Analytica Chimica Acta 936 (2016) 245–258

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

Combined untargeted and targeted fingerprinting with comprehensive two-dimensional chromatography for volatiles and ripening indicators in olive oil

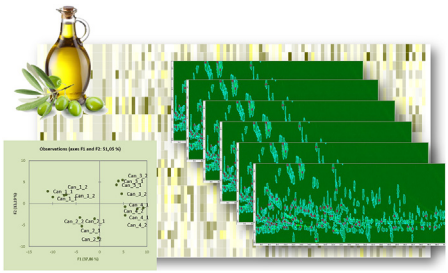
Federico Magagna [a, 1], Lucia Valverde-Som [b, 1], Cristina Ruíz-Samblás [b], Luis Cuadros-Rodríguez [b], Stephen E. Reichenbach [c], Carlo Bicchi [a], Chiara Cordero [a, *]

[a] Dipartimento di Scienza e Tecnologia del Farmaco, Università di Torino, Turin, Italy
[b] Department Analytical Chemistry, Faculty of Science, University of Granada, C/Fuentenueva s/n, 18071, Granada, Spain
[c] University of Nebraska – Lincoln, Lincoln, NE 68588-0115, USA

HIGHLIGHTS

- Fingerprinting of Extra Virgin Olive Oil volatiles by comprehensive two-dimensional gas chromatography mass spectrometry.
- Combined strategy for the most inclusive and effective fingerprinting of complex volatile fraction.
- Integrated and automated data processing for untargeted and targeted fingerprinting by peak-regions and visual features.
- Reliable and consistent retrospective analysis on 2D patterns and highly informative pair-wise comparison.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 13 May 2016
Received in revised form
27 June 2016
Accepted 1 July 2016
Available online 12 July 2016

Keywords:
Comprehensive two-dimensional gas chromatography-mass spectrometry
Untargeted and targeted fingerprinting
Extra Virgin Olive Oil
Olives ripening
Retrospective investigations

ABSTRACT

Comprehensive two-dimensional gas chromatography (GC × GC) is the most effective multidimensional separation technique for in-depth investigations of complex samples of volatiles (VOC) in food. However, each analytical run produces dense, multi-dimensional data, so elaboration and interpretation of chemical information is challenging.

This study exploits recent advances of GC × GC-MS chromatographic fingerprinting to study VOCs distributions from Extra Virgin Olive Oil (EVOO) samples of a single botanical origin (Picual), cultivated in well-defined plots in Granada (Spain), and harvested at different maturation stages. A new integrated work-flow, fully supported by dedicated and automated software tools, combines *untargeted* and *targeted* (*UT*) approaches based on peak-region features to achieve the most inclusive fingerprinting.

Combined results from untargeted and targeted methods are consistent, reliable, and informative on discriminant features (analytes) correlated with optimal ripening of olive fruits and sensory quality of EVOOs. The great flexibility of the *UT fingerprinting* here adopted enables retrospective analysis with great confidence and provides data to validate the transferability of ripening indicators ((*Z*)-3-hexenal, (*Z*)-2-hexenal, (*E*)-2-pentenal, nonanal, 6-methyl-5-hepten-2-one, octane) to external samples sets. Direct image comparison, based on *visual* features, also is investigated for quick and effective pair-wise

* Corresponding author. Dipartimento di Scienza e Tecnologia del Farmaco, Università di Torino, Via Pietro Giuria 9, I-10125 Torino, Italy.
E-mail address: chiara.cordero@unito.it (C. Cordero).
[1] Federico Magagna and Lucia Valverde-Som, listed in alphabetical order, equally contributed to this work.

http://dx.doi.org/10.1016/j.aca.2016.07.005
0003-2670/

investigations. Its implementation with reliable metadata generated by *UT fingerprinting* confirms the maturity of 2D data elaboration tools and makes advanced image processing a real perspective.

1. Introduction

Comprehensive two-dimensional gas chromatography (GC × GC) is the most effective multidimensional separation technique for in-depth investigations of complex samples of volatiles in food [1]. The combination, in a single analytical platform, of two separation dimensions with mass spectrometric detection and, when possible, automated sample preparation, delivers highly efficient sample profiling (detailed analysis of single molecular entities) and fingerprinting (rapid, high-throughput screening of samples for distinctive analytical signatures) [2].

Each analytical run produces dense, multi-dimensional data, so elaboration and interpretation of chemical information is a challenging task. In addition, food samples generally have a high-degree of chemical multidimensionality [3] thus creating highly complex analytical challenges. In this context, data elaboration strategies should implement smart and productive processes, preferably with a high degree of automation, to make cross-samples analysis efficient and informative.

Within the existing methodologies for GC × GC data elaboration [4,5], the approach based on *peak-region* features has been very effective because of its comprehensive and uniform treatment of information from each sample constituent, both knowns and unknowns. Each single chemical entity is characterized by its chromatographic and spectrometric parameters (retention time in both dimensions, detector response, and mass spectral information) and by its absolute and relative position within the pattern of all detectable constituents. As a consequence, the 2D peak-retention pattern of a sample is a diagnostic fingerprint, informative of its composition; and pattern recognition approaches can be successfully applied to improve effectiveness and productivity in multi-sample data elaboration.

Although these concepts are not new for the GC × GC community [6], the full automation of these procedures and their implementation in commercial software packages has been achieved only recently. This has limited both routine adoption of the technique for food analysis and investigative strategies for profiling [2,7].

Analysis of olive oil volatiles is a challenging and important problem and GC × GC can yield deeper knowledge of the composition of this fraction offering new perspectives for quality and authenticity assessment [8].

In spite of the great potential of GC × GC, few studies are available in this field. Vaz Freire et al. [9]first proposed an image-features approach, or more generally a pattern recognition methodology, to investigate the characteristic distribution of volatiles from oils. They adopted open-source image analysis software (Image J, National Institutes of Health) to extract information from small 2D regions located over the separation space and, by Principal Component Analysis (PCA), selected those regions with the highest discrimination potential. Then, they used targeted profiling to locate known analytes within informative 2D regions.

In 2010, Cajka and co-workers [10] exploited the targeted profiling potential of GC × GC-ToF-MS and identified 44 analytes able to discriminate samples of different geographical origin and production year. More recently, Purcaro et al. [8] combined targeted and untargeted analysis with the goal of a chemical blueprint of olive oil aroma defects. This inter-laboratory study confirmed the reliability of GC × GC for detailed profiling of olive oil volatile fractions and introduced an iterative strategy [11,12] to locate sensory-relevant analytes efficiently.

This study exploits the most recent advances of GC × GC-MS chromatographic fingerprinting to study VOC distributions from Extra Virgin Olive Oil (EVOO) samples of a single botanical origin (Picual), cultivated in well-defined plots in a single region (Granada, Spain), and harvested at different maturation stages. The principal interest in this application is the quality characteristics related to optimal ripening of olive fruits [13–21] and, as a consequence, olive oil classification and perceivable sensory quality [22,23]. In particular, this study proposes an integrated work-flow, fully supported by dedicated software tools, that performs cross-samples comparisons by contemporarily considering characteristic distributions (i.e., sample fingerprints) of both known and unknown compounds. This work-flow integrates both *untargeted* and *targeted* (*UT*) *fingerprinting* to realize the most comprehensive results, and so is termed *UT fingerprinting*. Challenges of retrospective analysis and immediacy of image fingerprinting also are discussed because of the advantages they offer in specific investigations.

2. Materials and methods

2.1. Reference compounds and solvents

Pure reference standards of α-thujone, used as Internal Standard (ISTD), at a concentration of 100 mg/L in dibuthyl phthalate, and *n*-alkanes (*n*-C9 to *n*-C25), used for linear retention index (I^T_S) determination, at a concentration of 100 mg/L in cyclohexane, were supplied by Sigma-Aldrich (Milan, Italy).

Solvents for *n*-alkanes dilution (toluene and cyclohexane HPLC-grade) and dibuthyl phthalate also were from Sigma-Aldrich.

2.2. Olive oil samples

Olive oil samples of *Picual* variety, harvested in 2014, were supplied by "GDR Altiplano de Granada" (Spain) and were obtained from olives harvested in three different plots in Granada: "812 *Caniles*" (organic production and drip irrigation); "233–234 *Baza*" (conventional production and drip irrigation); and "701 *Benamaurel*" (conventional production and drip irrigation).

Each sample was available in duplicate and obtained by mixing olives from at least five different trees in the same plot to have homogeneous and representative samples. Olives were harvested at four different ripening stages: November 10–12, 2014; November 24–28, 2014; December 16–17, 2014; and January 12–15, 2015.

Samples were analyzed by an accredited laboratory to define quality parameters: acidity (%), peroxide index (mEq O_2/kg), and UV absorption. Samples also were submitted to sensory evaluation by a recognized/official panel [24]. Sample descriptions and acronyms are reported in Table 1, together with quality assessments.

2.3. Head-space Solid Phase Micro Extraction sampling devices and conditions

Volatiles were sampled from the headspace (HS) by HS Solid

Table 1
List of analyzed samples together with plot denomination, field replicate, harvest stage, acronym, quality parameters according to COMMISSION REGULATION (EEC) No 2568/91 of 11 July 1991, sensory evaluation results, and commercial classification.

Plot	Field replicate	Harvest stage	Sample acronym	Acidity (%)	Peroxide index (mEq O_2/kg)	K_{232}	K_{270}	Δk	Md[a]	Mf[a]	Classification
Caniles	1	1	*Can_1_1*	0.2	4	2.02	0.3	0	0	7.9	EVOO
Caniles	1	2	*Can_2_1*	0.1	4	1.61	0.2	0	0	3.5	EVOO
Caniles	1	3	*Can_3_1*	0.3	4	1.41	0.2	0	0	3.6	EVOO
Caniles	1	4	*Can_4_1*	0.2	4	1.4	0.1	0	0	2.8	EVOO
Caniles	2	1	*Can_1_2*	0.2	4	2.23	0.2	0	0	6	EVOO
Caniles	2	2	*Can_2_2*	0.2	4	1.71	0.2	0	0	3	EVOO
Caniles	2	3	*Can_3_2*	0.3	4	1.38	0.2	0	0	3.4	EVOO
Caniles	2	4	*Can_4_2*	0.3	4	1.41	0.2	0	0	2.8	EVOO
Baza	1	1	*Baz_1_1*	0.2	5	1.84	0.2	0	0	5	EVOO
Baza	1	2	*Baz_2_1*	0.2	3	1.6	0.2	0	0	4.1	EVOO
Baza	1	3	*Baz_3_1*	0.2	5	1.17	0.2	0	>0.00	1.3	VOO
Baza	1	4	*Baz_4_1*	0.4	11	1.11	0.1	0	>0.00	0	LOO
Baza	2	1	*Baz_1_2*	0.2	4	1.92	0.2	0	0	5.2	EVOO
Baza	2	2	*Baz_2_2*	0.1	3	1.65	0.2	0	0	3.8	EVOO
Baza	2	3	*Baz_3_2*	0.2	6	1.28	0.1	0	>0.00	1.7	VOO
Baza	2	4	*Baz_4_2*	0.4	13	1.12	0.1	0	>0.00	0	LOO
Benamaurel	1	1	*Ben_1_1*	0.2	5	1.61	0.2	0	0	4.4	EVOO
Benamaurel	1	2	*Ben_2_1*	0.2	4	1.53	0.2	0	0	4.3	EVOO
Benamaurel	1	3	*Ben_3_1*	0.2	8	1.19	0.1	0	0	3.1	EVOO
Benamaurel	1	4	*Ben_4_1*	0.4	19	1.05	0.1	0	>0.00	0	LOO
Benamaurel	2	1	*Ben_1_2*	0.1	4	1.64	0.4	0	0	4.2	EVOO
Benamaurel	2	2	*Ben_2_2*	0.2	3	1.48	0.2	0	0	4.3	EVOO
Benamaurel	2	3	*Ben_3_2*	0.2	6	1.51	0.1	0	0	2.9	EVOO
Benamaurel	2	4	*Ben_4_2*	0.2	14	1.05	0.1	0	>0.00	0	LOO
Retrospective analysis – sampling from 2013 production											
Spain - blend	–		*R-EVO 1*	–	–	–	–	–	0	4.1	EVOO
PDO Monti Iblei - Sicily Italy			*R-EVO 2*	–	–	–	–	–	0	5.1	EVOO
PDO Dauno Gargano Puglia Italy			*R-EVO 3*	–	–	–	–	–	0	4.9	EVOO
Tuscany Italy - blend			*R-EVO 4*	–	–	–	–	–	0	4.1	EVOO
Tuscany Italy - blend			*R-EVO 5*	–	–	–	–	–	0	3.0	EVOO

