olive Oils Produced by Two Different Techniques in Portuguese Olive Varieties Galega Vulgar, Cobrançosa e Carrasquenha. *Anal. Chim. Acta* **2009**, *633* (2), 263–270. https://doi.org/10.1016/j.aca.2008.11.057.

(32) Stilo, F.; Liberto, E.; Reichenbach, S. E.; Tao, Q.; Bicchi, C.; Cordero, C. Untargeted and Targeted Fingerprinting of Extra Virgin Olive Oil Volatiles by Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry: Challenges in Long-Term Studies. *J. Agric. Food Chem.* **2019**, *67* (18), 5289–5302. https://doi.org/10.1021/acs.jafc.9b01661.

(33) EU Food Quality Labels.

(34) Progetto Ager. *Violin Project*; 2016.

(35) Neugebauer, A.; Granvogl, M.; Schieberle, P. Characterization of the Key Odorants in High-Quality Extra Virgin Olive Oils and Certified Off-Flavor Oils to Elucidate Aroma Compounds Causing a Rancid Off-Flavor. *J. Agric. Food Chem.* **2020**, *68* (21), 5927–5937. https://doi.org/10.1021/acs.jafc.0c01674.

(36) Kalua, C. M.; Allen, M. S.; Bedgood, D. R.; Bishop, A. G.; Prenzler, P. D.; Robards, K. Olive Oil Volatile Compounds, Flavour Development and Quality: A Critical Review. *Food Chem.* **2007**, *100* (1), 273–286. https://doi.org/10.1016/j.foodchem.2005.09.059.

(37) Stilo, F.; Cordero, C.; Sgorbini, B.; Bicchi, C.; Liberto, E. Highly Informative Fingerprinting of Extra-Virgin Olive Oil Volatiles: The Role of High Concentration-Capacity Sampling in Combination with Comprehensive Two-Dimensional Gas Chromatography. *Separations* **2019**, *6* (3), 34. https://doi.org/10.3390/separations6030034.

(38) Angerosa, F. Influence of Volatile Compounds on Virgin Olive Oil Quality Evaluated by Analytical Approaches and Sensor Panels. *Eur. J. Lipid Sci. Technol.* **2002**, *104* (9–10), 639–660. https://doi.org/https://doi.org/10.1002/1438-9312(200210)104:9/10<639::AID-EJLT639>3.0.CO;2-U.

(39) Morales, M. T.; Luna, G.; Aparicio, R. Comparative Study of Virgin Olive Oil Sensory Defects. *Food Chem.* **2005**, *91* (2), 293–301. https://doi.org/10.1016/j.foodchem.2004.06.011.

(40) Aprea, E.; Gasperi, F.; Betta, E.; Sani, G.; Cantini, C. Variability in Volatile Compounds from Lipoxygenase Pathway in Extra Virgin Olive Oils from Tuscan Olive Germoplasm by Quantitative SPME/GC-MS. *J. Mass Spectrom.* **2018**, *53* (9), 824–832. https://doi.org/10.1002/jms.4274.

(41) Cecchi, L.; Migliorini, M.; Mulinacci, N. Virgin Olive Oil Volatile Compounds: Composition, Sensory Characteristics, Analytical Approaches, Quality Control, and Authentication. *J. Agric. Food Chem.* **2021**. https://doi.org/10.1021/acs.jafc.0c07744.

(42) Wang, Y.; O'Reilly, J.; Chen, Y.; Pawliszyn, J. Equilibrium In-Fibre Standardisation Technique for Solid-Phase Microextraction. *J. Chromatogr. A* **2005**, *1072* (1), 13–17. https://doi.org/10.1016/j.chroma.2004.12.084.

(43) Vichi, S.; Guadayol, J. M.; Caixach, J.; Lopeztamames, E.; Buxaderas, S.; López-Tamames, E.; Buxaderas, S. Comparative Study of Different Extraction Techniques for the Analysis of Virgin Olive Oil Aroma. *Food Chem.* **2007**, *105* (3), 1171–1178. https://doi.org/10.1016/j.foodchem.2007.02.018.

(44) Cordero, C.; Cagliero, C.; Liberto, E.; Nicolotti, L.; Rubiolo, P.; Sgorbini, B.; Bicchi, C. High Concentration Capacity Sample Preparation Techniques to Improve the Informative Potential of Two-Dimensional Comprehensive Gas Chromatography-Mass Spectrometry: Application to Sensomics. *J. Chromatogr. A* **2013**, *1318*, 1–11. https://doi.org/10.1016/j.chroma.2013.09.065.

