



UNIVERSIDAD
DE GRANADA



Dpto. de Química Analítica
Prof. Fermín Capitán García

Uso de huellas dactilares instrumentales para
medidas analíticas de similitud.

Aplicación en problemas relacionados con la
calidad alimentaria.

Tesis Doctoral

Programa de Doctorado en Química

Fidel Ortega Gavilán

Directores:

Dr. Luis Cuadros Rodríguez

Dra. M^a Gracia Bagur González

Para optar al título de:

Doctor en Química

Granada 2022

Editor: Universidad de Granada. Tesis Doctorales
Autor: Fidel Ortega Gavilán
ISBN: 978-84-1117-459-6
URI: <http://hdl.handle.net/10481/76804>

PRESENTACIÓN

La presente Tesis Doctoral se encuentra regida por el Real Decreto 99/2011, del 28 de enero, por el que se regulan las enseñanzas oficiales de doctorado y que está modificado por el Real Decreto 5347/2013 en lo relativo a la evaluación y defensa de la misma.

El artículo 5 de dicho Real Decreto establece que los estudios de doctorado deben asegurar la adquisición de las siguientes competencias básicas:

- a) Comprensión sistemática de un campo de estudio y dominio de las habilidades y métodos de investigación relacionados con dicho campo.*
- b) Capacidad de concebir, diseñar o crear, poner en práctica y adoptar un proceso sustancial de investigación o creación.*
- c) Capacidad para contribuir a la ampliación de las fronteras del conocimiento a través de una investigación original.*
- d) Capacidad de realizar un análisis crítico y de evaluación y síntesis de ideas nuevas y complejas.*
- e) Capacidad de comunicación con la comunidad académica y científica y con la sociedad en general acerca de sus ámbitos de conocimiento en los modos e idiomas de uso habitual en su comunidad científica internacional.*
- f) Capacidad de fomentar, en contextos académicos y profesionales, el avance científico, tecnológico, social, artístico o cultural dentro de una sociedad basada en el conocimiento.*

La obtención del título de doctor debe proporcionar una alta capacitación profesional en diversos ámbitos como aquellos relacionados con la creatividad e innovación, de forma que los doctores puedan:

- a) Desenvolverse en contextos en los que hay poca información específica.*
- b) Encontrar las preguntas claves que hay que responder para resolver un problema complejo.*
- c) Diseñar, crear, desarrollar y emprender proyectos novedosos e innovadores en su ámbito de conocimiento.*
- d) Trabajar tanto en equipo como de manera autónoma en un contexto internacional o multidisciplinar.*
- e) Integrar conocimientos, enfrentarse a la complejidad y formular juicios con información limitada.*

f) *La crítica y defensa intelectual de soluciones.*

Además, en el artículo 13 de dicho Real Decreto se especifica que:

La tesis doctoral consistirá en un trabajo original de investigación elaborado por el candidato en cualquier campo del conocimiento. La tesis debe capacitar al doctorando para el trabajo autónomo en el ámbito de la I+D+i.

Esta Tesis Doctoral se enmarca dentro del Programa de Doctorado en Química, bajo la línea de investigación "Metodologías de obtención de información analítica en sistemas reales" dentro de la Escuela de Doctorado en Ciencias, Tecnologías e Ingenierías de la Universidad de Granada.

Gran parte de las actividades realizadas durante el desarrollo de la Tesis Doctoral se ha llevado a cabo en el seno del grupo de investigación "Análisis en Alimentación y Medio Ambiente (AnAMA)" (código PAIDI: FQM 232) perteneciente al Departamento de Química Analítica de la facultad de ciencias de la Universidad de Granada bajo la dirección y supervisión de los doctores Cuadros Rodríguez y Bagur González.

Las muestras empleadas dentro del trabajo realizado en la Universidad de Granada proceden de diversos proyectos científicos concedidos al grupo de investigación AnAMA, FQM-232, y en los que el candidato a doctor ha tenido la oportunidad de trabajar y colaborar:

- "Elaboración, certificación y distribución de nueve materiales de referencia certificados (MRC) para análisis organoléptico de aceite de oliva (SENSOLEO-MRC 2017)". Proyecto financiado por la Consejería de Agricultura, Pesca y Desarrollo Rural de la Junta de Andalucía (**Ref.: OTRI 3898-2017**).
- "Sistema de referencia y armonización de paneles de cata de aceite de oliva virgen de Andalucía – INTERPANEL". Proyecto financiado por el Fondo Europeo Agrícola de Desarrollo Rural (FEADER), la Asociación Europea de Innovación en materia de productividad y sostenibilidad agrícolas (AEI-AGRI) y a la Consejería de Agricultura, Ganadería, Pesca y Desarrollo Sostenible de la Junta de Andalucía (**Ref.: GOP3I-JA-16-0012**).

Además, una parte importante del estudio experimental que constituye la presente Tesis doctoral ha sido realizada durante el periodo de estancia (3 meses) en el "Dipartimento di Scienza e Tecnologia del Farmaco" perteneciente a la "Università

degli Studi di Torino" (Universidad de Turín) bajo la supervisión de la Dra. Cordero. Esta colaboración con la Universidad de Turín ha sido posible gracias a la financiación del Programa de Movilidad Internacional 2020 para Estudiantes de Programas de Doctorado de la Universidad de Granada, concedida en el curso académico 2019/2020.

Índice

ÍNDICE

	Pag
RESUMEN	1
Resumen	3
Abstract	4
Capítulo I Introducción	5
I.1. MARCO HISTÓRICO-LEGISLATIVO DE LA CALIDAD Y SEGURIDAD ALIMENTARIA	7
I.2. CONTROL DE LA CALIDAD EN LA INDUSTRIA ALIMENTARIA	11
I.2.1. Análisis sensorial	12
I.2.1.1. Materiales de referencia en análisis sensorial	13
I.2.1.2. Tipos de análisis sensorial.....	17
I.2.1.3. Los sentidos como sistemas de detección en análisis sensorial.....	24
I.2.1.3.1. El sentido del gusto	24
I.2.1.3.2. El sentido del olfato	31
I.2.2. Análisis instrumental	35
I.2.2.1. Técnicas espectrométricas	36
I.2.2.2. Técnicas cromatográficas y electroforéticas	42
I.2.2.2.1. Cromatografía de gases (GC)	45
I.2.2.2.2. Cromatografía de líquidos (LC)	51
I.2.2.2.3. Otras técnicas instrumentales	53
I.3. ANÁLISIS DIRIGIDOS Y NO DIRIGIDOS: DE LOS (CHEMO)MARCADORES A LA UTILIZACIÓN DE LA HUELLA INSTRUMENTAL	54
I.4. QUIMIOMETRÍA Y CALIDAD ALIMENTARIA	56
I.4.1. Técnicas para el reconocimiento no supervisado de pautas	56
I.4.2. Técnicas de reconocimiento supervisado de pautas	58

I.5. PROCESADO DE LA SEÑAL INSTRUMENTAL COMO REQUISITO PREVIO PARA APLICAR LA METODOLOGÍA "FINGERPRINTING"	61
I.5.1. Hacia el agnosticismo instrumental	62
I.6. JUSTIFICACIÓN E HIPÓTESIS.....	64
I.7. OBJETIVOS	67
REFERENCIAS.....	69
Capítulo II Evaluación de la homogeneidad en materiales de referencia certificados	81
II.a. RESUMEN (EN ESPAÑOL)	84
II.b. CONCLUSIONES DEL TRABAJO (EN ESPAÑOL).....	84
II.c. ESTUDIO BIBLIOMÉTRICO	85
II.d. ARTICULO CIENTÍFICO	87
Capítulo III Desarrollo y aplicación del agnosticismo de señales instrumentales	115
III.a. RESUMEN (EN ESPAÑOL)	118
III.b. CONCLUSIONES DEL TRABAJO (EN ESPAÑOL).....	118
III.c. ESTUDIO BIBLIOMÉTRICO	120
III.d. ARTICULO CIENTÍFICO	122
Capítulo IV Certificación de aceites de oliva como materiales de referencia	151
IV.a. RESUMEN (EN ESPAÑOL)	154
IV.b. CONCLUSIONES DEL TRABAJO (EN ESPAÑOL).....	154
IV.c. ESTUDIO BIBLIOMÉTRICO	155
IV.d. ARTICULO CIENTÍFICO	157
Capítulo V Detección de mezclas de aceite de oliva a partir de los cambios en el perfil de colorantes	201
V.a. RESUMEN (EN ESPAÑOL)	204
V.b. CONCLUSIONES DEL TRABAJO (EN ESPAÑOL).....	204
V.c. ESTUDIO BIBLIOMÉTRICO	205
V.d. ARTICULO CIENTÍFICO	207

Capítulo VI	Evaluación de la calidad de la avellana con diferentes herramientas quimiométricas.....	257
VI.a.	RESUMEN (EN ESPAÑOL)	260
VI.b.	CONCLUSIONES DEL TRABAJO (EN ESPAÑOL).....	260
VI.c.	ESTUDIO BIBLIOMÉTRICO	261
VI.d.	ARTICULO CIENTÍFICO	263
Capítulo VII	Discusión integrada	289
VII.1.	INTRODUCCIÓN	291
VII.2.	PLANTEAMIENTO DE LA PROBLEMÁTICA EN LA PRIMERA CAMPAÑA DE CERTIFICACIÓN DE MATERIALES DE REFERENCIA....	293
VII.3.	DESARROLLO Y PRIMERAS APLICACIONES DE LA AGNOSTIZACION INSTRUMENTAL EN CROMATOGRAFÍA DE GASES	297
VII.3.1.	Agnosticismo aplicado separaciones cromatográficas de baja temperatura	298
VII.3.2.	Agnosticismo aplicado separaciones cromatográficas de alta temperatura	301
VII.4.	NUEVAS APLICACIONES EN CROMATOGRAFÍA DE LÍQUIDOS EN FASE NORMAL.....	303
VII.5.	EXPANDIENDO LA METODOLOGÍA A NUEVAS MATRICES ALIMENTARIAS	313
Conclusiones		317
Anexo		323

Resumen

RESUMEN

La presente Tesis Doctoral titulada "Uso de huellas dactilares instrumentales para medidas analíticas de similitud. Aplicación en problemas relacionados con la calidad alimentaria " ha profundizado en la resolución de problemas analíticos relacionados con la calidad de los alimentos como el aceite de oliva y la avellana desde dos perspectivas diferentes: (i) el desarrollo y aplicación de Materiales de Referencia Certificados y (ii) el uso de huellas instrumentales agnotizadas en cuanto a su independencia del momento de adquisición y del estado del equipo utilizado en el análisis.

El objetivo de la misma se centra en la aplicación de la metodología de huella instrumental obtenida mediante técnicas cromatográficas (GC y LC) con diferentes sistemas de detección (FID, MS y DAD) de diferentes familias de compuestos presentes en alimentos tales como compuestos orgánicos volátiles de muestras de aceite y avellanas y triglicéridos, clorofilas, feofitinas y carotenos en muestras de aceite de oliva. El interés en estas familias responde a que están fundamentalmente asociados a propiedades/atributos organolépticos.

Las matrices de datos correspondientes a huellas instrumentales tratadas bajo el enfoque de agnotización instrumental, se han estudiado aplicando diversas técnicas de reconocimiento de pautas entre las que se encuentra el análisis de componentes principales (PCA) y el análisis jerárquico de clusters (HCA) para el estudio no supervisado y modelado independiente y flexible por analogía de clases (SIMCA), análisis discriminante lineal mediante regresión lineal de mínimos cuadrados parciales (PLS-DA) y máquinas de vectores soportados (SVM) para su clasificación/discriminación como ejemplo de herramientas para el reconocimiento supervisado de pautas.

También se ha tenido en cuenta el potencial de transferencia de los resultados de la investigación de parte de los contenidos de esta Tesis. Prueba de ello es que se ha propuesto el uso de índices y perfiles de similitud como herramientas adecuadas para: (i) la armonización de paneles de cata, elementos esenciales del análisis sensorial que se aplica al sector del aceite de oliva virgen y virgen extra, y (ii) como potencial herramienta para la trazabilidad de los mismos aceites.

ABSTRACT

The present Doctoral Thesis entitled "Use of instrumental fingerprints for analytical measures of similarity. Application in problems related to food quality " has delved into the resolution of analytical problems related to food quality such as olive oil and hazelnut from two different perspectives: (i) the development and application of Certified Reference Materials and (ii) the use of agnotized instrumental fingerprints in terms of their independence from the moment of acquisition and the state of the equipment used in the analysis.

The aim of this study is to apply the instrumental fingerprint methodology obtained using chromatographic techniques (GC and LC) with different detection systems (FID, MS and DAD) of different families of compounds present in foods such as volatile organic compounds from olive oil and hazelnuts samples, and triglycerides, chlorophylls, pheophytin and carotenes in olive oil samples. The interest in these families is due to the fact that they are fundamentally associated with organoleptic properties/attributes.

The data matrices corresponding to instrumental fingerprints processed under the instrumental agnostizing approach have been investigated by applying several pattern recognition techniques including principal component analysis (PCA) and hierarchical cluster analysis (HCA) for unsupervised study and Soft independent modelling by class analogy (SIMCA), partial least squares-discriminant analysis (PLS DA) and supported vector machines (SVM) for classification/discrimination as examples of tools for supervised pattern recognition.

The potential for transferring the research results of part of the contents of this thesis has also been taken into account. As proof of this, the use of similarity indices and similarity profiles has been proposed as suitable tools for: (i) the harmonisation of tasting panels, essential elements of the sensory analysis applied to the virgin and extra virgin olive oil sector, and (ii) as a potential tool for the traceability of the oils themselves.

Capítulo I

Introducción

I.1. MARCO HISTÓRICO-LEGISLATIVO DE LA CALIDAD Y SEGURIDAD ALIMENTARIA

Hoy en día tenemos muy presente el término calidad en nuestras vidas, puesto que para la sociedad es sinónimo de productos o servicios cuya excelencia está asegurada. Este concepto puede definirse de múltiples formas, siendo una de ellas la descrita en la norma UNE-EN-ISO 9000 que la define como el *grado con que un conjunto de características inherentes de un objeto cumple con una serie de requisitos* [1]. Este concepto se ha introducido prácticamente en todos los sectores de la industria, incluido el alimentario. En este contexto, encontramos una definición específica en la Ley 28/2015, que establece que la calidad alimentaria se refiere al conjunto de propiedades y características de un producto alimenticio o alimento relativas a las materias primas o ingredientes utilizados en su elaboración, a su naturaleza, composición, pureza, identificación, origen, y trazabilidad, así como a los procesos de elaboración, almacenamiento, envasado y comercialización utilizados y a la presentación del producto final, incluyendo su contenido efectivo y la información al consumidor final especialmente el etiquetado [2].

Aunque existe una percepción de que el control de la calidad, autenticidad y seguridad de alimentos es algo novedoso, lo cierto es que se lleva poniendo en práctica desde hace siglos. La prohibición de consumo de alimentos por motivos religiosos, la inspección de determinados alimentos con el fin de asegurar su higiene y conservación, o la creación de gremios especializados para el control de materias primas, procesos de elaboración y productos finales, se pueden considerar los primeros antecedentes en el control de la calidad, autenticidad y seguridad de los alimentos [3,4].

Realmente no es hasta mediados del siglo XX, tras el fin de la segunda guerra mundial, cuando empieza a gestarse las primeras políticas sobre el control de los alimentos con el fin de asegurar su calidad, autenticidad e inocuidad. Al finalizar el conflicto armado, la industria alimentaria empezó a trabajar con determinados aditivos con el fin de obtener productos más duraderos que pudiesen paliar las carencias producidas durante la guerra. El uso descontrolado de algunos de estos aditivos hizo que las recién creadas Organización de las Naciones Unidas para la Alimentación y la

Agricultura (FAO) y la Organización Mundial de la Salud (OMS) mostrasen una preocupación sobre la calidad y seguridad alimentaria.

Tras años de preocupación finalmente ambas organizaciones crearon la Comisión del Codex Alimentarius⁽¹⁾ destinado al desarrollo de estándares globales, directrices, normas y códigos de buenas prácticas para la producción de alimentos con el fin de garantizar la seguridad del consumidor, las prácticas comerciales justas y el desarrollo del comercio global [5,6,7].

En Europa, tras la creación de la Comunidad Económica Europea (CEE) en el año 1957 (entonces denominado Mercado Común Europeo), se comenzó la regularización en el control de alimentos tras la creciente preocupación originada por los sucesivos escándalos alimentarios que afectaron a varios de sus estados miembro [8]. Partiendo de la última modificación del Codex Alimentarius, donde ya se incorporan los procesos Análisis de Peligros y Puntos Críticos de Control (APPCC) empleados por la NASA, la Unión Europea (UE) publica el Libro Verde con los principios generales de seguridad alimentaria [9]. Al poco tiempo este sería transformado en el Libro Blanco de seguridad alimentaria que sirvió como base en la creación del Reglamento (CE) nº 178/2002 [10,11,12]. Dicho reglamento establece los requisitos generales de la legislación alimentaria y además se crea la Autoridad Europea de Seguridad Alimentaria (EFSA) la cual establece una política de higiene alimentaria única y transparente aplicable a todo tipo de alimento.

Originalmente existían 6 reglamentos recogidos en el denominado "paquete higiene" obligatorios para cualquier empresa del sector alimentario que deseara vender sus productos dentro de la Unión Europea [12,13,14,15,16,17]. Sin embargo, dos de estos reglamentos, en concreto el Reglamento 854/2004 y el 882/2004 fueron derogados y sustituidos por el Reglamento 2017/625 relativo actividades de control oficial en piensos [18]. En la Tabla 1 se recogen los principales hitos aportados por cada uno de estos documentos, así como el estatus de cada uno de ellos a día de hoy.

Poco antes de la publicación de estos Reglamentos, un conjunto de entidades dentro del ámbito privado identificó la necesidad de disponer de herramientas que permitiesen asegurar la implantación y el cumplimiento de las leyes de seguridad alimentaria. Estas herramientas se denominan genéricamente "referenciales", los cuales son programas de gestión basados en principios de trazabilidad y APPCC. En

(1) Web de consulta: <https://www.fao.org/fao-who-codexalimentarius/codex-texts/list-standards/es/> (web disponible para consulta a la fecha de entrega de este manuscrito)

este contexto nace la Iniciativa Global de Seguridad Alimentaria (GFSI) la cual se encarga de reconocer programas de gestión (referenciales) que cumplen con los criterios previamente establecidos. Esta iniciativa ampara el conjunto de referenciales tales como el IFS FOOD, el FSSC 22000 o el BRC Global Standard entre otros. No todos los conjuntos de normas tienen el mismo grado de restricciones, aunque todos ellos cumplen con los requisitos generales de la legislación internacional.

Tabla 1. Resumen de los principales cambios introducidos en el Codex Alimentarius tras la publicación de distintos Reglamentos o Decretos.

Reglamento / Decreto	Hito	Estatus	Ref.
Libro verde	Plantear la necesidad de control Alimentario en Europa y proponer una serie de principios generales de seguridad alimentaria.	Evolución a libro blanco	[9]
Libro blanco	Sentar la base de la política alimentaria en Europa.	-	[10]
Reglam. (CE) nº 178-2002	Primer reglamento europeo que proporciona una base para la salud de los consumidores. Crear la EFSA.	En vigor	[12]
Reglam. (CE) nº 852-2004	Regular las obligaciones generales de los operadores de empresas alimentarias.	En vigor	[13]
Reglam. (CE) nº 853-2004	Regular los parámetros de calidad e higiene en alimentos de origen animal.	En vigor	[14]
Reglam. (CE) nº 854-2004	Establecer controles oficiales de alimentos de origen animal.	Derogada por Reg. (CE) 2017/625	[15]
Reglam. (CE) nº 882-2004	Establecer controles oficiales en piensos y alimentos para animales. Primera norma sobre salud y bienestar de animales de granja.	Derogada por Reg. (CE) 2017/625	[16]
Reglam. (CE) nº 183-2005	Regular los parámetros de higiene en piensos. Primera norma en asegurar la trazabilidad de piensos.	En vigor	[17]
Reglam. (UE) nº 2017/625	Unificar los Reglamentos (CE) 854/2004 y 882/2004 y aclarar ambigüedades mediante enfoques específicos y listas concretas de ámbitos de aplicación, en los que se incluyen productos de origen vegetal.	En vigor	[18]
Real Decreto 1086/2020	Flexibilizar algunos requisitos de seguridad e higiene, facilitando la adecuación de técnicas tradicionales.	En vigor	[19]

La legislación internacional abarca de forma general todos los procesos contemplados en la industria alimentaria, sin embargo, no ampara aquellos productos que se han obtenido mediante procesos o técnicas tradicionales propias de cada país, así como de los establecimientos donde habitualmente se venden dichos productos. Este hecho dificulta que una empresa cuyo sistema de producción emplee técnicas tradicionales pueda lograr la implementación de estos referenciales de calidad. En estos casos los propios países pertenecientes a la UE son los encargados de elaborar leyes propias con el fin de regular o aclarar algunos aspectos de los reglamentos europeos. En el caso de España, es el Real Decreto 1086/2020 el que complementa y flexibiliza determinadas condiciones de aplicación de las disposiciones de la UE dentro de su territorio, de manera que productos obtenidos siguiendo técnicas tradicionales, o que se hayan producido de manera particular puedan seguir comercializándose bajo unos criterios de seguridad y calidad [19].

La Figura 1 recoge de forma esquematizada los acontecimientos remarcables que han generado los cambios y mejoras en materia de calidad, autenticidad y seguridad de los alimentos.

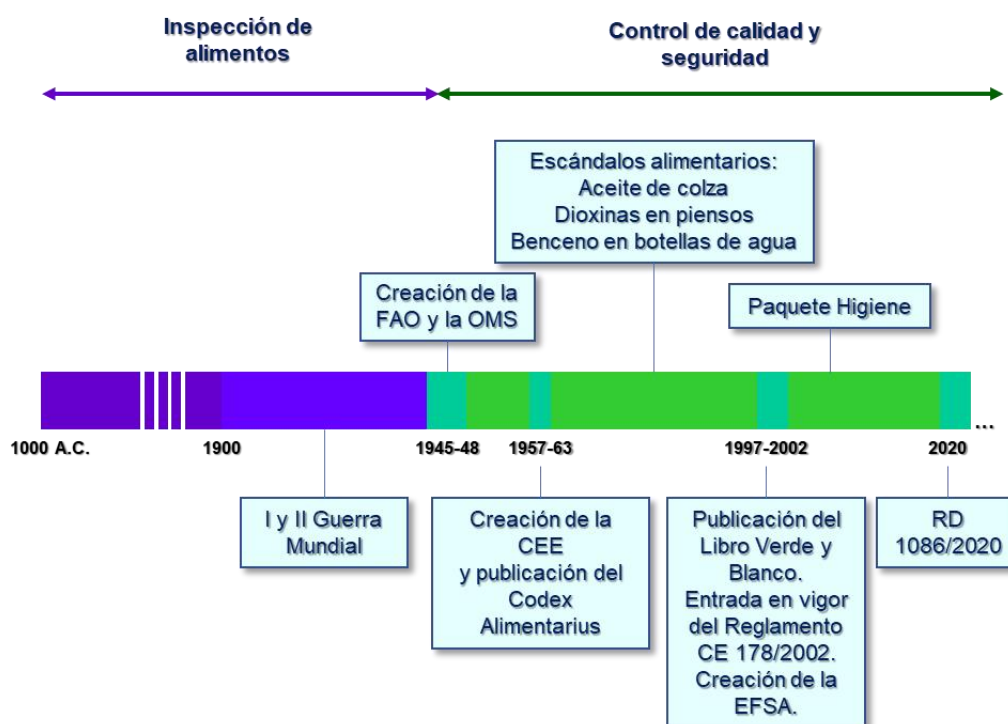


Figura 1. Línea temporal con los principales cambios acontecidos en el control de la calidad, autenticidad y seguridad alimentaria.

I.2. CONTROL DE LA CALIDAD EN LA INDUSTRIA ALIMENTARIA

Desde el punto de vista del consumidor, aún persiste el recuerdo de los escándalos alimentarios acontecidos en las últimas décadas y a día de hoy consideran que, pese a toda esta regularización, continúan apareciendo demasiados escándalos alimentarios. No es de extrañar que ante estas situaciones la sociedad experimente una creciente preocupación por conocer: i) la composición y los métodos de elaboración de los alimentos; ii) si estos cumplen con los requisitos legales considerados como seguros; y finalmente; y iii) si estos han sido o no adulterados [20].

Por su parte, la industria hace grandes esfuerzos por satisfacer las demandas de organismos y consumidores, aplicando nuevos métodos y estrategias analíticas que le permitan facilitar la información que se le exige, lo que hace del **Análisis de Alimentos** un área en continua evolución. La tipología de los análisis empleada para el control de alimentos es muy diversa, e incluye tanto métodos de carácter parcialmente subjetivo, como es el caso del análisis sensorial, o métodos basados en la determinación de propiedades biológicas, físicas o químicas que generan una señal analítica cuya medida se relaciona matemáticamente con la propiedad medida, siendo por tanto estos últimos métodos de carácter objetivo. La Figura 2 recoge un esquema general de los tipos de análisis empleados atendiendo a la objetividad del análisis.

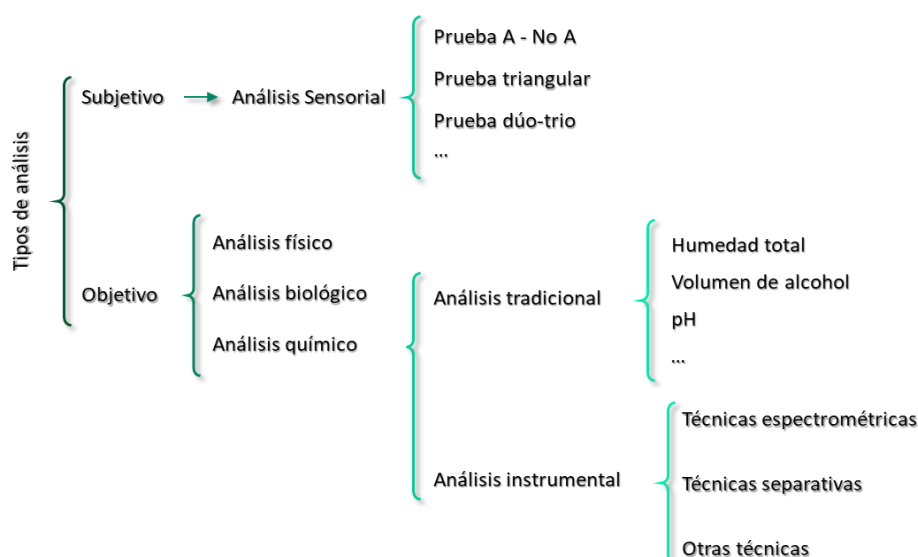


Figura 2. Esquema general de los tipos de análisis empleados la industria alimentaria para el control de la calidad y seguridad de los alimentos.

I.2.1. Análisis sensorial

El análisis sensorial consiste en la evaluación de una o varias propiedades organolépticas (olor, sabor, textura etc.) mediante el uso de los sentidos por parte de las personas que llevan a cabo dicho análisis, denominados catadores [21]. Según la norma ISO 6658, un **catador** es un asesor que evalúa una serie de atributos organolépticos, tanto positivos (parámetros de calidad) como negativos (parámetros que restan calidad al producto, también denominados defectos), con los sentidos, principalmente con la nariz y la boca. Dependiendo de la formación que ha recibido previamente el catador, se encuentran al menos tres rangos [21]:

- **Catador/asesor sensorial**, término que hace alusión a cualquier persona que participe en el análisis sensorial y que puede tener o no tener conocimientos previos sobre evaluaciones sensoriales.
- **Catador/asesor seleccionado**, relativo a personas previamente seleccionadas por sus habilidades para realizar el análisis sensorial.
- **Catador/asesor seleccionado experto**, el cual hace referencia a personas con una sensibilidad sensorial demostrada, con formación y experiencia considerable en materia de pruebas sensoriales que es capaz de proporcionar resultados coherentes y repetibles de diversos productos.

Para obtener resultados que describan de forma representativa a la muestra, el análisis sensorial requiere de la participación de más de un catador. Para ello, se constituyen grupos de trabajo denominados **paneles cata**. Si se correlaciona esta forma de trabajo con la seguida en el análisis instrumental, cada panel de cata se podría asimilar a un instrumento multisensor analítico capaz de realizar medidas cualitativas y cuantitativas de determinadas propiedades. Por tanto, al igual que sucede con cualquier instrumento analítico de medida, el panel de cata debe verificarse y/o calibrarse cada cierto tiempo para garantizar la veracidad y la precisión de los resultados. En este sentido, la verificación y calibración de paneles de cata se realiza mediante el entrenamiento de los catadores que constituyen dicho panel, utilizando en el entrenamiento patrones de referencia representativos de una o varias propiedades de interés a una intensidad definida [22].

El análisis sensorial ha demostrado ser una herramienta muy útil para entender qué características organolépticas son consideradas como parámetros de calidad en base

a los tiempos de madurez y la conservación de los alimentos; propiedades que suelen ser muy apreciados por los consumidores.

I.2.1.1. Materiales de referencia en análisis sensorial

La norma ISO Guide 30, define un material de referencia (MR) como un material lo suficientemente homogéneo y estable con respecto a una o más propiedades específicas y que se considera adecuado para su uso en proceso de medición [23]. En el caso del análisis sensorial este tipo de patrones suponen una herramienta para el control de la aptitud de los catadores que constituyen el panel y si además dichos patrones van acompañados de un certificado emitido por un organismo reconocido, donde se especifica la incertidumbre de la propiedad de interés, estos permiten mejorar la precisión de los resultados y asegurando la veracidad de los mismos [24].

En la misma norma, se define un Material de Referencia Certificado (MRC) como un MR que ha sido caracterizado por un procedimiento metrológicamente válido para una o varias propiedades especificadas y que viene acompañado de un certificado donde se proporciona el valor de la propiedad especificada, su incertidumbre asociada además de una declaración de trazabilidad metrológica [23].

De acuerdo con la propia definición de MR, el material que se postula como candidato a la certificación debe demostrar su aptitud de forma experimental, y por tanto se requiere de una cuidadosa planificación antes de llevar a cabo su producción y distribución. En esta planificación se definen los aspectos tecnológicos entre los que se establecen cuáles son las mejores condiciones de producción o de almacenamiento y en caso de matrices biológicas se definen los aspectos relacionados con las variedades o familias que puedan resultar de interés.

En la propia ISO Guide 35, también se recoge las pautas generales que deben ser tenidas en cuenta para poder certificar un material de referencia como MRC, aclarando en dicho documento que el proceso de certificación debe ser adaptado de acuerdo con la naturaleza y/o finalidad del material a certificar [25].

En cuanto a su naturaleza, estos materiales pueden encontrarse en cualquier estado de agregación, y en el caso de matrices biológicas, comúnmente utilizadas en análisis sensoriales, la propiedad de interés puede estar originada de forma natural por un proceso biológico o de forma artificial mediante la adición de sustancias químicas [24,26,27].

Cuando la propiedad de interés de un MR se ha originado mediante procesos biológicos (oxidación, fermentación, envejecimiento, etc.) que han tenido lugar de forma natural o que han sido forzados mediante formulaciones biotecnológicas se considera que dicho MR como una muestra de origen natural. Este tipo de materiales suelen ser muestras reales de alimentos, como semillas o frutos, o bien productos derivados de dichas muestras, como zumos o extractos, los cuales poseen una o varias propiedades de interés que se presentan de forma tan clara que dicha muestra se puede considerar como apta para su uso en el entrenamiento de catadores.

Las matrices biológicas también pueden ser certificadas como MR, pero antes estas deben ser sometidas a los estudios descritos en la norma ISO Guide 35 con el fin de asegurar que la propiedad de interés es homogénea en todo el conjunto y que a medida que este se utiliza, dicha propiedad permanece estable con el paso del tiempo [24,25,26]. Los MRC producidos de acuerdo con estas estrategias simulan a la perfección los atributos o defectos con que los catadores se encuentran habitualmente en las sesiones de análisis [26]. Sin embargo, la producción de MRC siguiendo esta estrategia se encuentra limitada por los propios procesos biológicos, y requiere de una búsqueda continua de nuevos candidatos que nuevamente deben ser sometidos a todo el proceso de certificación. Este cambio continuo impide tener MRC cuyas intensidades y propiedades de interés presentes en ellos sean idénticas o exactamente reproducibles año tras año [27,28].

Por otro lado, cuando la propiedad de interés se origina mediante la adición de una sustancia química a la matriz biológica durante el proceso de elaboración del candidato a MR, se dice que el mismo es de origen artificial o sintético. En la industria alimentaria, estas sustancias que modifican las percepciones sensoriales de un alimento, principalmente su flavor, se denominan aromatizantes [29].

Estos MR suelen prepararse de acuerdo a un procedimiento válido donde se estipula aspectos esenciales como el tipo de aromatizante que debe añadirse para originar una determinada sensación, la cantidad necesaria del mismo para alcanzar el nivel de intensidad deseado o el modo de preparación para lograr un material homogéneo. El uso de estas prácticas favorece la disponibilidad de MR para su uso en laboratorios, lo que facilita la incorporación de dichos materiales en el mercado internacional [27]. Sin embargo, aunque estas prácticas permiten obtener muestras reproducibles año tras año, las sensaciones producidas por la adición de aromatizantes no son exactamente idénticas a aquellas producidas como consecuencia de procesos biológicos. Este hecho repercute negativamente en el entrenamiento de catadores, ya que las

muestras que utilizan para su entrenamiento se alejan de los casos con los que habitualmente se encuentran durante el análisis.

Una vez definidos estos aspectos se procede a la obtención del material propiamente dicho, etapa en la que se incluye la adquisición, elaboración y envasado de los mismos. Finalmente, en la última etapa del proceso general es donde se pone de manifiesto la aptitud del candidato como MR mediante la ejecución de al menos tres estudios experimentales que pueden ser llevados a cabo mediante análisis sensorial o instrumental. Dichos estudios consisten en:

- (i) la evaluación de la uniformidad de la propiedad de interés dentro de cada una de las unidades envasadas como la uniformidad entre todas las unidades producidas y envasadas;
- (ii) la determinación de la estabilidad de la propiedad de interés bajo condiciones de degradación extremas o mediante la monitorización de los cambios producido por el paso del tiempo bajo las condiciones de almacenamiento estipuladas; y finalmente
- (iii) la determinación de la propiedad o propiedades de interés, comúnmente denominado como estudio de caracterización.

Existe un cuarto ensayo denominado estudio de conmutabilidad, y está relacionado con la capacidad que presenta un MR, caracterizado mediante un procedimiento de medición válido para un determinado fin, para actuar como material de control de calidad en otros ensayos en donde se utilizan procesos de medición diferentes al originalmente establecido. Estos ensayos son muy útiles para comprobar si un analito específico va a tener la misma respuesta independientemente de si la matriz en la que se encuentra es orgánica o inorgánica.

La Figura 3 muestra un esquema adaptado de la norma ISO Guide 35 donde se recoge el procedimiento genérico completo para llevar a cabo la certificación de materiales de referencia [25].

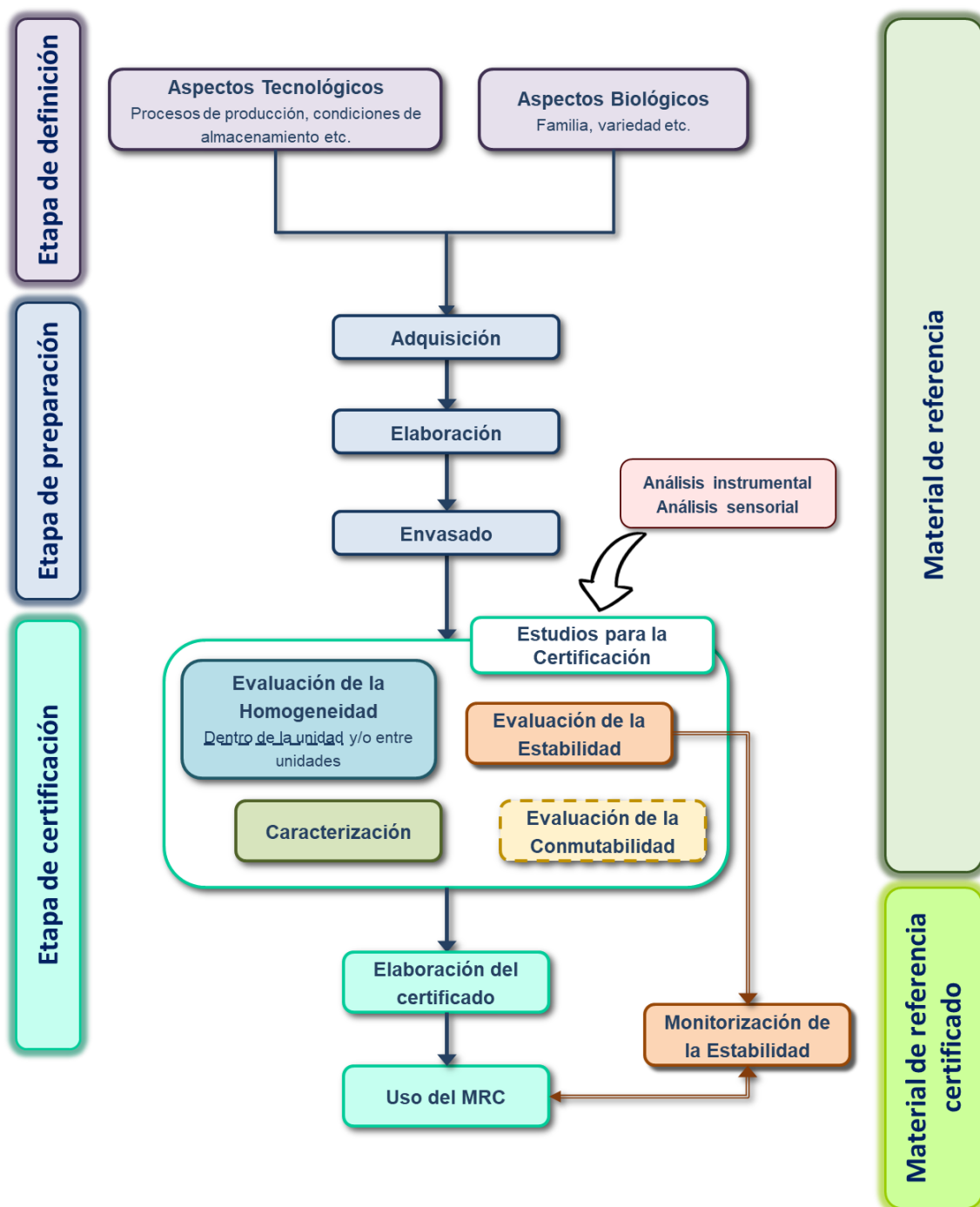


Figura 3. Procedimiento genérico del proceso para lograr la certificación de un MR. Las líneas discontinuas denotan aquellos procesos que no son obligatorios en el proceso.

I.2.1.2. Tipos de análisis sensorial

A día de hoy encontramos múltiples metodologías válidas para llevar a cabo un análisis sensorial. Tanto el número como nivel de formación de los integrantes del panel de cata dependerá de la finalidad con la que se realiza el análisis.

Si el análisis sensorial se realiza solo para conocer la preferencia del consumidor sobre un producto en concreto, los catadores que constituyen el panel no tienen por qué tener conocimientos previos sobre cómo realizar el análisis sensorial, por lo que el panel podría estar constituido por los propios consumidores. Sin embargo, cuando el análisis se aleja de la evaluación de preferencias por parte de los consumidores, se requiere que los paneles estén constituidos por catadores con cierto nivel de formación, ya sean catadores seleccionados o catadores expertos seleccionados. Esta situación se daría siempre que el análisis sensorial tenga fines [30,31]:

- a. **Discriminativos**, donde se determina la probabilidad de diferencia o similitud entre dos o más productos.
- b. **Descriptivos**, donde se caracteriza tanto cualitativa como cuantitativamente los atributos positivos o defectos específicos presentes en un producto.
- c. **Ordenamiento y/o categorización** de muestras, en donde se establece un orden a un conjunto de muestras dependiendo de la magnitud de la propiedad de estudio o de las categorías definidas previamente a las que las muestras pueden ser asignadas.

Existen múltiples test o ensayos para cada una de estas finalidades, por lo que la elección de un test frente a otro dependerá principalmente del objetivo del análisis, así como el nivel de sensibilidad estadístico deseado que determina el número de catadores necesarios para considerar válido el ensayo [31,32,33,34, 35,36,37,38,39,40,41,42]. La sensibilidad estadística es función de tres parámetros denominados como:

- (1) **riesgo α** , también conocido como ratio de falsos positivos o error tipo 1, establece la probabilidad de concluir la existencia de una diferencia perceptible cuando realmente no existe;
- (2) **riesgo β** , también conocido como ratio de falsos negativos o error tipo 2, establece la probabilidad de concluir la ausencia de una diferencia perceptible cuando realmente está presente y finalmente,

- (3) la proporción máxima en la que se detecta una diferencia perceptible entre dos productos viene determinado por el valor de P_d [43].

La Tabla 2 recoge los diferentes test que aparecen descritos en norma, y que pueden aplicarse en la evaluación sensorial, aportando en cada caso una breve descripción del método, el número mínimo de catadores expertos, así como el ámbito de aplicación de cada test.

En algunos casos, tanto el nivel de formación requerido como el número de participantes en el estudio se encuentra regulado por organismos gubernamentales con el fin de proteger la excelente calidad de sus productos y luchar contra posibles prácticas fraudulentas mediante la imposición de sanciones. Por ejemplo, la UE exige que, para la verificación de la calidad alimentaria por parte de Organismos de Control Oficial, el aceite de oliva sea sometido a diversos análisis, entre los que se incluyen el análisis sensorial, con el fin clasificar los aceites en sus diferentes categorías comerciales. El análisis sensorial de aceites de oliva se realiza de acuerdo con el método de 'Panel test' propuesto por el consejo Oleícola Internacional (COI). Dicho método estipula que la evaluación sensorial de atributos positivos y negativos detectados en nariz y boca debe ser realizada por un panel constituido por al menos 8-12 catadores específicamente entrenados para la detección de estas sensaciones [44,45].

Tabla 2. Conjunto de test o ensayos aplicados en análisis sensorial elaborados por organizaciones dedicadas a la elaboración de estándares.

Ensayos discriminativos			
Nombre del test (ISO/ASTM)	Descripción del método	Nº min cat (*)	Aplicación Ref.
Comparación por pares (ISO 5495 ASTM E2263)	Las muestras se presentan por pares para la comparación y detección de una diferencia o similitud sensorial perceptible basado en criterios definidos.	9	Test válido para llevar a cabo: - Determinación de la existencia de diferencias o similitudes en atributos/defectos específicos. - Selección, formación y monitorización de catadores - Comparación de dos productos en términos de preferencia del consumidor [32,33]
Triangular (ISO 4120 ASTME1885)	Involucra a tres muestras codificadas, siendo dos de ellas idénticas entre sí, las cuales son presentada de manera simultánea. Los catadores deben seleccionar la muestra que es diferente.	7	Test especialmente útil en los siguientes casos: - La naturaleza de la propiedad diferenciadora es desconocida - Para la selección y entrenamiento de catadores. Este test no debe ser utilizado para la determinación de preferencias. [34,35]
Dúo-trío (ISO 10399 ASTM 2618)	Los catadores trabajan con una triada de muestras, una de las cuales está etiquetada como referencia y las otras dos se encuentran codificadas, y donde una de estas es idéntica a la referencia. En base a esta información los catadores deben determinar que muestra es idéntica a la referencia.	12	Test aplicable en la determinación de diferencias o semejanzas respecto a una referencia dada, sobre todo cuando la muestra es bien conocida por los catadores. Este test también es aplicable en la selección, formación y monitorización de catadores. [36,37]

Tabla 2. (cont.) Conjunto de test o ensayos aplicados en análisis sensorial elaborados por organizaciones dedicadas a la elaboración de estándares.

Nombre del test (ISO/ASTM)	Descripción del método	Nº min cat (*)	Aplicación	Ref.
"2-out-of-5"	Los catadores deben agrupar en dos conjuntos las cinco muestras presentadas de manera simultánea, perteneciendo dos de ellas a un grupo y el resto a otro.	10	Test apropiado para establecer diferencias de la forma más económica posible. Uso frecuente en aplicaciones visuales, auditivas o táctiles. Como principal inconveniente, este test se ve afectado por la fatiga sensorial y/o efectos de memoria de los catadores.	[31]
"A" – "No A" (ISO 8588)	Los catadores deben de reconocer cuales de las muestras presentadas pertenecen a la clase "A" después de haber aprendido a reconocer esta. El resto de las muestras que no pertenezcan a la clase "A" se les denomina como "No A"	10	Test aplicable en la evaluación de muestras que presentan variaciones o que dejan sensaciones persistentes con posterioridad. El test es extremadamente útil cuando no se disponen de muestras repetidas siempre que no sean estrictamente similares, en cuyo caso se desaconseja su uso.	[38]
Tétrada (ASTM E3009)	Los catadores deben identificar que par de muestras pertenecen al mismo tipo. Para ello se presenta a los catadores cuatro muestras codificadas siendo iguales dos a dos.	11(**)	Test aplicable en la determinación de diferencias sensoriales perceptibles entre muestras o en la estimación de la magnitud de dicha diferencia cuando por ejemplo se realiza un cambio de ingrediente, método de elaboración, envasado, manipulación o almacenamiento.	[39]

Ensayos discriminativos

Tabla 2. (cont.) Conjunto de test o ensayos aplicados en análisis sensorial elaborados por organizaciones dedicadas a la elaboración de estándares.

Nombre del test (ISO/ASTM)	Descripción del método	Nº min cat. (*)	Aplicación	Ref.	
Ensayos descriptivos	Perfil sensorial cualitativo	El panel de cata entrenado lleva a cabo la descripción de los atributos sensoriales sin realizar un a medida de la intensidad organoléptica.	10	Se aplica en: - Identificación y descripción de atributos. - Establecimiento de las secuencias en que los atributos son percibidos.	[31,40]
	Perfil descriptivo cuantitativo	El panel de cata entrenado lleva a cabo la descripción de todos los atributos sensoriales cuantificando además la intensidad con la que estos aparecen. Los perfiles son específicos para cada panel y producto y no pueden ser interpretados por otro panel si previamente este no dispone de unas normas de referencia.	10	Se recomienda su uso en: - El desarrollo de nuevos productos - Establecimiento de la naturaleza de las diferencias entre productos - Control de la calidad - Para proporcionar datos sensoriales que se relacionen con datos instrumentales o proporcionados por consumidores	[31,40]
	Perfil de consenso	Los catadores evalúan individualmente las muestras y luego comparten sus opiniones individuales para lograr un consenso sobre los diferentes atributos, su orden de aparición y su intensidad	10	Se aplica en el desarrollo de nuevos productos	[31,40]

Tabla 2. (cont.) Conjunto de test o ensayos aplicados en análisis sensorial elaborados por organizaciones dedicadas a la elaboración de estándares.

Nombre del test (ISO/ASTM)	Descripción del método	Nº min cat. (*)	Aplicación	Ref.
Perfil de elección libre	Cada catador elige individualmente su propia lista de atributos sensoriales, cuantificando la intensidad mediante escalas predeterminadas por él mismo.	10	Especialmente útil para: -Elaboración de mapas perceptuales de los espacios de los productos -Estudios de mercado	[31,40]
Perfil flash	Cada catador establece los atributos para un conjunto de productos presentados simultáneamente, comparando estos productos mediante la intensidad de los atributos descritos.	10	Al igual que el Perfil de elección libre, este tipo de análisis se aplica en: -Elaboración de mapas perceptuales de los espacios de los productos - Estudios de mercado.	[31,40]
Desviación del perfil de referencia	En él los catadores comparan para cada atributo dos productos presentados de manera simultánea, calificando en cada uno de ellos el grado de diferencia.	10	Se aplica en control de calidad	[31,40]
Dominio temporal de las sensaciones	Los catadores indican de manera sucesiva y a partir de una lista común de atributos, cual es la sensación dominante y durante cuánto tiempo permanece en los sentidos.	16	Se aplica en la determinación de la duración de los efectos sensoriales independientemente de la complejidad de los mismos.	[31,40]

Ensayos descriptivos

Tabla 2. (cont.) Conjunto de test o ensayos aplicados en análisis sensorial elaborados por organizaciones dedicadas a la elaboración de estándares.

Nombre del test (ISO/ASTM)	Descripción del método	Nº min cat. (*)	Aplicación	Ref.
Clasificación (ISO 6658)	Los catadores deben ordenar las muestras en categorías previamente definidas.	-	Se aplica cuando se desea asignar diferentes muestras a la categoría que resulte más apropiada para dichas muestras.	[31]
Graduación (ISO 4121)	El panel de cata, constituido por catadores expertos, clasifica los productos en función de su calidad, evaluando atributos y defectos sensoriales percibidos en la muestra.	-	Se aplica cuando se desea asignar una clase que define cierto parámetro de calidad a las muestras analizadas.	[41]
Ranking (ISO 8587)	Los catadores deben clasificar las muestras colocándolas de acuerdo al orden de intensidad de un atributo específico.	-	Se aplica cuando es necesario realizar una caracterización rápida de un conjunto de muestras. Cuando las muestras tienen una cualidad persistente se recomienda trabajar con pequeños grupos de muestras, generalmente 6 por análisis.	[42]
Calificación y puntuación (ISO 6658)	El panel de cata asigna cada muestra a un punto de la escala ordinal, pudiendo coincidir varias muestras en el mismo punto de la escala. Este test ofrece una estimación de la magnitud independientemente de las muestras evaluadas, mientras que el método de ranking tan solo ordena las muestras en función de la intensidad.	-	Este método se aconseja antes que el test de Ranking. Se aplica en: - La evaluación de la intensidad del atributo o el grado de diferencia entre una muestra de ensayo y una muestra de referencia. - Control de la calidad.	[31]

(*) Número de catadores mínimos requeridos para los parámetros de sensibilidad α y $\beta = 0.2$ y $Pd = 0.5$

(**) De acuerdo con la norma ASTM E3009-15, se necesitan 11 catadores para los parámetros de sensibilidad α y $\beta = 0.2$ y $Pd = 1.5$

Nota. El guion "-" simboliza la ausencia de información sobre el número mínimo de catadores requeridos para llevar a cabo en dicho ensayo.

I.2.1.3. Los sentidos como sistemas de detección en análisis sensorial

Al retomar la idea ya planteada previamente, de que el panel de cata puede considerarse como un instrumento analítico de medida constituido por diferentes catadores, donde cada catador actúa como un sensor analítico que lleva a cabo la detección de las sustancias responsables de la propiedad a evaluar, se puede decir que de las diferentes propiedades que puede evaluar un panel de cata, el sabor, a menudo denominado como flavor, es la propiedad más compleja al ser el resultado de la combinación de varias propiedades sensoriales individuales.

De acuerdo con la norma ISO 5492 el flavor/sabor de un alimento se define como la combinación compleja de sensaciones olfativas, gustativas y trigeminales percibidas durante la degustación [21]. Desde el punto de vista químico existe una gran diferencia entre el conjunto de moléculas que producen las sensaciones olfativas y gustativas, las cuales quedan resumidas en la Tabla 3 [46].

Tabla 3. Propiedades generales de los compuestos químicos responsables del gusto y el olor de los alimentos.

Propiedad	Gusto	Olor (odorantes)
Punto de ebullición	Alto (baja volatilidad)	Bajo (alta volatilidad)
Solubilidad en agua	Solubles	Insolubles
Polaridad	Elevada polaridad	Poco polares
Peso molecular	Bajo – medio (entre 17 – 20000 Da)	Bajo (entre 17 – 300 Da)
Proporción en alimentos	Alta (% – ppm)	Baja (ppm – ppb)

I.2.1.3.1. El sentido del gusto

El gusto consiste en las sensaciones percibidas por la lengua como consecuencia del estímulo producido por los alimentos. Tradicionalmente estas percepciones se agrupan en cuatro aspectos básicos (agrio, amargo, dulce y salado) a partir de los cuales se constituyen el resto de sensaciones [46]. Durante gran parte del último siglo se mantuvo la creencia de que regiones concretas de la lengua se encontraban especializadas de manera específica en el reconocimiento de las diferentes sensaciones. Sin embargo, en las últimas décadas del siglo XX se publicaron numerosos artículos que rebatieron esta creencia, alegando que las sensaciones básicas se perciben en toda la superficie de la lengua de manera indiferente, aunque

existen receptores sensoriales localizados que muestran una ligera sensibilidad a determinadas sensaciones comparada con el resto y repartidas en toda la superficie de la lengua [47]. La Figura 4 muestra para cada sensación básica una distribución aproximada de los receptores en la superficie de la lengua.

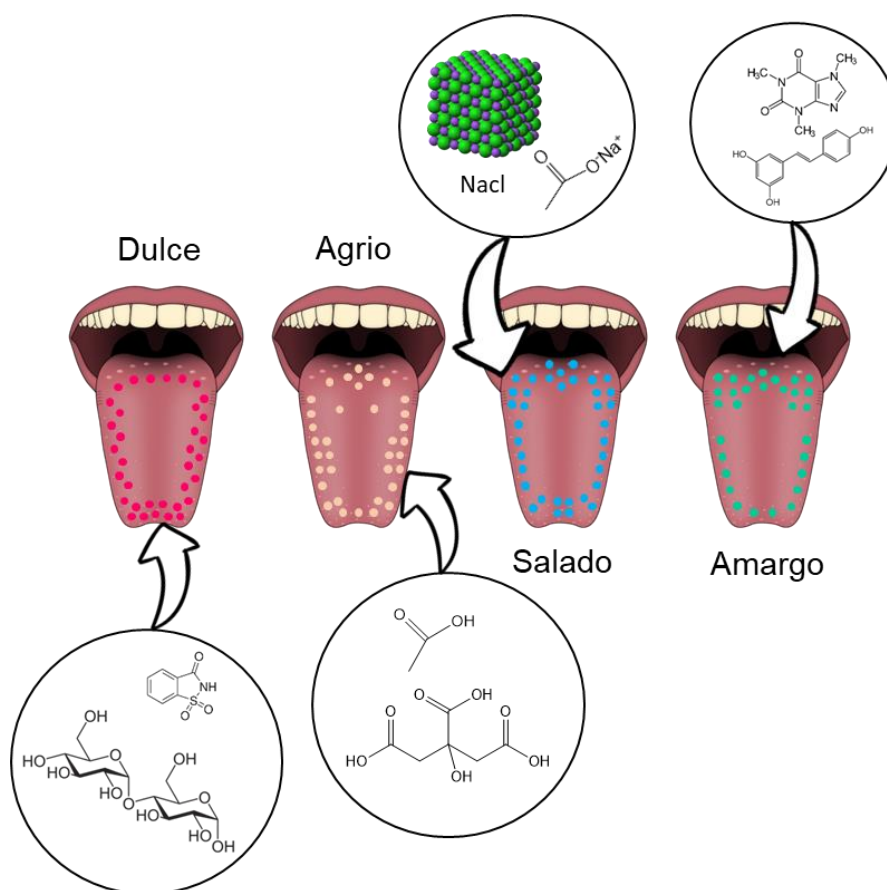


Figura 4. Distribución de receptores gustativos sensibles a las diferentes percepciones básicas en toda la superficie de la lengua.

En cuanto a la naturaleza composicional, estas sustancias están producidas por compuestos de diferente naturaleza (compuestos tanto orgánicos como inorgánicos). Como se especifica en la Tabla 3, el sabor se debe a sustancias solubles en agua con un elevado peso molecular y con una volatilidad bastante reducida comparada con otras sustancias [48].

➡ Dulzor en alimentos

El dulzor de los alimentos se debe principalmente a carbohidratos y sus derivados glucósidos que se adhieren a los receptores presentes en las papilas gustativas de

acuerdo a un modelo de enlace intermolecular antiparalelo del tipo (HA) – B, donde H se refiere a un átomo de hidrogeno y A y B pueden ser átomos de nitrógeno u oxígeno. En este modelo, recogido en la Figura 5, los grupos HA (defecto de carga positiva) y B (defecto de carga negativa) del azúcar interactúan con los correspondientes grupos B y HA respectivamente, presentes en los receptores gustativos, iniciando una reacción que se transmite hasta el cerebro [49].

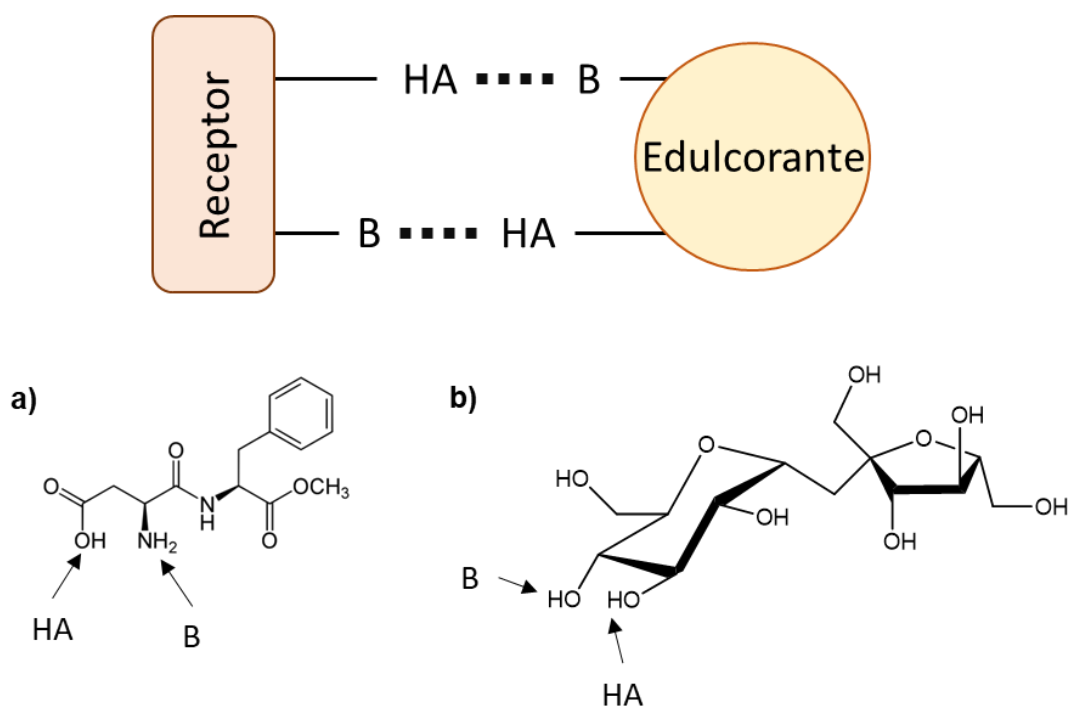


Figura 5. Modelo de unión intermolecular HA - B de edulcorantes empleados en alimentos a los receptores de dulzor presentes en la lengua, a) aspartamo y b) sacarosa.

La intensidad de dulzor percibida depende de la fortaleza de los enlaces que se forman cuando se produce la adhesión de una molécula a los receptores gustativos de acuerdo con este modelo [50]. Existe un gran número de compuestos de origen sintético, como por ejemplo el aspartamo o la sacarina, con una geometría tridimensional adecuada para formar este tipo de enlaces, transmitiendo así sensaciones de dulzor [51,52]. Es habitual que estas moléculas de origen sintético sean mucho más dulces que las de origen natural, debido a que la fortaleza de los enlaces formados durante la adhesión a los receptores gustativos es mucho mayor.

El dulzor también se ve afectado por la solubilidad y el tamaño de los compuestos encargados de transmitir esta sensación. Por ejemplo, la glucosa (monosacárido) mucho más dulce que la maltosa, (disacárido constituido por dos unidades de

glucosa), mientras que los polímeros de la glucosa, como el almidón no transmiten ningún tipo de dulzor. En la Figura 6 se recogen las estructuras de algunos compuestos responsables del dulzor de los alimentos.

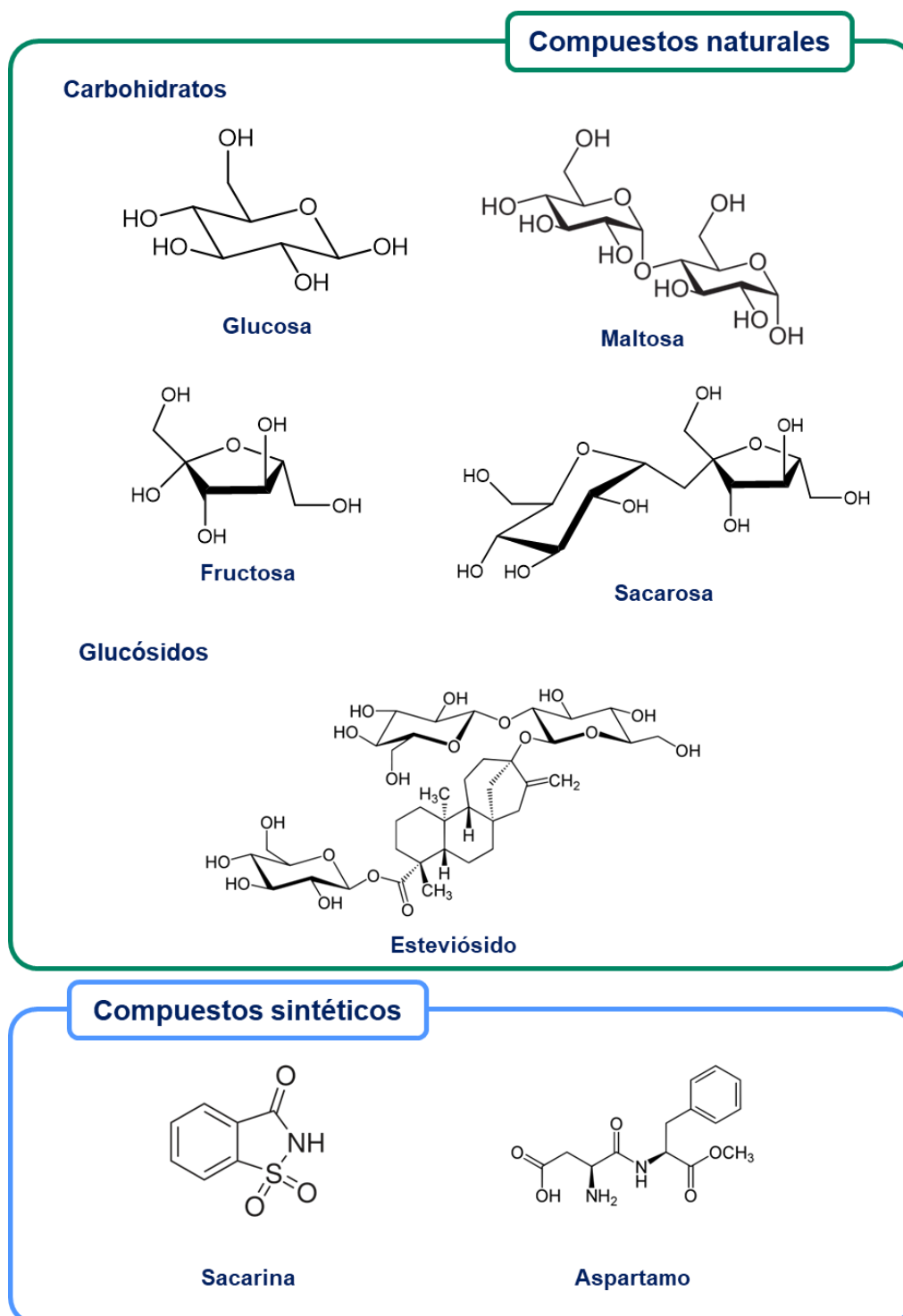


Figura 6. Estructuras de algunos compuestos responsables del dulzor de los alimentos.
(La sección recogida en el cuadro verde hace referencia a compuestos químicos de origen natural, mientras que la sección recogida en el cuadro azul hace referencia a compuestos sintéticos).

➡ Amargor en alimentos

El amargor se considera la sensación opuesta al dulzor por lo que a edades tempranas se considera como una sensación desagradable comparada con otras [53]. Este instinto se debe a la percepción del amargo como un sistema de alerta frente a la ingesta de alimentos tóxicos, por lo que es habitual que los mamíferos eviten el consumo de alimentos con un fuerte amargor [51]. El rechazo al amargor disminuye a medida que se observa que muchos alimentos amargos no son venenosos, aprendiendo a valorar esta sensación en productos como el pimienta, el café o la cerveza.

Desde el punto de vista químico, el amargor se debe principalmente a la presencia de compuestos alcaloides, compuestos heterociclos nitrogenados procedentes del metabolismo de aminoácidos en plantas, aunque otros compuestos como aminas, polifenoles, algunos aminoácidos e incluso algunos glucósidos complejos también son responsables del amargor [54]. Estas sustancias se adhieren a los 25 tipos de receptores especializados en el reconocimiento del amargor, aunque el ser humano tan solo es capaz reconocer una única sensación de amargor [51,53,55]. La Figura 7 muestra las estructuras de algunos compuestos responsables del amargor de los alimentos.

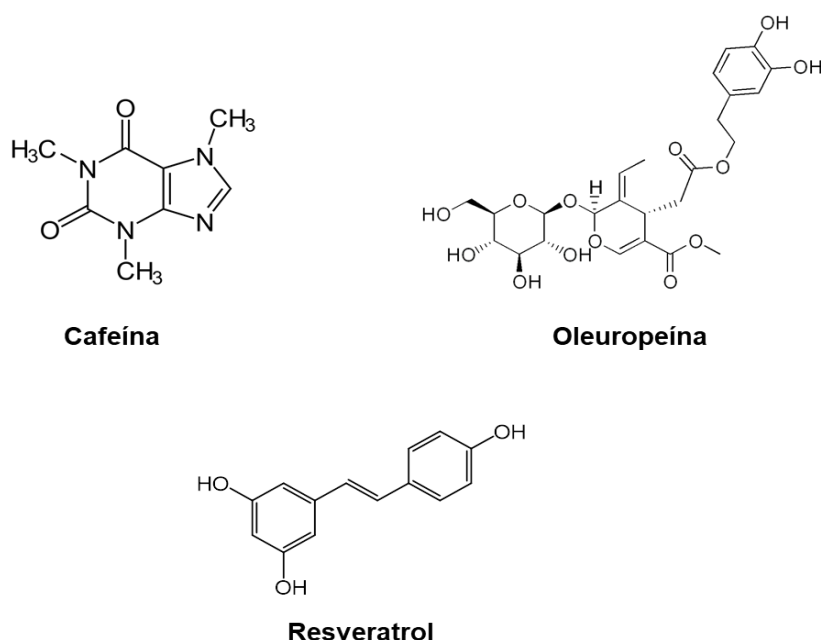


Figura 7. Estructuras químicas de compuestos responsables del amargor de alimentos como el café (cafeína), aceite (oleuropeína y resveratrol).

Comparado con el resto de sensaciones, el amargor cuenta con mayor variabilidad en la tipología de receptores, no para identificar diferentes tipos de amargor, sino para detectar con mayor rapidez la presencia de posibles sustancias que puedan ser venenosas.

➡ Acidez y salinidad de alimentos

La sensación agria/ácida de los alimentos coexiste con otras características principales como el amargor o la salinidad y se debe a la presencia de ácidos orgánicos en los mismos [56]. Esta sensación permite advertir la presencia de un alto contenido de ácidos que pueden resultar perjudiciales para la salud y que pueden estar originados por una falta de madurez en el caso de frutos o producidos por microorganismo como consecuencia de fenómenos de degradación de los alimentos. Sustancias como el ácido acético, cítrico, láctico o tartárico aparecen de forma natural en algunos alimentos en concentraciones moderadas, lo cual otorga al alimento una sensación ácida que resulta agradable al consumidor [56].

Originalmente se consideraba que la intensidad del sabor agrio estaba únicamente relacionada con la concentración de hidrogeniones liberados al medio. Sin embargo, varios estudios han demostrado que una solución de ácido acético provoca una mayor sensación de acidez que otra solución de ácido clorhídrico al mismo pH [57,58]. Esto se debe a que tanto la intensidad como el tipo de sensación agria percibida no se produce solo por una determinada concentración molar de hidrogeniones, sino que también influyen la concentración molar de todas las especies presentes en el medio con grupos carboxilo protonados, además de la hidrofobicidad de la cadena orgánica [56,59]. Con el aumento del grado de hidrofobicidad los ácidos orgánicos presentan una mayor facilidad para penetrar la membrana celular de las papilas gustativas. De hecho, son estas diferencias en la hidrofobicidad las que hacen que podamos diferenciar el gusto provocado por diferentes ácidos orgánicos.

Al igual que sucede con la acidez de los alimentos, la salinidad está causada por sustancias que en disolución se encuentran ionizadas [49]. Principalmente esta sensación se provoca por la presencia de sales orgánicas o inorgánicas de sodio y/o potasio las cuales juegan un papel fundamental en el metabolismo celular y en la transmisión de impulsos nerviosos. La detección se produce por la presencia de canales iónicos que detectan la presencia de los iones disueltos en la saliva [60]. La presencia de sales juega un papel importante sobre el resto de sensaciones percibidas ya que se pueden producir efectos de contraste, potenciando sabores como por

ejemplo el dulzor, o efectos de supresión que retardan la percepción de algunas sensaciones suavizando así los efectos en la lengua [51,61].

➡ Umami, la sensación sabrosa de los alimentos

A principios del siglo XX el químico japonés Kikunae Ikeda trató de identificar que compuestos eran los responsables de la sensación "deliciosa" que se percibía en el 'dashi', un tipo de sopa japonesa elaborada a partir de algas secas [62]. En sus estudios afirmó que existía una quinta sensación a la que denominó "umami" -término japonés que significa sabroso- diferente a las cuatro sensaciones básicas descritas hasta el momento. Dichos estudios le llevaron a concluir que este sabor estaba generado principalmente por la presencia de aminoácidos libres (consecuencia de la hidrólisis de proteínas durante la cocción o fermentación de alimentos), siendo la principal fuente de umami el ácido glutámico. Estudios posteriores revelaron que nucleótidos como el ácido inosínico o el ácido guanílico también eran capaces de generar esta nueva sensación. Aunque desde principios del siglo XX se realizaron muchos estudios sobre el ácido glutámico y su relación con el umami, no fue hasta el año 1985 cuando finalmente se reconoció como la quinta sensación básica provocada por los alimentos [46]. La Figura 8 muestra las estructuras de los tres principales compuestos a los que se les asocia esta quinta sensación.

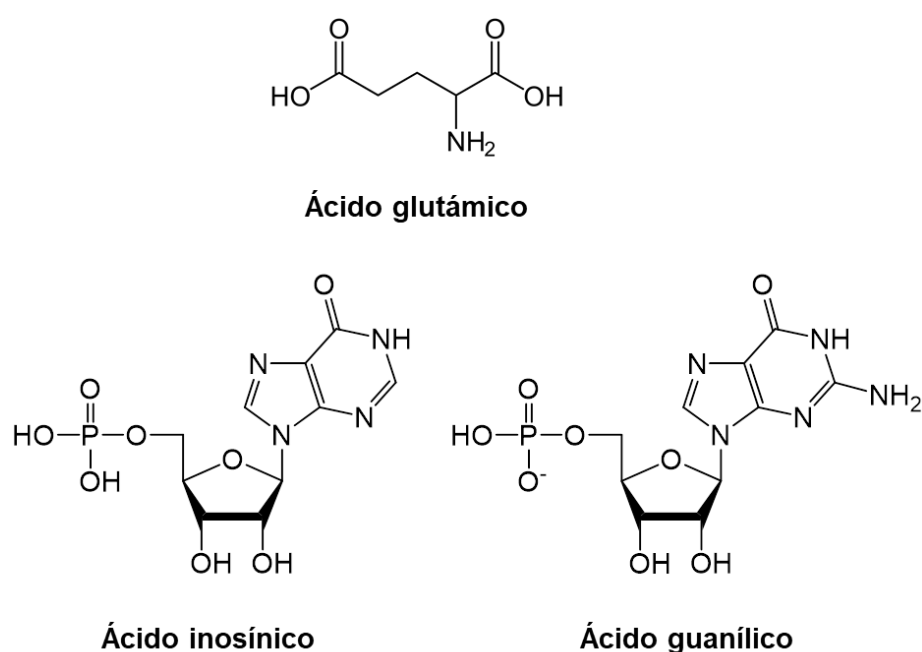


Figura 8. Estructura del ácido glutámico, principal responsable del gusto "umami", junto con otras moléculas responsables de esta sensación.

1.2.1.3.2. El sentido del olfato

Todas las sensaciones descritas, ya sean por sí solas o de forma simultánea, no son capaces de generar los diferentes sabores y/o matices que encontramos en los alimentos que consumimos. Los sabores únicos y característicos que se perciben en los alimentos se producen gracias a la combinación de miles compuestos orgánicos volátiles (COV), también denominados como odorantes [51], que se transfieren a la cavidad nasal directamente (olor) o a través de la garganta cuando se lleva a cabo la ingesta de alimentos, produciendo una sensación característica que ha recibido el nombre de flavor. En esta cavidad se encuentra el epitelio olfativo, una región de entre 1-3 cm³ donde tiene lugar el reconocimiento de este tipo de compuestos gracias a la presencia tanto de receptores olfativos y células neuroreceptoras [30]. De acuerdo con la información recogida en la Tabla 2, los odorantes son compuestos de pequeño tamaño y bajo peso molecular, con una elevada volatilidad y solubles en soluciones acuosas, como por ejemplo las mucosas nasales, para poder ser percibidos por los receptores olfativos.

Como norma general los odorantes se encuentran presentes en los alimentos en un amplio rango de concentraciones, aunque en proporción a la composición del alimento, estos compuestos se encuentran en niveles de trazas, ya que constituyen tan solo un 0.01% del total del alimento. El impacto de los odorantes no solo dependerá de la concentración, sino que también dependerá del umbral de olor/flavor de los compuestos responsables [63]. A pesar de estas bajas proporciones, la presencia de odorantes es indispensable para conformar los diferentes sabores propios de cada alimento [46]. Los diferentes tipos de olores se encuentran constituidos por un amplio abanico de moléculas con grupos funcionales muy diversos, donde dichas funciones son principalmente derivados de oxígeno, nitrógeno y azufre.

(A) Compuestos oxigenados

Los compuestos químicos con grupos funcionales principales basados en oxígeno son los más abundantes, causando todo tipo de sensaciones con diferente grado de intensidad dependiendo del tipo de unión que tenga el átomo de oxígeno a la cadena principal. Estas sustancias se encuentran de forma natural en los alimentos y durante su cocción da lugar a una parte importante de las sensaciones percibidas como consecuencia de la degradación térmica de aminoácidos y azúcares [64].

Los compuestos carbonílicos son los principales responsables de las sensaciones frescas en frutas y verduras, aunque no todos son igual de agradables. En concreto los **aldehídos saturados** de cadena corta (como por ejemplo el hexanal), así como los aldehídos insaturados (como es el caso del (Z)-hex-3-enal o el (E)-hex-2-enal) son responsables de sensaciones frescas o verdosas en la fruta, y pueden formarse por la desaminación de aminoácidos, la acción microbiana durante el proceso de fermentación o a partir de la oxidación de ácidos grasos [65,66]. La sensación generada por éstos compuestos carbonílicos cambia a medida que aumenta el tamaño de la cadena por encima de los 6 átomos de carbono, produciéndose sensaciones duales entre notas frutales o florales y olores grasos [48,63,65]. Este efecto tiene lugar independientemente de si la cadena principal presenta o no insaturaciones.

Las **cetonas** por su parte suelen aportar toques verdes cuando el grupo carbonilo se encuentra en el carbono 2, como la heptan-2-ona, mientras que cuando el grupo carbonilo se posiciona en carbono más internos de la cadena principal, el aroma empieza a virar hacia sensaciones terrosas o mohosas (octan-3-ona). Estos compuestos carbonílicos se producen principalmente por la oxidación de lípidos, así como a través del metabolismo del citrato y la glucosa [65,66]. Cabe destacar que las cetonas tienen menor importancia que los aldehídos en el flavor de los alimentos que contienen grasas. Esto se debe a que suelen encontrarse en menores cantidades y solo unas pocas cetonas se consideran clave en el aroma de determinados alimentos grasos [67].

Los **alcoholes** también se encuentran presentes en todos los alimentos, aunque su contribución al aroma es menor que los aldehídos. Las frutas y verduras son ricos en este tipo de compuestos, cuyo contenido va aumentando a medida que incrementa la madurez del mismo, momento en el que empiezan a aparecer ciertas sensaciones de dulzor [48,63]. Los alcoholes primarios y secundarios a menudo están relacionados con procesos de fermentación, aunque también pueden sintetizarse a partir del metabolismo de hidratos de carbono o aminoácidos, la reducción de grupos carbonilo o incluso la oxidación de ácidos grasos poliinsaturados [65]. La introducción de dobles enlaces en la cadena hace más atractivos a este tipo de compuestos ya que son capaces de generar característicos aromas a hoja verde ((z)-hex-3-en-1-ol) o incluso a setas (oct-1-en-3-ol).

Los **compuestos fenólicos** también aparecen en los alimentos y son particularmente olorosos. Los alquil-fenoles de pequeño tamaño causan olores no tan deseables en los alimentos como puede ser la sensación a ahumado generado por el *p*-cresol o a

estiércol provocado por el 2-etilfenol. Al contrario que sucede con los aldehídos, a medida que aumenta la complejidad del compuesto fenólico, las características aromáticas se hacen más deseables, como es el caso de la vanilina responsable del aroma característico del mismo nombre [20,48].

Los **ácidos orgánicos** son los principales responsables de las notas ácidas/agrias y picantes en algunos alimentos, los cuales resultan agradables en niveles bajos o moderados. Los ácidos también son responsables del olor rancio de aquellos alimentos con un elevado contenido graso. Este aroma aparece a medida que se produce la hidrólisis de grasas donde aparecen ácidos grasos libres que afectan al sabor y la calidad de los alimentos. Como norma general los ácidos de cadena corta suelen ser detectados fácilmente en el epitelio nasal mientras que a medida que aumenta la longitud de la cadena disminuye la intensidad de estos compuestos como odorantes. Sin embargo, la hidrólisis no es la única ruta sintética de ácidos, ya que los ácidos grasos lineales o ramificados de cadena corta pueden originarse mediante el metabolismo de aminoácidos, la oxidación de aldehídos o como subproductos de la fermentación alcohólica realizadas por determinados microorganismos [48,65,66,67].