[a] Median of defects (Md) - Median of fruity note (Mf).

Phase Micro Extraction (HS-SPME). The sampling protocol was optimized in a previous study [8] and employs a divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS) 50/30 μm, 2 cm length stableflex fiber from Supelco (Bellefonte, PA, USA).

The ISTD (α-tujone) was pre-loaded onto the fiber before sampling through the standard-in-fiber procedure. An ISTD solution, 2.0 μL, was placed into a 20 mL glass vial and submitted to HS-SPME at 50 °C for 15 min (min). The fiber then was exposed to the headspace of olive oil samples (1.500 g exactly weighted) in 20 mL glass vials, at 50 °C for 40 min. Last, the sampled analytes were recovered by introducing the fiber into the S/SL injection port of the GC × GC system at 260 °C and thermally desorbed for 5 min. Each sample was analyzed in duplicate.

2.4. Comprehensive two-dimensional gas chromatographic system (GC × GC-MS) set-up and analysis conditions

GC × GC analyses were performed on an Agilent 6890 unit coupled to an Agilent 5975C MS detector (Agilent, Little Falls, DE, USA) operating in EI mode at 70 eV. The GC transfer line was set at 270 °C and the MS scan range was 40–240 *m/z* with a scanning rate of 12,500 amu/s to obtain a spectra generation frequency of 30 Hz.

The system was equipped with a two-stage KT 2004 loop-type thermal modulator (Zoex Corporation, Houston, TX) cooled with liquid nitrogen and with the hot jet pulse time set at 250 ms with a modulation time of 4 s for all experiments. Fused silica capillary loop dimensions were 1.0 m long and 0.1 mm inner diameter. 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) (1 m × 0.1 mm d_c, 0.10 μm d_f) from Mega (Legnano, Milan, Italy).

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 250 °C. Carrier gas was helium at a constant flow of 1.8 mL/min. The temperature program was: from 40 °C (1 min) to 200 C at 3 °C/min and to 250 °C at 10 °C/min (5 min).

The *n*-alkanes liquid sample solution for I^T_S determination was analyzed under the following conditions: split/splitless injector in split mode, split ratio 1:50, injector temperature 280 °C, injection volume 1 μL.

2.5. Raw data acquisition and GC × GC data handling

Data were acquired by Agilent MSD ChemStation ver E.02.01.00 and processed using GC Image GC×GC Software version 2.5 (GC Image, LLC Lincoln NE, USA). Statistical analysis was performed by XLstat (Addinsoft, New York, NY USA) and the PLS Toolbox (Eigenvector Research Inc., West Eaglerock Drive, Wenatchee, WA, USA) for Matlab® software (The Mathworks Inc., Natick, MA, USA).

2.6. Profiling and advanced fingerprinting work-flow

The bi-dimensional chromatographic data elaboration proposed here was organized in a sequential work-flow illustrated in Fig. 1.

Untargeted and targeted analyses were performed by applying the *template matching* fingerprinting strategy, introduced by Reichenbach et al., in 2009 [6]. It uses the patterns of 2D peaks' metadata (retention times, MS fragmentation patterns, and detector responses) to establish reliable correspondences between the same chemical entities across multiple chromatograms. The output of template matching fingerprinting is a data matrix of aligned 2D

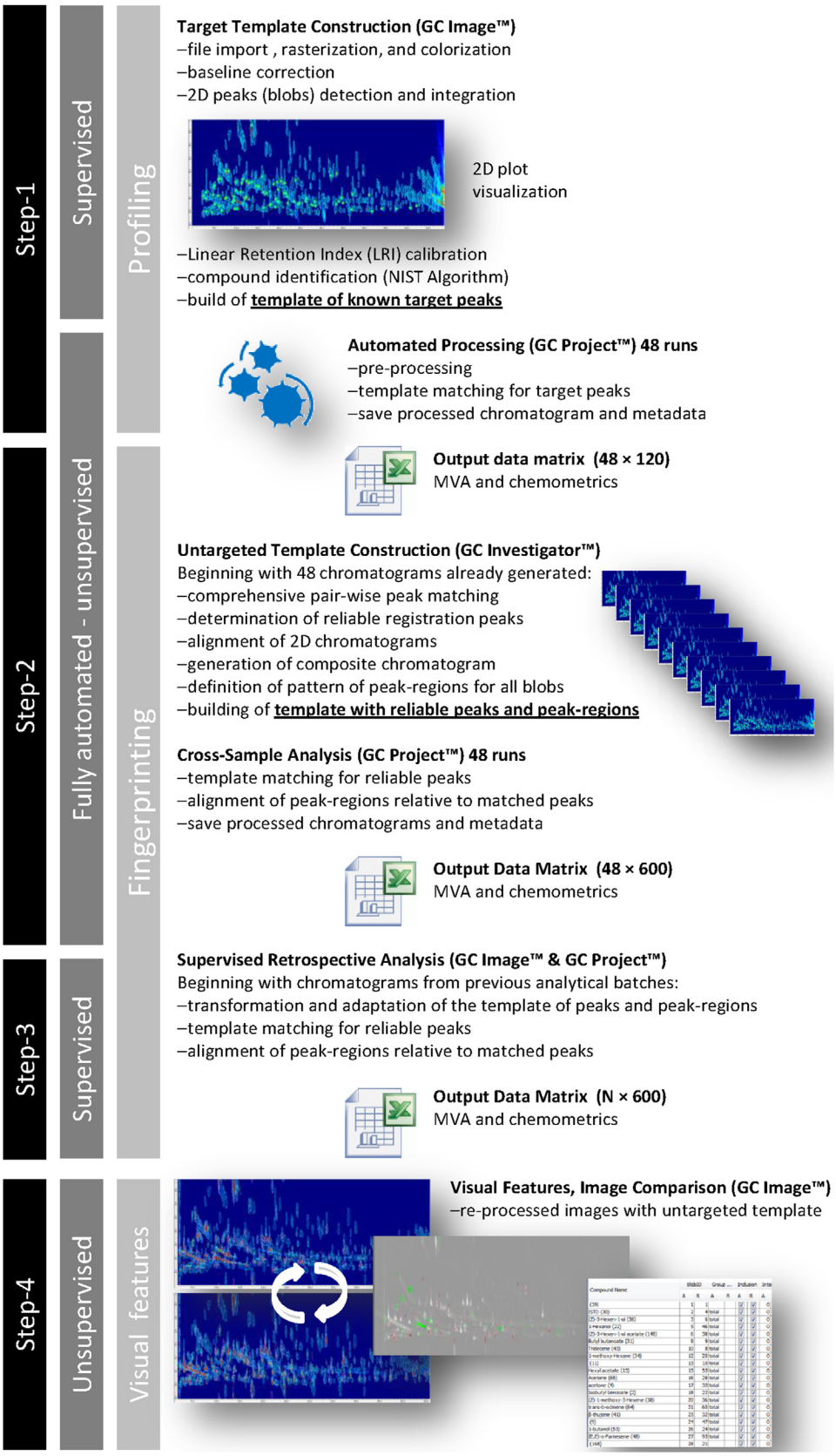


Fig. 1. Two-dimensional chromatographic data elaboration work-flow.

peaks and/or peak-regions, together with their related metadata (^{1}D and ^{2}D retention times, compound names for target analytes, fragmentation pattern, single ions or total ions response), that can be used for comparative purposes.

Targeted analysis (Step 1 of Fig. 1) focused on about 120 selected compounds, each reliably identified by matching their EI-MS fragmentation pattern (NIST MS Search algorithm, ver 2.0, National Institute of Standards and Technology, Gaithersburg, MD, USA, with Direct Matching threshold 900 and Reverse Matching threshold 950) with those collected in commercial (NIST2014 and Wiley 7n) and in-house databases. As a further parameter to support reliable identification, Linear Retention Indices (I^T_S) were considered and experimental values compared with tabulated ones.

Untargeted analysis (Step 2 of Fig. 1) was based on a *peak-regions* features approach [5] and was performed automatically by GC Image Investigator™ R2.5 (GC-Image LLC, Lincoln NE, USA). The untargeted analysis included all *peak-regions* above the arbitrarily fixed peak response threshold of 5000 counts together with target peaks from Step 1. This approach [25–28], briefly described in *Section 3.2*, re-aligned the 48 chromatograms using a set of *registration peaks*. The resulting data matrix was a 48 × 600 (samples × reliable *peak-regions*). Response data from aligned 2D *peak-regions* were used for PCA and results cross-compared to those obtained from target peaks distributions (See *Section 3.2* for the discussion of results.).

Visual features fingerprinting, performed as pair-wise image comparison, was the last step of the study (Step 4 of Fig. 1) and was rendered with "colorized fuzzy ratio" mode [cite Hollingsworth et al. JoCA 1105:51, 2006]. The algorithm computes the difference at each data point between pairs of TICs; a data point is the output of the detector at a point in time. These differences are mapped into Hue-Intensity-Saturation (HIS) color space to create an image for visualizing the relative differences between image pairs in the retention-times plane [29]. A detailed description is provided in Section 3.4.

3. Results and discussion

The goal of this study was to evaluate the potential of combining Untargeted and Targeted 2D data elaboration approaches based on untargeted *peak-region* features, target peaks, and *visual* features to approach to the most inclusive fingerprinting within EVOO volatiles: the *UT fingerprinting* strategy. Based on a sampling design focused on a single botanical variety and well-defined geographical locations, VOCs fingerprints were interpreted as a function of ripening stage and oil quality.