(45) Stilo, F.; Liberto, E.; Reichenbach, S. E.; Tao, Q.; Bicchi, C.; Cordero, C. Exploring the Extra-Virgin Olive Oil Volatilome by Adding Extra Dimensions to Comprehensive Two-Dimensional Gas Chromatography and Time of Flight Mass Spectrometry Featuring Tandem Ionization: Validation of Ripening Markers in Headspace Linearity Conditio. *J. AOAC Int.* **2020**. https://doi.org/10.1093/jaoacint/qsaa095.

(46) Kolb, B.; Ettre, L. S. *Static Headspace-Gas Chromatography : Theory and Practice*; Wiley-VCH: New York, 2006.

(47) Nicolotti, L.; Mall, V.; Schieberle, P. Characterization of Key Aroma Compounds in a Commercial Rum and an Australian Red Wine by Means of a New Sensomics-Based Expert System (SEBES)—An Approach To Use Artificial Intelligence in Determining Food Odor Codes. *J. Agric. Food Chem.* **2019**, *67* (14), 4011–4022. https://doi.org/10.1021/acs.jafc.9b00708.

(48) Bicchi, C.; Ruosi, M. R.; Cagliero, C.; Cordero, C.; Liberto, E.; Rubiolo, P.; Sgorbini, B. Quantitative Analysis of Volatiles from Solid Matrices of Vegetable Origin by High

Concentration Capacity Headspace Techniques: Determination of Furan in Roasted Coffee. *J. Chromatogr. A* **2011**, *1218* (6), 753–762. https://doi.org/10.1016/j.chroma.2010.12.002.

(49) Pawliszyn, J.; Ross, C. F. Headspace Analysis. In *Comprehensive Sampling and Sample Preparation*; Elsevier, 2012; Vol. 2, pp 27–50. https://doi.org/10.1016/B978-0-12-381373-2.10036-5.

(50) Sgorbini, B.; Cagliero, C.; Liberto, E.; Rubiolo, P.; Bicchi, C.; Cordero, C. Strategies for Accurate Quantitation of Volatiles from Foods and Plant-Origin Materials: A Challenging Task. *J. Agric. Food Chem.* **2019**, acs.jafc.8b06601. https://doi.org/10.1021/acs.jafc.8b06601.

(51) Cordero, C.; Guglielmetti, A.; Sgorbini, B.; Bicchi, C.; Allegrucci, E.; Gobino, G.; Baroux, L.; Merle, P. Odorants Quantitation in High-Quality Cocoa by Multiple Headspace Solid Phase Micro-Extraction: Adoption of FID-Predicted Response Factors to Extend Method Capabilities and Information Potential. *Anal. Chim. Acta* **2019**, *1052*, 190–201. https://doi.org/10.1016/j.aca.2018.11.043.

(52) Kolb, B.; Ettre, L. S. Theory and Practice of Multiple Headspace Extraction. *Chromatographia* **1991**, *32* (11–12), 505–513. https://doi.org/10.1007/BF02327895.

(53) Suzuki, M.; Tsuge, S.; Takeuchi, T. Gas Chromatographic Estimation of Occluded Solvents in Adhesive Tape by Periodic Introduction Method. *Anal. Chem.* **1970**, *42* (14), 1705–1708. https://doi.org/10.1021/ac50160a035.

(54) MCAULLIFE C. GC Determination of Solutes by Multiple Phase Equilibration. *Chem. Tech.* **1971**, No. 1, 46–51.

(55) Sgorbini, B.; Bicchi, C.; Cagliero, C.; Cordero, C.; Liberto, E.; Rubiolo, P. Herbs and Spices: Characterization and Quantitation of Biologically-Active Markers for Routine Quality Control by Multiple Headspace Solid-Phase Microextraction Combined with Separative or Non-Separative Analysis. *J. Chromatogr. A* **2015**, *1376*, 9–17. https://doi.org/10.1016/j.chroma.2014.12.007.

(56) Costa, R.; Tedone, L.; De Grazia, S.; Dugo, P.; Mondello, L. Multiple Headspace-Solid-Phase Microextraction: An Application to Quantification of Mushroom Volatiles. *Anal. Chim. Acta* **2013**, *770*, 1–6. https://doi.org/10.1016/j.aca.2013.01.041.