Independientemente de su origen, estos compuestos dan lugar a reacciones de esterificación cuando se produce el aumento del contenido de alcoholes en los alimentos. Al originarse como consecuencia del envejecimiento del alimento las sensaciones provocadas por los ésteres (generalmente ésteres etílicos) suelen ir dirigidas hacia notas maduras o curadas, e incluso mohosas si aparecen en altas concentraciones. Por otro lado, si la reacción de esterificación se produce vía intramolecular (condensación intramolecular) a partir de un hidroxilácido se obtiene la correspondiente lactona. Dependiendo del tamaño del anillo formado durante la reacción de esterificación, aparecen dos nuevas familias de lactonas, las γ -lactonas y las δ -lactonas. Las primeras, γ -lactonas, se caracterizan por formar anillos furánicos (5 miembros) y además presentan un mayor poder odorante que las segundas (δ -lactonas) caracterizadas por formar anillo piránicos (6 miembros) y que son más abundantes en alimentos grasos [48,63,65,66,67].

Los **terpenos y terpenoides** son los principales constituyentes de los aceites esenciales y son responsables del perfil aromático de muchas frutas, hierbas y especias. Este tipo de sustancias se biosintetizan a partir de unidades de isopreno que pueden ser lineales, cíclicas o policíclicas. Una de las estructuras más comunes es el limoneno presente en la piel de frutos cítricos, aunque su poder odorante no es tan fuerte como el de otros terpenos como por ejemplo el (-)-mentol, que no solo aporta la

sensación mentolada, sino que también causa una sensación de frescor al activar los receptores sensibles al frío dentro de la cavidad bucal [48,65]. Algunas de las estructuras comentadas se muestran en la Figura 9.

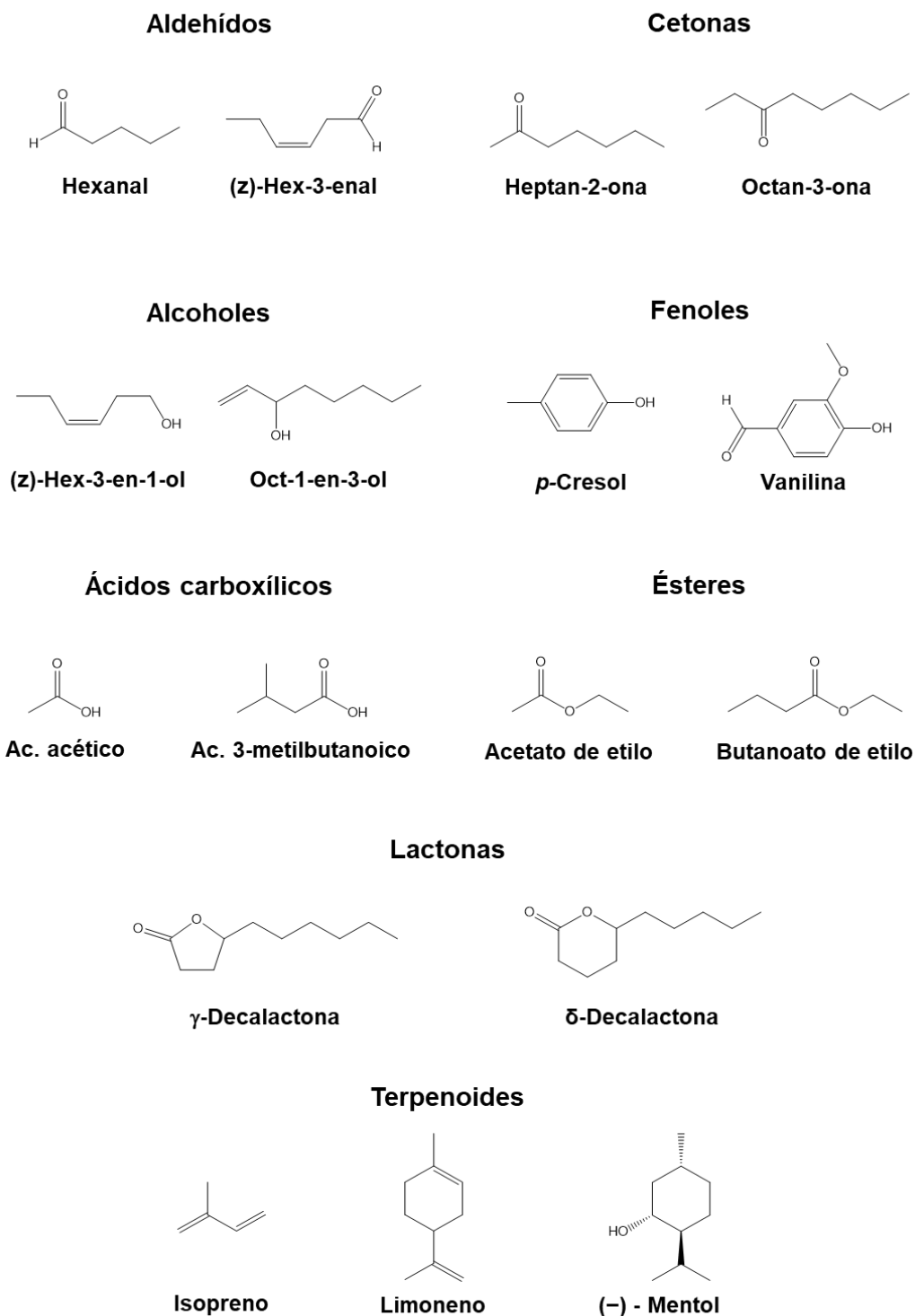


Figura 9. Algunas de las estructuras responsables de determinados olores en alimentos.

(B) Compuestos nitrogenados y derivados de azufre

Los compuestos con funciones nitrogenadas también aparecen de forma natural en los alimentos, aunque en cantidades considerablemente inferiores. En elevadas concentraciones estas sustancias causan sensaciones desagradables como aromas fecales o carnes en avanzados estados de descomposición, mientras que en concentraciones bajas su olor denota toque frutado, alcohólico e incluso a barniz [48,65]. Durante el cocinado, la concentración de compuestos nitrogenados aumenta considerablemente como consecuencia de la reacción de Maillard en la que participan grupos amino libres procedente de aminoácidos o proteínas y azúcares reductores como aldosas o cetosas. El aroma producido difiere en gran medida dependiendo del tipo de sustancia formada, por ejemplo, las aminas de cadena corta que se generan durante la fritura de pescados hacen que en ocasiones pueda detectarse cierto aroma a amoníaco, mientras que el pirrol y sus derivados son los responsables del color y olor característico de productos cocidos, tostados y quemados.

Finalmente, los compuestos químicos con azufre como función principal son excepcionalmente activos en cuanto a olor, tanto que pese a encontrarse en cantidades inferiores a los umbrales de detección analítica, estos son percibidos por el olfato. En bajas concentraciones este tipo de compuestos aportan sensaciones agradables como las que aparecen en frutas tropicales, café o en diferentes tipos de carnes, aunque no son responsables de ningún aroma característico. Los compuestos de azufre también son responsables de olores desagradables cuando se encuentran en elevadas concentraciones como es el caso del sulfuro de hidrógeno responsables del aroma a huevos podridos en muchos alimentos [48,63,65,66,67].

I.2.2. Análisis instrumental

Aunque su utilidad está más que demostrada, son pocas las industrias que aplican a diario el análisis sensorial para determinar la calidad de los alimentos que producen, y opta por un enfoque más objetivo basado en el análisis de propiedades biológicas, físicas y/o químicas de los alimentos. Precisamente, el análisis químico es uno de los más extendidos en la industria debido a la gran cantidad de información composicional que ofrecen los resultados de estos análisis. Al conocer la composición de los alimentos se pueden explicar muchas de las propiedades de los mismos e incluso que efectos van a tener estos sobre la salud de los consumidores a corto, medio y largo plazo.

Los primeros métodos químicos (análisis clásico) utilizados para el análisis de alimentos se basaban en la determinación de propiedades o variables genéricas como, por ejemplo, la humedad total, el volumen de alcohol, el pH, el contenido total de grasas, etc. Sin embargo, con el paso del tiempo, y gracias a la aparición sucesiva de equipos/instrumentos capaces de separar, medir y cuantificar compuestos de forma rápida, segura y sencilla garantizando la calidad de los resultados [68,69], quedaron relegados a un plano secundario pasando a ser utilizados en situaciones muy concretas.

Los métodos basados en la utilización de estos "equipos de medida y ensayo", y por ende, métodos instrumentales, se basan en la utilización de **técnicas espectroscópicas** –en las que se incluyen la espectroscopía de infrarrojo cercano (NIR) o la espectroscopia de absorción ultravioleta-visible (UV-Vis)–, **técnicas separativas** como la cromatografía o la electroforesis capilar y otras técnicas entre las que se incluyen las técnicas electroanalíticas como la Voltamperometría o el uso de sensores entre otras muchas.

1.2.2.1. Técnicas espectrométricas

Bajo este epígrafe se recogen aquellas técnicas instrumentales que permiten el registro de la propiedad medida en forma de lo que habitualmente se recoge en un sentido amplio como "**espectro**", aplicado a cualquier señal que se pueda medir o descomponer a lo largo de una variable continua, como por ejemplo la absorbancia (A), la transmitancia (%T) o la intensidad de emisión ($I_{em.}$) en algunas espectroscopias ópticas o la relación de masa a carga (m/q) en la espectrometría de masas.

No obstante, existe una diferencia significativa entre ellas, que radica en **la fuente que se utiliza para interactuar con la muestra a analizar**. Así, las técnicas espectroscópicas ópticas requieren de una fuente de radiación que emita radiación procedente de zonas específicas del espectro electromagnético (espectro REM) para provocar una alteración en el estado natural del analito/materia/material en estudio que puede ser relacionada con un parámetro de interés (hecho que a su vez permite subdividirlas atendiendo a la energía/longitud de onda de la radiación utilizada para provocar la interacción), mientras que la espectrometría de masas se basa en el empleo de un haz de electrones para ionizar la muestra de interés provocando la aparición de fragmentos iónicos que permiten registrar su distribución en la muestra analizada – denominado '*total ion current*' (TIC)– utilizando como variable de medida la

relación entre la masa y la carga de cada uno de los fragmentos generados como consecuencia de la interacción, de ahí que se les denomine espectros de masas.

Por otro lado, el hecho de utilizar una fuente de excitación integrada por electrones que origina la ionización de parte de los componentes de la muestra a estudiar, y que, como consecuencia de su interacción con ella, genera la formación de iones procedentes de la misma, hace necesario que en los instrumentos de medida (espectrómetros de masas) se requiera de fuentes de excitación específicas y que sea necesaria la inclusión de componentes no presentes en los espectrofotómetros utilizados con las técnicas espectroscópicas. Esto hace que para cualquier espectrómetro de masas se consideren tres componentes fundamentales:

- i) **Fuente de ionización**, región destinada a la producción de iones en estado gaseoso a partir de las sustancias en estudio. Existen diversos tipos de fuentes de ionización como las de ionización por electrones o impacto electrónico (EI), ionización química (CI), ionización química a presión atmosférica (APCI), fotoionización a presión atmosférica (APPI), ionización por electrospray (ESI) o la desorción/ionización láser asistida por matriz (MALDI).
- ii) **Analizador**, región en donde tiene lugar la separación del tanto del ion molecular cargado como sus diferentes fragmentos ionizados en base a la relación masa/carga (m/z). Se considera la parte más característica del espectrómetro de masas y existen diferentes tipos como el cuadrupolo (Q), tanto simple como triple (tQ), la trampa de iones (IT), el sistema Orbitrap, o el tiempo de vuelo (ToF).
- iii) **Detector**, región en la que se recolecta los diferentes fragmentos registrando la abundancia relativa de cada una de las especies iónicas presentes, siendo los más comunes los detectores electrón multiplicadores, la copa de Faraday y los platos multicanal.

En general, todas ellas se caracterizan por ser técnicas no destructivas que permiten llevar a cabo la medida tanto de forma directa (muestra que no requiere de ningún tipo de tratamiento previo al análisis) como en disolución, proporcionando resultados de forma rápida y sencilla. Otra ventaja añadida de estas técnicas es que no requieren de grandes cantidades de reactivos y disolventes para llevar a cabo las determinaciones por lo que se consideran técnicas de bajo coste respetuosas con el medio ambiente.

Para terminar, la Tabla 4 resume las técnicas espectrométricas más utilizadas, así como algunos métodos desarrollados para su uso en la industria de la alimentación [70,71,72,73,74,75,76,77,78,79,80].

Tabla 4. Técnicas espectroscópicas/espectrométricas más utilizadas en la industria de la alimentación.

Técnicas Espectroscópicas			
Zona REM utilizada	Denominación genérica	Observaciones	Ref.
Radiofrecuencias (RF) $\lambda > 30$ cm	Resonancia magnética nuclear	Técnica ampliamente extendida para la elucidación estructural compuestos, aunque también puede emplearse para fines cuantitativos.	[70,71]
Microondas 30 cm – 0.1 cm	Espectroscopia rotacional o de microondas	Las técnicas basadas únicamente en espectroscopia rotacional apenas tienen aplicación dentro del ámbito de alimentos.	[72]
	Espectroscopia de terahercios	Técnica emergente, aunque no se encuentra ampliamente extendida. No aporta una excesiva información estructural.	[73]

Espectroscopias

Técnicas Espectroscópicas

Tabla 4. (cont.) Técnicas espectroscópicas/espectrométricas más utilizadas en la industria de la alimentación.

Técnicas Espectroscópicas				
Zona REM utilizada	Denominación genérica	Observaciones	Aplicación	Ref.
Espectroscopias	Infrarrojo Lejano (FIR) 1000 µm a 50 µm	Es la región de menos energía y la menos accesible. No existen muchas correlaciones útiles entre el espectro generado y la información estructural en esta región.	—	[74]
	Infrarrojo Medio (MIR) 50 µm a 2.5 µm	Es el intervalo de mayor interés debido a la gran cantidad de información estructural-funcional que aporta. La espectroscopia de infrarrojos por transformada de Fourier es la modalidad más utilizada ya que previene fenómenos de fluorescencia durante el proceso de medida.	Estas técnicas pueden se utilizan frecuente mente con fines tanto cualitativos como cuantitativos. Entre la aplicabilidad de dichas técnicas destaca:	[74,75,76]
	Infrarrojo Cercano (NIR) 2500 nm a 780 nm	Este intervalo de mayor energía y su uso también se encuentra ampliamente extendido en el ámbito de la alimentación. Los equipos NIR también suelen hacer uso de la transformada de Fourier para evitar fenómenos de fluorescencia.	➤ Determinación del origen geográfico de vinos, aceites, frutos secos, mieles etc. ➤ Discriminación de especies y variedades de productos de origen animal (carne, queso, leche, etc.) y vegetal (vinos, aceites, frutos secos). ➤ Detección de adulteraciones en aceites vegetales, mieles, productos lácteos.	
	Espectroscopias Infrarrojas			
Espectroscopias	Infrarrojo (IR) 1000 µm a 0.78 µm			[74,75,76]
	Espectroscopia Raman	Tipo de espectroscopia vibracional basada en el efecto Raman (dispersión inelástica de la luz incidente sobre la materia). Esta técnica espectroscópica aporta información complementaria al NIR y MIR sobre los modos de vibración moleculares.	➤ Control del proceso de producción. ➤ Determinación de propiedades físico-químicas como el contenido total de antioxidantes en alimentos.	

Tabla 4. (cont.) Técnicas espectroscópicas/espectrométricas más utilizadas en la industria de la alimentación.

Técnicas Espectroscópicas		Zona REM utilizada	Denominación genérica	Observaciones	Aplicación	Ref.
Espectroscopias	Espectrometría	Ultravioleta-Visible (UV-Vis)	780 nm a 190 nm	Técnica de uso más frecuente en los laboratorios dedicados al análisis de alimentos, ya sea midiendo la intensidad de la radiación absorbida o emitida una vez la radiación (en el rango del UV-Vis) ha atravesado la muestra. No es aconsejable su uso cuando la concentración del analito de interés es o muy alta o muy baja. En la modalidad de absorción los resultados pueden verse afectados por el pH, la temperatura o presencia de contaminantes.	Ampliamente extendida en el análisis de aceites vegetales comestibles para determinar la presencia de adulteraciones con otros aceites o con aditivos alimentarios. Otras industrias también utilizan esta técnica en: ➤ Seguridad alimentaria: para la determinación de metales pesado o presencia de hidrocarburos policíclicos aromáticos. ➤ Prevención de la adulteración de carnes y todo tipo de bebidas ➤ Determinación de la composición de frutas, aceites comestibles, productos lácteos, carnes y pescados.	[74,77,78]
			Espectroscopia UV-Vis	Única técnica espectrométrica que no se debe a una interacción de una REM con la muestra. Existen diferentes tipos de analizadores para establecer la relación masa/carga de las moléculas. Es posible trabajar en al menos dos modalidades: i) análisis directo o ii) en tándem, habitualmente acoplado a un cromatógrafo o a otro espectrómetro de masas (MS/MS). Cualquiera de estas modalidades presenta límites de detección y cuantificación superiores a la mayoría de las técnicas espectroscópicas.	Ampliamente extendido el uso de análisis dirigido (targeted) y no dirigido (non-targeted) de analitos en: ➤ Prevención de fraudes. ➤ Presencia de pesticidas y otros contaminantes en frutas y verduras ➤ Evaluación de la calidad mediante la determinación del contenido de polifenoles o la presencia de determinados compuestos volátiles orgánicos en frutos secos, aceites comestibles, vinos etc. ➤ Determinación del origen botánico y/o geográfico de alimentos.	[79,80]

I.2.2.2. Técnicas cromatográficas y electroforéticas

Las técnicas cromatográficas (incluidas dentro de un grupo más amplio generalmente conocido como técnicas separativas de análisis) se basan en la interacción de los compuestos presentes en una mezcla con la fase estacionaria, con la fase móvil o con las dos para lograr su separación, purificación, identificación y/o cuantificación gracias a la diferencia de propiedades físico-químicas de cada compuesto. Estas técnicas se encuentran ampliamente extendidas dentro de la industria alimentaria por varias razones, siendo las principales: i) la posibilidad de acoplamiento con todo tipo de sistemas de detección, y ii) por aportar mayor cantidad de información acerca de la composición de las muestras analizadas comparadas con las técnicas espectroscópicas.

Precisamente, por estas razones las técnicas en el proceso cromatográfico tienen una gran aplicabilidad en la resolución de problemas analíticos en el campo de la alimentación. Una de las aplicaciones más interesantes de estas técnicas es su capacidad para la resolución de problemas analíticos relacionados con la autenticidad de alimentos, el cual como se ha indicado con anterioridad es uno de las mayores preocupaciones tanto por los consumidores como por la propia industria es la autenticidad de alimentos. Otras técnicas, como la **electroforesis capilar** (CE) también se basan en la separación de compuestos (en este caso en presencia de un campo eléctrico). En concreto la CE se presenta como un método alternativo a la cromatografía por su bajo coste, mínimo volumen de muestra requerida y cortos tiempos de análisis. El principal problema es que no todo tipo de sustancias pueden ser estudiadas mediante esta técnica.

Al tratarse de un tema de actualidad y con gran interés durante el transcurso de esta tesis doctoral surgió la posibilidad de participar en la redacción de un capítulo de libro basado en la aplicación de diferentes técnicas cromatográficas, así como los diferentes sistemas de detección más utilizados en problemas relacionados con la autenticidad y calidad de alimentos. Una captura de dicho capítulo se muestra en la Figura 10. Dicho capítulo se incluye de forma íntegra dentro del "Anexo" de esta Tesis.

En la Tabla 5 se muestran algunas de las técnicas separativas más utilizadas en la industria de la alimentación, así como los sistemas de detección más habituales [81,82,83,84,85,86,87,88,89,90,91].

3

Chromatographic methods

M. Gracia Bagur-González, Ana M. Jiménez-Carvelo, Fidel Ortega-Gavilán, Antonio González-Casado

DEPARTMENT OF ANALYTICAL CHEMISTRY, FACULTY OF SCIENCES, UNIVERSITY OF GRANADA, GRANADA, SPAIN

3.1 Introduction

The current global profile of food consumers shows a growing demand to incorporate the adjective “authentic” to food. This means that both producers and legal authorities need to carry out the authentication of food. *Food authentication* involves the verification of the claims of the label, which may include many aspects such as the analytical determination of characteristic components, adulterants, and/or contaminants, the verification of differentiated quality requirements, or the method of preparation or processing of such foodstuffs. If authentication is aimed at detecting and/or quantifying adulterants, then food authentication becomes *food adulteration* control. If authentication is aimed at checking declared specifications, it becomes *food fraud* control (Fig. 3–1).

Food authentication involves confirmation that the stated specifications are accurate. There are three different chemical approaches to obtain this information: analytical determination of (bio)markers, component profiles (*profiling*), and instrumental fingerprints (*fingerprinting*). In all these approaches chromatographic methods are the most commonly used when dealing with food fraud and authenticity. The main applications can be found in relation with the authenticity of fruit, vegetables, cereals, honey, fats and oils, juices, wine, hot drinks such as coffee, tea, and other beverages (Danezis et al., 2016).

In relation with the analytical determination of (bio)markers, Ballin and Laursen (2019) quote that (bio)markers are predefined analytical targets (normally one) linked to the issue of authentication. Thus when the (bio)marker provides a result directly addressing a specific authenticity issue/extension of suspect food, it is defined as a *primary (bio)marker*; meanwhile when it provides indirect information about the authenticity, for example, geographic origin, agricultural production methods, etc. it is denoted as a *secondary (bio)marker*.

As far as the two main analytical approaches recently accepted are concerned (Fig. 3–2), that is, targeted and nontargeted approaches (Nuñez, 2020), the (bio)markers and

Figura 10. Captura de la primera página del capítulo de libro "Chromatographic methods" publicado en el libro "Food Authentication and Traceability".

(Bagur González, M.G., Jiménez Carvelo, A.M., Ortega Gavilán, F. & González Casado A. (2021) Chromatographic methods. In M. Galanakis (Ed) Food Authentication and Traceability, (pp 65–99). London, Elsevier).

Tabla 5. Descripción de las técnicas cromatográficas y electroforéticas.

Técnicas	Sistemas de detección más habituales	Aplicación	Ref.
Electroforesis capilar	Detectores ópticos (UV-Vis, fluorescencia), detectores electroquímicos (amperométricos, conductimétricos y potenciométricos).	Determinación de nutrientes (péptidos, aminoácidos, vitaminas, carbohidratos, iones esenciales), compuestos fenólicos y aditivos alimentarios en té, vino café y aceites vegetales.	[81,82]
Gases	Detector de conductividad térmica (TCD), detector de ionización de llama (FID) , detector nitrógeno-fosforo (NPD), Detector de captura electrónica (ECD), detector olfatométrico (OD), espectrómetro de masas (MS) , espectrómetro de movilidad iónica (IMS).	Análisis de compuestos volátiles orgánicos, fracción lipídica, pesticidas, toxinas, nitrosaminas o aditivos alimentarios para: ➢ Evaluación de la calidad ➢ Presencia de sustancias tóxicas ➢ Determinación de origen botánico y geográfico ➢ Autenticidad y prevención de fraudes	[83,84,85,86,87,88]
Cromatografía	Líquidos Detector ultravioleta visible (UV-Vis), detector de diodos en fila (DAD) , detector de fluorescencia, detector electroquímico, detectores de índice de refracción, detector de corona de descarga (CAD), espectrómetro de masas (MS).	Análisis de compuestos relacionados con el color, fracción grasa, proteínas, vitaminas, antioxidantes naturales y/o artificiales, así como todo tipo de aditivos alimentarios para: ➢ Evaluación de la calidad ➢ Presencia de sustancias tóxicas ➢ Determinación de origen botánico y geográfico ➢ Autenticidad y prevención de fraudes	[87,88,89,90,91]

Nota. En negrita se marcan las técnicas separativas, así como los sistemas de detección empleados en el transcurso de esta Tesis Doctoral.

1.2.2.2.1. Cromatografía de gases (GC)

La GC se emplea en la separación de compuestos térmicamente estables en la temperatura requerida para su vaporización en función de las interacciones de dichos compuestos con el relleno de la columna capilar (mucho más eficiente que las columnas de tipo compacto) que es el que actúa como fase estacionaria [92]. La separación se produce considerando que los compuestos con baja volatilidad interaccionan en menor medida esta fase siendo arrastrados por el gas portador a través de la columna capilar hasta el detector, a diferencia de los compuestos de mayor volatilidad que interacciona de forma más intensa con la fase estacionaria y por tanto sufren mayor retención llegando a tiempos mayores al detector.

A continuación, se detalla el funcionamiento de algunos de los sistemas de medida de la señal (generalmente denominados detectores) más habituales en laboratorios de análisis de alimentos y que han sido utilizados para el desarrollo de este trabajo [86].

(A) Detección mediante ionización de llama

De todos los sistemas de detección recopilados en la Tabla 5, el detector de ionización de llama (FID) es el más habitual en los cromatógrafos de gases debido a los bajos límites de detección, su amplio rango dinámico, su asequibilidad y su fiabilidad.

Este detector se encuentra constituido por un horno que se puede calentar entre 300 y 350 °C donde se mezcla el gas portador procedente de la columna capilar, el cual contiene al analito, con hidrógeno antes de llegar a la base del mechero. La llama se inicia mediante una descarga eléctrica en un entorno que contiene exceso de aire introducido en compartimento de detección a través de un canal separado. Prácticamente la totalidad de los compuestos orgánicos (salvo el ácido fórmico y el metano) se pirolizan al entrar en la llama, produciéndose iones que quedan recogidos en el colector. Al pasar al colector, los iones producen una pequeña corriente proporcional a la cantidad de compuesto quemado que se mide como señal del detector y se amplifica gracias a un electrómetro. Esta relación de la señal con la cantidad de sustancia ionizada hace que este detector se considere sensible a la masa (específicamente al número de átomos de carbono que sales de la columna) más que a la concentración, por lo que un cambio de flujo no afecta demasiado al resultado final. En la Figura 11 se muestra un esquema de los componentes del detector FID [83,85,93,94,95].

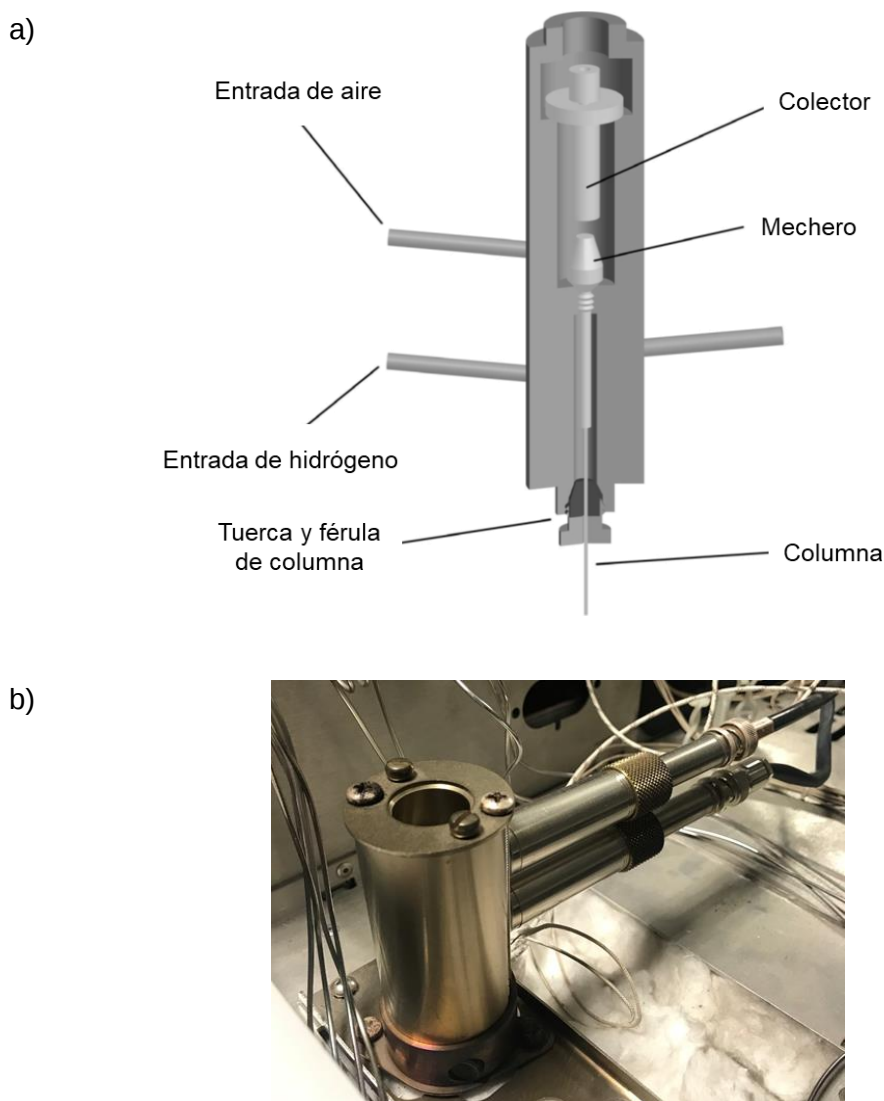


Figura 11. Detector de ionización de llama: a) componentes del detector; b) detector FID instalado en el equipo Varian GC 3800 empleado en el desarrollo de la tesis doctoral. (La Figura a) está adaptada de Rood, D. (2007) Detectors. In D. Rood The troubleshooting and maintenance guide for gas chromatographers (pp 139–185) Wiley-VCH Weinheim, Germany).

(B) Detección mediante espectrometría de masas

Otra configuración posible es el acoplamiento de un cromatógrafo de gases a un espectrómetro de masas, de forma que una vez separados los analitos, estos puedan ser identificados en función de su relación masa carga (m/z). Como se indicaba anteriormente las partes fundamentales de los espectrómetros de masas son: la fuente de ionización, el analizador y el sistema de detección.

La **fente de ionización** más habitual en espectrómetros de masas acoplados a cromatógrafos de gases son aquellas basadas en **ionización por impacto electrónico (EI)**. La ionización de los analitos neutros en fase gaseosa se produce gracias a un haz de electrones emitidos por un filamento metálico (habitualmente volframio o renio). Estos electrones se someten a un proceso de aceleración aplicando para ello un potencial eléctrico habitualmente comprendido entre los 50 y los 70 eV. Este rango de energía es el más habitual, ya que aumenta la probabilidad de transferencia cuantitativa de energía a las moléculas neutras para dar lugar a la formación de iones. Durante dicha aceleración los electrones producidos en el filamento se movilizan hacia el polo positivo de la cámara de ionización, extrayendo durante este proceso electrones de las moléculas de la muestra presentes en la cámara de ionización, ionizando así las moléculas de analito. Una vez ionizadas las moléculas, los altos valores de energía interna originan la fragmentación de las moléculas en otras de menor tamaño. Aunque todo este proceso es conocido como impacto electrónico, los electrones emitidos rara vez colisionan con las moléculas de muestra. En la Figura 12 se muestra un esquema de la ionización por impacto electrónico en la fuente homónima [80,96].

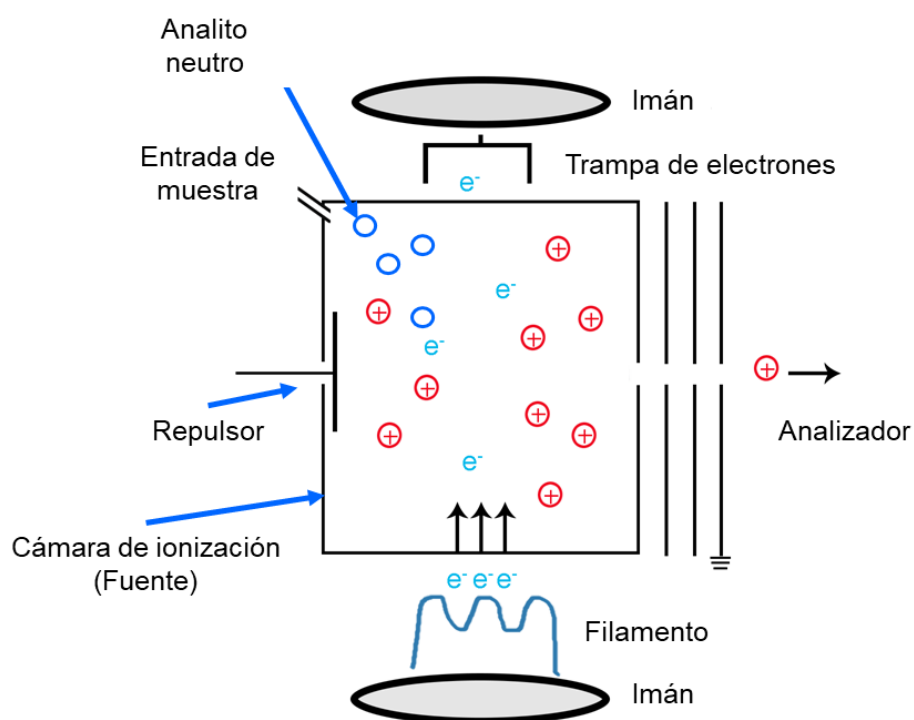


Figura 12. Esquema del proceso de ionización en una fuente de impacto electrónico.

(Fuente: elaboración propia).

Existe una gran variedad de analizadores, así como posibles combinaciones entre ellos para poder evaluar las mínimas diferencias existentes entre los iones logrando así una mejor identificación de los mismos. De los diferentes tipos de analizadores, los más habituales en laboratorios de rutina son el **cuadrupolo simple (Q)** y la **trampa de iones (IT)**.

El **cuadrupolo simple** es el tipo de analizador más común en los laboratorios de rutina debido a los bajos costes que suponen su adquisición, son compactos y emplean bajos tiempos de barrido que permiten tener información de la separación cromatográfica en tiempo real, siendo capaces de generar cincuenta espectros por segundo. Este analizador consiste en cuatro barras cilíndricas o hiperbólicas (electrodos) paralelas entre sí las cuales actúan como filtro de masas. Estas cuatro barras se utilizan para generar dos potenciales eléctricos iguales pero desfasados a su vez, siendo uno de ellos de corriente alterna mientras que el otro potencial es de corriente continua. La diferencia de potencial varía para generar un campo eléctrico oscilante entre dos barras opuestas, lo que hace que estas tengan la misma carga, pero opuesta a su vez. El ajuste de los potenciales de las barras del cuadrupolo permite seleccionar un rango de masas concreto, el tipo de iones seleccionados o incluso ofrece la posibilidad de detectar un único ion.

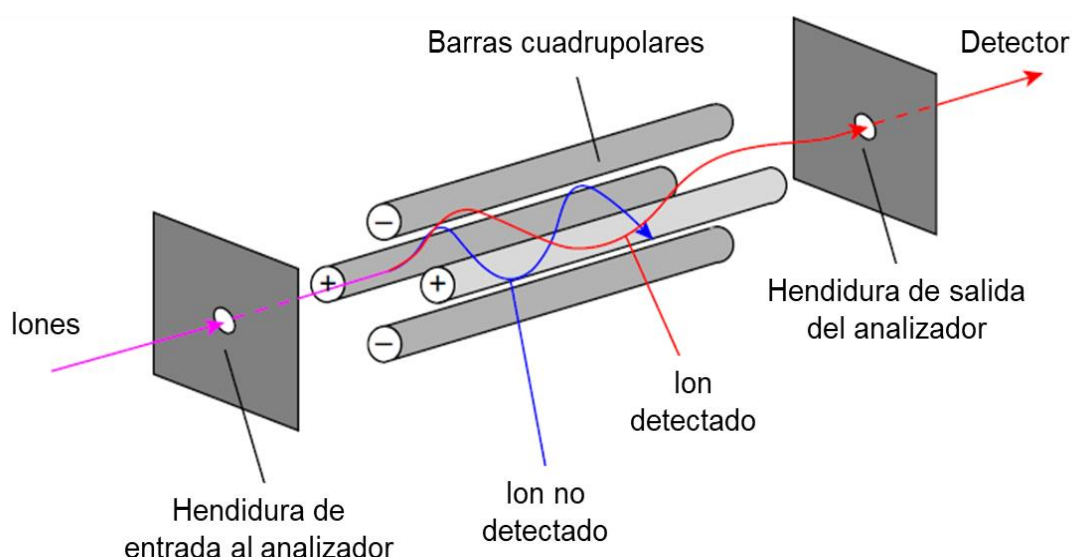


Figura 13. Esquema de configuración y funcionamiento del analizador Cuadrupolo simple (Q).

(Adaptada de: Kaklamanos, G., Aprea, E. & Theodoridis, G. (2012). En Y. Picó (Ed) Chemical analysis of food techniques and application 1ª ed (249–283) Elsevier, Londres, Reino Unido).

Cuando los iones entran dentro del campo electrónico producido por el cuadrupolo, instantáneamente se verán atraídos hacia las barras con carga opuesta. Si el potencial de esta barra vuelve a cambiar antes de que el ion impacte con ellas, dicho ion será desviado hacia la barra que presente carga opuesta. De esta forma, cada ion con una trayectoria estable, describe un movimiento de tipo senoidal desde su entrada en la región del cuadrupolo hasta su llegada al transductor. El resto de iones con trayectorias inestables colisionan con las barras del cuadrupolo y se eliminan del campo oscilante gracias al vacío que se realiza en el interior del analizador. La Figura 13 muestra un esquema tanto de la configuración del analizador de cuadrupolo como el movimiento descrito por un ion desde su salida de la cámara de ionización hasta su llegada al detector.

Los analizadores de **trampa de iones** son dispositivos que presentan una alta sensibilidad y resolución espectral comparados con los analizadores de tipo cuadrupolo simple. La trampa de iones es especialmente útil cuando se dispone de dos espectrómetros de masas acoplados en tándem para repetir fragmentaciones determinadas de los iones producidos. A pesar de las ventajas que ofrece la trampa de iones, la rápida aceptación del cuadrupolo simple, así como el desarrollo del triple cuadrupolo hacen que el analizador de trampa de iones no sea tan habitual en los laboratorios. La trampa de iones es esencialmente un analizador cuadrupolar multidimensional que almacena los iones en una cavidad interna denominada "trampa" para luego expulsarlos de la misma atendiendo a la relación masa carga. Una vez atrapado, los iones pueden someterse a otra separación (espectrometría de masas en tándem) para lograr un aumento de la resolución de las masas y mejorar la sensibilidad. Los iones se introducen en la trampa a través del electrodo de entrada quedando retenidos en la misma gracias a un campo de radiofrecuencias variables al describir trayectorias estables tanto en la dirección axial como radial. A medida que varía la radiofrecuencia parte de los iones "atrapados" se vuelven inestables y se expulsan de la trampa a través del electrodo de salida en dirección al detector.

Un esquema de este analizador se muestra a continuación:

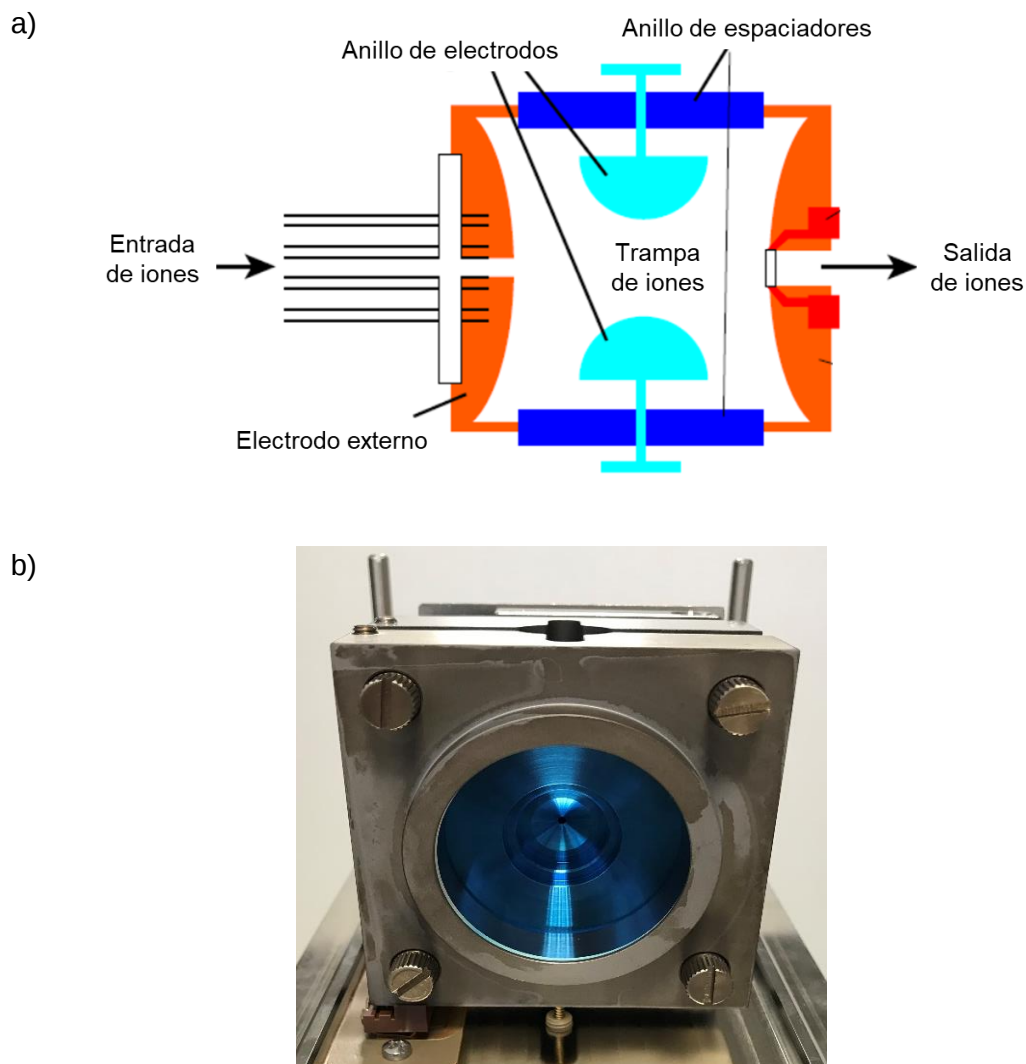


Figura 14. a) Esquema de configuración y funcionamiento de un analizador de trampa de iones; b) analizador IT integrado en el espectrómetro de masas Varian 4000ms.

(La Figura a) está adaptada de: Kaklamanos, G., Aprea, E. & Theodoridis, G. (2012). En Y. Picó (Ed) Chemical analysis of food techniques and application 1ª ed (249–283) Elsevier, Londres, UK).

La principal diferencia entre una trampa de iones y un cuadrupolo simple radica en la estabilidad de los iones que llegan al detector. En la trampa de iones, son los iones que presentan una trayectoria inestable los que salen de la trampa para ser detectados, mientras que en el cuadrupolo los iones estabilizados son los que alcanzan el detector. Para este analizador se emplean helio como gas amortiguador para reducir la energía de los iones, por tanto, es habitual el uso de este mismo gas como gas portador en la separación cromatográfica [68,96,97].

1.2.2.2.2. Cromatografía de líquidos (LC)

La LC es otra modalidad bastante extendida en la industria de la alimentación al presentar una alta sensibilidad, selectividad y resolución que permite la separación de compuestos no volátiles y térmicamente estables. En este caso la separación se realiza atendiendo a propiedades físico-químicas como la diferencia de polaridad, las propiedades ácido-base o la estructura tridimensional entre muchas otras propiedades y se produce como consecuencia de las interacciones de los analitos tanto con la fase estacionaria como con la móvil [91,98].

Puede clasificarse en dos grupos atendiendo a la naturaleza de la fase móvil en: fase normal (NPLC) y fase invertida (RPLC). La fase normal presenta grandes ventajas sobre la fase invertida en cuanto a los altos niveles de estabilidad al emplear fases no acuosas, la mejor solubilidad de la mayoría de compuestos orgánicos en las fases empleadas y por trabajar a presiones inferiores a la fase reversa. Sin embargo, el coste de los disolventes, el bajo número de platos de separación y el desarrollo de fases enlazadas han contribuido a la rápida expansión de la fase invertida en la industria [99].

Hoy en día la LC con sistemas de detección basados en la absorción de radiación en la zona del ultravioleta-visible o en la utilización de espectrómetros de masas se consideran las técnicas más potentes para abordar problemas de seguridad alimentaria y garantizar la autenticidad de alimentos con el fin de evitar el fraude [100].

Los **detectores ultravioleta-visible (UV-Vis)** son los más comunes en cromatografía de líquidos y hacen uso de una determinada longitud de onda para determinar la concentración a través de la ley de Lamber-Beer. Los avances en tecnología han derivado en el desarrollo de detectores de longitud de onda múltiple que permiten registrar de manera simultánea la respuesta de la muestra a diferentes longitudes de onda en cada punto temporal con una mínima pérdida de sensibilidad y resolución. A su vez, este hecho permite identificar de manera simultánea diferentes familias de compuestos [101].

El **detector de fila de diodos (DAD)**, también conocido como detector de fila de fotodiodos (PDA), es el más utilizado en la actualidad debido a su bajo coste y a la gran cantidad de información que facilita. Este tipo de sistema de detección enfoca el espectro completo generado por la lámpara dentro de la celda de flujo. La luz dispersada incide sobre una rejilla de difracción, la cual distribuye el espectro completo

sobre los fotodiodos dispuestos en fila. La cantidad de fotodiodos en estos detectores varía en función de la marca y modelo, aunque estos suelen estar comprendidos entre 512 y 1024 fotodiodos. Las señales individuales registradas en estos fotodiodos son procesadas para generar los espectros de absorbancia individuales o una combinación de todas ellas para generar un mapa de contornos donde se representa la intensidad de la señal registrada a cada longitud de onda y a cada tiempo de análisis [102,103]. Un esquema del detector de arreglo de diodos, así como las lámparas empleadas como fuente de radiación se muestra en la Figura 15.

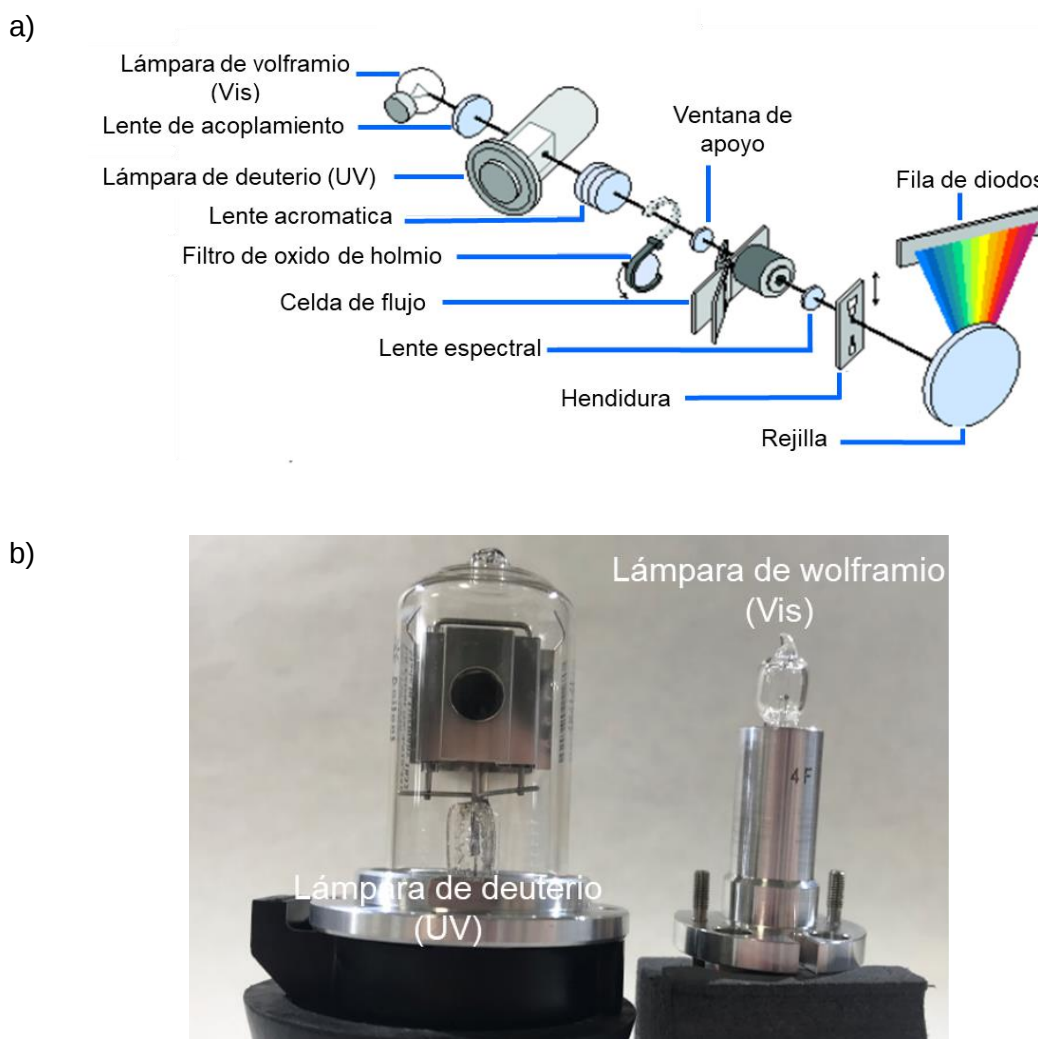


Figura 15. a) Esquema del detector de fila de diodos (DAD) empleado en el módulo de HPLC Agilent 1260 DAD WR: b) lámparas empleadas en el módulo HPLC Agilent 1260 DAD WR. (La Figura a) ha sido adaptada de Agilent 1260 Infinity Diode Array and Multiple Wavelength Detector).

1.2.2.3. Otras técnicas instrumentales

Pese a las ventajas que presentan las técnicas espectroscópicas y la versatilidad de las técnicas separativas, también se han indagado en otros tipos de técnicas para la monitorización de alimentos en términos de calidad y procesos de producción. Las nuevas técnicas, alejadas de la espectroscopia y la necesidad de separación de los compuestos que integran la muestra, emplean sensores para poner de manifiesto la presencia de microorganismos y sustancias químicas de manera similar al comportamiento neurofisiológico de los sentidos humanos [104,105].

En la industria de la alimentación, los sensores se pueden agrupar en dos categorías atendiendo a la naturaleza de la sustancia/sustancias de interés: los sensores de gases (compuestos volátiles orgánicos en el caso de alimentos) comúnmente denominados como **nariz electrónica (e-nose)** y los sensores químicos y bioquímicos que habitualmente se denominan como **lengua electrónica (e-tongue)**. En ambos casos la presencia de las sustancias de interés se pone de manifiesto a través de cambios reversibles en las propiedades eléctricas de dichos sensores [106].

En cuanto a su configuración, la *e-nose* consiste en un conjunto de sensores dispuestos en una cámara de reacción que se encuentra conectada a una bomba de aire que lleva a cabo el muestreo y limpieza de dicha cámara. Los sensores a su vez se encuentran conectados a un dispositivo electrónico para el control del equipo y la adquisición de datos. Estos dispositivos suelen ser compactos y se pueden modificar fácilmente para obtener funciones únicas para distintas muestras. Sin embargo, comercialmente no existe una nariz electrónica que pueda analizar todos los compuestos volátiles presentes en los alimentos [104,105,106]. Por su parte la *e-tongue* consiste en un conjunto de sensores químicos los cuales se sumergen sobre un recipiente de reacción para poner de manifiesto la presencia de sustancias. Estos sensores se conectan directamente a un transductor de señal que se encuentra conectado a su vez a un ordenador que permite tanto el control instrumental como la recolección de datos [104,105,106,107].

Estos dispositivos tienen un bajo coste, son fáciles de usar y proporcionan resultados de forma rápida sin necesidad de llevar a cabo una destrucción de la muestra. Sin embargo, son sistemas muy sensibles a la humedad, la presión y la temperatura por lo que se requiere de un control estricto de las condiciones ambientales para su uso, además en la mayoría de los casos se requiere de un tratamiento de muestra con el fin de reducir o eliminar el efecto matriz de los resultados. La principal limitación de este

tipo de sistemas es la reducida vida útil de la reversibilidad de los sensores, especialmente los biomateriales que constituyen algunos sensores. Aunque prometedores, estos instrumentos aún requieren de mejoras para que su uso sea comparable con las técnicas espectroscópicas y separativas [105,106,108,109].

I.3. ANÁLISIS DIRIGIDOS Y NO DIRIGIDOS: DE LOS (CHEMO)MARCADORES A LA UTILIZACIÓN DE LA HUELLA INSTRUMENTAL

Las señales obtenidas a partir de todos estos instrumentos quedan almacenadas en conjuntos de datos, denominados en la terminología matemática como tensores, en forma de vectores (tensor de orden 1) o matrices (tensor de orden 2). El tipo de tensor vendrá determinado por la complejidad, por el tipo y por la cantidad de información que recopile el instrumento empleado. A partir de la información obtenida surgen dos estrategias dependiendo de la cantidad de información utilizada para la resolución del problema analítico planteado y que queda esquematizada en la Figura 16 [88,110]:

- i) El enfoque dirigido (*targeted*), donde se utiliza una parte reducida de la señal específica directamente relacionada con compuesto(s) químico(s) (marcador(es) químico(s) o *chemomarker(s)*) o con familias de compuestos (perfil específico o *compound profile*) identificados. Por su parte, los marcadores químicos (también denominados biomarcadores si son resultado de procesos biológicos) pueden estar directamente o no relacionados con el problema analítico estudiado, como por ejemplo la autenticidad de alimentos. Cuando la relación entre el marcador y el problema analítico es directa se denomina como (bio)marcador químico primario, mientras que, si la relación es indirecta, el marcador pasa a denominarse (bio)marcador químico secundario [111].
- ii) El enfoque no dirigido (*untargeted*), donde se utiliza la señal inespecífica completa obtenida durante el análisis (huella instrumental o *instrumental fingerprint*) sin asumir ningún conocimiento previo de los componentes, ya sean dichos componentes relevantes o irrelevantes, que forman parte de la muestra analizada.



Figura 16. Posibles estrategias a seguir dependiendo de la cantidad de información utilizada en la obtención de conclusiones.

(Adaptada de: Bagur González, M.G., Jiménez Carvelo, A.M., Ortega Gavilán, F. & González Casado A. (2021) Chromatographic methods. In M. Galanakis (Ed) Food Authentication and Traceability, (pp 65–99). London, Elsevie).

Comparando la cantidad de información facilitada por los (bio)marcadores, perfiles y huellas instrumentales, las dos últimas pueden ser consideradas como una extensión de la primera, entendiendo que el perfil es la combinación de sucesivos (bio)marcadores específicos de una misma familia, mientras que la huella instrumental es una combinación de todas las señales o marcadores relevantes e irrelevantes obtenidas en el análisis.

Aunque los tres tipos de señales codifican información acerca de la identidad de los alimentos, la huella instrumental se considera la de mayor interés al codificar por completo toda la información acerca de la composición e identidad de la muestra en estudio, en este caso los alimentos.

En el ámbito analítico el uso de este tipo de señales ha dado lugar a la "**metodología de huellas dactilares instrumentales**" (*instrumental fingerprinting methodology*), la cual al igual que ocurre con la huella dactilar humana, puede emplearse en este caso en la identificación de alimentos sin necesidad de identificar y/o cuantificar ningún componente concreto. Esta metodología puede ser empleada para otros fines como la

evaluación de la calidad, la determinación del origen geográfico, botánico o animal e incluso para la detección de adulteraciones en alimentos [112,113].

La complejidad de la información recogida hace necesario el uso de **quimiometría** para poder extraer la información útil con la que obtener conclusiones acerca de la muestra de estudio [114,115].

I.4. QUIMIOMETRÍA Y CALIDAD ALIMENTARIA

De acuerdo con la IUPAC, la **quimiometría** se define como la ciencia que relaciona las mediciones realizadas en un sistema o proceso químico con el estado del sistema mediante la aplicación de métodos matemáticos o estadísticos [116].

A día de hoy existe un gran número de herramientas quimiométricas destinadas al diseño de experimentos, el reconocimiento de pautas, etc. [117,118,119]. En concreto las técnicas para el reconocimiento de pautas están ampliamente extendidas en el ámbito de la alimentación por su capacidad para identificar relaciones, vínculos o semejanzas entre muestras que permitan su agrupamiento y/o clasificación en función de las propiedades que tienen en común. Las técnicas para el reconocimiento de pautas pueden englobarse en dos grandes grupos: i) técnicas de reconocimiento de pautas no supervisadas, y ii) técnicas para el reconocimiento de pautas supervisadas.

I.4.1. Técnicas para el reconocimiento no supervisado de pautas

Las técnicas de reconocimiento de pautas no supervisadas se centran en el estudio del comportamiento natural de las muestras sin tener en cuenta ningún tipo de información acerca de la naturaleza, composición o procedencia de las muestras. De este modo puede evaluarse las relaciones naturales existentes entre los datos sin disponer de información previa (en caso de que esta exista). Dentro de este grupo encontramos: el análisis de componentes principales (PCA), análisis factorial (FA) y análisis jerárquico de clases (HCA).

En la Tabla 6 se recogen las particularidades de cada una de estas herramientas [119,120,121,122].

Tabla 6. Descripción de las técnicas de reconocimiento de pautas no supervisadas.

Herramienta	Descripción	Ref.
Reconocimiento de pautas no supervisadas	Es probablemente la herramienta quimiométrica más extendida. Esta técnica lleva a cabo una representación de las observaciones descritas por variables dependientes intercorrelacionadas entre si a un nuevo sistema de ejes ortogonales, denominados 'Componentes Principales' que coinciden con la máxima variabilidad de las observaciones iniciales. La proyección de los datos originales sobre estas nuevas componentes se denomina 'scores', mientras que la importancia o "peso" que tiene cada una de las variables originales en la obtención de dichos 'scores' se denomina 'loadings'. La proyección en cada PC de los datos originales contenidos en la matriz X se realiza de acuerdo a la siguiente ecuación: $X = T \times P^T + E$ Donde T se corresponde con la matriz de puntuaciones ('Scores'); P^T se corresponde con la matriz de cargas ('loadings') traspuesta y E se corresponde con la matriz de residuales. Esta herramienta trata de caracterizar las variables presentes en la matriz X mediante un pequeño número de factores comunes 'p' y sus correspondientes errores o factores específicos 'e'. A cada factor 'p' se le asigna una propiedad que no puede ser observada de forma directa. El FA sigue una estrategia similar al PCA, ya que lleva a cabo una transformación de los datos originales a un nuevo sistema mediante métodos de rotación, sin embargo, difieren en cómo se determinan los factores o componentes. El PCA busca componentes que expliquen la mayor parte de la varianza total mientras que el FA busca factores que expliquen la mayor parte de la varianza común compartida con otras variables.	[119,120,122]
	Análisis de componentes principales (PCA)	
	Análisis factorial (FA)	[114,117]
Análisis jerárquico de clúster (HCA)	El análisis jerárquico de clúster lleva a cabo un agrupamiento de muestras paso a paso según las distancias o similitudes existente entre las variables de las muestras ya sean por procedimientos aglomerativos o divisivos. El procedimiento más común es el aglomerativo, donde el modelo parte de objetos individuales (muestras, compuestos etc.) que se fusionan en grupos (clústeres) de objetos más grandes. El procedimiento división parte de un grupo de objetos que se van dividiendo en nuevos clústeres individuales. El resultado es un dendrograma donde se visualiza la estructura obtenida en la agrupación. El método de Ward de mínima varianza es altamente eficiente en el cálculo de las distancias de todos los posibles clústeres, razón por la que es el método de anidamiento más extendido.	[114,117,121,122]

Nota. En negrita se indican las técnicas de reconocimiento de pautas no supervisadas empleadas en el desarrollo de esta Tesis Doctoral.

I.4.2. Técnicas de reconocimiento supervisado de pautas

Las técnicas para el reconocimiento de pautas supervisadas tienen en consideración la información previa de la muestra durante la construcción del modelo para después poder clasificar o cuantificar una o varias propiedades de muestras desconocidas.

Mientras que los modelos de clasificación relacionan muestras con al menos una característica, atributo o cualidad en clases o categorías, los modelos de cuantificación establecen relaciones entre propiedades cuantitativas de la muestra con múltiples variables independientes.

El desarrollo de estos modelos consta de dos etapas:

1. Una primera **etapa de entramiento del modelo** en la que se selecciona un conjunto de muestras perteneciente a una o varias clases conocidas. Las muestras pueden ser seleccionadas de forma aleatoria por el operador o mediante la aplicación de algoritmos informáticos de selección como el propuesto por Kennar-Stone o el método "Onion".
2. Una segunda **etapa de validación del modelo** en la que se evalúa el poder clasificador del mismo, llevando a cabo una validación cruzada interna (*cross-validation*) o una validación externa. La validación cruzada interna se lleva a cabo sobre el conjunto de datos de entrenamiento para: i) evaluar el modelo cuando se dispone de pocas muestras, ii) evaluar el poder de clasificación empleando las muestras de entrenamiento, o iii) optimizar la selección del número de componentes necesarias para establecer el modelo. Por su parte, la validación externa hace uso de un conjunto alternativo de muestras diferente al empleado para llevar a cabo el entrenamiento.

Dentro de este grupo se incluyen los análisis de similitud, k-vecinos cercanos (k-nearest neighbors o kNN), modelo independiente y flexible por analogía de clases (soft independent modeling by class analogy o SIMCA), análisis discriminante lineal por mínimos cuadrados parciales (partial least squares-discriminant analysis o PLS-DA) sistema de aprendizaje automático soportado por vectores (supported vectors machine o SVM) entre otros.

Algunas de estas técnicas de reconocimiento de pautas se describen a continuación en la Tabla 7 [114,119,121,123, 124,125,126,127,128].

Tabla 7. Descripción de las técnicas de reconocimiento de pautas supervisadas.

Herramienta	Descripción	Ref.
Reconocimiento de pautas supervisadas	Se trata de una herramienta quimiométrica que lleva a cabo la comparación de dos vectores de datos que contienen la información de interés elemento por elemento para obtener un parámetro informativo que expresa como de semejantes o diferentes son estos dos vectores. Existen diferentes tipos de índices de similitud que pueden agruparse en función del rango de distancias en índices de distancias no limitadas , que evalúan la distancia desde cero hasta infinito, e índices de distancias limitadas donde la distancia medida va desde cero hasta un valor máximo impuesto por un límite superior. Independientemente del tipo de distancia medida es habitual el cálculo del índice de similitud medio cuyo valor es independiente de las p variables que describen a las muestras, y que habitualmente toma valores comprendidos entre 0 y 1.	[125,126]
	Los índices de similitud más utilizados son el coseno del ángulo formado entre dos vectores (Cosθ) y el coeficiente de determinación (R²), aunque otros índices como el índice de proximidad (Nearness) han demostrado ser más sensibles que los tradicionalmente utilizados.	
	Se trata de un método de clasificación no paramétrico basado en la medida de la distancia local relacionada con la similitud entre las muestras. Se emplea cuando los grupos de muestras no pueden separarse en un mismo plano. La pertenencia de la muestra a una clase se realiza mediante la búsqueda de la distancia euclídea más corta a número determinado de vecinos representado por "k". El valor óptimo de k se determina a partir de la validación cruzada, seleccionando el número de vecinos que den menor error de clasificación.	[114,117,119]
Soft independent modeling by class analogy (SIMCA)	SIMCA se considera como una aplicación más profunda del PCA donde se produce una evolución de un método de reconocimiento de pautas no supervisado a una técnica de aprendizaje supervisado. Se trata de una herramienta de clasificación basada en la construcción de componentes principales que mejor se ajusta alrededor de cada clase modelada de forma independiente en un conjunto de datos. Para cada clase se define un límite de aceptación basado en los residuos máximos de las muestras de ese grupo.	[114,121, 127,128]
K-nearest neighbors (K-NN)	Cuando se introduce una muestra desconocida en el modelo, esta se asigna a la clase con la menor varianza residual durante la validación externa. Puede darse el caso que una muestra desconocida pertenezca a más de una clase. Este hecho puede ser o bien por que el modelo desarrollado carece de potencia suficiente para discriminar entre las clases en cuestión o porque haya tenido lugar un error durante la medida.	

Tabla 7. Descripción de las técnicas de reconocimiento de pautas supervisadas.

Herramienta	Descripción	Ref.
Reconocimiento de pautas supervisadas		
Linear Discriminant Analysis (LDA)	LDA es un método paramétrico basado en PCA que establece la combinación lineal más óptima de las variables originales. Las observaciones transformadas se denominan funciones discriminantes las cuales maximizan la varianza dentro de los grupos. La separación entre distintas clases se establece mediante la selección de las direcciones que consiguen una máxima separación entre ellas. La pertenencia de una muestra desconocida a una clase se establece evaluando la distancia euclídea al centroide del grupo.	[117,121]
Partial Least Squares Discriminant Analysis (PLS-DA)	PLS-DA es el método de más popular y más eficiente para tratar una matriz de covarianzas singular. PLS es un método de regresión que maximiza la covarianza entre las variables independientes y las correspondientes variables dependientes de conjuntos de datos altamente dimensionales, permitiendo la predicción de nuevas muestras en base a un número reducido de variables latentes. Los límites de las clases se pueden establecer mediante un análisis discriminante, el cual optimiza el número de variables latentes del modelo PLS.	[117,119,121,122,123,124,125]
Support Vector Machine (SVM)	La clasificación se basa en construir un modelo de calibración con diferentes clases asociadas a un valor numérico como 0 o 1 en el caso de notación binaria, llevando a cabo la asignación de clases de nuevas muestras mediante la predicción de estos valores numéricos. SVM consiste en un método de aprendizaje desarrollado para problemas de clasificación binaria que emplea hiperplanos para establecer/definir los límites de decisión que separan los datos de diferentes clases. SVM presenta la ventaja de establecer tanto clasificaciones simples mediante modelos lineales como problemas de clasificación complejos donde el modelo es no lineal. El modelo se construye mediante el mapeo de los datos originales en un nuevo espacio de alta dimensión de forma que el problema de clasificación se simplifique en este nuevo espacio.	[117,119,121,122,125]

Nota. En negrita se indican las técnicas de reconocimiento de pautas no supervisadas empleadas en el desarrollo de esta Tesis Doctoral.

I.5. PROCESADO DE LA SEÑAL INSTRUMENTAL COMO REQUISITO PREVIO PARA APLICAR LA METODOLOGÍA "FINGERPRINTING"

En el caso específico de la metodología de huellas dactilares es necesaria una etapa de pre-procesado de datos con el objetivo de linealizar la respuesta y eliminar cualquier fuente de variabilidad que pueda tener un efecto negativo en los resultados finales. Estas herramientas se aplican en función del tipo de datos de los que se dispone una vez que estos se han agrupado en matrices o cubos de datos dependiendo del orden del tensor [124,129]. Habitualmente el pre-procesado de la señal suele constar de los siguientes pasos:

1. Una **corrección de la línea base** con la finalidad de eliminar los efectos de fondo provocados por la propia variabilidad instrumental entre diferentes análisis y que dan lugar a desviaciones sistemáticas sobre la línea base de la señal. Durante esta corrección también se produce el ajuste del valor cero, de forma que el mínimo valor registrado en la señal sea siempre cero.
2. Un **filtrado del ruido**, también conocido como suavizado de la señal, con la finalidad de eliminar toda señal que es independiente de la información analítica. Esta eliminación se refiere al ruido que acompaña a la señal analítica de interés y que no aporta ningún tipo de información relativa a la muestra. Es importante controlar el suavizado muy estricto podría dar lugar a pérdida de información analítica. Existen múltiples algoritmos que suavizan la señales, sin embargo, el más utilizado es el sugerido por Savitzky y Golay [130].
3. Un **escalado o normalización de la intensidad y la dispersión** de la señal. La normalización en intensidad se lleva a cabo con respecto a una señal de referencia. La señal de referencia puede ser el valor de intensidad máximo independiente de la información química relativa a la señal o bien el valor de intensidad de una señal concreta relacionada con una información química previamente conocida. Este último caso sería, por ejemplo, la normalización con respecto al valor de intensidad de un patrón interno añadido en la muestra antes de llevar a cabo el análisis. También es habitual restar a cada valor la media, lo que se denomina un centrado en la media, con lo que se consigue que la señal se desarrolle alrededor del cero.

La normalización de la dispersión se hace generalmente dividiendo cada valor por la desviación estándar, y es habitual aplicarlo después de un centrado en la media. La operación en su conjunto recibe la denominación de auto-escalado.

4. Una **reducción de variables** (también conocido como decimación) con la finalidad de reducir las dimensiones de la señal, por ejemplo, seleccionando una de cada dos variables. Esta etapa es especialmente útil cuando hay una gran cantidad de datos disponibles como datos de orden 2 o superior los cuales consumen por completo los recursos informáticos dando lugar a errores en el procesamiento del cálculo o incluso incapacitando el desarrollo del modelo quimiométrico. Es aconsejable que la proporción empleada no sea superior a 10, ya que de esta manera el perfil de la huella instrumental permanece "inalterado", y con ello la información de interés.
5. En algunos casos es necesario llevar a cabo un **alineamiento de picos** para corregir los desplazamientos en las puntuaciones de posición, como por ejemplo las señales procedentes de instrumentos de cromatografía. Los posibles desplazamientos se corrigen mediante la aplicación de algoritmos de alineamiento que determinan una función de deformación que lleve a cabo el alineamiento de picos hacia una señal concreta denominada como objetivo. Existen múltiples algoritmos de alineamiento algunos de los más empleados los algoritmos 'Correlation Optimized Warping' (COW), 'Semi-parametric Time Warping' (STW) 'Interval Correlation Optimised Shifting' (ICOSHIFT) [131,132,133].

Como estrategia general los 4 primeros pasos suelen aplicarse independientemente del tipo de técnica instrumental aplicada, mientras que el quinto paso, el alineamiento, tan solo se lleva a cabo en señales donde los valores de posición varían con el paso del tiempo.

I.5.1. Hacia el agnosticismo instrumental

En el caso de no existir dicha dependencia, los resultados obtenidos tras el pre-procesado son altamente reproducibles, además en este tipo de señales existe la posibilidad de desarrollar metodologías de transferencia calibración, como en el caso de los instrumentos basados en NIR. Esta metodología permite transferir el

comportamiento de un primer instrumento denominado como padre, maestro o anfitrión a un segundo instrumento denominado hijo, esclavo o invitado, de forma que las señales obtenidas por ambos instrumentos son totalmente comparables unas con otras, independientemente del lugar y del momento de su adquisición. Esta metodología implica: i) el análisis de un conjunto de muestras externas de referencia en el instrumento maestro para generar un modelo de calibración inicial; ii) transferencia del modelo inicial al instrumento esclavo; iii) medida tanto del conjunto de muestras como las referencias externas en el instrumento esclavo; y iv) aplicación del modelo de transferencia con el que directamente se lleva a cabo la corrección de las diferencias residuales de las señales [134,135].

Por su parte, las medidas tomadas en sistemas en flujo, como por ejemplo aquellas que provienen de técnicas separativas, si presentan una fuerte dependencia de las puntuaciones de posición con respecto al tiempo. El motivo es la propia dependencia de la señal tanto del instrumento como de las pequeñas desviaciones producidas por las condiciones de funcionamiento del mismo, como por ejemplo la temperatura, ligeros cambios en el caudal del gas portados o en los porcentajes de mezcla de fases móviles, la degradación de la fase estacionaria, etc. Esta dependencia hace que el uso de algoritmos de alineamiento sea más que necesario para lograr disminuir el efecto que tiene la deformación cromatográfica sobre los métodos multivariantes [136]. Sin embargo, a diferencia de lo que sucede con las señales procedentes del NIR, los resultados de pre-procesado no son completamente reproducibles. El motivo es que los algoritmos de alineamiento suelen aplicarse internamente en conjunto de señales previamente establecidos sin hacer uso de referencias externas. La falta de referencias externas dificulta la posibilidad de transferencia de los parámetros de alineamiento inicialmente considerados a conjuntos de señales que previamente no han sido tenidos en cuenta antes de la aplicación de dichos algoritmos. Como consecuencia dos conjuntos de señales de la misma naturaleza y alineados bajo los mismos parámetros, pero procedentes de diferentes equipos podrían generar resultados diferentes que dificultan su comparación. Otra dificultad añadida aparece con la presencia de artefactos o nuevos picos cromatográficos, los cuales pueden generar confusión con picos vecinos haciendo que el alineamiento no se lleve a cabo de forma correcta.

Fuera de la aplicación de algoritmos de alineamiento, a menudo se lleva a cabo la transformación del tiempo de retención a un nuevo sistema de tiempos de retención relativos. Estos valores se obtienen dividiendo el eje de tiempos de retención entre el

valor de tiempo de retención experimental de un patrón interno, practica que compensa ligeras desviaciones y que haría de las señales comparables hasta cierto punto. Sin embargo, los retrasos producidos durante los primeros minutos, el uso de diferentes programas de temperatura/porcentajes de mezcla de disolventes y el mal ajuste que se produce a medida que aumenta la diferencia entre el tiempo de retención del patrón interno con los compuestos finales del análisis hacen que la compensación no sea tan eficaz como cabría esperar.

Otra opción que surge en GC es la propuesta por Kovats y ampliada por Van Den Dool para llevar a cabo la identificación de compuestos desconocidos la cual hace uso de índices de retención (I_R) obtenidos a partir de una serie homóloga de n-alcanos [137,138,139]. Los I_R constituyen una serie de parámetros empíricos obtenidos por la interpolación del tiempo de retención del compuesto de interés con los tiempos de retención de los patrones o estándares que eluyen inmediatamente antes y después de dicho compuesto de interés. De esta forma se produce el cambio del dominio del tiempo de retención dependiente del instrumento a un nuevo dominio basado en índices empíricos que es independiente ajeno al instrumento y al tiempo de retención. Si existiese una metodología capaz de combinar la estrategia propuesta por Kovats y Van Den Dool, con herramientas informáticas para el pre-procesado de señales sería posible establecer un sistema de transferencia similar al desarrollado en NIR que permitiese obtener huellas dactilares cromatográficas reproducibles tanto en el tiempo como en cualquier laboratorio de análisis.

I.6. JUSTIFICACIÓN E HIPÓTESIS

La huella dactilar cromatográfica contiene una gran cantidad de información no especificada acerca de la identidad y/o estado del alimento analizado. Por tanto, cualquier cambio significativo observado en el perfil de la huella cromatográfica de una misma muestra obtenida en dos momentos diferentes podría ser:

- i) indicativo de una posible mezcla no autorizada de dicho alimento con otras sustancias (ya sea sustancias químicas o mezcla de otros alimentos de peor calidad);
- ii) degradación del alimento como consecuencia del propio paso del tiempo.

El principal inconveniente de la comparación de dos huellas cromatográficas mediante análisis de la similitud, es la dependencia de las mismas del momento en que fueron

obtenidas o del estado del instrumento entre otras causas. Esta dependencia podría solventarse mediante el empleo de patrones químicos que corrigiesen algunos aspectos de la señal como la intensidad de la misma o, para el caso de métodos cromatográficos, los tiempos de retención, siendo esta última la que causa mayores diferencias entre señales. Para el caso de los tiempos de retención una estrategia similar a la empleada por Kovats y Van Den Dool sería la solución idónea para transferir el comportamiento de una serie de patrones externos al resto de muestras analizadas en la misma secuencia de análisis.

En base a esta hipótesis y sumado al creciente interés de los consumidores por disponer de mayor información sobre los alimentos que consumen, se plantea hacer uso de la comparación de huellas cromatográficas estandarizadas para la resolución de problemas relacionados con la calidad de alimentos, poniendo en el punto de mira dos alimentos apreciados como pueden ser el **aceite de oliva** o la **avellana**. El primero por ser un ingrediente importante en la dieta mediterránea y el segundo por ser un ingrediente muy utilizado en repostería (chocolates, bombones, praliné, helados, galletas, etc.)

El **aceite de oliva virgen** y en especial el **aceite de oliva virgen extra**, es uno de los alimentos de mayor interés, no solo por sus excelentes propiedades organolépticas, sino también por los beneficios que reporta sobre la salud [140]. Además, es en el único alimento en el que se obliga a los organismos de control oficial a que realicen un análisis sensorial para verificar su denominación de venta para su comercialización. En este caso, el análisis sensorial se lleva a cabo mediante la aplicación del método de 'panel test' propuesto por el Consejo Oleícola Internacional (COI) y reconocido por la Unión Europea de acuerdo al Reglamento (CEE) 2568/91 [44,45]. Aunque se trata de un método robusto y estadísticamente fiable, hay evidencias de la existencia de discrepancias entre los resultados emitidos por diferentes paneles de cata para una misma muestra.

Estas discrepancias pueden tener un origen muy diverso y se solucionarían fácilmente mediante el uso, por parte de los paneles de cata, de materiales de referencia certificados (MRC). Aunque existen numerosos métodos analíticos para la certificación de materiales de referencia de diferente naturaleza, apenas se dispone de información y/o métodos analíticos que permitan la certificación de MR destinados al análisis sensorial de aceite de oliva.

Independientemente de su categoría comercial, el aceite de oliva virgen tiene un precio de venta superior a otros aceites vegetales debido a que el proceso de producción del mismo es puramente mecánico. El elevado coste de este producto lo hace un candidato idóneo a ser adulterado con otros aceites de peor calidad con el fin de obtener un mayor margen de beneficios de forma ilícita. Estas mezclas no autorizadas podrían darse en cualquier etapa del proceso de producción, incluido el transporte en grandes cisternas. En este contexto, disponer de una metodología basada en la comparación de resultados de una alícuota de muestra tomada y analizada en origen y destino, permitiría poner de manifiesto aquellos casos en los que tienen lugar este tipo de prácticas fraudulentas. De este modo los organismos gubernamentales dispondrían de métodos rigurosos que les ayuden a combatir y penalizar este tipo de prácticas.

Por otra parte, las **avellanas** son el segundo fruto seco más relevante en el mercado mundial, siendo sus principales productores Turquía e Italia. Este fruto seco se encuentra ampliamente extendido debido a su uso como base en chocolates y otros tipos de productos horneados [141]. Dado que más del 90% de la producción mundial de avellanas se destinan a la elaboración de productos de confitería, los cultivos suelen seleccionarse y adecuarse para la obtención de frutos secos uniformes y de alta calidad, cualidades que a menudo vienen establecidas por la propia industria de la confitería [142,143,144]. Sin embargo, no todas las avellanas presentan las mismas propiedades física y organolépticas, ya que estas dependen en gran medida de la variedad y de la región en donde se producen.

Además, estos alimentos pueden sufrir procesos de degradación en el propio cultivo, durante su cosecha o durante su almacenamiento lo cual afecta negativamente a los parámetros de calidad sensoriales de los mismos [141]. Por ello, el análisis de compuestos orgánicos volátiles de avellanas de diferentes cosechas y cultivos, almacenadas en diferentes condiciones y analizadas periódicamente ayudaría a comprender mejor que parámetros se encuentran relacionados con la calidad de las avellanas, así como los cambios que se producen durante la conservación de los mismos.