This strategy was inspired by a previous study focused on olive oil aroma defects [8], in which results clearly indicated that the informative potential of GC × GC-MS to delineate specific fingerprints for sensory quality classification of oils. These results showed that a "*strictly structured experimental design (considering more variables, such as cultivar, geographical origin, etc.)*" would be mandatory to "*robustly and reliably characterize specific markers and related characteristics concentration windows*" to support, or even replace, sensory evaluation [8]. In addition, it was clear that larger numbers of "external variables" affecting VOCs pattern reduce the effectiveness of untargeted approaches.

In the present study, with fewer sample-set variables, and a data elaboration process that combines untargeted and targeted approaches (i.e., *peak-region features, peaks,* and *visual features* methods), we achieve highly effective fingerprinting. The proposed work-flow is comprehensive yet efficient and fully supported by new commercial software. In addition, we validate both the data elaboration strategy and the informative role of some targets by a retrospective investigation on VOCs patterns from EVOO, VOO, and *lampante* oils (LOO) analyzed in previous studies.

Following this scheme, we first present and discuss results from targeted analysis, focusing on peaks for known informative chemicals and selected VOCs strictly related to the olives' geographical location, ripening stage, and product (oils) quality (presence/absence of sensory defects). Next, untargeted analysis based on peak-regions features, implemented in the second step, is discussed from the perspective of: (a) confirming sample classification results; (b) indicating new potential targets; (c) defining chemical indexes of ripening and quality through the ratio between informative analytes; and (d) validating the role of informative ratios through retrospective elaboration of samples analyzed in previous studies by adopting the *UT* template created on the current sample set. The last part of the study aims at determining if classification based on peak-regions features could be replaced by direct image comparison without losing information about the chemical composition of this fraction. The following paragraphs illustrate the research steps and critically discuss results.

3.1. Targeted analysis and samples discrimination

The sample set is reported in Table 1 with their quality parameters, sensory evaluation results, and commercial classification. Quality metrics (acidity %, peroxide index, UV absorbance, and organoleptic assessment) indicated that 6 of the 24 samples were not compliant with *Extra-Virgin* classification [30]. These samples, classified as *Virgin* (VOO) and *Lampante* (LOO), were from the late ripening stages of the *Baza* and *Benamaurel* plots. This quality classification was confirmed by replicate sampling (i.e., *Baz_3_1/_2 and Baz_4_1/_2; Ben_4_1/_2*) and was related to sensory defects revealed by the panel (Median of defects − Md >0.00). In addition, low-quality samples were connoted by a higher peroxide index and acidity %.

From the available literature [31–39], analytes detected in the GC × GC data were identified by their EI-MS fragmentation pattern and Linear Retention Indices (I^T_S) (see section 2.6 for details). Following the work-flow in Fig. 1, *template-matching* fingerprinting (see Paragraph *2.6*) with 119 target peaks was used to map these informative chemicals across samples.

Fig. 2A-B shows the pseudocolorized GC × GC chromatogram of an EVOO sample from the *Benamaurel* plot harvested at stage 4 (in January 2015). Fig. 2B locates the 119 known target peaks (empty light green circles) linked to the ISTD (α-tujone black circle) by red lines.

The quali-quantitative distribution of VOCs changed with the harvest stages. The number of detectable peaks above a Volume threshold of 4000 (arbitrarily fixed on the Total Ion Current signal) was about 270–280 at the first stage and reached about 360 at the final stage (data not shown).

The effectiveness of GC × GC, in both peak-capacity and overall chromatographic resolution, plays a critical role in isolating the information for compounds with similar retention times in the ^{1}D dimension. A zoomed region, highlighted in Fig. 2C, emphasizes the retention area of highly volatile compounds in which some branched hydrocarbons (eluting later in the ^{2}D) are separated along the ^{2}D from saturated and unsaturated and carbonyl compounds (e.g., pentanal, hexanal, (*E*)-2-butenal), and 1-penten-3-one, an odor-active volatile deriving from linolenic acid degradation), and alcohols (e.g., 1-propanol, 2-butanol, and 2-methyl-2-propanol).

The relative distribution (Normalized 2D Peak Volumes) of the 119 peaks is illustrated as heat-map in Supplementary Figure 1 (SF1). Columns are ordered left-to-right by ^{1}D retention indices (polar phase column, 100% polyethylene glycol). The logarithmic color map is based on 2D-Peak Volumes divided by ISTD response and is normalized by dividing single values by row standard deviations.

Table 2 reports the 119 target compounds together with their ^{1}D and ^{2}D retention times, I^T_S, sensory descriptors, and correlation

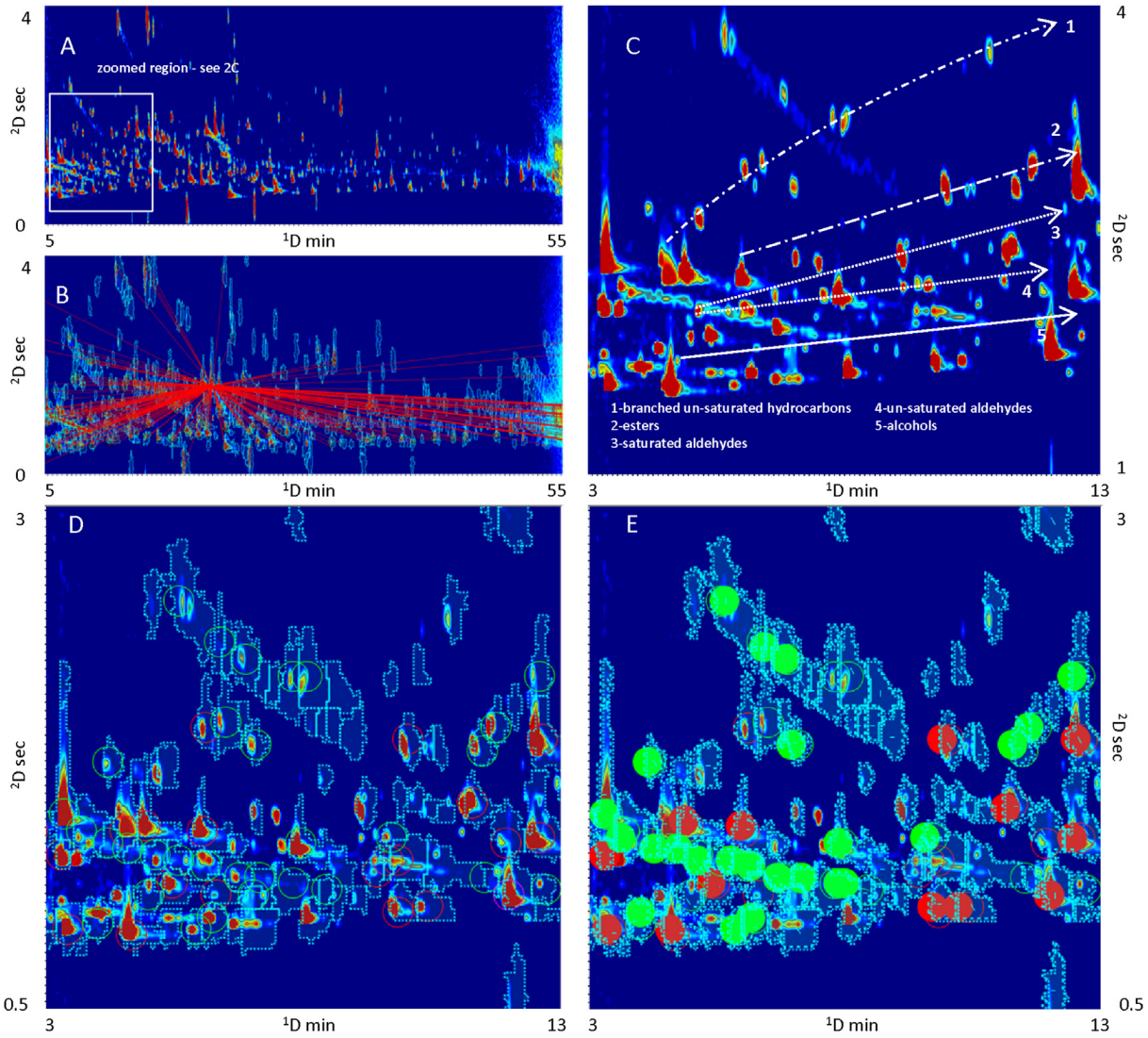


Fig. 2. A-E: (**A**) Pseudocolorized GC × GC chromatogram of *Ben_4_1* harvested at stage 4 (in January 2015). (**B**) Position of the 119 known target peaks (empty light green circles) linked to the ISTD (α-tujone black circle) by red lines. (**C**) Retention area of highly volatile compounds referred to the white rectangle of A. (**D**) Peak-regions delineated by light blue graphics together with targeted peaks (empty light blue circles). (E) Results of comprehensive template matching for peak-regions, target peaks (green circles) and registration peaks (red circles). For details see text. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

with oil defects as reported in reference literature [30–38]. 2D Peak Normalized Volumes (average values of two analytical replicates) are provided as Supplementary information (Supplementary Table 1- ST1).

The target analytes distribution (Normalized 2D volumes) was adopted as an informative fingerprint for possible discrimination of samples within different harvest stages and, in particular, to locate and validate specific indicators of ripeness, and, when feasible, odor-active compounds related to sensory quality.

Principal Component Analysis (PCA) maps the natural, unsupervised conformation of samples' groups and sub-groups [40]. Fig. 3A shows the scores plot on the first two principal components (F1–F2 plane), based on the 48 × 119 matrix (samples × targets). The variance from the first principal component (F1) was 30.64% while for the second principal component (F2) was 10.06%. Autoscaling and mean centering were applied as pre-processing methods, because baseline correction already was applied for 2D data elaboration by GC Image. The corresponding loading plots are available as Supplementary information (Supplementary Figure SF2A).

The PCA shows a clear discrimination between EVOO and VOO (clustered together in the right side) and *LOO* samples. Additionally, a further sub-classification according to harvesting stage is evident along F2 and within EVOO samples (see arrow).

The samples' structure/classification over the PCA loading plot (see Supplementary Information SF2A) indicates those analytes that are effectively responsible for the discrimination of *lampante* oils (stage 4 of *Benamaurel* and *Baza*). These analytes include saturated (e.g., heptanal, octanal, and nonanal) and unsaturated (e.g., (*E*)-2-heptenal) aldehydes, well-known from the literature to be correlated with specific sensory defects of olive oils. Moreover, the separation of *lampante* oils is also driven by other compounds, including some alcohols (e.g., propan-1-ol, 1-octen-3-ol, heptan-1-ol, and octan-1-ol), ketones (e.g., heptan-2-one and octan-2-one) and esters (e.g., ethyl acetate, ethyl-2-methyl butanoate, and ethyl-3-methyl butanoate).