(57) Ezquerro, Ó.; Ortiz, G.; Pons, B.; Tena, M. T. Determination of Benzene, Toluene, Ethylbenzene and Xylenes in Soils by Multiple Headspace Solid-Phase Microextraction. *J. Chromatogr. A* **2004**, *1035* (1), 17–22. https://doi.org/10.1016/j.chroma.2004.02.030.

(58) Xie, W.-Q.; Gong, Y.-X.; Yu, K.-X. Enhancing the Sensitivity of Full Evaporation Technique Using Multiple Headspace Extraction Analysis. *Chromatographia* **2017**, *80* (8), 1263–1268. https://doi.org/10.1007/s10337-017-3343-x.

(59) Wenzl, T.; Haedrich, J.; Schaechtele, A.; Robouch, P.; Stroka, J. *Guidance Document on the Estimation of LOD and LOQ for Measurements in the Field of Contaminants in Feed and Food. EUR 28099 EN*; 2016. https://doi.org/10.2787/8931.

(60) Cordero, C.; Kiefl, J.; Schieberle, P.; Reichenbach, S. E.; Bicchi, C. Comprehensive Two-Dimensional Gas Chromatography and Food Sensory Properties: Potential and Challenges. *Anal. Bioanal. Chem.* **2015**, *407* (1), 169–191. https://doi.org/10.1007/s00216-014-8248-z.

(61) De Saint Laumer, J. Y.; Leocata, S.; Tissot, E.; Baroux, L.; Kampf, D. M.; Merle, P.; Boschung, A.; Seyfried, M.; Chaintreau, A. Prediction of Response Factors for Gas Chromatography with Flame Ionization Detection: Algorithm Improvement, Extension to Silylated

Compounds, and Application to the Quantification of Metabolites. *J. Sep. Sci.* **2015**, *38* (18), 3209–3217. https://doi.org/10.1002/jssc.201500106.

(62) Cecchi, L.; Migliorini, M.; Giambanelli, E.; Rossetti, A.; Cane, A.; Mulinacci, N. New Volatile Molecular Markers of Rancidity in Virgin Olive Oils under Nonaccelerated Oxidative Storage Conditions. *J. Agric. Food Chem.* **2019**, *67* (47), 13150–13163. https://doi.org/10.1021/acs.jafc.9b05809.

(63) Angerosa, F.; Servili, M.; Selvaggini, R.; Taticchi, A.; Esposto, S.; Montedoro, G. Volatile Compounds in Virgin Olive Oil: Occurrence and Their Relationship with the Quality. *J. Chromatogr. A* **2004**, *1054* (1–2), 17–31. https://doi.org/10.1016/j.chroma.2004.07.093.

(64) Genovese, A.; Caporaso, N.; Leone, T.; Paduano, A.; Mena, C.; Perez-Jimenez, M. A.; Sacchi, R. Use of Odorant Series for Extra Virgin Olive Oil Aroma Characterisation. *J. Sci. Food Agric.* **2019**, *99* (3), 1215–1224. https://doi.org/10.1002/jsfa.9293.

(65) Quintanilla-Casas, B.; Bertin, S.; Leik, K.; Bustamante, J.; Guardiola, F.; Valli, E.; Bendini, A.; Toschi, T. G.; Tres, A.; Vichi, S. Profiling versus Fingerprinting Analysis of Sesquiterpene Hydrocarbons for the Geographical Authentication of Extra Virgin Olive Oils. *Food Chem.* **2020**, *307* (March 2019). https://doi.org/10.1016/j.foodchem.2019.125556.

(66) Cavalli, J. F.; Fernandez, X.; Lizzani-Cuvelier, L.; Loiseau, A. M. Characterization of Volatile Compounds of French and Spanish Virgin Olive Oils by HS-SPME: Identification of Quality-Freshness Markers. *Food Chem.* **2004**, *88* (1), 151–157. https://doi.org/10.1016/j.foodchem.2004.04.003.

(67) Aparicio, R.; Morales, M. T.; Aparicio-Ruiz, R.; Tena, N.; García-González, D. L. Authenticity of Olive Oil: Mapping and Comparing Official Methods and Promising Alternatives. *Food Res. Int.* **2013**, *54* (2), 2025–2038. https://doi.org/10.1016/j.foodres.2013.07.039.