I.7. OBJETIVOS

La presente Tesis Doctoral tiene por objetivo general la aplicación de la metodología de huella instrumental obtenida mediante técnicas cromatográficas (GC y LC) de diferentes familias de compuestos presentes en alimentos, fundamentalmente asociados a propiedades/atributos organolépticos, combinando dicha metodología con técnicas quimiométricas multivariantes que permitan la resolución de problemas relacionados con la calidad de los alimentos.

De este objetivo general derivan los siguientes objetivos específicos:

- Obtener huellas cromatográficas independientes (agnostizadas) del momento de adquisición de la señal y/o del estado del equipo utilizado para ello.
- En caso de ser necesario, desarrollar o modificar la metodología analítica necesaria para su obtención.
- Desarrollar e implementar procedimientos que se basen en el uso combinado del análisis sensorial e instrumental para facilitar la certificación de muestras reales de aceite de oliva virgen como MR.
- Evaluar la similitud de la huella cromatográfica de la fracción de compuestos volátiles orgánicos para asegurar la homogeneidad de las muestras de aceite y monitorizar su evolución con el paso del tiempo.
- Evaluar y proponer el uso de herramientas estadísticas complementarias (i.e. más robustas) a la propuesta por el COI que permitan minimizar los efectos de valores anómalos, buscando un valor de consenso más representativo de los datos proporcionados por los paneles de cata.
- Desarrollar métodos analíticos rápidos que no requieran de tratamientos complicados de muestras para poner de manifiesto la presencia de mezclas de aceites no autorizadas que hayan podido producirse durante el transporte de aceites en grandes cisternas.
- Evaluar la agnostización de huellas bidimensionales cromatográficas (espectro-cromatogramas 2D) como herramienta para obtener perfiles de similitud que permitan evaluar posibles mezclas fraudulentas en base a los cambios producidos en las familias de compuestos relacionadas con el color del aceite de oliva.
- Desarrollar diversos modelos de clasificación junto con diferentes opciones de pre-procesado que permitan evaluar la calidad de las avellanas

cultivadas en el Piamonte (variedad *Tonda Gentile Trilobata*) frente a muestras procedentes de otras regiones de Italia y Turquía. Para ello se estudia el uso de las propias muestras para obtener huellas instrumentales agnostizadas que permitan corregir la dependencia de la señal no solo con el tiempo de adquisición sino también el estado del instrumento.

REFERENCIAS

- [1] UNE-EN-ISO 9000 (2015). Sistemas de gestión de la calidad. Fundamentos y vocabulario, Madrid, *Agencia Española de Normalización*.
- [2] Ley 28/2015, de 30 de julio (2015), para la defensa de la calidad alimentaria. *Boletín Oficial del Estado*, 182, 65889.
- [3] Torres Saumeth, K.M., Solís Ospino, L., Ruiz Afanador, T.S. (2012). Calidad y su evolución: una revisión, *Dimensión Empresarial*, 10(2), pp.100–107.
- [4] Winterhalter, P. (2007). Authentication of Food and Wine. In S.E. Ebeler, G.R. Takeoka & P. Winterhalter (Eds.), *Authentication of Food and Wine*, (pp. 2–12). Washington DC, American Chemical Society.
- [5] Programa Conjunto FAO/OMS sobre Normas Alimentarias (2005). CODEX ALIMENTARIUS. Alimentos producidos orgánicamente, Roma.
- [6] Schieber, A. (2018). Introduction to Food Authentication. In D.W. Sun (Ed.), *Modern Techniques for Food Authentication*, 2nd ed. (pp. 1–21). Amsterdam, Elsevier.
- [7] Singhal, S.R., Kulkarni, P.R. & Rege, D.V. (1997). The Development of the Concept of Food Quality, Safety and Authenticity. In R.R. Singhal, P.R. Kulkarni & D.V. Rege (Eds.), *Handbook of Indices of Food Quality and Authenticity*, (pp. 9–31). Cambridge, Woodhead Publishing Limited.
- [8] Bánáti, D. (2011). Consumer response to food scandals and scares. *Trend in Food Sciences & Technology*, 22(2-3), 56–60.
- [9] Comisión Europea (1997), Libro verde de la comisión. Principios generales de la legislación alimentaria de la Unión Europea. Bruselas, 30.04.1997 COM(97) 176 final.
- [10] Comisión Europea (2000), Libro blanco sobre seguridad alimentaria. Bruselas, 12.1.2000 COM(1999) 719 final.
- [11] Brimer, L. (2011). The EU with EFSA and EMEA. in L. Brimer (Ed.) *Chemical Food Safety*, (pp. 75–79). Massachusetts, CABI.
- [12] Comisión Europea (2002). Reglamento (CE) nº 178/2002 del Parlamento Europeo y del Consejo, de 28 de enero de 2002 por el que se establecen los principios y los requisitos generales de la legislación alimentaria, se crea la Autoridad Europea de Seguridad Alimentaria y se fijan los procedimientos relativos a la seguridad alimentaria. *Diario Oficial de la Unión Europea*, L31/1
- [13] Comisión Europea (2004). Reglamento (CE) nº 853/2004 del Parlamento Europeo y del Consejo, de 29 de abril de 2004 relativo a la higiene de los productos alimentarios. *Diario Oficial de la Unión Europea*, L139/1

- [14] Comisión Europea (2004). Reglamento (CE) nº 853/2004 del Parlamento Europeo y del Consejo, de 29 de abril de 2004 por el que se establecen normas específicas de higiene de los alimentos de origen animal destinados al consumo humano. *Diario Oficial de la Unión Europea*, L139/55
- [15] Comisión Europea (2004). Reglamento (CE) nº 854/2004 del Parlamento Europeo y del Consejo, de 29 de abril de 2004 por el que se establecen normas específicas para la organización de controles oficiales de los productos de origen animal destinados al consumo humano. *Diario Oficial de la Unión Europea*, L139/206
- [16] Comisión Europea (2004). Reglamento (CE) nº 882/2004 del Parlamento Europeo y del Consejo, de 29 de abril de 2004 sobre los controles oficiales efectuados para la verificación del cumplimiento de la legislación en materia de piensos y alimentos y la normativa sobre salud animal y bienestar de los animales. *Diario Oficial de la Unión Europea*, 165/1
- [17] Comisión Europea (2005). Reglamento (CE) nº 183/2005 del Parlamento Europeo y del Consejo, de 12 de enero de 2005, por el que se fijan requisitos en materia de higiene de los piensos. *Diario Oficial de la Unión Europea*, L35/1
- [18] Comisión Europea (2017). Reglamento (UE) nº 625/2017 del Parlamento Europeo y del Consejo, de 15 de marzo de 2017, relativo a los controles y otras actividades oficiales realizados para garantizar la aplicación de la legislación sobre alimentos y piensos, y de las normas sobre salud y bienestar de los animales, sanidad vegetal y productos fitosanitarios. *Diario Oficial de la Unión Europea*, L95/1
- [19] Real Decreto 1086/2020, de 9 de diciembre por el que se regulan y flexibilizan determinadas condiciones de aplicación de las disposiciones de la Unión Europea en materia de higiene de la producción y comercialización de los productos alimenticios y se regulan actividades excluidas de su ámbito de aplicación. *Boletín Oficial del Estado Núm.* 322
- [20] Zhang, A., Mankad, A., & Ariyawardana, A. (2020). Establishing confidence in food safety: is traceability a solution in consumers' eyes? *Journal of Consumer Protection and Food Safety*, 15(2), 99–107.
- [21] ISO 5492 (2008). Sensory analysis – Vocabulary, Geneva: International Organization for Standardization.
- [22] ISO 8586 (2012) Sensory analysis – General guidelines for the selection, training and monitoring of selected assessors and expert sensory assessors, Geneva: International Organization for Standardization.
- [23] ISO Guide 30 (2015). Reference materials – Selected terms and definitions, Geneva: International Organization for Standardization.

-
- [24] Cont L., Bendini A., Valli E., Lucci P., Moret S., Maquet A., Lacoste F., Brereton P., Gacía-González D.L., Moreda W. & Toschi T. (2019). Olive oil quality and authenticity: A review of current EU legislation, standards, relevant methods of analyses, their drawbacks and recommendations for the future, *Trends in Food Science & Technology*, 105, 483–493.
- [25] ISO Guide 35 (2017) Reference materials — Guidance for characterization and assessment of homogeneity and stability, Geneva: International Organization for Standardization.
- [26] Ortega Gavilán F., García Mesa J.A., Marzal Fernández J.C., Moreno Ballesteros F.J., Rodríguez García F.P., González Casado A. & Cuadros Rodríguez L. (2022). Innovative and thorough practice to certify reference materials for sensory defects of olive oil, *Food Chemistry*, 380, 132195.
- [27] Barbieri S., Brkic Bubola K., Bendini A., Bucar-Miklavcic M., Lacoste F., Tibet U., Winkelmann O., García González D.L., & Gallina Toschi T. (2020). Alignment and proficiency of virgin olive oil sensory panels: The OLEUM approach, *Foods*, 9, 355.
- [28] Aparicio Ruiz R., Barbieri S., Gallina Toschi T. & García González D.L. (2020). Formulations of rancid and winey-vinegary artificial olfactory reference materials (AORMs) for virgin olive oil sensory evaluation, *Foods*, 9, 1870.
- [29] Comisión Europea (2008). Reglamento (CE) No 1334/2008 sobre los aromas y determinados ingredientes alimentarios con propiedades aromatizantes utilizados en los alimentos y por el que se modifican el Reglamento (CEE) n o 1601/91 del Consejo, los Reglamentos (CE) n o 2232/96 y (CE) n o 110/2008 y la Directiva 2000/13/CE (Versión consolidada 02008R1334). *Diario Oficial de la Unión Europea*, L354
- [30] Lewkowska, P., Dymerski, P. & Namieśnik, J. (2015). Use of sensory analysis methods to evaluate the odor of food and outside air. *Critical Reviews in Environmental Science and Technology*, 45, 2208–2244.
- [31] ISO 6658 (2017), Sensory analysis — Methodology — General guidance, Genova: International Organization for Standardization.
- [32] ISO 5495 (2005), Sensory analysis — Methodology — Paired comparison test, Genova: International Organization for Standardization.
- [33] ASTM E2263 – 12 (2018), Standard test method for paired preference test, West Conshohocken, USA, ASTM International.
- [34] ISO 4120 (2004), Sensory analysis — Methodology — Triangle test, Genova: International Organization for Standardization.
- [35] ASTM E1885 – 04 (2011), Standard test method for sensory analysis — Triangle test, West Conshohocken, USA, ASTM International.
-

- [36] ISO 10399 (2017), Sensory analysis — Methodology — Duo-trio test, Genova: International Organization for Standardization.
- [37] ASTM E2618 – 18 (2018) Standard test method for sensory analysis — Duo-trio test, West Conshohocken, USA, ASTM International.
- [38] ISO 8588 (2017), Sensory analysis — Methodology — “A” – “not A” test, Genova: International Organization for Standardization.
- [39] ASTM E3009 – 15e1 (2016), Standard test method for sensory analysis—Tetrad test, West Conshohocken, USA, ASTM International.
- [40] ISO 13299 (2016), Sensory analysis — Methodology — General guidance for establishing a sensory profile, Genova: International Organization for Standardization.
- [41] ISO 4121 (2003), Sensory analysis — Guidelines for the use of quantitative response scales, Genova: International Organization for Standardization.
- [42] ISO 8587 (2006), Sensory analysis — Methodology — Ranking, Genova: International Organization for Standardization.
- [43] Schlich, P. (1993). Risk tables for discrimination tests. *Food Quality and Preference*, 4, 141–151.
- [44] COI/T.20/Doc. No 15 / Rev.10 (2018). Sensory analysis of olive oil. Method for the organoleptic assessment of virgin olive oil. Madrid: International Olive Council (IOC).
- [45] European Commission (2016). Regulation (EEC) No. 2568/91 on the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis (consolidated version 01991R2568). *Official Journal of the European Union*, L248, 1128.
- [46] Choi, N.E. & Han, J.H. (2015) What is the taste? In N.E. Choi & J.H. Han (Eds) *How flavor works: the science of taste and aroma*, 1st ed., (pp.1–14). Chichester: John Wiley & Sons.
- [47] Smith, D.V. & Margolskee, R.F. (2001). Making Sense of Taste. *Scientific American*, 284, pp. 32-39.
- [48] Parker, J.K. (2015) Introduction to aroma compounds in foods. In J.K. Parker, J.S. Elmore & L. Methven (Eds.) *Flavour development, Analysis and Perception in Food and Beverages*, 1st ed. (pp. 15–46). Cambridge: Woodhead Publishing.
- [49] Shallenberger, R.S. (1997). Taste recognition chemistry. *Pure and Applied Chemistry*, 69, 659–666.
- [50] Hayes, J.E. (2008). Transdisciplinary Perspectives on Sweetness. *Chemosensory Perception*, 1, 48–57.
- [51] Choi, N.E. & Han, J.H. (2015) The origin of the taste: Why do we taste? In N.E. Choi & J.H. Han (Eds) *How Flavor Works: the Science of Taste and Aroma*, 1st ed. (pp.15–46). Chichester: John Wiley & Sons.

- [52] Laffitte, A., Neiers, F. & Briand, L. (2017) Characterization of taste compounds: chemical structures and sensory properties. In E. Guichard, C. Salles, M. Morzel & A.M. Le Bon (Eds) *Flavour: From Food to Perception*, 1st ed. (pp. 154–191). Chichester: John Wiley & Sons.
- [53] Mannella, J.A. & Bobowski, N.K. (2015). The sweetness and bitterness of childhood: Insights from basic research on taste preferences. *Physiology & Behavior*, 152, 502–507.
- [54] Gehrens, M. & Meyerhof, W. (2015) Taste receptors. In J.K. Parker, J.S. Elmore & L. Methven (Eds). *Flavour Development, Analysis and Perception in Food and Beverages*, 1st ed. (pp. 296–329). Cambridge: Woodhead Publishing.
- [55] Grassin-Delybe, S., Salvator, H., Mantov, N., Abrial, C., Brollo, M., Faisy, C., Naline, E., Couderc, L.J. & Deliver, P. (2019). Bitter taste receptors (TAS2Rs) in human lung macrophages: receptor expression and inhibitory effects of TAS2R agonists. *Frontiers in Physiology*, 10, 1267.
- [56] Neta, E.R.C., Johanningsmeier, S.D., Drake, M.A. & McFeeters, R.F. (2007). A chemical basis for sour taste perception of acid solutions and fresh-pack dill pickles. *Journal of Food Science*, 72, S352–S359.
- [57] Richter, T.A., Caicedo, A. & Roper, S.D. (2003). Sour taste stimuli evoke Ca^{2+} and pH responses in mouse taste cells. *The Journal of Physiology*, 547, 475–483.
- [58] Lugaz, O., Pillias, A.M., Boireau-Ducept, N. & Faurion, A. (2005). Time–Intensity evaluation of acid taste in subjects with saliva high flow and low flow rates for acids of various chemical properties. *Chemical Senses*, 30, 89–103.
- [59] Gardner, R.J. (1980). Lipid solubility and the sourness of acids: implications for models of the acid taste receptor. *Chemical Senses*, 5, 185–194.
- [60] Breslin, P.A. & Spector, A.C. (2008). Mammalian taste perception. *Current biology*, 18, R148–R155.
- [61] Gaudette, N.J., Pietrasik, Z., Johnston, S.P. (2019). Application of taste contrast to enhance the saltiness of reduced sodium beef patties. *LWT – Food Science and Technology*, 116, 108585.
- [62] Mouritsen, O.G., Duelund, L., Petersen, M.A., Hartmann, A.L. & Frost, M.B. (2019). Umami taste, free amino acid composition, and volatile compounds of brown seaweeds. *Journal of Applied Phycology*, 31, 1213–1232.
- [63] Jelen, H. & Gracka, A. (2017). Characterization of aroma compounds: structure, physico-chemical and sensory properties. In E. Guichard, C. Salles, M. Morzel & A.M. Le Bon (Eds) *Flavour: From Food to Perception*, 1st ed. (pp. 126–153), Chichester: John Wiley & Sons.

- [64] Choi, N.E. & Han, J.H. (2015) What creates smell? In N.E. Choi & J.H. Han (Eds), *How Flavor Works: the Science of Taste and Aroma*, 1st ed. (pp.91–108). Chichester: John Wiley & Sons.
- [65] Tylewicz, U. Inchingolo, R. & Rodriguez-Estrada, M.T. (2021). Food aroma compounds. In C. M. Galanakis (Ed) *Nutraceutical and Functional Food Components*, 2nd ed. (pp.363–409). Amsterdam, Elsevier.
- [66] Longo, M.A. & Sanromán, M.A. (2006). Production of Food Aroma Compounds: Microbial and Enzymatic Methodologies. *Food Technology and Biotechnology*, 44, 335–353.
- [67] Frérot, E. (2017) Fats and oils. In A Buettner (Ed) *Handbook of Odor*, 1st ed. (pp. 223-260). New York, Springer.
- [68] Zaroual, H., Chénè, C., El Hadrami, E.M. & Karoui, R. (2021) Application of new emerging techniques in combination with classical methods for the determination of the quality and authenticity of olive oil: a review, *Critical Reviews in Food Science and Nutrition*, 1–21.
- [69] Skoog, D.A., Holler, F.J. & Crouch, S.R. (2019) *Principios de Análisis Instrumental* (7th ed.) Boston, Cengage Learning.
- [70] Kamal, G. M., Uddin, J., Tahir, M.S., Khalid, M., Ahmad, S. and Hussain, A. I. (2021). Nuclear magnetic resonance spectroscopy in food analysis. In M. S. Khan & M. S. Rahman (Eds) *Techniques to Measure Food Safety and Quality* (pp. 137–168). Cham, Springer.
- [71] Janovick, J., Spyros, A., Dais, P. & Hatzakis, E. (2020). Nuclear magnetic resonance. In Y. Picó (Ed), *Chemical Analysis of Food: Techniques and Applications*, 2nd Ed (pp. 135–175) London, Elsevier.
- [72] Penner, M.H. (2017). Basic principles of spectroscopy. In S.S Nielsen (Ed) *Food Analysis* 5th Ed (pp. 79–88). Cham, Springer.
- [73] Feng, C.H., Otani, C. (2021). Terahertz spectroscopy technology as an innovative technique for food: Current state-of-the-Art research advances. *Critical Reviews in Food Science and Nutrition*. 61, 2523–2543.
- [74] Nawrocka, A. & Lamorska, J. (2013). Determination of food quality by using spectroscopic methods. In S. Grundas, & A. Stępniewski, (Eds.). *Advances in Agrophysical Research* (pp.347-367). London, IntechOpen.
- [75] Abbas, O., Dardenne, P. & Baeten, V. (2020). Near-Infrared, Mid-Infrared, and Raman Spectroscopy. In Y. Picó (Ed), *Chemical Analysis of Food: Techniques and Applications*, 2nd Ed (pp. 77–134). London, Elsevier.
- [76] Rodríguez-Saona, L. (2017) Infrared and Raman spectroscopy. In S.S Nielsen (Ed) *Food Analysis* 5th Ed (pp. 89–106). Cham, Springer.

-
- [77] Haque, F., Bubli S.Y. & Khan M.S. (2021). UV–Vis spectroscopy for food analysis. In M. S. Khan & M. S. Rahman (Eds) *Techniques to Measure Food Safety and Quality* (pp. 169–193). Cham, Springer.
- [78] Penner, M.H. (2017). Ultraviolet, visible, and fluorescence spectroscopy. In S.S Nielsen (Ed) *Food Analysis*, 5th Ed (pp. 89–106). Cham, Springer.
- [79] Khan, M.S. & Rahman, S. (2021). Introduction on techniques to measure food safety and quality. In M. S. Khan & M. S. Rahman (Eds), *Techniques to Measure Food Safety and Quality* (pp. 1–10). Cham, Springer.
- [80] Smith, J.S. & Thakur, R.A. (2017). Mass Spectrometry. In S.S Nielsen (Ed), *Food Analysis*, 5th Ed (pp. 165–181). Cham, Springer.
- [81] do Lago, C.L., Daniel, D., Lopes, F.S., Cieslarová, Z. (2020) Electroforesis. In Y. Picó (Ed), *Chemical Analysis of Food: Techniques and Applications*, 2nd Ed (pp. 135–175) London, Elsevier.
- [82] Chen, G. (2016) Capillary electrophoresis with electrochemical detection. In A. Escarpa, M.C. González & M.A. López (Eds), *Agricultural and Food Electroanalysis* (pp. 117–133). Chichester, John Wiley & Sons Ltd.
- [83] Poole, C.F. (2021) Conventional detectors for gas chromatography. In C.F. Poole (Ed) *Gas Chromatography*, 2nd ed (pp. 343–370). London, Elsevier.
- [84] Lundanes, E., Reubsaet, L. & Greibrokk, T. (2014) Gas chromatography. In E. Lundanes, L. Reubsaet & T. Greibrokk, (Eds), *Chromatography: Basic Principles, Sample Preparations and Related Methods* (pp. 17–46). Weinheim, Wiley-VCH.
- [85] Feng, T., Sun, M., Song, S., Zhuang, H. & Yao L. (2019) Gas chromatography for food quality evaluation. In J. Zhong & X. Wang (Eds), *Evaluation Technologies for Food Quality* (pp 219–265). Duxford, Woodhead Publishing.
- [86] Poole, C.F. (2019), Conventional gas chromatography: basic principles and instrumental aspects. In P.Q. Tranchida (Ed) *Advanced gas chromatography in food analysis* (pp. 83–130). Cambridge, Royal Society of Chemistry.
- [87] Esteki, M., Simal-Gandara, J., Shahsavari, Z., Zandbaaf, S., Dashtaki, E. & Vander Heyden, Y. (2018) A review on the application of chromatographic methods, coupled to chemometrics, for food authentication. *Food Control*, 93, 165–182
- [88] Bagur González, M.G., Jiménez Carvelo, A.M., Ortega Gavilán, F. & González Casado A. (2021) Chromatographic methods. In M. Galanakis (Ed) *Food Authentication and Traceability*, (pp 65–99). London, Elsevie.
- [89] Lundanes, E., Reubsaet L. & Greibrokk, T. (2014) High performance liquid chromatography. In Lundanes, E., Reubsaet L. & Greibrokk, T. (Eds) *Chromatography:*

- Basic Principles, Sample Preparations and Related Methods (pp. 47–105). Weinheim, Wiley-CH.
- [90] Fanali, C., Dugo, P., Hermnandes Borges, J., Mondello L., Aturki, Z. & Fanali, S. (2015) High-Performance Liquid Chromatography: An Established Separation Technique in Food Chemistry. In L.M.L. Nollet & F. Toldra (Eds) *Handbook of Food Analysis*, 3rd ed (pp 459–479). Boca Raton, CRC Press.
- [91] Esteki, M., Shahsavari, Z. & Simal-Gandara. J. (2019) Food identification by high performance liquid chromatography fingerprinting and mathematical processing. *Food Research International*, 122, 303–317.
- [92] Dorman, F.L. & Dawes, P. (2021) Column technology: open-tubular columns. In C.F. Poole (Ed), *Gas Chromatography*, 2nd ed (pp. 99–116). London, Elsevier.
- [93] Rood, D. (2007) Detectors. In D. Rood, *The Troubleshooting and Maintenance Guide for Gas Chromatographers* (pp 139–185). Weinheim, Wiley-VCH.
- [94] McNair, H.M. & Miller, J.M. Detectors. In H.M. McNair & J.M. Miller (Eds) *Basic gas chromatography*, 2nd ed (pp 104–128). Chichester, John Wiley & Sons Ltd.
- [95] Quian, M.C., Peterson, D.G. & Reineccius, G.A. (2017), *Gas Chromatography*. In S.S. Nielsen (Ed) *Food analysis*, 5th ed (pp. 227–253). Cham, Springer.
- [96] Janssen, H.G., García Cicourel, A. & Tranchida, P.Q. (2019), *Conventional gas chromatography: mass spectrometry hyphenation and applications in food analysis*. In P.Q. Tranchida (Ed). *Advanced Gas Chromatography in Food Analysis* (pp. 131–168). Cambridge, Royal Society of Chemistry.
- [97] Kaklamanos, G., Apreb, E. & Theodoridis G. (2020) Mass spectrometry: principles and instrumentation. In Y. Picó (Ed) *Chemical analysis of food: techniques and applications* 2nd Ed (pp. 525–552). London, Elsevier.
- [98] Bajkacz, S. & Kycia-Slocka, E. (2020) Liquid chromatography in food analysis. In Y. Picó (Ed) *Chemical analysis of food: techniques and applications* 2nd Ed (pp. 391–455). London, Elsevier.
- [99] Fanali, C., Dugo, P., Mondello, L., D'Orazio, G. & Fanali, S. (2013) Recent developments in high-performance liquid chromatography. In L.M.L. Nollet & F. Toldra (Eds) *Food Analysis by HPLC*, 3rd ed. (pp. 1–32). Boca Raton, CRC Press, Florida.
- [100] Nuñez, O. & Lucci P. (2020). Application of Liquid Chromatography in Food Analysis, *Foods*, 9, 1277.
- [101] Morgan, N.Y. & Smith, P.D. (2011) HPLC Detectors. In Corradini D. (Ed) *Handbook of Food Analysis* 2nd ed (pp. 207–231), Florida, CRC Press.

-
- [102] Snyder, L.R., Kirkland, J.J. & Dolan J.W. (2010) Detection. In L.R. Snyder, J.J. Kirkland & J.W. Dolan (Eds) *Introduction to Modern Liquid Chromatography* 3rd ed (pp. 147–197) New Jersey, John Wiley & Sons.
- [103] Shackman, J.G. (2013) General instrumentation. In S. Fanili, P. Haddad, C. Poole, P. Schoenmakers & D. Lloyd (Eds) *Liquid Chromatography. Fundamental and Instrumentation* (pp. 283–306) London, Elsevier.
- [104] di Rosa, A.R., Leone, F. & Chiofalo, V. (2020). Electronic noses and tongues. In Y. Picó (Ed) *Chemical Analysis of Food: Techniques and Applications* 2nd Ed (pp. 353–389), London, Elsevier.
- [105] Persaud, K. (2016) Electronic noses and tongues in the food industry. In V. Preedy & M.L. Rodríguez Méndez (Eds) *Electronic Noses and Tongues in Food Science* (pp.1–12) London, Elsevier.
- [106] Tan, J. & Xu, J. (2020). Applications of electronic nose (e-nose) and electronic tongue (e-tongue) in food quality-related properties determination: A review. *Artificial Intelligence in Agriculture*. 4, 104–115.
- [107] Nowshad, F. & Khan, M.S. (2021) Electronic tongue for food safety and quality assessment. In M. S. Khan & M. S. Rahman (Eds) *Techniques to Measure Food Safety and Quality* (pp. 229–247) Cham, Springer.
- [108] Hui, G. (2016) Apple Analysis and the Electronic Nose. In V. Preedy & M.L. Rodríguez Méndez (Eds) *Electronic Noses and Tongues in Food Science* (pp.15–19) London, Elsevier.
- [109] Lvova, L. (2016) Electronic tongue principles and applications in the food industry. In V. Preedy & M.L. Rodríguez Méndez (Eds) *Electronic Noses and Tongues in Food Science* (pp.151–160) London, Elsevier.
- [110] Ballin, N.Z. & Laursen, K.H. (2019) To target or not to target? Definitions and nomenclature for targeted versus non-targeted analytical food authentication, *Trends in Food Science & Technology*, 86; 537-543.
- [111] Nuñez, O. (2020) Trends of Gas Chromatography-Mass Spectrometry Techniques. In P. Henricho (Ed). *Food Authentication in Gas Chromatography*, New York, Nova Science Publishers Inc.
- [112] Cuadros Rodriguez, L., Ruiz Samblás, C., Valverde Som, L., Pérez Castaño, E. & González Casado, A. (2016) Chromatographic fingerprinting: An innovative approach for food 'identification' and food authentication – A tutorial. *Analytica Chimica Acta*, 909, 9–23.
- [113] Bagur González, M.G., Pérez Castaño, E., Sanchez Viñas, M. & Gázquez Evangelista, D. (2015) Using the liquid-chromatographic-fingerprint of sterols fraction to discriminate virgin olive from other edible oils, *Journal of Chromatography A*, 1380, 64–70
-

- [114] Otto, M. (2017) *Chemometrics. Statistics and Computer Application in Analytical Chemistry* 3rd ed. Weinheim, Wiley-VCH.
- [115] Oliveri, P. Malegori, C. & Casale, M. (2020) *Chemometrics: multivariate analysis of chemical data*. In Y. Picó (Ed) *Chemical Analysis of Food: Techniques and Applications* 2nd Ed (pp. 33–75) London, Elsevier.
- [116] Hibbert, D.B. (2016) Vocabulary of concepts and terms in chemometrics (IUPAC recommendations 2016). *Pure and Applied Chemistry*, 88, 407–443
- [117] Brereton, R.G. (2007). *Applied Chemometrics for Scientists*, Chichester, John Wiley and Sons Ltd.
- [118] Buvé, C., Saeys, W., Rasmussen, M.A., Neckebroek, B., Hendricks, M., Grauwet, T. & Van Loey, B. (2022). Application of multivariate data analysis for food quality investigations: An example-based review. *Food Research International*, 151, 110878.
- [119] Granato, D., Putnik, P., Kovacevic, D. B., Santos, J. S., Calado, V., Rocha, R. S., Gomes Da Cruz, A., Jarvis, B., Rodionova, O. Y. & Pomerantsev, A. (2018), Trends in chemometrics: food authentication, microbiology, and effects of processing. *Comprehensive Reviews in Food Science and Food Safety*, 17, 663–677.
- [120] Abdi, H. & Williams, L.J. (2010) Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2, 433–459.
- [121] Rodríguez Saona, L.E., Giusti, M.M. & Shotts, M. (2016) Advances in infrared spectroscopy for food authenticity testing. In G. Downey (Ed) *Advances in food authenticity testing*, (pp. 71–117) London, Elsevier.
- [122] Oliveri, P. & Simonetti, R. (2016) Chemometrics for food authenticity applications. In G. Downey (Ed) *Advances in food authenticity testing*, (pp. 702–728) London, Elsevier.
- [123] Pérez Castaño, E., Medina Rodríguez, S. & Bagur González, M.G. (2019) Discrimination and classification of extra virgin olive oil using a chemometric approach based on TMS-4,4'-desmethylsterols GC(FID) fingerprints of edible vegetable oils, *Food Chemistry*, 274, 518–525
- [124] Pérez Castaño, E., Sánchez Viñas, E., Gázquez Evangelista, D. & Bagur González, M.G. (2018) Quantification of extra virgin olive oil in dressing and edible oil blends using the representative TMS-4,40-desmethylsterols gas-chromatographic-normalized fingerprint, *Food Chemistry*, 239, 1192–1199
- [125] Brereton, R.G. (2009). *Chemometrics for Pattern Recognition*, Chichester, John Wiley and Sons, Ltd.
- [126] Todeschini, R., Ballabio, D. & Consonni, V. (2020) Distances and similarity measures in chemometrics and chemoinformatics. *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation*, 1–40.

- [127] Bevilacqua, M., Nescatelli, R., Bucci, R., Magrì, A.D., Magrì, A.L. & Marini, F. (2014). Chemometric classification techniques as a tool for solving problems in analytical chemistry, *Journal of AOAC International*, 97, 19–28.
- [128] Livingstone, D. (2009). *A Practical Guide to Scientific Data Analysis*, Chichester, John Wiley and Sons Ltd.
- [129] Pérez Castaño, E. Ruiz-Samblás C., Mediona Rodríguez, S., Quirós Rodríguez, V., Jiménez Carvelo, A.M., Valverde Som, L., González Casado, A. & Cuadros Rodríguez, L. (2015) Comparison of different analytical classification scenarios: application for the geographical origin of edible palm oil by sterolic (NP) HPLC fingerprinting, *Analytical Methods*, 7, 4192–4201
- [130] Savitzky, A. & Golay, M.J.E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36, 1627–1639.
- [131] Nielsen, N., Carstensen, J. & Smedsgaard, J. (1998) Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping, *Journal of Chromatography A*, 805, 17–35.
- [132] Daszykowski, M., Vander Heyden, Y., Boucon, C. & Walczaka, B., (2010). Automated alignment of one-dimensional chromatographic fingerprints, *Journal of Chromatography A*, 1217, 6127–6133.
- [133] Savorani, F., Tomasi, G. & Engelsen, S.B. (2010). Icosift: A versatile tool for the rapid alignment of 1D NMR spectra, *Journal of Magnetic Resonance*, 202 190–202.
- [134] Wang, Y., Veltkamp, D.J. & Kowalski, B.R. (1991). Multivariate instrument standardization, *Analytical Chemistry*, 63, 2750–2756.
- [135] Workman, J. & Howard, M. (2020). One real challenge that still remains in applied chemometrics, *Spectroscopy* 35, 24–26.
- [136] Johnson, K.J., Wright, B.W., Jarman, K.H. & Synovec, R.E. (2003). High-speed peak matching algorithm for retention time alignment of gas chromatographic data for chemometric analysis, *Journal of Chromatography A*, 996, 141–155.
- [137] Kovats, E. (1958) Gas-chromatographische charakterisierung organischer verbindungen. Teil 1: retentions indices aliphatischer halogenide, alkohole, aldehyde und ketone, *Helvetica Chimica Acta*, 41, 1915–1932.
- [138] Van Den Dool, H. & Kratz, P.D. (1963). A generalization of the retention index system including linear temperature programmed gas liquid partition chromatography, *Journal of Chromatography*, 11, 463–471.
- [139] Tarjan, G., Nyiredy, Sz., Gyor, M., Lombosi, E.R., Lombosi, T.S., Budahegyi, M.V., Meszaros, S.Y. & Takacs, J.M. (1989). Thirtieth anniversary of the retention index

- according to Kováts in gas-liquid chromatography, *Journal of Chromatography A*, 472, 1–92.
- [140] Willett, W. C., Scacks, F., Trichopoulou, A., Drescher, G., Ferro Luzzi, A., Helsing, E., & Trichopoulos, D. (1995). Mediterranean diet pyramid: A cultural modern for healthy eating. *The American Journal of Clinical Nutrition*, 61, 1402S–1406S.
- [141] Stilo, F., Liberto, E., Spigolon, N., Genova S., Rosso, G., Fontana, M., Reichenbach S.E., Bicchi, C. & Cordero, C. (2021). 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*, 340, 128135.
- [142] Caligiani, A., Coisson J.D., Travaglia, F., Acquotti, D., Palla, G., Palla, L. & Arlorio, M. (2014). Application of ¹H NMR for the characterisation and authentication of “Tonda Gentile Trilobata” hazelnuts from Piedmont (Italy). *Food Chemistry*, 148, 77–85.
- [143] Göncüoğlu Tas, N. & Gokmen, V. (2017). Maillard reaction and caramelization during hazelnut roasting: A multiresponse kinetic study. *Food Chemistry*, 221, 1911–1922.
- [144] Yuan, B., Lu, M., Eskridge, K.M., Isom, L.D. & Hanna, M.A. (2018). Extraction, identification, and quantification of antioxidant phenolics from hazelnut (*Corylus avellana* L.) shells. *Food Chemistry*, 244, 7–15.

Capítulo II

**Evaluación de la homogeneidad en materiales de
referencia certificados**

Homogeneity assessment of reference materials for sensory analysis of liquid foodstuffs. The virgin olive oil as case study.

Food Chemistry 322 (2020) 126743



Contents lists available at ScienceDirect

Food Chemistry

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

Homogeneity assessment of reference materials for sensory analysis of liquid foodstuffs. The virgin olive oil as case study



Fidel Ortega-Gavilán^{a,*}, Lucía Valverde-Som^{a,b}, Francisco P. Rodríguez-García^c,
Luis Cuadros-Rodríguez^a, M. Gracia Bagur-González^a

^a Department of Analytical Chemistry, Faculty of Science, University of Granada, C/ Puentenueva s/n, E-18071 Granada, Spain

^b Department of Chemistry, Faculty of Science, Universidad de Burgos, Plaza Misael Bañuelos s/n, E-09001 Burgos, Spain

^c Institute of Agricultural and Fishery Research & Training (IFAPA) Consejería de Agricultura, Pesca y Desarrollo Sostenible, Junta de Andalucía, Sevilla, Spain

ARTICLE INFO

Keywords:

Homogeneity assessment
Reference materials
Sensory analysis
Virgin olive oil
Instrumental chromatographic analysis

ABSTRACT

Sensory properties are critical characteristics that determine quality and can be evaluated by trained tasting panels. The panels function as multi-sensor measuring instrument and need the use of reference materials (RMs) for training.

The homogeneity between units packaged from a batch of RM can be evaluated by gas chromatography coupled to flame ionization detection (GC-FID), using this instrumental technique as an alternative to sensory analysis. For this purpose, the fingerprint methodology is applied, taking into account that the homogeneity assessment will be based on evaluating the similarity between the fingerprints of the fraction of volatile organic compounds acquired from samples representative of the batch.

The proposed methodology is applied with good results to evaluate the homogeneity of several RMs for sensory analysis of virgin olive oil samples, using similarity indices, control charts and exploratory analysis of multivariate data to observe the grouping RM and fingerprint regions representative of each defect.

1. Introduction

A reference material (RM) is defined as a *sufficiently homogeneous and stable material with respect to one or more specified properties, which has been established to be fit for its intended use in a measurement process* (ISO Guide 30, 2015). The best option to improve the accuracy of the results from analysis is the using of certificated reference materials (CRMs) (Aparicio, Lanfranco, & Flebly, 2013), which are *RMs characterized by a metrologically valid procedure for one or more specified properties, accompanied by an RM certificate that provides the value of the specified property, its associated uncertainty, and a statement of metrological traceability* (ISO Guide 30, 2015). Performing three experimental studies, homogeneity, characterization and stability, is needed in order to ensure compliance with requirements for RM qualifying.

The homogeneity is defined as the *uniformity of a specified property value throughout a defined portion of a RM* (ISO Guide 30, 2015). The homogeneity study is carried out to assure that the value of the property or properties of interest regarding all packaged units of the batch are sufficiently similar (Pauwels, Lamberty, & Schimmel, 1998; Linsinger, Pauwels, Van der Veen, Schimmel, & Lamberty, 2001; Van der Veen, Linsinger, & Pauwels, 2001; Rutkowska et al., 2018).

Therefore, the packing process is one of the most critical stages since it has a higher impact on the homogeneity. The homogeneity study requires a high-precision measurement method in order to be able to detect minor and/or systematic changes between material units. Furthermore, stability study is necessary to ensure that the value of the property or properties previously characterized remains unchanged over time. The stability is defined as *characteristic of a reference material to maintain a specified property value within specified limits for a specified period of time* (ISO Guide 30, 2015). When the RM is produced from a biological matrix, as the case of foodstuffs, the stability study is carried out at the same time that the RM is being used; since these matrices are 'live' materials whose composition changes over the time. Characterization of a RM is the *determination of the property values or attributes of a reference material, as part of the production process* (ISO Guide 30, 2015). The characterization should be ideally carried out applying a set of high metrologically analytical methods. All the information related to the characterization study should be included in the RM certificate or provided to users by any other appropriate means (ISO Guide 35, 2017). In the case of a CRM, these values must be accompanied by uncertainty associated with the measurement processes applied. Each experimental study, homogeneity, stability and characterisation,

* Corresponding author.

E-mail address: fog@ugr.es (F. Ortega-Gavilán).

<https://doi.org/10.1016/j.foodchem.2020.126743>

Received 13 January 2020; Received in revised form 30 March 2020; Accepted 2 April 2020

Available online 03 April 2020

0308-8146/ © 2020 Elsevier Ltd. All rights reserved.

II.a. RESUMEN (EN ESPAÑOL)

Las propiedades sensoriales son características críticas que determinan la calidad y que pueden ser evaluadas por paneles de cata debidamente entrenados. Dichos paneles funcionan como un instrumento de medida multisensor y por ello necesitan el uso de materiales de referencia (MR) para su entrenamiento.

La homogeneidad entre las unidades envasadas de un lote de MR puede evaluarse mediante cromatografía de gases acoplada a un sistema de detección por ionización de llama (GC-FID), utilizando esta técnica instrumental como alternativa al análisis sensorial. Para ello se aplica la metodología de la huella instrumental, teniendo en cuenta que la evaluación de la homogeneidad se basará en la valoración de la similitud entre las huellas instrumentales correspondientes a la fracción de compuestos orgánicos volátiles, adquiridas de muestras representativas del lote.

La metodología propuesta se aplica con buenos resultados para evaluar la homogeneidad de varios MR para el análisis sensorial de muestras de aceite de oliva virgen, mediante el uso de índices de similitud, gráficos de control y análisis exploratorio de datos multivariantes con el objetivo de observar la agrupación de los MR, así como de las regiones de huellas instrumentales que resultan representativas de cada defecto.

II.b. CONCLUSIONES DEL TRABAJO (EN ESPAÑOL)

En este artículo se establece una metodología novedosa y precisa para la evaluación de la homogeneidad de muestras de aceite de oliva mediante el uso de huellas instrumentales obtenidas mediante GC-FID correspondientes a la fracción de compuestos orgánicos volátiles (COV), previamente aislados por microextracción en fase sólida.

Dicha estrategia cromatográfica sustituye el uso del análisis sensorial cuando la finalidad del estudio se encuentra orientada a la evaluación de la homogeneidad, facilitando la certificación de MR de aceites de oliva destinados a verificar las categorías de comercialización de este alimento asignadas por elaboradores y envasadores, y que, además, resultan ser herramientas clave para el entrenamiento y armonización de paneles sensoriales.

La estrategia propuesta, basada en la similitud de señales, permite revelar la falta de homogeneidad entre unidades muestreadas del MR candidato. Los gráficos de control obtenidos a partir de los índices de similitud permiten comprobar, de forma sencilla, cuándo se superan los criterios de aceptación de los valores límite, de forma similar a lo dispuesto en el proceso de certificación.

Además, utilizando una aproximación quimiométrica mediante métodos multivariantes de reconocimiento de pautas no supervisados (como el HCA y el PCA) se podría poner de manifiesto la agrupación natural de las unidades pertenecientes a cada uno de los materiales de referencia e identificar posibles marcadores químicos.

II.c. ESTUDIO BIBLIOMÉTRICO

De acuerdo con la información obtenida de la web *Journal Citation Reports* (JCR) (<https://jcr.clarivate.com/jcr/home>), la revista *Food Chemistry* se encuentra indexada dentro de tres categorías diferentes:

- 1) *Nutrition & dietetics* – SCIE. En el ámbito de la medicina clínica, en 2020 la revista se encuentra indexada dentro del primer decil (D1) y por tanto dentro del primer cuartil (Q1) del total de revistas indexadas en dicha categoría, ocupando el puesto 6 del ranking constituido por un total de 86 revistas.
- 2) *Chemistry, applied* – SCIE. En el ámbito de la química, de nuevo en 2020 la revista se encuentra indexada dentro del primer decil (D1) y por tanto también dentro del primer cuartil (Q1) de la categoría, ocupando el puesto 7 de un total de 74 revistas que constituyen la categoría *Chemistry, applied* – SCIE.
- 3) *Food science & technology* – SCIE. En el ámbito de la biología, biotecnología, química e ingenierías, en 2020 coincide con la tendencia observada en las categorías anteriores, indexándose dentro del primer decil (D1) y por tanto dentro del primer cuartil (Q1) en la posición 7 de un total de 144 revistas pertenecientes a dicha categoría.

La propia web del JCR recoge el número total de citas de los artículos que han sido publicados hasta ese mismo año en la revista *Food Chemistry*, acumulando un total de 156.901 citas, donde se incluyen las citas realizadas por un mismo autor de sus trabajos previos (autocitas) en esta revista. De acuerdo a esta información se obtiene como factor de impacto de la revista un valor de 7,514, el cual se reduce a un valor de 6,981 si no se tienen en cuenta las autocitas realizadas durante el año 2020.

Si se evalúa la tendencia del factor de impacto de la revista desde su creación en el año 1997, se observa un aumento gradual a lo largo de los años. En la Figura 17 se muestra una captura de pantalla del gráfico de evolución del factor de impacto realizado por la web JCR.

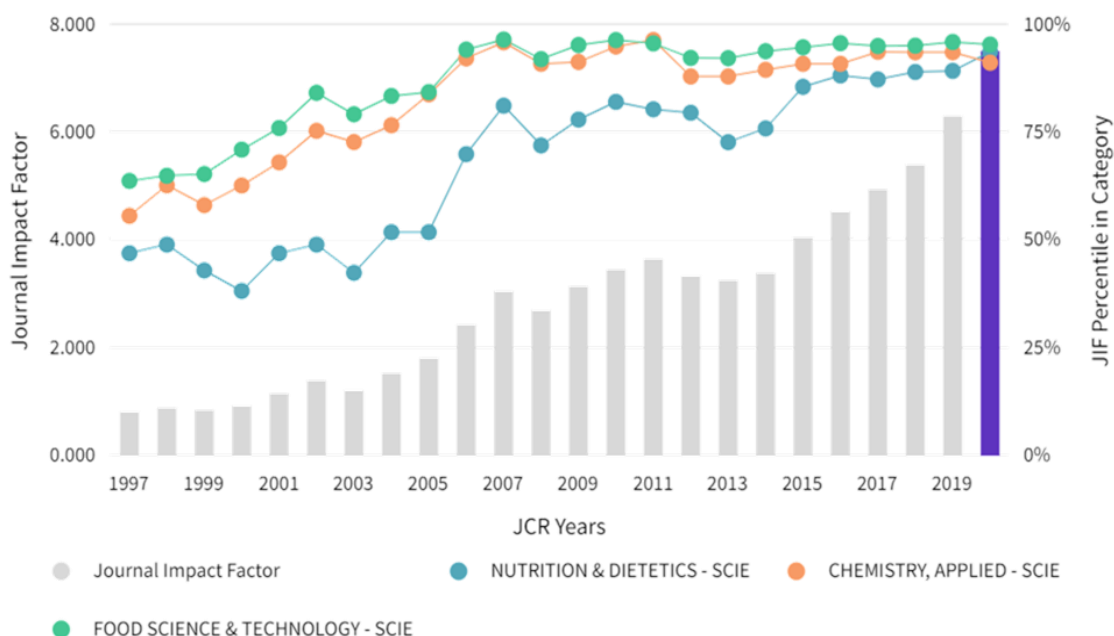


Figura 17. Gráfico de tendencia de la evaluación del factor de impacto de la revista *Food Chemistry* desde su creación en el año 1997 tomada de la información suministrada por el *Journal Citation Reports* (fecha de acceso 17 de enero de 2022).

<https://jcr.clarivate.com/jcr-jp/journal-profile?journal=FOOD%20CHEM&year=2020&fromPage=%2Fjcr%2Fhome>

Desde su publicación y hasta el cierre de la edición de esta Memoria, el artículo que se presenta en este Capítulo ha sido citado en 9 artículos científicos publicados en diversas revistas indexadas en el JCR. De estas 9 publicaciones 7 se encuentran publicadas en revistas del primer cuartil (Q1) de las categorías a las que pertenecen.

II.d. ARTICULO CIENTÍFICO

Homogeneity assessment of reference materials for sensory analysis of liquid foodstuffs. The virgin olive oil as case study

Fidel Ortega-Gavilán ^(1,✉), Lucía Valverde-Som^(1,2), Francisco P. Rodríguez-García⁽³⁾, Luis Cuadros-Rodríguez⁽¹⁾, M. Gracia Bagur-González⁽¹⁾

⁽¹⁾ *Department of Analytical Chemistry, Faculty of Science, University of Granada, C/ Fuentenueva s/n, E-18071, Granada, Spain*

⁽²⁾ *Department of Chemistry, Faculty of Science, Universidad de Burgos, Plaza Misael Bañuelos s/n, E-09001, Burgos, Spain*

⁽³⁾ *Institute of Agricultural and Fishery Research & Training (IFAPA) Consejería de Agricultura, Pesca y Desarrollo Sostenible, Junta de Andalucía, Sevilla, Spain*

HIGHLIGHTS

- Strategy for homogeneity assessment of RMs for sensory analysis has been proposed
- Fingerprint of olive oil has been used for establishing the similarity indices
- HCA and PCA have been applied for checking natural grouping of samples

KEYWORDS

Homogeneity assessment, reference materials, sensory analysis, virgin olive oil, instrumental chromatographic analysis.

✉ Corresponding author, e-mail: fog@ugr.es; Phone: +34 958240797.

Abstract

Sensory properties are critical characteristics that determine quality and can be evaluated by trained tasting panels. The panels function as multi-sensor measuring instrument and need the use of reference materials (RMs) for training.

The homogeneity between units packaged from a batch of RM can be evaluated by gas chromatography coupled to flame ionization detection (GC-FID), using this instrumental technique as an alternative to sensory analysis. For this purpose, the fingerprint methodology is applied, taking into account that the homogeneity assessment will be based on evaluating the similarity between the fingerprints of the fraction of volatile organic compounds acquired from samples representative of the batch.

The proposed methodology is applied with good results to evaluate the homogeneity of several RMs for sensory analysis of virgin olive oil samples, using similarity indices, control charts and exploratory analysis of multivariate data to observe the grouping RM and fingerprint regions representative of each defect.

1. INTRODUCTION

A reference material (RM) is defined as a *sufficiently homogeneous and stable material with respect to one or more specified properties, which has been established to be fit for its intended use in a measurement process* (ISO Guide 30, 2015). The best option to improve the accuracy of the results from analysis is the using of certificated reference materials (CRMs) (Aparicio, Lanfranco, & Fiebig, 2013), which are *RMs characterized by a metrologically valid procedure for one or more specified properties, accompanied by an RM certificate that provides the value of the specified property, its associated uncertainty, and a statement of metrological traceability* (ISO Guide 30, 2015). Performing three experimental studies, homogeneity, characterization and stability, is needed in order to ensure compliance with requirements for RM qualifying.

The homogeneity is defined as the *uniformity of a specified property value throughout a defined portion of a RM* (ISO Guide 30, 2015). The homogeneity study is carried out to assure that the value of the property or properties of interest regarding all packaged units of the batch are sufficiently similar (Pauwels, Lamberty, & Schimmel, 1998; Linsinger, Pauwels, Van der Veen, Schimmel, & Lamberty, 2001; Van der Veen, Linsinger, & Pauwels, 2001; Rutkowska et al., 2018). Therefore, the packing process is one of the most critical stages since it has a higher impact on the homogeneity. The homogeneity study requires a high-precision measurement method in order to be able to detect minor and/or systematic changes between material units. Furthermore, stability study is necessary to ensure that the value of the property or properties previously characterized remains unchanged over time. The stability is defined as *characteristic of a reference material to maintain a specified property value within specified limits for a specified period of time* (ISO Guide 30, 2015). When the RM is produced from a biological matrix, as the case of foodstuffs, the stability study is carried out at the same time that the RM is being use; since these matrices are 'live' materials whose composition changes over the time. Characterization of a RM is the *determination of the property values or attributes of a reference material, as part of the production process* (ISO Guide 30, 2015). The characterization should be ideally carried out applying a set of high metrologically analytical methods. All the information related to the characterization study should be included in the RM certificate or provided to users by any other appropriate means (ISO Guide 35, 2017). In the case of a CRM, these values must be accompanied by uncertainty associated with the measurement processes applied. Each experimental study, homogeneity, stability and

characterisation, informs about a particular uncertainty value which must be taken into account.

The olive oil (OO) obtained by olive fruit milling is the only foodstuff that must be subjected to sensory analysis to verify if the labelling correctly addresses the different marketing categories: extra virgin olive oil (EVOO) and virgin olive oil (VOO) (Mariotti, 2014). The European Commission Regulation (EEC) No 2568/91 describes the methods for the physicochemical and organoleptic assessment for the OO. The organoleptic or sensory assessment is defined as the perception and description of both its qualitative and quantitative flavour characteristics using human senses (mainly smell and taste) and its classification on the basis of those characteristics, as determined by a group of tasters selected, trained and monitored as a panel (COI/T.20/Doc. No 14, 2018; ISO 5492, 2008). 'Panel test' is a sensory method proposed by International Olive Council (IOC) to classify olive oil depending upon its organoleptic characteristics, which are measured by a group among 8 to 12 trained tasters who work as a multisensor analytical instrument (COI/T.20/Doc. No 15, 2018). The trained panellists evaluate the flavour of the olive oil samples, perceiving negative (or defective) and/or positive attributes of the OO samples. The identity and intensity of each detected attribute are marked on a profile sheet by each one of the tasters. Although the 'panel test' is a robust and statistically reliable method, there is a certain lack of reproducibility in the results obtained by different panels due to the inherent subjectivity of the tasters (Morales, Aparicio-Ruiz, & Aparicio, 2013). The best way to achieve the harmonization among different panels, as well as between tasters in the same panel, would be through the use of recognised references which should be obtained from the remaining samples of a proficiency interlaboratory testing or from particular RMs properly produced for sensory analysis (COI/T.20/Doc. No 14, 2018; Aparicio-Ruiz, Morales, & Aparicio, 2019).

In spite of there is extensive information on how food-based RMs should be evaluated regarding physico-chemical properties (Lee et al., 2019; Olivares, Souza, Nogueira, Toledo, & Marcki, 2018), little directives about the production of RMs for olive oil analysis are available. As far as the authors know only a previous study has been reported by Cuadros-Rodríguez, Bosque-Sendra, Mata-Espinosa, Gonzalez-Casado, & Rodríguez-García, (2008) about the certification of physicochemical parameters of OO. In respect of sensory analysis, a battery of ISO methods has been described mainly devoted to classify or describe matters regarding sensory attributes (ISO 8588, 2017; ISO 4120, 2004; ISO 10399, 2017; ISO 5495, 2005; ISO 16820, 2019; ISO 8587,

2006). Nevertheless, none of these methods are directly applicable to carry out a homogeneity study and the procedures must be suitably designed to this end. Only a reported study applying multivariate cluster analysis to assess physicochemical and sensory homogeneity of traditional French dry sausages deserves to be considered as a valid precedent (Rason, Laguët, Berge, Dufour, & Lebecque, 2007).

Homogeneity assessment in VOO samples can be carried out from the experimental data obtained by two independent analytical ways: sensory analysis or instrumental analysis. The first one focuses on the application of different statistical test for comparing the measured organoleptic intensities and the second one is based on the similarity analysis of unspecific instrumental signals or fingerprints related to the active odorant chemical compounds. In both cases, a minimum number of units are required in order to obtain results statistically reliable; this number is the maximum of the two values set out in the expression (ISO Guide 35, 2017):

$$N_h = \max (10, \sqrt[3]{N_{\text{prod}}})$$

where N_h is the number of unitary samples to be analysed for homogeneity testing, and N_{prod} is the total number of units available from the same batch.

The sensory analysis way requires the participation of a high-precise panel consisted of qualified tasters who have been previously trained and harmonised in the sensory detection and quantitation of the considered attributes. Precisely, the availability of panel meeting these requirements is the main drawback for implanting this way.

Two sensory strategies may be applied to evaluate the homogeneity of a sample set: ranking tests and discrimination tests. The ranking testing is designed to order a set of similar samples which differ only in the perceivable intensity of an organoleptic attribute. When identical samples are evaluated, a random ranking established by the different tasters is a clear indication of homogeneity. Nonetheless, if a number of tasters recognise a significant ranking tendency, the samples should not be considered as similar units and the batch should be stated as non-homogeneous. Some statistical test could be applied in order to detect ranking as the Page test, (Page, 1963) and the 'run' tests (Wald, & Wolfowitz, 1940). Page test is a non-parametric hypothesis test, which allows checking if there are differences between several levels of an independent variable. In this sense, the test checks if the testing samples are ordered

in a specific sequence. 'Run' test allows establishing the existence of a trend in the results of the organoleptic evaluation. In this test, a sequence of values lower or higher than the median is called 'run', with a length consisting on the number of consecutive observations according with a specific property.

Discrimination testing is focussed to check perceivable differences between two or more batch units regarding certain attribute of the material. Panellists or tasters receive at least two types of samples, one of which is belonging to the target or reference group (homogeneous units) and another that is associated to the possible source of non-homogeneity i.e. heterogeneity. Then, the results are statistically evaluated. The samples are accepted as homogeneous when the number of poorly classified samples is not statistically significant. There are different tests to carry out the sensory analysis discrimination such as the 'A-not A' test (ISO 8588, 2017), the 'triangular' test (ISO 4120, 2004) or the 'duo-trio' test (ISO 10399 2017) among many others. In the case of OO, the origin of heterogeneity may affect the intensity of one or more attributes being very difficult to provide the right source of representative heterogeneous samples (non-homogeneous class) in a reliable way. For instance, if the chemical compositions of the samples are altered by adding odorant substances to change the intensity of one or more attributes, the tasters are able to recognise that the difference is caused by an exogenous chemical compound and therefore these samples do not match the real samples. This is the main reason why the CRMs for OO sensory analysis should be produced from real samples taken from native product already holding the attribute or attributes of interest.

The problems arisen from the difficulty to find the non-homogeneous class to assess the homogeneity of the samples by means of the tests developed in the previous section can be avoided if instrumental analysis is used. For this purpose, the homogeneity would be evaluated through the analytical determination of the substance or substances (markers) responsible for the certified property (negative attributes). Nevertheless, an added difficulty with marker selection is that there are not a single marker for each OO attribute but, the intensity sensory perception is the result of a complex and non-linear combination of the type and concentration of odorants. That is why this strategy cannot be applied as a general protocol for homogeneity assessing, although it would be possible only when the OO batch were previously characterised and well-known.

There are different instrumentals analytical techniques that would be valid to homogeneity assessment such as near infrared (NIR) or nuclear magnetic resonance

spectrometries (NMR) (Rason et al., 2007; Tangpaisarnkul, Tuchinda, Wilairat, Siripinyanond, Shiowattana, & Nobsathian, 2018); gas or liquid chromatographies coupled with detection systems as mass spectrometry (GC-MS, LC-MS) (Elordui-Zapatarietxe et al., 2016; Olivares et al., 2018); diode array UV-Vis absorption (LC-DAD) (Shehata, Rizk, & Rend, 2016; Tangpaisarnkul et al., 2018). In all these instances, instrumental techniques have been used for the analytical determination of a specific compound whose concentration was used as critical property of the RM.

As far as the authors concern, an equivalent methodology addressed to RMs to be used in sensory analysis has not yet been described. Taking advantages from chromatographic techniques, it would be possible to assess the homogeneity of the type of RMs under considered, studying in an unspecific way the chemical profile of odorant compounds responsible of the flavour such as volatile organic compounds (VOC). Under this assumption, the use of instrumental fingerprints to ensure homogeneity in CRM for sensory analysis is proposed. They have been obtained from the fraction of VOCs presents in OO using gas chromatography coupled with a flame ionization detector (GC-FID) (Linsinger et al., 2001; García-González, Morales, & Aparicio, 2010; Taticchi, Esposto, & Servili, 2014; Valverde-Som, Ruiz-Samblás, Rodríguez-García, & Cuadros-Rodríguez, 2018b). The fingerprinting methodology does not need a good resolution of chromatographic peaks; therefore, the analysis time could be shortened. In general, fingerprinting could be used for: (i) developing pattern recognition methods for classification and regression, (ii) laboratory or industrial monitoring of chemical process, (iii) controlling degradation and ageing process of material systems, or (iv) applying similarity analysis between related specimens; this last goal has been employed in this study. Nevertheless, the fingerprints do not offer evident information, and therefore it is necessary to apply chemometrics for selecting and extracting the useful information. The run-to-run chromatograms are often affected of profile shifts and warping, so it is essential to carry out a processing of the raw acquired fingerprint in order to remove instrumental noise, aligns retention times and corrects intensity variations.

As a viable alternative, this paper proposes the use of a chromatography-based methodology to assess the homogeneity of three olive oil RMs produced each one at three intensity levels of a single sensory defective attribute (considered 9 independent RMs as a whole). The methodology is focused on the analysis of VOCs using gas chromatography coupled to flame ionization detection (GC-FID). Note that the identification and quantification of the chemical odorant compounds are not the

purpose of the present study and it is not carried out. A similar methodology has been already reported by some of the authors aimed at monitoring the stability of OO reference materials for sensory analysis (Valverde-Som, Ruiz-Samblás, Rodríguez-García, & Cuadros-Rodríguez, 2018a; Valverde-Som, 2018b).

2. FUNDAMENTALS OF THE HOMOGENEITY STATEMENT

As it is stated in the introduction section, to establish the homogeneity of a RM batch, a number N_h of representative units covering the whole batch, is systematically sampled and analysed. Similarities indices are then calculated from the volatile chromatographic fingerprints and used to obtain reliable metrics for homogeneity assessing. The chromatograms obtained are firstly transformed in data vectors representing each one the individual fingerprint of the selected N_h units for the study. Following, every fingerprint is pre-processed in order to: i) remove the first and last parts of each fingerprint that do not contain any useful information, ii) filter and smooth for noise removing, iii) correct the baseline drifting, iv) align time shifting and warping, and v) normalize the intensities in respect to a suitable internal standard (IS).

The strategy to be applied to each RM involves two steps: (1) several aliquots of the first unit are analysed in order to have a set of replicate fingerprints which are combined to obtain an average fingerprint used as reference; (2) each one of the fingerprints obtained from the remaining $(N_h - 1)$ units are compared with the reference fingerprint and consequently $(N_h - 1)$ similarity indices can be calculated in order to study the similarity between them. According to Valverde-Som (Valverde-Som et al., 2018a & 2018b) three indices have been considered for each comparison:

- The coefficient of determination, R^2 , signifies the square of the Pearson correlation coefficient (Brereton, 2009). The R^2 index takes values from 0 to 1.

$$R^2(x_r, x_c) = \frac{(\sum(x_{ci} - \bar{x}_c) \times \sum(x_{ri} - \bar{x}_r))^2}{\sum(x_{ci} - \bar{x}_c)^2 \times \sum(x_{ri} - \bar{x}_r)^2}$$

- The cosine of the angle between the considered and reference fingerprints, $\cos\theta$. The $\cos\theta$ will be 1 when the angle is 0° , i.e. the two fingerprints have the same orientation. It is obtained by applying the expression (Andríc & Hérberger, 2017):

$$\cos\theta = \frac{\sum (x_{ci} \times x_{ri})}{\sqrt{\sum x_{ci}^2 \times \sum x_{ri}^2}}$$

- The nearness index, NEAR, is a measurement of the normalized Euclidean distance between the elements of two vectors and it is used to describe the equivalence in the interval 0 to 1 (Pérez-Robles, Navas, Medina-Rodríguez, & Cuadros-Rodríguez, 2017). This index is calculated from the following equation:

$$\text{NEAR}(x_c, x_r) = 1 - \frac{\sqrt{\sum (x_{ci} - x_{ri})^2}}{\sqrt{\sum (x_{ci} + x_{ri})^2}}$$

Where, x_{ci} and x_{ri} symbolise each element of the considered and reference fingerprints, respectively.

In all the cases, a high level of similarity can be assumed when not differences between the fingerprints had been found. In addition, the closer to 1 are the values of the indices, the greater is the similarity between the evaluated and the reference fingerprints. These values can also be expressed in %. This is, 100% involves similar fingerprints.

The homogeneity statement of the materials was established through the application of acceptance and rejection criteria fixing a threshold value for each similarity index. In order to visualize easily whether the acceptance criteria are met, suitable control charts, plotting the obtained similarity indices, can be used (ISO 7870-1, 2019).

3. MATERIALS AND METHODS

3.1. Olive oil reference materials (OO-RMs)

The OO-RMs specimens were provided through 'SensOLEO-CRM 2017' project, which was funded by the Andalusian Regional Government and aimed to develop certified reference materials for sensory analysis of OO. Table SM1 (in Supplementary Material) gathers the main characteristics of the nine OO-RM batches produced. The whole RM set can be described as olive oil samples representative of three defective attributes

prepared at three levels of increasing intensity which are symbolised by the number 1, 2 and 3 (note that the OO-RMs at 2 and 1 levels were prepared blending different proportions of the OO-RM at level 3 with a undefective olive oil base which is specifically selected to be compatible).

Each OO candidate to RM was produced in a single batch and bottled in 350 units of 250 mL. In order to evenly cover the batch, ten representative units were systematically sampled from each OO-RM batch (9 RMs $\times$ 10 bottles sampled) obtaining a set sample of 90 bottles. The selected samples, which will be used to test the homogeneity of the RM batch, were cool-stored (0-8 °C) in darkness until being analysed. Figure 1 shows graphically a scheme of the whole process.

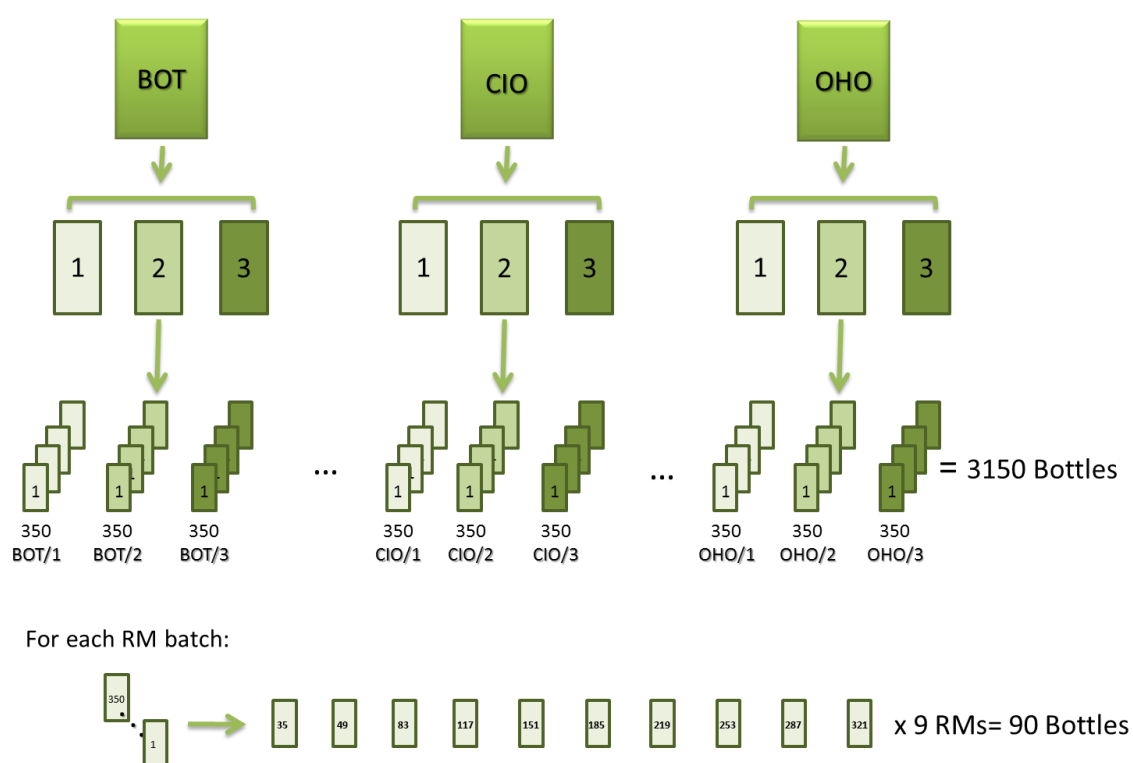


Figure 1. Applied process to select the sample set from each RM batch to test the homogeneity.

Notice that only between-bottles homogeneity is studied, since within-bottle homogeneity is always negligible in the case of intrinsically homogeneous materials such as solutions. In order to ensure the within-bottle homogeneity, each OO bath has been filtered prior to bottling and the bottles are always conveniently shaken before analysis; this is a well-known precaution that is always routinely applied in the field of vegetable oils.

3.2. Sample preparation

In a 20 mL headspace vial, 0.250 g of IS solution ($0.100 \text{ mg}\cdot\text{g}^{-1}$ solution of fluorobenzene (99%; Sigma-Aldrich, St Louis, MO, USA) prepared in refined olive oil) were accurately weighted; then 4.750 g of olive oil RM was carefully added until that the final mass was 5.000 g.

3.3. Chromatographic analysis

The chromatographic separation of the samples was carried out in a Varian 3800 GC (Varian Inc., Palo Alto, CA, USA) equipped with a split/splitless injector coupled to flame ionization detector. The instrument uses an auto-sampler system (CombiPal; CTC Analytics, Zwingen, Switzerland). The analytical column was a capillary column covered with polyethylene glycol (Agilent HP-INNOWAX; $30\text{m} \times 0.25\text{mm}$ internal diameter $\times 0.25 \text{ }\mu\text{m}$ film thickness; maximum temperature $260 \text{ }^{\circ}\text{C}$; Agilent Technologies, Santa Clara, CA, USA).

As previous step to the chromatographic analysis, the VOCs were extracted using a 50/30 μm divinylbenzene/carboxen/polydimethylsiloxane StableFlex 24 Ga fibre (Supelco Analytical, Bellefonte, PA, USA). The headspace solid-phase microextractions, as well as, GC-FID experimental conditions employed were those described by Valverde-Som et al., 2018b. Briefly, the samples were incubated in a shaker oven at $40 \text{ }^{\circ}\text{C}$ for 2 min and the extraction was carried out for 30 min. Fibre thermal desorption into GC was conducted at $250 \text{ }^{\circ}\text{C}$ for 2 min in splitless injection mode, and then the fibre was cleaned into the GC with a 1:100 split ratio for 6 min. The column oven temperature of $80 \text{ }^{\circ}\text{C}$ was held for 2.5 min, then programmed at $40 \text{ }^{\circ}\text{C}\cdot\text{min}^{-1}$ to $240 \text{ }^{\circ}\text{C}$, maintaining this temperature for 3.5 min. Hydrogen (99.999%; Air Liquid, Madrid, Spain) and air (99.999%; Air Liquid) were used as the flame (flow rate 30 and $300 \text{ mL}\cdot\text{min}^{-1}$, respectively), whereas helium (99.999%; Air Liquid) was used as the carrier gas (flow rate $1.5 \text{ mL}\cdot\text{min}^{-1}$).

As has been quoted previously, the first unit of each RM was chosen as reference and it was analysed in triplicate, the remaining samples were analysed once.

To assure the quality control of the instrument, $0.100 \text{ mg}\cdot\text{g}^{-1}$ IS solution was analysed at the beginning and end of each chromatographic run using the same operational conditions.

3.4. Data processing

MS Workstation (version 6.9) software data system (Varian Inc., Palo Alto, CA, USA) was used for chromatographic data acquisition and recording. In order to obtain the VOC chromatographic fingerprints, each chromatogram was exported to CSV format. Data pre-processing was performed using MATLAB, version 8.2 (Mathworks Inc., Natick, MA, USA) and 'Medina' function (version 14) (Pérez-Castaño et al., 2015) programmed in MATLAB environment. This function uses a number of algorithms of MATLAB Bioinformatics Toolbox, followed by 'icoshift' (interval correlation optimized shifting) algorithm for the alignment of chromatographic peaks (Tomasí, Savorani, & Engelsen, 2011).

The subsequent steps in order to assess the similarity of chromatographic fingerprints were described in '*Fundamentals of the homogeneity statement*' section. For each comparison the three similarity indices were calculated. Besides, the values of the index were plotted in control charts to verify when the threshold acceptance limits imposed were exceeded.

4. RESULTS AND DISCUSSION

4.1. Pre-processing of chromatographic data

The chromatograms obtained from the VOCs analysis were properly pre-processed using the 'Medina' function (Pérez-Castaño et al., 2015). Considering the OHO/3 reference material as guiding example, Figure 2a shows the three overlapped chromatograms from the three replicated analysed prior the pre-processing, meanwhile Figure 2b shows the three corresponding VOC GC-FID fingerprints after the pre-processing. Although, at first glance the chromatograms are similar, small differences could be noted in the zoomed region from 4 to 5 min running time (Figure 2a); after data pre-processing the differences are negligible (Figure 2b). Finally, an average of the three chromatographic fingerprints was calculated and used as reference fingerprint; it was constituted by (1×3480) variables (Figure 2c). The chromatograms from the other eight RMs were treated in the same fashion.

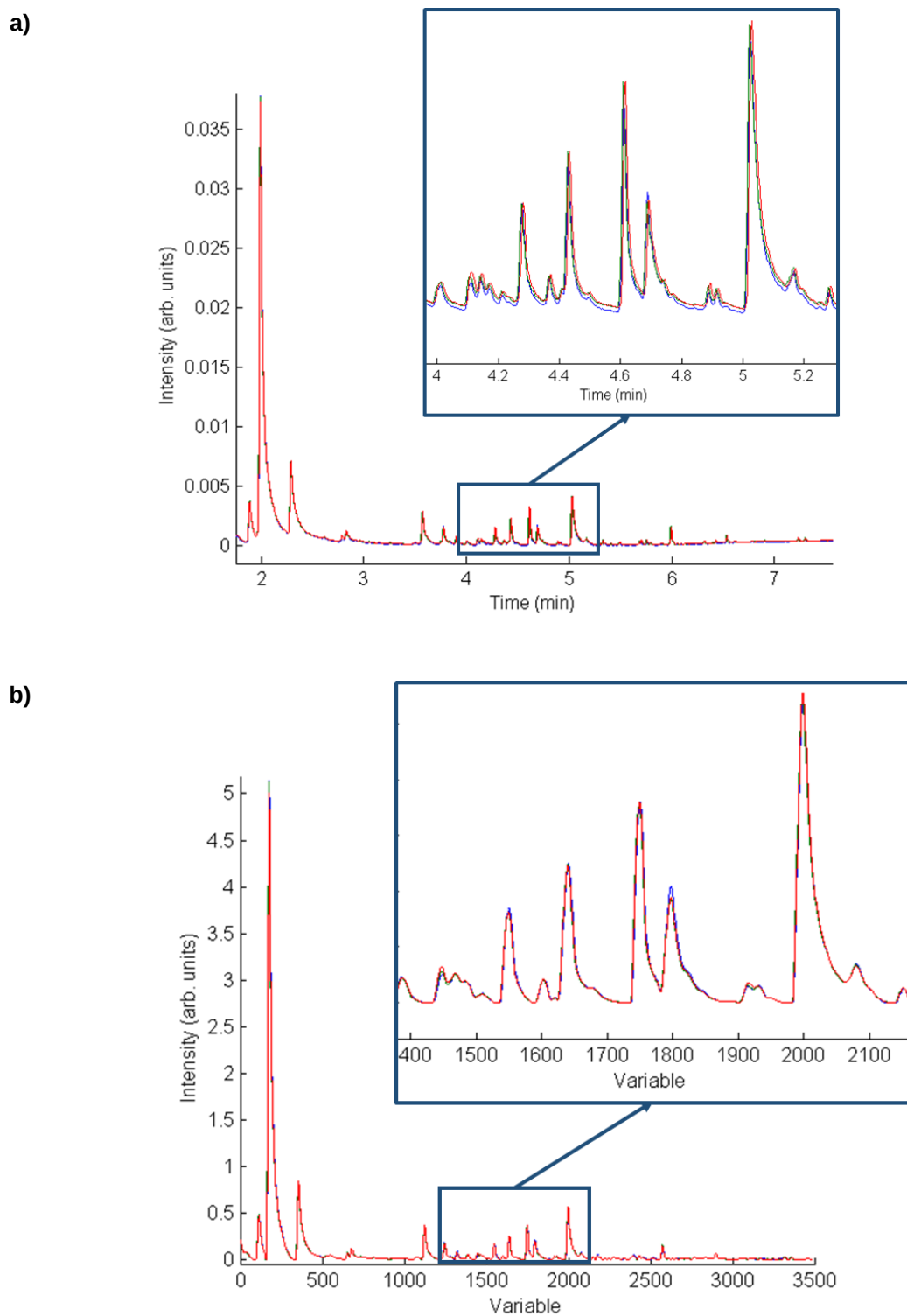


Figure 2. Instrumental outcomes obtained from OHO/3 material: a) original chromatogram of the three replicates before alignment; b) three reference fingerprints after alignment; c) average fingerprint used as reference.

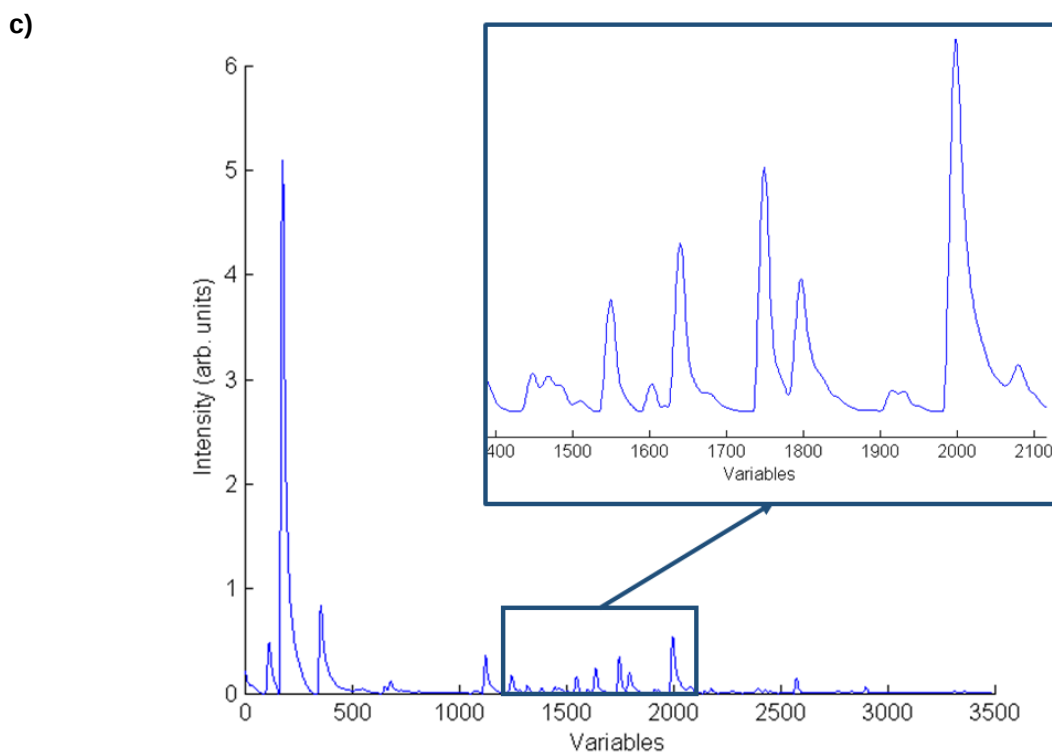


Figure 2. (cont.) Instrumental outcomes obtained from OHO/3 material: a) original chromatogram of the three replicates before alignment; b) three reference fingerprints after alignment; c) average fingerprint used as reference.