Additional insight on the targets' distribution as a function of harvest time was obtained by independently processing single subsets of samples by geographical location. In this way, all variables related to the pedoclimatic conditions and field treatments

Table 2
List of the 119 target analytes together with ^{1}D and ^{2}D retention times, I^{T}_{S} and sensory descriptors as reported in reference literature [30–38]. The 2D Peak Volume data is provided as Supplementary information in Supplementary Table 1 - ST1.

ID	Compound name	1D (min)	2D (sec)	Exp ITS	Odor descriptor
1	1,4-Pentadiene	2.69	1.08	870	–
2	Octane	3.29	0.97	893	Solvent, unpleasant
3	Acetone	3.40	0.53	897	Pungent
4	Ethyl acetate	4.02	0.59	920	Fruity, aromatic, ethereal,
5	3-Methyl butanal	4.42	0.69	935	Malty
6	2-Propanol	4.61	1.11	942	–
7	Ethanol	4.68	0.49	945	Ethanol-like
8	1-Methoxy-Hexane	4.95	1.14	955	–
9	2-Ethylfuran	4.98	0.70	956	–
10	3,4-Diethyl-1,5-hexadiene (RS + SR)[a]	5.19	1.44	964	–
11	3,4-Diethyl-1,5-hexadiene (meso)[a]	5.35	1.41	970	–
12	Pentanal	5.47	0.79	974	Almond-like, pungent, malt
13	2-Butanone	5.49	0.59	975	Ethereal, fruity
14	(*Z*)-1-Methoxy-3-Hexene	6.08	1.11	997	–
15	(5*Z*)-3-Ethyl-1,5-octadiene[a]	6.14	1.73	1000	–
16	1-Penten-3-one	6.37	0.73	1008	Green
17	(5*E*)-3-Ethyl-1,5-octadiene[a]	6.55	1.77	1015	–
18	1-Propanol	6.69	0.55	1020	Pungent, pineaple
19	Ethyl butanoate	6.69	1.00	1020	Fruity
20	Tetrahydro-Furan	6.98	0.55	1031	–
21	(*E*)-2-Butenal	6.99	0.66	1031	–
22	Ethyl 2-methylbutyrate	7.14	1.22	1037	Fruity
23	Ethyl 3-methylbutyrate	7.55	1.21	1053	Fruity
24	Butyl acetate	7.74	1.01	1059	Pear
25	(*E*,*Z*)-3,7-Decadiene	7.91	2.01	1066	–
26	Hexanal	8.03	1.01	1070	Tallowy, leaf-like
27	2-Methyl-2-propanol	8.17	0.59	1076	–
28	(*E*,*E*)-3,7-Decadiene	8.21	2.04	1077	–
29	3-Methylbutyl acetate	9.28	1.24	1117	–
30	(*E*)-2-Pentenal	9.54	0.87	1127	Pungent, apple-like
31	Xylene	9.82	1.08	1137	Plastic
32	(*Z*)-3-Hexenal	9.87	0.93	1139	Leaf-like
33	1-Butanol	9.98	0.63	1144	Winey
34	(*E*)-3-Hexenal	10.01	0.93	1145	–
35	Butyl 2-methylpropanoate	10.19	1.63	1151	–
36	ethyl-Benzene	10.33	1.33	1157	–
37	1-Penten-3-ol	10.53	0.64	1164	Pungent
38	2-Heptanone	11.41	1.26	1197	Sweet, fruity
39	Heptanal	11.50	1.27	1201	Oily, fatty, woody
40	2-Ethyl-hexanal	11.60	1.59	1204	–
41	Limonene	11.92	1.73	1216	Citrus, mint
42	(*Z*)-2-Hexenal	12.17	1.04	1225	Fruity
43	3-Methyl-1-butanol	12.28	0.69	1230	–
44	(*E*)-2-Hexenal	12.75	1.07	1211	Bitter almond, green
45	Butyl butanoate	12.82	1.65	1213	–
46	δ-3-carene	13.46	1.59	1231	–
47	Tridecane	13.87	3.77	1243	–
48	1-Pentanol	13.98	0.69	1246	Sweet, pungent
49	Styrene	14.09	0.93	1249	Balsamic, gasoline
50	Trans-b-ocimene	14.09	1.63	1249	–
51	Hexyl acetate	14.99	1.50	1274	Fruity
52	3-Hydroxy-2-butanone	15.19	0.66	1280	Buttery
53	2-Octanone	15.47	1.46	1288	Mould, green
54	Octanal	15.64	1.48	1293	Fatty, sharp
55	2-Ethyl-2-hexenal	15.99	1.47	1303	–
56	(*E*)-4,8-Dimethyl-1,3,7-nonatriene	16.40	1.86	1301	–
57	(*E*)-2-Penten-1-ol	16.47	0.64	1303	Mushroom
58	1-Penten-3-ol	16.49	0.66	1303	–
59	(*Z*)-3-Hexen-1-ol acetate	16.73	1.29	1309	Sweet
60	(*E*)-3-Hexen-1-ol acetate	16.80	1.26	1311	–
61	(*Z*)-2-Penten-1-ol	16.81	0.66	1311	Green, plastic, rubber
62	(*E*)-2-heptenal	17.04	1.22	1317	Fatty, almond-like
63	6-Methyl-5-hepten-2-one	17.56	1.22	1330	Pungent, green
64	1-Hexanol	18.38	0.75	1351	Fruity, banana, soft
65	4-Hydroxy-4-methylpentanone	18.50	0.78	1354	–
66	(*E*)-3-Hexen-1-ol	18.62	0.72	1357	–
67	cis-β-Ocimene	19.04	1.59	1367	Sweat, herb
68	(*Z*)-3-Hexen-1-ol	19.44	0.73	1377	Leaf-like
69	Benzyl methyl ether	19.62	1.00	1382	–
70	Nonanal	19.95	1.64	1390	Fatty, waxy, pungent
71	(*E*,*Z*)-2,4-Hexadienal	20.11	0.89	1394	Green
72	1-Hepten-3-ol	20.12	1.36	1394	–

(*continued on next page*)

Table 2 (*continued*)

ID	Compound name	1D (min)	2D (sec)	Exp ITS	Odor descriptor
73	(*E,E*)-2,4-Hexadienal	20.37	0.89	1400	Green
74	(*E*)-2-Hexen-1-ol	20.37	0.73	1400	Green grass, leaves
75	(*Z*)-2-Hexen-1-ol	20.51	0.71	1404	Green grass, leaves
76	α-Thujone (ISTD1)	20.86	1.69	1406	–
77	1-ethenyl-3-ethyl-Benzene	21.57	1.22	1422	–
78	β-Thujone (ISTD2)	21.65	1.61	1424	–
79	1-ethenyl-4-ethyl-Benzene	21.98	1.22	1432	–
80	1-Octen-3-ol	22.27	0.89	1439	Mould, earthy
81	Acetic acid	22.34	0.48	1440	Sour, vinegary
82	1-Heptanol	22.47	0.85	1443	Herb
83	2-Ethyl-1-hexanol	23.87	0.94	1476	–
84	Decanal	24.26	1.73	1485	Penetrating, sweet, waxy
85	5-ethyl-2(3H)-Furanone	24.67	0.87	1495	–
86	Benzaldehyde	24.91	0.79	1505	Almond, burnt sugar
87	1,2-Dimethoxy propanol	25.23	0.85	1513	–
88	2,3-Butanediol	25.69	0.61	1525	–
89	Linalool	26.38	1.06	1542	Citrus
90	1-Octanol	26.63	0.91	1548	Moss, nut, mushroom
91	Dimethyl Sulfoxide	26.96	0.70	1556	–
92	Formic acid	27.22	0.59	1562	–
93	Propanoic acid	27.55	0.54	1570	Fruity, pungent
94	5-Methyl-4-hexen-3-one	28.01	0.79	1582	–
95	3-Methyl-2-hexen-4-one	28.18	0.76	1586	–
96	Undecanal	28.39	1.83	1591	Sweet, fatty, waxy-floral-citrus
97	Methyl benzoate	28.77	0.91	1604	–
98	2-Ethoxy-ethanol	28.82	0.76	1605	–
99	2-(2-Ethoxyethoxy)ethanol	28.88	0.75	1607	–
100	Butanoic acid	29.38	0.52	1620	Sweaty, rancid
101	Dodecanal	32.52	1.88	1701	Fatty, citrus-like
102	Pentanoic acid	33.58	0.56	1730	Sweaty
103	5-ethyl-2(5H)-Furanone	33.79	0.82	1736	–
104	(E,E)-α-Farnesene	33.87	2.03	1739	–
105	Methyl salicylate	34.43	0.92	1754	–
106	2-Butoxy ethanol	35.35	0.87	1779	–
107	Hexanoic acid	37.41	0.59	1837	Goat-like, sweaty
108	(*Z*)-6,10-dimethyl-5,9-Undecadien-2-one	37.55	1.60	1842	–
109	(E)-Geranylacetone	37.61	1.52	1843	–
110	Isobutyl benzoate	37.64	1.23	1844	–
111	Benzyl alcohol	37.92	0.66	1852	Sweet, fruity
112	2-Phenyl ethanol	39.11	0.72	1885	Honey-like
113	Heptanoic acid	41.00	0.60	1941	–
114	Phenol	42.24	0.58	1977	–
115	Octanoic acid	44.52	0.64	2047	Sweaty
116	2-Phenoxy ethanol	46.53	0.69	2110	–
117	Nonanoic acid	47.86	0.66	2152	Sweaty, waxy
118	Decanoic acid	51.09	0.70	2259	Soap-like, fatty
119	γ-lactone	52.60	0.78	2311	–

[a] Ref. [40].

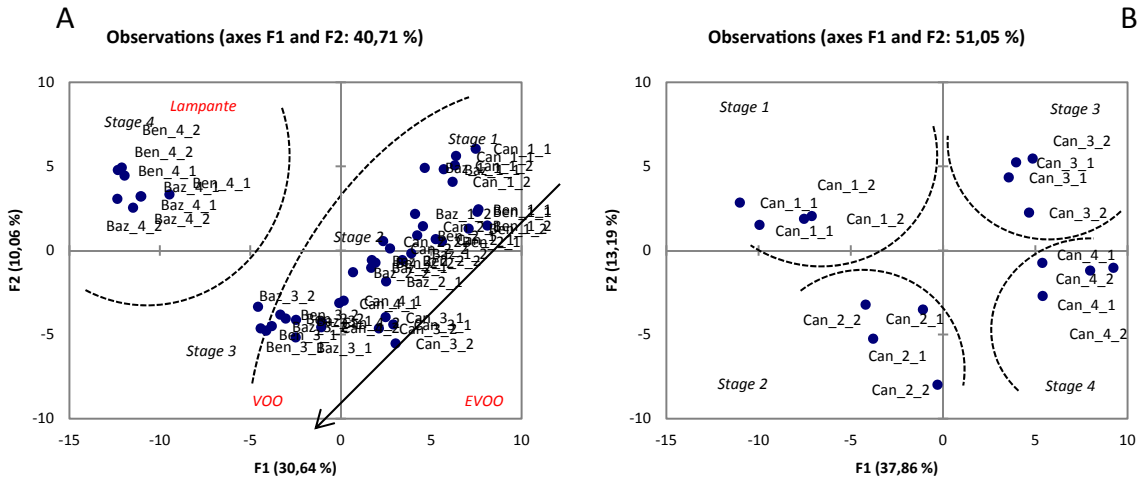


Fig. 3. A-B: PCA results. (**A**) Scores plot on the first two principal components (F1–F2 plane), based on targets distribution across all samples (48 × 119 matrix - samples × targets). (**B**) Scores plot the first two principal components (F1–F2 plane) based on targets distribution across *Caniles* EVOO subset.