(68) Hachicha Hbaieb, R.; Kotti, F.; Gargouri, M.; Msallem, M.; Vichi, S. Ripening and Storage Conditions of Chétoui and Arbequina Olives: Part I. Effect on Olive Oils Volatiles Profile. *Food Chem.* **2016**, *203*, 548–558. https://doi.org/10.1016/j.foodchem.2016.01.089.

(69) Ros, A. Da; Masuero, D.; Riccadonna, S.; Bubola, K. B.; Mulinacci, N.; Mattivi, F.; Lukić, I.; Vrhovsek, U. Complementary Untargeted and Targeted Metabolomics for Differentiation of Extra Virgin Olive Oils of Different Origin of Purchase Based on Volatile and Phenolic Composition and Sensory Quality. *Molecules* **2019**, *24* (16), 1–17. https://doi.org/10.3390/molecules24162896.

Conclusions

The aim of this Doctoral Thesis was the development of advanced and innovative analytical strategies for the characterization of the Italian olive oil through the in-depth investigation of their complex volatile fraction.

In this context, the role of the reviews was on the conceptualization of peculiar investigation approaches (*i.e.*, chromatographic fingerprinting by GC×GC), on the critical evaluation of some trends and strategies adopted in the field of food-omics.

The contribution entitled "*Chromatographic fingerprinting by comprehensive two-dimensional chromatography: fundamentals and tools*" reviewed the state-of-the art literature dealing with *chromatographic fingerprinting* concept. The approaches were described and discussed for their potentials and limitations to give access to a higher level of information about sample(s)' composition. Different type of features available for chromatographic fingerprinting (*i.e.,* datapoint, region, peak and peak-region features) were rationalized and related strategies analyzed in depth. Challenging scenarios with parallel detection, translation of methods from thermal to differential-flow modulated GC×GC and variable ionization energy MS were also accounted for their relevance in specific applications. Machine learning was briefly introduced highlighting its fundamental role in supporting fingerprinting workflows.

The focus of the contribution entitled "*Untargeted approaches in food-omics: the potential of comprehensive two-dimensional gas chromatography/mass spectrometry*" was on untargeted processing approaches currently adopted to explore the complex data matrices produced by the GC×GC-MS when adopted in food-omics investigations. In this review, it was found that the increased separation space offered by GC×GC coupled with MS feature alignment can detect/identify "unknown – knowns" while providing the means to obtain the total, detectable metabolome in agricultural crops and food.

The latest contribution, entitled "*Comprehensive two-dimensional gas chromatography as a boosting technology in food-omics investigations*" was focused on the role that GC×GC can play within the investigation workflows of food-omics and related disciplines and sub-disciplines, including food metabolomics, nutrimetabolomics, sensomics, and food safety, which require a truly comprehensive approaches to capture compositional complexity of samples and to establish robust correlations to better understand complex biological phenomena. The potential of GC×GC to tackle compositional challenges and dedicated data processing were analyzed, leading to the modern concepts of individual/personalized investigations.

Research papers followed a rational progression by investigating critical steps of the whole analytical process matching for full automation, robustness and an increased informative potential: sample preparation, separation and detection, data processing and quantitative analysis.

In particular, the study entitled "*Highly informative fingerprinting of extra-virgin olive oil volatiles: The role of high concentration-capacity sampling in combination with comprehensive two-dimensional gas chromatography*" showed how high concentration capacity HS sampling could successfully be integrated in a GC×GC-TOF MS platform for highly informative fingerprinting of the complex EVOO volatilome. The influence of different variables on extraction effectiveness was shown, focusing on potent odorants and/or on key-markers known to be correlated with oil sensory defects. SPME with a multi-component fiber confirmed its good quali-quantitative coverage of the different chemical dimensions present in the EVOO volatilome. Moreover, it was indicated that, to derive consistent and accurate quantitative considerations, HS linearity should be accomplished at the sampling stage.

The paper entitled "*Chromatographic fingerprinting by template matching for data collected by comprehensive two-dimensional gas chromatography*" delineated the operative protocol for the application of the template matching approach, enabling 2D pattern recognition in a very effective, specific, semi-automatic and intuitive way. The visual approach, moreover, is able to completely change the perspective on chromatograms interpretation and elaboration, and it was applied to demonstrate its effectiveness to compare EVOOs from different Italian regions. This contribution will be soon available as a "video protocol" freely accessible (open-access).