4.2. Similarity indices

In all cases the RM fingerprints are regarded as similar to the reference fingerprint when the values of R^2 , $\cos\theta$ and NEAR (expressed as percentages) are larger than 98%, 99% and 92% respectively. Values lower than these ones indicate non-similarity. Figure 3 shows the similarity control charts of the nine OO-RMs studied. As can be seen, all the similarity indices exceed the limits of acceptance, thus the selected units from each OO-RM have a high degree of similarity and therefore the nine OO-RM batches would be considered as homogeneous.

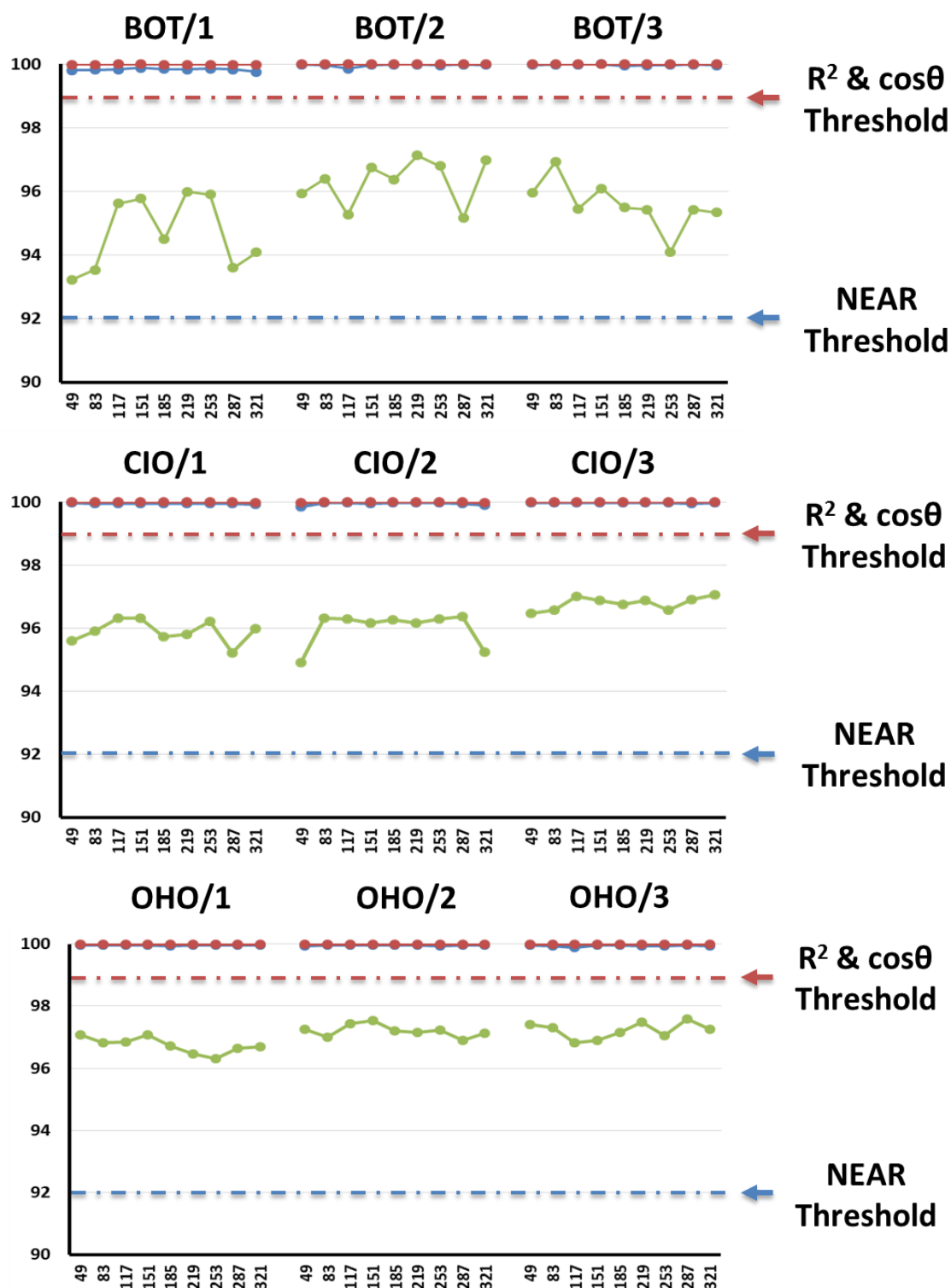


Figure 3. Similarity indices control charts (percentage of similarity vs numerical identifying code of samples) obtained for the nine RMs analysed by the proposed method.

It also can be observed that in all cases the $\cos\theta$ index is 100%, these results can be explained considering that small variations in the fingerprints do not have a significant influence in the calculation of the angle formed among them. Also, the results of R^2

index are very close to 100% and always larger than 99.90%, it can be assumed that there is a high correlation between each pair of variables in the compared fingerprints.

Regarding the NEAR index, it can be observed that its value is always upper than threshold value considered but not all the RMs have similar values. Therefore, all samples for OHO attribute have NEAR index values larger than 96.89%, all the CIO attribute samples have values larger than 94.91 % for NEAR index and finally the samples for BOT attributed have NEAR index value larger than 93.24%. Consequently, NEAR index presents a higher sensitivity to small variations in comparison with the other ones; that is why a lower threshold value is required. Due to this, a change in the value of one of the other two indices ($\cos\theta$ or R^2) does not necessarily implies that the other one must be affected in the same way; being mandatory to use them in conjunction with a third index.

4.3. Exploratory analysis

In addition to the similarity analysis of the selected units from each OO-RM, the homogeneity could be also studied analysing the natural grouping of the VOO GC-FID fingerprint from each OO-RM batch using unsupervised pattern recognition methods as hierarchical cluster analysis (HCA) and principal components analysis (PCA). To this end, a (108×3480) fingerprint matrix was built for this study (note that the 108 value is obtained considering the individual 90 RMs fingerprints and two additional replicates of the 9 first units selected as reference). In addition, a mean centring was applied on the fingerprint matrix previously to the chemometric approach.

4.3.1. Hierarchical Cluster analysis (HCA)

HCA were performed by selecting Manhattan distance and Ward method as distance metric and linkage criterion respectively. Figure 4 shows the dendrogram obtained; it can be observed that, regardless of the number of detected clusters, each one of the initial nesting group only includes all the samples of a particular OO-RM according the intensity of the representative attribute. This fact corroborates the homogeneity results previously obtained from the similarity analysis.

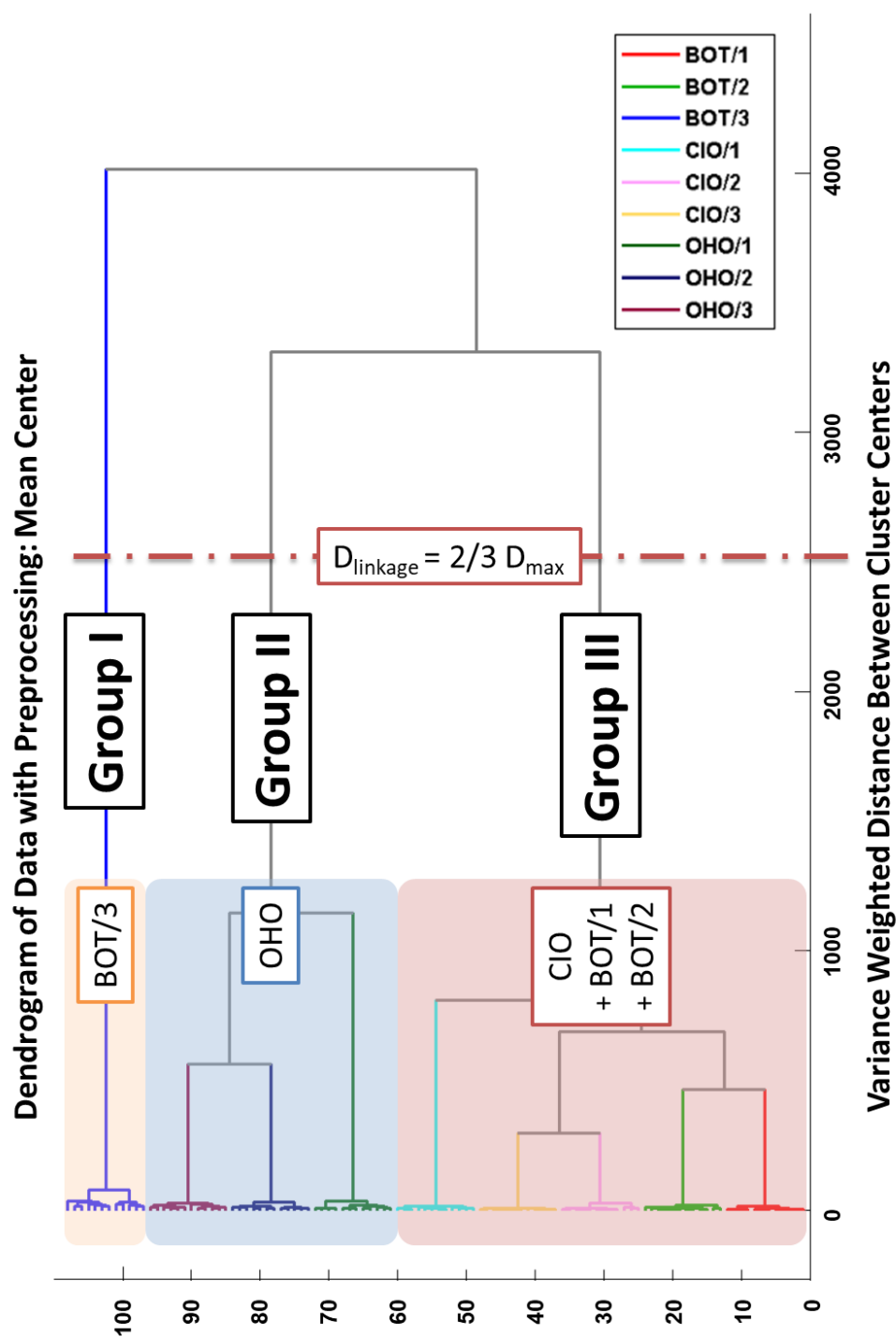


Figure 4. Dendrogram of olive oils reference materials analysed (OO-RMs) according its VOCs GC-FID fingerprints. HCA: Mahalanobis distance and Ward method as distance metric and linked criterion respectively.

Considering $D_{\text{linkage}} = 2/3 D_{\text{max}}$ as a criterion to define the natural clustering, three groups or clusters can be defined: Group I constituted by the BOT/3 units, this group is corresponding to the OO-RM having the most 'intense' or concentrated fusty/muddy

attribute. Group II constituted by OHO, *i.e.* all the musty olive oils internally ordered attending their intensities. Finally, Group III integrated by the CIO and both BOT/2 and BOT/1 OO-RMs, this is, olive oils having rancid defect, and medium and low intensity of the fusty/muddy attribute, respectively.

4.3.2. Principal Components analysis (PCA)

The selected three principal components explained 99.13% of the total variance. Figures 5 and 6 show the projection of scores and the loadings plots on the PC1-PC2 (86.03% and 10.27% of variance) and PC3-PC2 plane (2.83% and 10.27% of variance), respectively.

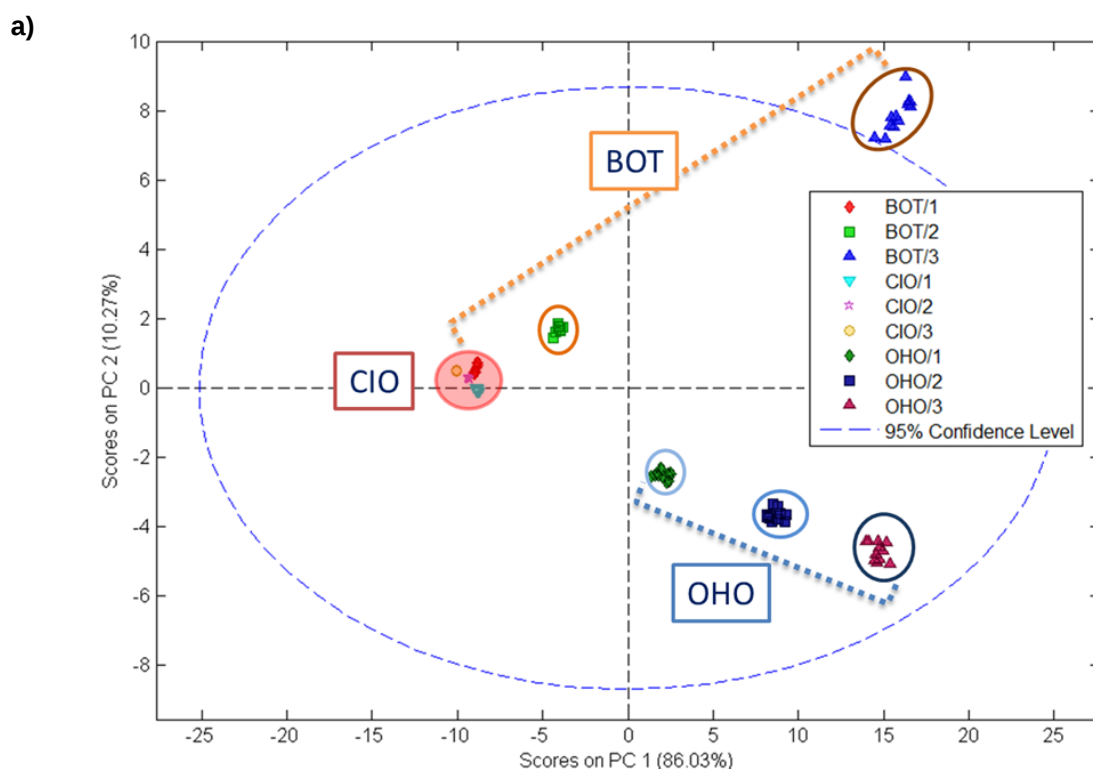


Figure 5. Scatter plot obtained from PCA of OO-RMs fingerprint matrix (108×3480). a) PC1 vs PC2, b) PC3 vs PC2.

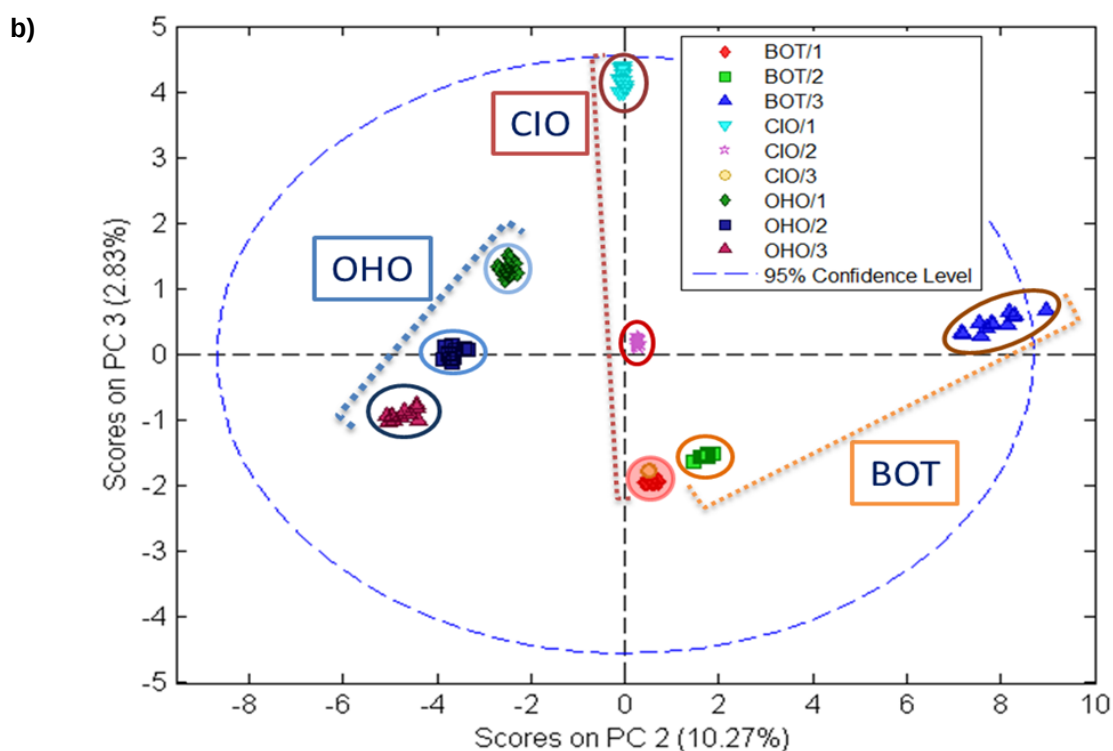


Figure 5. (cont.) Scatter plot obtained from PCA of OO-RMs fingerprint matrix (108×3480). a) PC1 vs PC2, b) PC3 vs PC2.

The scores plot on the PC1-PC2 plane (Figure 5a) clearly distinguished six sample groups. From these, five are coincident with five of the nine OO-RMs studied and could be described as following: i) BOT/3 with positive score values on both PC1 and PC2; ii) OHO/1 to OHO/3, all of them are characterized by positive scores on PC1 and negative on PC2 (note that the more intensity of the musty attribute the more negative the score value on PC2); iii) BOT/2 constituted by the samples having fusty attribute at medium intensity which are characterised by negative scores on PC1 and positives on PC2. Finally, the last group jointly constituted by the rancid (all intensities) and the fusty/muddy (low intensity) OO-RMs, these samples have negative scores on PC1 and positives, close to zero, on PC2.

From the scores plot on the PC3-PC2 plane (Figure 5b) it can be seen a better separation of OO-RMs with CIO code. On the other hand, the 8 natural groupings observed are coincident with 7 of the 9 OO-RMs studied. The last group is characterized by negative values of PC3 and close to 0 positive values of PC2, constituted by CIO/3 (rancid defect at maximum concentration) and BOT/1 (fusty/muddy defect at the lowest concentration). Near this group appears BOT/2

(medium intensity of the fusty/muddy defect) characterized by negative values of PC3 and positive values of PC2. The proximity between these groups is coincident with the results obtained in HCA.

Regarding the loading plot from Figure 6a, only variables 473 and 1234 affect negatively to the PC1 (they correspond to retention times of 2.48 and 3.76 min in the original chromatogram). The variables 172, 350, 1236, 1549, 1746 and 1999 (retention times: 1.98 min, 2.28 min, 3.76 min, 4.28 min, 4.61 min and 5.02 min, respectively) have a negative contribution on the PC2, and they would be responsible of the separation of the musty samples from the remaining ones (Figure 6b). In relation with PC3 (Figure 6c), the variable 172 has a negative contribution, on the contrary variables 1236, 1549, 1645, 1753, 1810 and 2005 (retention times: 1.98 min, 3.76 min, 4.28 min, 4.44 min, 4.62 min, 4.72 min and 5.04 min, respectively) have a positive contribution on this PC, would being responsible of the separation in the different intensity levels of all OO-RMs samples. In general terms, and independently of the positive/negative influence that would affect to the PCs, the considered variable could be used as possible chemical markers in future studies.

Finally, it could be observed that two of the RMs do not have major differences in the VOCs present in those samples with fusty/muddy defect at its lowest intensity. This suggests that the percentage of mixture of oils used to elaborate the OO-RMs would not be the most suitable. This fact would affect to the scores assigned by the tasters when these attributes at different intensities would be sensory analysed.

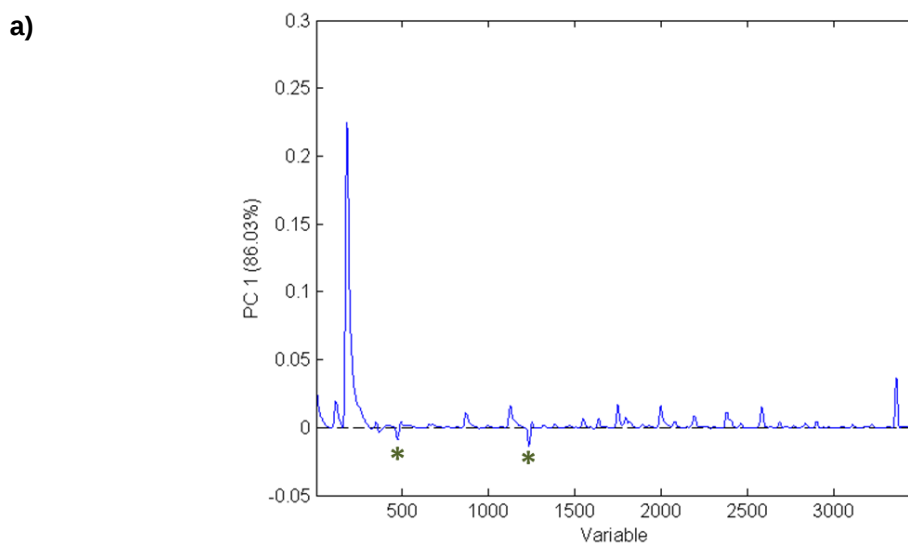


Figure 6. Loadings plots obtained from PCA of OO-RMs fingerprint matrix. a) Loadings on LV1, b) Loadings on LV2, c) Loadings on LV3. Potential chemical markers are denoted by asterisk (*).

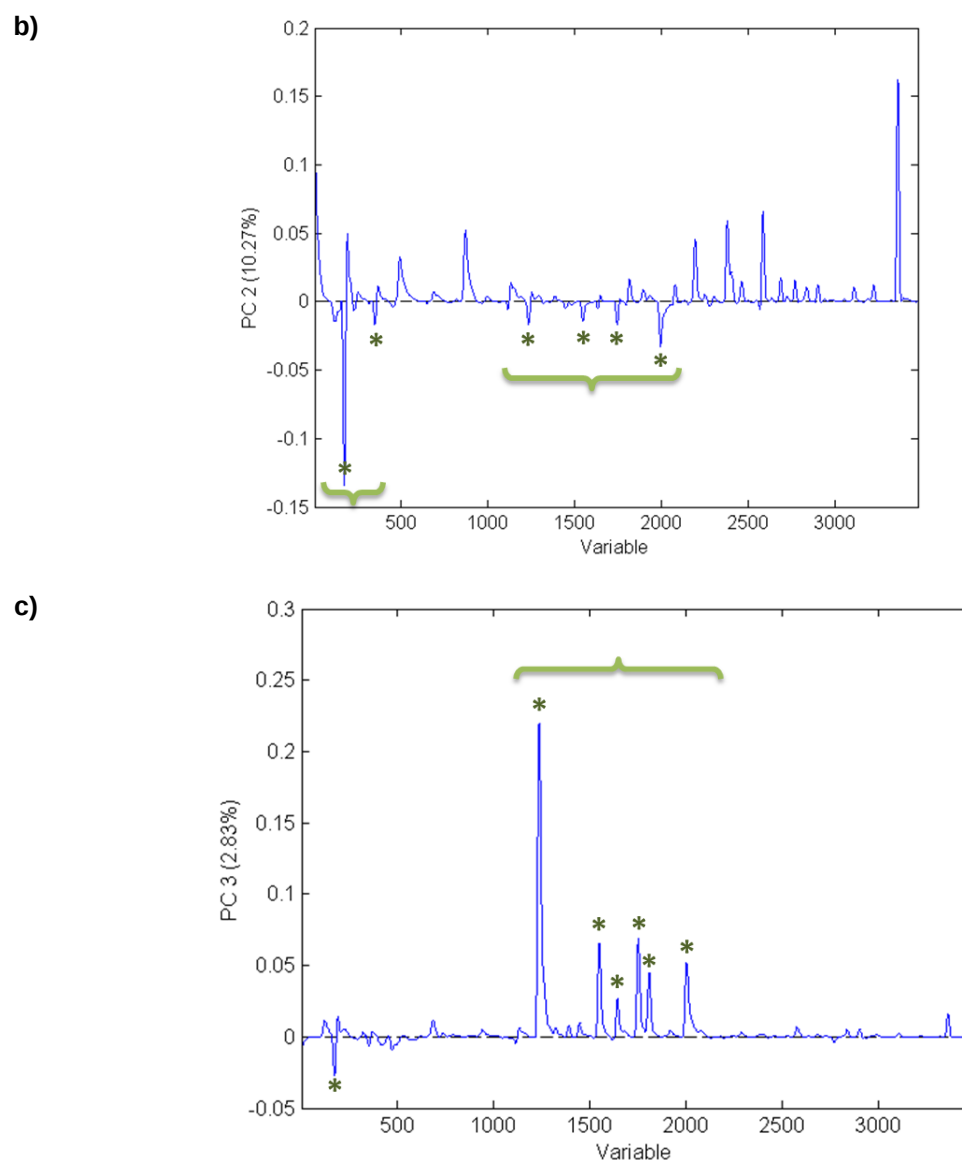


Figure 6. (cont.) Loadings plots obtained from PCA of OO-RMs fingerprint matrix. a) Loadings on LV1, b) Loadings on LV2, c) Loadings on LV3. Potential chemical markers are denoted by asterisk (*).

5. CONCLUSIONS

This paper establishes a valuable and rigorous methodology for the homogeneity assessment of olive oil samples using GC-FID fingerprints corresponding to the volatile organic compounds fraction (VOCs), previously isolated by solid phase microextraction. This chromatographic strategy for homogeneity testing overcomes the use of sensory analysis and enables the certification process of RMs of olive oils to be

used as a key tool for training and harmonising sensory panel aimed to verify the olive oil marketing categories assigned by the manufacturers and packagers.

The proposed similarity-based strategy allows revealing the non-homogeneity between sampled units of the candidate RM. The similarity indices control charts obtained permit to check, in an easy way, when threshold limit value acceptance criteria are overreached in a similar way as in the certification process.

Furthermore, using a chemometric approach by unsupervised multivariate pattern recognition methods (as HCA and PCA) the natural grouping of the units belonging to each one of the OO-RM could be highlighted and possible chemical markers could be identified.

6. ACKNOWLEDGMENTS

The authors would like to thank the Andalusian Regional Government (Consejería de Agricultura, Pesca y Desarrollo Sostenible, Spain) and the initiative of the 'Dirección General de Industrias, Innovación y Cadena Agroalimentaria' for financial assistance through the project SensOLEO-2017. The authors are also very grateful to Aceites MAEVA SL. Granada, Spain, for the production and packaging of the OO-RMs batches. In addition, one of the authors, FOG wants to express his sincere gratitude to the Andalusian Regional Government (Consejería de Agricultura, Pesca y Desarrollo Sostenible, Spain) for a technical assistant contract (project GOP3-JA-16-0012).

SUPPLEMENTARY MATERIAL

Table SM1. Description of the reference materials of olive oil used in this study (note that BOT, CIO and OHO are encoded names and they do not relate to the material composition).

Material code	Blend of olive oils	Description
BOT/3	100% concentrated	Olives of 'picual' variety were subjected to a fermentation process to generate the fusty/muddy defect in the olive oil (concentrated).
BOT/2	25% concentrated, 75% base	Blend of 25% BOT/3 and 75% of virgin olive oil and refined olive oil mix (base).
BOT/1	2% concentrated, 98% base	Blend of 2% BOT/3 and 98% of virgin olive oil and refined olive oil mix (base).
CIO/3	100% concentrated	Blend of olive oil subjected to an oxidative process to generate the rancid defect in the olive oil (concentrated).
CIO/2	75% concentrated, 25% base	Blend of 75% CIO/3 and 25% of virgin olive oil of 'arbequina' variety (olive oil was obtained from ripe olives harvested directly from the tree) (base).
CIO/1	25% concentrated, 75% base	Blend of 25% CIO/3 and 75% of virgin olive oil of 'arbequina' variety (olive oil was obtained from ripe olives harvested directly from the tree) (base).
OHO/3	100% concentrated	Olives of 'picual' variety were stored for a period of time in humid conditions to generate the musty defect in the olive oil (concentrated).
OHO/2	75% concentrated, 25% base	Blend of 75% OHO/3 and 25% of virgin olive oil of 'picual' variety (olive oil was obtained from ripe olives harvested directly from the tree) (base).
OHO/1	25% concentrated, 75% base	Blend of 25% OHO/3 and 75% of virgin olive oil of 'picual' variety (olive oil was obtained from ripe olives harvested directly from the tree) (base).

REFERENCES

- Andríc, F., & Héberger, K. (2017), How to compare separation selectivity of high-performance liquid chromatographic columns properly? *Journal of Chromatography A*, 1488, 45–56.
<https://doi.org/10.1016/j.chroma.2017.01.066>
- Aparicio, R., Lanfranco, S.C., & Fiebig, H.J. (2013), Olive Oil Authentication, in R. Aparicio, & J. Harwood, (eds), *Handbook of Olive Oil: Analysis and Properties* (2nd ed.) (pp. 590–653), New York: Springer.
- Aparicio-Ruiz, R., Morales, M.T., & Aparicio, R. (2019), Does authenticity of virgin olive oil sensory quality require input from chemistry? *European Journal of Lipid Science and Technology*, 1900202.
<https://doi.org/10.1002/ejlt.201900202>
- Brereton, R.G. (2009), Exploratory data analysis, in R.G. Brereton (ed), *Chemometrics for Pattern Recognition* (pp. 47–105). Chichester: John Wiley & Sons.
- Commission Delegated Regulation (EU) 2016/2095 of 26th September 2016: amending Regulation (EEC) No. 2568/91 on the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis (consolidated version 01991R2568), Official Journal of the European Union L326, 1–128.
- Cuadros-Rodríguez, L., Bosque-Sendra, J.M., Mata-Espinosa, A.P., Gonzalez-Casado, A., & Rodríguez-García, F.P. (2008), Elaboration of four olive oil certified reference materials: InterOleo-CRM 2006 Certification Study. *Food Analytical Methods*, 1, 259–269.
<https://doi.org/10.1007/s12161-008-9037-0>
- Elordui-Zapatarietxe, S., Fettig, I., Richter, J., Vanermen, R.P.G., Monteyne, E., Boom, G. & Emteborg, H. (2016), Preparation and evaluation of sufficiently homogeneous and stable reference materials for priority hazardous substances in whole water. *Accreditation and Quality Assurance*, 21, 113–120.
<https://doi.org/10.1007/s00769-015-1189-1>
- García-González, D.L., Morales, M.T., & Aparicio, R. (2010), Olive and olive oil, in Y.H. Hui (ed), *Handbook of Fruit and Vegetable Flavors* (pp. 821–847). Chichester: John Wiley & Sons.
- COI/T.20 / Doc. No 14 / Rev.5 (2018). Sensory Analysis of Olive Oil. Guide for the

selection, training and quality control of virgin olive oil taster-qualifications of taster, panel leaders and trainers. Madrid: International Olive Council (IOC).

COI/T.20 / Doc. No 15 / Rev. 10 (2018). Sensory Analysis of Olive Oil. Method for the organoleptic assessment of virgin olive oil. Madrid: International Olive Council (IOC).

ISO 10399: Sensory analysis – Methodology – Duo-trio test, Geneva, Switzerland 2017.

ISO 16820: Sensory analysis – Methodology – Sequential analysis, Geneva, Switzerland, 2019.

ISO 4120: Sensory analysis – Methodology – Triangle test, Geneva, Switzerland, 2004.

ISO 5492: Sensory analysis – Vocabulary, Geneva Switzerland, 2008.

ISO 5495: Sensory analysis – Methodology – Paired comparison test, Geneva, Switzerland, 2005.

ISO 7870-1: Control charts — Part 1: General guidelines, Geneva, Switzerland, 2019.

ISO 8587: Sensory analysis – Methodology – Ranking, Geneva, Switzerland, 2006.

ISO 8588: Sensory analysis – Methodology – "A" - "not A" test, Geneva, Switzerland, 2017.

ISO Guide 30: Reference Materials – Selected Terms and Definitions, Geneva, Switzerland, 2015.

ISO Guide 35: Reference Materials – Guidance for characterization and assessment of homogeneity and stability, Geneva, Switzerland, 2017.

Lee, J., Kim, B., Lee, S.Y., Choi, J., Kang, D., Lee, H. Sim, H-J., Baek, S.Y., Lee, H., Hyung, S.W., Ahn, S., Seo, D., Hwang, J., Park, J.S., Kwak, B.M., & Won, J. (2019), Development of an infant formula certified reference material for the analysis of organic nutrients. *Food Chemistry*, 298, 1–8.

<https://doi.org/10.1016/j.foodchem.2019.125088>

Linsinger, T.P.J., Pauwels, J., Van der Veen, A.M.H., Schimmel, H., & Lamberty, A. (2001), Homogeneity and stability of reference materials. *Accreditation and Quality Assurance*, 6, 20–25.

<https://doi.org/10.1007/s007690000261>

Mariotti, M. (2014), Virgin olive oil: definition and standards, in C. Peri (ed), *The Extra-Virgin Olive Oil Handbook* (pp. 11-20). Chichester: Wiley.

- Morales, M.T., Aparicio-Ruiz, R. & Aparicio, R. (2013), Chromatographic methodologies: compounds for olive oil odor issues, in R. Aparicio & J. Harwood (eds), *Handbook of Olive Oil: Analysis and Properties*, 2nd ed. (pp. 261–309). New York: Springer.
- Olivares, I.R.B., Souza, G.B., Nogueira, A.R.A., Toledo, G.T.K. & Marcki, D.C. (2018), Trends in developments of certified reference materials for chemical analysis - Focus on food, water, soil, and sediment matrices, *Trends in Analytical Chemistry*, 100, 53–64.
<https://doi.org/10.1016/j.trac.2017.12.013>
- Page, E. (1963), Ordered hypotheses for multiple treatments: a significance test for linear ranks. *Journal of the American Statistical Association*, 58, 261-230. DOI version not available.
- Pauwels, J., Lamberty, A., & Schimmel, H. (1998), Homogeneity testing of reference materials. *Accreditation and Quality Assurance*, 3, 51–55.
<https://doi.org/10.1007/s007690050186>
- Pérez-Castaño, E., Ruiz-Samblás, C., Medina-Rodríguez, S., Quirós-Rodríguez, V., Jiménez-Carvelo, A.M., Valverde-Som, L., González-Casado, A., & Cuadros-Rodríguez, L. (2015), Comparison of different analytical classification scenarios: application for the geographical origin of edible palm oil by sterolic (NP) HPLC fingerprinting. *Analytical Methods*, 7, 4192–4201.
<https://doi.org/10.1039/C5AY00168D>
- Pérez-Robles, R., Navas, N., Medina-Rodríguez, S., & Cuadros-Rodríguez, L. (2017), Method for the comparison of complex matrix assisted laser desorption ionization-time of flight mass spectra. Stability of therapeutical monoclonal antibodies. *Chemometrics and Intelligent Laboratory Systems*, 170, 58–67.
<https://doi.org/10.1016/j.chemolab.2017.09.008>
- Rason, J., Laguet, A., Berge, P., Dufour, E., & Lebecque, A. (2007), Investigation of the physicochemical and sensory homogeneity of traditional French dry sausages. *Meat Science*, 7, 359–370.
<https://doi.org/10.1016/j.meatsci.2006.06.004>
- Rutkowska, M., Kochańska, K., Kandel, L., Bajger-Nowak, G., Słomińska, M., Marć, M. Chojnacka, K., Polkowska-Motrenko, H., Zabiegała, B., Namieśnik, J. & Konieczka, P. (2018), Homogeneity study of candidate reference material

(contaminated soil) based on determination of selected metals, PCBs and PAHs. *Measurement*, 128, 1–12.

<https://doi.org/10.1016/j.measurement.2018.06.021>

Shehata, A.B., Rizk, M.S., & Rend, E.A. (2016), Certification of caffeine reference material purity by ultraviolet/visible spectrophotometry and high-performance liquid chromatography with diode-array detection as two independent analytical methods. *Journal of Food and Drug Analysis*, 24, 703–715.

<https://doi.org/10.1016/j.jfda.2016.06.009>

Tangpaisarnkul, N., Tuchinda, P., Wilairat, P., Siripinyanond, A., Shiowattana, J., & Nobsathian, S. (2018), Development of pure certified reference material of stevioside. *Food Chemistry*, 255, 75–80.

<https://doi.org/10.1016/j.foodchem.2018.02.029>

Taticchi, A., Esposto, S., & Servili, M. (2014), The basis of sensory properties of virgin olive oil, in E. Monteleone & S. Langstaff (eds), *Olive Oil Sensory Science* (pp. 33–54). Chichester: John Wiley & Sons.

Tomasi, G., Savorani, F., & Engelsen, S.B. (2011), Icoshift: an effective tool for the alignment of chromatographic data. *Journal of Chromatography A*, 1218, 7832–7840.

<https://doi.org/10.1016/j.chroma.2011.08.086>

Valverde-Som, L., Ruiz-Samblás, C., Rodríguez-García, F.P., & Cuadros-Rodríguez, L. (2018, a), Multivariate approaches for stability control of the olive oil reference materials for sensory analysis – Part I: Framework and fundamentals. *Journal of the Science of Food and Agriculture*, 98, 4237–4244.

<https://doi.org/10.1002/jsfa.8948>

Valverde-Som, L., Ruiz-Samblás, C., Rodríguez-García, F.P., & Cuadros-Rodríguez, L. (2018, b), Multivariate approaches for stability control of the olive oil reference materials for sensory analysis – Part II: Applications. *Journal of the Science of Food and Agriculture*, 98, 4245–4252.

<https://doi.org/10.1002/jsfa.8946>

Van der Veen, A.M.H., Linsinger, T.P.J., & Pauwels, J. (2001), Uncertainty calculations in the certification of reference materials. 2. Homogeneity study. *Accreditation and Quality Assurance*, 6, 26–30.

<https://doi.org/10.1007/PL00010459>

Wald, A., & Wolfowitz, J. (1940), On a Test Whether Two Samples are from the Same Population. *The Annals of Mathematical Statistics*, 11, 147-162.

<https://doi.org/10.1214/aoms/1177731909>

Capítulo III

**Desarrollo y aplicación del agnosticismo de
señales instrumentales**

Standardization of chromatographic signals – Part I: Towards obtaining instrument-agnostic fingerprints in gas chromatography

Journal of Chromatography A 1641 (2021) 461983



Contents lists available at ScienceDirect

Journal of Chromatography A

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



Standardization of chromatographic signals – Part I: Towards obtaining instrument-agnostic fingerprints in gas chromatography

Luis Cuadros-Rodríguez^a, Fidel Ortega-Gavilán^{a,*}, Sandra Martín-Torres^a,
Santiago Medina-Rodríguez^b, Ana M. Jimenez-Carvelo^a, Antonio González-Casado^a,
M. Gracia Bagur-González^a

^a Department of Analytical Chemistry, Faculty of Science, University of Granada, C/ Fuentenueva s/n, E-18071, Granada, Spain

^b By Techdesign S.L., C/ Tomas Edison 5, E-28500, Arganda del Rey, Madrid, Spain

ARTICLE INFO

Article history:

Received 13 December 2020

Revised 29 January 2021

Accepted 7 February 2021

Available online 9 February 2021

Keywords:

Instrument-agnostic chromatograms
Standardized chromatographic fingerprint
Empirical retention indices
Intensity scaling
Signal alignment

ABSTRACT

One of the main causes for the sparse use of multivariate analytical methods in routine laboratory work is the dependency on the measuring instrument from which the analytical signal is acquired. This issue is especially critical in chromatographic equipment and results in limitations of their applicability. The solution to this problem is to obtain a standardized instrument-independent signal –or instrument-agnostic signal– regardless of the measuring instrument or of the state of the same instrument from which it has been acquired. The combined use of both internal and external standard series, allows us to have external and transferable references for the normalization of both the intensity and the position of each element of the data vector being arranged from the raw signal. From this information, a simple mathematical data treatment process is applied and instrument-agnostic signals can be secured. This paper describes and applies the proposed methodology to be followed for obtaining standardized instrumental fingerprints from two significant fractions of virgin olive oil (volatile organic compounds and triacylglycerols), obtained by gas chromatography coupled to mass spectrometry (GC-MS) and analysed with two temperature conditions (conventional and high-temperature, respectively). The results of both case studies show how the instrument-agnostic fingerprints obtained are coincidental, regardless of the state of the chromatographic system or the time of acquisition.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Since the early 2000's, fingerprinting methodologies have been growing in popularity. An instrumental fingerprint refers to any characteristic and pre-processed instrumental analytical signal obtained from an analysed material reflecting its complex chemical composition in a non-obvious way. This is related to certain physico-chemical properties and thus may be applied for the authentication of the material under consideration [1]. Note that an instrumental fingerprint is different to a raw instrumental signal as a pre-processing step could have been applied in order to clean the signal and remove noise and other disturbances not directly associated with features of the sample measured. Under these considerations, an instrumental fingerprint recalls a comprehensive description of a test material that has been carried out in a non-selective (non-targeted) way.

Instrumental signals or instrumental fingerprints have a mathematical structure that can be associated with an intensity data tensor **INT**¹ (a mathematical object that can hold multivariate data) where each *i*-th element of the tensor may be symbolised by $INT_i = f(x_{1i}, x_{2i}, x_{3i}, \dots)$ in which INT_i is the signal intensity of the *i*-th element from the tensor and $x_{1i}, x_{2i}, x_{3i}, \dots$ are the coordinates or scores which describe the position of the same element within the tensor. Note that the number of scores defines the 'order' of the tensor and the number of 'ways' that outline the dimensionality of the data [2,3]. Analytical instruments capable of generating multiple measurements at one time or for one sample are considered first-order instruments. Hence a single index suffices to organize the data, and the measurements can be put into an ordered array referred to as a data vector $\langle INT_1, INT_2, INT_3, INT_4, \dots \rangle$ or first-order tensor, where each element represents a signal intensity that depends on a single score, $INT_i = f(x_{1i})$ (note that the number of elements depends on the acquisition time of

* Corresponding author.

E-mail address: fog@ugr.es (F. Ortega-Gavilán).

¹ Normally represented in bold.

III.a. RESUMEN (EN ESPAÑOL)

Una de las principales causas del uso limitado de métodos analíticos multivariable en el trabajo rutinario de un laboratorio es la dependencia de las señales medidas con el equipo empleado para su adquisición. Este problema es especialmente crítico en equipos basados en una separación cromatográfica, lo que provoca limitaciones en su aplicabilidad.

La solución a este problema pasaría por poder obtener una señal estandarizada e independiente –o señal agnóstica del instrumento– que no tenga en cuenta el propio instrumento o el estado en que dicho instrumento se encuentra durante el uso del mismo. El uso conjunto de patrones externos e internos, permite la obtención de información transferible al resto de señales, que se aplica para normalizar la intensidad y la posición de cada elemento del vector de datos obtenido a partir de una señal bruta. Tomando como base esta información, puede establecerse un sencillo proceso de tratamiento matemático de los datos con el fin de obtener señales agnósticas de los instrumentos.

En este artículo se propone y aplica la metodología que debe seguirse para la obtención de huellas instrumentales estandarizadas de dos fracciones significativas del aceite de oliva virgen (compuestos orgánicos volátiles y triacilglicerolos), obtenidas por cromatografía de gases acoplada a espectrometría de masas (GC-MS) y analizadas con dos condiciones de temperatura (rango de temperatura convencional y alta temperatura, respectivamente). Los resultados de ambos casos de estudio demuestran cómo las huellas instrumentales agnostizadas obtenidas son coincidentes, independientemente del estado del sistema cromatográfico o del momento de la adquisición.

III.b. CONCLUSIONES DEL TRABAJO (EN ESPAÑOL)

La metodología innovadora de agnosticismo instrumental propuesta en este trabajo establece una mejora significativa en la obtención de señales cromatográficas armonizadas y comparables con el fin de conseguir nuevas huellas instrumentales o señales analíticas estandarizadas, sin importar el instrumento utilizado o la naturaleza de los datos brutos adquiridos. De acuerdo con los resultados obtenidos en los dos casos de estudio, las huellas Instrumentales son prácticamente coincidentes independientemente del momento de adquisición y del estado cromatográfico del

instrumento en el periodo de análisis. El uso de un conjunto de patrones adecuados, tanto internos como externos, permite transferir la influencia del instrumento, principalmente relacionada con disminuciones de intensidades de señal, o desplazamientos como consecuencia de la degradación de la fase estacionaria, al resto de cromatogramas, independientemente de si estos se han obtenido en la misma tanda cromatográfica o en diversas tandas analizadas en días diferentes. De este modo, es posible obtener señales reproducibles en las que un posible cambio en el perfil de la señal cromatográfica no suponga un problema a la hora de alinearlas.

Este nuevo enfoque permitiría, por ejemplo, que las huellas cromatográficas de una misma muestra alimentaria procedentes de dos instrumentos diferentes tuvieran un alto grado de similitud. Además, trabajar con huellas cromatográficas estandarizadas abriría la posibilidad de crear un modelo multivariante único, tanto para su uso en aplicaciones cualitativas (clasificación) como cuantitativas (cuantificación), en el que poder incorporar los resultados obtenidos por diferentes instrumentos en distintos laboratorios.

Por último, y no menos importante, utilizando la metodología planteada, sería posible generar bases de datos de huellas cromatográficas obtenidas a partir de diferentes alimentos, que podrían ser compartidas por la comunidad científica, facilitando la resolución de muchos problemas. De este modo, se podría disponer de señales de referencia (huellas cromatográficas) de todo tipo de alimentos, que podrían ser utilizadas en cualquier parte del mundo y facilitar así la trazabilidad de los alimentos. Estas bases de datos podrían ser la solución ideal al tema de interés número 5 recogido en el documento "Horizonte 2020 -Temas prioritarios de investigación de la EFSA" dedicado al *Desarrollo de un sistema europeo integrado de recogida de datos, gestión de datos, análisis de datos e información sobre seguridad alimentaria basado en la nube*, en donde se propone el desarrollo de una base de datos para armonizar, recoger, analizar y compartir información comparable de análisis de alimentos. Es evidente que el agnosticismo instrumental es una metodología ideal para aplicar en el caso de señales analíticas que requieran una separación cromatográfica previa, ya que éstas podrían depender del tiempo y, por tanto, de todas las variables no controladas que puedan afectarles en el transcurso de la obtención de la señal. Además, la metodología propuesta abriría el camino para estandarizar también señales cromatográficas bidimensionales 2D o señales GC-IMS. En este sentido, los autores están actualmente trabajando en el desarrollo de una metodología dirigida a señales de orden 2.

III.c. ESTUDIO BIBLIOMÉTRICO

El artículo científico que se corresponde con el Capítulo III fue publicado en la revista '*Journal of Chromatography A*' la cual, según la información obtenida de la web JCR, se encuentra indexada en dos categorías diferentes:

- 1) *Biochemical research methods* – SCIE. Categoría que se encuentra incluida dentro del ámbito de la biología y la bioquímica. En el año 2020 esta revista se encuentra indexada en el primer cuartil (Q1) del total de revistas que conforman la categoría, ocupando el puesto 13 de un total de 78 revistas.
- 2) *Chemistry, analytical* – SCIE. Categoría incluida en el ámbito de la química. En el año 2020, esta revista está indexada dentro del primer cuartil (Q1) de las revistas que conforman dicha categoría, ocupando en esta ocasión el puesto 17 de un total de 87 revistas.

Durante 2020 la revista *Journal of Chromatography A* acumula un total de 67.823 citas, cifra en la que quedan incluidas las autocitas de autores en la misma revista, información que ha sido obtenida de la web JCR. Esta información se traduce en un factor de impacto de 4,759 mientras que dicho valor se reduce a 3,950 si se excluyen las autocitas.

En cuanto a la tendencia en el factor de impacto, en la Figura 18 se observa como desde su creación en el año 1997 la revista *Journal of Chromatography A* tuvo un aumento progresivo hasta el año 2012. En los siguientes 4 años la revista experimentó una caída del factor de impacto hasta valores cercanos a 3,750, mientras que, en los últimos años, ha experimentado un auge en su impacto dentro de la comunidad científica, superando en 2020 los valores máximos previamente alcanzados en el año 2012.

Desde su publicación y hasta el cierre de la edición de esta Memoria, el artículo que se presenta en este Capítulo ha sido citado en 5 artículos científicos publicados en diversas revistas indexadas en el JCR. En este caso, dichos trabajos también se encuentran publicadas en revistas del primer cuartil (Q1) de las categorías a las que pertenecen.

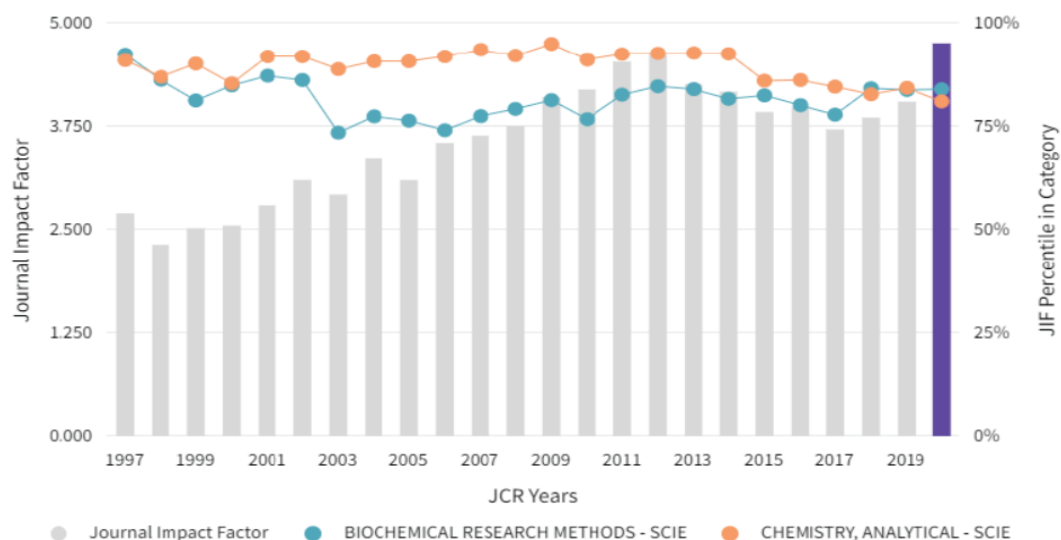


Figura 18. Gráfico de tendencia de la evaluación del factor de impacto de la revista *Journal of Chromatography A* desde su creación en el año 1997 tomada de la información suministrada por el *Journal Citation Reports* (fecha de acceso 24 de enero de 2022).

<https://jcr.clarivate.com/jcr-jp/journal-profile?journal=J%20CHROMATOGR%20A&year=2020&fromPage=%2Fjcr%2Fhome>

III.d. ARTICULO CIENTÍFICO

Standardization of chromatographic signals – Part I: Towards obtaining instrument-agnostic fingerprints in gas chromatography

Luis CUADROS-RODRÍGUEZ⁽¹⁾, Fidel ORTEGA-GAVILÁN^(1,✉), Sandra MARTÍN-TORRES⁽¹⁾, Santiago MEDINA-RODRÍGUEZ⁽²⁾, Ana M. JIMENEZ-CARVELO⁽¹⁾, Antonio GONZÁLEZ-CASADO⁽¹⁾, M. Gracia BAGUR-GONZÁLEZ⁽¹⁾

⁽¹⁾ *Department of Analytical Chemistry, Faculty of Science, University of Granada, C/ Fuentenueva s/n, E-18071, Granada, Spain*

⁽²⁾ *By Techdesign S.L., C/ Tomas Edison 5, E-28500, Arganda del Rey, Madrid, Spain*

HIGHLIGHTS

- Signal standardization is mandatory to extract reliable information from fingerprints
- Instrument agnostizing is proposed as a methodology for chromatogram standardization
- Retention indices similar to Kovats' proposal are used as external references for alignment
- VOCs and TGs instrument-agnostic fingerprints are obtained showing valuable results

KEYWORDS

Instrument-agnostic chromatograms; standardized chromatographic fingerprint; empirical retention indices; intensity scaling; signal alignment

✉ Author for correspondence e-mail: Fidel Ortega Gavilán, fog@ugr.es; phone: +34 958240797.

Abstract

One of the main causes for the sparse use of multivariate analytical methods in routine laboratory work is the dependency on the measuring instrument from which the analytical signal is acquired. This issue is especially critical in chromatographic equipment and results in limitations of their applicability. The solution to this problem is to obtain a standardized instrument-independent signal –or instrument-agnostic signal– regardless of the measuring instrument or of the state of the same instrument from which it has been acquired. The combined use of both internal and external standard series, allows us to have external and transferable references for the normalization of both the intensity and the position of each element of the data vector being arranged from the raw signal. From this information, a simple mathematical data treatment process is applied and instrument agnostic signals can be secured. This paper describes and applies the proposed methodology to be followed for obtaining standardized instrumental fingerprints from two significant fractions of virgin olive oil (volatile organic compounds and triacylglycerols), obtained by gas chromatography coupled to mass spectrometry (GC-MS) and analysed with two temperature conditions (conventional and high-temperature, respectively). The results of both case studies show how the instrument agnostic fingerprints obtained are coincidental, regardless of the state of the chromatographic system or the time of acquisition.

1. INTRODUCTION

Since the early 2000's, fingerprinting methodologies have been growing in popularity. An instrumental fingerprint refers to any characteristic and pre-processed instrumental analytical signal obtained from an analysed material reflecting its complex chemical composition in a non-obvious way. This is related to certain physico-chemical properties and thus may be applied for the authentication of the material under consideration [1]. Note that an instrumental fingerprint is different to a raw instrumental signal as a pre-processing step could have been applied in order to clean the signal and remove noise and other disturbances not directly associated with features of the sample measured. Under these considerations, an instrumental fingerprint recalls a comprehensive description of a test material that has been carried out in a non-selective (non-targeted) way.

Instrumental signals or instrumental fingerprints have a mathematical structure that can be associated with an intensity data tensor **INT**¹ (a mathematical object that can hold multivariate data) where each *i*-th element of the tensor may be symbolised by $INT_i = f(x_{1i}, x_{2i}, x_{3i}, \dots)$ in which INT_i is the signal intensity of the *i*-th element from the tensor and $x_{1i}, x_{2i}, x_{3i}, \dots$ are the co-ordinates or scores which describe the position of the same element within the tensor. Note that the number of scores defines the 'order' of the tensor and the number of 'ways' that outline the dimensionality of the data [2,3]. Analytical instruments capable of generating multiple measurements at one time or for one sample are considered first-order instruments. Hence a single index suffices to organize the data, and the measurements can be put into an ordered array referred to as a data vector $\langle INT_1, INT_2, INT_3, INT_4, \dots \rangle$ or first-order tensor, where each element represents a signal intensity that depends on a single score, $INT_i = f(x_{1i})$ (note that the number of elements depends on the acquisition time of the measuring device, and may consist of several thousands). Some more complex instruments can generate a data matrix (second-order tensor) per sample that has two position scores for each element, $INT_i = f(x_{1i}, x_{2i})$. Only first-order and second-order tensors may be graphed by plotting the intensity on the vertical axis (Y-axis) and the x-scores values on the horizontal axis (X-axis) or plane (X_1 - X_2 -plane), resulting in 2D or 3D graphics respectively.

In order for the fingerprinting methodology to be successfully applied it is necessary that the instrumental signals acquired from the same measuring equipment or analytical method to be equal or very similar. This is relatively simple when the

¹ Normally represented in bold

fingerprint is acquired in a static mode directly from the material system without being affected by the time at which it is measured, e.g., spectroscopic fingerprints, which depend only on the optical system of the instrument.

On the contrary, when the measurements are performed with a flow system, such as chromatographic fingerprints, the position scores are time-dependent. The variations in both x-scores (retention times) and y-intensities result in dissimilarities that can be large, consequently applications are often restricted to the laboratory where they were generated, and signal acquired from other instruments cannot be transferred. The retention times are strongly influenced by the instrument and by small deviations in the operating conditions that produce shifts, warps or drifts in time-scores. Similarly, flow systems also suffer from fluctuations in measuring intensities that may be caused by several factors, such as the sample introduction or the flow rate. This means that the characteristic values from the raw signal data vector need to be corrected, not only when the chromatographic method is run on a different gas chromatograph but also over long time periods, since multivariate methods are intrinsically sensitive to chromatogram drift.

To overcome some of these limitations, an approach based on the use of relative retention times is often applied, i.e., dividing the retention time of the compound of interest by the retention time of an internal standard. Thus, slight variations in relative retention times could be compensated for, which to some extent may even be comparable between different systems. However, initial acquisition delays or entirely different temperature programmes cannot be compensated for with this methodology, and a large number of standards would be necessary to cover the entire time range of the chromatogram [4].

A widely applied strategy aimed at mitigating the adverse problems derived from x-score shifting is the signal alignment. The first attempts to correct this phenomenon for first-order analytical signals, in which chromatographic signals are included, emerged relatively early [5,6]. Alignment of chromatographic signals requires finding a suitable transformation, of their time axis (x-score in tensor notation), the so-called alignment or warping function developed through an alignment algorithm. The main issues to address in providing a useful retention time alignment algorithm are: (i) it must preserve the chemical selectivity differences between samples whilst minimizing run-to-run retention time differences, (ii) it must provide a significantly better precision than that initially provided by instrumentation, and (iii) it must be fast enough to provide high sample throughput. Then, the warping function is determined in order to optimize

a certain objective function by aligning signals with a particular signal called a target. The success of signal alignment is mostly dependent on the type of signal shifts observed (linear or non-linear) and a sufficient flexibility of the warping function. In chromatography, signal shifts are often of a non-linear type and thus their correction may require the application of a highly non-linear warping function, for example, by applying a number of spline functions of a given degree. A comprehensive description of the main methods of chromatogram alignment has recently been reviewed [7].

However, none of these approaches can be considered as a standard procedure from which reproducible universal analytical signals could be obtained in order to solve universal standardization of data. All of them are initially applied to previously selected signal sets, but if new signals are further added to the existing set, these will not be aligned with the others and could be identified as outliers. The use of an external reference allows the transfer of a settled alignment to the new signals. Furthermore, they only perform properly for signals with similar shapes, since the presence of artefacts or new chromatographic peaks may cause the algorithm to confuse the unusual peak with one of the neighbouring ones and the alignment result would not be correct.

In 1958 E. Kovats proposed the use of the retention index (RI) using a n-alkanes homologous series as reference peaks to try to obtain GC retention data in a universal reference system for identifying unknown organic compounds [8,9]. Currently, the RI is an empirical parameter obtained by interpolation, relating the adjusted retention time or the retention factor of the sample component to the adjusted retention times of two standards eluted both before and after the peak of the sample component [10]. Since Kovats proposed the use of n-alkanes, several alternative homologous series of organic compounds (e.g., haloalkanes, n-alkanols, 2-alkanones, alkyl ethers, alkyl acetates, ethyl esters, fatty acid methyl esters and ethyl ester or polynuclear aromatic hydrocarbons, amongst others) have been used in order to avoid the inconvenience associated to the influence of stationary phase polarity, the characteristic matrix sample to be analysed or the development of compound-specific detectors [11,12].

The fact that improvements have been made during decades of the development of a retention index system (RIS), is the reason why the generalized use of GC coupling with MS has become established as the typical strategy for the identification of compounds in complex mixtures relegating to oblivion the previous RIS. However, the idea of using an external reference system of retention scores could be applied to solve new challenges.

In order to be able to use data indistinctly obtained from pairs of near-infrared spectroscopic (NIR) instruments a new methodology was developed in 1991 [13] called 'calibration transfer methodology'. This consists of measuring a set of reference samples in a first instrument, termed parent, master, primary or host, and a multivariate calibration model is then developed. Next, this model is applied to a second one term child, slave, secondary or guest, without the need for further processing. The multivariate model transfer has been mainly applied to optical instruments based on molecular spectroscopy because it is assumed that the x-scores (optical features such as wavelengths, wavenumbers, or frequencies) and intensities (absorbances, transmittances, etc.) only depend on the instrument performance and are invariant throughout successive analyses, even when spaced by a long interval of time. Precise information on instrumental strategies for calibration transfer and correction methods can be found in specialised literature [14,15]. However, although this approach could be considered a promising strategy for addressing the dependence on the chromatographic system, no studies have been reported in which this methodology has been applied to chromatographic techniques.

Finally, retention time locking (RTL) is an instrumental innovation implemented in Agilent GC systems equipped with electronic pneumatic control [16]. Briefly, RTL is based on the development of an empirical function between retention times and head pressure that is fitted from an external standard of each of the analytes under consideration and in this way the chromatograms are recorded and keep the retention times of these analytes fixed at a value previously set by the user.

As far as we are concerned, a new methodology is necessary for the standardization of chromatographic signals against a stable and widely recognized external reference in order to obtain comparable fingerprints, regardless of time, column state and even the equipment used. The aim of this paper is to describe the methodology to be followed for standardizing the signals collected by GC, so that the same kind of samples analysed in different GC equipment in different laboratories and at different times lead to similar instrumental signals (chromatographic fingerprints) which could be considered as equivalent. The paper is structured into three general sections, first of all an overview about instrument-agnosticism methodology and how this can be performed using Matlab is described. Secondly, the application of this methodology for two examples and finally some concluding remarks are presented.

2. BACKGROUND: INSTRUMENT AGNOSTICISM

Significantly different chromatographic fingerprints obtained from different instruments or from the same instrument but in different time periods may occur, even if a single analytical method is applied. This could have a negative effect on the harmonization of results. This fact involves many areas where this separative analytical technique is used and requires the development of a reliable methodology for providing comparable chromatographic fingerprints.

The term instrument-agnostic has already been used to refer to a property that is independent of the instrument [17]. In this sense, instrument agnostizing is the ability to provide standardized and comparable chromatographic fingerprints regardless of the instrument, the particular instrumental conditions or the time of analysis, in order to ensure better reproducibility. The influence of the analytical equipment on the chromatographic fingerprint may be minimized by establishing a proper instrument agnostizing methodology to obtain chromatographic fingerprints when the same analytical method is performed. Instrument-agnostic chromatographic fingerprints are achieved using a common set of both internal and external chemical standards as references which are analysed in the same chromatographic batch as the samples under study. Each type of chemical standard set is used to normalize a signal feature, such as the intensity values and the retention scores. A general description of the whole standardization procedure is shown in a flowchart in Figure 1 and described in the following sub-sections.

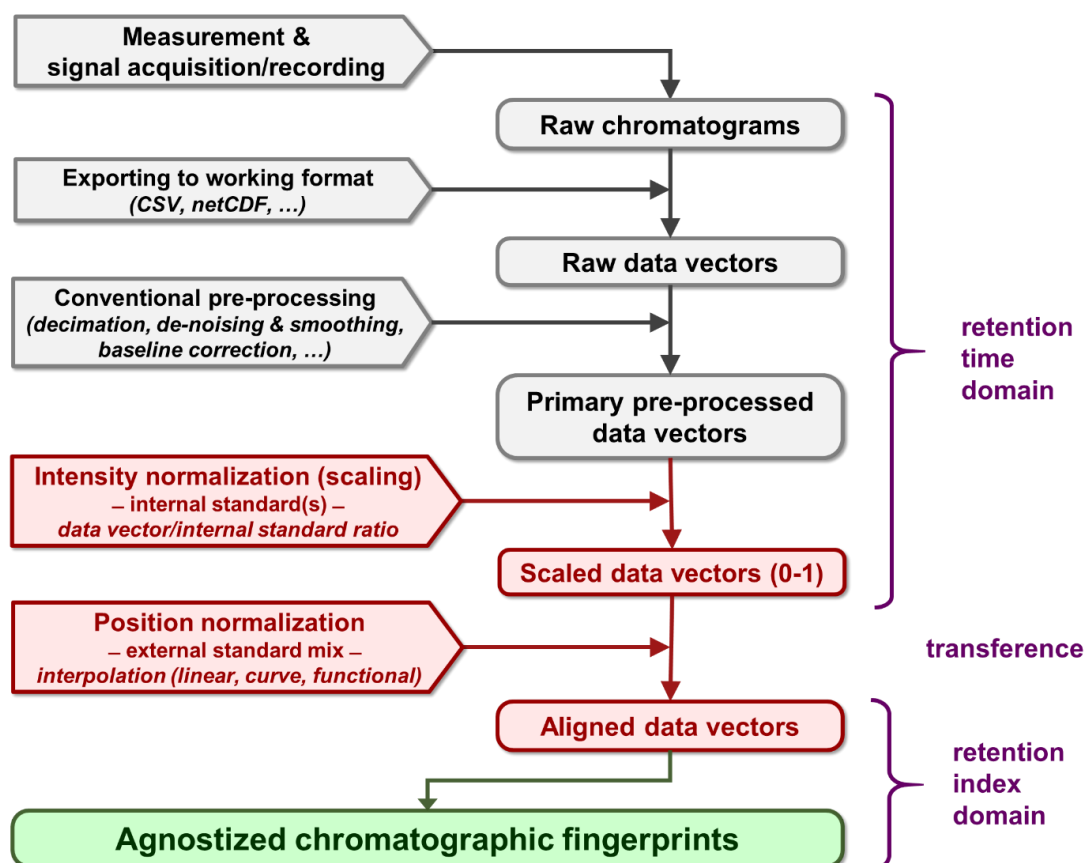


Figure 1. Flowchart showing the main processes and partial results related to the chromatographic signal agnostization.

2.1. Intensity normalization: scaling

The intensity normalization involves the scaling of the intensity data vector in order to provide values within the same intensity range for all the chromatographic fingerprints. A reference 0–1 range is then chosen. The value 0 refers to the baseline in the original chromatogram whilst the value 1 is assigned to the intensity (peak height) of the reference internal standard(s) previously added to the sample test solution at a well-defined and constant concentration. Ideally, the highest signal intensity element in the sample chromatogram data vector should correspond to the internal standard(s), so that once the intensity-normalized chromatogram has been obtained, all the intensity values are within the range 0–1 [18] (see Figure 2(c)). However, it is difficult to define such a concentration beforehand and some elements of the data vector may have an intensity value greater than 1.

The intensity normalization process is based on calculating a relative intensity for each

element of the intensity data vector from the samples (INT_i^{rel}) by dividing the original intensity value (INT_i) by the corresponding reference intensity value of the internal standard(s), denoted by the symbol INT_i^{ref} , using the following expression:

$$INT_i^{norm} = INT_i^{rel} = \frac{INT_i}{INT_i^{ref}}$$

The addition of a single internal standard provides limited information about the occasional fluctuations in the detector during the analysis. Therefore, whenever possible the selection of at least two internal standards is strongly recommended so that they elute into each of the two halves of the sample chromatogram, although it is sometimes difficult to find a suitable chemical compound that appears in the second half of the chromatogram. A good strategy might be to use two of the constituents of the external standard mixture whose characteristics meet the conditions of an internal standard. As a reference intensity the following could be used:

- The intensity of one of these two internal standards, preferably the one that elutes first. In this particular case, the INT_i^{ref} is constant and equal for all elements of the intensity data vector (INT^{ref}). Therefore, the second internal standard is used as a quality control to check the reliability of the chromatographic run.
- The pooled intensity from both internal standards (INT_i^{pooled}) which is specifically obtained for each i -th element of the data vector as a distance-weighted average, compared to the retention time values (RT), for each internal standard. The INT_i^{ref} for each i -th element is different and needs be calculated by applying the following expression:

$$INT_i^{ref} = INT_i^{pooled (IS_1, IS_2)} = \frac{(INT_{IS_1} \times |\Delta t|_{IS_2, i}) + (INT_{IS_2} \times |\Delta t|_{IS_1, i})}{|\Delta t_{IS_2, IS_1}|}$$

where INT_{IS_1} and INT_{IS_2} are the intensity values (height peaks) of both internal standards IS_1 and IS_2 , and $|\Delta t|_{IS_1, i}$, $|\Delta t|_{IS_2, i}$ and $|\Delta t|_{IS_1, IS_2}$ are the absolute values of the differences between the retention times for IS_1 and the i -th element, IS_2 and the i -th element, and IS_1 and IS_2 respectively.

In addition, the ratio-intensity from the intensities of the two internal standards could be used as a useful and valuable quality control metric for each sample chromatogram run (QC metrics = INT_{IS1} / INT_{IS2}). For this purpose, this ratio-intensity must remain constant in successive analyses within a pre-determined tolerance range. A significant deviation from the established value suggests a problem throughout the chromatographic run, thus the result should be discarded and the analysis repeated.

2.2. Retention normalization: signal alignment

Retention normalization is based on verifying the deviations in the experimental values of the retention times of a set of chemical standards from certain empirical values previously assigned, in order to transfer the identical effect to all the chromatograms acquired from the samples analysed in the same chromatographic batch. The set of chemical standards constitutes an external standard reference mix which is analysed in a test solution independent of the sample test solutions. Therefore, the external standard mix should be analysed at least at the beginning and at the end of each chromatographic batch.

Ideally, the standard mix should be made up of a homologous series of well identified and sufficiently pure chemical compounds with the same functional group, so that the differences in their physical-chemical properties are only based on the increasing mass or chain size. A series of well-defined chromatographic peaks evenly distributed on the retention time axis are assigned a specific empirical retention index which is linked in a similar way to the methodology proposed by Kovats [8,9], e.g., the number of carbon atoms of the fatty molecular chain.

A critical aspect of the retention normalization is the selection of the family of chemical compounds to be used as the reference chemical set since not all potential candidates are consistent with the operating conditions of a particular method. The reference standard mix must be composed of thermally stable compounds at the chromatographic working temperature and preferably have a molecular mass near to the analytes being separated. There are a lot of homologous series candidates to constitute the reference standard mixes, which have been commented about in the introduction section, although the series of n-alkanes or linear fatty acid methyl esters (FAMES) could be considered as the most suitable for GC conventional methods and are also commercially affordable. In addition, the series of triglycerides consisting of three equal fat chains would be suitable for methods based on high-temperature gas

chromatography.

The number of chemical compounds which make up the standard reference mix must be sufficient to cover the whole retention time range in which the target constituents of the samples appear. This number could be higher than necessary, e.g., if a commercial standard mixture is used, but only the sub-set of compounds that elute at times similar to the target constituents of the samples should be considered. A minimum of 4 or 5 compounds would be advisable, although the larger the number of compounds in the same range, the better the available information on the actual behaviour of the chromatographic system to provide a correct normalization from the linear interpolation.

The chromatographic separation between each compound depends on the analytical method concerned. Two analytical options for applying chromatographic agnostizing are considered, so that the method used in the chromatographic separation of the external standard is:

- An identical analytical method to that used for sample analysis is applied. This is the ideal option since chromatographic retention is similar for the external reference constituents and sample compounds.
- A different but closely related analytical method to that used for the analysis of samples is applied in this instance. A constant linear gradient is selected by considering the initial and final temperatures specific to the analytical method for the samples. Such a method is easily reproducible and could be used as a representative standard method for other methods working in the same or similar temperature range.

In either case, and regardless of the analytical method applied, the two stages are successively carried out in the agnostizing process. Figure 2 describes a graphical overview of both stages.

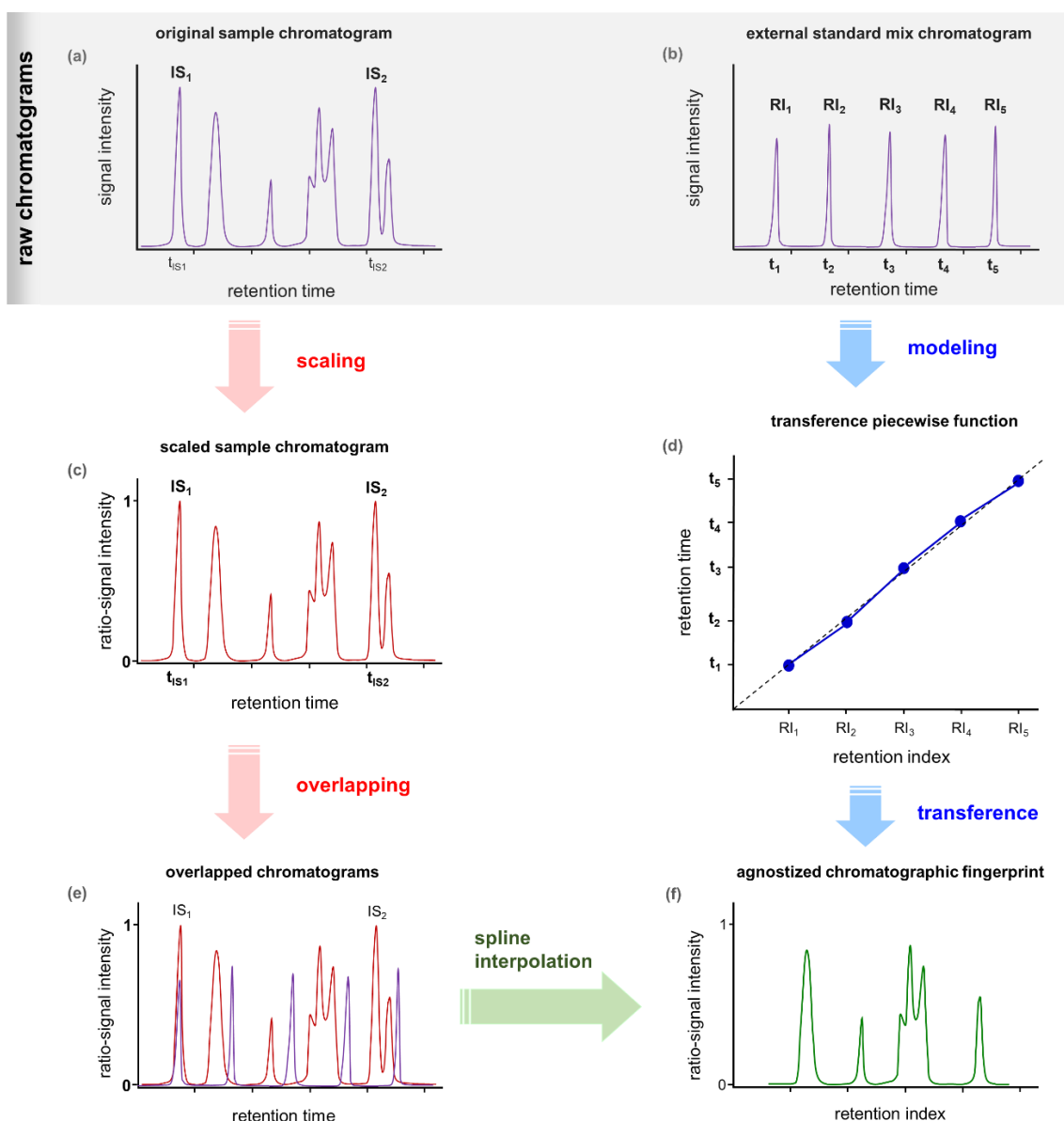


Figure 2. Graphical overview of the chromatographic agnostizing process: (a) raw chromatogram of a random sample; (b) raw chromatogram of the external standard mix; (c) intensity-normalized sample chromatogram (scaled within the 0–1 range); (d) plot of experimentally obtained retention time values for the compounds constituting the reference standard mix versus empirically retention indices (RIs) assigned to each compound (i.e., Kovats indices); (e) both sample and external standard superposed chromatograms (the height peaks of the external standards shown are random and have no influence on the interpolation process; (f) final instrument-agnostic chromatographic fingerprint (the peaks of the internal standards have been removed, as they are not inherent to the sample).

In the first stage, the experimentally obtained retention time values for the compounds constituting the reference standard mix are plotted against the retention indices empirically assigned to each compound, for example, the Kovats indices in the case of

n-alkanes. Usually, a nearly linear behaviour is found, although in some cases a curved pattern appears, then a log-linearization could be applied although this step is not essential. This retention times or indices relationship is defined by a piecewise function which is representative of the chromatographic batch and it may then be applied to each of the retention time values that define the position of the elements of the data vectors obtained from the chromatograms of the samples which are chromatographed in the same batch (see Figure 2(d)). This retention-score normalization process constitutes the second stage and involves the transference of the chromatographic intensity vectors from the instrumental-dependent retention time domain to an instrument-agnostic retention index domain.

The transference may be carried out by applying different mathematical strategies on the retention times of each element in each of the sample data vectors such as: (i) linear spline interpolation between the two values of external standard data vector that it includes, (ii) non-linear spline interpolation (quadratic or cubic) using three or four elements of the external standard data vector, (iii) interpolation from a previously fitted (regressed) polynomial function representative of the whole interval defined by the selected constituents of the external standard mix. Although all of these interpolation procedures allow the assignation of retention indices, the linear spline interpolation between pairs is the easiest way to simplify the calculation process required. This was the option selected by Kovats to estimate the retention indices (RI). Accordingly, the retention index for each i -th element of the sample data vector (RI_i) corresponding to the time t_i is obtained from the following expression:

$$RI_i = RI_{ES(n)} + (t_i - t_{ES(n)}) \times \left(\frac{RI_{ES(n+1)} - RI_{ES(n)}}{t_{ES(n+1)} - t_{ES(n)}} \right)$$

where RI_{ES} and t_{ES} represent each one of the retention indices and retention times, respectively, assigned to the chemical constituents of the external standard mix, and n and $n+1$ symbolize the two chemical compounds belonging to the external standard mix whose retention times bracketed the time of the i -th element under consideration ($t_{ES(n)} \leq t_i < t_{ES(n+1)}$) so that interpolation can be carried out.

Linear spline interpolation provides satisfactory results when the spacing between consecutive retention times of the standard mix constituents is small and the piecewise linear model is a good approximation of the real chromatographic retention

performance. A spacing of no more than 20% of the chromatographic runtime should be satisfactory.

2.3. Programming the agnostizing process: 'GCnostic' Matlab script

The entire mathematical background described in the previous sections should be programmed using proper programming environment software such as Matlab in order to automate and simplify the operational procedure described in Figures 1 and 2 (see electronic supplementary information for more details of the 'GCnostic' script).

The raw data vectors obtained from raw chromatograms are exported in a format compatible with the Matlab environment (for this study, 'Network Common Data Form' (netCDF) file format was used) and recorded in a particular path, then a set of Matlab instructions is executed. These instructions carry out both the pre-processing and intensity and position normalization. A homemade Matlab script, the so-called 'GCnostic' (version 01), has been *ad hoc* developed for this application, which uses some Matlab functions as well as a set of functions included with the Bioinformatics Toolbox™ software to correct potential defects in the data vectors.

The pre-processing carried out by GCnostic is as follows:

- (1) Conversion of the raw data files (from netCDF format to Matlab environment). It makes use of the Matlab function 'ncread' [19,20]. This function reads all the data from the variable *varname* contained in the netCDF file, where the *varname* could be: 'scan_index', 'mass_values', 'intensity_values', 'total_intensity', etc. In particular, a matrix of intensity values (of size $M \times T$), a vector of retention times ($T \times 1$) and a vector of masses ($M \times 1$) are read for each sample file.
- (2) Selection and truncation of the region of interest of the original chromatographic signals (range of retention times and mass-to-charge intervals). In this case, the matrix of intensity values and vectors with the retention time and mass values of each sample are clipped to a certain interval (e.g., [1.7-8.0 min] for retention times and [30-400 Da] for m/z values).
- (3) Setting up the total ion chromatogram (TIC) from the clipped matrix of intensity values. For this, all the intensity values of each of the columns of the intensity matrix are added, obtaining as a result a row vector of the sums of each column.