(organic or conventional) were excluded, and indications on variables (markers) correlated with ripening stage are more clearly evidenced. Fig. 3B shows the score plot for the *Caniles* EVOO subset and the corresponding loadings plot is provided as Supplementary information (Supplementary Figure 2B -SF2B).

Compounds that contribute most to discriminating harvest stage 1 are: (*Z*)-2-hexenal, connoted by a *fruity* note; (*Z*)-3-hexenal, with *green* odor; and (*E*,*E*)-2,4-hexadienal, contributing a *fresh* note to the overall perception. A group of unsaturated hydrocarbons, tentatively identified from the literature [41], was found to be distinctive in the discrimination of the earlier harvest stages (1–2): 3,4-diethyl-1,5-hexadiene (RS + SR), 3,4-diethyl-1,5-hexadiene (meso), (5*Z*)-3-Ethyl-1,5-octadiene, (5*E*)-3-Ethyl-1,5-octadiene, (*E*,*Z*)-3,7-decadiene, (*E*,*E*)-3,7-decadiene, and (*E*)-4,8-Dimethyl-1,3,7-nonatriene. As expected, all these markers decrease in later ripening stages.

Interestingly, but not surprisingly, the evolution of (*Z*)-2-hexenal and (*Z*)-3-hexenal, which provide a fruity note (Mf), through time is in accordance with the sensory evaluation of the panel (as reported in Table 1). The relative abundance of these analytes shows a marked decrease from samples harvested in November (2014) to late January (2015). This observation is confirmed by data from *Baza* oils where (*Z*)-3-hexenal falls below method Limit of Detection (LOD) at stages 3 and 4 consistent with the perception of defects (Md > 0.00) leading to their classification as *lampante* oils, while (*Z*)-2-hexenal in *Benamaurel* samples was not detected even at first harvesting time.

On the other hand, some other target analytes, for example octane (*sweety*, *alcane*), nonanal (*fatty*, *waxy*), and 6-methyl-5-hepten-2-one (*pungent*, *green*), showed an opposite trend by increasing their relative abundance from stage 1 to the 4. Their presence was not revealed by the panel, possibly because of their relatively high odor thresholds (octane 940 μg/kg, nonanal 150 μg/kg, and 6-methyl-5-hepten-2-one 1000 μg/kg [22]).

These results are consistent with those reported by Aparicio and Morales [19], Raffo et al. [42] and other researchers [22,43] who hypothesized an increase of autoxidation products (e.g., octane and 6-methyl-5-hepten-2-one) accompanied by a decrease of lipoxygenase pathway products (e.g., (*Z*)-2-hexenal, (*Z*)-3-hexenal, and (*E*,*E*)-2,4-hexadienal) with later harvest times.

PCA carried out on *Baza* and *Benamaurel* oils (scores and loadings provided as Supplementary information, Supplementary Figures 3A and 3B - SF3A and SF3B) confirms, with some exceptions (e.g., in *Benamaurel* samples), the distribution of the samples and the trend of these specific chemicals over time from stage 1–4.

3.2. Untargeted analysis

Untargeted analysis was performed to extend the comparative process to the entire pattern of detected VOCs. The unsupervised fingerprinting was based on the peak-region feature approach and implemented by Image Investigator in the GC Image software package. This data elaboration step was made more informative by considering the 2D peaks included in the targeted template built within the Step 1 of the work-flow (illustrated in Fig. 1), thus preserving all information about known analytes within the fingerprinting.

The fully automated procedure of *peak-regions* fingerprinting delineates a small 2D retention-times window (or *region*) per 2D *peak* over the chromatographic space. Regions are shown in Fig. 2B and D, delineated with light blue graphics. In this context, the process approaches "one-feature-to-one-analyte" selectivity, typical of peak features methods, with all the advantages of regional features matching [5,27,28]. These advantages includes unambiguous cross-detection/matching of trace peaks that may be detected in some samples but not in others and co-eluting analytes that may be resolved in some chromatograms but not in others.

The unsupervised procedure is:

1. Detect and record 2D peaks in individual chromatograms.
2. Locate *registration* peaks, i.e., peaks that reliably match across all chromatograms (connoted by red circles in Fig. 2D and E). This is verified for a sub-group of targeted peaks.
3. Align and combine all chromatograms to create a composite chromatogram [5].
4. Define a pattern of *region* features around every 2D peak detected in the composite chromatogram.
5. Create a combined targeted and untargeted template from:
   a. the registration peaks from Step 2,
   b. the peak-regions from Step 4, and
   c. the targeted peaks.

The programmed output of the Image Investigator is the template that includes only (a) and (b). An innovation of this work is the addition of (c) targeted peaks.

Once the resulting template, as shown in Fig. 2B superimposed on the image of the *Baz_4_1* sample - analytical replicate 1, is matched to a target chromatogram, the analysis includes peak-regions (light blue graphics), targeted peaks (green circles), and registration peaks (red circles). Feature regions are aligned relative to corresponding peaks, and the characteristics of those features including all metadata (retention times in both chromatographic dimensions, detector response, relative/absolute intensity, peaks' EI-MS fragmentation pattern, response factors, etc.) are computed to create a feature vector for the target chromatogram to be adopted for cross-sample analysis. The final output is a data matrix where peak-regions and template peaks are cross-aligned within all samples' chromatograms and the response data are available for further chemometrics.

Results based on 180 reliable peak-regions (i.e., those that matched in all-but-one chromatogram of the set) are shown in Fig. 2B, and visualized by PCA of Fig. 4A. They confirmed what already was evidenced by the known targets distribution: a clear discrimination of *lampante* oils from VOO and EVOO while maintaining the sub-classification based on harvesting period. These results account for a total variability of 42%, in line with previous elaborations.

Targeted peak-regions cross-validate the classification based on PCA: (*Z*)-2-hexenal, (*Z*)-3-hexenal, (*E*,*E*)-2,4-hexadienal, 1,4-pentadiene, (5*Z*)-3-ethyl-1,5-octadiene, and (*E*)-4,8-dimethyl-1,3,7-nonatriene contribute to the discrimination of stages 1 and 2 against the others, as obtained in the previous elaboration. Untargeted analysis does not discover additional informative roles of unidentified features and confirms the coverage of the targeted peaks.

One interesting and positive aspect of these results is the strong accordance between targeted and untargeted fingerprinting in terms of sample discrimination effectiveness. This result was not observed when, for example, sampling conditions included too many variables known to impact the VOCs fingerprint (e.g., cultivars, geographical origin, harvesting period/year, technological process, bad practices etc.) [8]. In those less-controlled cases, the sensitivity and effectiveness of untargeted methodologies were lower and targeted analysis gave better results. In such cases with more experimental variables, much larger numbers of samples may be required for effective discrimination.

Another interesting outcome, in line with previous studies on flavor blueprint [8], is the accordance between sensory quality scores and samples sub-classes. Because sensory profiles by descriptive analysis were not available, a direct correlation between odor-active compounds distribution and sensory quality was not

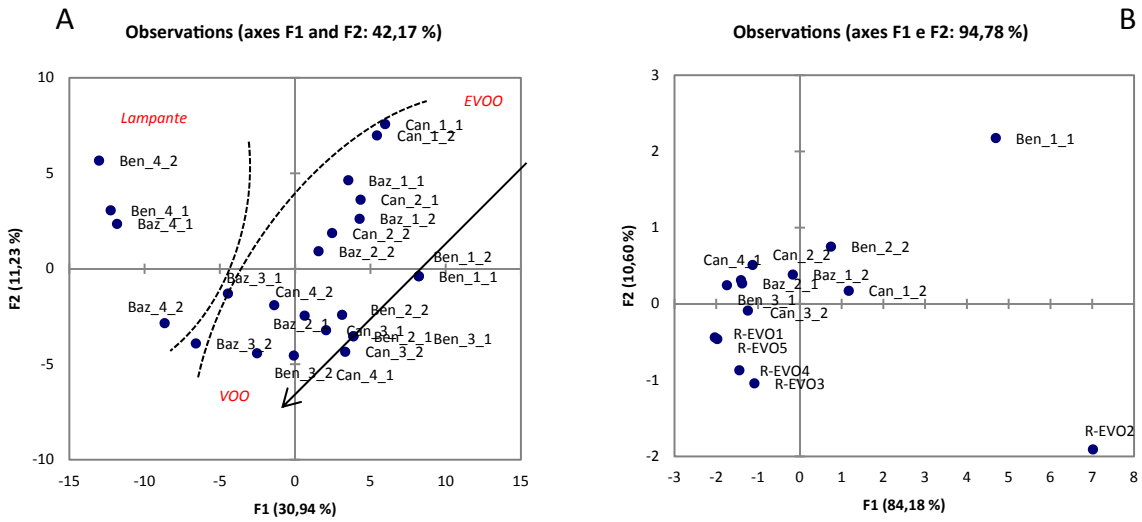


Fig. 4. A-B: PCA results. (**A**) Scores plot on the first two principal components (F1–F2 plane), based on reliable *peak-regions* distribution across all samples (48 × 180 matrix - samples × reliable *peak-regions*). (**B**) Scores plot the first two principal components (F1–F2 plane) based on informative ratios between ripening markers. For details see text.

possible. However, positive attributes (Mf in Table 1) had high scores for samples harvested at stages 1 and 2 that rapidly decreased at stages 3–4. Along the same Principal Component (e.g., F2) samples discrimination is in accordance with both variables (i.e., quality score and ripening stage).

Cross-validation of fingerprinting results reinforces and confirms the role played by some ripening markers responsible for positive attributes (green, fruity and fresh) [22,33]. These compounds appear at stage 1, last up to the stage 2, and then start to decrease. From these results, and in agreement with quality parameters (Table 1), the optimal harvest period to obtain a product with high sensory quality from *Picual* variety appears to have been November within stages 1 and 2.

Several informative analytes positively and/or negatively correlated with ripening and oil quality, were therefore selected and their ratio profiled as a function of harvest stages. In addition, a retrospective analysis on EVOO samples' pattern acquired during a previous study [8] was performed to verify the reliability and consistency of these indicators.

3.3. Retrospective analysis and definition of reliable chemical indexes of ripening

Relative ratios (based on 2D Peak Volumes) from the informative chemicals highlighted by the *UT fingerprinting* were calculated and trends observed along harvest stages.

These ratios are functions of sampling parameters (phase ratio, β; sampling temperature; and time), but derive from analyses conducted under highly standardized and head-space linearity conditions. These VOCs fingerprints are therefore informative and replicable, and these ratios could be transferred to other studies/batches and considered as chemical indices of ripening.