The publication "*Untargeted and Targeted Fingerprinting of Extra Virgin Olive Oil Volatiles by Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry: Challenges in Long-Term Studies*" focuses on the complex volatile fraction of EVOO and addresses 2D-peak patterns variations, including MS signal fluctuations, as they may occur in long-term studies where pedo-climatic, harvest year or shelf-life changes are studied. 2D-pattern misalignments were forced and simulated by changing chromatographic settings and MS acquisition, and all procedural steps, preceding pattern recognition by template matching, were analyzed defining a rational workflow to accurately re-align patterns and analytes metadata. S/R detection threshold, reference spectra extraction, and similarity match factor threshold are critical to avoid false-negative matches, while distance thresholds and polynomial transform parameters are key for effective template matching. The accuracy reached for the combined *untargeted* and *targeted (UT) fingerprinting* was 97.9 %.

"*A step forward in the equivalence between thermal and differential-flow modulated comprehensive two-dimensional gas chromatography methods*" investigated the performances of differential flow modulation as valid alternative to thermal modulation, especially in routine applications. Six different columns configurations were systematically evaluated for the flow-modulated counterpart and the set-up providing best performances (20 m x 0.18 mm d_c x 0.18 µm d_f + 1.8 m x 0.18 mm d_c x 0.18 µm d_f) was further evaluated to assess method sensitivity, linearity, accuracy, and pattern reliability. The experimental results convincingly showed that the method translation procedure was effective, allowing a successful transfer of target analytes template metadata. However, the need for higher flows to the ^{2}D of a FM system, at least to achieve adequate separation power, slightly limits system performances resulting either in an equivalent sensitivity and quantitation consistency at the cost of ~20 % separation power.

The study "*Exploring the extra-virgin olive oil volatilome by adding extra dimensions to comprehensive two-dimensional gas chromatography and time of flight mass spectrometry featuring tandem ionization: validation of ripening markers in headspace linearity conditions*" explored the complex EVOO volatilome by combining HS-SPME, applied under HS linearity conditions, to GC×GC-TOF MS featuring hard and soft

ionization in tandem. Multiple analytical dimensions were combined in a single run and evaluated in terms of chemical dimensionality, method absolute and relative sensitivity, identification reliability provided by spectral signatures acquired at 70 and 12 eV, and dynamic and linear range of response provided by soft ionization. Method effectiveness was validated on a sample set of oils from *Picual* olives (Altipiano de Granada, Spain) harvested at different ripening stages. Markers (3,4-diethyl-1,5-hexadiene (*RS*/*SR*), 3,4-diethyl-1,5-hexadiene (meso), *(5Z)*-3-ethyl-1,5-octadiene, *(5E)*-3-ethyl-1,5-octadiene, *(E,Z)*-3,7-decadiene and *(E,E)*-3,7-decadiene, *(Z)*-2-hexenal, *(Z)*-3-hexenal and *(Z)*-3-hexenal, *(E)*-2-pentenal, *(Z)*-2-pentenal, 1-pentanol, 1-penten-3-ol, 3-pentanone, and 1-penten-3-one) and quality indexes (*(Z)*-3-Hexenal/Nonanal, *(Z)*-3-Hexenal/Octane, *(E)*-2-Pentenal/Nonanal, and *(E)*-2-Pentenal/Octane) were confirmed for their validity in HS linearity conditions. For the complex olive oil volatilome, the proposed approach offered concrete advantages for the validation of the informative role of existing analytes while suggesting new potential markers to be studied in larger sample sets.

"*Chromatographic fingerprinting enables effective discrimination and identitation of high-quality Italian extra-virgin olive oils*" was focused on the high-quality food authentication, a challenging process that takes advantages by highly informative chromatographic fingerprinting and its *identitation* potential. In this study, the unique chemical traits of the complex volatile fraction of EVOO from Italian production, was captured by GC×GC-TOF MS and explored by pattern recognition algorithms. The consistent re-alignment of untargeted and targeted features over 73 samples, including oils obtained by different olives cultivar (n=24), harvest years (n=3) and processing technologies, provided solid foundation for samples identification and discrimination based on production Region (n=6). Through a dedicated multivariate statistics workflow, *identitation* was achieved by two-level PLS regression, that highlighted Region diagnostic patterns accounting between 58 and 82 of untargeted and targeted compounds, while samples classification was by sequential application of SIMCA models, one for each production Region. Samples were correctly classified in five of the six single-class models and quality parameters (*i.e.*, sensitivity, specificity, precision, efficiency, and area under the receiver operating characteristic curve (AUC)) were equal to 1.00.