(4) Resampling of each TIC signal using the 'msresample' function from the Matlab Bioinformatics Toolbox. `[Xout, Intensitiesout] = msresample(X, Intensities, N)` resamples raw noisy signal data. The output signal has N elements ($N = 1000$ in our case) with a spacing that increases linearly within the range $[\min(X) \max(X)]$. X can be a linear or a quadratic function of its index (in this case, it is a vector with retention time values). When a down-sampling takes place the 'msresample' applies a low-pass filter before resampling to minimize aliasing. For the anti-alias filter, the 'msresample' uses a linear-phase FIR filter with a least-squares error minimization. The cut off frequency is set by the largest down-sampling ratio when comparing the same regions in the X and $Xout$ vectors.

(5) Baseline correction using the 'arPLS' algorithm. This algorithm computes the baseline of an input signal using an asymmetric reweighted penalized least square smoothing method (arPLS) which is successfully applied to various spectral analyses. This code is simply an implementation of the algorithm proposed by Baek, et al. [21] who proposed a new weighting scheme based on the generalized logistic function. This method estimates the noise level iteratively and adjusts the weights correspondingly, obtaining slightly better results than other baseline correction methods. The method replaces the weights iteratively by estimating a baseline. If a signal is below a previously fitted baseline, a large weight is given. On the other hand, no weight or a small weight is given when a signal is above a fitted baseline as it could be assumed to be a part of peak. However, in practice, noise can be distributed both above and below the baseline, so it is usually desirable to give the same or similar weights in both cases.

(6) De-noising and smoothing of the chromatographic signals using a Savitzky-Golay least-squares polynomial filter. In this case, the 'smoothdata' function of Matlab is used, selecting the following parameters: a 'sgolay' smoothing method which smoothes according to a Savitzky-Golay filter with a window size of 5 points and a second-order polynomial that is fitted over each window of the data vector. Savitzky-Golay filters are used to smooth out noisy signals with a large frequency span. Savitzky-Golay smoothing filters tend to filter out less of the signal's high-frequency content than the standard averaging 'finite impulse response' (FIR) filters, which makes them particularly interesting for those applications where it is desirable to preserve the height of the raw signal peaks. However, they are less successful at rejecting noise when noise levels are particularly high [22,23,24,25].

Additionally, 'GCnostic' normalizes the intensity values of the pre-processed primary

vectors by programming the set of expressions described in section 2.1. This section incorporates a decision parameter in order to apply the required process according to the internal standard(s) added to each sample.

Finally, the transfer function and spline interpolation for retention normalization is carried out by applying linear interpolation, where a linear behaviour in the relationship between retention time and retention index is assumed in each interpolation range. As a preliminary step a reduced data vector is created containing only the elements of the retention time and empirically assigned retention index (e.g., certain number of carbons) for each compound present in the external standard mix. The data matrices (intensity-retention time) of the different samples are then fully run through, with each time element being replaced by the corresponding retention index obtained after applying the linear spline interpolation equation described in section 2.2.

3. EXPERIMENTAL APPLICATIONS

Two analytical examples have been developed to test the applicability of the proposed agnostizing methodology. For this, six virgin olive oil samples were analysed in duplicate in different time periods using GC-MS in order to obtain the characteristic chromatographic signal from two significant fractions: volatile organic compounds and triglycerides.

A Varian 3800 GC (Varian Inc., Palo Alto, CA, USA) equipped with a split/splitless injector coupled to a Varian 4000 ion trap mass spectrometer (Varian Inc., Palo Alto, CA, USA) was used. The instrument includes an auto-sampler system (CombiPal; CTC Analytics, Zwingen, Switzerland) Helium (99.999%; Air Liquid) was used as the carrier gas (flow rate $1.5 \text{ mL} \cdot \text{min}^{-1}$).

In order to have the corresponding data vectors, each recorded chromatogram, originally in the particular proprietary format of the instrument software, was exported to the netCDF format using the OpenChrom® version 1.3 software (Lablicate & Scientific community, Hamburg, Germany), freely available at www.lablicate.com/.

3.1. Case study 1: Fingerprints from the volatile organic compound fraction of virgin olive oil using the conventional temperature gas chromatographic method

A working solution of $100\ \mu\text{g}\cdot\text{mL}^{-1}$ was prepared and analysed as the external standard mix by dilution in n-hexane of the commercial mixture of 34 n-alkanes standard solution from C7 to C40 ($1000\ \mu\text{g}\cdot\text{mL}^{-1}$ in hexane, Supelco Analytical, Bellefonte, PA, USA). In addition, 0.1 g of a mixture solution of fluorobenzene (99%; Sigma-Aldrich, St Louis, MO, USA) and 4-fluorobiphenyl (95%, Sigma-Aldrich, St Louis, MO, USA) of 0.1 and $0.5\ \text{mg}\cdot\text{g}^{-1}$ respectively, prepared in recently refined olive oil, was added to 4.9 g of each virgin olive oil sample as internal standards. A more detailed explanation of both sample preparation and chromatographic conditions are described in a previous paper [26]. The chromatographic method is based on a conventional temperature range from 80 to 250°C to obtain a characteristic and suitable chromatogram. The chromatographic conditions for sample analysis are shown in Table 1.

The method used for the analysis of the external standard mix differs slightly from the one used for sample analysis which is described in Table 1. Instead of SPME fibre, a 10 μL syringe was used to introduce the standard mix solution directly into the chromatographic system (injection volume 2 μL) and the time of analysis was extended to 25 minutes. It should be noted that in this case the same temperature programme was applied as that used to analyse the samples (see Table 1); however, a single linear gradient could also have been applied according to the considerations described in section 2.2.

The raw total ion chromatograms obtained are shown in Figures 3 and 4. Figure 3a shows the chromatograms of the same aliquot of the external standard mix analysed at the beginning and end of the same analytical batch. Conversely, Figure 3b shows the chromatograms of the external standard mix analysed in different month. It is clear how successive analyses over time have a visible effect on retention times. Although the external standard mix is composed of 34 n alkanes, only 25 of them are shown in Figure 3a (C10 – C34). It should be noted that the first three n alkanes (C7 to C9) elute into the solvent peak (hexane, C6) and they cannot be shown. From the 25 compounds shown in Figure 3, only the 13 orange-highlighted compounds (from C10 to C22) were used for the retention normalization. For this case study, the differences in the chromatograms are only noticeable in the second half of the chromatograms, regardless of the acquisition date.

Table 1. Operational condition for gas chromatographic fingerprinting of volatile organic compound in virgin olive oil – conventional temperature gas chromatographic method.

SPME fibre	
Composition	DVB/CAR/PDMS (divinylbenzene/carboxen/polidimetilsiloxano)
Needle size	24 gauge needle number
Fibre coating	50/30 µm
Fibre length	2 cm
Microextraction	
Incubation temperature	40 °C
Pre-incubation time	2 min
Extraction time	30 min
Injector	
Desorption time	2 min
Fibre cleaning time	6 min
GC column	
Commercial name	HP-INNOWAX
Stationary phase	Polyethylene glycol
Dimensions	30 m × 0.25 mm × 0.25 µm
GC oven programme	
Injector temperature	250 °C
Column temperature ramp	80 °C (2.5 min), 240 °C (40 °C/min), 240 °C (7.5 min)
Chromatographic runtime	14 min
Mass spectrometer conditions	
Solvent delay	1.7 min
Transfer line temperature	250 °C
Ion source temperature	230 °C
Ionization electron energy	70 eV
Emission current	10 µA
Scan speed	1.2 s
Acquisition mode	Full-scan, m/z range 10-400

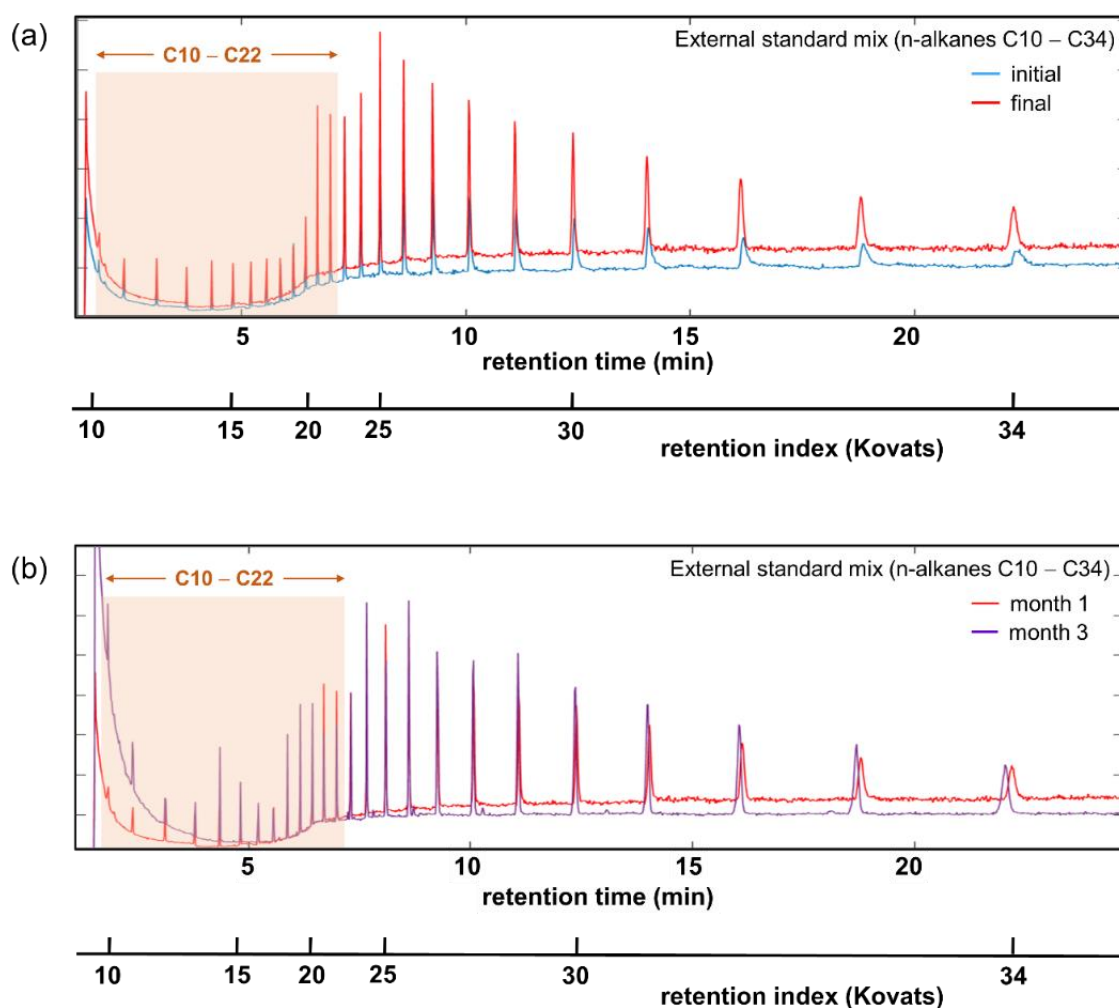


Figure 3. Raw total ion chromatograms acquired in full scan mode from the external standard mix constituted by the n-alkanes series (C10 to C34) applying a conventional temperature gas chromatographic method in the domains of retention times and retention indices, obtained: (a) at the beginning and at the end of the same chromatographic batch, and (b) from two different analysis (two months apart).

Figure 4 shows the chromatogram of the external standard mix (only the C10 – C22 interval is now considered), the raw chromatograms and the instrument-agnostic fingerprints obtained two months apart from one of six samples analysed.

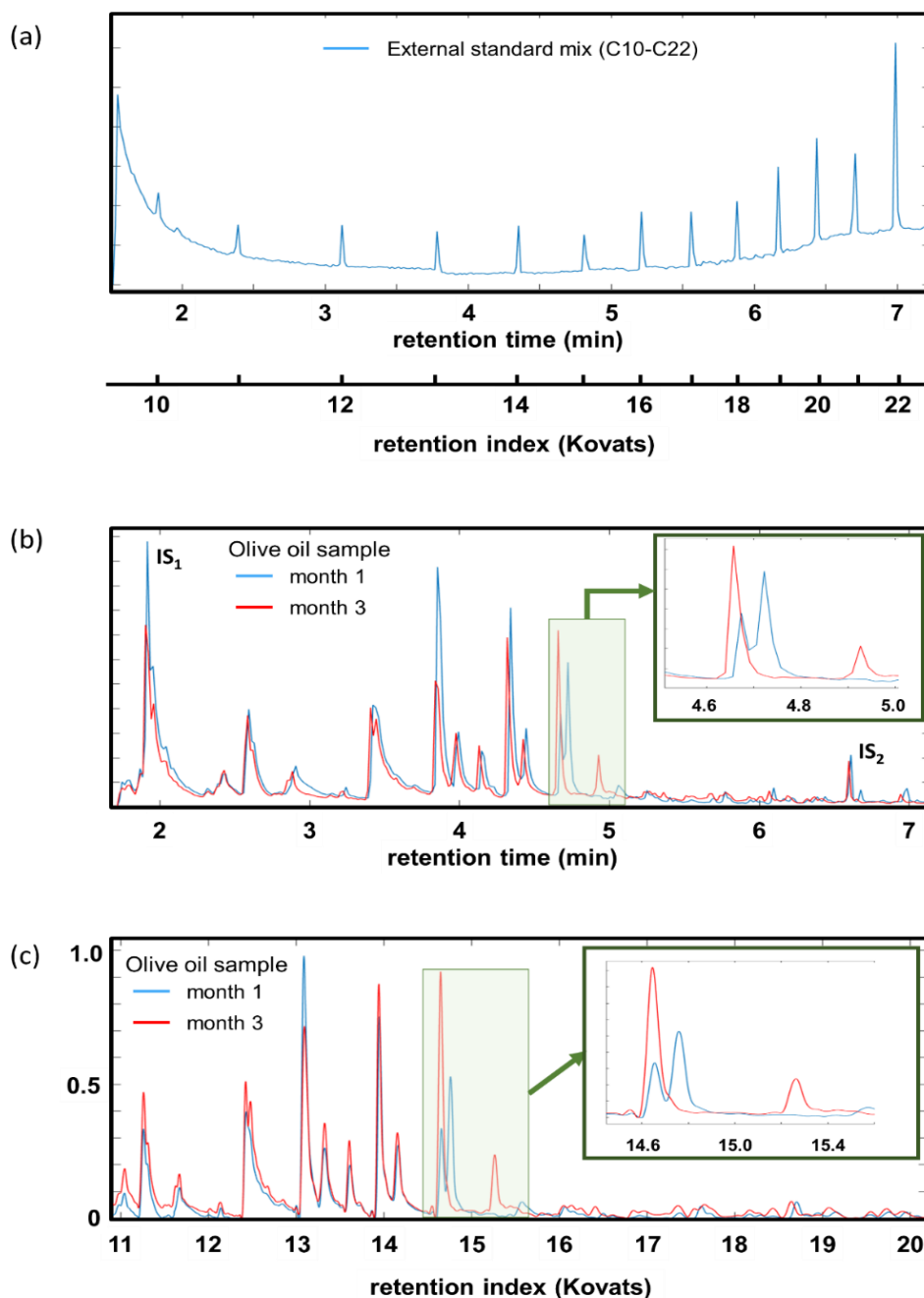


Figure 4. Raw total ion chromatograms and instrument-agnostic fingerprints obtained from the volatile organic compound fraction present in virgin olive oil samples. (a) Chromatographic region of the n-alkane mix (C10 to C22) used for retention normalizing (x-axis: retention time and retention indices domain); (b) raw chromatograms from the same olive oil sample analysed two months apart (month 1 and month 3); and (c) instrument-agnostic chromatographic fingerprints.

The differences in the intensity and retention time values for the raw chromatogram are shown in Figure 4b. It can be seen how the IS_1 intensity in the chromatogram for month 3 is significantly lower than the same peak for month 1, although the same amount of

internal standard solution was added to all the samples. A significant difference in the retention time for the same peak and the occurrence of a new peak is observed in the zoomed region from the same figure (region framed in light green). These differences result from chemical changes during the storage period between the different analyses. Moreover, Figure 4c shows the corresponding chromatographic fingerprints once the agnostizing methodology has been applied.

Even though two internal standards were added to carry out the intensity normalization, only IS₁ (fluorobenzene) was able to be used for this purpose. After several analyses, it was observed that within the same analytical batch the intensity value of IS₂ (4-fluorobiphenyl) showed a high dispersion in the intensity values, which were not consistent with the IS₁ intensity whose values remained constant. This behaviour might be caused by the high boiling point that hinders the phase change, so the recovery of this compound on the fibre is not repeatable. Note that, after intensity normalization, all the ratio-intensity values are scaled in the 0–1 range. In addition, it can be seen how each chromatogram peak is located on the same retention index, regardless of the original retention time in the raw chromatograms. The correction and alignment carried out is particularly evident in the highlighted region, where there was a clearly visible lengthening of the raw chromatogram.

Therefore, once the instrument-agnostic fingerprints are obtained, the differences observed between the processed chromatographic signals are only due to compositional changes in the samples.

3.2. Case study 2: Fingerprinting of triglyceride in vegetable olive oil using the high-temperature gas chromatographic method

Here we describe a second case study which is typified by using high-temperatures for the chromatographic separation (range 250–400°C). For this purpose, a standard mix of trimyristin (95% Tokyo Chemical Industry UK Ltd., Oxford, UK), tripalmitin (99%, Sigma-Aldrich, St Louis, MO, USA), tristearin (99%, Sigma-Aldrich, St Louis, MO, USA) and triarachidin (98% Tokyo Chemical Industry UK Ltd., Oxford, UK) was externally analysed at the beginning and the end of the chromatographic batch. This standard mix was prepared by solving these compounds directly in chloroform to a final concentration of 1 mg·mL⁻¹ for each triglyceride.

The samples were prepared according to the procedure described in a previous study

[27]. Both the external standard mix and olive oil samples were analysed four days apart according to the chromatographic conditions described in Table 2.

Table 2. Operational conditions for gas chromatographic fingerprinting of triglycerides in vegetable olive oil – the high temperature gas chromatographic method.

Syringe	
Volume	10 μ L
GC column	
Commercial name	Rtx-65T
Stationary phase	Crossbond® 65% diphenyl; 35% dimethyl polysiloxane
Dimensions	30 m $\times$ 0.32 mm $\times$ 0.1 μ m
GC oven programme	
Injection volume	2 μ L
Injector temperature	370 °C
Column temperature ramp	320 °C to 370 °C (4 °C/min), 370 °C (3.5 min)
Chromatographic runtime	16 min
Mass spectrometer conditions	
Solvent delay	2 min
Transfer line temperature	350 °C
Ion source temperature	250°C
Ionization electron energy	70 eV
Emission current	10 μ A
Scan speed	1.2 s
Acquisition mode	Full-scan, m/z range 200-1000

No internal standard(s) were added to the sample test solution, and only the applicability of retention normalization was verified. Note that the Kovats retention indices are not used in this case study but rather alternative empirical retention indices based on the number of carbons of the fatty chains are proposed by the authors. These integers are intuitive and easier to use, and the changeover does not affect the intended purpose of chromatogram alignment.

The raw total ion chromatograms acquired in full scan mode and the instrument-agnostic fingerprints obtained after application of the methodology proposed are shown in Figure 5.

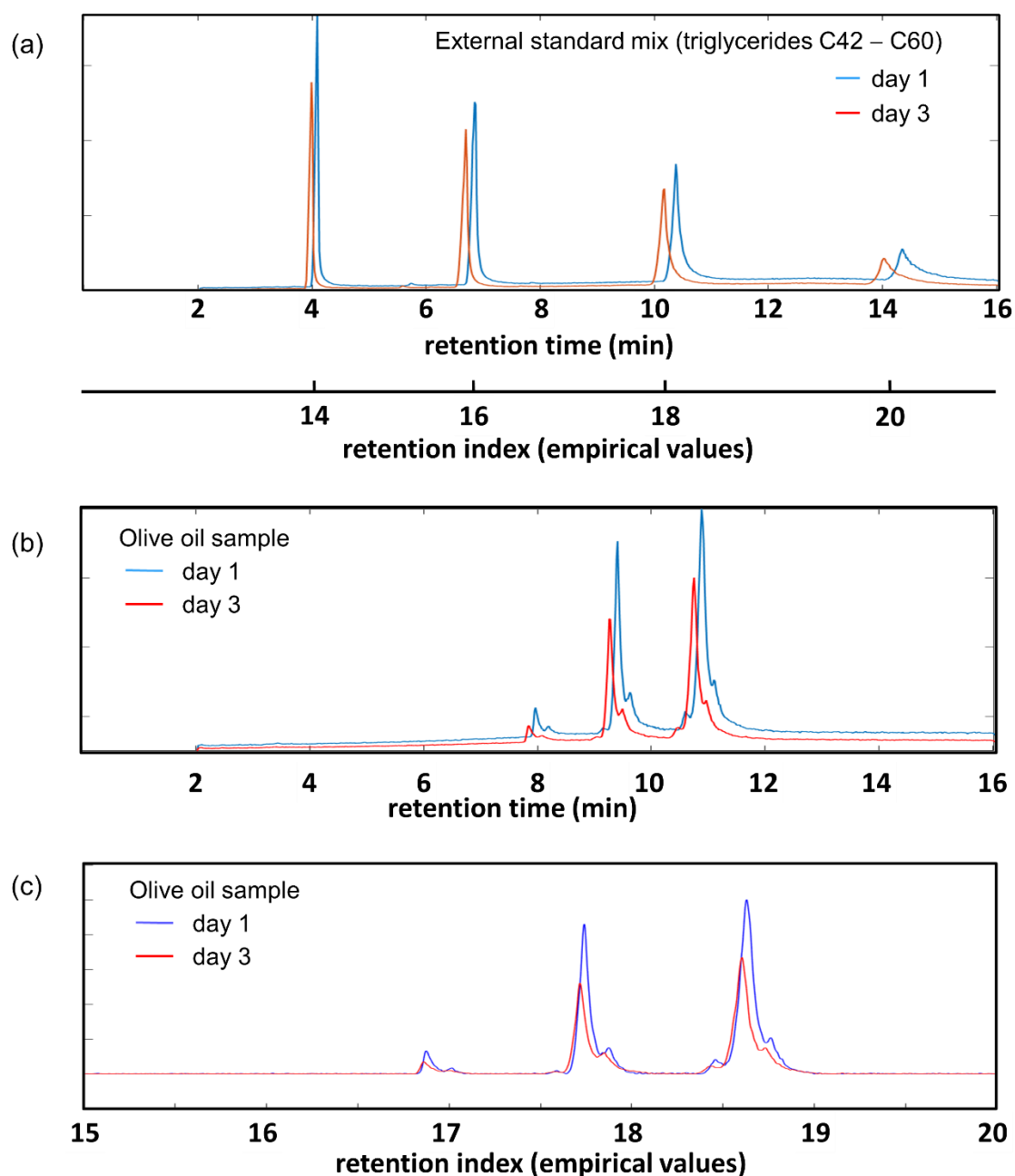


Figure 5. Raw total ion chromatograms and instrument-agnostic fingerprints obtained from the triglycerides in olive oil samples acquired on different days. (a) Raw chromatogram of the triglyceride standard mix ($3\times C14$ to $3\times C20$) used as external standard (x-axis: retention time and retention indices domain); (b) raw chromatogram from the same olive oil sample obtained on different days; and (c) instrument-agnostic chromatographic fingerprints.

Figure 5a shows how all retention times from the four chemical constituents of the external standard mix are shifted to higher values with respect to the initial values. Furthermore, Figures 5b and 5c show the chromatograms from the same sample before and after applying the agnostizing methodology. Particularly, Figure 5c shows

that the same retention index values were assigned to the whole chromatograms after the retention normalization.

4. CONCLUSIONS & FINAL REMARKS

The innovative instrument agnostizing methodology we propose establishes a significant improvement in obtaining harmonised and comparable chromatographic signals in order to achieve new standardized fingerprints or analytical signals regardless of the instrument or the raw acquired data used. According to the results in both case studies, the fingerprints are nearly coincidental regardless of the time and the chromatographic state of the same instrument throughout the analysis. The use of a proper external standard mix allows the transfer of the instrument influence to the remaining chromatograms, whether these have been obtained in the same chromatographic batch or from diverse batches analysed on different days. Thus, reproducible signals are obtained in which a change in the chromatographic shape is not a problem when they are aligned. This new approach would enable, for example, fingerprints of the same food sample from two different instruments to have a high degree of similarity. Furthermore, working with standard chromatographic fingerprints would open up the opportunity of creating a unique multivariate model, either for use in qualitative (classification) or quantitative (quantification) applications, which would allow the introduction of results obtained by different instruments in different laboratories.

Last but not least, by using the methodology we have proposed, it will be possible to generate databases of chromatographic fingerprints obtained from different foods, which can be shared by the scientific community, facilitating the resolution of many problems. It will be possible to have reference signals (chromatographic fingerprints) of all types of foods, which could be used anywhere in the world and will facilitate the instruments traceability. These databases would respond to priority topic number 5 of the "Horizon 2020: EFSA's Priority Research Topics": "Development of European Cloud-based integrated data collection, data management, data analysis and reporting system for food safety" [28], which proposes the development of a database to harmonize, collect, analyse and share comparable food analysis information. It is clear that the agnosticism of the instrument is an ideal methodology to apply in the case of analytical signals that require previous chromatographic separation, since these could depend on time and, therefore, on all the uncontrolled variables that may affect them in the course of obtaining the signal. In addition, the proposed methodology would open

up the way to also standardize comprehensive 2D chromatographic or GC-IMS signals. We are currently working on developing a methodology to address these second-order signals.

Furthermore, using a chemometric approach by unsupervised multivariate pattern recognition methods (as HCA and PCA) the natural grouping of the units belonging to each one of the OO-RM could be highlighted and possible chemical markers could be identified.

FUNDING

This research was partially supported by the European Agricultural Fund for Rural Development (EAFRD) by the European Innovation Partnership for Agricultural Productivity and Sustainability (EIP-AGRI), and the 'Consejería de Agricultura, Ganadería, Pesca y Desarrollo Sostenible' (Department of Agriculture, Livestock, Fisheries and Sustainable Development, Regional Government of Andalusia, Spain) for the funding of the Operational Group project and 'INTERPANEL' (GOP3-JA-16-0012). In addition, the authors acknowledge technical and partial funding support from the 'Organización Interprofesional del Aceite de Oliva Español' (Interbranch Organisation of Spanish Olive Oil, <https://www.oliveoilsfromspain.org/>).

REFERENCES

- [1] A.M. Jiménez Carvelo, S. Martin Torres, L. Cuadros Rodríguez, A. Gozález Casado, Non-targeted fingerprinting approaches, in: C.M. Galanakis (Ed), Food Traceability and Authentication, Academic Press / Elsevier, Oxford, 2021, ch. 6, pp. 163–193.
<https://doi.org/10.1016/B978-0-12-821104-5.00010-6>
- [2] K.S. Booksh, B.R. Kowalski, Theory of analytical chemistry, Anal. Chem. A 66 (1994) 782A–791A.
<https://doi.org/10.1021/ac00087a718>
- [3] K. Danzer, J.F. van Staden, D. Thorburn Burns, Concepts and applications of the term “dimensionality” in analytical chemistry (IUPAC Technical Report), Pure Appl. Chem. 74 (2002) 1479–1487.
<https://doi.org/10.1351/pac200274081479>
- [4] M. Fortini, M. Migliorini, C. Cherubini, L. Cecchi, L. Calamai, Multiple internal standard normalization for improving HS-SPME-GC-MS quantitation in virgin olive oil volatile organic compounds (VOO-VOCs) profile, Talanta 165 (2017) 641–652.
<https://doi.org/10.1016/j.talanta.2016.12.082>
- [5] K.J. Johnson, B.W. Wright, K.H. Jarman, R.E. Synovec, High-speed peak matching algorithm for retention time alignment of gas chromatographic data for chemometric analysis, J. Chromatogr. A 996 (2003) 141–155.
[http://dx.doi.org/10.1016/S0021-9673\(03\)00616-2](http://dx.doi.org/10.1016/S0021-9673(03)00616-2)
- [6] M. Daszykowski, Y. Vander Heyden, C. Boucon, B. Walczak, Automated alignment of one-dimensional chromatographic fingerprints, J. Chromatogr. A 1217 (2010) 6127–6133.
<https://doi.org/10.1016/j.chroma.2010.08.008>
- [7] R.H. Jellema, A. Folch-Fortuny, M. MWB Hendriks, Variable shift and alignment, in: S.D. Brown, R. Tauler, B. Walczak (Eds), Comprehensive Chemometrics: Chemical and Biochemical Data Analysis, second edition, Elsevier B.V., Amsterdam, 2020, vol. 3, ch. 3.05, pp. 115–136.
<https://doi.org/10.1016/B978-0-12-409547-2.14886-3>

- [8] E. Kovats, Gas-chromatographische charakterisierung organischer verbindungen. Teil 1: retentions indices aliphatischer halogenide, alkohole, aldehyde und ketone, *Helv. Chim. Acta* 41 (1958) 1915–1932.
<https://doi.org/10.1002/hlca.19580410703-2>
- [9] G. Tarjan, Sz. Nyiredy, M. Gyor, E.R. Lombosi, T.S. Lombosi, M.V. Budahegyi, S.Y. Meszaros, J.M. Takacs, Thirtieth anniversary of the retention index according to Kováts in gas-liquid chromatography, *J. Chromatogr. A* 472 (1989) 1–92.
[https://doi.org/10.1016/S0021-9673\(00\)94099-8](https://doi.org/10.1016/S0021-9673(00)94099-8)
- [10] IUPAC, Compendium of Chemical Terminology, 2nd ed. (the "Gold Book"). Compiled by A. D. McNaught and A. Wilkinson, Blackwell Scientific Publications, Oxford (1997). Online version (2019-) created by S. J. Chalk. ISBN 0-9678550-9-8.
<https://doi.org/10.1351/goldbook.R05360>
- [11] B. d'Acampora Zellner, C. Bicchi, P. Dugo, P. Rubiolo, G. Dugo, L. Mondello, Linear retention indices in gas chromatographic analysis: a review, *Flavour Fragr. J.* 23 (2008) 297–314.
<https://doi.org/10.1002/ffj.1887>
- [12] V.I. Babushok, Chromatographic retention indices in identification of chemical compounds, *Trends Anal. Chem.* 69 (2015) 98–104.
<https://doi.org/10.1016/j.trac.2015.04.001>
- [13] Y. Wang, D.J. Veltkamp, B.R. Kowalski, Multivariate instrument standardization, *Anal. Chem.* 63 (1991) 2750–2756.
<https://doi.org/10.1021/ac00023a016>
- [14] J. Workman Jr., A review of calibration transfer practices and instrument differences in spectroscopy, *Appl. Spectrosc.* 72 (2018) 340–365.
<https://doi.org/10.1177/0003702817736064>
- [15] S.D. Brown, Transfer of multivariate calibration model, in: S.D. Brown, R. Tauler, B. Walczak (Eds), *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*, second edition, Elsevier B.V., Amsterdam, 2020, vol. 3, ch. 3.18, pp. 359–391.
<https://doi.org/10.1016/B978-0-12-409547-2.00644-2>

-
- [16] N. Etxebarria, O. Zuloaga, M. Olivares, L.J. Bartolomé, P. Navarro, Retention-time locked methods in gas chromatography, *J. Chromatogr. A* 1216 (2009) 1624–1629.
<http://dx.doi.org/10.1016/j.chroma.2008.12.038>
- [17] K. Georgouli, M.T. Osorio, J. Martinez del Rincon, A. Koidis, Data augmentation in food science: synthesising spectroscopic data of vegetable oils for performance enhancement, *J Chemometr* 32: e3004 (2018) 1-15.
<https://doi.org/10.1002/cem.3004>
- [18] L. Cuadros Rodríguez, A. González Casado, A.M. García Campaña, J.L. Vílchez, Ensuring both normality and homoscedasticity of chromatographic data-ratios for internal-standard least-squares calibration, *Chromatographia* 47 (1998) 550–556.
<https://doi.org/10.1007/BF02467494>
- [19] K.R. Murphy, P. Wenig, G. Parcsi, T. Skov, R.M. Stuetz, Characterizing odorous emissions using new software for identifying peaks in chemometric models of GC-MS datasets. *Chem Intel Lab Sys* 118 (2012) 41–50.
<https://doi.org/10.1016/j.chemolab.2012.07.006>
- [20] T. Skov, R. Bro, Solving fundamental problems in chromatographic analysis, *Anal Bioanal Chem*, 390 (1) (2008) 281–285.
<https://doi.org/10.1007/s00216-007-1618-z>
- [21] S.J. Baek, A. Park, Y.J. Ahn, J. Choo, Baseline correction using asymmetrically reweighted penalized least squares smoothing, *Analyst* 140 (2015) 250–257.
<https://doi.org/10.1039/C4AN01061B>
- [22] A. Savitzky, M. Golay, Smoothing and differentiation of data by simplified least squares procedures, *Anal. Chem.* 36 (1964) 1627-1639.
<https://doi.org/10.1021/ac60214a047>
- [23] S.J. Orfanidis, *Introduction to Signal Processing*. Englewood Cliffs, Prentice Hall INC., New Jersey, 1995.
ISBN: 978-0-13-209172-5
- [24] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, New York, 1992.

- [25] R.W. Schafer, What Is a Savitzky-Golay filter? [Lecture Notes], IEEE Signal Proc Mag, 28 (2011) 111–117.
<https://doi.org/10.1109/MSP.2011.941097>
- [26] F. Ortega Gavilán, L. Valverde Som, F.P. Rodríguez García, L. Cuadros Rodríguez, M.G. Bagur González. Homogeneity assessment of reference materials for sensory analysis of liquid foodstuffs. The virgin olive oil as case study, Food Chem, 322 (2020) 126743.
<https://doi.org/10.1016/j.foodchem.2020.126743>
- [27] C. Ruiz Samblás, F. Marini, L. Cuadros Rodríguez, A. González Casado, Quantification of blending of olive oils and edible vegetable oils by triacylglycerol fingerprint gas chromatography and chemometric tools, J. Chromatogr. B, 910 (2012) 71–77.
<https://doi.org/10.1016/j.jchromb.2012.01.026>
- [28] G. Kass, J. Moon, T. Robinson, Horizon 2020: EFSA's Priority Research Topics, Technical Report EFSA, European Food Safety Authority, 2017.
<http://dx.doi.org/10.2903/sp.efsa.2017.EN-1166>

Capítulo IV

**Certificación de aceites de oliva como materiales
de referencia**

Innovative and thorough practice to certify reference materials for sensory defects of olive oil

Food Chemistry 380 (2022) 132195



Contents lists available at ScienceDirect

Food Chemistry

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

Innovative and thorough practice to certify reference materials for sensory defects of olive oil

Fidel Ortega-Gavilán^{a,*}, José Antonio García-Mesa^b, José Carlos Marzal-Fernández^c,
Francisco J. Moreno-Ballesteros^d, Francisco P. Rodríguez-García^b, Antonio González-Casado^a,
Luis Cuadros-Rodríguez^a

^a Department of Analytical Chemistry, Faculty of Science, University of Granada, C/ Fuentenueva s/n, E-18071 Granada, Spain

^b Institute of Agricultural and Fishery Research & Training (IFAPA) "Venta del Llano", Crta. Nacional Balén-Moril Km 18.5, E-23620 Mengibar, Jaén, Spain

^c Foundation for the Promotion and Development of the Olive Grove and Olive Oil from Andalusia, GEOLIT Technological Park, C/ Sierra Morena 13a, E-23620 Mengibar, Jaén, Spain

^d Regulatory Council of the Protected Designation of Origin "Sierra de Segura", Industrial Park "Llanos de Anillo" s/n, E-23360, La Puerta de Segura, Jaén, Spain

ARTICLE INFO

Keywords:

Certified reference material
Sensory analysis
Taster training
Gas chromatographic fingerprinting
Similarity indices
Robust statistics

ABSTRACT

An important problem in the olive sector is the occasional mismatch of results obtained by different tasting panels when the same olive oil sample is analysed. These discrepancies could be minimised by using reference materials (RM) for taster training. A comprehensive protocol based on the combined use of sensory and instrumental analysis for the certification of olive oil batches as RMs, developed within the framework of the project 'Operational Group INTERPANEL', is proposed. Similarity indices (R^2 , $\cos\theta$ and NEAR) applied on GC-MS fingerprints, allow a successful homogeneity and stability assessment of produced batches. Furthermore, the use of robust statistics combined with a set of instructions developed to remove outliers were applied with excellent results on sensory data set provided by supra-panel composed by more than 100 qualified tasters. This work is the first to provide a comprehensive protocol for certification of real olive oil samples as RM for sensory analysis.

1. Introduction

Olive oil is one of the major characteristic foods of the Mediterranean diet, with a high added value for the consumer due to its high health benefits (Willet et al., 1995; Barjol, 2013). Indeed, the olive oil obtained by milling olive fruits is the only foodstuff that must be subjected to sensory analysis to verify the proper labelling of the two highest quality commercial categories: extra virgin olive oil (EVOO) and virgin olive oil (VOO) (Mariotti, 2014). The analytical methods employed for the physicochemical and organoleptic assessment of olive oils are detailed described in the updated version of Regulation (EC) No 2568/91 (European Commission, 2016). The intensity of particular sensory attributes is determined by a tasting panel which is related to a group of tasters, who have been specifically selected and trained, and who gather to perform the sensory analysis of a product under controlled conditions (COI/T.20/Doc. No 4 / Rev.1, 2007; COI/T.20/Doc. No 14 / Rev. 7, 2021). The 'Panel Test' is a sensory method proposed by the

International Olive Council (IOC) to classify olive oil attending to their organoleptic characteristic measured by a panel of 8–10 tasters who works as a multisensor analytical instrument (COI/T.20/Doc. No 15 / Rev.10, 2018). This is a robust and statistically reliable method, but suffers from a certain lack of reproducibility in the results obtained by different panels due to the inherent subjectivity in the taster (Morales et al., 2013). One part of the industrial sector claims a solution to the differences among results, which have a negative effect, thus causing significant economic losses (Aparicio Ruiz, Morales, & Aparicio, 2019; Circi et al., 2017). In fact, this concern has been highlighted in a recent report requested by the European Commission (European Commission, 2020) in order to seek the origin of the problem and the solutions needed to improve the organoleptic evaluation of VOO and EVOO as conformity control systems. Despite the fact that some European Member States assume that the organoleptic assessment should be excluded from the analysis set carried out on VOO according current legislation, the above-mentioned report concludes that the organoleptic assessment should be

* Corresponding author.

E-mail address: fog@ugr.es (F. Ortega-Gavilán).

<https://doi.org/10.1016/j.foodchem.2022.132195>

Received 14 February 2021; Received in revised form 5 January 2022; Accepted 16 January 2022

Available online 20 January 2022

0308-8146/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

IV.a. RESUMEN (EN ESPAÑOL)

Un problema importante en el sector oleícola es la discrepancia ocasional en los resultados obtenidos por diferentes paneles de cata cuando se analiza la misma muestra de aceite de oliva. Estas discrepancias podrían minimizarse utilizando Materiales de Referencia (MR) para la formación de los catadores. En este trabajo se propone un protocolo integral basado en el uso combinado de análisis tanto sensorial como instrumental para la certificación como MR de diferentes lotes de aceite de oliva, desarrollado en el marco del proyecto "Grupo Operativo INTERPANEL".

El uso de índices de similitud (R^2 , $\cos\theta$ y NEAR) aplicados sobre las huellas instrumentales obtenidas mediante cromatografía de gases acoplada a espectrometría de masas GC-MS, permiten una evaluación fiable de la homogeneidad y la estabilidad de los lotes producidos. Además, el uso de estadísticos robustos combinados con un conjunto de filtros desarrolladas para eliminar los valores anómalos, se aplicó con excelentes resultados sobre un conjunto de datos sensoriales proporcionados por un suprapanel compuesto por más de 100 catadores cualificados. Este trabajo es el primero que proporciona un protocolo completo para la certificación de muestras reales de aceite de oliva como MR para el análisis sensorial.

IV.b. CONCLUSIONES DEL TRABAJO (EN ESPAÑOL)

El uso combinado del análisis sensorial e instrumental, desarrollado dentro del Grupo Operativo "Sistema de referencia y armonización de paneles de cata de aceites de oliva vírgenes de Andalucía – INTERPANEL", se presenta como una metodología rigurosa para la certificación de la intensidad organoléptica de atributos sensoriales, utilizando como caso de estudio ocho lotes de aceites de oliva reales. Las técnicas instrumentales permitieron la evaluación de la homogeneidad y estabilidad mediante la evaluación de la similitud entre las huellas cromatográficas de la fracción de compuestos orgánicos volátiles (COV) obtenidas mediante SPME-GC-MS. Además, se han utilizado técnicas de reconocimiento de pautas no supervisadas (HCA y PCA) para examinar la agrupación natural de cada material producido como paso previo al cálculo de los índices de similitud. Este análisis proporciona información previa del conjunto de datos sobre la posible heterogeneidad dentro de un lote de material concreto.

El conjunto de filtros estadísticos de calidad desarrollado para el tratamiento de los datos de las diferentes sesiones de cata permitió obtener resultados estadísticamente fiables de

forma rápida y sencilla. Este conjunto de filtros de calidad no se limita a la caracterización del aceite de oliva, ya que también puede aplicarse con otras finalidades, como en la armonización de paneles. Aunque los resultados de los tres estimadores del valor central aplicados en el suprapanel de catadores fueron similares, las desviaciones estándar robustas obtenidas a partir del estimador estadístico H15 de Huber fueron menores que las restantes. Teniendo en cuenta estos resultados, se seleccionó el método H15 de Huber para la asignación del valor certificado y la evaluación de la incertidumbre.

Aquellos materiales diseñados para representar atributos sensoriales complejos, como es el caso de los materiales obtenidos a partir de frutos maduros de la variedad botánica "picual", ponen de manifiesto la necesidad de utilizar referencias válidas, como los materiales de referencia certificados (CRM), para la formación de los catadores, ya sean obtenidas mediante la adición de compuestos químicos odorantes, o mediante mezclas de muestras reales. Aunque estas últimas deban ser sustituidas en cada campaña, los catadores reconocen que el defecto se ha producido por procesos naturales. Además, estos materiales requieren un estudio exhaustivo de sus características sensoriales, lo que afecta directamente al tiempo necesario para conseguir la certificación. Mediante el desarrollo del sistema de intercomunicación de paneles desarrollado dentro del proyecto INTERPANEL, se pretende reducir el tiempo necesario para la certificación de cada RM de aceite de oliva. Es por ello que tanto los objetivos del proyecto INTERPANEL como los de este trabajo se encuentran en sintonía con las soluciones propuestas en un reciente informe contratado por la UE (Comisión Europea, 2020).

IV.c. ESTUDIO BIBLIOMÉTRICO

Dentro de la categoría *Food science & technology* – SCIE, muchas revistas incluyen dentro de su ámbito de interés los resultados obtenidos mediante el análisis sensorial de alimentos. Sin embargo, la revista *Food Chemistry* es la única revista perteneciente al primer cuartil (Q1) de esta categoría que incluye de manera específica entre sus temas de interés los resultados obtenidos en estudios de intercomparación, así como el desarrollo de materiales de referencia alimentarios para su uso en ensayos con alimentos, lo que convierte a *Food Chemistry* como la revista más adecuada para la publicación del trabajo que constituye el presente Capítulo.

Los datos relativos al estado de la revista en diferentes categorías recogidas en *Journal Citation Reports* (JCR) ya han sido descritos en el Capítulo II de esta Memoria. No

obstante, es de destacar una particularidad ya que este trabajo ha sido publicado en la modalidad *Gold OA* (Open Access) de la revista *Food Chemistry* gracias a la iniciativa "Apoyo a la Publicación Científica" (APC) que ofrece la Universidad de Granada a sus investigadores. Dicha iniciativa subvenciona el 100% de las tasas de publicación en revistas cuya editorial dispone de un acuerdo previo con la Universidad de Granada, entre las que se encuentra la editorial Elsevier.

IV.d. ARTÍCULO CIENTÍFICO

Innovative and thorough practice to certify reference materials for sensory defects of olive oil

Fidel ORTEGA-GAVILÁN^(a,✉), José Antonio GARCÍA-MESA^(b), José Carlos MARZAL-FERNÁNDEZ^(c), Francisco J. MORENO-BALLESTEROS^(d), Francisco P. RODRÍGUEZ-GARCÍA^(b), Antonio GONZÁLEZ-CASADO^(a), Luis CUADROS-RODRÍGUEZ^(a)

^(a) Department of Analytical Chemistry, Faculty of Science, University of Granada, C/ Fuentenueva s/n, E-18071, Granada (Spain)

^(b) Institute of Agricultural and Fishery Research & Training (IFAPA) "Venta del Llano", Crta. Nacional Bailén-Motril Km 18.5, E-23620, Mengíbar, Jaén (Spain)

^(c) Foundation for the Promotion and Development of the Olive Grove and Olive Oil from Andalusia, GEOLIT Technological Park, C/ Sierra Morena 13a, E-23620, Mengíbar, Jaén (Spain)

^(d) Regulatory Council of the Protected Designation of Origin "Sierra de Segura", Industrial Park "Llanos de Armijo" s/n, E-23360, La Puerta de Segura, Jaén (Spain)

Highlights

- Olive oil certified reference materials for training of tasting panels are produced
- Chromatographic fingerprinting is used for homogeneity and stability verifying
- Exploratory data analysis is applied to screen the grouping of batch materials
- Directives are offered to detect & remove outliers from the characterisation study
- Robust statistics is applied on supra-panel data to decide the reference values

✉ Corresponding author, e-mail: fog@ugr.es; Phone: +34 958240797.

Keywords

Certified reference material, sensory analysis, taster training, gas chromatographic fingerprinting, similarity indices, robust statistics.

Abstract

An important problem in the olive sector is the occasional mismatch of results obtained by different tasting panels when the same olive oil sample is analysed. These discrepancies could be minimised by using reference materials (RM) for taster training. A comprehensive protocol based on the combined use of sensory and instrumental analysis for the certification of olive oil batches as RMs, developed within the framework of the project 'Operational Group INTERPANEL', is proposed. Similarity indices (R^2 , $\cos\theta$ and NEAR) applied on GC-MS fingerprints, allow a successful homogeneity and stability assessment of produced batches. Furthermore, the use of robust statistics combined with a set of instructions developed to remove outliers were applied with excellent results on sensory data set provided by supra-panel composed by more than 100 qualified tasters. This work is the first to provide a comprehensive protocol for certification of real olive oil samples as RM for sensory analysis.

1. INTRODUCTION

Olive oil is one of the major characteristic foods of the Mediterranean diet, with a high added value for the consumer due to its high health benefits (Willett et al., 1995; Barjol, 2013). Indeed, the olive oil obtained by milling olive fruits is the only foodstuff that must be subjected to sensory analysis to verify the proper labelling of the two highest quality commercial categories: extra virgin olive oil (EVOO) and virgin olive oil (VOO) (Mariotti, 2014). The analytical methods employed for the physicochemical and organoleptic assessment of olive oils are detailed described in the updated version of Regulation (ECC) No 2568/91 (European Commission, 2016). The intensity of particular sensory attributes is determined by a tasting panel which is related to a group of tasters, who have been specifically selected and trained, and who gather to perform the sensory analysis of a product under controlled conditions (COI/T.20/Doc. No 4 Rev. 1, 2007; COI/T.20/Doc. No 14 Rev. 5, 2018). The 'Panel Test' is a sensory method proposed by the International Olive Council (IOC) to classify olive oil attending to their organoleptic characteristic measured by a panel of 8-10 tasters who works as a multisensor analytical instrument (COI/T.20/Doc. No 15 Rev. 5, 2018). This is a robust and statistically reliable method, but suffers from a certain lack of reproducibility in the results obtained by different panels due to the inherent subjectivity in the taster (Morales, Aparicio-Ruiz & Aparicio, 2013). One part of the industrial sector claims a solution to the differences among results, which have a negative effect, thus causing significant economic losses (Aparicio-Ruiz, Morales & Aparicio, 2019; Circi et al., 2017). In fact, this concern has been highlighted in a recent report requested by the European Commission (European Commission, 2020) in order to seek the origin of the problem and the solutions needed to improve the organoleptic evaluation of VOO and EVOO as conformity control systems. Despite the fact that some European Member States assume that the organoleptic assessment should be excluded from the analysis set carried out on VOO according current legislation, the above-mentioned report concludes that the organoleptic assessment should be continued, albeit improved. The same report proposes the following solutions: (i) the development of reference samples for olive oil taste training, and (ii) the organization of regular tasting sessions with tasters from different countries. These coincide with the solutions proposed by several authors (COI/T.20/Doc. No 14 Rev. 7, 2021; Aparicio-Ruiz et al., 2019; Aparicio, Conte & Fiebig, 2013) which consider the use of certified reference materials as the best way to harmonise tasting panels, as well as tasters from the same panel, which translates in improved results of the analysis.

A certified reference material (CRM) is defined as *a sufficiently homogeneous and stable material with respect to one or more specified properties, which has been established to be fit for intended use in a measurement process and is accompanied by a certificate that provides the value of the specified property, its associated uncertainty, and a statement of metrological traceability* (ISO Guide 30, 2015). According to ISO Guide 35 (ISO Guide 35, 2017) a reference material must be subjected to three experimental studies before being certified. The first study involves the homogeneity assessment of the produced RM batch. The packaging process is one of the most critical stages in RM production, so it is necessary to verify that the values of one or several targeted properties are similar enough regarding all packaged units of the batch. Furthermore, a stability study is necessary to ensure that the targeted property or properties remains unchanged over the time. RMs produced from biological matrices as perishable foodstuffs which may evolve over time, the stability study is carried out at the same time as the RM is being use. Finally, the characterisation study determines the value of one or several targeted properties or attributes that is used as the reference value. The results of these studies are gathered in the RM certificate, in which the certified value as well as the uncertainty value involved in the whole certification process are specified.

Regarding the experimental procedure, the characterisation study of organoleptic attributes should be performed by sensory analysis, but homogeneity and stability studies might be carried out using instrumental analysis (Rason, Laguët, Berge, Dufour & Lebecque, 2007; Valverde-Som, Ruíz-Samblás, Rodríguez-García & Cuadros-Rodríguez, 2018b). The chromatographic techniques coupled to different signal acquisition systems, e.g. gas chromatography coupled with flame ionization detector (GC-FID) or mass spectrometry (GC-MS), allow to study in an unspecific way the signal profile of the volatile organic compounds (VOCs) fraction responsible of the food flavour. Some recent studies use GC-FID to obtain VOC chromatographic fingerprints of olive oils in order to assess the similarity among these analytical signals by means of similarity indices (Valverde-Som et al., 2018b, Valverde-Som, Ruíz-Samblás, Rodríguez-García & Cuadros-Rodríguez, 2018a; Ortega-Gavilán, Valverde-Som, Rodríguez-García, Cuadros-Rodríguez & Bagur-González, 2020). A high level of similarity can be assumed when no differences between successive fingerprints are found.

Nevertheless, alternative homogeneity studies use the intensity of an instrumental signal, from a specific chemical marker, to monitor the between-samples variability. This approach has been applied within the OLEUM project (www.oleumproject.eu/), a project funded by the EU's Horizon 2020 programme, for homogeneity assessment in samples

intended for physical-chemical analysis (OLEUM project, 2020), although this strategy could also be applied in sensory analysis if a suitable flavouring is used as a reference marker.

The following approaches may be considered in the production of reference materials for sensory analysis of virgin olive oil:

- Design of formulation types to prepare olive oils with a certain organoleptic attribute and with a precise intensity by adding flavourings –substances which cause a specific flavour– on an olive oil free of any perceptible negative attribute, called base. The previous knowledge of the corresponding odour threshold (or recognition threshold) of the substance to be used is always required. This is the strategy being addressed in the framework of OLEUM project (Barbieri et al., 2020). At least two artificial olfactory reference materials for sensory analysis of olive oil having rancid and winey-vinegary defects have been produced within the project and their characteristics and production details discussed in Aparicio-Ruiz et al., 2020.
- Use of olive oil types already produced with the required characteristics or production of these olive oil types from olive fruits or olive oils by applying controlled degradation processes. These olive oils, called concentrates, have a high intensity of a negative attribute so that the final products are obtained by mixing them with a compatible base in a certain percentage. The materials developed following this technique simulate the possible defects present in the olive oil much better than the previous approach. However, this strategy requires the search, development and certification of new materials in each campaign.

This paper describes a certification practice of olive oil reference materials for sensory analysis, which have been developed by applying the second approach previously described within the framework of the project 'Reference and harmonization system of virgin olive oil tasting panels from Andalusia – INTERPANEL' (www.interpanel.es; <https://ec.europa.eu/eip/agriculture/en/find-connect/projects/sistema-de-referencia-y-armonización-de-paneles-de>). The protocol described in this paper involves the combined application of instrumental techniques with sensory analysis to achieve the certification. The homogeneity and stability studies are based on the analysis of volatile organic compounds (VOCs) by gas chromatography coupled to mass spectrometry (GC-MS), while the characterisation study is achieved after the application of the 'Panel Test' method proposed by the IOC. The main novelty of this work is the development and application of an innovative approach which is applied on the characterisation study. It is

based on several quality filters to detect and remove outlier values, combined with the use of robust statistics applied to the sensory data provided by a supra-panel consisting of more than a hundred tasters which belong to different laboratories.

2. MATERIALS AND METHODS

2.1. Olive oil materials

A total of eight olive oil batches, each with a total volume of 50 L, were produced in this work. The materials were designed to be representative of the main attributes and defects (fruity, fusty/muddy, musty and winey-vinegary) proposed by the IOC. The different olive oil batches were produced at the Institute of Agricultural and Fishery Research & Training (IFAPA) "Venta del Llano" using a "Il Molinetto" compact plant (Pieralisi, Jesi, Italy, http://www.pieralisi.com/Viewdoc?co_id=373). The only exception was the wine/vinegary batch, which was provided by the Regulatory Council of the Protected Designation of Origin "Sierra de Segura". Only three of the five attributes or defects considered in this study were prepared at two levels of decreasing intensity symbolised by 2 (medium intensity) and 1 (low intensity), by blending an attribute-concentrated oil with a dilution base. The official agri-food quality control laboratories of Córdoba and Granada (Department of Agriculture, Livestock, Fisheries and Sustainable Development, Regional Govern of Andalusia, Spain) were involved in the setting and selection of the different mixing percentages. The main characteristics and composition are described in Table SM1 (in Supplementary Material).

Note that the material coded as PIC/1 was used as dilution base for materials with fusty/muddy and musty defect.

Each olive oil material was produced in a single batch and PET-amber bottled in units of 200 mL. Both the first and last five bottles from each material batch were then discarded. In order to evenly cover the entire batch, ten representative units were evenly sampled and recoded with numbers from 1 to 10 according to the order of packaging from each olive oil material batch, obtaining a set sample of 80 bottles which were reserved for the homogeneity study. In addition, twelve consecutive bottles of each olive oil material batch were also reserved for a quarterly analytical control of the material stability once the batch has been declared homogeneous. Finally, three additional bottles of each material batch were set aside to further perform the characterisation of the materials. All batches were bottled in July 2019 and were kept in the dark and below 0°C until use, in order to ensure

the volatile profile stability. After the homogeneity assessment, all samples were stored in the dark at 7-8°C, in order to keep the typical storage conditions in a real sensory analysis laboratory.

2.2. Standard solutions

A mixture of n-alkanes standard solution from C7 to C40 (1 mg mL⁻¹ in hexane, Supelco Analytical, Bellefonte, PA, USA) was purchased. A working external standard mix solution of 100 µg mL⁻¹ was prepared by dilution of the commercial mixture in n-hexane.

Furthermore, a mixture solution of fluorobenzene (99%; Sigma-Aldrich, St Louis, MO, USA) and 4-fluorobiphenyl (95%, Sigma-Aldrich, St Louis, MO, USA) of 0.1 and 0.5 mg g⁻¹ respectively, prepared in recent refined olive oil, was used as internal standard.

2.3. Chromatography analysis

The chromatographic analysis of the samples was carried out using a Varian 3800 GC (Varian Inc., Palo Alto, CA, USA) equipped with a split/splitless injector coupled to a Varian 4000 ion trap mass spectrometer (Varian Inc., Palo Alto, CA, USA). The instrument includes an auto-sampler system (CombiPal; CTC Analytics, Zwingen, Switzerland). The analytical column was a capillary column covered with polyethylene glycol (Agilent HP-INNOWAX; 30 m × 0.25 mm internal diameter × 0.25 µm film thickness; maximum temperature 260 °C, Agilent Technologies, Santa Clara, CA, USA).

As previous step, the VOCs were extracted using a 50/30 µm divinylbenzene/carboxen/polydimethylsiloxane StableFlex 24 Ga fibre (Supelco Analytical, Bellefonte, PA, USA). 0.100 g of the IS mixture solution was accurately weighted in a 20 mL headspace vial; then 4.900 g of olive oil RM was carefully added until that the final mass was 5.000 g. The samples were incubated in a shaker oven at 40°C for 2 min and the extraction was carried out for 30 min. Fibre thermal desorption into GC was conducted at 250°C for 2 min in splitless injection mode, and then the fibre was cleaned into the GC with a 1:100 split ratio for 6 min. The column oven temperature of 80 °C was held for 2.5 min, then programmed at 40°C min⁻¹ to 240°C, then this last temperature was kept for 7.5 min. Helium (99.999%; Air Liquid) was used as the carrier gas (flow rate 1.5 mL min⁻¹). The mass spectrometric conditions were as follow: the ion source temperature was held isothermally at 230°C during the chromatographic separation and the transfer-line

temperature at 250°C was maintained through the analysis. The ionization electron energy was 70 eV and the emission current 10 μ A. The chromatograms were recorded in full-scan mode and the average spectra were acquired in the m/z range of 10-400.

Chromatographic analysis was carried out for both homogeneity and stability studies. For homogeneity assessment, which was carried out in October 2019, the first unit from each homogeneity sample set was chosen as reference and analysed in triplicate, whereas the remaining nine units were analysed once. The stability study is carried out in parallel with the use of each reference material by the different laboratories. Once the homogeneity of the material has been verified and the certified property has been characterised, the stability study is carried out. Every three months, one unit of each material is taken from the set of samples reserved for this study. Each unit is analysed in triplicate to obtain the representative fingerprint of the month concerned. In order to control the performance of the instrument, 0.100 mg g⁻¹ IS solution was analysed at the beginning and end of each chromatographic run using the same operational conditions.

2.4. Sensory analysis

Sensory analysis was carried out throughout November 2019, by applying the EU 'Panel Test' method for the characterisation study of RM candidates. A total of 22 laboratories, where each panel was composed by a group of 8-10 tasters trained in accordance with IOC guidelines (COI/T.20/Doc. No 14 Rev.5, 2018), participated in the characterisation study. This group consisted of 19 private laboratories authorised by the Regional Govern of Andalusia for quality control and 3 official ones, all of them located in Andalusia with the exception of an official laboratory located in Madrid.

The results obtained in each tasting session were provided through the panels' intercommunication platform called INTERPANEL, on which the statistical data treatment was also performed.

2.5. Data treatment

2.5.1. Chromatographic data treatment

MS Workstation (version 6.9) software data system (Varian Inc., Palo Alto, CA, USA) was used for chromatographic data acquisition and recording. The chromatograms obtained are firstly transformed in data vectors, each one representing the individual fingerprint of each selected material units for the involved study. In order to obtain the VOC chromatographic fingerprints, each chromatogram was exported to netCDF format using the OpenChrom® version 1.3 software (Lablicate & Scientific community, Hamburg, Germany), available free at www.lablicate.com/.

Afterwards, every chromatogram was pre-processed in order to: (i) remove the lower and upper range values of each fingerprint that did not contain any useful information; (ii) filter and smooth using a Savitzky-Golay least-squares polynomial filter with a window size of 5 points and a second-order polynomial for noise removing; (iii) correct the baseline drifting; (iv) align time shifting and warping with respect to the mix external standard solution; and (v) normalise the intensities by using a suitable internal standard as reference. Data pre-processing was performed using MATLAB, version 9.3 (Mathworks Inc., Natick, MA, USA) and an *ad-hoc* home-programmed MATLAB script, so-called 'GCnostic' (version 1), that runs a number of algorithms of MATLAB Bioinformatics Toolbox, and a specific algorithm for the alignment of chromatographic peaks based on Kovats indices (d'Acampora Zellner et al., 2008; Cuadros-Rodríguez et al., 2021), instead of applying a self-alignment as usual. This function has a new alignment algorithm incorporated to correct the shifting and warping in retention time of chromatographic fingerprint observed both in the same or different sequences. For this purpose, the retention time values are transferred to a new empirical retention index obtained from the analysis of n-alkane mix used as external standard analysed at the beginning and the end of each chromatographic run.

After pre-processing, similarity indices were then calculated by comparing the fingerprint data vectors directly with MATLAB's capabilities. These indices were used to obtain reliable metrics for homogeneity and stability assessing (Valverde-Som et al., 2018a; Ortega-Gavilán et al., 2020).

In addition, an exploratory data analysis was performed in order to evaluate the natural grouping of the homogeneity samples analysed by GC-MS by the means of PLS_Toolbox software version 8.6.1 (Eigenvector Research Inc., Manson, WA, USA). Furthermore, a mean centring was applied on the fingerprint matrix prior to chemometric treatment.

2.5.2. Sensory analysis data treatment

The statistical data processing of the sensory analysis for the characterisation of materials was carried out through the INTERPANEL platform. Based on experience gained by the authors in processing data from sensory analysis, robust statistics were programmed and applied for the estimation of reference values. For each material, the arithmetic mean and some robust estimations of the central value (median and Huber's average H15), and the corresponding standard deviations (SD) (interquartile range SD_{IQR} , equivalent median absolute deviation (MADe) and Huber's $SD(H15)$) were calculated in order to establish the characterisation value (Thompson & Lowthian, 2008). Furthermore, in order to identify and remove inconsistent biased values that may affect the quality of the results; the following quality statistical filters were designed and applied:

- First filter: rejecting the panels whose robust coefficient of variation (CV_{IQR}), obtained by dividing the interquartile range standard deviation (SD_{IQR}) by the median, exceeds 20% (COI/T.20/Doc. No 22, 2005).
- Second filter: rejecting inconsistent panels that have passed the first filter by means of the robust z^* -score (z_p^*), using the Equation 1.

$$z_p^* = \frac{m_p - m_{overall}}{MADe_{overall}} \quad (1)$$

where m_p is the median obtained by the panel p for the attribute under consideration, $m_{overall}$ is the median of the panel medians and $MADe_{overall}$ is the equivalent median absolute deviation taking into account all panels, both for the same attribute. A tasting panel is considered inconsistent if the corresponding z^* -score has an absolute value greater than 2; in this case, the panel is also removed of the study. Once this filter is applied, a single supra-panel is conformed, including all the tasters that belong to the non-eliminated panels.

- Third filter: rejecting inconsistent values of sensory data provided by each taster within the supra-panel. The statistical test is again based on a robust $z_t^*(H15)$ -score, specific to each taster t , which is calculated using the robust Huber's H15 estimators applying the following Equation 2.

$$z_t^*(H15) = \frac{x_t - Av_{sp}(H15)}{SD_{sp}(H15)} \quad (2)$$

where x_t is the organoleptic intensity observed by the taster t , and $Av_{sp}(H15)$ and $SD_{sp}(H15)$ are the estimated average and standard deviation Huber's H15 parameters of the organoleptic intensities for the attribute under consideration from the supra-panel results. Again, a single sensory datum is considered inconsistent when the corresponding $z_t^*(H15)$ has an absolute value greater than 2; in this case, the datum is removed.

The results are considered statistically reliable if the number of panels removed by the first and second filters does not exceed 25% of the total number of participants. Regarding the third filter, the number of values removed should not exceed 20% of the total number of supra-panel tasters. If any of these limits are exceeded, the material is directly qualified as 'non-certifiable'.

After the application of the filters, the suitability of the candidate materials to be certified is checked depending on the value of $SD_{sp}(H15)$ value. Three possible settings are covered:

- (i) $SD_{sp}(H15)$ has a value lower than 1. The material under study is fitted for certification purpose.
- (ii) $SD_{sp}(H15)$ has a value between 1 and 2.5. There is sufficient consensus among the tasters to establish a reference value, but not enough to be certified. The material is then labelled as 'consensus guiding material'.
- (iii) $SD_{sp}(H15)$ has a value higher than 2.5. Due to the high data scattering, the candidate material is qualified as 'non-certifiable'.

The reference value assigned to the organoleptic intensity of the attribute under consideration for each material is defined by the $Av_{sp}(H15)$ estimation from the results provides by the valid tasters that constitute the final supra-panel.

According to ISO Guide 30 and 35 (ISO Guide 30, 2015; ISO Guide 35, 2017), the reference value of materials suitable for certification should be accompanied by an uncertainty value associated with the certification process. A section in ISO Guide 35 (2017) describes how the uncertainty value of a certified reference material (u_{CRM}) is calculated by the Equation 3.

$$u_{CRM} = \sqrt{u_{char}^2 + u_{homog}^2 + u_{stab}^2} \quad (3)$$

where u_{char} is the uncertainty calculated in the characterisation study, and u_{homog} and u_{stab} are respectively the uncertainty components derived from the possible lack of homogeneity and stability under storage conditions. The same section of ISO Guide 35 (2017) specifies that, depending on the nature of the material to be certified, the uncertainty sources are different, thus the above equation could be modified by adding or removing terms. In the case of olive oil, the main uncertainty source comes from the sensory characterisation study and consequently the above Equation 3 is simplified in Equation 4.

$$u_{\text{CRM}} = u_{\text{char}} = \frac{SD_{\text{sp}}(\text{H15})}{\sqrt{n}} \quad (4)$$

where n refers to the number of data from which the standard deviation has been obtained; in this case, the number of tasters not removed in any of the filters described above that constitute the final supra-panel. For each material and attribute under consideration, the certification results are then summarised by the Equation 5.

$$\text{Certified value (CRM)} = X_{\text{CRM}} \pm U_{\text{CRM}} = X_{\text{CRM}} \pm (k \cdot u_{\text{CRM}}) \quad (5)$$

where U_{CRM} is expanded uncertainty obtained by multiplying the u_{CRM} value by the coverage factor k ($k=2$). Finally, the results obtained during the sensory and instrumental analysis are included in the certificate of the reference material according to ISO Guide 31 (ISO Guide 31, 2015).

3. RESULT AND DISCUSSION

3.1. Chemometric exploratory analysis

Unsupervised pattern recognition methods such as hierarchical cluster analysis (HCA) and principal component analysis (PCA) were applied. A fingerprint matrix (96×2000) was built for the study (note that the 96 value were obtained from considering 80 samples and two additional replicates of the first 8 units selected as reference, and 2000 is the number of numerical values acquired by the measuring device and which conform each chromatogram).

3.1.1. Hierarchical cluster analysis (HCA)

HCA were performed by selecting Manhattan distance and Ward method as a distance metric and linkage criterion respectively. Figure 1 shows the dendrogram obtained, where each nesting group corresponds to the olive oil materials which have a specific defect independently of the intensity levels. This fact suggests that the mixing percentages used between intensity levels did not create large differences in the chromatograms.

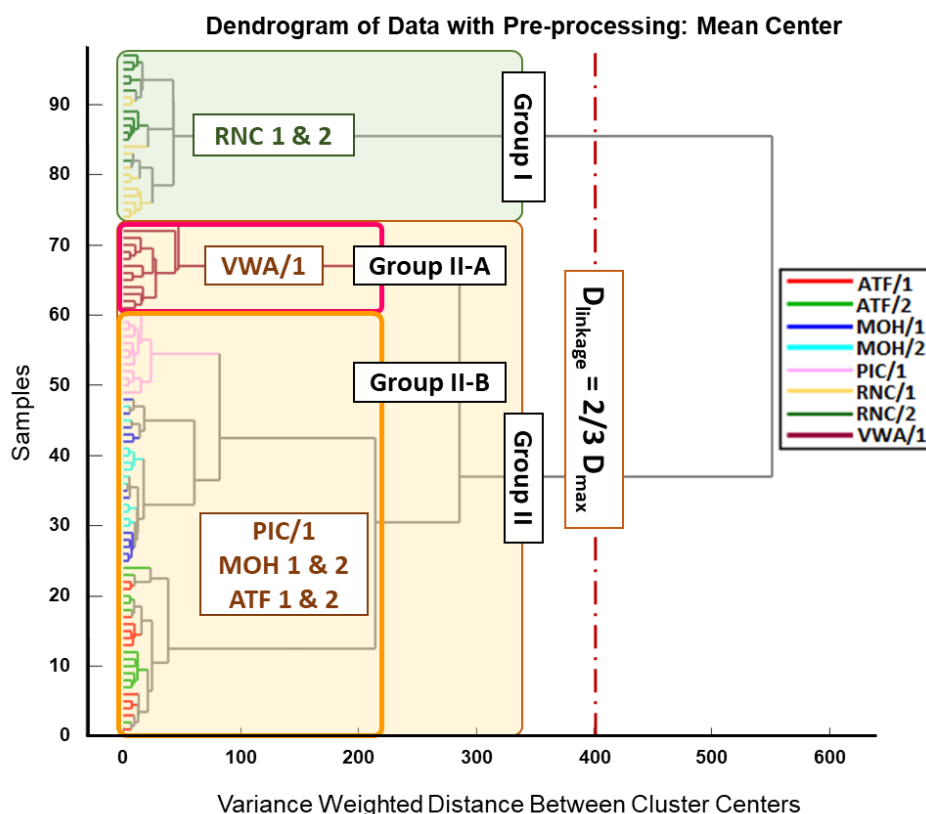


Figure 1. Dendrogram of olive oils candidates to reference materials analysed according its VOCs fingerprints. HCA: Mahalanobis distance and Ward method as distance metric and linked criterion, respectively.

Considering $D_{\text{linkage}} = 2/3 D_{\text{max}}$ as a criterion to define the natural grouping clustering, two groups can be defined. Group I gathers the materials coded as RNC/1 and RNC/2, which were characterised by rancid defect. As detailed in Table SM1, these olive oil materials were obtained from fruits of 'arbequina' botanical variety. Group II is defined by the rest of materials, which were obtained from olives of the 'picual' botanical variety. In addition, this group is divided into two subgroups ($D_{\text{linkage}} = 2/5 D_{\text{max}}$) according to the fruit ripeness degree. Group II-A is composed by the material VWA/1, which was obtained from olives with a medium degree of ripeness, while group II-B collects the remaining materials that

was milled from ripe fruits. It should also be noted that the PIC/1 material was used as a base for diluting other materials, which justifies the grouping with the ATF and MOH materials.

3.1.2. Principal component analysis (PCA)

Three principal components (PCs) explaining 91.9% of the total variance were selected. Figure 2 shows the 3D scores plot on PC1-PC2-PC3 (58.5%, 19.1% y 14.3% of variance, respectively).

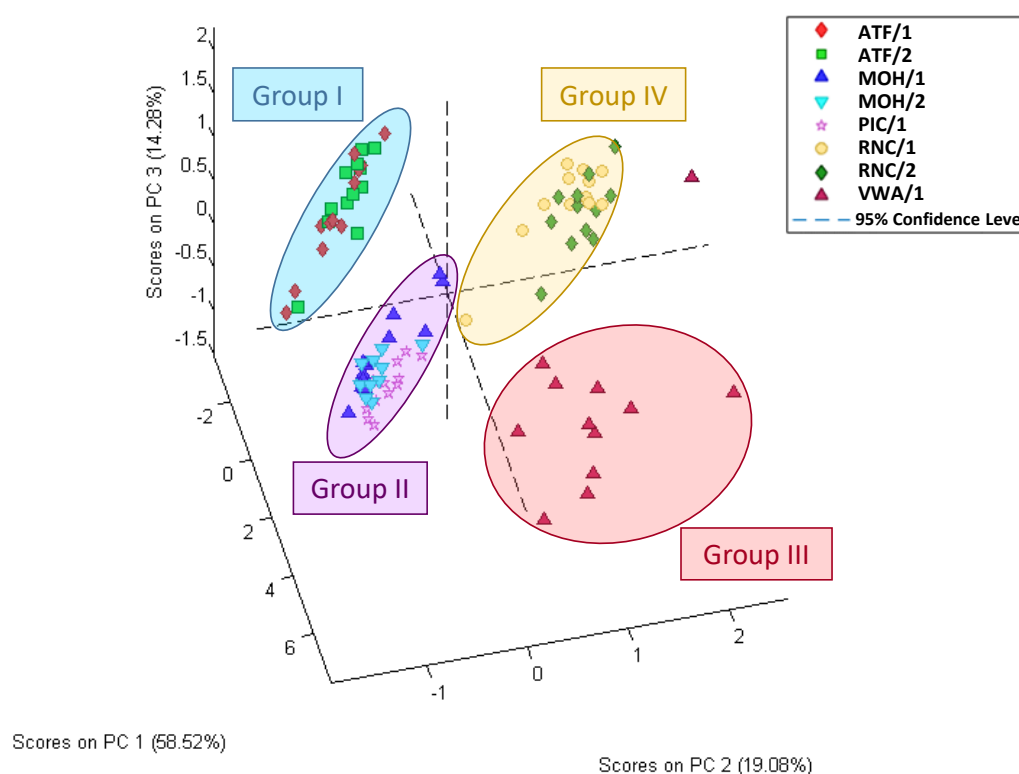


Figure 2. 3D scores plot of PC1-PC2-PC3 obtained from PCA of fingerprint matrix (96×2000).

The 3D scores plot (Figure 2) clearly shows four groups. Three of them are consistent with the materials having the same defect while the fourth group is composed by miscellaneous materials. The grouping is described as follows: (I) ATF/1 and ATF/2 materials having fusty/muddy defect at two intensity levels, are characterised by values close to zero in PC1, negative values of PC2 and positive values of PC3; (II) MOH/1 and MOH/2 materials having musty defect at two intensity levels, as well as PIC/1 material, the undefective base, are characterised by values close to zero in PC1 and negative values

for PC2 and PC3; (III) VWA/1 material, distinctive of wine/vinegary defect, shows positive values of PC1 and PC2 and near or below zero in PC3 (note the presence of possible outlier in this material); and (IV) RNC/1 and RNC/2 materials, which are qualified as rancid defectives at two intensity levels, have negative values in PC1 and PC3 and positive values in PC2.

Therefore, PC1 separates each group according to the type of olive oil used as a dilution base in the production of the different RM. Groups I and II, whose dilution base was obtained from ripe fruits of the picual variety, have similar values in PC1 but very different from the values of Group III, consisting of olive oil obtained from fruits of a different picual variety to the previous groups, and Group IV, whose base was obtained from olives of the arbequina variety. PC2 separates the materials according to the ripeness of the fruit. Groups I and II, RM from ripe fruit, present negative values in PC2 clearly opposed to those presented by Groups III and IV, consisting of RM whose fruit ripeness degree is lower than the previous ones. Furthermore, PC3 differentiates each group according to the defect which the materials were designed when the same type of base and maturity of the fruit is involved. In this way, PC3 allows a clear separation between Groups I and II, whose organoleptic defects were different from each other. It should be noted that the closeness of PIC/1 to MOH/1 and MOH/2 could be due to the low mixing percentages used in their production process, which implies small dissimilarity of the chromatographic signals.

3.2. Similarity analysis for verifying homogeneity and stability of the reference material

The similarity analysis for the homogeneity and stability studies were performed according to methodology described in previous papers (Valverde-Som et al., 2018a and 2018b; Ortega-Gavilán et al., 2020). Briefly, the similarity values in the homogeneity study were obtained by comparing the fingerprint of each batch unit analysed with a reference fingerprint. For each material the reference fingerprint was the average of the results of the analysed unit in triplicate. Both the reference agnostized chromatographic fingerprint and each of the single unit agnostized chromatographic fingerprint for each RM candidate obtained within the homogeneity study are displayed in the supplementary material (Figure SM1 - Figure SM8). Chromatographic fingerprints were considered to be similar when the values of the similarity indices, namely R^2 , $\cos\theta$ and NEAR (Brereton, 2009; Andrić & Hérberger, 2017; Pérez-Robles, Navas, Medina-Rodríguez & Cuadros-

Rodríguez, 2017) (expressed in percentage), were higher than the threshold imposed for this study, 95% for the first two indices and 90% for the third one. The thresholds were defined by the authors based on the experience gained from several studies. Similarity values below these decision thresholds suggest a potential heterogeneity among the bottled units of the packaged batch. Figure 3 shows the similarity control charts of the eight olive oil materials developed. In this figure the relationship between the behaviour of the three similarity indices could be verified, as their values increase and decrease following the same trend.

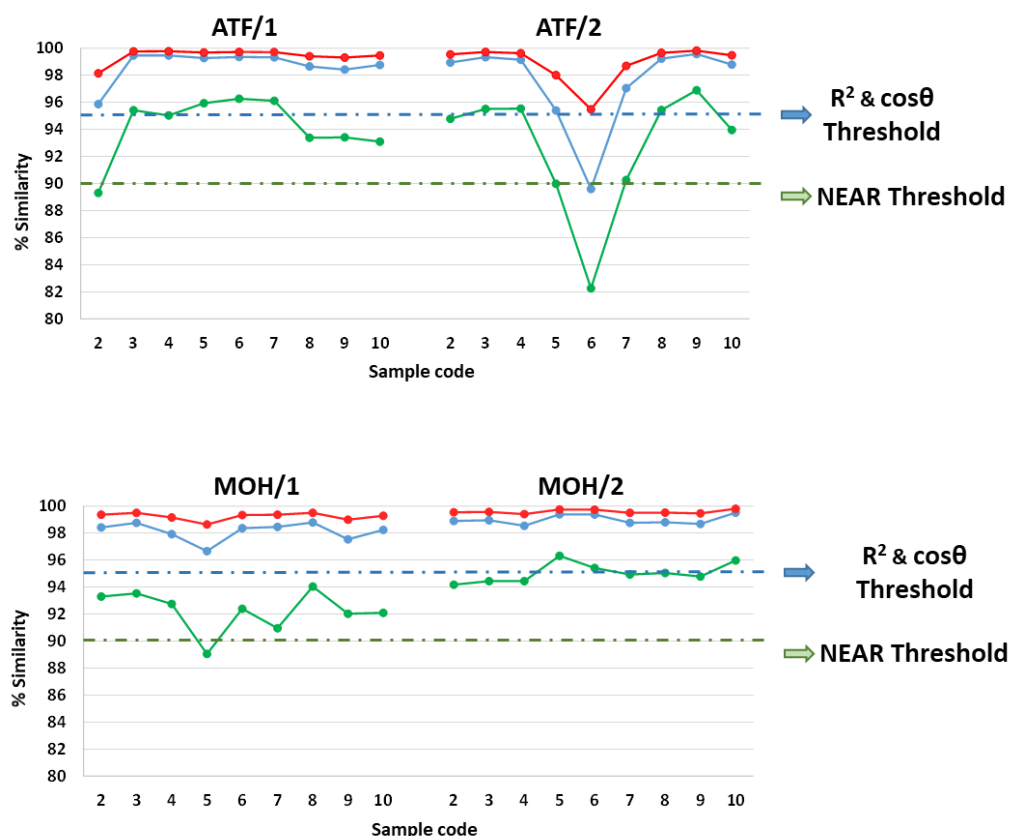


Figure 3. Similarity-based control charts (percentage of similarity vs numerical identifying code of samples) obtained for the eight RMs candidates analysed for homogeneity assessment. The values of the similarity indices, R^2 , $\cos\theta$ and NEAR, are plotted in blue, red and green respectively.