Analytes chosen to discriminate samples at stages 1 and 2 were: (*Z*)-2-Hexenal, (*Z*)-3-Hexenal, (*E*)-2-Pentenal, and (*E,E*)-2,4-

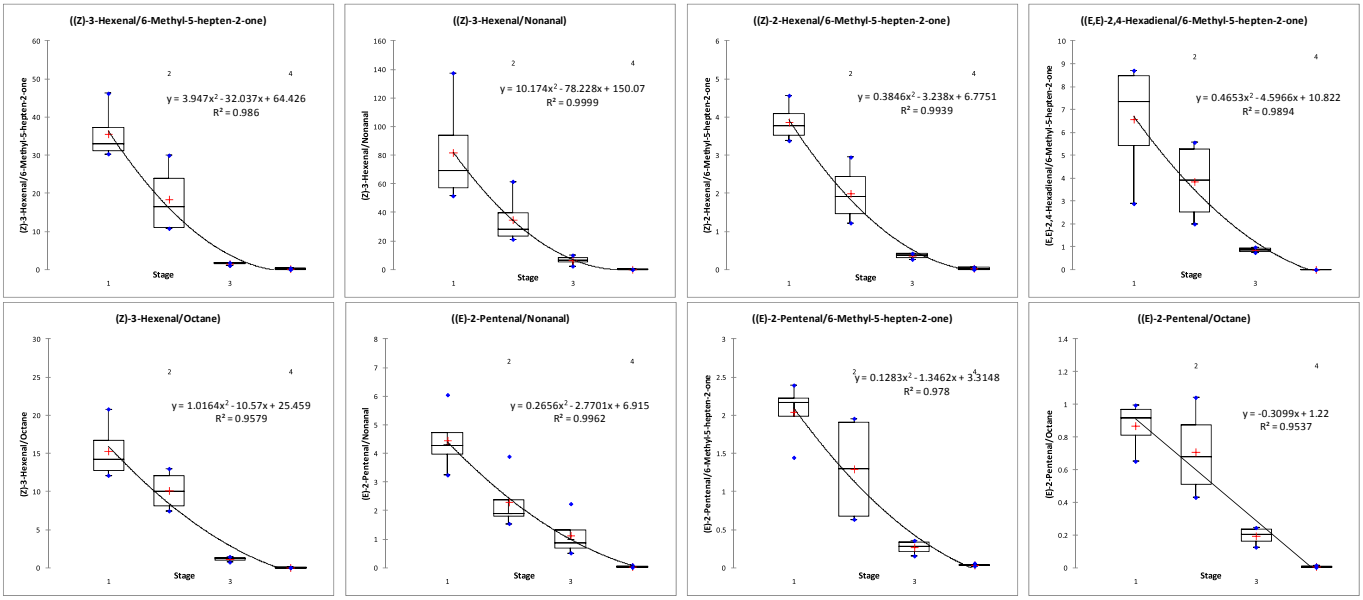


Fig. 5. Box-plot graphics showing the evolution of different informative ratios between ripening markers along harvest stages for Baza plot samples.

Table 3
Ripening informative markers evolution trends along harvest stages. The quality of fittings is referred as Coefficient of Determination (R^2).

Informative ratios	Benamaurel	R^2	Baza	R^2	Caniles	R^2
(Z)-3-Hexenal/6-Methyl-5-hepten-2-one	$y = 31.09x2 - 227.3x + 413.1$	0.999	$y = 3.947x2 - 32.03x + 64.42$	0.986	$y = 103.8e - 0.59x$	0.939
(Z)-3-Hexenal/Nonanal	$y = -4.729x2 - 21.60x + 158.3$	0.956	$y = 10.17x2 - 78.22x + 150.0$	0.999	$y = 167.7e - 0.53x$	0.956
(Z)-2-Hexenal/6-Methyl-5-hepten-2-one	$y = 2.606x2 - 18.63x + 32.95$	0.998	$y = 0.384x2 - 3.238x + 6.775$	0.993	$y = 11.66e - 0.59x$	0.959
(E.E)-2.4-Hexadienal/6-Methyl-5-hepten-2-one	$y = 236.1e - 1.71x$	0.969	$y = 0.465x2 - 4.596x + 10.82$	0.989	$y = 2.283x2 - 15.30x + 26.81$	0.998
(Z)-3-Hexenal/Octane	$y = 5.22x2 - 46.37x + 100.9$	0.987	$y = 1.016x2 - 10.57x + 25.45$	0.957	$y = -4.758x + 25.81$	0.862
(E)-2-Pentenal/Nonanal	$y = 27.60e - 1.28x$	0.794	$y = 0.265x2 - 2.770x + 6.915$	0.996	$y = x2 - 6.839x + 13.15$	0.995
(E)-2-Pentenal/6-Methyl-5-hepten-2-one	$y = 42.31e - 1.58x$	0.962	$y = 0.128x2 - 1.346x + 3.314$	0.978	$y = 5.705e - 0.51x$	0.910
(E)-2-Pentenal/Octane	$y = -0.601x + 2.250$	0.951	$y = -0.309x + 1.22$	0.953	$y = -0.303x + 1.630$	0.971

Hexadienal; those chosen that contributed to discriminate the late harvest stage were: octane, 6-methyl-5-hepten-2-one, and nonanal. Their ratios for the *Baza* samples set are illustrated by the box-plot graphics in Fig. 5 and correspond to: (*Z*)-3-Hexenal/6-Methyl-5-hepten-2-one, (*Z*)-3-Hexenal/Nonanal, (*Z*)-2-Hexenal/6-Methyl-5-hepten-2-one, (*E,E*)-2,4-Hexadienal/6-Methyl-5-hepten-2-one, (*Z*)-3-Hexenal/Octane, (*E*)-2-Pentenal/Nonanal, (*E*)-2-Pentenal/6-Methyl-5-hepten-2-one, and (*E*)-2-Pentenal/Octane. Trends were estimated by fittings with exponential, polynomial, or linear functions to delineate their evolution along harvest stage, resulting functions are reported in Table 3. The accuracy of fittings is as assessed by the determination coefficient (R^2).

As a general consideration, most of the ratios followed an exponential or second order polynomial trend with the exception of (*E*)-2-pentenal/octane index whose evolution was relatively linear. In addition, non-linear trends are connoted by higher

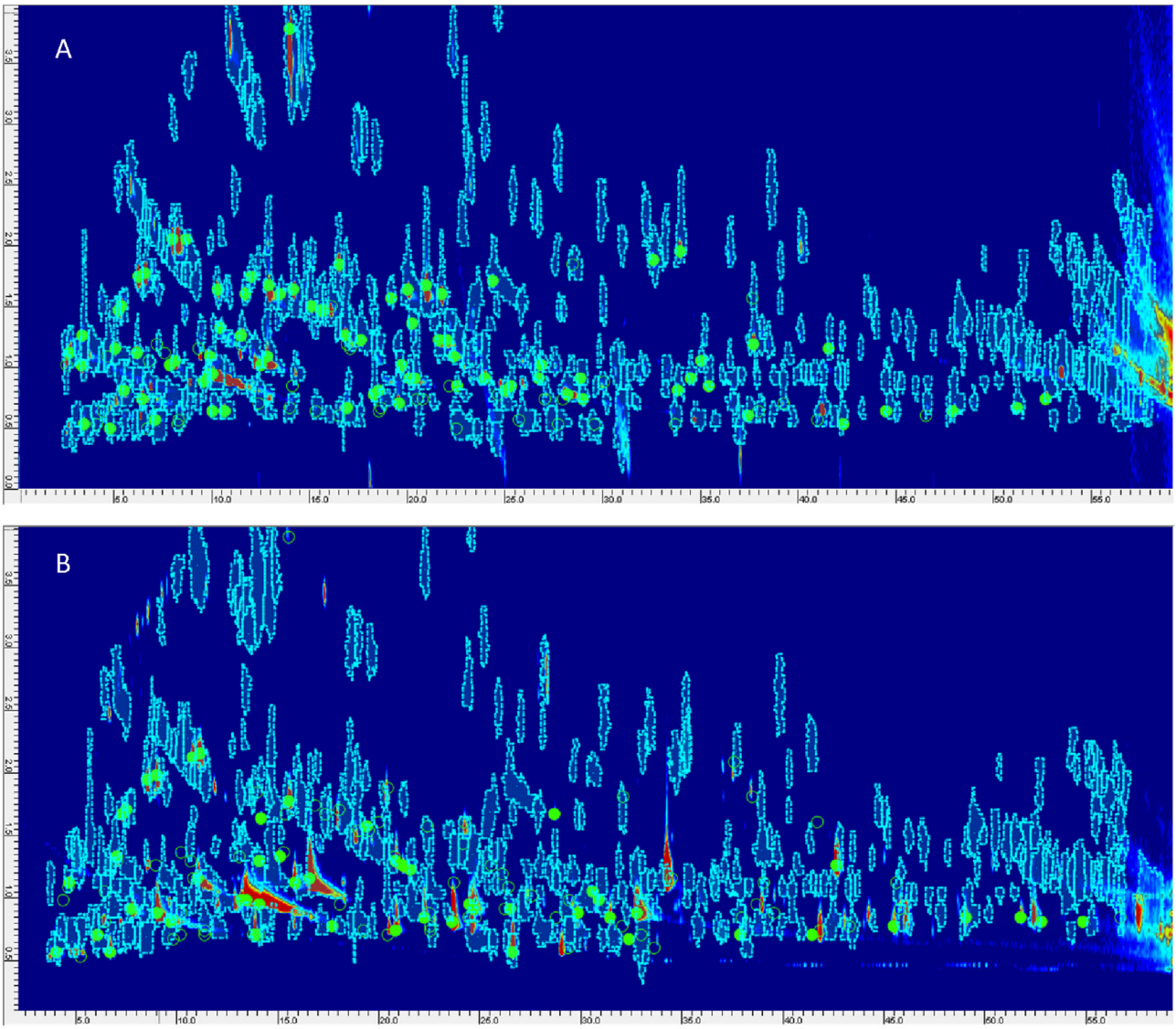


Fig. 6. A-B: (**A**) 2D chromatogram of *Can_1_2* sample together with the reliable peak-regions template. (**B**) 2D chromatogram of *R-EVOO 2* sample (PDO Monti Iblei – Sicily Italy) together with the reliable peak-regions template transformed and adapted with a supervised approach.

informative potential because of their sudden changes between optimal and non-optimal ripening stages. Notably, their numeric values decreased one order-of-magnitude between harvest stages where oil quality changed from EVO to VO or *lampante*.

The usefulness of such ratios also might be evaluated from a wider perspective where, for example, VOCs fingerprints are adopted for quality classification of EVO oils. Within selected volatiles produced during the climacteric stage of ripening [22], (*Z*)-3-hexenal is a product of the lipoxygenase pathway and, in EVO and VO oils, it contributes to the fresh aroma perception thanks to its relatively low odor threshold [20,21,34,37]. This compound is also a cultivar-specific marker for the *Picual* variety [34], as is 6-methyl-5-hepten-2-one that, as a counterpart, is connoted by a negative odor perception and an incremental trend along ripening stages. Nonanal provides information about oxidation state as well as octane [36,37,44,45].