"*Delineating the extra-virgin olive oil aroma blueprint by multiple headspace solid phase microextraction and differential-flow modulated comprehensive two-dimensional gas chromatography*" was focused on the adoption of comprehensive two-dimensional gas chromatography with parallel mass spectrometry and flame ionization detection (GC×GC-MS/FID) to enable effective chromatographic fingerprinting of complex samples by comprehensively mapping untargeted and targeted components. The complementary characteristics of MS and FID opened the possibility of performing multi-target quantitative profiling with great accuracy. In this study, untargeted/targeted (*UT*) fingerprinting of EVOO volatile fractions was combined with accurate quantitative profiling by multiple headspace solid phase microextraction (MHS-SPME), highlighting and measuring the inaccuracy of the largely adopted approach that uses single internal standardization for quantitative purposes. Subsequently, external calibration on fifteen pre-selected analytes and FID predicted relative response factors (RRFs) enabled the accurate quantification of forty-two analytes in total, including key-aroma compounds, potent odorants, and olive oil geographical markers. Finally, through odor activity values (OAVs), the sensory contribution of the quantified analytes was integrated in the process, and the MHS-SPME-GC×GC-MS/FID method was here adopted as an *artificial intelligence smelling machine* to delineate EVOOs aroma blueprints.

Scientific publications

Stilo, F., Jiménez-Carvelo, A.M., Liberto, E., Bicchi, C., Reichenbach, S.E., Cuadros-Rodríguez, L, Cordero C. Chromatographic fingerprinting enables effective discrimination and *identitation* of high-quality Italian extra-virgin olive oils (Submitted to *Journal of Agricultural and Food Chemistry)*

Gabetti, E., Sgorbini, B., **Stilo, F**., Bicchi, C., Rubiolo, P., Chialva, F., Reichenbach, S.E., Bongiovanni, V., Cordero, C., Cavallero. A. Chemical fingerprinting strategies based on comprehensive two-dimensional gas chromatography combined with gas chromatography-olfactometry to capture the unique signature of Piemonte peppermint essential oil (Mentha x piperita var Italo-Mitcham), **2021**, 1645, 462101, https://doi.org/10.1016/j.chroma.2021.462101

Cialiè Rosso, M., **Stilo, F**., Mascrez, S., Bicchi, C., Purcaro, G., Cordero, C. Shelf-Life Evolution of the Fatty Acid Fingerprint in High-Quality Hazelnuts (Corylus avellana L.) Harvested in Different Geographical Regions. *Foods*, **2021**, 10(3), 685; https://doi.org/10.3390/foods10030685

Stilo, F., Bicchi, C., Reichenbach, S. E., Cordero, C. Comprehensive two-dimensional gas chromatography as a boosting technology in food-omic investigations. *Journal of Separation Science*, **2021**, 44, 8, 1592-1611, https://doi.org/10.1002/jssc.202100017

Stilo, F., Bicchi, C., Robbat, A., Reichenbach, S. E., Cordero, C. Untargeted approaches in food-omics: The potential of comprehensive two-dimensional gas chromatography/mass spectrometry. *TrAC - Trends in Analytical Chemistry.*, **2021**, 135, 116162. https://doi.org/10.1016/j.trac.2020.116162

Cialiè Rosso, M., **Stilo, F.**, Bicchi, C., Charron, M., Rosso, G., Menta, R., Reichenbach, S. E., Weinert, C. H., Mack, C. I., Kulling, S. E., & Cordero, C. Combined Untargeted and Targeted Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography to Track Compositional Changes on Hazelnut Primary Metabolome during Roasting. *Applied Sciences*, **2021**, 11(2), 525. https://doi.org/10.3390/app11020525

Stilo, F., Bicchi, C., Jimenez-carvelo, A. M., Cuadros-rodriguez, L., Reichenbach, S. E., & Cordero, C. Chromatographic fingerprinting by comprehensive two-dimensional chromatography : Fundamentals and tools. *TrAC - Trends in Analytical Chemistry*, **2021**, 134, 116-133. https://doi.org/10.1016/j.trac.2020.116133