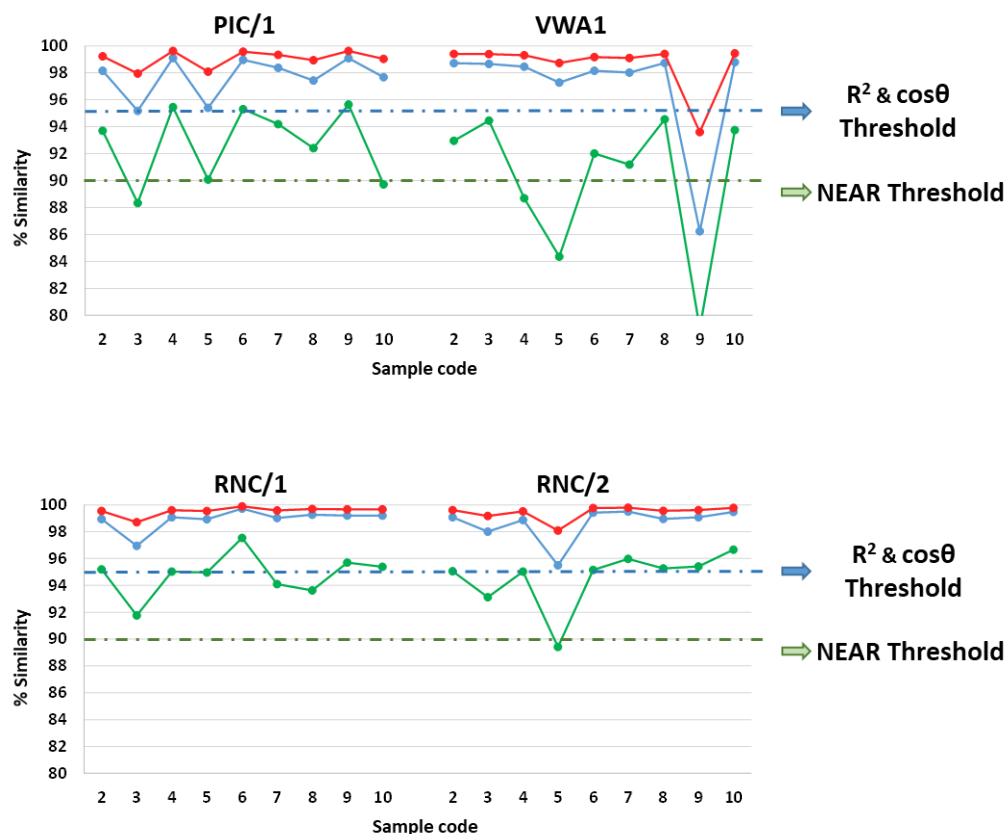


Figure 3. (cont.) Similarity-based control charts (percentage of similarity vs numerical identifying code of samples) obtained for the eight RMs candidates analysed for homogeneity assessment.

The values of the similarity indices, R^2 , $\cos\theta$ and NEAR, are plotted in blue, red and green respectively.

All materials show values of the similarity indices above or very close to the decision threshold. Only two units (bottle 6 of the ATF/2 material and bottle 9 of the VWA/1 material) have values below the threshold for the three indices. Regarding to the ATF/2 material, Figures SM2 a and b reveal that the main differences in the similarity indices for the unit 6 correspond to a considerable decrease in the intensity values of the signals highlighted by the green arrows. The control chart of ATF/2 samples shows a significant deviation that is corrected throughout the analysis. This fact suggests a possible lack of homogeneity among the bottled units 4 and 8 that may have been caused by the bulk container used to package the different units having been refilled several times. However, in order to establish the homogeneity of the material batch, the units within this interval were discarded. Regarding the control chart of the VWA/1 material, no clear trend is observed and sample 9 shows a value in all three indices that is evidently lower than the others. In order to confirm this behaviour, unit 9 was retested and the same results were

obtained again. In fact, it can be seen in Figures SM6 a and b how the signals highlighted by the green arrows are responsible for the lowest values of similarity indices. The lack of homogeneity may also be caused by the reasons mentioned in the case of ATF material. Consequently, all the bottled units from the original batch that were between the tested samples 8 and 10 of this material were rejected in order to declare the material batch as homogeneous. It should be noted that such low values of similarity for this unit were anticipated in the PCA, where the outlier corresponded indeed to unit 9 of this material. As for the remaining samples, the oscillations in the similarity values could be indicative of a lack of mixing throughout the packaging in the different units composing the batch. The zoomed regions of Figures SM1 to SM8 bring forward the signals responsible of the fluctuations observed in the similarity indices, highlighting the differences among different samples for remaining RM candidates.

In contrast, the similarity indices of the stability study were obtained by comparing the representative material fingerprint corresponding to the month of study with the reference fingerprint at time zero. When the homogeneity of the different batches is ensured and all units are stored under the same conditions, it is assumed that the fingerprint obtained for each stability analysis is representative of the remaining units stored. The representative fingerprint is the average fingerprint of the three replicates from the analysed material, while the fingerprint at time zero is the average from the results of the homogeneity study of RM. The agnostized chromatographic fingerprints obtained in the stability study for each one of the different RM candidate are also included in the supplementary material (Figures SM9 - SM16).

Furthermore, for stability study, the samples are considered under control when the values of R^2 , $\cos\theta$ and NEAR are higher than the threshold imposed for this study: 80% for the first two indices and 70% for the third one. Again, the thresholds were defined by the authors. It should be noted that the chromatographic analyses for the homogeneity study are carried out under repeatability conditions, whereas the stability study is carried out under quasi-reproducibility conditions. Therefore, the criterion to be applied in the latter should be less restrictive than in the former. A material is considered out of control when two consecutive stability analyses carried out in different months show values below the established threshold. Figure 4 shows the stability control chart of the eight olive oil materials for different analysis months.

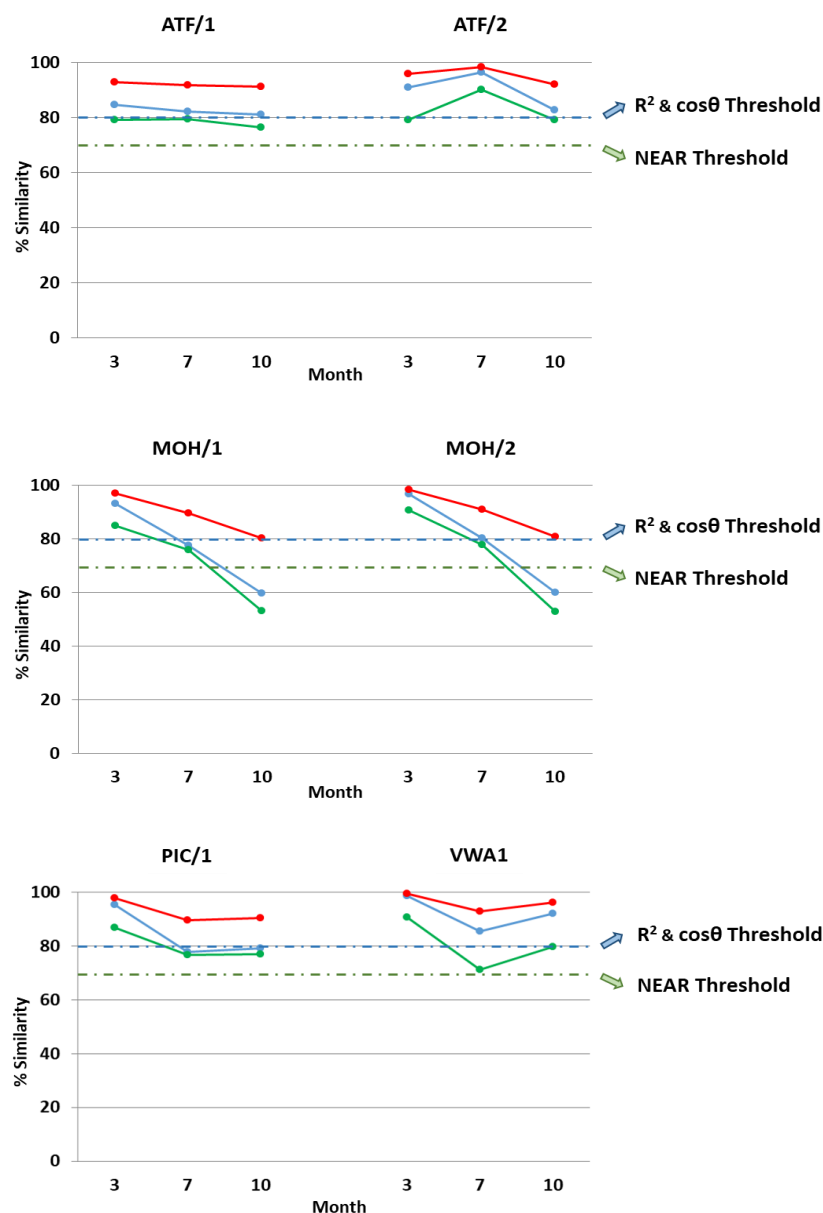


Figure 4. Similarity-based control charts (percentage of similarity vs month analysis) obtained during the stability study for the eight RMs candidates. The values of the similarity indices used in this work, R^2 , $\cos\theta$ and NEAR, are plotted in blue, red and green respectively. Note: the second stability analysis, corresponding with month 6, had to be postponed due to the COVID-19 pandemic.

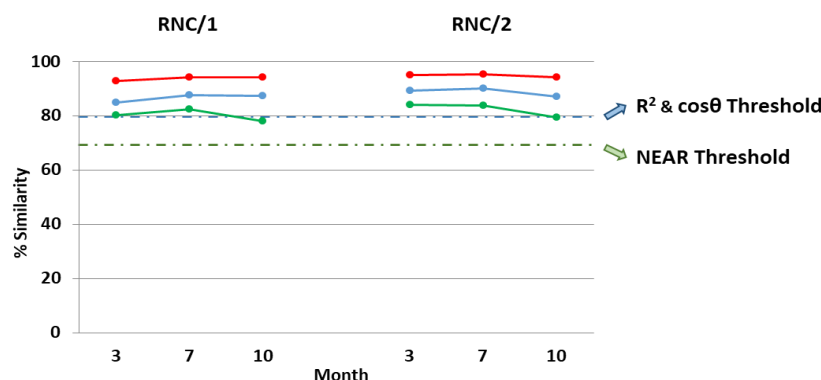


Figure 4. (cont.) Similarity-based control charts (percentage of similarity vs month analysis) obtained during the stability study for the eight RMs candidates. The values of the similarity indices used in this work, R^2 , $\cos\theta$ and NEAR, are plotted in blue, red and green respectively. Note: the second stability analysis, corresponding with month 6, had to be postponed due to the COVID-19 pandemic.

Currently, the stability study is in progress and only three units of each RM candidate reserved for it have been analysed. It can be observed in Figure 4 how, throughout the different months of the study, the similarity indices of all the materials were higher than the established thresholds. This fact suggests that the remaining units that were cold stored during this period of time keep their organoleptic characteristics unchanged and therefore are considered to be under control. On the contrary, the samples coded as MOH/1 and MOH/2 were the only ones with similarity indices (R^2 , $\cos\theta$ and NEAR) lower than the threshold for month 10. In order to discard any possible error in the instrumental measurement, these samples were analysed again, obtaining similar results. The decreasing trend observed in both materials is associated with changes to the VOC fraction as a consequence of some degradation process. In fact, in Figures SM11 and SM12 (a and b) it can be seen how some of the signals have been disappearing since the first month of analysis while new signals have been appearing and increasing their intensity in the successive analyses. Therefore, the decreasing of similarity indices is caused by differences between the reference fingerprint (month 0) and the fingerprint considered as representative of all units of these materials at month 10. For the same reason, these compositional changes could suggest a shift in the certified intensity. If similar results are obtained for both materials in subsequent stability analyses, they should be subjected again to sensory analysis to confirm the change in the certified defect. However, if the tasters determine a change in intensity or the appearance of a new defect, the materials will be declared as out of stability, thus invalidating their certificate.

3.3. Sensory analysis for characterising the reference materials

According to COI/T.20 Doc. No. 15, the assessment of sensory characteristics of an olive oil should be performed by sensory analysis. Therefore, once the eight material batches were declared as homogeneous, they were subjected to the EU panel test method in order to characterise the organoleptic intensity of the attribute/defect to be certified. Note that the characterisation study was carried out between the homogeneity and stability studies. For this study, three tasting sessions were organised, in which 22 recognised tasting panels participated. The sensory analysis was conducted as part of the routine tasting of the panels, so the tasters were not aware that the oil samples were candidates for reference material. In all cases, the results were uploaded to the INTERPANEL platform by the corresponding panel leaders, where the statistical data processing was performed.

The first preliminary session was aimed at determining the most intensely perceived defect by the tasters in each material. In five of the eight materials (coded as ATF/1 & ATF/2, MOH/1 & MOH/2 and VWA/1) the selected attribute corresponds to the original designed defects (see Table SM1). In case of the RNC/1 material (low intensity of rancid), at least 2/3 of the tasters did not observe any defect in the samples. Conversely, 25% of the tasters considered the RNC/2 material to remain without any appreciable defects, while the rest of tasters divided their scores equally between musty and rancid defect. The PIC/1 material is a particular case as it was developed from ripe fruit of 'picual' variety. Although no panel had difficulty in identifying the fruity attribute, some of them perceived the ripeness of the fruit as a difficult defect to classify. These results highlight the problem of olive oils obtained from ripe fruit of this variety, even though being recognised by governmental institution as an EVOO, some of the panels are not able to evaluate it properly.

The second session was planned to carry out a sensory analysis targeted to the defective attribute selected in the first session, in order to obtain a representative value of the organoleptic intensity. The tasters' score was the sum of all the sensory perceptions detected in the tasting session. As a result, the obtained intensity was about twice the expected value with which the materials were designed and thus the result of the second sensory analysis was discarded.

Therefore, the third and definitive session was devoted to establish the organoleptic intensity to be certified. Only 20 of the 22 panels presented results through the INTERPANEL platform. The defects and intensities with which the materials were

designed were correctly identified on this occasion. Thus, for the material coded as RNC/1, at least 2/3 of the tasters perceived the rancid defect, while the remaining third did not notice any defect. In contrast with first tasting session, more than 95% of the tasters identified the rancid defect correctly in the material coded as RNC/2, while the remaining 5% did not identify any defect. Although all tasters identified correctly the fusty defect in the ATF/1 material, the scores given by the taster were distributed between 0 and 4. Again the material coded as PIC/1 similar results were obtained to the first tasting session. All tasters identified correctly the fruity attributed but considered the ripeness of the fruit as a difficult defect to be classified.

After applying the statistical filters described in section 2.5.2, it was found that more than 70% of the tasters were considered suitable for establishing the certified value. The results of the three central value estimators as well as their corresponding deviation estimators are shown in Table 1.

It can be seen how, after detecting and removing inconsistent values for each material, the results of the different central tendency and deviation estimators were similar to each other. Only the materials with a lower consensus –ATF/1 and RNC/1– show poor similarity among the different estimated standard deviations. In both cases, the lowest standard deviation value corresponded to the statistics proposed by Huber (Thompson et al., 2008). With the use of an iterative process, Huber's H15 estimation minimises the influence of outliers on the data set, which considerably reduces the deviation from the final result. Therefore, in this study, Huber's estimators were selected to establish the certified and uncertainty value.

Table 1. Results of the three central estimators calculated for the assignment of the certified value
(Note that the value of PIC/1 refers only to the intensity of the fruity attribute).

Material code	Central value estimators			Deviation estimators		
	Mean	Median	Av(H15)	SIQR	MADe	SD(H15)
ATF/2	4.2	4.2	4.2	0.9	0.7	0.8
ATF/1	2.5	2.6	2.5	2.6	2.0	1.9
MOH/2	4.2	4.2	4.3	1.2	1.0	1.0
MOH/1	2.8	2.8	2.8	1.1	0.9	0.9
PIC/1	3.0	3.0	3.1	0.8	0.7	0.6
RNC/2	2.5	2.4	2.5	0.9	0.9	0.8
RNC/1	1.2	1.3	1.2	1.8	1.8	1.2
VWA/1	2.5	2.4	2.5	0.8	0.6	0.6

The units of each estimation are expressed in cm, according to COI/T.20 Doc. Nº 15, 2018.

The final results of the certification study are shown in Table 2. According to the criteria established in previous sections, five of the eight materials produced (codified as ATF/2, MOH/1, MOH/2, RNC/2 and VWA/1) were certified, assigning each reference value the corresponding uncertainty value. The materials coded as ATF/1 and RNC/1 presented standard deviation values (SD(H15)) higher than the unit, so they did not satisfy the criteria to be declared as CRM. However, the final results showed sufficient consensus among the tasters to use these candidates to RM as consensus guiding materials. The materials with this designation were also suitable for the tasters' training, although their certificate did not provide any uncertainty value. Finally, the material coded as PIC/1 was discarded as a training reference. Although there was sufficient consensus for the fruity attribute, the discrepancies produced by considering the maturity of the fruit as a defect made the material unsuitable for certification.

Table 2. Certified values of the olive oil reference materials obtained in according the characterisation study (*Note that the values of the certified reference materials are in bold; in addition, the 'consensus guiding values' are in italics and in parentheses*).

Material code	Attribute/Defect	Status	Certified value (cm)	$\pm U_{CRM} (k = 2) ^*$
ATF/2	Fusty/muddy	Certified	4.2	0.1
ATF/1	Fusty/muddy	Consensus	(2.5)	–
MOH/2	Musty	Certified	4.3	0.2
MOH/1	Musty	Certified	2.8	0.2
PIC/1	Fruity	Non-Certified	–	–
RNC/2	Rancid	Certified	2.5	0.2
RNC/1	Rancid	Consensus	(1.2)	–
VWA/1	Wine/vinegary	Certified	2.5	0.2

* The U_{CRM} values have been obtained from equation 4 and 5.

4. CONCLUSION

The combined use of sensory and instrumental analysis, developed within the 'Reference and harmonization system of virgin olive oil tasting panels from Andalusia – INTERPANEL', is presented as a rigorous methodology for the certification of the organoleptic intensity of sensory attributes using eight real olive oils batches as reference materials. The instrumental techniques allowed the homogeneity and stability assessment by evaluating the similarity among chromatographic fingerprints of VOC fraction obtained by SPME-GC-MS. In addition, unsupervised pattern recognition techniques (HCA & PCA) have been used to screen the natural grouping of each material produced as a preliminary step to the calculation of similarity indices. This analysis provides prior information from the data set about possible heterogeneity within a particular material batch.

The set of quality statistical filters developed for data processing of the different tasting sessions allowed to obtain statistically reliable results quickly and easily. This set of quality filters is not restricted to olive oil characterisation, since it may also be applied for panel harmonisation. Although the results of the three central value estimators applied on the supra-panel of tasters were similar, the standard deviations obtained when the Huber's H15 statistical estimator was applied were lower than the remaining ones.

Considering these results, Huber's H15 method was selected for the assignment of the certified value and uncertainty assessment.

Complex olive oil materials such as those obtained from ripe fruit of the 'picual' botanical variety reflect the need to use valid references, such as CRM, for the training of tasters, whether obtained by the addition of flavourings or through mixtures of real samples. Even though the latter need to be replaced with each campaign, tasters recognise that the defect has been produced by natural processes. In addition, these materials require an exhaustive study of their sensory characteristics, which directly affects the time required to achieve certification. Through the development of the panel intercommunication system developed within the INTERPANEL project, it is intended to reduce the time required for certification of each olive oil RM. Therefore, both the objectives of the INTERPANEL project and those of this work are in tune with the solutions proposed in a recent report contracted by the EU (European Commission, 2020).

5. ACKNOWLEDGMENTS

The authors would like to thank the European Agricultural Fund for Rural Development (EAFRD) by the European Innovation Partnership for Agricultural Productivity and Sustainability (EIP-AGRI), and the "Consejería de Agricultura, Ganadería, Pesca y Desarrollo Sostenible" (Department of Agriculture, Livestock, Fisheries and Sustainable Development, Regional Govern of Andalusia, Spain) for the funding of the Operational Group project 'INTERPANEL' (GOP3-JA-16-0012).

SUPPLEMENTARY MATERIAL

Table SM1. Description of the starting materials used in the production of the olive oil batches candidates to reference materials (*Note that ATF, MOH, PIC, RNC and VWA are the encoded names and they are not related to the material composition*).

Material code	Blend of olive oils	Description
ATF/2	6.0% concentrate #1 94.0% base #1	Freshly harvested olive fruits of 'picual' botanical variety were subjected to a fermentation process, stacking the fruits on the earthy soil to generate the fusty/muddy defect. Then the olives are milled and a fusty/muddy defective olive oil is produced (concentrate #1). This was blended with EVOO from ripe olive fruits of the 'picual' variety (base #1).
ATF/1	1.6% concentrate #1 98.4% base #1	Olive oil similar to the ATF/2 material, but resulting from a higher dilution of concentrate #1 into base #1.
MOH/2	2.0% concentrate #2 98.0% base #1	Freshly harvested olive fruits of botanical 'picual' variety were stored in a humid environment for long enough to generate the musty defect. Then the olives are milled and a musty defective olive oil is produced (concentrate #2). This was blended with EVOO from ripe olive fruits of the 'picual' variety (base #1).
MOH/1	0.45% concentrate #2 99.55% base #1	Olive oil similar to the MOH/2, but resulting from a higher dilution of concentrate #2 into base #1.
PIC/1	100% base #1	EVOO from ripe olive fruits of the 'picual' botanical variety (base #1).
RNC/2	20.0% concentrate #3 80.0% base #2	EVOO from olive fruits of 'arbequina' botanical variety (base #2), packed in glass containers, were left outdoors in a sunny area in order to develop an oxidative degradation process to generate the rancid defect (concentrate #3). This was blended with original non-degraded EVOO (base #2).
RNC/1	5.0% concentrate #3 95.0% base #2	Olive oil similar to the RNC/2, but resulting from a higher dilution of concentrate #3 into base #2.
VWA/1	100% concentrate #4	VOO from olive fruits of 'picual' botanical variety with wine/vinegary defect, provided by PDO 'Sierra de Segura' (concentrate #4). No further information about the production process of the defect in question is available.

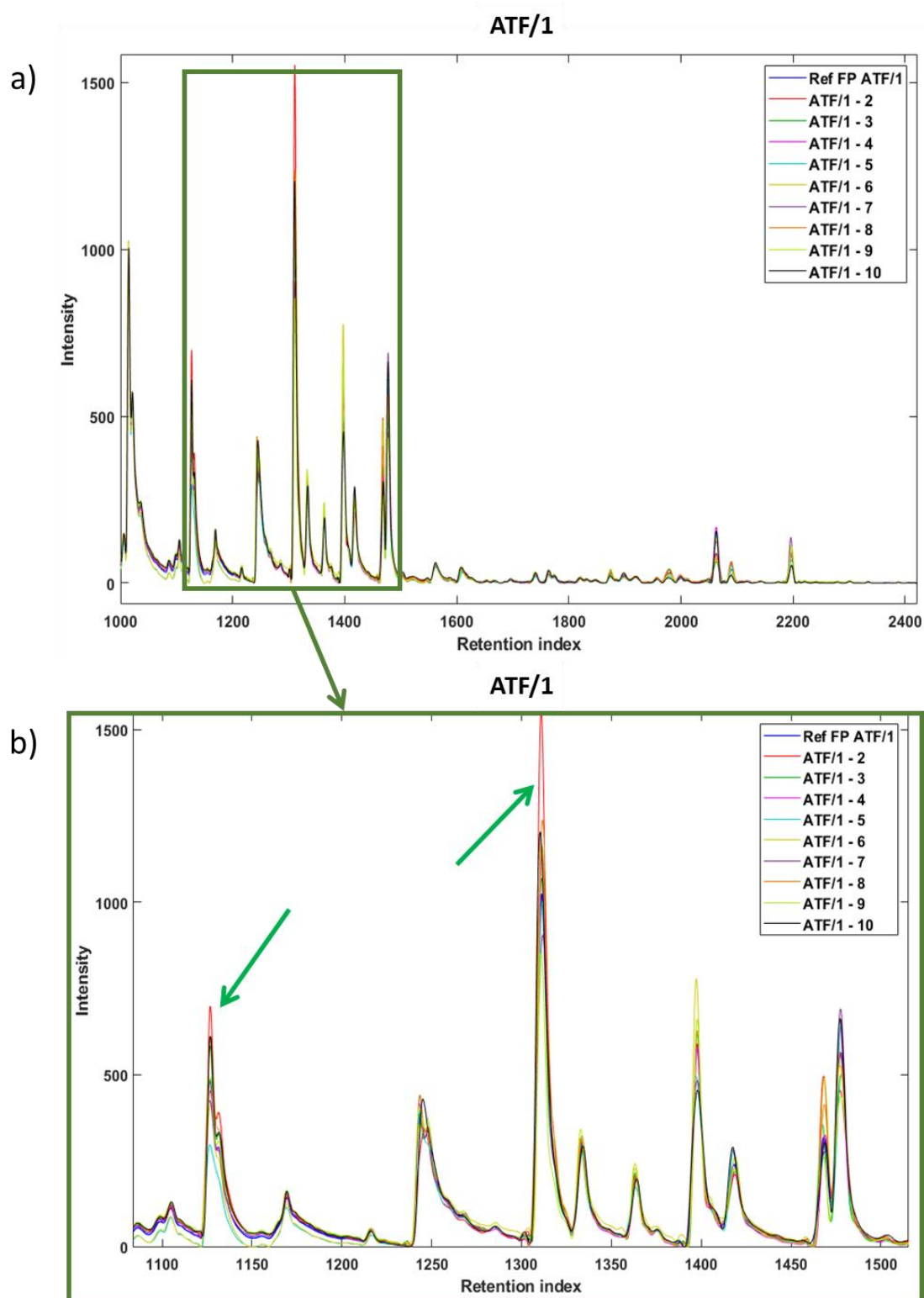


Figure SM1. a) Agnostized chromatographic fingerprints (reference and individual units) of the RM candidate encoded as ATF/1 (low intensity of fusty/muddy defect) used to obtain the similarity indices in the homogeneity study. b) Zoomed region where the larger differences appear in the agnostized chromatographic fingerprints (variables highlighted with green arrows).

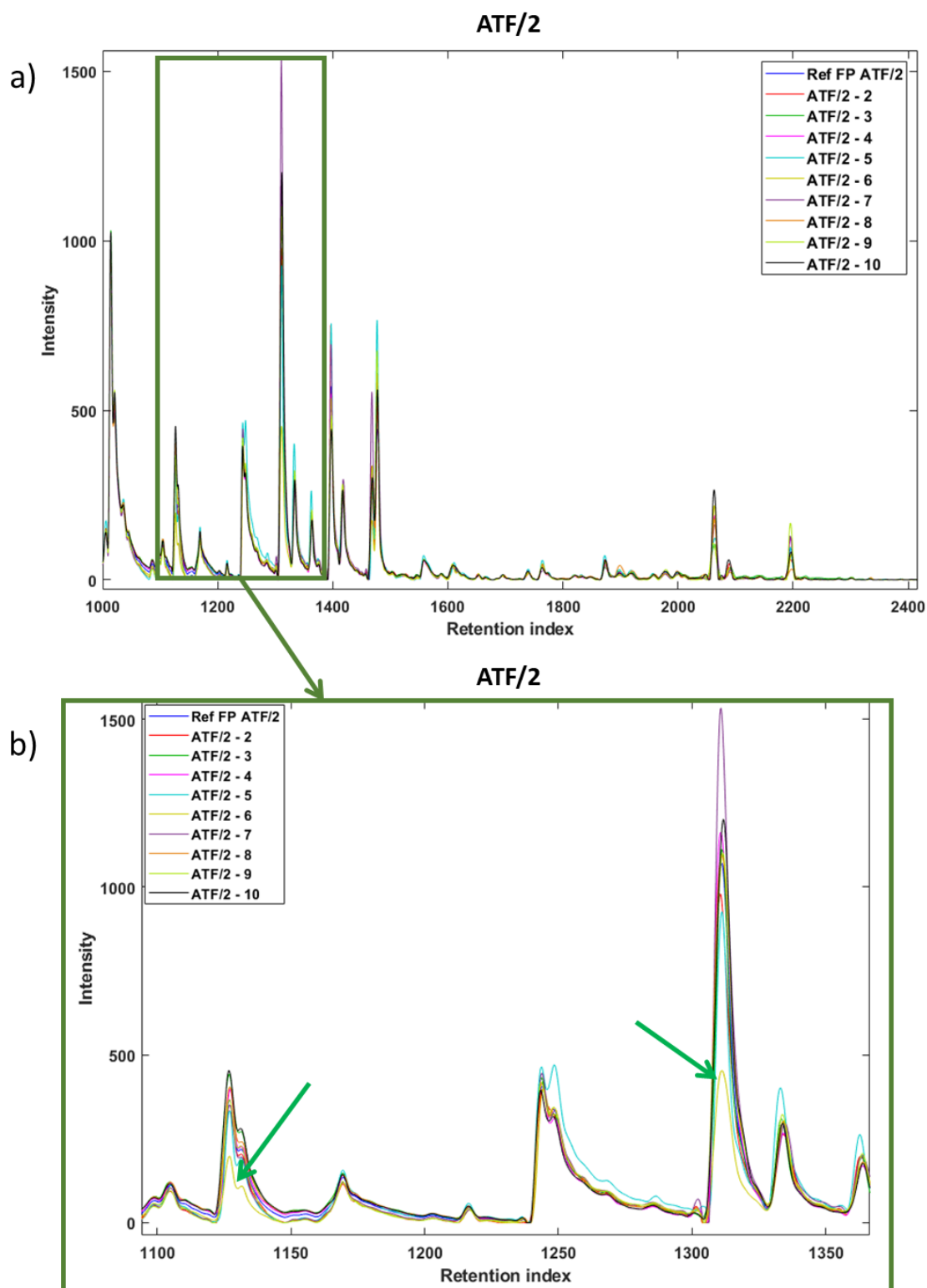


Figure SM2. a) Agnostized chromatographic fingerprints (reference and individual units) of the RM candidate encoded as ATF/2 (medium intensity of fusty/muddy defect) used to obtain the similarity indices in the homogeneity study. b) Zoomed region where the larger differences appear in the agnostized chromatographic fingerprints (variables highlighted with green arrows).

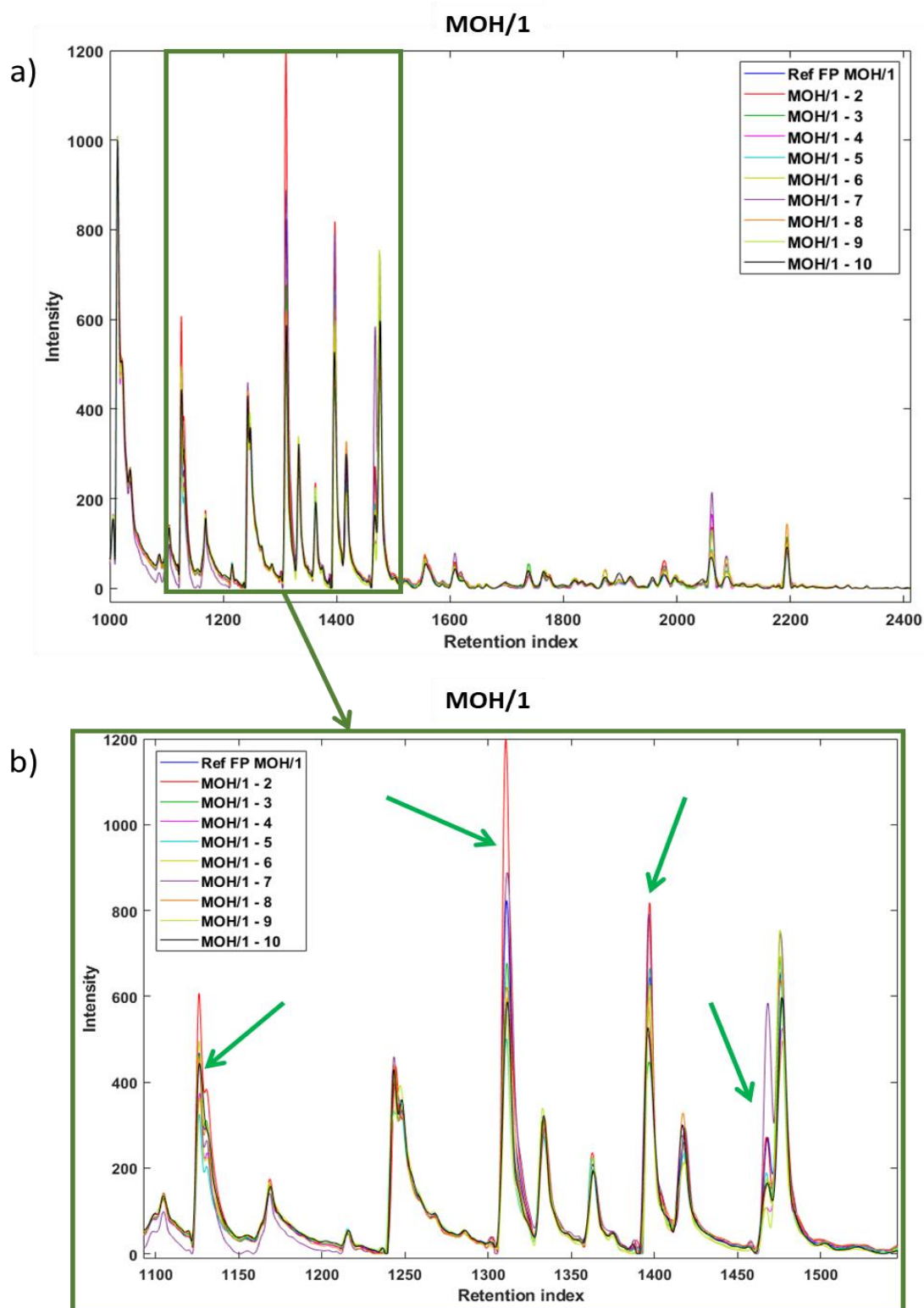


Figure SM3. a) Agnostized chromatographic fingerprints (reference and individual units) of the RM candidate encoded as MOH/1 (low intensity of musty defect) used to obtain the similarity indices in the homogeneity study. b) Zoomed region where the larger differences appear in the agnostized chromatographic fingerprints (variables highlighted with green arrows).

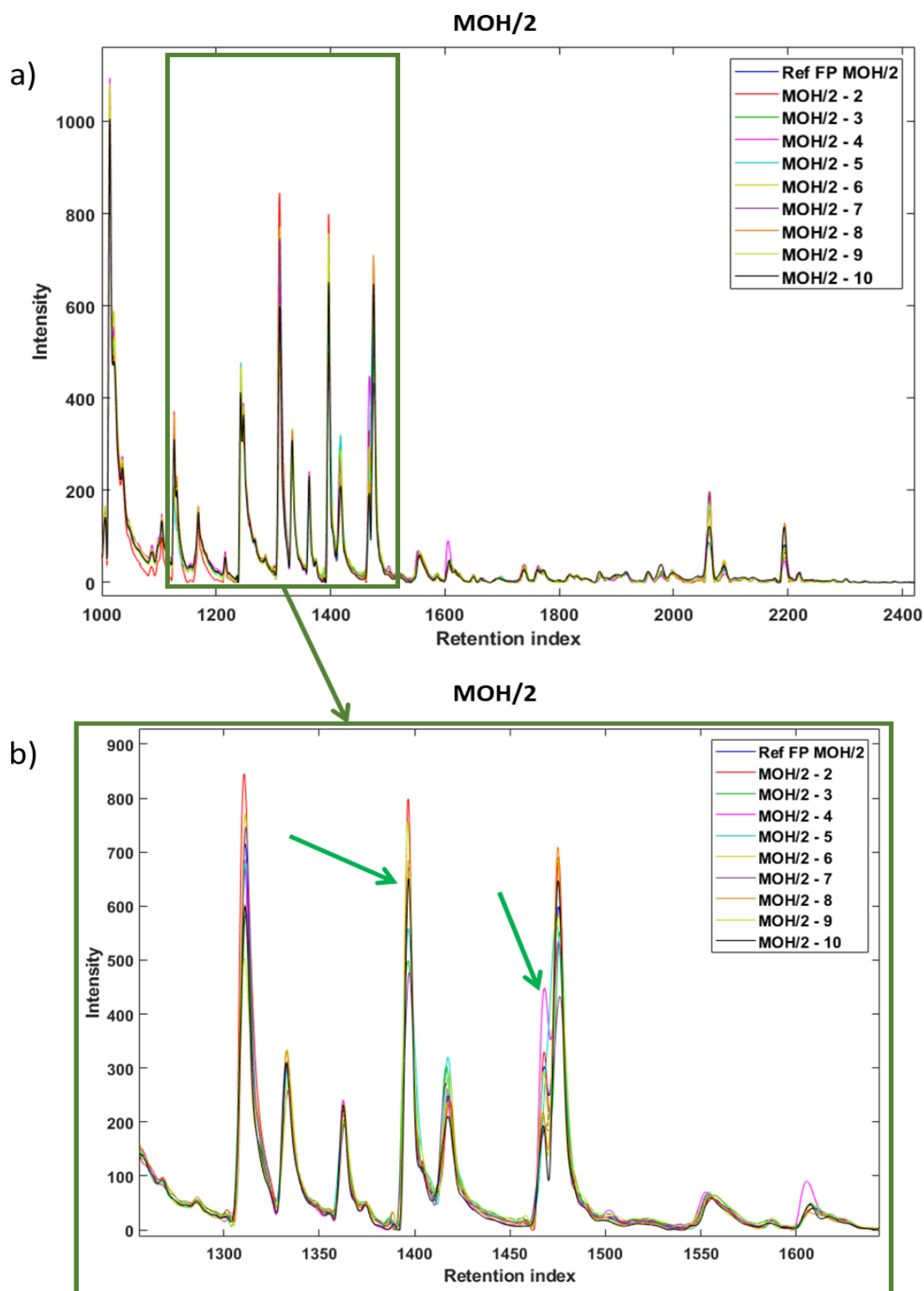


Figure SM4. a) Agnostized chromatographic fingerprints (reference and individual units) of the RM candidate encoded as MOH/2 (medium intensity of musty defect) used to obtain the similarity indices in the homogeneity study. b) Zoomed region where the larger differences appear in the agnostized chromatographic fingerprints (variables highlighted with green arrows).

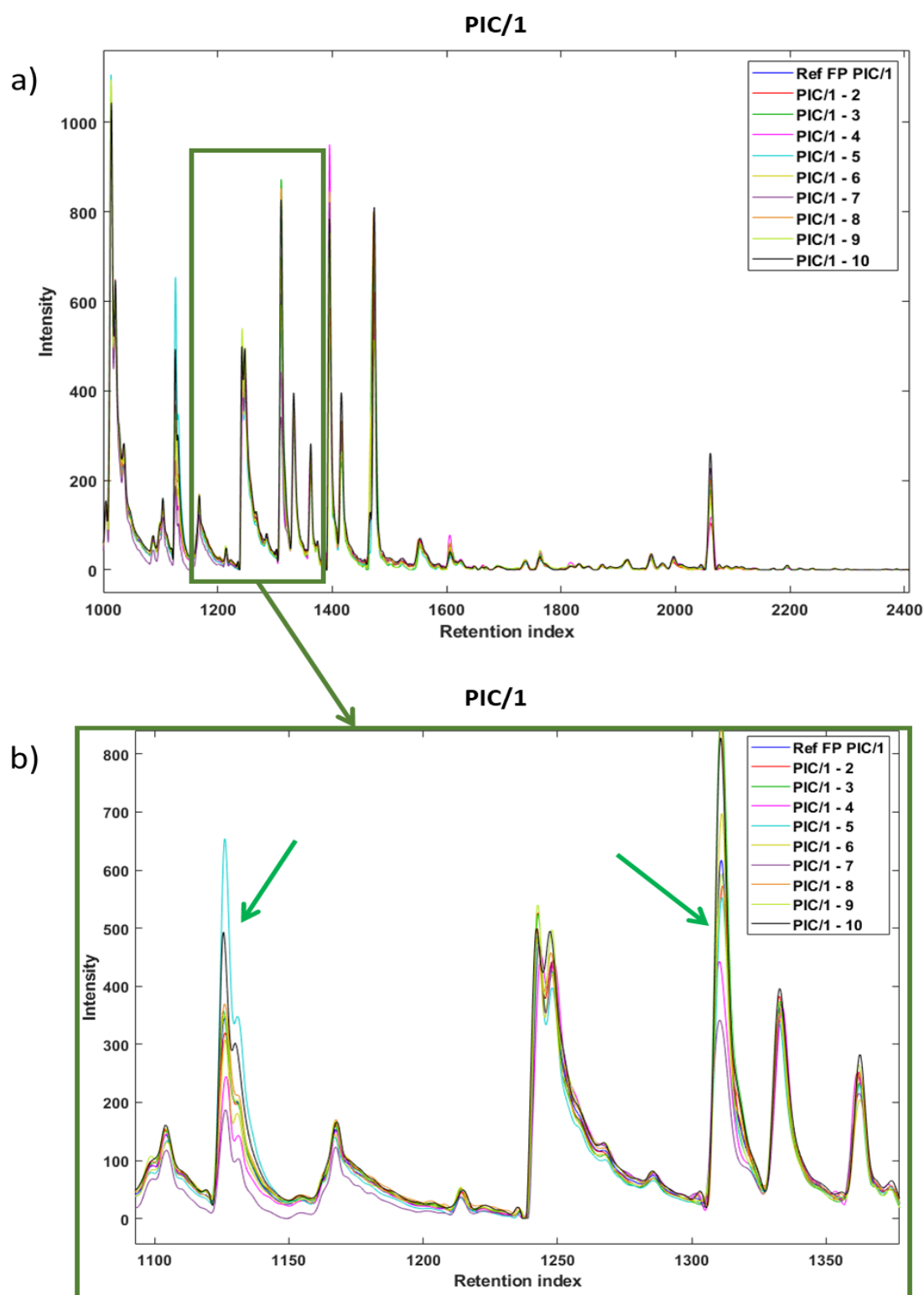


Figure SM5. a) Agnostized chromatographic fingerprints (reference and individual units) of the RM candidate encoded as PIC/1 (fruity attribute) used to obtain the similarity indices in the homogeneity study. b) Zoomed region where the larger differences appear in the agnostized chromatographic fingerprints (variables highlighted with green arrows).

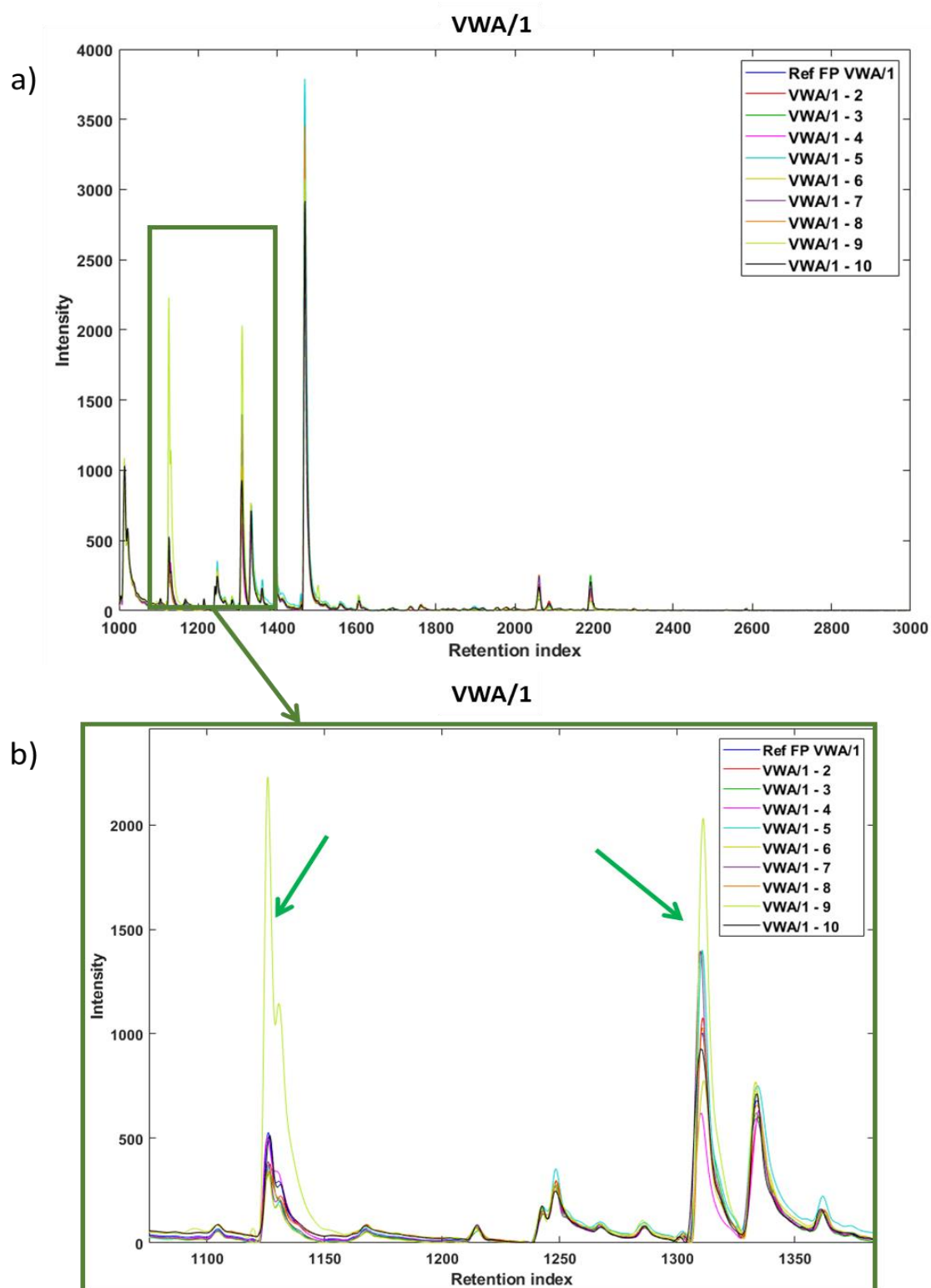


Figure SM6. a) Agnostized chromatographic fingerprints (reference and individual units) of the RM candidate encoded as VWA/1 (wine/vinegary defect) used to obtain the similarity indices in the homogeneity study. b) Zoomed region where the larger differences appear in the agnostized chromatographic fingerprints (variables highlighted with green arrows).

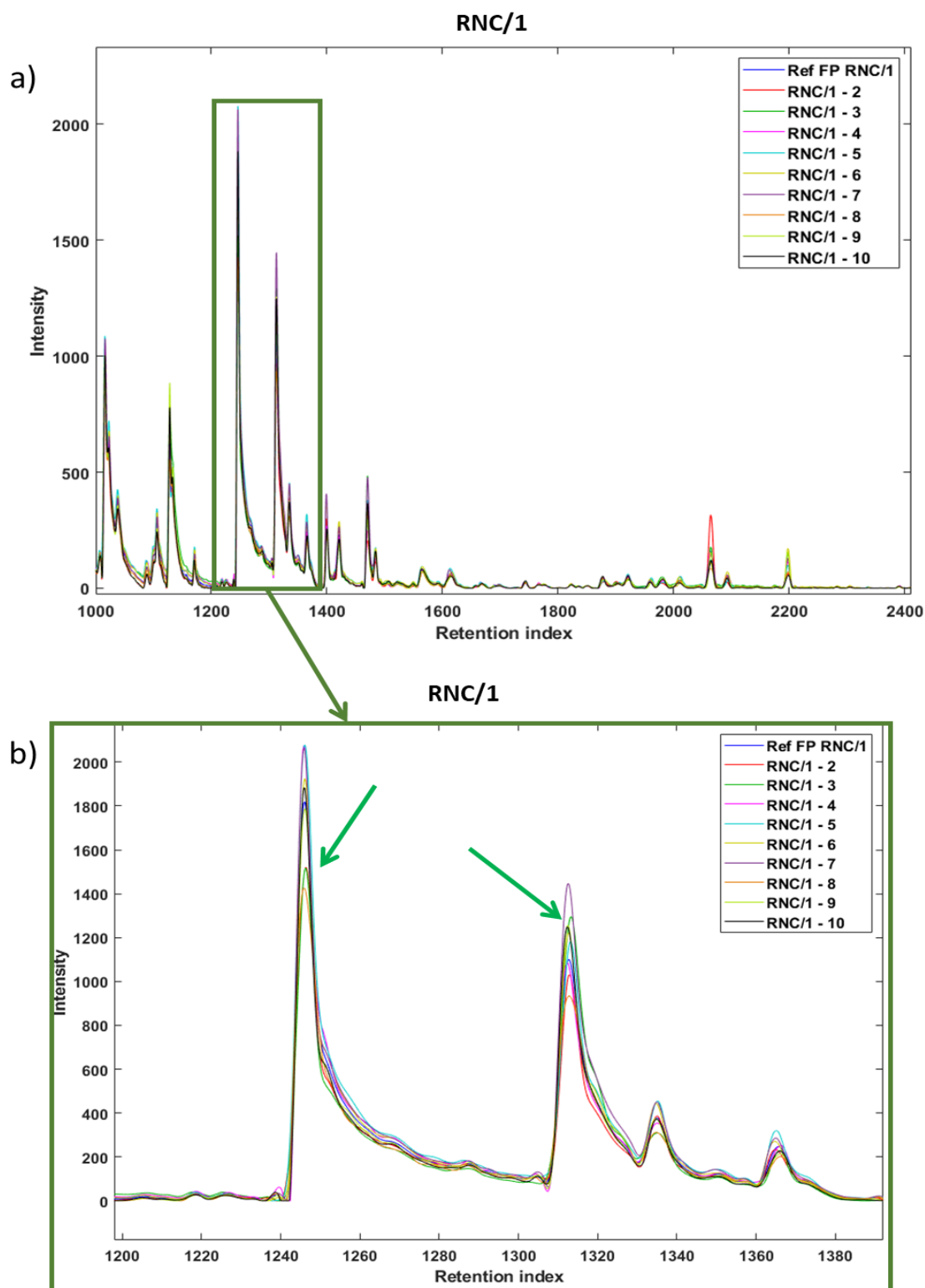


Figure SM7. a) Agnostized chromatographic fingerprints (reference and individual units) of the RM candidate encoded as RNC/1 (low intensity of rancid defect) used to obtain the similarity indices in the homogeneity study. b) Zoomed region where the larger differences appear in the agnostized chromatographic fingerprints (variables highlighted with green arrows).

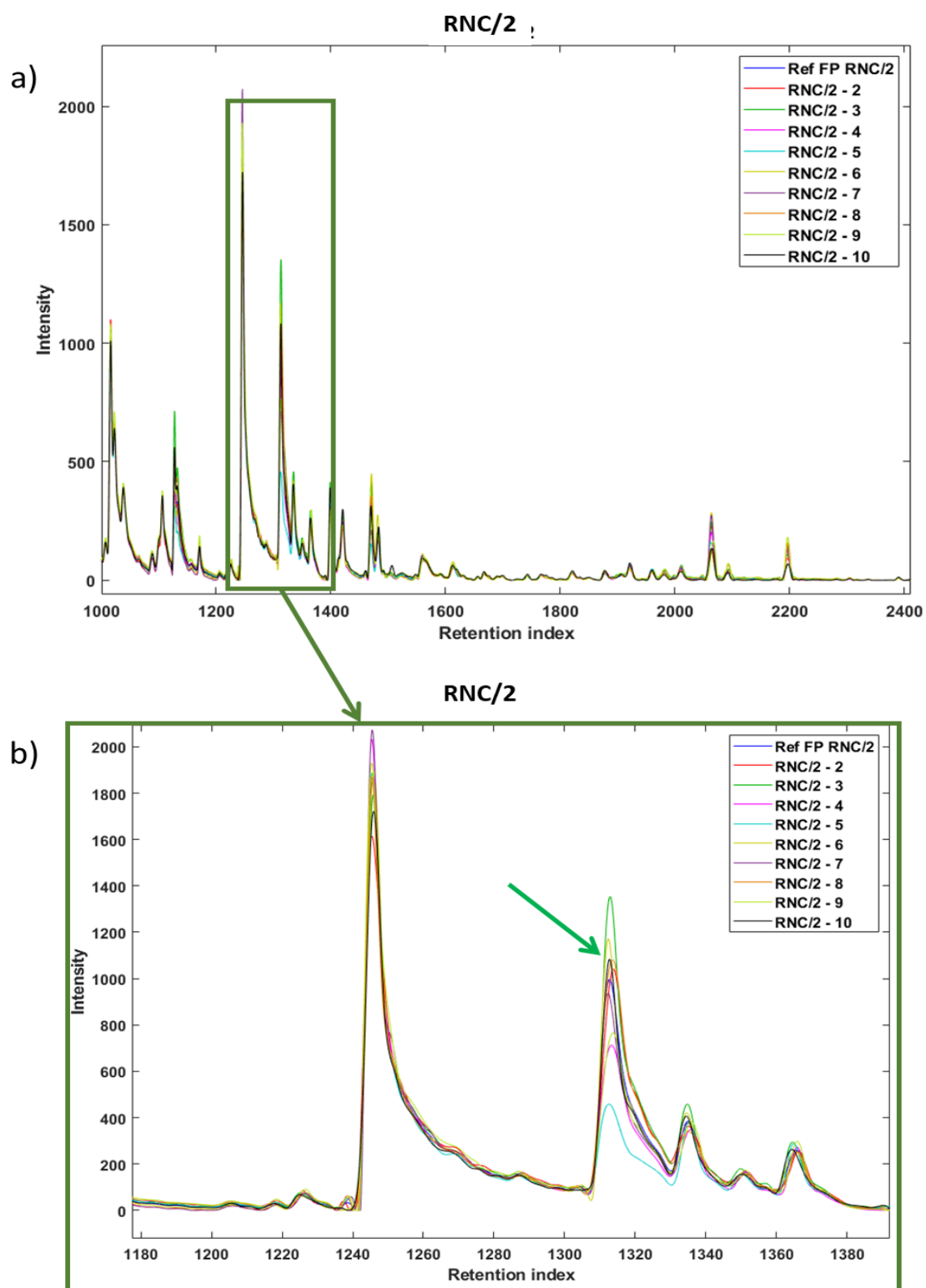


Figure SM8. a) Agnostized chromatographic fingerprints (reference and individual units) of the RM candidate encoded as RNC/2 (medium intensity of rancid defect) used to obtain the similarity indices in the homogeneity study. b) Zoomed the region where the larger differences appear in the agnostized chromatographic fingerprints (variables highlighted with green arrows).

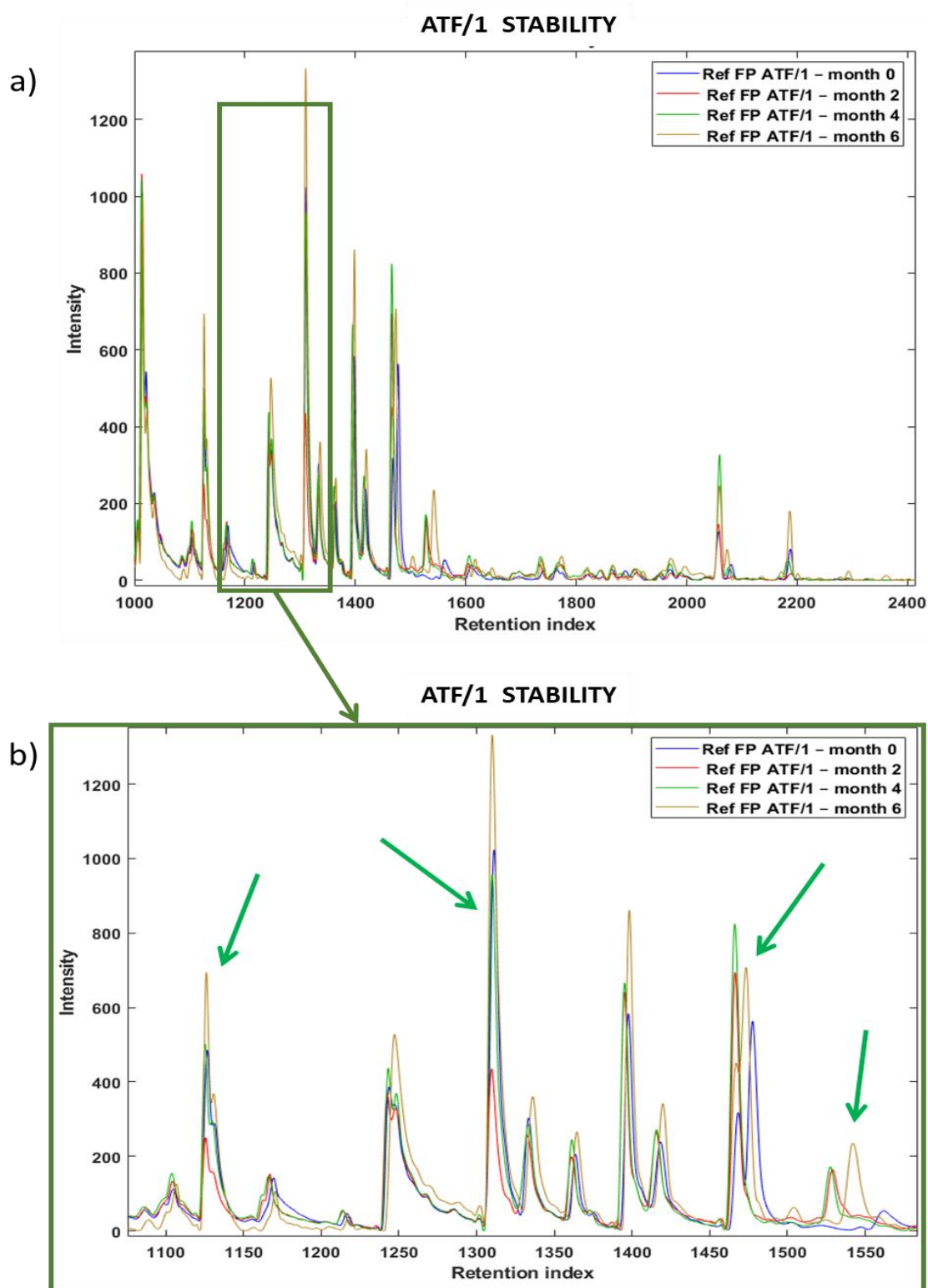


Figure SM9. a) Agnostized chromatographic fingerprints of the RM candidate encoded as ATF/1 (low intensity of fusty/muddy defect) used to obtain the similarity indices in the stability study. b) Zoomed the region where the larger differences appear in the agnostized chromatographic fingerprints (variables highlighted with green arrows) during the different months.

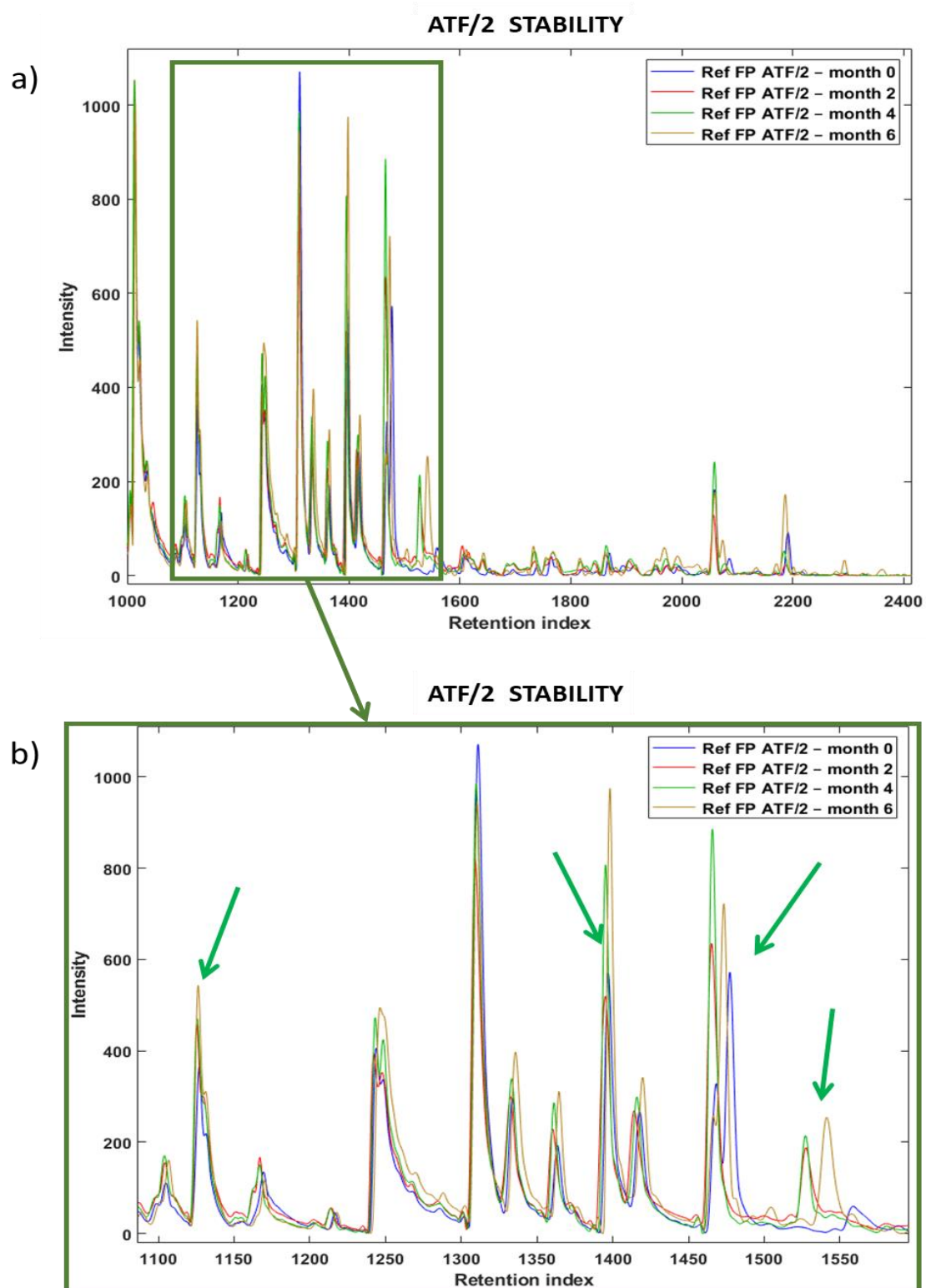


Figure SM10. a) Agnostized chromatographic fingerprints of the RM candidate encoded as ATF/2 (medium intensity of fusty/muddy defect) used to obtain the similarity indices in the stability study.

b) Zoomed the region where the larger differences appear in the agnostized chromatographic fingerprints (variables highlighted with green arrows) during the different months.

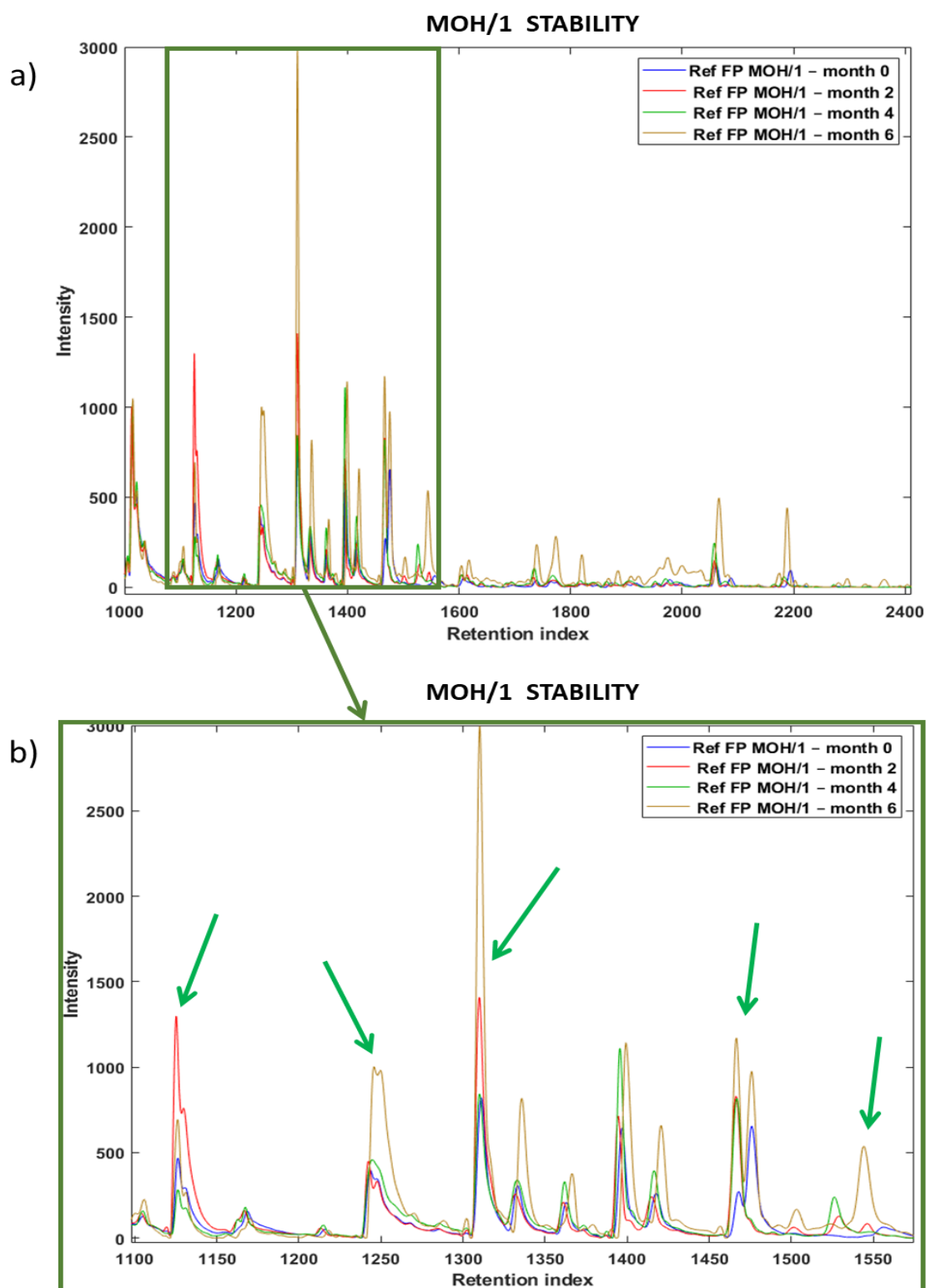


Figure SM11. a) Agnostized chromatographic fingerprints of the RM candidate encoded as MOH/1 (low intensity of musty defect) used to obtain the similarity indices in the stability study. b) Zoomed the region where the larger differences appear in the agnostized chromatographic fingerprints (variables highlighted with green arrows) during the different months.

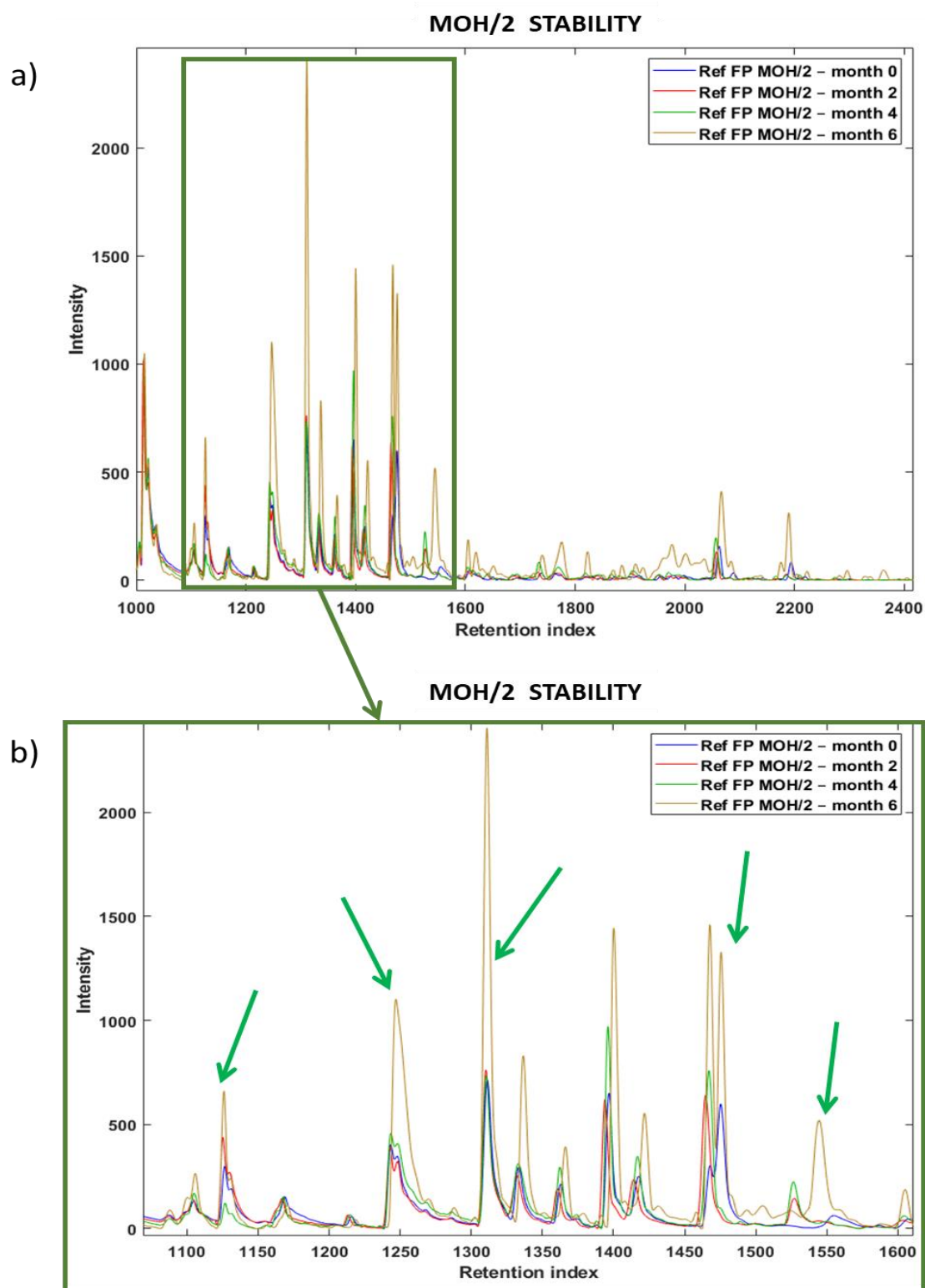


Figure SM12. a) Agnostized chromatographic fingerprints of the RM candidate encoded as MOH/2 (medium intensity of musty defect) used to obtain the similarity indices in the stability study. b) Zoomed the region where the larger differences appear in the agnostized chromatographic fingerprints (variables highlighted with green arrows) during the different months.

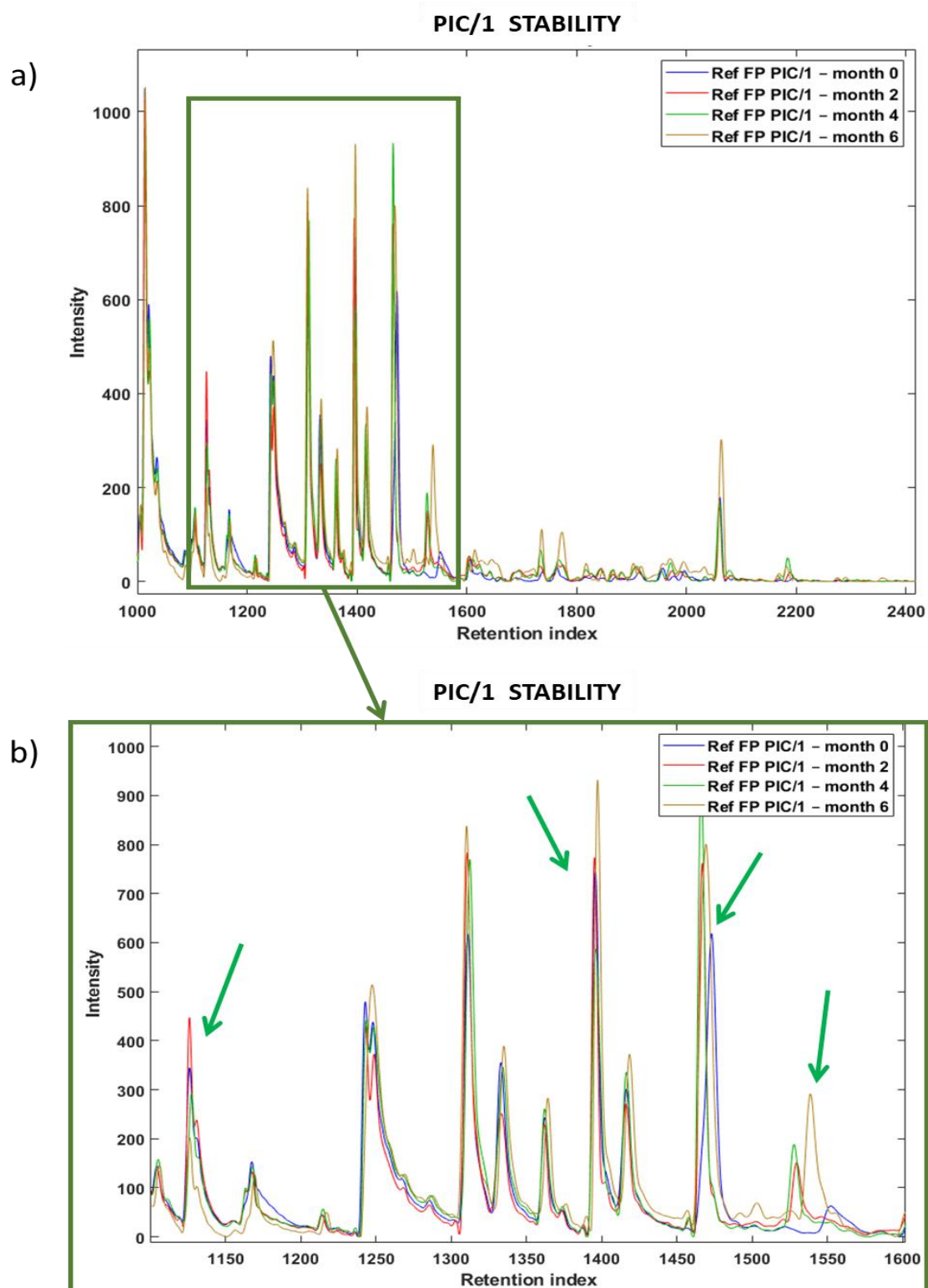


Figure SM13. a) Agnostized chromatographic fingerprints of the RM candidate encoded as PIC/1 (fruity attribute) used to obtain the similarity indices in the stability study. b) Zoomed the region where the larger differences appear in the agnostized chromatographic fingerprints (variables highlighted with green arrows) during the different months.

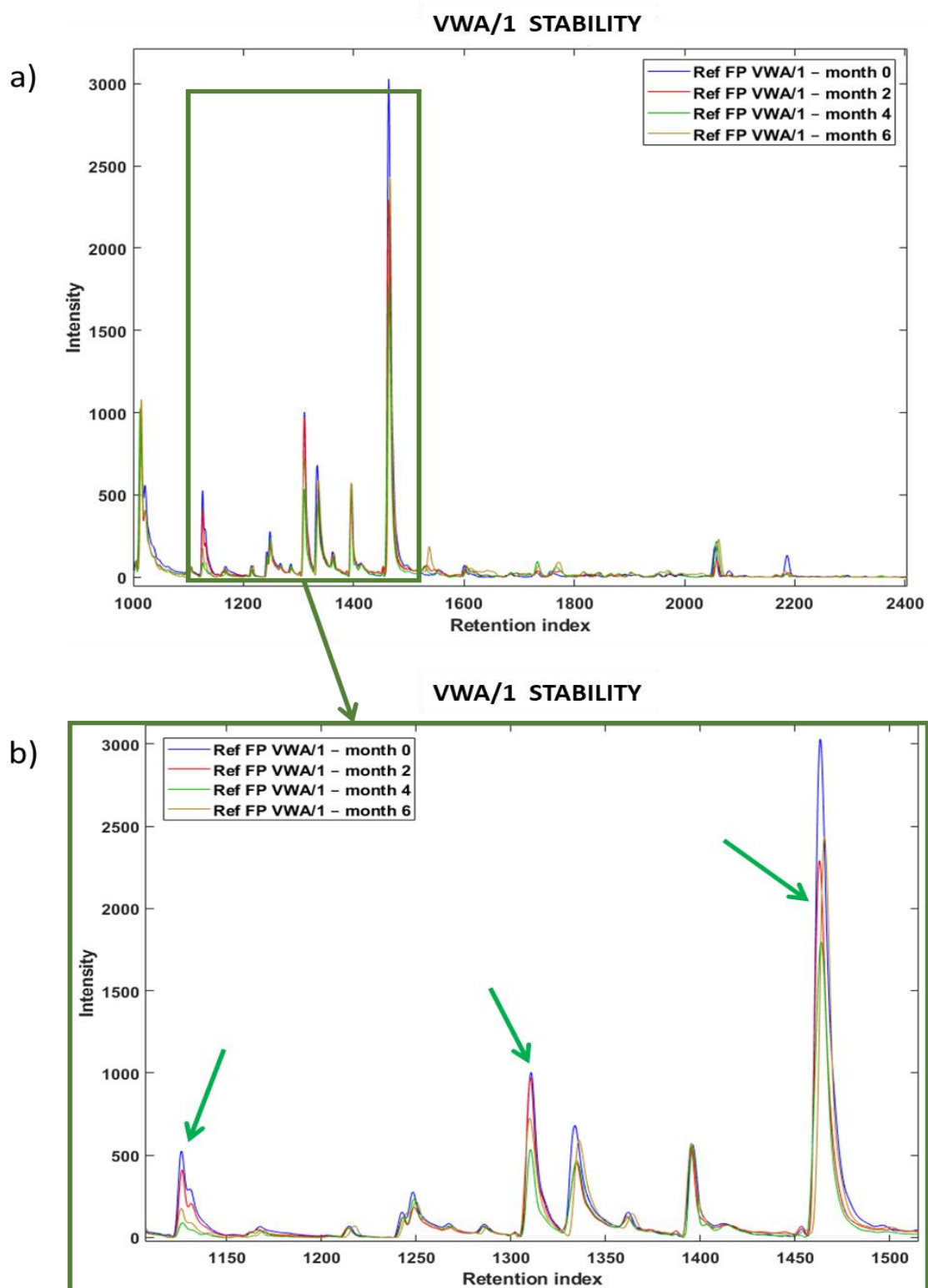


Figure SM14. a) Agnostized chromatographic fingerprints of the RM candidate encoded as VWA/1 (wine/vinegary defect) used to obtain the similarity indices in the stability study. b) Zoomed the region where the larger differences appear in the agnostized chromatographic fingerprints (variables highlighted with green arrows) during the different months.