To evaluate the consistency and the transferability of this approach for informative chemical indexes, a retrospective analysis was attempted by re-processing chromatograms from a previous study [8]. Samples consisted of EVOO from different botanical/geographical origins and technological processes and from olives harvested in 2013. They were analyzed previously, in the authors' laboratory, with the same nominal HS-SPME sampling protocol and GC × GC-MS conditions. Details are reported in Table 1.

The peak-regions template created in this study (and shown in Fig. 6A with the *Can_1_2* sample) was matched to these older, GCxGC chromatograms (as shown in Fig. 6B with the EVOO oil from Sicily PDO Monti Iblei) after a supervised transformation of the template to compensate for non-linear retention times differences in both dimensions [46].

These chromatographic inconsistencies are not infrequent because, in a time frame of two years, column sets were replaced and/or columns have altered retention behaviour (in particular the 1D PEG polar phase) producing minimal, but not negligible, pattern alterations. However, thanks to the specificity of the matching methods, 2D peaks that positively match are just those with EI-MS fragmentation pattern similarity above 700 (direct match) or 900 (reverse match). Cross-aligned results are in consequence reliable and consistent, making possible retrospective investigations.

Ratios between informative markers for the EVOO samples from *Baza*, *Caniles,* and *Benamaurel*, plus five samples from the previous study (R-EVOO from 1 to 5), were analyzed by PCA and the results are shown in Fig. 4B. The discrimination power of the first two PCs reaches 94%, confirming the informative power of the combination of variables. Picual samples (*Baz*, *Can* and *Ben*) are clustered together, with the exception of the *Benamaurel* EVOO at the earliest harvest stage, whereas along F2, there is evident discrimination for ripening stage. On the other hand, R-EVOO samples clustered together close to stages 2 and 3, with the only exception of R-EVOO 2 PDO (Monti Iblei Sicily, Italy), which showed a very high value for (Z)-3-Hexenal/6-Methyl-5-hepten-2-one because of the high abundance of (Z)-3-Hexenal (which accounted for 12% of Total Volume). The corresponding loading plot for informative rations is provides as Supplementary information Supplementary Figure 4B - SF4B.

The proposed ratios are consistent within Picual variety, but to

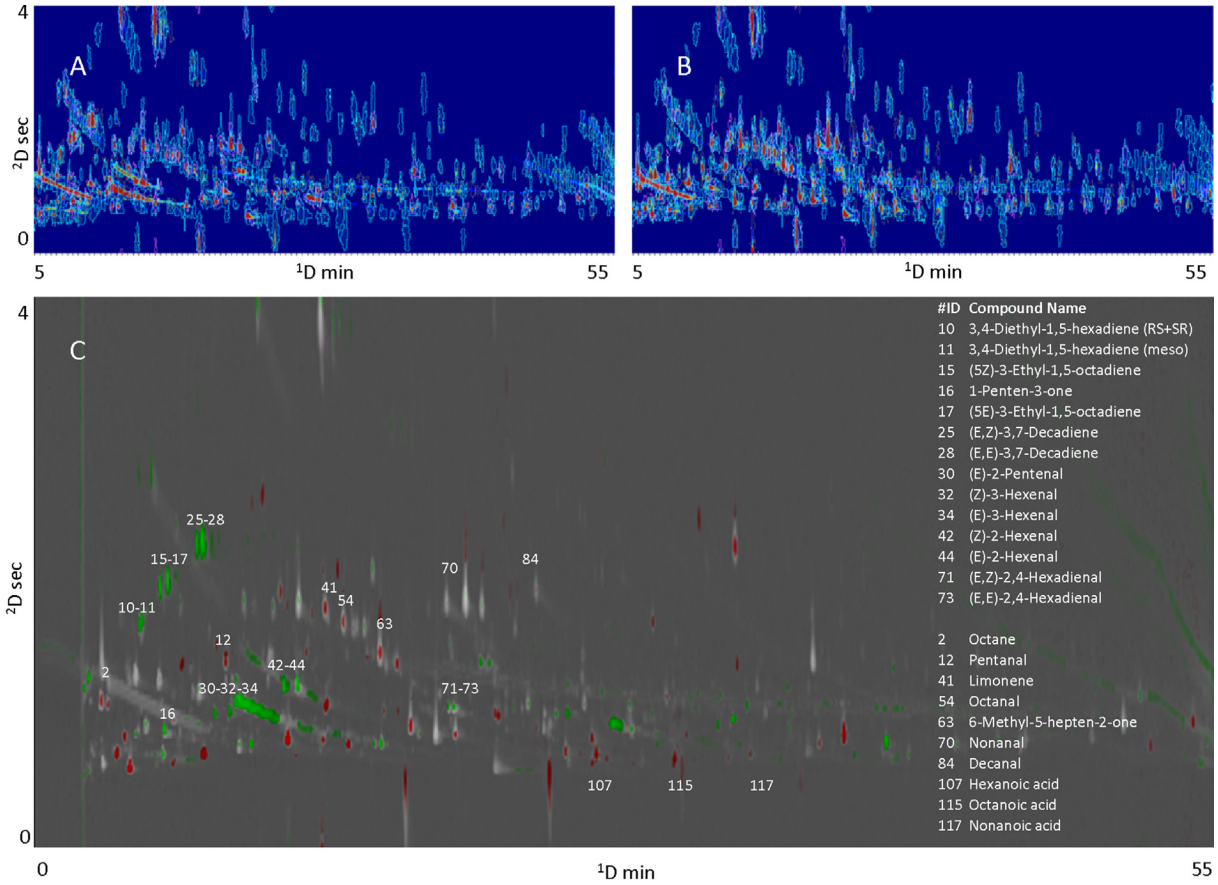


Fig. 7. A-C: (**A**) Averaged 2D-chromatogram of *Benamaurel* oil samples (field replicates and analytical replicates) obtained at stage 1 and (**B**) at stage 4. (**C**) Image comparison results between average image of harvest stage 1 (**A**) and stage 4 (**B**). The resulting image is rendered as "colorized fuzzy ratio". Analytes that varied between stages are listed together with their unique ID numbering (ref. Table 2).

be considered as general indices for ripening classification their reliability should be verified and validated by analyzing samples from different harvest years and location, and their transferability to other botanical origins and geographical locations should be investigated ex-novo by screening samples after a rigorous sampling design.

3.4. Fingerprinting by image features approach

The last part of this study focuses on a fingerprinting approach based on *visual* features and it is suitable for rapid and effective pair-wise pattern comparisons. The approach is one of the earliest introduced in GC × GC data elaboration [5], and is still adopted when distinctive patterns have to be compared on an untargeted basis to immediately reveal compositional differences.

Previous studies demonstrated the potentials of this simple and intuitive approach by exploring the volatile fraction of roasted coffee and juniper [47], volatiles emitted from *Chrysolina herbacea* bugs fed by *Mentha* spp. leaves [48], and primary metabolites distribution in mice urine after dietary manipulation [26]. The same approach was used iteratively, by cross matching sample pairs, to reveal a chemical blueprint of odor active compounds responsible of sensory defects [8].

In this application, where VOCs variations are mainly related to harvest/ripening stage, the visual approach would be effective to immediately highlight 2D peaks and/or analytes that have significantly different relative distributions between sample pairs. In addition, by comparing samples within the same production plot, the effect of fruit maturation is magnified while keeping constant the effect of local pedoclimatic changes.

This fully automated approach, namely Image Comparison (GC Image v2.5b), if implemented with peak-regions fingerprinting template, provides immediate information about targeted or untargeted peak-regions variations between pair-wise compared samples.

The example here illustrated refers to a *Benamaurel* oil sample obtained at stage 1 (averaged normalized image from *Ben_1_1 and Ben_1_2*) arbitrarily considered as the *analyzed* image *versus* the stage 4 samples (averaged normalized image from *Ben_4_1 and Ben_4_2*) arbitrarily considered as the *reference* image.

Fig. 7 shows the image comparison results between the average image of harvest stage 1 (Fig. 7A) and stage 4 (Fig. 7B) obtained by averaging the 2D chromatograms from two replicate locations and two analytical runs. The resulting image (Fig. 7C) is rendered as "colorized fuzzy ratio" that uses the Hue-Intensity-Saturation (HIS) color space to color each pixel in the retention-times plane. The algorithm computes the difference at each data point between aligned pair-wise images. If a pixel is colored green, then the difference is positive, indicating a larger detector response in the *analyzed* image (*Ben_1_1 and Ben_1_2*). If a pixel is colored red, then the difference is negative, indicating a larger detector response in the *reference* image (*Ben_4_1 and Ben_4_2*). Brightness depends on the magnitude of the difference, and so white saturation indicates pixels at which peaks have detector responses that are nearly equal in the analyzed and reference images.

Because the 2D chromatograms submitted to the image comparison were already pre-processed by *peak-region* fingerprinting, results are implemented with the information about 2D peaks' identity (if known) or unique identification numbering (#) for unknowns.

Results of *visual* features fingerprinting are intuitive and promptly give information on discriminant peaks. Green colored regions in the upper part of the 2D plot at lower ^{1}D retention correspond to unsaturated alkanes [40] (#ID 10, 11, 15, 17, 21), unsaturated aldehydes (#30 (E)-2-Pentenal, #32 (Z)-3-Hexenal, #34 (E)-3-Hexenal, #42 (Z)-2-Hexenal, #44 (E)-2-Hexenal, #71 (E,Z)-2,4-Hexadienal, and #73 (E,E)-2,4-Hexadienal), whereas red areas corresponds to limonene (#41), short chain fatty acids (#107 hexanoic, #115 octanoic and #117 nonanoic acid), linear saturated aldehydes (#12 pentanal, #54 octanal and #70 decanal), and some ketones, such as 6-methyl-5-hepten-2-one (#63).

4. Conclusions

This study evidences and emphasizes the potentials of fingerprinting based on GC × GC-MS separations and highlights the synergism between untargeted and targeted methodologies to investigate complex fractions of volatiles in depth. Their combination enables to achieve the most inclusive/comprehensive fingerprinting (*UT fingerprinting*) and if compared to previous studies, the degree of automation implemented in the data elaboration work-flow is promising. Experimental results on EVOO volatiles definitely confirm the maturity of available software tools to exploit dense and multi-level data set effectively.

The consistency and reliability of cross-sample analysis results in revealing informative/discriminant features is confirmed by matching results from different approaches, and is of interest in this challenging application field where accurate fingerprinting can be very useful: (a) to support studies aimed at improving product quality; (b) to define a distinctive chemical fingerprint to discriminate samples of a certain botanical/geographical origin; and (c) to re-investigate, on a retrospective projection, samples in light of new informative features.

Acknowledgments

The authors are grateful to "GDR Altiplano de Granada" (Spain) for olives samples.