Stilo, F., Liberto, E., Spigolon, N., Genova, G., Rosso, G., Fontana, M., Reichenbach, S. E., Bicchi, C., & Cordero, C. An effective chromatographic fingerprinting workflow based on

comprehensive two-dimensional gas chromatography - mass spectrometry to establish volatiles patterns discriminative of spoiled hazelnuts (Corylus avellana L.). *Food Chemistry*, **2020**, 340, 128-135. https://doi.org/10.1016/j.foodchem.2020.128135

Stilo, F., Cordero, C., Bicchi, C., Peroni, D., Tao, Q., & Reichenbach, S. E. (2020). Chromatographic fingerprinting by template matching for data collected by comprehensive two-dimensional gas chromatography. *Journal of Visualized Experiments*, **2020** (163), 1–20. https://doi.org/10.3791/61529

Cialiè Rosso, M., **Stilo, F.**, Squara, S., Liberto, E., Mai, S., Mele, C., Marzullo, P., Aimaretti, G., Reichenbach, S. E., Collino, M., Bicchi, C., & Cordero, C. Exploring extra dimensions to capture saliva metabolite fingerprints from metabolically healthy and unhealthy obese patients by comprehensive two-dimensional gas chromatography featuring Tandem Ionization mass spectrometry. *Analytical and Bioanalytical Chemistry*. **2020**, 413, 403–418 https://doi.org/10.1007/s00216-020-03008-6

Stilo, F., Liberto, E., Reichenbach, S. E., Tao, Q., Bicchi, C., & Cordero, C. (2020). Exploring the extra-virgin olive oil volatilome by adding extra dimensions to comprehensive two-dimensional gas chromatography and time of flight mass spectrometry featuring tandem ionization: validation of ripening markers in headspace linearity conditio. *Journal of AOAC INTERNATIONAL*. **2020**. https://doi.org/10.1093/jaoacint/qsaa095

Stilo, F., Gabetti, E., Bicchi, C., Carretta, A., Peroni, D., Reichenbach, S. E., Cordero, C., & McCurry, J. A step forward in the equivalence between thermal and differential-flow modulated comprehensive two-dimensional gas chromatography methods. *Journal of Chromatography A*, **2020**, *1627*, 461396. https://doi.org/10.1016/j.chroma.2020.461396

Stilo, F., Tredici, G., Bicchi, C., Robbat, A., Morimoto, J., & Cordero, C. Climate and Processing Effects on Tea (Camellia sinensis L. Kuntze) Metabolome: Accurate Profiling and Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography/Time-of-Flight Mass Spectrometry. *Molecules*, **2020**, *25*(10), 2447. https://doi.org/10.3390/molecules25102447

Stilo, F., Liberto, E., Reichenbach, S. E., Tao, Q., Bicchi, C., & Cordero, C. Untargeted and Targeted Fingerprinting of Extra Virgin Olive Oil Volatiles by Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry: Challenges in Long-Term Studies. *Journal of Agricultural and Food Chemistry*, **2019**, *67*(18), 5289–5302. https://doi.org/10.1021/acs.jafc.9b01661

Stilo, F., Liberto, E., Bicchi, C., Reichenbach, S. E., & Cordero, C. GC×GC–TOF-MS and Comprehensive Fingerprinting of Volatiles in Food: Capturing the Signature of Quality. *LCGC Europe*, **2019**, *32*,5,234–242. http://www.chromatographyonline.com/gc-gc-tof-ms-and-comprehensive-fingerprinting-volatiles-food-capturing-signature-quality

Stilo, F., Cordero, C., Sgorbini, B., Bicchi, C., & Liberto, E. Highly Informative Fingerprinting of Extra-Virgin Olive Oil Volatiles: The Role of High Concentration-Capacity Sampling in Combination with Comprehensive Two-Dimensional Gas Chromatography. *Separations*, **2019**, *6*(3), 34. https://doi.org/10.3390/separations6030034

Oral contributions to congresses

Stilo, F., Bicchi C., Reichenbach S.E., McCurry J., Peroni D., Cordero C. Chromatographic fingerprinting and accurate quantitative profiling by multiple headspace solid phase microextraction and differential-flow modulated comprehensive two-dimensional gas chromatography: the aroma blueprint of extra virgin olive oil – *1° European Sample Preparation e-Conference* (11-12/03/2021)