REFERENCES

- Andrić, F., & Héberger, K. (2017). How to compare separation selectivity of high-performance liquid chromatographic columns properly? *Journal of Chromatography A*, **1488**, 45–56.
<https://doi.org/10.1016/j.chroma.2017.01.066>
- Aparicio, R., Conte, L.S., & Fiebig, H.J. (2013). Olive oil authentication. In R. Aparicio, & J. Harwood (Eds), *Handbook of Olive Oil: Analysis and Properties* (pp. 590-653), (2nd ed.). New York: Springer.
https://doi.org/10.1007/978-1-4614-7777-8_16
- Aparicio-Ruiz, R., Morales, M.T., & Aparicio, R. (2019). Does authenticity of virgin olive oil sensory quality require input from chemistry? *European Journal of Lipid Science and Technology*, **121**, 1900202.
<https://doi.org/10.1002/ejlt.201900202>
- Aparicio Ruiz, R., Barbieri, S., Gallina Toschi, T., & García González, D.L. (2020). Formulations of rancid and winey-vinegary artificial olfactory reference materials (AORMs) for virgin olive oil sensory evaluation. *Foods*, **9**, 1870.
<https://doi.org/10.3390/foods9121870>
- Barbieri, S., Bubola, K. B., Bendini, A., Bucar Miklavcic, M., Lacoste, F., Tibet, U., ... Gallina Toschi, T. (2020). Alignment and proficiency of virgin olive oil sensory panels: the OLEUM approach, *Foods*, **9**, 355.
<https://doi.org/10.3390/foods9030355>
- Barjol, J.L. (2013). Introduction. In R. Aparicio, & J. Harwood (Eds.), *Handbook of Olive Oil: Analysis and Properties* (pp. 1–15), (2nd ed.). New York: Springer.
https://doi.org/10.1007/978-1-4614-7777-8_1
- Brereton, R.G. (2009), Exploratory data analysis, in R.G. Brereton (ed), *Chemometrics for Pattern Recognition* (pp. 47–105). Chichester: John Wiley & Sons.
- Circi, S., Capitani, D., Randazzo, A., Ingallina, C., Mannina, L., & Sobolev, A.P. (2017). Panel test and chemical analyses of commercial olive oils: a comparative study. *Chemical and Biological Technologies in Agriculture*, **4**, 1-10.
<https://doi.org/10.1186/s40538-017-0101-0>

COI/T.20/Doc. No 4 / Rev.1 (2007). Sensory analysis of olive oil. General basic vocabulary. Madrid: International Olive Council (IOC).

COI/T.20/Doc. No 14 / Rev. 7 (2021). Sensory analysis of olive oil. Guide for the selection, training and quality control of virgin olive oil taster-qualifications of taster, panel leaders and trainers. Madrid: International Olive Council (IOC).

COI/T.20/Doc. No 15 / Rev.10 (2018). Sensory analysis of olive oil. Method for the organoleptic assessment of virgin olive oil. Madrid: International Olive Council (IOC).

COI/T.20 / Doc. No 22 (2005). Organoleptic assessment of extra virgin olive oil applying to use a designation of origin. Madrid: International Olive Council (IOC).

Cuadros Rodríguez, L., Ortega Gavilán, F., Martín Torres, S., Medina Rodríguez, S., Jiménez Carvelo, A. M., González Casado, A., & Bagur González, M. G. (2021). Standardization of chromatographic signals – Part I: Towards obtaining instrument-agnostic fingerprints in gas chromatography. *Journal of Chromatography A*, 1641, 461983.

<https://doi.org/10.1016/j.chroma.2021.461983>

d'Acampora Zellner, B., Bicchi, C., Dugo, P., Rubiolo, P., Dugo, G., & Mondello, L. (2008). Linear retention indices in gas chromatographic analysis. *Flavour and Fragrance Journal*, **23**, 297–314.

<https://doi:10.1002/ffj.1887>

European Commission (2016). Regulation (EEC) No. 2568/91 on the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis (consolidated version 01991R2568). Official Journal of the European Union, L326, 1–128.

European Commission (2020). *Study on the implementation of conformity checks in the olive oil sector throughout the European Union*. Publication office of European Union. Retrieved from: <https://op.europa.eu/en/publication-detail/-/publication/606555af-46ff-11ea-b81b-01aa75ed71a1#>. Accessed April 27, 2020.

ISO Guide 30 (2015). Reference Materials – Selected Terms and Definitions, Geneva: International Organization for Standardization.

ISO Guide 31 (2015). Reference materials — Contents of certificates, labels and accompanying documentation, Geneva: International Organization for Standardization.

- ISO Guide 35 (2017). Reference Materials – Guidance for characterisation and assessment of homogeneity and stability, Geneva: International Organization for Standardization.
- Mariotti, M. (2014). Virgin olive oil: definition and standards. In C.Peri (Ed.), *The Extra-Virgin Olive Oil Handbook* (pp. 11-20). Chichester: Wiley.
- Morales, M.T., Aparicio Ruiz, R., & Aparicio, R. (2013). Chromatographic methodologies: compounds for olive oil odor issues. In R. Aparicio, & J. Harwood (Eds.), *Handbook of Olive Oil: Analysis and Properties*, (pp. 261-309), (2nd ed.). New York: Springer.
https://doi.org/10.1007/978-1-4614-7777-8_8
- OLEUM Project (2016), *Protocol for the standardization of a procedure in terms of homogeneity and representativeness of the samples to be analysed in T4.2 and T4.3*. Retrieved from: http://www.oleumproject.eu/images/uploads/files/D4.1_def.pdf. Accessed April 27, 2020.
- Ortega Gavilan, F., Valverde Som, L., Rodríguez García, F.P., Cuadros Rodríguez, L., & Bagur González, M.G. (2020). Homogeneity assessment of reference materials for sensory analysis of liquid foodstuffs. The virgin olive oil as case study. *Food chemistry*, **322**, 126743.
<https://doi.org/10.1016/j.foodchem.2020.126743>
- Pérez Robles, R., Navas, N., Medina Rodríguez, S., & Cuadros Rodríguez, L. (2017), Method for the comparison of complex matrix assisted laser desorption ionization-time of flight mass spectra. Stability of therapeutical monoclonal antibodies. *Chemometrics and Intelligent Laboratory Systems*, **170**, 58–67.
<https://doi.org/10.1016/j.chemolab.2017.09.008>
- Rason, J., Laguet, A., Berge, P., Dufour, E., & Lebecque, A. (2007). Investigation of the physicochemical and sensory homogeneity of traditional French dry sausages. *Meat Science*, **7**, 359–370.
<https://doi.org/10.1016/j.meatsci.2006.06.004>
- Thompson, M., & Lowthian, P.J. (2011) *Notes on Statistics and Data Quality for Analytical Chemists*, (pp. 135–139). Imperial College Press, London.

Valverde Som, L., Ruiz Samblás, C., Rodríguez García, F. P., & Cuadros Rodríguez, L. (2018a). Multivariate approaches for stability control of the olive oil reference materials for sensory analysis – Part I: Framework and fundamentals. *Journal of the Science of Food and Agriculture*, **98**, 4237–4244.

<https://doi.org/10.1002/jsfa.8948>

Valverde Som, L., Ruiz Samblás, C., Rodríguez García, F.P., & Cuadros Rodríguez, L. (2018b). Multivariate approaches for stability control of the olive oil reference materials for sensory analysis – Part II: Applications. *Journal of the Science of Food and Agriculture*, **98**, 4245–4252.

<https://doi.org/10.1002/jsfa.8946>

Willett, W. C., Scacks, F., Trichopoulou, A., Drescher, G., Ferro Luzzi, A., Helsing, E., & Trichopoulos, D. (1995) Mediterranean diet pyramid: a cultural modern for healthy eating. *The American Journal of Clinical Nutrition*, **61**, 1402S–1406S.

<https://doi.org/10.1093/ajcn/61.6.1402S>

Capítulo V

**Detección de mezclas de aceite de oliva a partir de
los cambios en el perfil de colorantes**

Innovative and thorough practice to certify reference materials for sensory defects of olive oil

Journal of Chromatography A

The chromatographic similarity profile – an innovative methodology to detect fraudulent blends of virgin olive oils --Manuscript Draft--

Manuscript Number:	JCA-22-35
Article Type:	Full length article
Keywords:	Direct analysis of olive oil; chromatography-agnostic methodology; similarity profile; olive oil blend detection.
Corresponding Author:	Fidel Ortega-Gavilán University of Granada Granada, SPAIN
First Author:	Fidel Ortega-Gavilán
Order of Authors:	Fidel Ortega-Gavilán
	Ana María Jiménez Carvelo, PhD
	Luis Cuadros Rodríguez, PhD
	María Gracia Bagur González, PhD
Abstract:	Extra virgin olive oil is a potentially vulnerable foodstuff that can be mixed with other vegetal edible oils including poorer quality olive oils in order to obtain illicit profits. These unauthorized operations may take place at any stage of the production process and radically affect the chemical composition. In this paper, the analysis of different virgin olive oil samples before and after blending with other lower-grade olive oils in different proportions were performed. The direct analysis of the samples by (NP)HPLC-DAD in a wavelength range between 190 and 700 nm allowed the simultaneous analysis of several compound families responsible of the colour including chlorophylls, pheophytins, carotenes and tocopherols, the first three responsible for the olive oil colour. Unsupervised pattern recognition techniques applied on the chromatography-agnostic fingerprints of unblended virgin olive oil samples clearly showed the occurrence of groupings according to the sample hue (green and yellow). Two strategies, based on revealing changes in the spectrum-chromatographic fingerprints, are tested in order to detect the occurrence of such fraudulent blends: two-input class classification methods (SIMCA) and similarity analysis. The SIMCA strategy was effective only for detecting blends carried out on virgin olive oils with a greenish hue (high chlorophyll/pheophytin content). Furthermore, the similarity profile, developed and applied for the first time in this study evidences the blending in all cases irrespective of the original olive oil hue.

V.a. RESUMEN (EN ESPAÑOL)

El aceite de oliva virgen extra genuino es un alimento potencialmente vulnerable que puede mezclarse con otros aceites vegetales comestibles, incluidos los de peor calidad, para obtener beneficios ilícitos. Estas operaciones no autorizadas pueden tener lugar en cualquier fase del proceso de producción y afectan radicalmente a la composición química. En este estudio se ha llevado a cabo el análisis de diferentes muestras de aceite de oliva virgen antes y después de su mezcla en diferentes proporciones con otros aceites de oliva de menor calidad. El análisis directo de las muestras por (NP)HPLC-DAD en un rango de longitudes de onda comprendidas entre 190 y 700 nm permitió la caracterización simultánea de varias familias de compuestos entre las que se incluyen: clorofilas, feofitinas, carotenos y tocoferoles, siendo los tres primeros, responsables del color del aceite de oliva.

Las técnicas para el reconocimiento de pautas no supervisadas aplicadas sobre las huellas instrumentales espectro-cromatográficas bidimensionales (2D) de muestras de aceite de oliva virgen sin mezclar, mostraron claramente la aparición de agrupaciones de las muestras según la tonalidad de las mismas (verde y amarillo). Para detectar la aparición de estas mezclas fraudulentas, se probaron dos estrategias basadas en poner de manifiesto cambios en las huellas cromatográficas espectrales: (i) modelos de clasificación de dos clases de entradas (SIMCA) y (ii) análisis de similitudes. La estrategia SIMCA sólo fue eficaz para detectar las mezclas realizadas sobre aceites de oliva vírgenes con un tono verdoso (alto contenido en clorofila/feofitina). Sin embargo, el uso de perfiles de similitud entre huellas instrumentales, desarrollado y aplicado por primera vez en este estudio, pone de manifiesto de forma correcta la presencia de mezclas en todos los casos en que estas se dieron, independientemente de la tonalidad original del aceite de oliva.

V.b. CONCLUSIONES DEL TRABAJO (EN ESPAÑOL)

El análisis directo de muestras de aceite de oliva por (NP)HPLC-DAD ha permitido la detección simultánea de diferentes familias de compuestos (algunos de ellos relacionados con el color de los aceites de oliva) cuya composición cambia cuando se produce una mezcla con cualquier otro aceite vegetal. La metodología propuesta ha permitido no sólo una reducción considerable del tiempo de análisis y del procesamiento de datos necesarios para obtener un resultado fiable, sino también una reducción del

coste del análisis. Además, la dependencia de la señal con respecto del sistema cromatográfico se ha minimizado gracias al uso de la metodología de agnostización instrumental. Cabe destacar que el procedimiento desarrollado se utiliza por primera vez con un conjunto de estándares químicos externos fiables, lo que permite la correcta aplicación de esta metodología cuando se aplica cromatografía líquida en fase normal.

El empleo de herramientas quimiométricas mediante métodos de reconocimiento de pautas supervisados y no supervisados ha revelado que las mayores diferencias entre las muestras analizadas se deben a la tonalidad presente en las mismas. El modelo de clasificación SIMCA desarrollado ha sido capaz de poner de manifiesto las mezclas en aquellas muestras que originalmente tenían una fuerte tonalidad verdosa, pero no detecta las mezclas producidas cuando la tonalidad del AOVE original comercializado es diferente.

El perfil de similitud desarrollado y aplicado por primera vez en este estudio ha demostrado ser una herramienta sencilla para evitar los problemas de clasificación errónea anteriormente descritos, además de ser útil para indicar cuándo se ha producido una mezcla de aceites de oliva, independientemente de la tonalidad de la muestra. Dado que el perfil de similitud no requiere una reducción de variables, el uso de toda la información espectro-cromatográfica obtenida en el análisis, permite una reducción significativa del tiempo requerido para el procesamiento de datos y la toma de decisiones, siendo además dichos resultados más fiables que los facilitados por el modelo de clasificación.

Por último, la metodología global propuesta y aplicada en este estudio podría considerarse como una herramienta útil para la prevención y/o detección de posibles mezclas fraudulentas que pudieran haberse producido durante el transporte del aceite de oliva virgen extra en grandes cisternas, mediante el análisis de una alícuota de aceite obtenida tanto en origen como en destino.

V.c. ESTUDIO BIBLIOMÉTRICO

El trabajo recogido en el presente Capítulo describe una metodología integra para la resolución de problemas analíticos relacionados con mezclas no autorizadas analizadas mediante técnicas cromatográficas, utilizando el aceite de oliva como caso de estudio. En este trabajo además se describe una nueva aplicación de la metodología de agnosticismo instrumental descrita en el Capítulo III y que fue publicado en la revista *Journal of*

Chromatography A, pero en esta ocasión centrado en su aplicación en cromatografía de líquidos en fase normal.

Al ser una nueva ampliación de la metodología anteriormente descrito, se consideró como primera opción para la publicación del trabajo la propia revista *Journal of Chromatography A*, la cual, como se ha descrito en el estudio bibliográfico recogido en el Capítulo III, se trata de una revista de reconocido prestigio en el ámbito de la química analítica.

Es necesario señalar que, en la fecha de entrega de esta Memoria, el artículo tiene asignado número de manuscrito (JCA-22-35) y actualmente se encuentra en el status de "under review" en la web de la revista, por lo que próximamente se espera la recepción de una respuesta sobre la aceptación o rechazo del artículo en dicha revista (Figura 19).

Manuscript Number ▲	Title ▲	Initial Date Submitted ▼	Status Date ▲	Current Status
JCA-22-35	The chromatographic similarity profile – an innovative methodology to detect fraudulent blends of virgin olive oils	08 Jan 2022	11 May 2022	Under Review

Figura 19. Captura de pantalla de la web de la revista *Journal of Chromatography A* con información acerca del estado del manuscrito a fecha de depósito de esta Memoria.

V.d. ARTÍCULO CIENTÍFICO

The chromatographic similarity profile – an innovative methodology to detect fraudulent blends of virgin olive oils

Fidel ORTEGA-GAVILÁN , Ana M^a JIMÉNEZ-CARVELO, Luis CUADROS-RODRÍGUEZ, M^a. Gracia BAGUR-GONZÁLEZ

Department of Analytical Chemistry, Faculty of Science, University of Granada, C/ Fuentenueva s/n, E-18071, Granada (Spain)

Highlights

- Simultaneous analysis of different family compounds related to olive oil colour
- Instrument-agnostizing methodology to obtain chromatography-independent results
- Use of pattern recognition methods to identify possible blends
- Use of similarity profiles for fake olive oil detection regardless colouring

Keywords

Direct analysis of olive oil, chromatography-agnostizing methodology, similarity profile, olive oil blend detection.

Abstract

Extra virgin olive oil is a potentially vulnerable foodstuff that can be mixed with other vegetal edible oils including poorer quality olive oils in order to obtain illicit profits. These unauthorized operations may take place at any stage of the production process and radically affect the chemical composition. In this paper, the analysis of different virgin olive oil samples before and after blending with other lower-grade olive oils in different

 Corresponding author, e-mail: fog@ugr.es; Phone: +34 958240797.

proportions were performed. The direct analysis of the samples by (NP)HPLC-DAD in a wavelength range between 190 and 700 nm allowed the simultaneous analysis of several compound families responsible of the colour including chlorophylls, pheophytins, carotenes and tocopherols, the first three responsible for the olive oil colour.

Unsupervised pattern recognition techniques applied on the chromatography-agnostic fingerprints of unblended virgin olive oil samples clearly showed the occurrence of groupings according to the sample hue (green and yellow). Two strategies, based on revealing changes in the spectrum-chromatographic fingerprints, are tested in order to detect the occurrence of such fraudulent blends: two-input class classification methods (SIMCA) and similarity analysis. The SIMCA strategy was effective only for detecting blends carried out on virgin olive oils with a greenish hue (high chlorophyll/pheophytin content). Furthermore, the similarity profile, developed and applied for the first time in this study evidences the blending in all cases irrespective of the original olive oil hue.

1. INTRODUCTION

Virgin olive oil (VOO) is one of the most prestigious products in the food sector. Apart from the hedonic character, this food has exceptional properties for the health of the consumers [1,2]. In contrast to other edible vegetable oils, VOO is produced exclusively using mechanical processes avoiding the use of solvents. As it is a totally mechanical process, the production costs and therefore the selling price of VOO, in particular the highest quality grade, referred as extra virgin olive oil (EVOO), are higher than those of other vegetable oils [3,4,5]. Such differences in the quality and selling price make EVOO a foodstuff potentially susceptible to being mixed with other edible vegetal oils, including olive oils of poorer quality, in order to obtain an illicit profit [6]. In this last case, the blending percentages are higher than those usually used for blending with other vegetable oils. Fake EVOO could appear in any stage of the production process, including the large-scale transport and delivery from the oil mill to the packaging factory.

The fraud risk is the main concern for consumers and certain sectors of the industry that are negatively affected [6]. This concern has reached European governmental bodies, which are attempting to ensure regular control of olive oils in order to prevent possible counterfeiting by publishing various regulations [7,8]. Some member States have adopted these measures in their own legislation, as in the case of the Spanish government, which in 2021 approved a law to guarantee the traceability of olive oils [9].

Indeed, a number of analytical methods, mainly spectroscopic and chromatographic, have been developed and applied over the last decades to reveal non-authorized blending. Spectroscopic techniques such as near-infrared (NIR), Fourier transform infrared (FTIR), Raman, or nuclear magnetic resonance (NMR), always supported by chemometrics have demonstrated their applicability for this purpose. For example, Yuanpeng Li et al. applied interval and synergy interval partial least squares regression (iPLS and siPLS respectively) to create quantification models of EVOO samples mixed with frying oils by analysing the samples directly with a portable Raman spectrometer [10]. Other studies such as those developed by Jiménez-Carvelo et al. involve the development of different classification models for VOO blended with other edible vegetable oils, analysing the methyl trans-esterified fraction of the samples using FTIR and Raman spectroscopies [11]. Although these methods require low cost and little sample preparation, do not provide the same amount of compositional information as chromatographic methods.

In recent years many chromatographic methods are focused on the quantification of specific chemical markers to identify and quantify the blends of EVOO with other types of

vegetable oils [12]. Both the analysis of major components such as triglycerides or the fatty acids of which the triglycerides are composed, and other minority fractions such as phytosterols, volatile organic compounds or phenolic compounds, among many others, are carried out. Some of these families of compounds such as triglycerides, fatty acids methyl esters, phytosterols or volatile compounds can be determined by gas chromatography, being the main disadvantages of these methods: (i) the use of high temperatures of analysis, (ii) the long incubation times of the samples, or (iii) the need for previous sample pre-treatment as fractionation or derivatisation [13,14,15,16,17]. Furthermore, substances such as triglycerides, phytosterols, phenolic compounds, tocopherols/tocotrienols, chlorophylls/pheophytins or carotenoids are determined by liquid chromatography. The use of the reverse phase liquid chromatography ((RP)HPLC) for the determination of these compounds has become widespread, and in most cases a prior extraction/clean-up step is necessary, which complicates the analysis [13,18,19,20].

Nevertheless, some of these compounds such as tocopherols/tocotrienols and compound related with olive oil (OO) hue can be determined by normal phase liquid chromatography ((NP)HPLC) by directly introducing the olive oil sample into the chromatographic system [21,22,23,24]. However, it should be noted that the main disadvantage of the colour-responsible compounds is their low stability over time. Despite this, the colorant fraction has been used to reveal fake VOO, for example by the addition of other exogenous colouring agents of both natural and artificial origin [25]. In addition, Gandul Rojas et al. report that the colorant composition depends significantly on the variety and the production area, which would allow the botanic authentication of the olive fruit to be assessed [20]. Similarly, the tocopherol profile can also be exploited. Tocopherols are natural antioxidants related to the stability of olive oil. This group of compounds is closely related to chlorophylls as they prevent the degradation of sensitive compounds in oxidative conditions, i.e., light or oxygen presence [21]. The concentration of this chemical family is reduced during the olive oil refining process, which could be used to detect blending by addition of this type of olive oil [26].

The blending of vegetal oils of different types or categories always results in a change in the proportion and content of colorants or tocopherols compounds that could imply significant differences in the related chromatographic fingerprints. The similarity assessment between two instrumental fingerprints has the potential to be used as a tool to detect non-authorised blending as long as the chromatographic fingerprint is available before and after the fraudulent mixture occurs. The similarity study by closeness analysis involves the evaluation of the algebraic (Euclidean) distance in two-dimensional space

that lies among different elements of two data matrices. In other words, the less distance between the elements being compared, the greater similarity between the two matrices involved. Recent research focusing on olive-oil-based reference material certification studies of has demonstrated the usefulness of this kind of strategy to assess the differences between samples of the same batch –homogeneity assessment– or to monitor the compositional changes that could occur over time –stability assessment– [27,28,29]. The only disadvantage in proximity analysis arises when instrumental fingerprints from the same sample analysed on different analytical measuring devices are compared. Nevertheless, the instrument-agnostizing methodology proposed by Cuadros-Rodríguez et al. [30,31] would provide a workable solution to this drawback since it allows obtaining chromatographic fingerprints that are independent of the state of the equipment and/or the time in which they were acquired. The agnostizing process is based on the use of two specific sets of chemical standards (internal and external sets) that must be previously fitted for the purpose in order to normalise both the intensities and the time scores of the whole chromatographic signal. Up to date, this methodology has not been applied in (NP)HPLC, and therefore no suitable chemical compound set to be used for the external correction of the signal shift has been described in the literature.

In this context, this study proposes the simultaneous analysis of different family compounds (mainly chlorophylls, pheophytins, carotenoids and tocopherols), the three first ones related to the colour of VOO in order to prevent and detect possible fraudulent mixtures during transport in large tankers. Using a (NP)HPLC-DAD method, different EVOO and in-lab-prepared blends samples are analysed in order to emulate the problem described above. The blends are prepared by mixing EVOO with others lower quality olive oils in proportions that yield a maximum illicit profit. Apart from pattern recognition techniques, a new application of similarity indices, called similarity profiling, which is able to identify fraudulent blending, has been established and firstly applied in this work.

2. MATERIALS AND METHODS

2.1. Chemicals and standard solutions

HPLC quality solvents such as n-hexane (Panreac Quimica S.L.U., Barcelona, Spain) and 2-propanol (Honeywell, Deutschland, Germany) were used in this study.

Chemical standards such as trimethylbenzene (99%, Sigma-Aldrich, St Louis, MO, USA), acetophenone (99%, Alfa Aesar, Kandel, Germany), propiophenone (99%, Alfa Aesar, Kandel, Germany) 2,5 di-tert-butylhydroquinone (99%, Sigma-Aldrich, St Louis, MO, USA), 2,5-dimethylphenol (99%, Sigma-Aldrich, St Louis, MO, USA), 3,4-dimethylphenol (99%, Sigma-Aldrich, St Louis, MO, USA), cholesterol (99%, Sigma-Aldrich, St Louis, MO, USA), uvaol (99%, Sigma-Aldrich, St Louis, MO, USA), ethyl p-hydroxybenzoate (99%, Fluka Chemika, Buch, Germany), trimethylhydroquinone (99%, Sigma-Aldrich, St Louis, MO, USA), tert-butylhydroquinone (99%, Sigma-Aldrich, St Louis, MO, USA) and 4-ethylresorcinol (99%, Sigma-Aldrich, St Louis, MO, USA), USA) were used for this study. The ordering of these compounds corresponds to the experimental elution order under the analytical conditions applied.

A solution of 2,4-dimethylphenol (99%, Sigma-Aldrich, St Louis, MO, USA) in a 99/1 (v/v) hexane/2-propanol mixture of 5 mg/L final concentration was used as internal standard solution.

The external standard mixture containing 5 mg/L for trimethylbenzene, acetophenone and propiophenone; 10 mg/L of 2,5 dimethylphenol and 3,4 dimethylphenol; 15 mg/L of 2,5 di-tert-butylhydroquinone and 4-ethylresorcinol; 25 mg/L of ethyl p-hydroxybenzoate and tert-butylhydroquinone; 100 mg/L of cholesterol and 200 mg/L of uvaol, necessary to apply the chromatography-agnostic methodology, was prepared in n-hexane/2-propanol 99/1 (v/v).

2.2. Olive oil samples

The olive oil sample bank was constituted by two sample sets. The first set includes marketed olive oils of different quality categories having a wide range of colour hues, from deep green to golden yellow: 34 EVOOs, 4 VOOs, a sample of lampante olive oil (LOO) selected for its high green hue and an uncoloured sample of refined olive oil (ROO). The sample codes, quality categories and colour are shown in Table 1.

Table 1. Description of the marketed olive oils included in the sample bank.

Sample code	Quality category	Colour
V_001	EVOO	Green
V_002	EVOO	Green
V_003	EVOO	Green
V_004	EVOO	Green
V_005	LOO	Green
V_006	EVOO	Green
V_007	EVOO	Green
V_008	EVOO	Green
V_009	EVOO	Green
V_010	EVOO	Green
V_011	EVOO	Green
VM_012	EVOO	Medium Green
VM_013	EVOO	Medium Green
VM_014	EVOO	Medium Green
VM_015	EVOO	Medium Green
VM_016	EVOO	Medium Green
MA_017	EVOO	Medium Yellow
MA_018	EVOO	Medium Yellow
MA_019	EVOO	Medium Yellow
MA_020	EVOO	Medium Yellow
MA_021	EVOO	Medium Yellow
MA_022	EVOO	Medium Yellow
MA_023	EVOO	Medium Yellow
MA_024	EVOO	Medium Yellow
MA_025	EVOO	Medium Yellow
A_026	EVOO	Yellow
A_027	EVOO	Yellow
A_028	EVOO	Yellow
A_029	EVOO	Yellow
A_030	EVOO	Yellow
A_031	EVOO	Yellow
A_032	EVOO	Yellow
A_033	EVOO	Yellow
A_034	EVOO	Yellow
A_035	EVOO	Yellow

Table 1. (cont.) Description of the marketed olive oils included in the sample bank.

a_036	VOO	Yellow
b_037	VOO	Yellow
c_035	VOO	Yellow
d_039	VOO	Yellow
e_040	ROO	Bleached

Note. In bold are marked the poorest quality olive oils used to produce the fraudulent blends.

The second one set was constituted by 70 in-lab prepared blends resulting from mixing the original marketed olive oil samples with 25% or 50% w/w of other poorer quality olive oils. The blended sample codes, colours, composition, both actual and initial quality categories and colour are shown in supplementary material (Table SM1).

For sample preparation, 0.400 g of the marketed or blended olive oil was weighed into 8 mL glass vials containing 100 μ L of internal standard solution. On the same vial, 4.9 mL of a 99/1 (v/v) n-hexane/2-propanol mixture was added to dilute the sample to a final oil concentration of 8% (w/v). The samples were filtered with PTFE filters with a pore size of 22 μ m before analysis.

2.3. Chromatography analysis

Chromatographic analysis was performed on an Agilent 1260 Infinity II LC System (Santa Clara, CA) equipped with a CH30 column thermostat (Eppendorf, Hamburg, Germany), a quaternary pump, a degasser and an autosampler. A G7115A infinity II diode array UV-Vis molecular absorption detector (DAD) was used as measuring device, which was programmed to cover the wavelength range between 190 and 700 nm. This range allows the detection of tocopherols, chlorophylls, pheophytins and carotenoids in a single chromatographic run.

A column containing silica as stationary phase (Konic Extrasil SI60, 250 $\times$ 4 mm i.d., 10 μ m) coupled to a pre-column with the same stationary phase and diameter (10 $\times$ 4 mm i.d.) were used for all analyses. Chromatographic separation was performed using a mixture n-hexane/2-propanol 99/1 (v/v) as solvent A and 2-propanol as solvent B, maintaining a constant flow rate of 0.8 mL/min and keeping the column temperature at 25°C.

The chromatographic analysis of the external standard mixture was repeated 30 times over 4 days in order to take into account as much variability as possible in the calculation of the standard retention scores (SRS) [31]. Additionally, the external standard mixture was analysed at the beginning, middle and end of each chromatographic batch for the subsequent calculation of the SRS. The chromatographic separation of the external standard mixture was carried out as follows: 100% phase A at min 0, from min 0 to min 15 a progressive increase of phase B until reaching 10%, maintaining this percentage until min 17, while from min 17 to min 20 the initial conditions were restored. In the case of oil samples, chromatographic separation was performed according to the following gradient: 100% phase A from min 0 to min 4, increase to 10% phase B from min 4 to min 8, maintain 10% phase B until min 15, return to initial conditions in two min (100% phase A) and a between-analyses time of 3 min. The injection volume was 20 μ L for both external standard mixture and olive oil samples.

2.4. Data treatment

2.4.1. Obtaining of chromatography-agnostic fingerprints

Agilent ChemStation OpenLab CDS software (rev C.01.09) was used for instrumental control, as well as data visualisation and export. For each analysis, a rectangular matrix of data intensities was obtained containing the information related to the chromatographic separation (columns) for each wavelength recorded in the analysis (rows).

Data processing was carried out within the MATLAB environment, version 9.6 (Mathworks Inc., Natick, MA, USA), applying the ad-hoc script called LCnostic-2D (version 1), which makes use of different functions of the Matlab Bioinformatics Toolbox. Prior to running the script, the different rectangular data matrices were exported to the MATLAB environment. Each data matrix was processed according to the following procedure:

- (1) Subtraction of the blank signal in order to remove the interferences produced by the mobile phase in the runs.
- (2) Selection and truncation of the 2D-spectrum-chromatographic fingerprint to remove regions that do not contain relevant information. The selected range of interest was that corresponding to those peaks appearing between minute 3 and 16 in the original chromatograms.

- (3) Signal filtering and smoothing by applying the Savitzky-Golay filter with a window size of 5 points and a second order polynomial fitting over each window.
- (4) Intensity normalisation taking as a reference the maximum intensity value of the internal standard, which is assigned the value 1 in the new intensity scale. In this case the maximum absorption of the internal standard appears between minutes 6.7 and 7.7 in the original chromatogram at a wavelength range 200-250 nm.
- (5) Baseline correction by applying the 'msbackadj' function using 150 units of separation for the 'windowsize' parameter and 37.5 units of separation for the 'stepsize' parameter.
- (6) Determination of standard retention scores (SRS) in accordance with the procedure provided by Cuadros-Rodriguez et al. [31].
- (7) Replacement of the retention time values on the whole x-axis by the estimated SRS values for each chromatographic signal, in order to correct potential warping/drift times.
- (8) Resampling in order to retain the same number of elements in all data matrices using the function 'resample'.

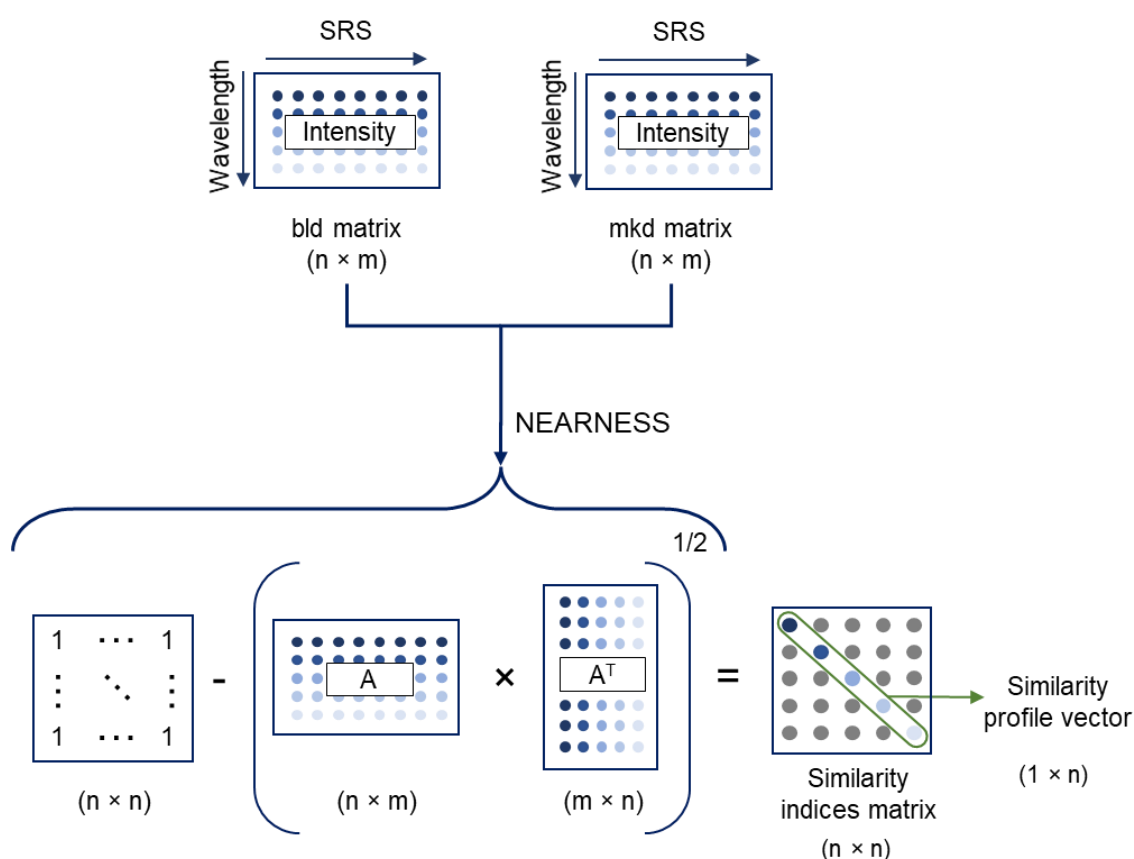
As a result of this script, 110 data matrices were obtained, where each matrix was the 2D spectrum-chromatographic instrument-agnostic fingerprint (2D SCIA-FP) related to each olive oil sample. By applying the instrument-agnostizing methodology, the influence of the instrument state on the acquired chromatographic signal were eliminated, so that the differences in the instrument-agnostic fingerprints were only due to the differences in the composition of the sample.

2.4.2. Similarity profiling

Once the 2D SCIA-FP were obtained for each analysed samples, the similarity between each blended OO sample and the corresponding marketed olive oil was evaluated by applying the 'nearness matrix' [27] according to the following matrix expression:

$$\text{NEAR} (M_{\text{bld}}, M_{\text{mkd}})_{(n \times n)} = M(1) - \left[\sqrt{\left(\frac{(M_{\text{bld}} - M_{\text{mkd}})}{(M_{\text{bld}} + M_{\text{mkd}})} \right) \cdot \left(\frac{(M_{\text{bld}} - M_{\text{mkd}})^T}{(M_{\text{bld}} + M_{\text{mkd}})^T} \right)} \right]$$

where the 'bld' and 'mkd' subscripts denotes the matrix related to the blended and marketed olive oil samples, respectively. $M(1)$ is a $n \times n$ matrix where all the elements of such a matrix have the value 1, M_{bld} is a $n \times m$ matrix refers to 2D SCIA-FP of an EVOO blended sample and M_{mkd} is a $n \times m$ matrix related to 2D SCIA-FP of the corresponding starting marketed olive oil sample. In accordance with Matlab script, the dot present in the formula indicates that the matrices division is carried out element by element, i.e. each element of the numerator matrix is divided by the corresponding element of the denominator matrix. The output of this equation is a new square matrix $n \times n$ of values between 0 and 1, where each element of a row is the similarity index obtained by comparing all the elements of that row in the M_{bld} matrix with all the elements of the same column in the transposed M_{mkd} matrix. The main diagonal of the square matrix could be defined as the similarity profile vector or lambda dependant near indices for each blended and starting marketed olive oil samples. An illustrative scheme of how to obtain the similarity profile vector is shown in Figure 1.



Note. Matrix $A = (M_{bld} - M_{mkd}) / (M_{bld} + M_{mkd})$

Figure 1. Schematic graphical layout describing how the similarity profile is obtained.

3. RESULT AND DISCUSSION

3.1. Chemometric analysis

Each 2D SCIA-FP matrix was subjected to unfolding and variable reduction to compact all chromatographic information into a single data vector containing the total intensity signal at each SRS index. A brief schematic representation of the process is shown in the Figure SM1. As a result, a total of 110 data vectors were obtained related to the 1D unfolded-reduced instrument-agnostic fingerprints (1D URIA-FP), which were used to set up three different 1D URIA-FP data matrices as follow: the first one containing the marketed olive oils (40 samples $\times$ 1700 variables), the second one containing the 25% blended olive oils (35 samples $\times$ 1700 variables) and the last one containing the 50% blended olive oils (35 samples $\times$ 1700 variables).

Unsupervised pattern recognition analysis (PCA & HCA) was first performed on the marketed olive oil 1D URIA-FP matrix in order to assess the main structure of the data. In a second stage, a supervised pattern recognition analysis (SIMCA) was conducted on the marketed olive oil 1D URIA-FP matrix as training set and the 25% and 50% blended OO 1D URIA-FP matrices as external prediction sets to test the ability to achieve an effective classification between marketed and blended olive oils. In all cases, a mean centred pre-processing step was applied prior to the successive analysis.

3.1.1. Unsupervised Pattern Recognition Analysis

Principal component analysis was carried out by selecting the first 20 principal components (PCs) which explained 95.58% of the accumulated variance. Figure 2 shows the scores plot of PC1 (37.41% of the variance) vs PC2 (18.3% of the variance). Initially, two main groups (A and B) can be observed, being the Group A constituted by marketed EVOO samples with negative values of PC1 and Group B constituted by lower-quality olive oils, characterised by positive scores on PC1. In addition, examining the PC1 and PC2 scores of the Group A in more detail, 2 additional subgroups can also be observed: Subgroup I with negative values for both PC1 and PC2 is constituted by deep-green marketed EVOO samples; and Subgroup II with negative values close to zero for PC1 and positive or close to zero for PC2 is composed by yellowish hue EVOO (plotted by red diamonds) and an adjacent one characterised by an intermediate green-yellow hue (plotted by a blue triangle).

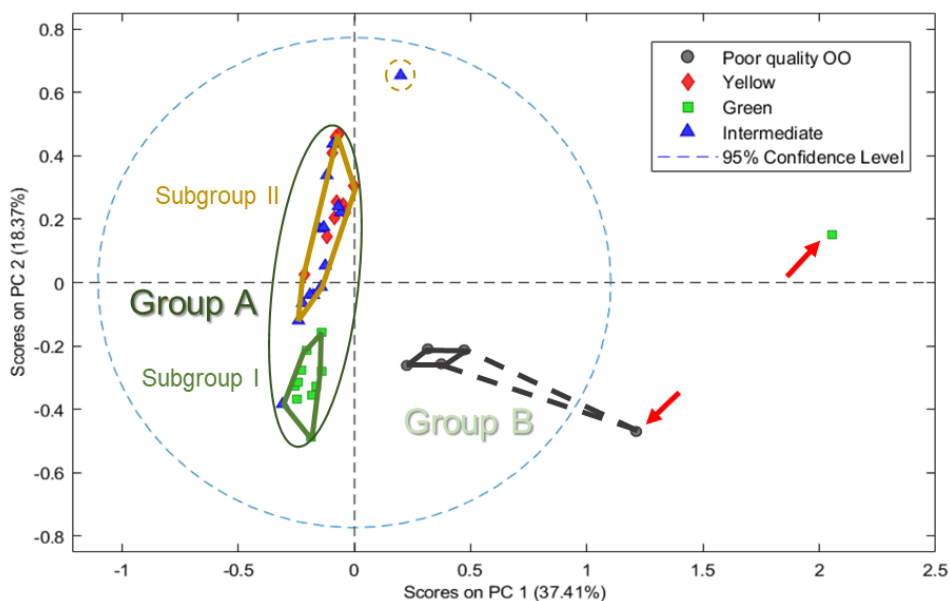


Figure 2. Score plot of PC1 vs PC2 from PCA using the 1D unfolded-reduced instrument-agnostic fingerprints (1D URIA-FP) matrix (40 samples $\times$ 1700 variables) from marketed olive oil.

Furthermore, Figure 2 shows the occurrence of outliers, which are identified by an arrow and correspond to the samples V_005 (green square highlighted), e_040 (grey circle highlighted).

In order to verify the natural grouping, regarding only the EVOO colour, a new PCA analysis was run using a new reduced EVOO 1D URIA-FP data matrix (34 samples $\times$ 1700 variables) in which six samples (lower quality olive oils) were excluded. The first 4 PCs were then extracted which explained 96.25% of the accumulated variance. Figure 3 shows the new projection of the EVOO samples on different PCs.

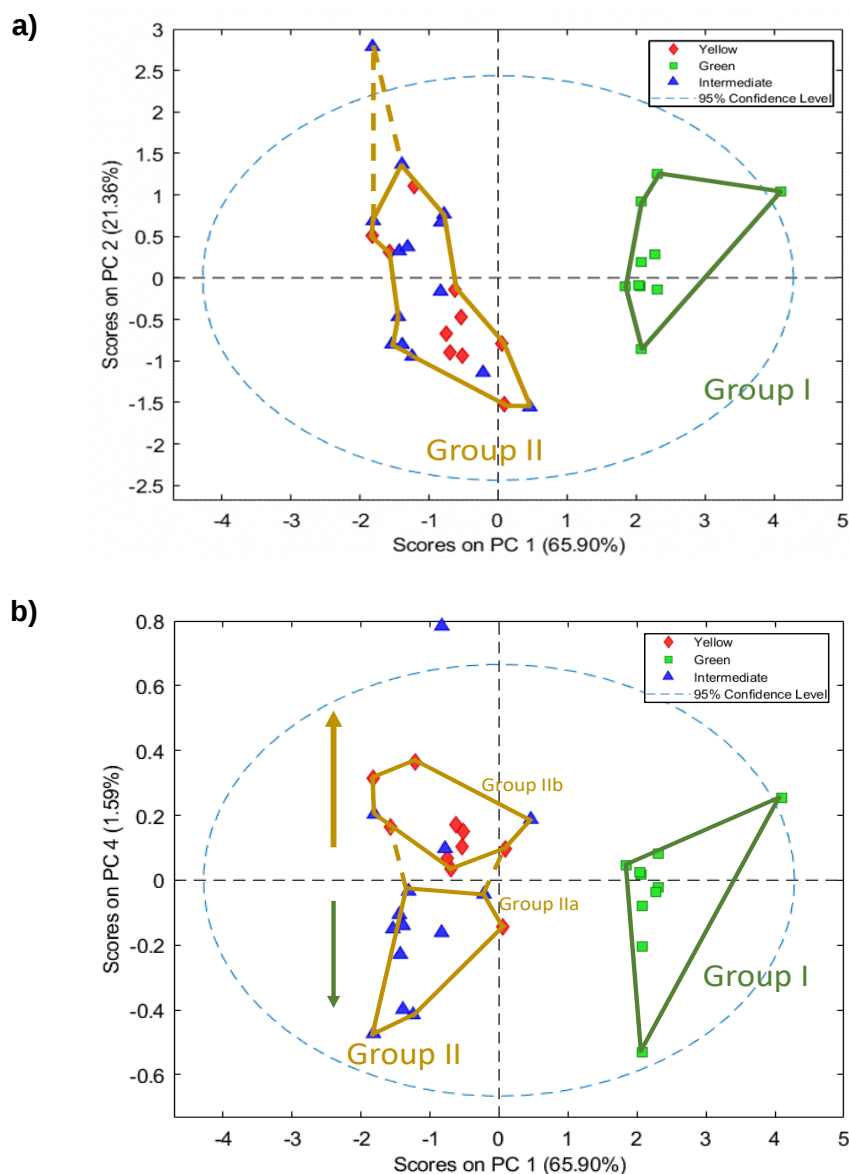


Figure 3. Score plots from PCA using the 1D unfolded-reduced instrument-agnostic fingerprints (1D URIA-FP) matrix (34 samples $\times$ 1700 variables) from the marketed olive oil. a) PC1 vs PC2. b) PC1 vs PC4.

Figure 3a shows clearly two groups of samples according to their colour. As can be seen, PC1 separates the samples according to the 'Green-Not green' colour of the EVOO, distinguishing between deep green samples (Group I, plotted at positive PC1 score region) and the remaining samples (Group II, plotted at negative or near zero PC1 scores region). It also can be seen how scores on PC2 establishes some differences between the diverse EVOO hues that compose Group II, although the trend is not clear.

These differences are more evident in Figure 3b, which shows the scores plot of PC4 vs PC1. PC4 separates Group II into two subgroups, Group IIa, plotted on the negative region of both PC1 and PC4 scores, collects the EVOO samples having an intermediate green-yellow hue. On the contrary, Group IIb, showing positive PC4 scores and negative PC1 scores, groups the EVOO samples with yellowish hue. The chlorophyll/pheophytin and carotenoid content in the olive oils is the responsible for the grouping, having a higher chlorophyll/pheophytin content in Group I. Notice that Group I cover a larger region than Group 2. This is a result of the higher ability of PC4 to detect differences between similar greenish hues. Furthermore, the image also reveals the presence of two EVOO samples (V_010 and MA_017) outside the confidence limit of the model. This is probably due to the fact that these samples are not fully transparent as they were not filtered at the factory before packaging.

To corroborate the results found in PCA, a hierarchical cluster analysis (HCA) was carried out, selecting the Manhattan distance and Ward's method as distance metric and nesting criterion respectively. Figure 4 shows the dendrogram resulting from this analysis.

Considering $D_{\text{linkage}} = 2/3 D_{\text{max}}$ as a criterion to define the natural grouping of the cluster, two groups can be defined again: Group I gather all the green EVOO labelled accordingly, while Group II is constituted by the remaining yellow or intermediate green-yellow olive oil samples.

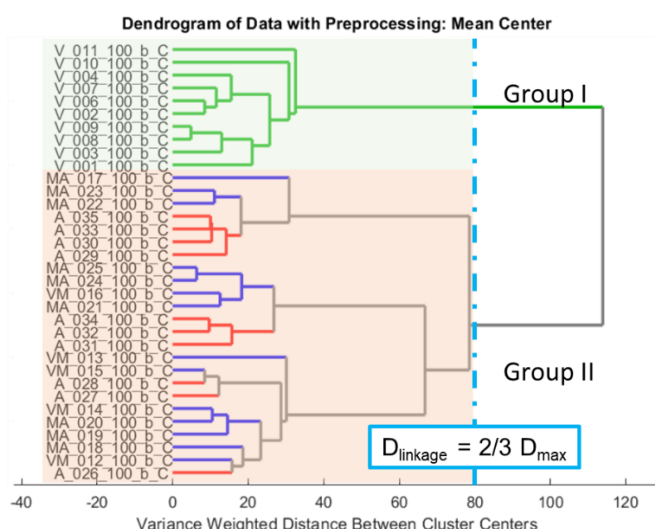


Figure 4. Dendrogram of pure olive oils samples analysed obtained from 1D unfolded-reduced instrument agnostic fingerprints (1D URIA-FP) matrix (34 samples $\times$ 1700 variables) of the marketed EVOO samples. Mahalanobis distance and Ward method was applied as distance metric and linked criterion, respectively.

3.1.2. Supervised Pattern Recognition Analysis

In order to identify non-authorised blended olive oils, a two-input class SIMCA classification model was then developed. The input classes of the model were established according to the previous results obtained from PCA and HCA by getting the 'Green-Not green' labels (see table SM1 in supplementary material). A different validation approach to that traditional one used in the development of chemometric models was applied in this study. The model is considered successful if the samples of the external prediction set are misclassified or stated as inconclusive ones (i.e., not belonging to any of the modelled classes), so that the presence of mixtures is revealed.

The SIMCA model, which were developed from the marketed olive oil 1D URIA-FP matrix, included 29 samples labelled as 'Not green' and the remaining 11 samples labelled as 'Green', in accordance with the information provided in table 1. The model was created retaining 4 and 5 PCs for the first ('Not green') and the second ('Green') class, respectively, which explained 96.60% and 98.14% of the total variance. In order to validate the method, the SIMCA model was tested using two datasets independently composed by 1D URIA-FP matrices corresponding to the 25% or 50% blended olive oil previously described. Figure 5a and 5b show the corresponding Cooman's plots for both datasets.

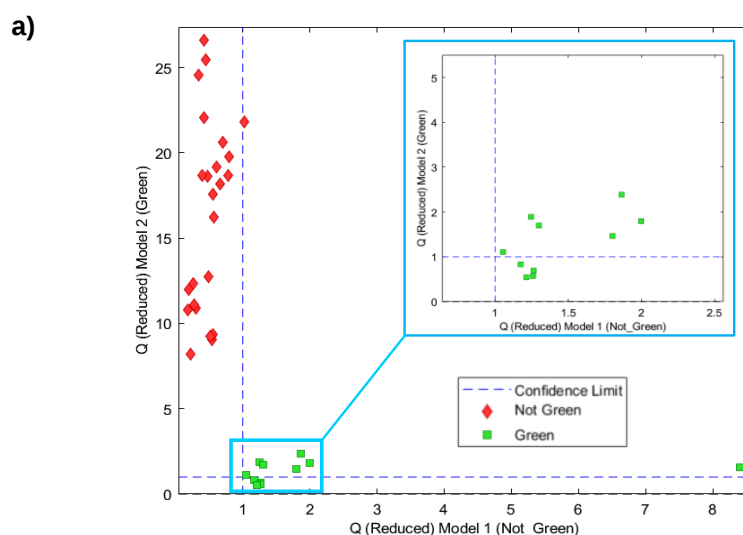


Figure 5. Cooman's classification plots using as prediction set 1D the unfolded-reduced agnostic fingerprints (1D URIA-FP) matrices from the blended olive oil samples related to: a) 25% blends, and b) 50% blends.

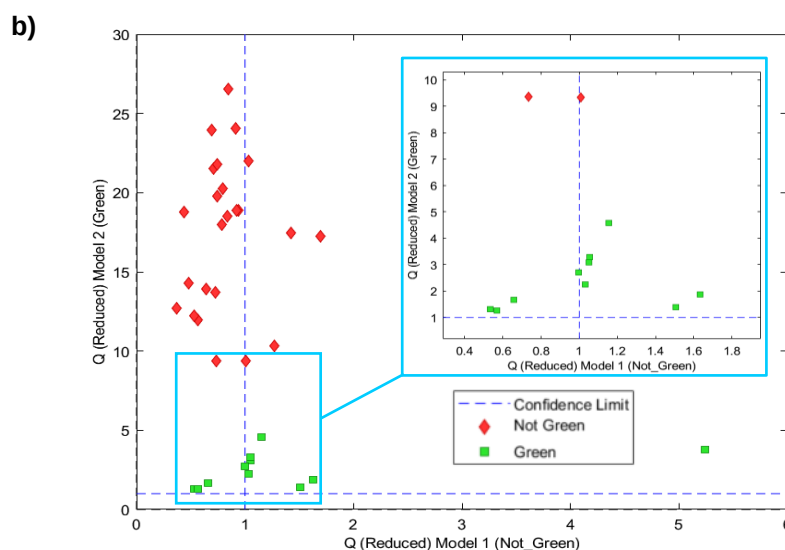


Figure 5. (cont.) Cooman's classification plots using as prediction set 1D the unfolded-reduced agnostic fingerprints (1D URIA-FP) matrices from the blended olive oil samples related to: a) 25% blends, and b) 50% blends.

The zoomed region in Figure 5a shows how only 4 'Green' samples are contained within the bottom-right quadrant, which is representative of that class. The remaining samples of the 'Green' class were not recognised as belonging to any class. Only one 'Not green' sample, coded as MA_017_25e, is located in the top-right quadrant, although it is very close to the classification limit.

In the same way, the zoomed region from the Figure 5b shows that 3 'Green' samples are located in the top-left quadrant, which is representative of the 'Not green' class. Indeed, an additional 'Nor green' sample is found within the top-left quadrant. The remaining 'Green' samples are non-recognised since they are located within the top-right quadrant, although most of them are very close to the limits. Concerning the 'Nor green' samples, 19 of them are located within the top-left quadrant, while the remaining 5 are non-recognised (top-right quadrant). The contingencies of both classifications are shown in Figures SM2 and SM3 respectively. These results allowed the determination of the classification performance metrics, which are shown in Table 2.

Table 2. Summary of classification performance metrics obtained for SIMCA model.

Classification performance metrics	Class 1 (Green)	
	25% adulterated	50% adulterated
Sensitivity (SENS)	0.36	0.00
Specificity (SPEC)	0.96	0.68
False positive rate (FPR)	0.04	0.32
False negative rate (FNR)	0.64	1.00
Positive predictive value (precision) (PPV)	1.00	0.00
Negative predictive value (NPV)	1.00	1.00
Youden index (YOUN)	0.32	-0.32
Positive likelihood rate (LR (+))	8.73	0.00
Negative likelihood rate (LR (-))	0.66	1.47
Classification odds ratio (COR)	13.14	0.00
F-measure (F)	0.53	–
Discriminant power (DP)	0.62	–
Efficiency (or accuracy) (EFFIC)	0.77	0.54
Misclassification rate (MR)	0.23	0.46
AUC (correctly classified rate)	0.66	0.34
Gini coefficient (Gini)	0.32	-0.32
G-mean (GM)	0.59	0.00
Matthews correlation coefficient (MCC)	0.59	–
Chance agreement rate (CAR)	0.49	0.46
Chance error rate (CER)	0.43	0.32
Kappa coefficient (KAPPA)	0.55	0.16
Bayes' PROB (1/1)	0.80	0.00
Bayes' PROB (2/2)	0.77	0.73
Bayes' PROB (1/2)	0.23	0.27
Bayes' PROB (2/1)	0.20	1.00

Note. The hyphen "–" refers to metrics that cannot be determined since a division by zero is involved.

Sensitivity values for both validation sets were rather poor indicating a low agreement with the class 1 ('Green'). By contrast, specificity values show that for 25% blended samples set there is a high concordance with the class 2 ('Not green'), while for 50% blended samples set this performance parameter shows poorer values. The precision values suggest that from the 25% blended samples set, those classified as 'Green' are really belonging to the 'Green' class. On the contrary, in the 50% blended samples set, none of the them were classified as belonging to the 'Green' class, which explains why the

precision value is equal to zero. Although the efficiency and misclassification rate values indicate that the developed model does not correctly classify class 1, it fulfils the purpose for which the model was developed. Indeed, blended olive oil samples, labelled as 'Green' samples, were either classified as 'Not green' or were not recognised as belonging to any class. Nevertheless, this behaviour is not as evident in the 'Not green' samples as only 5 samples are within the not recognised region.

3.2. Evaluation of similarity profile

Since yellowish blended olive oils were not efficiently classified by SIMCA, a similarity analysis was then performed between the 2D SCIA-FP matrices from blended olive oils and the corresponding marketed olive oils of the starting batches. To this end, the corresponding similarity profiles were evaluated according to the procedure described in subsection 2.4.2. In this approach, for each marketed olive oil sample concerned, two similarity profiles were obtained as a result of comparisons with the 25% and 50% blends. As representative examples, Figures 6a-6d show a graphical display of the two similarity profiles for four arbitrarily selected comparisons.

A_026 marketed EVOO with blended (d_039, 25%) olive oil (Figure 6a). It can be seen how the similarity values decrease in the wavelength range of 390 and 450 nm. Chlorophylls a and b, pheophytins a and b as well as carotenoids show a maximum absorbance in this wavelength interval. The values suggest that the colourant composition of the blended olive oils changes slightly but not enough to ensure fraudulent blending of the sample. Nevertheless, the similarity values in the range of 600 and 660 nm, where chlorophylls a and b show a second absorption maximum, are below the established threshold value. In addition, this second absorption maximum reveals the fraudulent blending.

A_026 marketed EVOO with blended (d_039, 50%) olive oil (Figure 6b). The same trend observed in Figure 6a is now enhanced as the blend percentage increases. In this case, the non-authorized blending would be also confirmed regarding the 390-450 nm range, since all similarity values have exceeded the threshold limit by far.

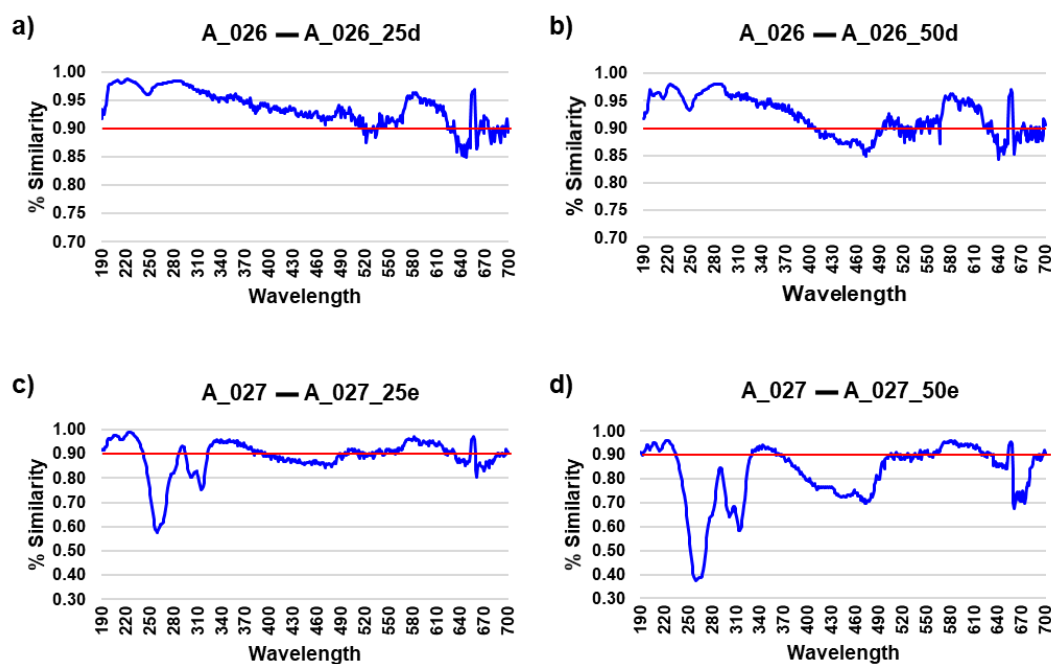


Figure 6. Similarity profile plots obtained by comparing: a) marketed EVOO sample A_026 with blended VOO sample A_026_25d; b) marketed EVOO sample A_026 with blended VOO sample A_026_50d; c) marketed EVOO sample A_027 with blended OO sample A_027_25e; and d) marketed EVOO sample A_027 with blended OO sample A_027_50e. The red line denotes the similarity threshold imposed by the authors.

A_027 marketed EVOO with blended (e_040, 25%) olive oil (Figure 6c). In this case, two additional ranges of interest can be observed, the former between 210 and 270 nm and the later between 290 and 330 nm. According to the literature [32], the first range could be related to the composition of unsaturated fatty acids, which produce a signal measurable by a UV detector while the second range is due to changes in the tocopherol profile. In this case, the fact evidences how the composition of these chemical families is modified when refined olive oil is used as a blending adulterant.

A_027 marketed EVOO with blended (e_040, 50%) olive oil (Figure 6d). The same behaviour as in the previous 25% blend becomes more remarkable as the blend percentage increases.

The blending occurrence was also easily evident for the blended olive oils showing green-yellow intermediate hue. Figure SM15b (supplementary material) plots the similarity profiles corresponding to the blended olive oils obtained from VM_012 EVOO sample. It can be seen how the similarity values of the 50% blended olive oil sample in the

wavelength range relative to the natural colourant fraction (390 to 450 nm) are reduced to a value of 0.7.

In order to graphically show the differences revealed by the similarity analysis, Figure 7 shows the 2D SCIA-FPs heat map from the VM_012 EVOO and for the blended olive oils as a result of 25% and 50% their blending with a_036 olive oil. The changes caused by blending could be identified with the naked eye—and are evidenced by an increase in the signal intensity (enclosed in a green rectangle or marked with arrows in the same colour) or a decrease (and even absence) of signal (enclosed in a red rectangle or marked with arrows in the same colour).

Finally, as can be seen in the supplementary material (Figures SM4-SM14) for the deep green olive oil samples, the occurrence of fraudulent blending with other oils is clearly evident due to the compositional change concerning tocopherols, chlorophylls, pheophytins and carotenoids.

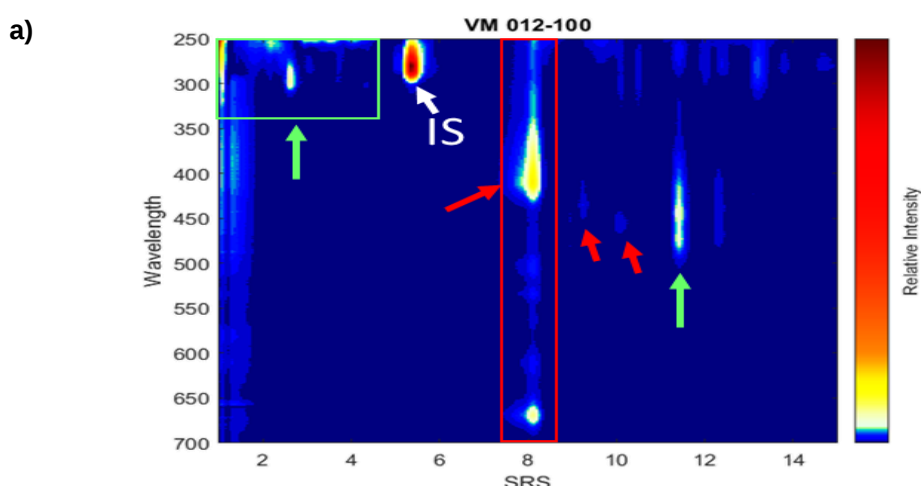


Figure 7. Heat maps corresponding to the 2D spectrum-chromatographic instrument-agnostic fingerprint (2D SCIA-FP) from samples: **a)** marketed EVOO VM_012; **b)** blended VOO sample VM_012_25a constitutes by VM_012 + a_036 (25%); and **c)** blended VOO sample VM_012_50a constitutes by VM_012 + a_036 (50%). Note: The meaning of the arrows is explained throughout the text of the article.

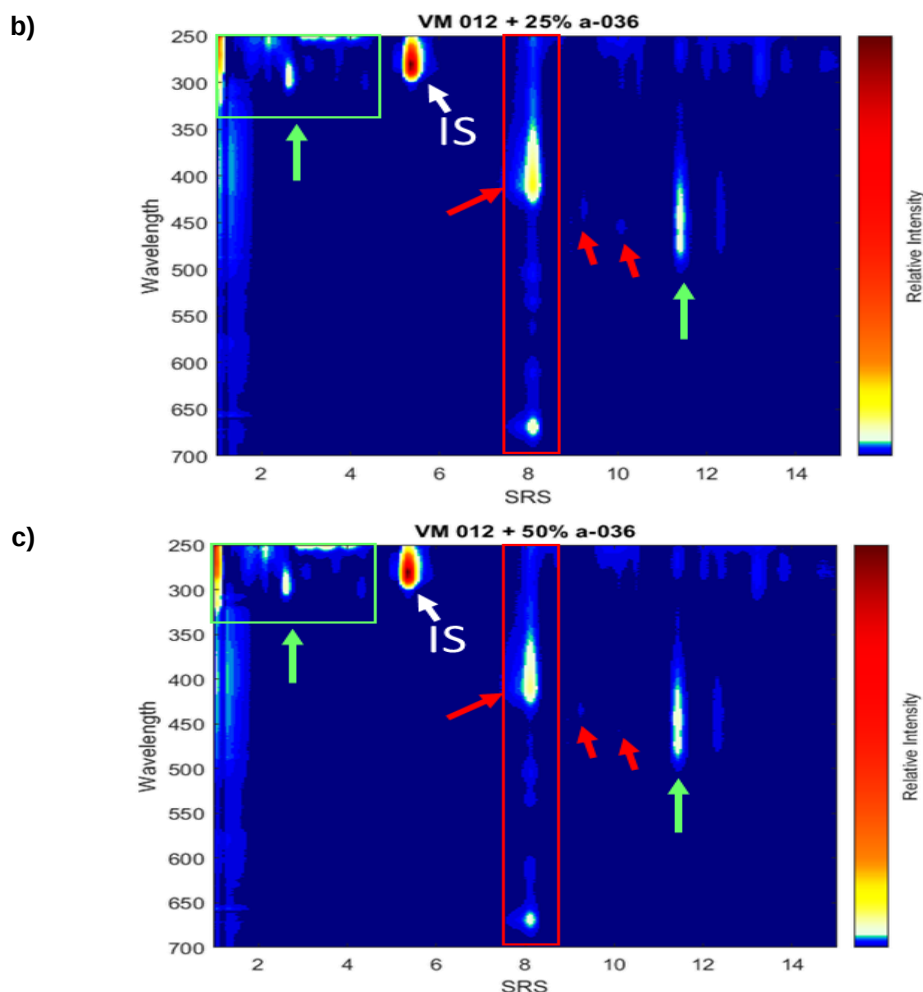


Figure 7. (cont.) Heat maps corresponding to the 2D spectrum-chromatographic instrument-agnostic fingerprint (2D SCIA-FP) from samples: **a)** marketed EVOO VM_012; **b)** blended VOO sample VM_012_25a constitutes by VM_012 + a_036 (25%); and **c)** blended VOO sample VM_012_50a constitutes by VM_012 + a_036 (50%). Note: The meaning of the arrows is explained throughout the text of the article.

4. CONCLUSION

The direct analysis of olive oil samples by (NP) HPLC DAD has allowed the simultaneous detection of different families of compounds (some of them related to the colour of olive oils) whose composition changes when a mixture with any other vegetable oil is present. The proposed methodology has allowed not only a considerable reduction in time analysis and data processing required to obtain a reliable result but also a reduction in cost analysis. In addition, the prevention of signal dependence over the state of the chromatographic system has been minimized by using the instrument-agnostizing methodology. It should be noted that the developed procedure is used for the first time

with reliable external standards set, which permits the correct application of this methodology in the normal phase.

The use of a chemometric approach by means of Unsupervised and Supervised methods has revealed that the greatest differences between the samples analysed has been due to the samples' hue. The developed SIMCA calibration model has been able to detect the blends in those samples that originally had a strong greenish hue, not being capable to detect the blending produced when the hue of the original marketed EVOO is not greenish.

The similarity profile developed and applied in this study for the first time has proven to be a simple tool to avoid the previous misclassification problems, as well as being useful to indicate when a blend of olive oils has occurred, regardless of the sample hue. Since the similarity profile does not require a variable reduction, the use of all spectrum-chromatographic information obtained in the analysis, permits a significant reduction in the time required for data processing and decision-taking and furthermore, most reliable results compared with the previous ones.

Finally, the overall methodology proposed and applied in this study could be considered as a useful tool for the prevention and/or detection of possible fraudulent mixtures that may have occurred during the transport of EVOO in large tanks, by analysing an aliquot of oil obtained both at origin and destination.

SUPPLEMENTARY MATERIAL

Table SM1. Description of the in-lab prepared blends included in the sample bank.

Sample code	Commercial category	Fake commercial category	Colour*	Mixed with
V_001_25a	VOO	EVOO	Green	a_036
V_001_50a	VOO	EVOO	Green	a_036
V_002_25b	VOO	EVOO	Green	b_037
V_002_50b	VOO	EVOO	Green	b_037
V_003_25c	VOO	EVOO	Green	c_038
V_003_50c	VOO	EVOO	Green	c_038
V_004_25d	VOO	EVOO	Green	d_039
V_004_50d	VOO	EVOO	Green	d_039
V_005_25e	LOO	LOO	Green	e_040
V_005_50e	LOO	LOO	Green	e_040
V_006_25a	VOO	EVOO	Green	a_036
V_006_50a	VOO	EVOO	Green	a_036
V_007_25b	VOO	EVOO	Green	b_037
V_007_50b	VOO	EVOO	Green	b_037
V_008_25c	VOO	EVOO	Green	c_038
V_008_50c	VOO	EVOO	Green	c_038
V_009_25d	VOO	EVOO	Green	d_039
V_009_50d	VOO	EVOO	Green	d_039
V_010_25e	OO	EVOO	Green	e_040
V_010_50e	OO	EVOO	Green	e_040
V_011_25a	VOO	EVOO	Green	a_036
V_011_50a	VOO	EVOO	Green	a_036
VM_012_25a	VOO	EVOO	Medium Green	a_036
VM_012_50a	VOO	EVOO	Medium Green	a_036
VM_013_25b	VOO	EVOO	Medium Green	b_037
VM_013_50b	VOO	EVOO	Medium Green	b_037

Table SM1. (cont.) Description of the in-lab prepared blends included in the sample bank.

Sample code	Commercial category	Fake commercial category	Colour*	Mixed with
VM_014_25c	VOO	EVOO	Medium Green	c_038
VM_014_50c	VOO	EVOO	Medium Green	c_038
VM_015_25d	VOO	EVOO	Medium Green	d_039
VM_015_50d	VOO	EVOO	Medium Green	d_039
VM_016_25a	VOO	EVOO	Medium Green	a_036
VM_016_50a	VOO	EVOO	Medium Green	a_036
MA_017_25e	OO	EVOO	Medium Yellow	e_040
MA_017_50e	OO	EVOO	Medium Yellow	e_040
MA_018_25a	VOO	EVOO	Medium Yellow	a_036
MA_018_50a	VOO	EVOO	Medium Yellow	a_036
MA_019_25b	VOO	EVOO	Medium Yellow	b_037
MA_019_50b	VOO	EVOO	Medium Yellow	b_037
MA_020_25c	VOO	EVOO	Medium Yellow	c_038
MA_020_50c	VOO	EVOO	Medium Yellow	c_038
MA_021_25b	VOO	EVOO	Medium Yellow	b_037
MA_021_50b	VOO	EVOO	Medium Yellow	b_037
MA_022_25c	VOO	EVOO	Medium Yellow	c_038
MA_022_50c	VOO	EVOO	Medium Yellow	c_038
MA_023_25d	VOO	EVOO	Medium Yellow	d_039
MA_023_50d	VOO	EVOO	Medium Yellow	d_039
MA_024_25e	OO	EVOO	Medium Yellow	e_040
MA_024_50e	OO	EVOO	Medium Yellow	e_040
MA_025_25a	VOO	EVOO	Medium Yellow	a_036
MA_025_50a	VOO	EVOO	Medium Yellow	a_036

Table 1. (cont.) Description of the in-lab prepared blends included in the sample bank.

Sample code	Commercial category	Fake commercial category	Colour*	Mixed with
A_026_25d	VOO	EVOO	Yellow	d_039
A_026_50d	VOO	EVOO	Yellow	d_039
A_027_25e	VOO	EVOO	Yellow	e_040
A_027_50e	EVOO	EVOO	Yellow	e_040
A_028_25a	VOO	EVOO	Yellow	a_036
A_028_50a	VOO	EVOO	Yellow	a_036
A_029_25b	VOO	EVOO	Yellow	b_037
A_029_50b	VOO	EVOO	Yellow	b_037
A_030_25b	VOO	EVOO	Yellow	b_037
A_030_50b	VOO	EVOO	Yellow	b_037
A_031_25c	VOO	EVOO	Yellow	c_038
A_031_50c	VOO	EVOO	Yellow	c_038
A_032_25d	VOO	EVOO	Yellow	d_039
A_032_50d	VOO	EVOO	Yellow	d_039
A_033_25e	OO	EVOO	Yellow	e_040
A_033_50e	OO	EVOO	Yellow	e_040
A_034_25a	VOO	EVOO	Yellow	a_036
A_034_50a	VOO	EVOO	Yellow	a_036
A_035_25b	VOO	EVOO	Yellow	b_037
A_035_50b	VOO	EVOO	Yellow	b_037

* This color category was assigned supposing that the OO packer do not show any evidence of the fraud.

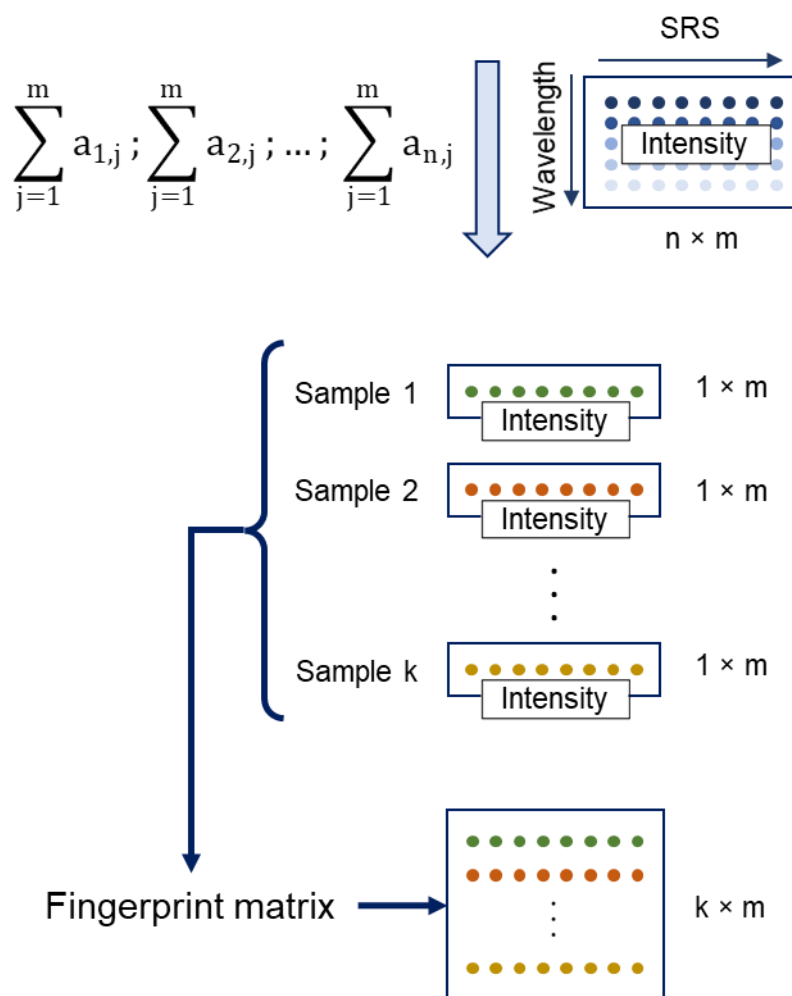


Figure SM1. Schematic procedure for the variable reduction and fingerprint matrices development.

		11	24	35
Asignation	Not Recognized	7 (20,00%)	1 (0,03%)	8
	Inconclusive	0	0	0
	Class 2 (Not Green)	0	23 (65,71%)	23
	Class 1 (Green)	4 (11,43%)	0	4
		Class 1	Class 2	
		Actual		

Figure SM2. Contingency values for the SIMCA classification model using the 25% blended 1D URIA-FP matrix as external prediction set.

		7	28	35
Asignation	Not Recognized	7 (20,00%)	5 (14,29%)	12
	Inconclusive	0	0	0
	Class 2 (Not Green)	0	19 (54,29%)	19
	Class 1 (Green)	0	4 (11,43%)	4
		Class 1	Class 2	
		Actual		

Figure SM3. Contingency values for the SIMCA classification model using the 50% blended 1D URIA-FP matrix as external prediction set.