The research was carried out thanks to the support "International mobility program for young researchers (PhD)" by University of Granada and CEI BioTic Granada.

Note: S. E. Reichenbach has a financial interest in GC Image, LLC.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.aca.2016.07.005.

References

[1] P.Q. Tranchida, P. Donato, F. Cacciola, M. Beccaria, P. Dugo, L. Mondello, Potential of comprehensive chromatography in food analysis, TrAC Trends Anal. Chem. 52 (2013) 186–205.
[2] C. Cordero, J. Kiefl, P. Schieberle, S.E. Reichenbach, C. Bicchi, Comprehensive two-dimensional gas chromatography and food sensory properties: potential and challenges, Anal. Bioanal. Chem. 407 (2015) 169–191.
[3] C. Giddings, Sample dimensionality: a predictor of order-disorder in component peak distribution in multidimensional separation, J. Chromatogr. A 703 (1995) 3–15.
[4] Z. Zeng, J. Li, H.M. Hugel, G. Xu, P.J. Marriott, Interpretation of comprehensive two-dimensional gas chromatography data using advanced chemometrics, TrAC Trends Anal. Chem. 53 (2014) 150–166.
[5] S.E. Reichenbach, X. Tian, C. Cordero, Q. Tao, Features for non-targeted cross-sample analysis with comprehensive two-dimensional chromatography, J. Chromatogr. A 1226 (2012) 140–148.
[6] S.E. Reichenbach, P.W. Carr, D.R. Stoll, Q. Tao, Smart templates for peak pattern matching with comprehensive two-dimensional liquid chromatography, J. Chromatogr. A 1216 (2009) 3458–3466.
[7] P.Q. Tranchida, G. Purcaro, M. Maimone, L. Mondello, Impact of comprehensive two-dimensional gas chromatography with mass spectrometry on food analysis, J. Sep. Sci. 39 (2016) 149–161.
[8] G. Purcaro, C. Cordero, E. Liberto, C. Bicchi, L.S. Conte, Toward a definition of blueprint of virgin olive oil by comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1334 (2015) 101–111.
[9] L.T. Vaz-Freire, M.D.R.G. da Silva, A.M.C. Freitas, 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 633 (2009) 263–270.
[10] T. Cajka, K. Riddellova, E. Klimankova, M. Cerna, F. Pudil, J. Hajslova, Traceability of olive oil based on volatiles pattern and multivariate analysis, Food Chem. 121 (2010) 282–289.
[11] C. Cordero, E. Liberto, C. Bicchi, P. Rubiolo, S.E. Reichenbach, X. Tian, Q. Tao, Targeted and non-targeted approaches for complex natural sample profiling by GC×GC -qMS, J. Chromatogr. Sci. 48 (2010) 251–261.
[12] C. Cordero, E. Liberto, C. Bicchi, P. Rubiolo, P. Schieberle, S.E. Reichenbach, Q. Tao, 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 origins, J. Chromatogr. A 1217 (2010) 5848–5858.
[13] M.N. Franco, J. Sánchez, C. De Miguel, M. Martínez, D. Martín-Vertedor, Influence of the fruit's ripeness on virgin olive oil quality, J. Oleo Sci. 64 (3) (2015) 263–273.
[14] L.C. Matos, S.C. Cunha, J.S. Amaral, J.A. Pereira, P.B. Andrade, R.M. Seabra, B.P.P. Oliveira, Chemometric characterization of three varietal olive oils (Cvs. Cobrançosa, Madural and Verdeal Transmontana) extracted from olives with different maturation indices, Food Chem. 102 (2007) 406–414.
[15] L. Martínez Nieto, G. Hodaifa, J.L. Lozano Peña, Changes in phenolic compounds and rancimat stability of olive oils from varieties of olives at stages of ripeness, J. Sci. Food Agric. 90 (2010) 2393–2398.
[16] T. Gallina-Toschi, L. Cerretani, A. Bendini, M. Bonoli-Carbognin, G. Lercker, Oxidative stability and phenolic content of virgin olive oil: an analytical approach by traditional and high resolution techniques, J. Sep. Sci. 28 (2005) 859–870.
[17] L. Nasini, P. Proietti, Olive harvesting, in: C. Peri (Ed.), The Extra-virgin Olive Oil Handbook, John Wiley Blackwell, Chichester (UK), 2014, pp. 89–105.
[18] M.N. Franco, J. Sánchez, C. De Miguel, M. Martínez, D. Martín-Vertedor, Influence of the fruit's ripeness on virgin olive oil quality, J. Oleo Sci. 64 (3) (2015) 263–273.
[19] R. Aparicio, M.T. Morales, Characterization of olive ripeness by green aroma compounds of virgin olive oil, J. Agric. Food Chem. 46 (1998) 1116–1122.
[20] M.T. Morales, R. Aparicio, J.J. Calvente, Influence of olive ripeness on the concentration of green aroma compounds in virgin olive oil, Flav. Frag. J. 11 (1996) 171–178.
[21] R. Aparicio, G. Luna, Characterisation of monovarietal virgin olive oils, Eur. J. Lipid Sci. Technol. 104 (2002) 614–627.
[22] C.M. Kalua, M.S. Allen, Jr Bedgood, A.G. Bishop, P.D. Prenzler, K. Robards, Olive oil volatile compounds, flavour development and quality: a critical review, Food. Chem. 100 (2007) 273–286.
[23] A.M. Inarejos-García, M. Santacatterina, M.D. Salvador, G. Fregapane, S. Gómez-Alonso, PDO virgin olive oil quality-minor components and organoleptic evaluation, Food Res. Int. 43 (2010) 2138–2146.
[24] Regulation EU No 1348/2013 on the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis.
[25] S.E. Reichenbach, D.W. Rempe, Q. Tao, D. Bressanello, E. Liberto, C. Bicchi, S. Balducci, C. Cordero, Alignment for comprehensive two-dimensional gas chromatography with dual secondary columns and detectors, Anal. Chem. 87 (19) (2015) 10056–10063.
[26] D. Bressanello, E. Liberto, M. Collino, S.E. Reichenbach, E. Benetti, F. Chiazza, C. Bicchi, C. Cordero, 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 1361 (2014) 265–276.
[27] S.E. Reichenbach, X. Tian, A.A. Boateng, C.A. Mullen, C. Cordero, Q. Tao, Reliable peak selection for multisample analysis with comprehensive two-dimensional chromatography, Anal. Chem. 85 (10) (2013) 4974–4981.
[28] S.E. Reichenbach, X. Tian, Q. Tao, E.B. Ledford Jr., Z. Wu, O. Fiehn, Informatics for cross-sample analysis with comprehensive two-dimensional gas chromatography and high-resolution mass spectrometry (GCxGC-HRMS), Talanta 83 (4) (2011) 1279–1288.
[29] B. Hollingsworth, S. Reichenbach, Q. Tao, A. Visvanathan, Comparative visualization for comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1105 (1–2) (2006) 51–58.
[30] COMMISSION REGULATION (EEC) No 2568/91 of 11 July 1991 on the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis and successive emendments.
[31] K. Kotsiou, M. Tasioula-Margari, Changes occurring in the volatile composition of Greek virgin olive oils during storage: oil variety influences stability, Eur. J. Lipid Sci. Technol. 117 (2015) 514–522.
[32] I. Romero, D.L. G arcía-González, R. Aparicio-Ruiz, M.T. Morales, Validation of SPME-GCMS method for the analysis of virgin olive oils volatiles responsible for sensory defects, Talanta 134 (2015) 394–401.
[33] L. Cerretani, M. Desamparados Salvador, A. Bendini, G. Fregapane, Relationship between sensory evaluation performed by Italian and Spanish official panels and volatile and phenolic profiles of virgin olive oils, Chem. Percept. 1 (2008) 258–267.
[34] G. Luna, M.T. Morales, R. Aparicio, Characterisation of 39 varietal virgin olive oils by their volatile composition, Food Chem. 98 (2006) 243–252.
[35] R. Iraqi, C. Vermeulen, A. Benzekri, A. Bouseta, S. Collin, Screening for key odorants in Moroccan green olives by gas chromatography-olfatometry/aroma extract dilution analysis, J. Agric. Food Chem. 53 (2005) 1179–1184.
[36] M.T. Morales, G. Luna, R. Aparicio, Comparative study of virgin olive oil sensory defects, Food Chem. 91 (2005) 293–301.
[37] J. Reiners, W. Grosch, Odorants of virgin olive oils with different flavor profiles, J. Agric. Food Chem. 46 (1998) 2754–2763.
[38] M. Rychlik, P. Schieberle, W. Grosch, Compilation of Odor Thersholds, Odor Qualities and Retention Indeces of Key Food Odorants, Deutsche Forschungsanstalt für Lebensmittelchemie and Institut für Lebensmittelchemie der Technischen Universität München, Garching, Germany, 1998.
[39] http://www.flavornet.org/flavornet.html [Check: 02 May 2016].
[40] R. Leardi, Chemometrics in data analysis, in: M. Lees (Ed.), Food Authenticity and Traceability, Woodhead, Cambridge, 2003.
[41] F. Angerosa, L. Camera, N. D'Alessandro, G. Mellerio, Characterization of seven new hydrocarbon compounds present in the aroma of virgin olive oils, J. Agric. Fodd. Chem. 46 (2) (1998) 648–653.
[42] A. Raffo, R. Bucci, A. D'Aloise, G. Pastore, Combined effects of reduced malaxation oxygen levels and storage time on extra-virgin olive oil volatiles investigated by a novel chemometric approach, Food Chem. 182 (2015) 257–267.
[43] M.D.R. Gomes da Silva, A.M. Costa Freitas, M.J.B. Cabrita, Garcia Raquel. Olive oil composition: volatile compounds, in: D. Boskou (Ed.), Olive Oil-constituents, Quality, Health Properties and Bioconversions, Rijeka, Croatia, 2011, pp. 17–46.
[44] V. Messina, A. Sancho, N. Walsöe de Reca, Monitoring odour of heated extra-virgin olive oils from Arbequina and Manzanilla cultivars using an electronic noise, Eur. J. Lipid Sci. Technol. 117 (2015) 1–6.
[45] A. Kanavouras, P. Hernandez-Münoz, F. Coutelieris, S. Selke, Oxidation-derived flavor compounds as quality indicators for packaged olive oil, J. Am. Oil Chem. Soc. 81 (2004) 251–257.
[46] GC Image, LLC, GC Image GCxGC Edition Users' Guide, R2.6, 2016.
[47] C. Cordero, E. Liberto, C. Bicchi, P. Rubiolo, S.E. Reichenbach, X. Tian, Q. Tao, Targeted and non-targeted approaches for complex natural sample profiling by GCxGC-qMS, J. Chromatogr. Sci. 48 (2010) 251.
[48] C. Cordero, S.A. Zebelo, G. Gnavi, A. Griglione, C. Bicchi, M.E. Maffei, P. Rubiolo, HS-SPME-GC×GC-qMS volatile metabolite profiling of Chrysolina herbacea frass and Mentha spp. leaves, Anal. Bioanal. Chem. 402 (2012) 1941–1952.