Stilo, F., Liberto E., Reichenbach S.E., Qingping T., Bicchi C., Cordero C.E. Long-term studies on virgin olive oil volatiles: untargeted and targeted fingerprinting by comprehensive two-dimensional gas chromatography with mass spectrometry - *EVOO Research's Got Talent 2020* (Bari, 19-22/01/2020)

Stilo, F., Liberto E., Reichenbach S.E., Qingping T., Bicchi C., Cordero C.E. Extending over long-term studies the untargeted and targeted fingerprinting of extra-virgin olive oil volatiles by comprehensive two-dimensional gas chromatography with mass spectrometry – *9°International Symposium on Recent Advances in Food Analysis* (Prague, 05-08/11/2019)

Poster contributions to congresses

Stilo, F., Liberto, E., Spigolon, N., Genova, G., Fontana, M., Reichenbach, S.E., Bicchi, C., Cordero, C. An effective chromatographic fingerprinting approach based on image and peak-region features generated by comprehensive two-dimensional gas chromatography - mass spectrometry: food quality applications – *12° Multidimensional Chromatography Virtual Workshop* (01-03/02/2021)

Stilo, F., Franchina, F. A., Cordero, C., Focant, J. Comprehensive two-dimensional gas chromatography coupled to mass spectrometry (GC×GC-TOF MS): discrimination of Italian extra virgin olive oils (EVOOs) from different regions and exploration of high resolution (HR) mass spectrometry information – *12° Multidimensional Chromatography Virtual Workshop* (01-03/02/2021)

Stilo, F., Gabetti E., Spigolon N., Genova G., Fontana M., Bicchi C., Cordero C.E. Raw hazelnut volatiles: challenges in defining odorant patterns related to sensory defects by comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry - *9° International Symposium on Recent Advances in Food Analysis* (Prague, 05-08/11/2019)

Stilo, F., Gabetti E., Spigolon N., Genova G., Somenzi M., Fontana M., Bicchi C., Cordero C.E. Chromatographic fingerprinting of hazelnuts volatiles by Comprehensive Two-Dimensional Gas Chromatography coupled with Time of Flight Mass Spectrometry: Challenges in defining odorant patterns related to sensory defects - *43° International Symposium on Capillary Chromatography* (13-17/05/2019)

Stilo, F., Gabetti E., Spigolon N., Genova G., Somenzi M., Fontana M., Bicchi C., Cordero C.E. High-informative chromatographic fingerprinting of raw hazelnuts volatiles by comprehensive two-dimensional gas chromatography coupled with Time of flight mass spectrometry: challenges in defining odorant patterns related to sensory defects - *XII Italian Food Chemistry Congress* (Camerino, 24-27/09/2018)

Stilo, F., Cialiè Rosso M., Liberto E., Carretta A., Cobelli G., Miliazza A., Giardina M., Reichenbach S.E., Qingping T., Bicchi C., Cordero C.E. High-throughput chromatographic fingerprinting of Extra Virgin Olive Oil volatiles by GC×GC-MS/FID and differential flow modulation - *XII Italian Food Chemistry Congress* (Camerino, 24 27/09/2018)

Stilo, F., Liberto E., Cagliero C., Rubiolo P. Sgorbini B., Reichenbach S.E., Qingping T., Bicchi C., Cordero C.E. Extra Virgin Olive oil volatiles a mine of chemical information: challenges in

chromatographic data alignment and response normalization for reliable fingerprinting by comprehensive two-dimensional gas chromatography coupled with Time-Of-Flight mass spectrometry - *XII Italian Food Chemistry Congress* (Camerino, 24-27/09/2018)

Stilo, F., Liberto E., Cagliero C., Rubiolo P. Sgorbini B., Reichenbach S.E., Qingping T., Bicchi C., Cordero C.E. GC×GC- tandem ionization-tofms with combined untargeted and targeted (UT) fingerprinting of volatiles from extra virgin olive oil: challenges for data alignment and response normalization - *42° International Symposium on Capillary Chromatography* (Riva del Garda, 13-18/05/2018)

Stilo, F., Cialiè Rosso M., Liberto E., Carretta A., Cobelli G., Miliazza A., Giardina M., Reichenbach S.E., Qingping T., Bicchi C., Cordero C.E. Fingerprinting of extra virgin olive oil volatiles by GC×GC: method translation and metadata transfer from thermal to differential flow modulated platforms - *42° International Symposium on Capillary Chromatography* (Riva del Garda, 13-18/05/2018)

Acknowledgements