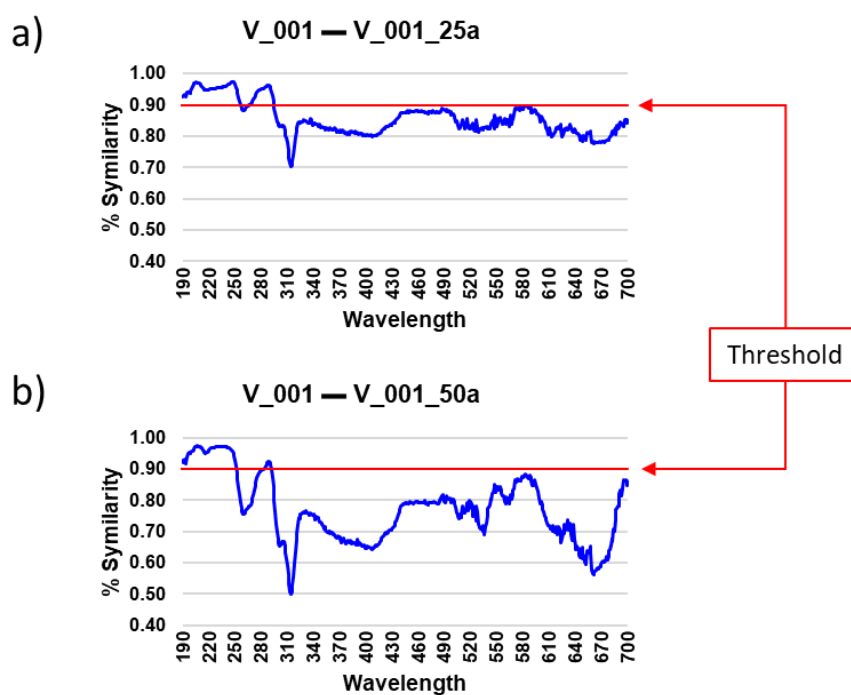


Figure SM4. Similarity profile plots obtained by comparing marketed EVOO sample V_001 with: a) blended VOO sample V_001_25a; b) blended VOO sample V_001_50a.

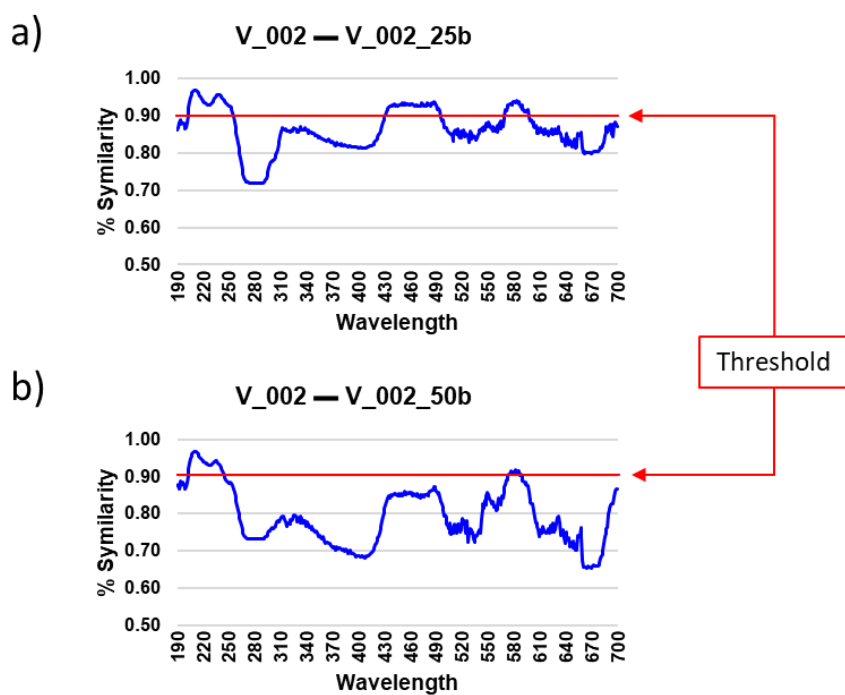


Figure SM5. Similarity profile plots obtained by comparing marketed EVOO sample V_002 with: a) blended VOO sample V_002_25b; b) blended VOO sample V_002_50b.

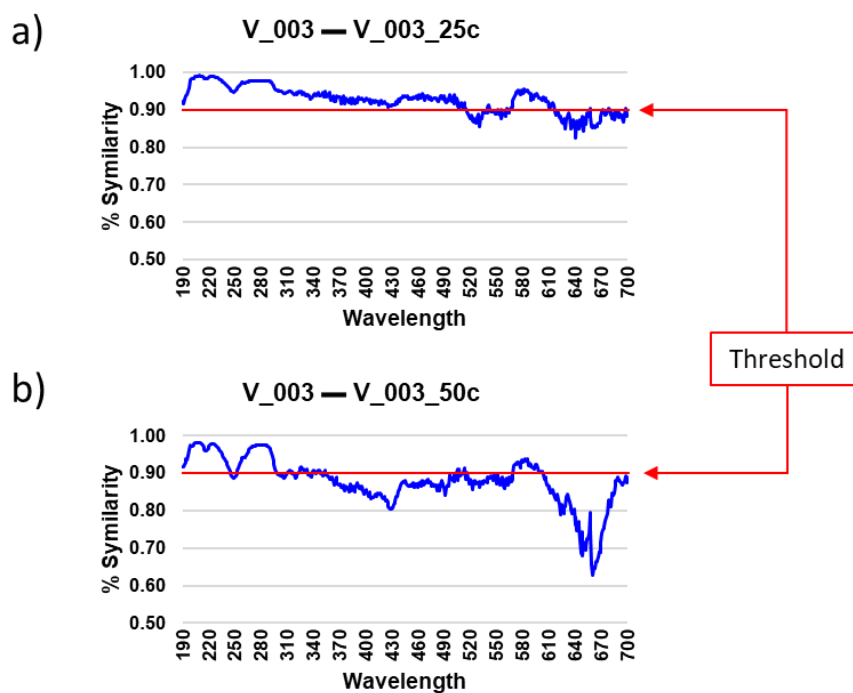


Figure SM6. Similarity profile plots obtained by comparing marketed EVOO sample V_003 with: a) blended VOO sample V_003_25c; b) blended VOO sample V_003_50c.

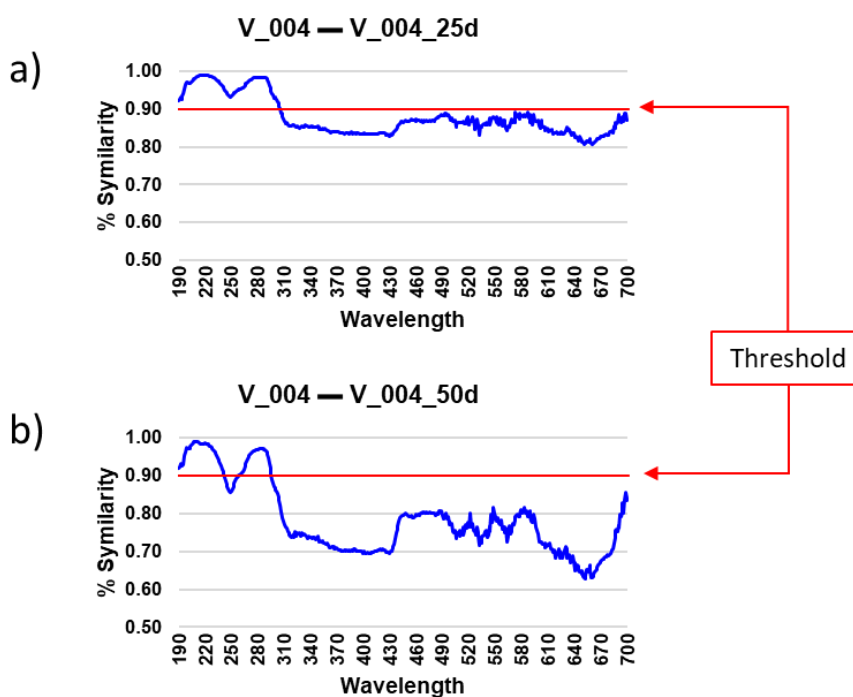


Figure SM7. Similarity profile plots obtained by comparing marketed EVOO sample V_004 with: a) blended VOO sample V_004_25d; b) blended VOO sample V_004_50d.

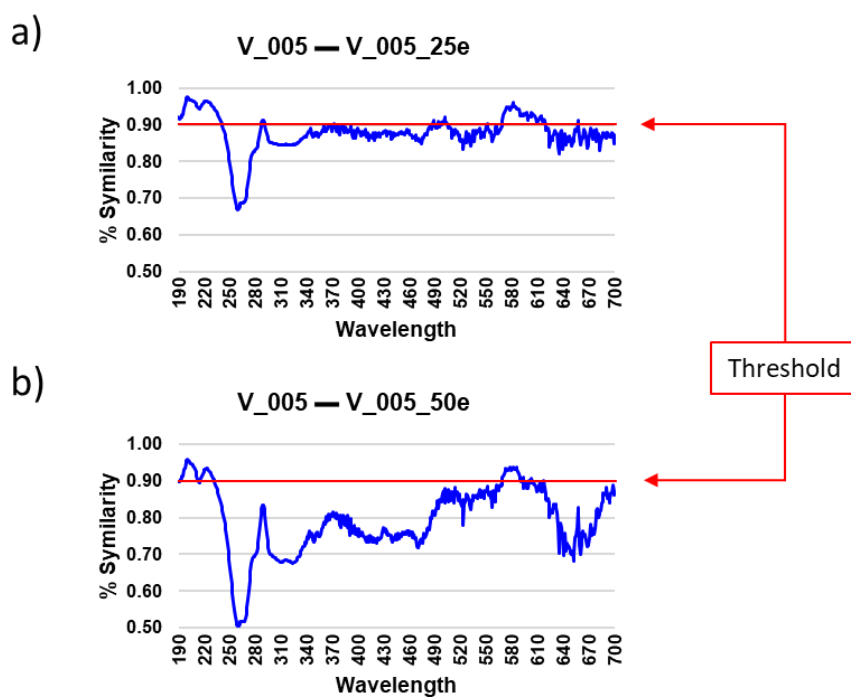


Figure SM8. Similarity profile plots obtained by comparing marketed EVOO sample V_005 with: a) blended OO sample V_005_25e; b) blended OO sample V_005_50e.

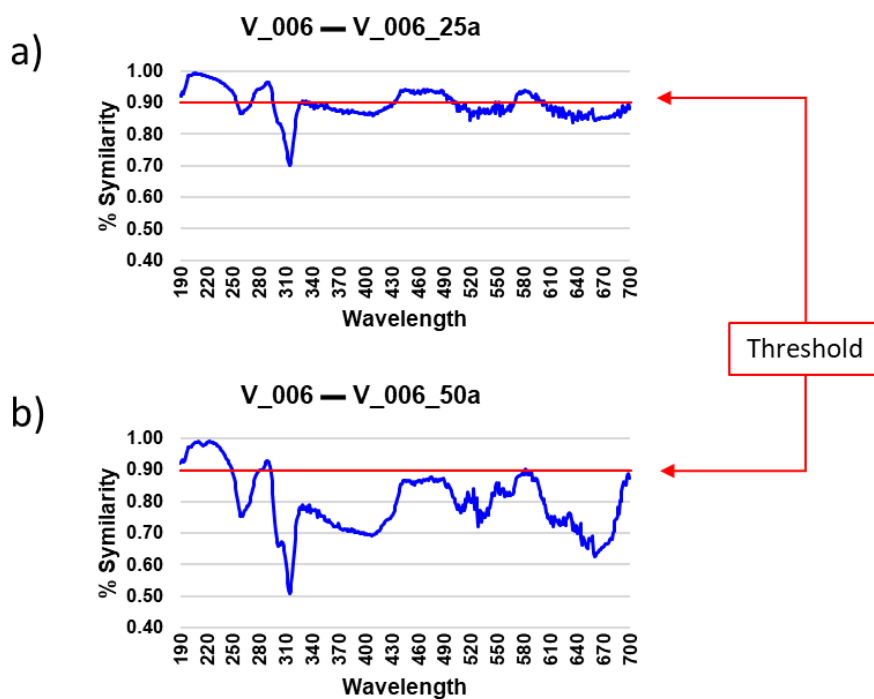


Figure SM9. Similarity profile plots obtained by comparing marketed EVOO sample V_006 with: a) blended VOO sample V_006_25a; b) blended VOO sample V_006_50a.

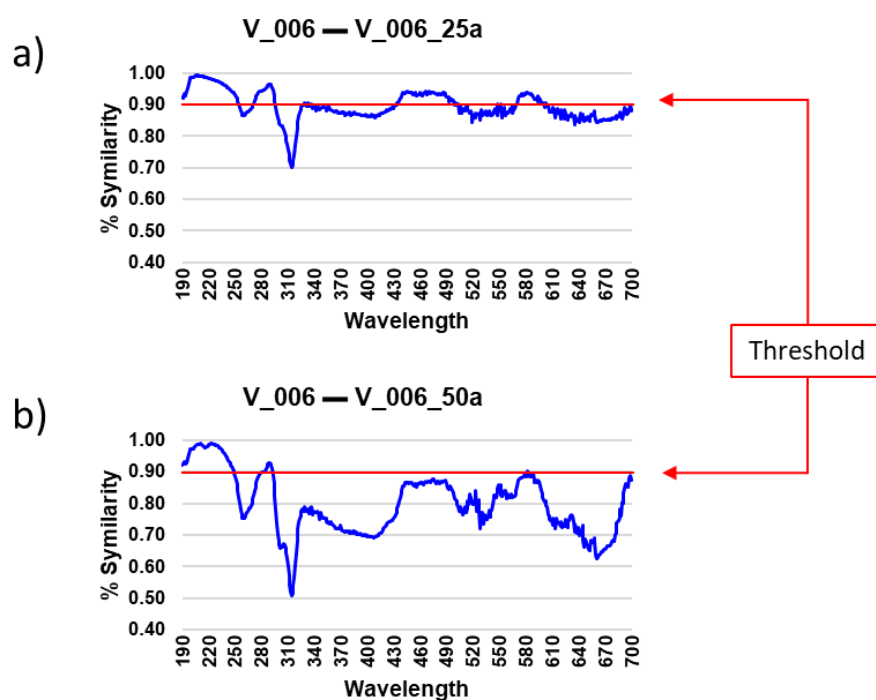


Figure SM9. Similarity profile plots obtained by comparing marketed EVOO sample V_006 with: a) blended VOO sample V_006_25a; b) blended VOO sample V_006_50a.

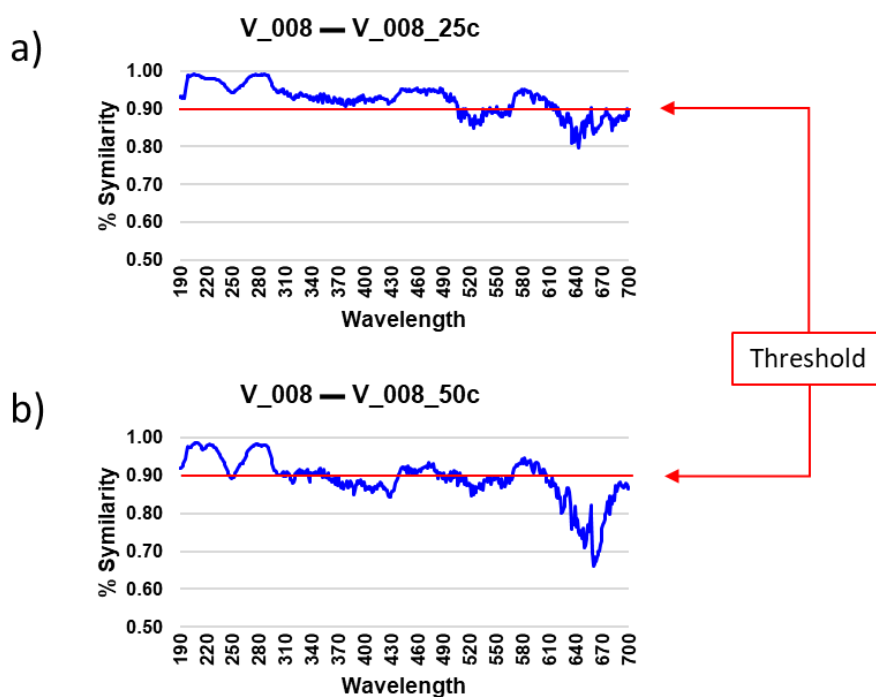


Figure SM11. Similarity profile plots obtained by comparing marketed EVOO sample V_008 with: a) blended VOO sample V_008_25c; b) blended VOO sample V_008_50c.

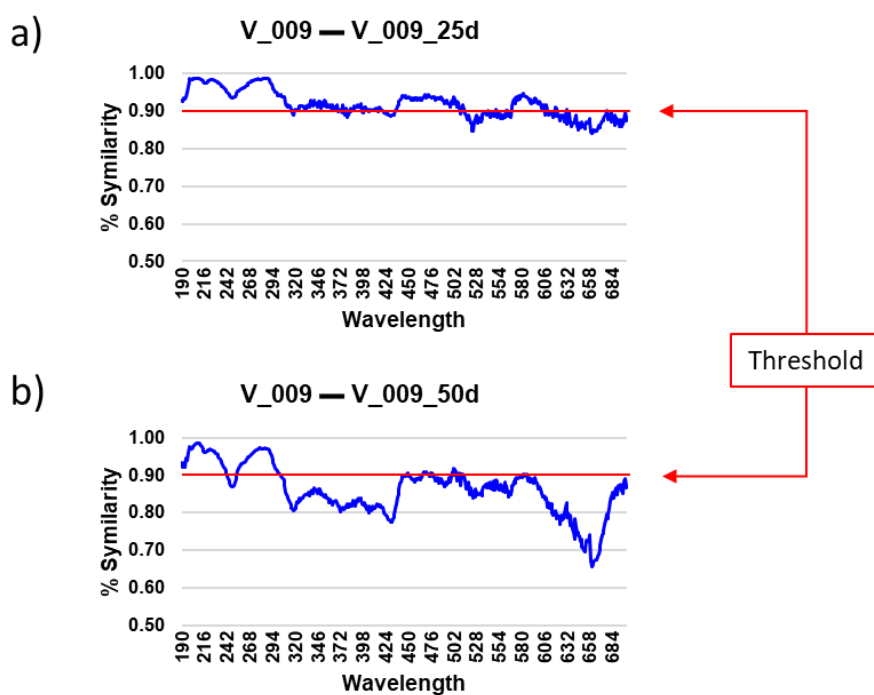


Figure SM12. Similarity profile plots obtained by comparing marketed EVOO sample V_009 with: a) blended VOO sample V_009_25d; b) blended VOO sample V_009_50d.

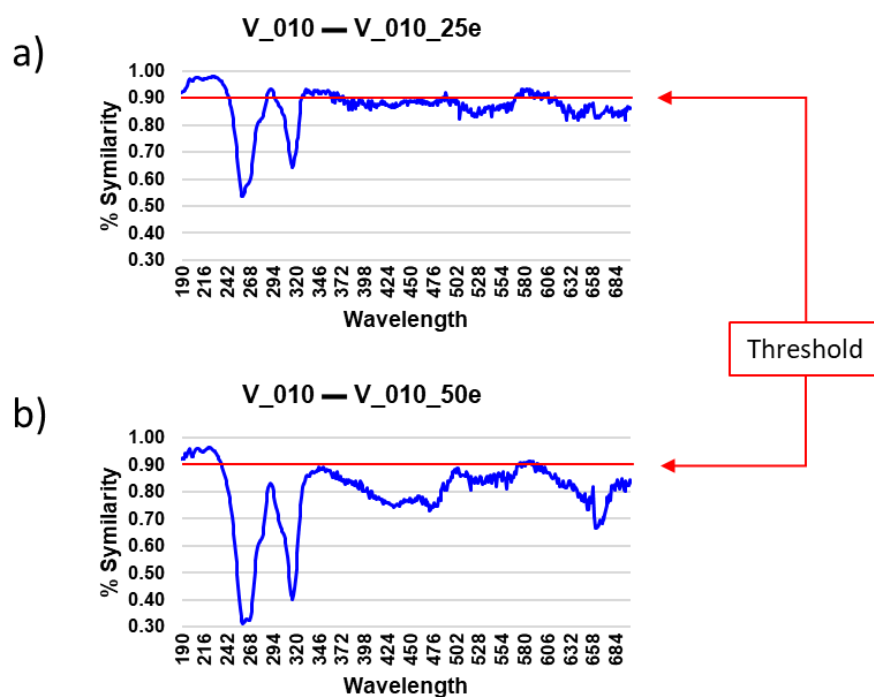


Figure SM13. Similarity profile plots obtained by comparing marketed EVOO sample V_010 with: a) blended OO sample V_010_25e; b) blended OO sample V_010_50e.

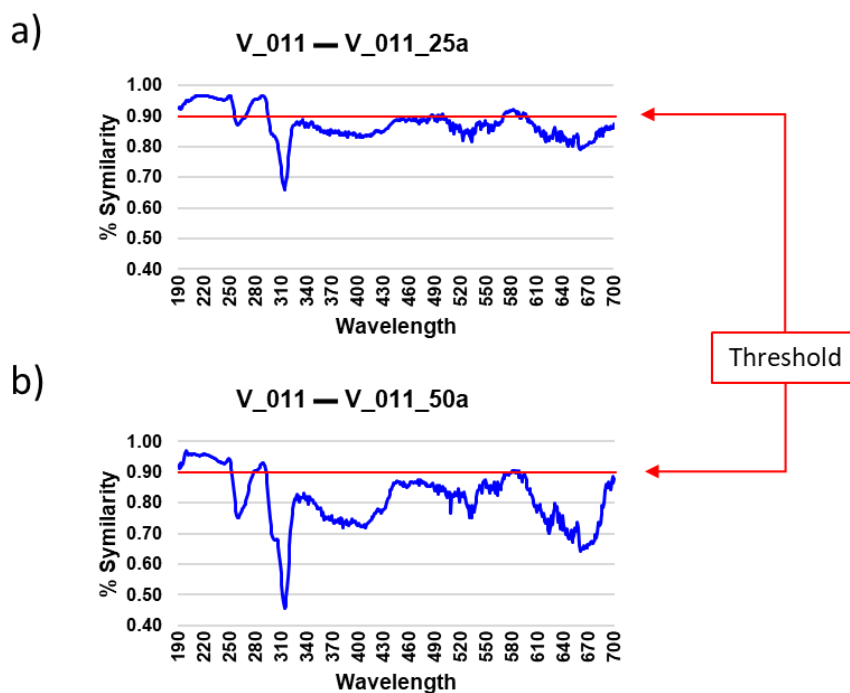


Figure SM14. Similarity profile plots obtained by comparing marketed EVOO sample V_011 with: a) blended VOO sample V_011_25a; b) blended VOO sample V_011_50a.

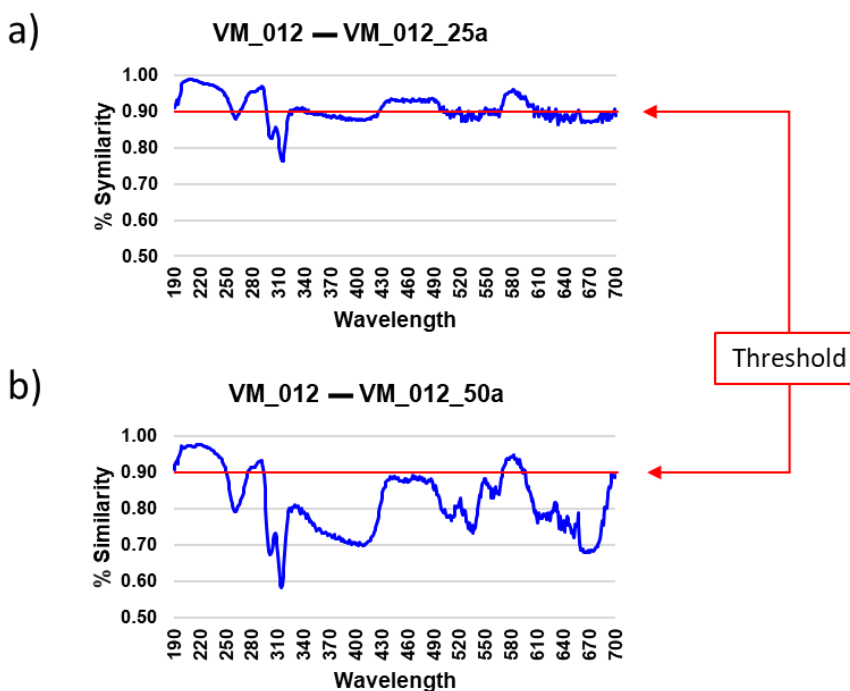


Figure SM15. Similarity profile plots obtained by comparing marketed EVOO sample VM_012 with: a) blended VOO sample VM_012_25a; b) blended VOO sample VM_012_50a.

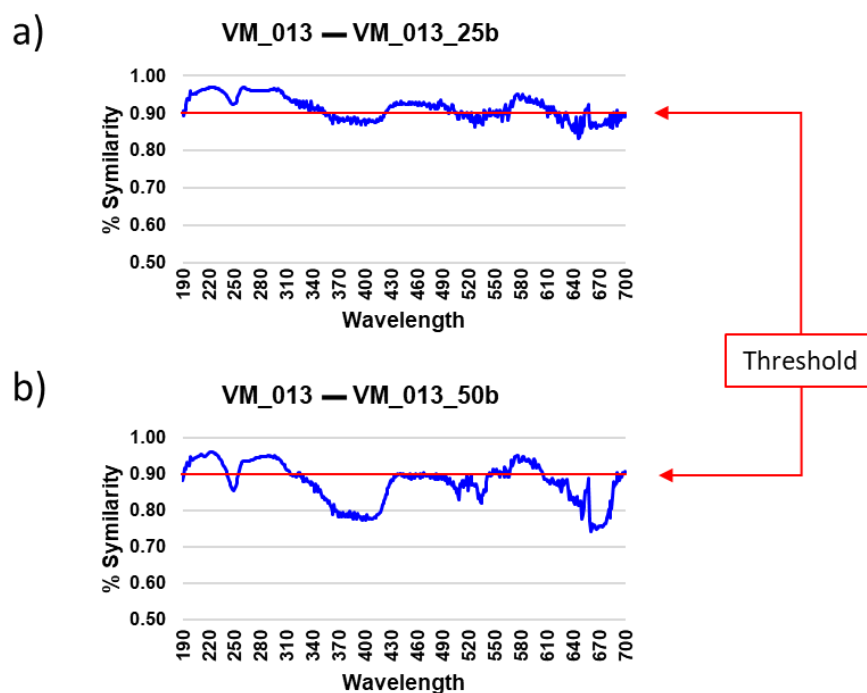


Figure SM16. Similarity profile plots obtained by comparing marketed EVOO sample VM_013 with: a) blended VOO sample VM_013_25b; b) blended VOO sample VM_013_50b.

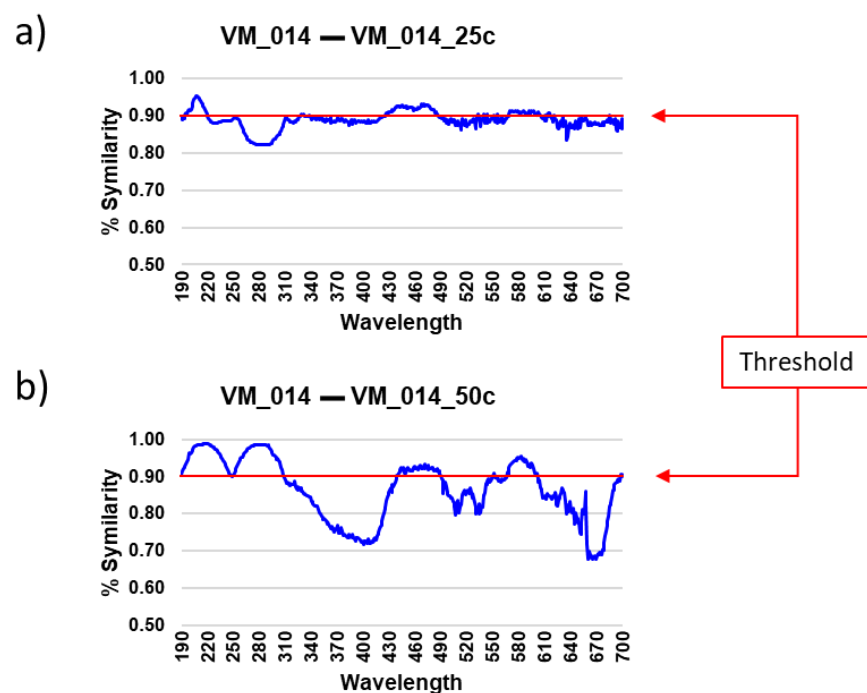


Figure SM17. Similarity profile plots obtained by comparing marketed EVOO sample VM_014 with: a) blended VOO sample VM_014_25c; b) blended VOO sample VM_014_50c.

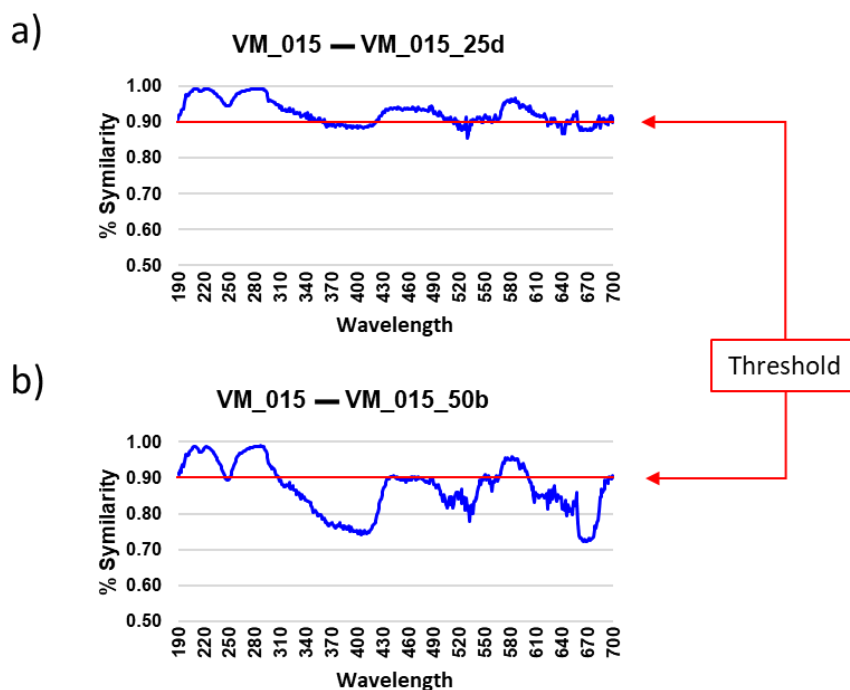


Figure SM18. Similarity profile plots obtained by comparing marketed EVOO sample VM_015 with: a) blended VOO sample VM_015_25d; b) blended VOO sample VM_015_50d.

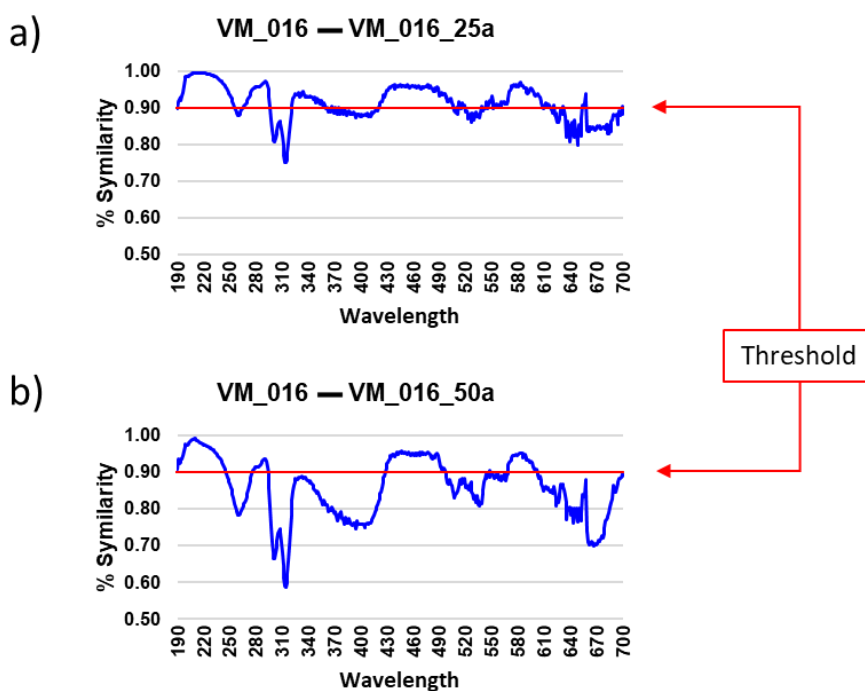


Figure SM19. Similarity profile plots obtained by comparing marketed EVOO sample VM_016 with: a) blended VOO sample VM_016_25a; b) blended VOO sample VM_016_50a.

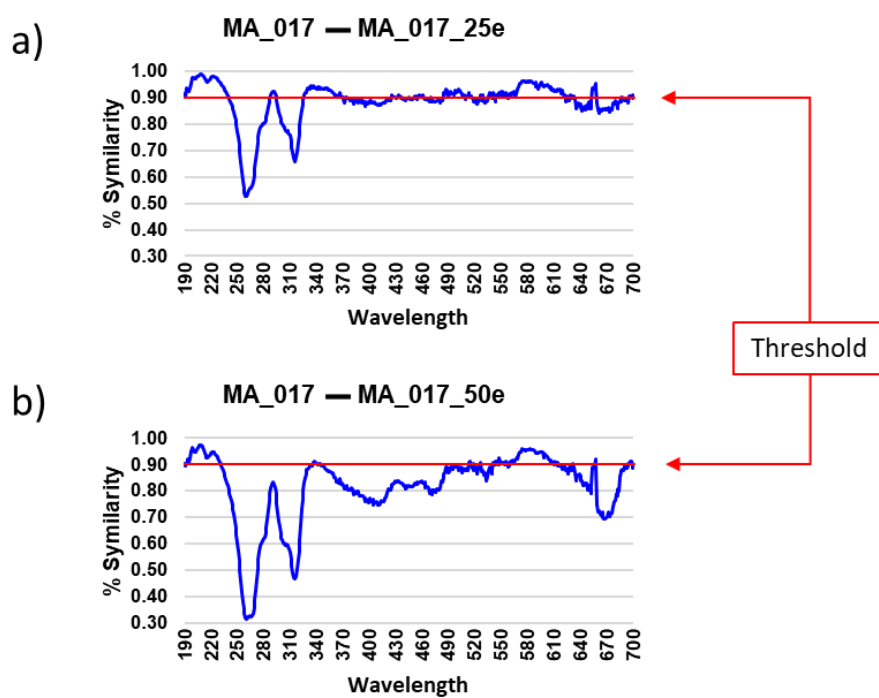


Figure SM20. Similarity profile plots obtained by comparing marketed EVOO sample MA_017 with: a) blended OO sample MA_017_25e; b) blended OO sample MA_017_50e.

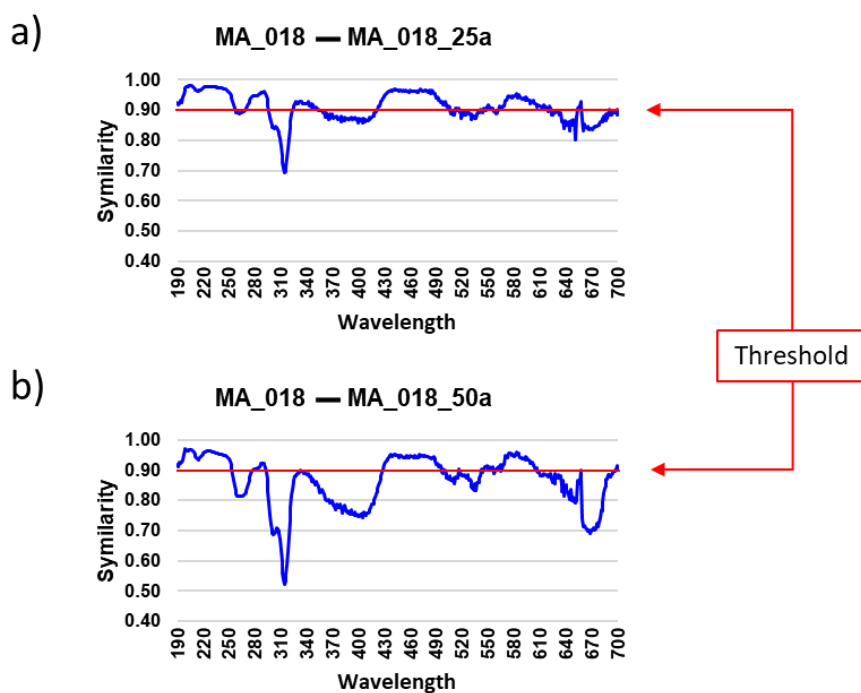


Figure SM21. Similarity profile plots obtained by comparing marketed EVOO sample MA_018 with: a) blended VOO sample MA_018_25a; b) blended VOO sample MA_018_50a.

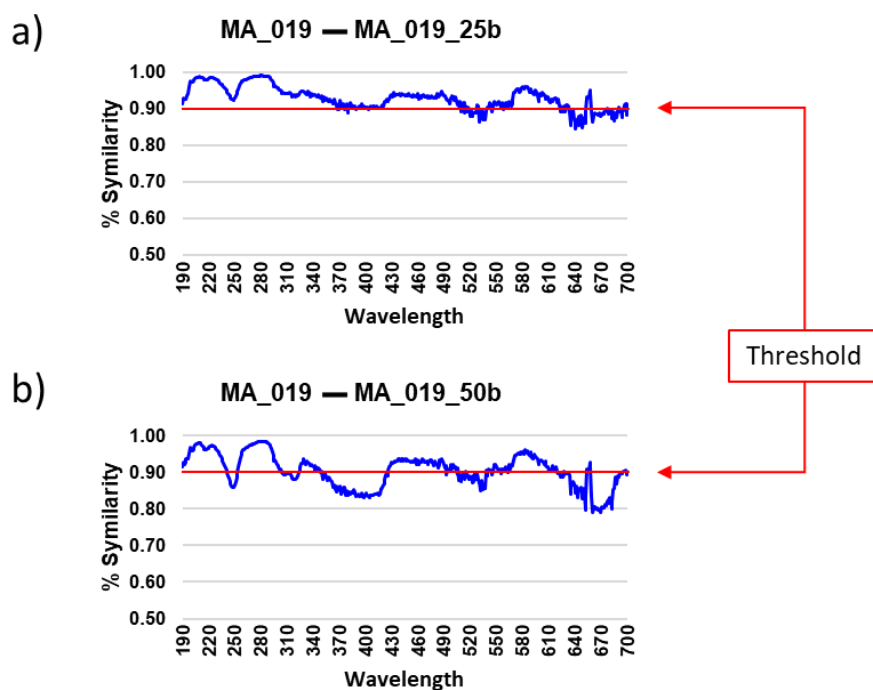


Figure SM22. Similarity profile plots obtained by comparing marketed EVOO sample MA_019 with: a) blended VOO sample MA_019_25b; b) blended VOO sample MA_019_50b.

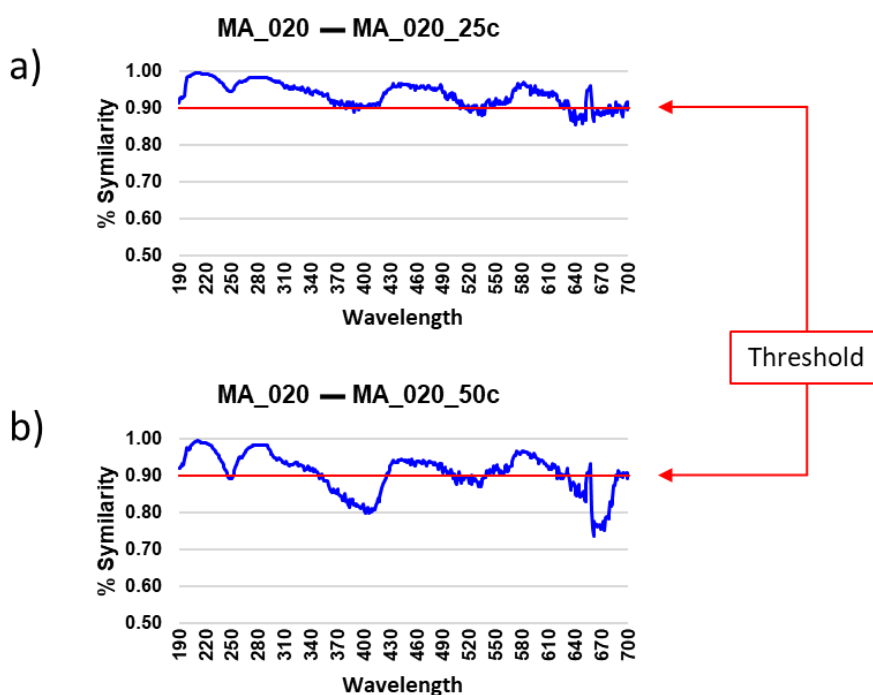


Figure SM23. Similarity profile plots obtained by comparing marketed EVOO sample MA_020 with: a) blended VOO sample MA_020_25c; b) blended VOO sample MA_020_50c.

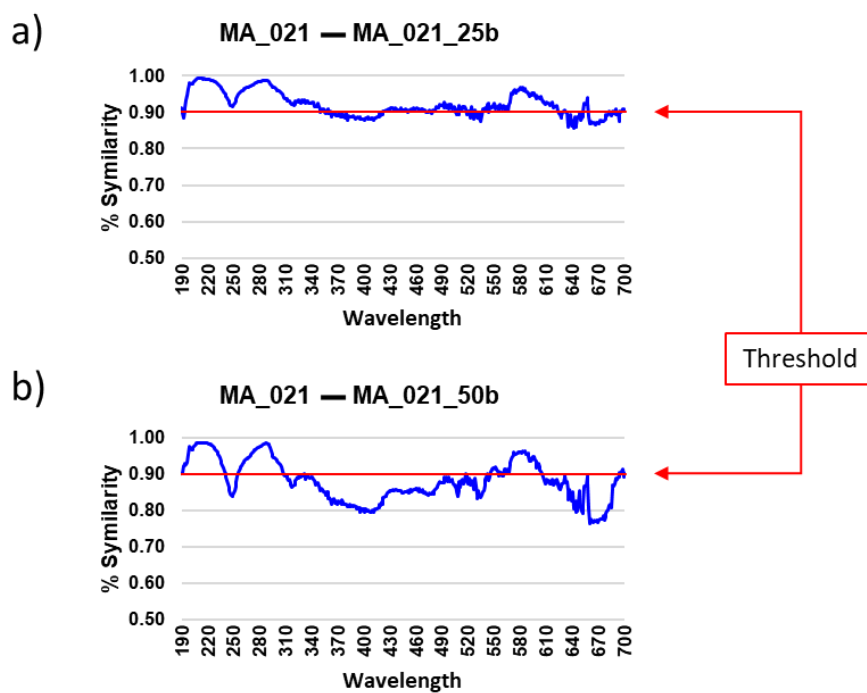


Figure SM24. Similarity profile plots obtained by comparing marketed EVOO sample MA_021 with: a) blended VOO sample MA_021_25b; b) blended VOO sample MA_021_50b.

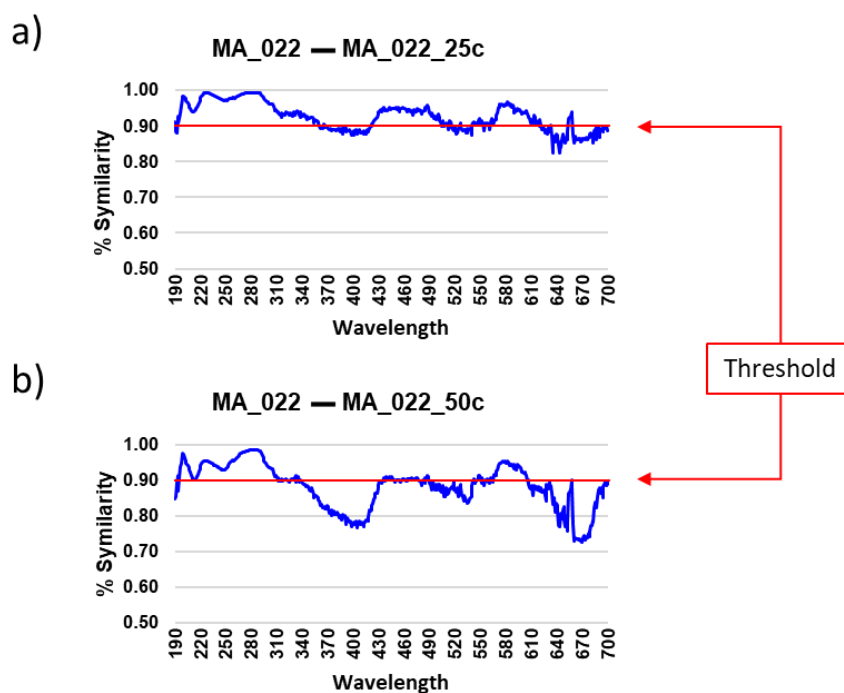


Figure SM25. Similarity profile plots obtained by comparing marketed EVOO sample MA_022 with: a) blended VOO sample MA_022_25c; b) blended VOO sample MA_022_50c.

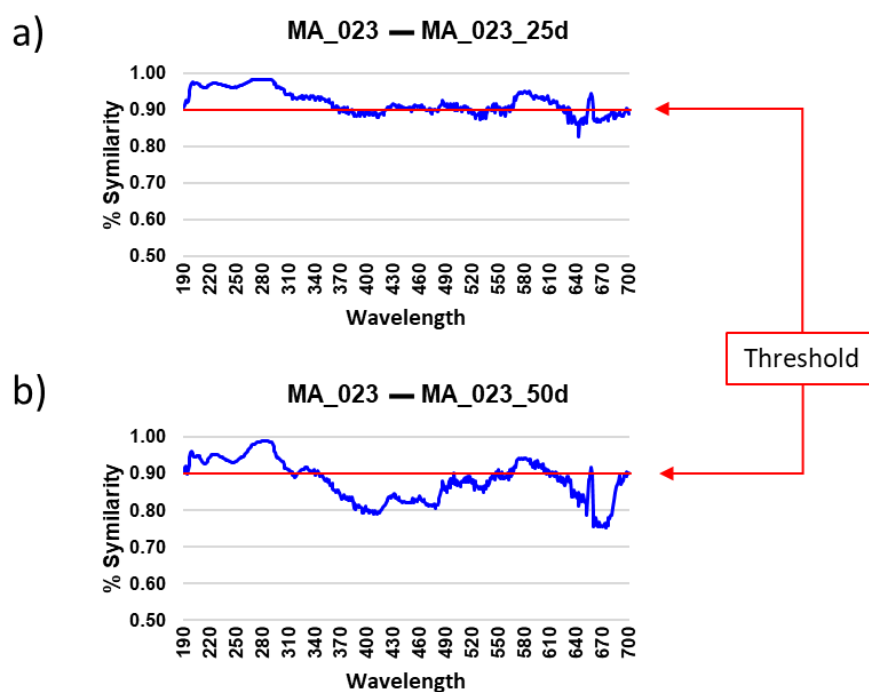


Figure SM26. Similarity profile plots obtained by comparing marketed EVOO sample MA_023 with: a) blended VOO sample MA_023_25d; b) blended VOO sample MA_023_50d.

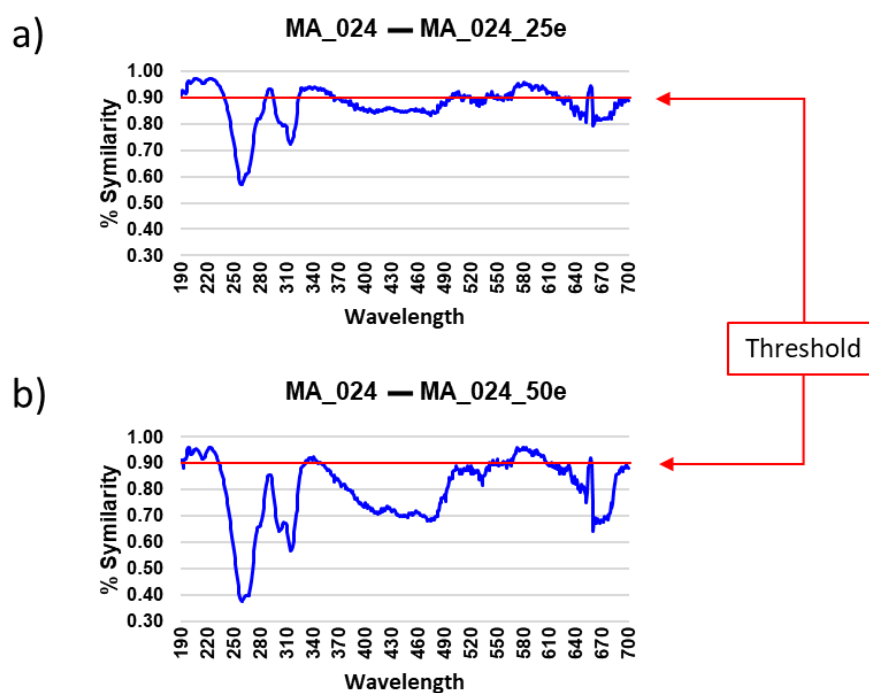


Figure SM27. Similarity profile plots obtained by comparing marketed EVOO sample MA_024 with: a) blended OO sample MA_024_25e; b) blended OO sample MA_024_50e.

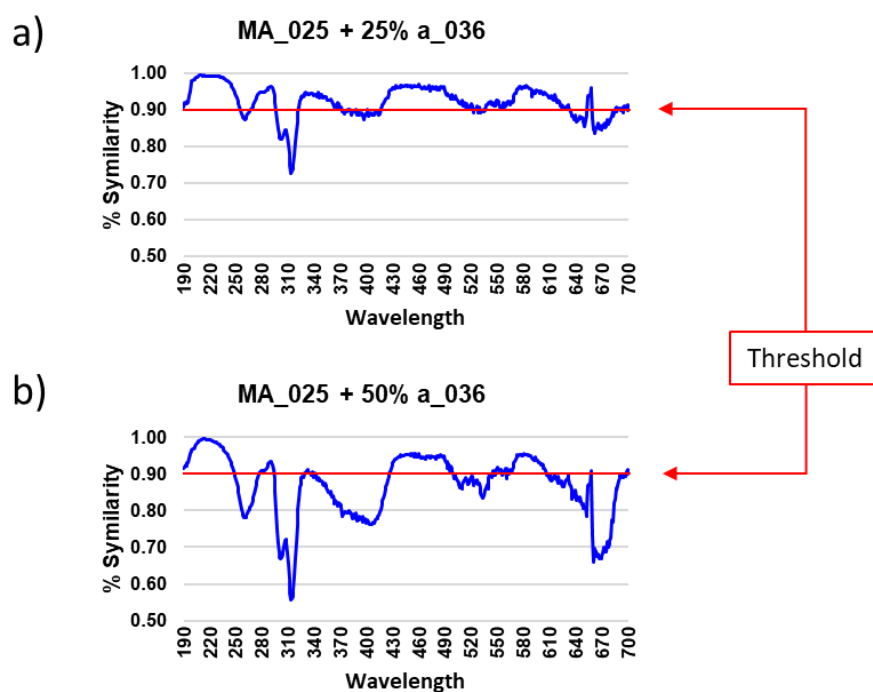


Figure SM28. Similarity profile plots obtained by comparing marketed EVOO sample MA_025 with: a) blended VOO sample MA_025_25a; b) blended VOO sample MA_025_50a.

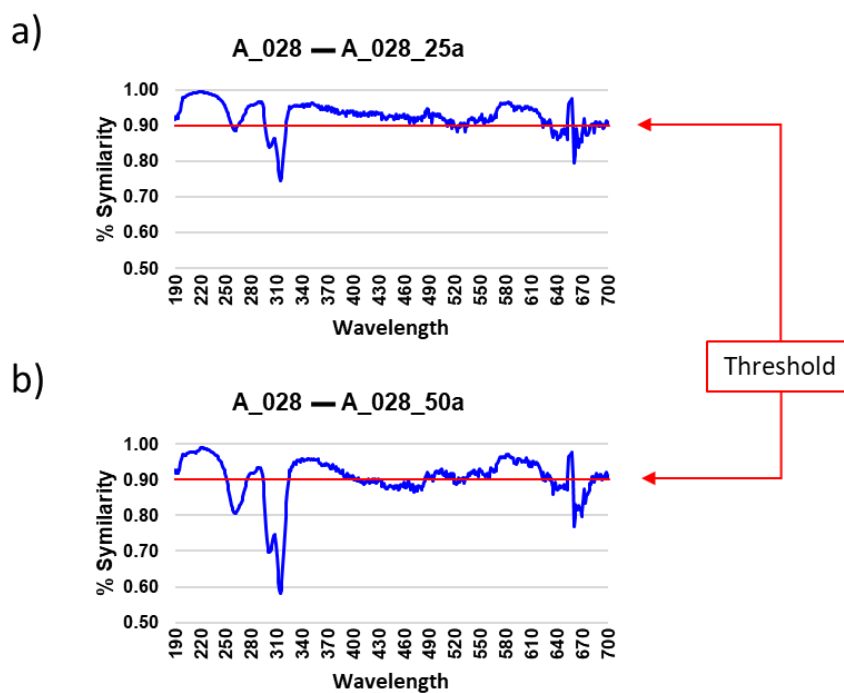


Figure SM29. Similarity profile plots obtained by comparing marketed EVOO sample A_028 with: a) blended VOO sample A_028_25a; b) blended VOO sample A_028_50a.

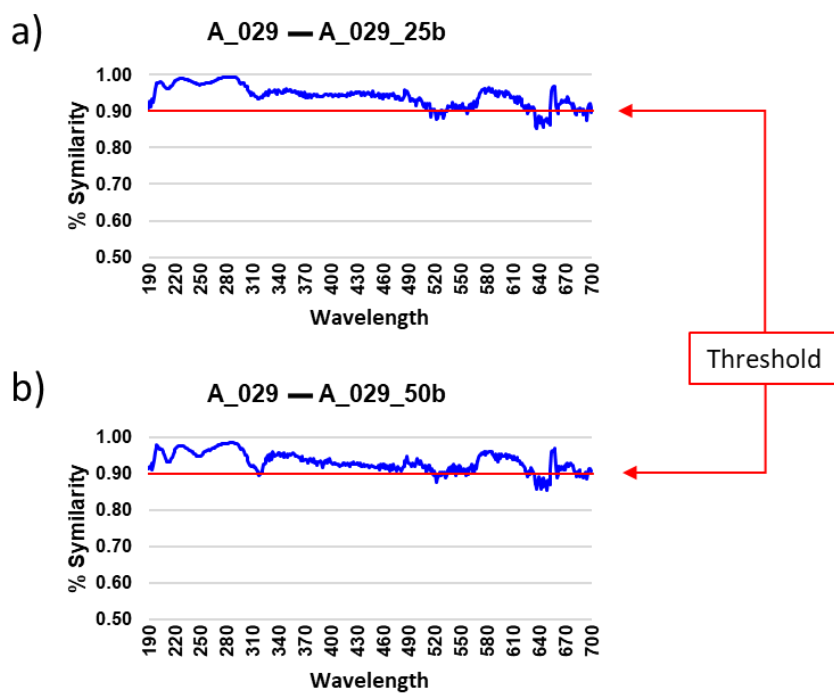


Figure SM30. Similarity profile plots obtained by comparing marketed EVOO sample A_029 with: a) blended VOO sample A_029_25b; b) blended VOO sample A_029_50b.

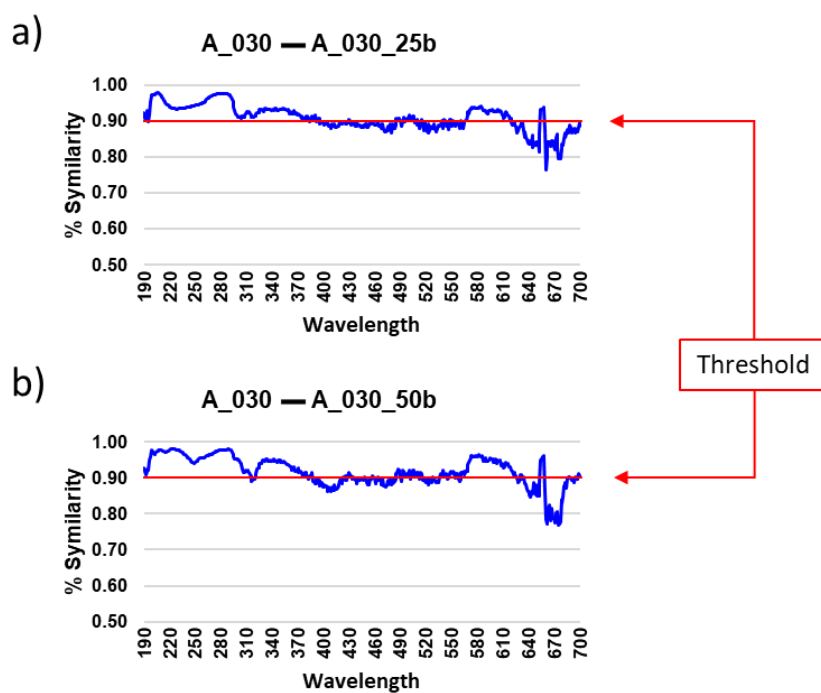


Figure SM31. Similarity profile plots obtained by comparing marketed EVOO sample A_030 with: a) blended VOO sample A_030_25b; b) blended VOO sample A_030_50b.

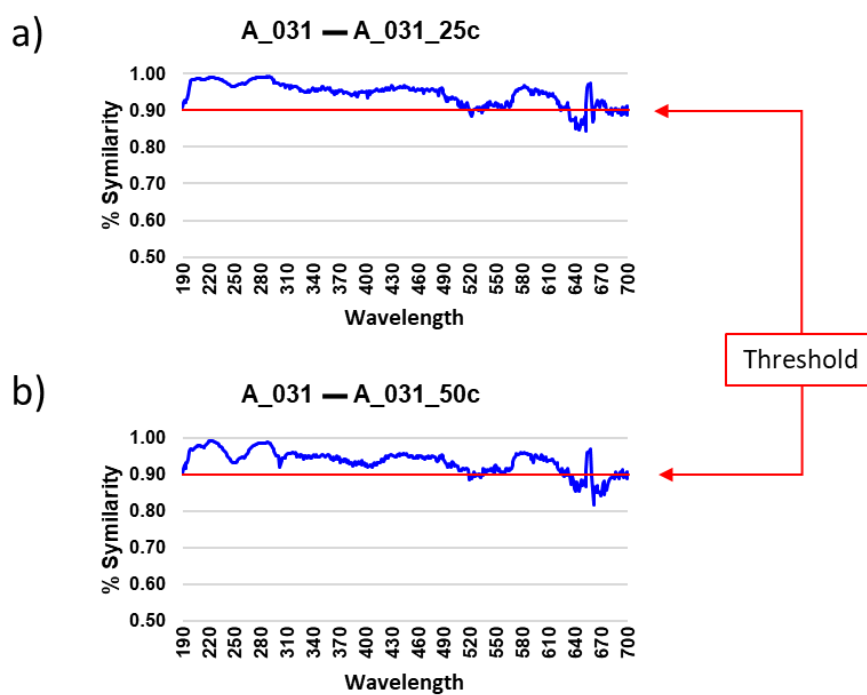


Figure SM32. Similarity profile plots obtained by comparing marketed EVOO sample A_031 with: a) blended VOO sample A_031_25c; b) blended VOO sample A_031_50c.

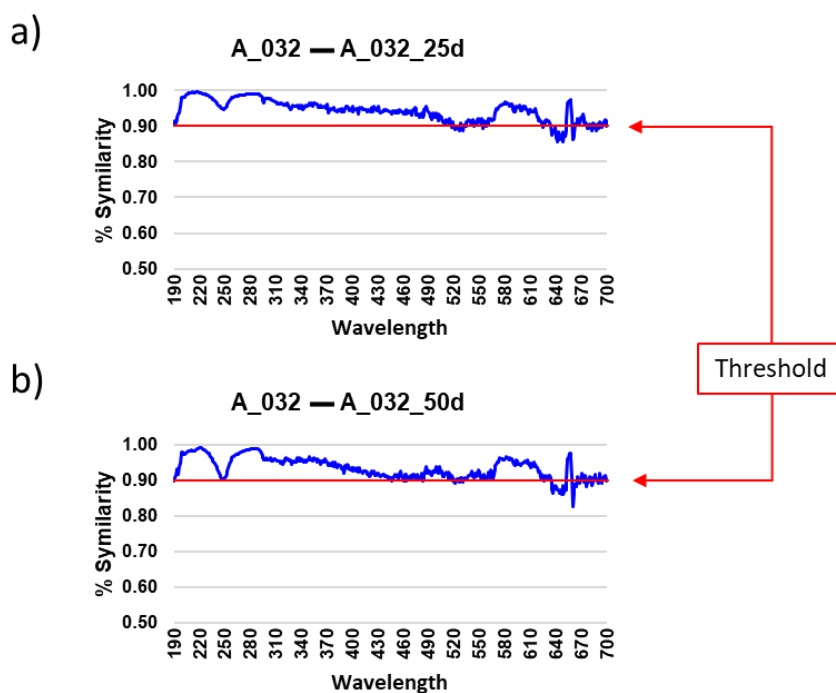


Figure SM33. Similarity profile plots obtained by comparing marketed EVOO sample A_032 with: a) blended VOO sample A_032_25d; b) blended VOO sample A_032_50d.

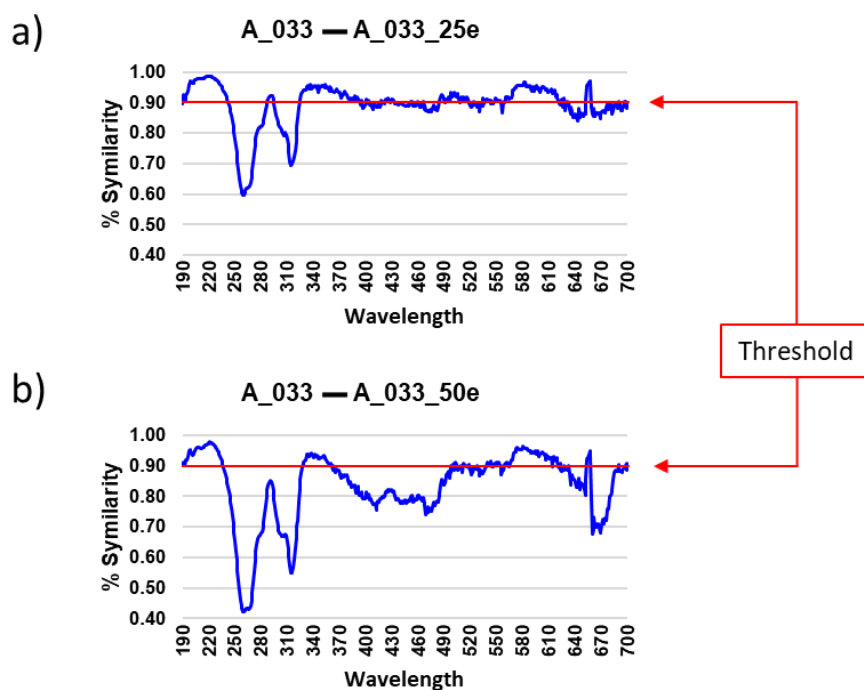


Figure SM34. Similarity profile plots obtained by comparing marketed EVOO sample A_033 with: a) blended OO sample A_033_25e; b) blended OO sample A_033_50e.

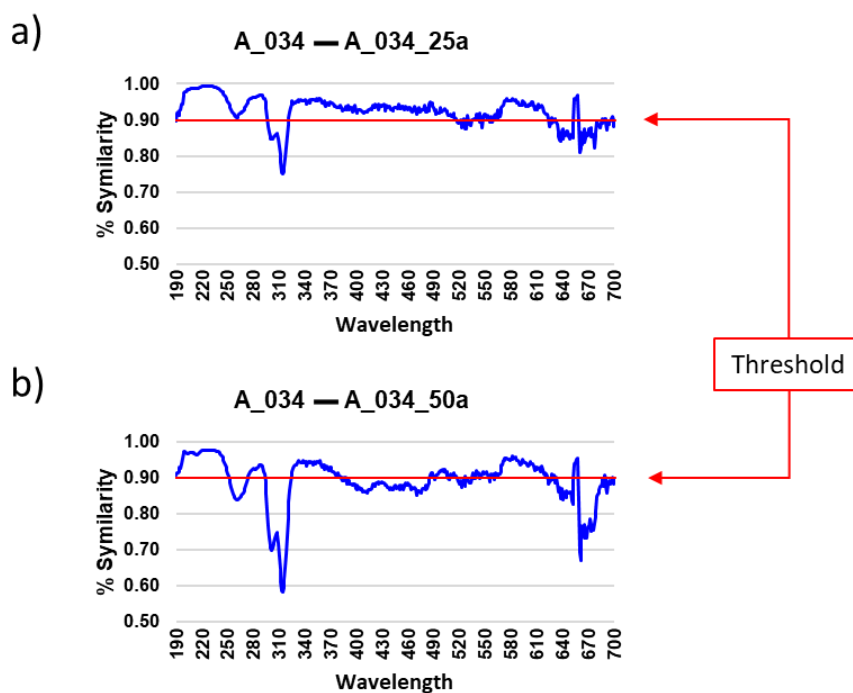


Figure SM35. Similarity profile plots obtained by comparing marketed EVOO sample A_034 with: a) blended VOO sample A_034_25a; b) blended VOO sample A_034_50a.

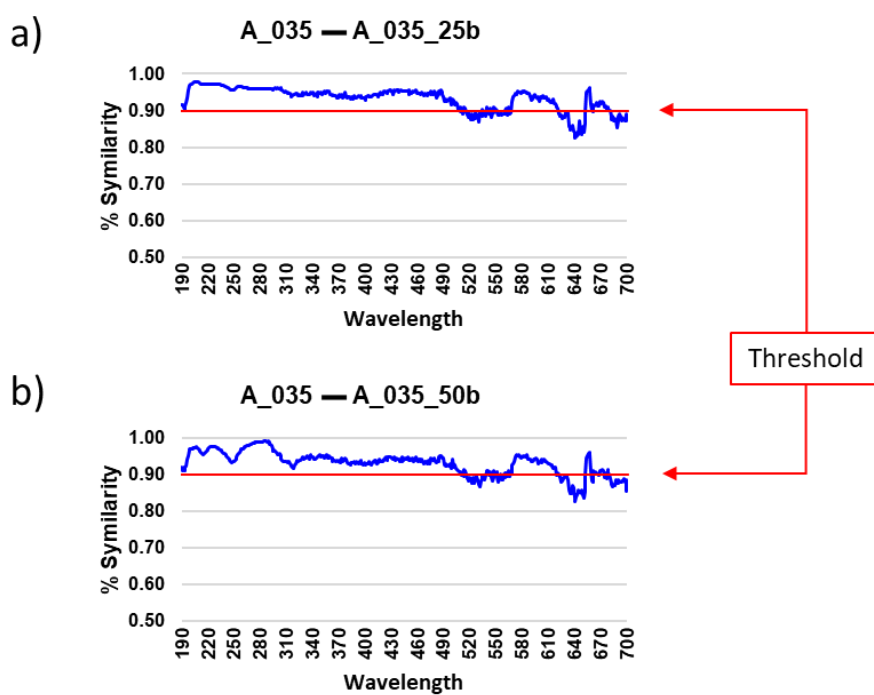


Figure SM36. Similarity profile plots obtained by comparing marketed EVOO sample A_035 with: a) blended VOO sample A_035_25b; b) blended VOO sample A_035_50b.

REFERENCES

- [1] W. C. Willett, F. Scacks, A. Trichopoulou, G. Drescher, A. Ferro Luzzi, E. Helsing, D. Trichopoulos, Mediterranean diet pyramid: a cultural modern for healthy eating, *Am. J. Clin. Nutr.*, 61 (1995), 1402S–1406S.
<https://doi.org/10.1093/ajcn/61.6.1402S>
- [2] J.L. Barjol, Introduction, in: R. Aparicio, J. Harwood (Eds.), *Handbook of olive oil: analysis and properties* (2nd ed.), Springer, New York, 2013, pp. 1–15.
https://doi.org/10.1007/978-1-4614-7777-8_1
- [3] M. Mariotti, Virgin olive oil: definition and standards, in: C. Peri (Ed.), *The extra-virgin olive oil handbook*, Wiley, Chichester, 2014, pp. 11–19.
<https://doi.org/10.1002/9781118460412.ch2>
- [4] C. Peri, The extra-virgin olive oil chain, in: C. Peri (Ed.), *The extra-virgin olive oil handbook*, Wiley, Chichester, 2014, pp. 3–10.
<https://doi.org/10.1002/9781118460412.ch1>
- [5] E. Bertolotti, The production cost of extra-virgin olive oil, in: C. Peri (Ed.), *The extra-virgin olive oil handbook*, Wiley, Chichester, 2014, pp. 303–319.
<https://doi.org/10.1002/9781118460412.ch23>
- [6] E. Casadei, E. Valli, F. Panni, J. Donarski, J. F. Gubern, P. Lucci, L. Conte, F. Lacoste, A. Maquet, P. Brereton, A. Bendini, T. G. Toschi, Emerging trends in olive oil fraud and possible countermeasures, *Food Control*, 124 (2021), 107902.
<https://doi.org/10.1016/j.foodcont.2021.107902>
- [7] Commission Implementing Regulation (EU) No 29/2012 of 13 January 2012 on marketing standards for olive oil, *Official Journal of the European Union* (2012), L 12/14–12/21 (consolidated version 02012R0029 – EN – 06.02.2019 – 007.001).
<https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02012R0029-20190206&from=EN>
- [8] Commission Regulation (EEC) No 2568/91 on the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis. *Official Journal of the European Communities*, (1991) L 248/1 -248/83 (consolidated version 01991R2568 – EN – 20.10.2019 – 032.001).
<https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:01991R2568-20191020&rid=2>
- [9] Real Decreto 760/2021, de 31 de agosto, por el que se aprueba la norma de calidad de los aceites de oliva y de orujo de oliva, *Boletín Oficial del Estado*, 209 (2021), 106518–106532 (in Spanish).
<https://www.boe.es/eli/es/rd/2021/08/31/760>

-
- [10] Y. Li, T. Fang, S. Zhu, F. Huang, Z. Chen, Y. Wang, Detection of olive oil adulteration with waste cooking oil via Raman spectroscopy combined with iPLS and SiPLS, *Spectrochim. Acta. A Mol. Biomol. Spectrosc.*, 189 (2018), 37–43.
<http://dx.doi.org/10.1016/j.saa.2017.06.049>
- [11] A.M. Jiménez Carvelo, M.T. Osorio, A. Koidis, A. González Casado, L. Cuadros Rodríguez, Chemometric classification and quantification of olive oil in blends with any edible vegetable oils using FTIR-ATR and Raman spectroscopy, *LWT-Food Sci. Technol.*, 86 (2017), 174-184.
<http://dx.doi.org/10.1016/j.lwt.2017.07.050>
- [12] M. Meenu, Q. Cai, B. Xu, A critical review on analytical techniques to detect adulteration of extra virgin olive oil, *Trends Food. Sci. Technol.*, 91 (2019), 391–408.
<https://doi.org/10.1016/j.tifs.2019.07.045>
- [13] R. Aparicio Ruiz, I. Romero, D.L. García González, C. Oliver Pozo, R. Aparicio, Soft-deodorization of virgin olive oil: Study of the changes of quality and chemical composition, *Food Chem.*, 220 (2017), 42–50.
<http://dx.doi.org/10.1016/j.foodchem.2016.09.176>
- [14] Y. Yang, M. Duarte Ferrero, I. Cavaco, Y. Liang, Detection and identification of extra virgin olive oil adulteration by GC-MS combined with chemometrics, *J. Agric. Food Chem.*, 61 (2013), 3693–3702.
<http://dx.doi.org/10.1021/jf4000538>
- [15] C. Ruiz Samblas, F. Marini, L. Cuadros Rodríguez, A. González Casado, Quantification of blending of olive oils and edible vegetable oils by triacylglycerol fingerprint gas chromatography and chemometric tools, *J. Chromatogr. B*, 910 (2012), 71–77
<https://doi.org/10.1016/j.jchromb.2012.01.026>
- [16] K.M. Al-Ismail, A.K. Alsaed, R. Ahmad, M. Al-Dabbas, Detection of olive oil adulteration with some plant oils by GLC analysis of sterols using polar column, *Food Chem.*, 121 (2010), 1255–1259.
<https://doi.org/10.1016/j.foodchem.2010.01.016>
- [17] S. Mildner-Szkadlarz, H.H. Jelen, The potential of different techniques for volatile compounds analysis coupled with PCA for the detection of the adulteration of olive oil with hazelnut oil, *Food Chem.*, 110 (2008), 751–761.
<https://doi.org/10.1016/j.foodchem.2008.02.053>
- [18] N. Carranco, M. Farrés Cerbián, J. Saurina, O. Núñez, Authentication and quantitation of fraud in extra virgin olive oils based on HPLC-UV fingerprinting and multivariate calibration, *Foods*, 7 (2018), 1–15.
-

- <https://doi.org/10.3390/foods7040044>
- [19] A.M. Jimenez Carvelo, A. González Casado, E. Pérez Castaño, L. Cuadros Rodríguez, Fast-HPLC fingerprinting to discriminate olive oil from other edible vegetable oils by multivariate classification methods, *J. AOAC Int.*, 100 (2017), 345–350.
<https://doi.org/10.5740/jaoacint.16-0411>
- [20] B. Gandul-Rojas, M.R.L. Cepero, I. Mínguez Mosquera, Use of chlorophyll and carotenoid pigment composition to determine authenticity of virgin olive oil, *J. Am. Oil. Chem. Soc.*, 77 (2000), 853–858.
<https://doi.org/10.1007/s11746-000-0136-z>
- [21] C.M., Seppanen, M. Rahmani, A.S. Csallany, Simultaneous determination of chlorophylls, pheophytins, β -carotene, tocopherols, and tocotrienols in olive and soybean oils by high performance liquid chromatography. *J. Food Sci.*, 68 (2003), 1644–1647.
<https://doi.org/10.1111/j.1365-2621.2003.tb12306.x>
- [22] E. Psomiadou, M. Tsimidou, Simultaneous HPLC determination of tocopherols, carotenoids, and chlorophylls for monitoring their effect on virgin olive oil oxidation. *J. Agric. Food Chem.*, 46 (1998), 5132–5138.
<https://doi.org/10.1021/jf980250n>
- [23] E. Psomiadou, M. Tsimidou, Pigments in Greek virgin olive oils: occurrence and levels, *J. Sci. Food Agric.*, 81 (2001), 640–647.
<https://doi.org/10.1002/jsfa.859>
- [24] S.C. Cunha, J.S. Amaral, J.O. Fernandes, M.B.P.P. Oliveira, Quantification of tocopherols and tocotrienols in portuguese olive oils using HPLC with three different detection systems, *J. Agric. Food Chem.*, 54 (2006), 3351–3356.
<https://doi.org/10.1021/jf053102n>
- [25] A. Giuliani, L. Cerretani, A. Cichelli, Chlorophylls in olive and in olive oil: chemistry and occurrences, *Crit. Rev. Food Sci. Nutr.*, 51 (2011), 678–690.
<https://doi.org/10.1080/10408391003768199>
- [26] P. Lucci, V. Bertoz, D. Pacetti, S. Moret, L. Conte, Effect of the refining process on total hydroxytyrosol, tyrosol, and tocopherol contents of olive oil, *Foods* 9 (2020), 292.
<https://doi.org/10.3390/foods9030292>
- [27] L. Valverde Som, C. Ruiz Samblás, F. P. Rodríguez García, L. Cuadros Rodríguez, Multivariate approaches for stability control of the olive oil reference materials for sensory analysis – Part I: Framework and fundamentals. *J. Sci. Food Agric.*, 98 (2018), 4237–4244.
<https://doi.org/10.1002/jsfa.8948>

-
- [28] L. Valverde Som, C. Ruiz Samblás, F. P. Rodríguez García, L. Cuadros Rodríguez, Multivariate approaches for stability control of the olive oil reference materials for sensory analysis – Part II: Applications. *J. Sci. Food Agric.*, 98 (2018), 4245–4252.
<https://doi.org/10.1002/jsfa.8946>
- [29] F. Ortega Gavilán, L. Valverde Som, F.P. Rodríguez García, L. Cuadros Rodríguez, M.G. Bagur González, Homogeneity assessment of reference materials for sensory analysis of liquid foodstuffs. The virgin olive oil as case study. *Food Chem.*, 322 (2020), 126743.
<https://doi.org/10.1016/j.foodchem.2020.126743>
- [30] L. Cuadros Rodríguez, F. Ortega Gavilán, S. Martín Torres, S. Medina Rodríguez, A. M. Jiménez Carvelo, A. González Casado, M. G. Bagur González, Standardization of chromatographic signals – Part I: Towards obtaining instrument-agnostic fingerprints in gas chromatography, *J. Chromatogr. A*, 1641 (2021), 461983.
<https://doi.org/10.1016/j.chroma.2021.461983>
- [31] L. Cuadros Rodríguez, S. Martín Torres, F. Ortega Gavilán, A. M. Jiménez Carvelo, R. López Ruiz, A. Garrido Frenich, M. G. Bagur González, A. González Casado, Standardization of chromatographic signals – Part II: Expanding instrument-agnostic fingerprints to reverse phase liquid chromatography, *J. Chromatogr. A*, 1641 (2021), 461973.
<https://doi.org/10.1016/j.chroma.2021.461973>
- [32] S. Bikrani, A.M. Jiménez Carvelo, M. Nechar, M.G. Bagur González, B. Souhail, L. Cuadros Rodríguez, Authentication of the geographical origin of margarines and fat-spread products from liquid chromatographic UV-absorption fingerprints and chemometrics, *Foods*, 8 (2019), 588.
<https://doi.org/10.3390/foods8110588>

Capítulo VI

**Evaluación de la calidad de la avellana con
diferentes herramientas quimiométricas**

Evaluation of different chemometric tools combined with the GC-agnostized fingerprint for the quality assessment of hazelnut

VI.a. RESUMEN (EN ESPAÑOL)

La avellana es un fruto seco que se utiliza principalmente en la industria de la confitería para la obtención de dulces y cremas de chocolate, siendo Turquía, Italia y Estados Unidos los principales productores de dicho alimento. Entre todas las variedades/orígenes conocidas, la *Tonda Gentile Trilobata* procedente del Piamonte (Italia) es altamente apreciada por la industria alimentaria debido a sus excelentes propiedades organolépticas y consideradas como "referencia Gold".

Aunque el uso de (bio)marcadores está muy extendido en la evaluación de la calidad de los alimentos, la metodología de la huella instrumental podría proporcionar más información relacionada con la calidad de la avellana, ya que hace uso de la información no visible de los propios (bio)marcadores pero que se encuentra contenida en la huella instrumental.

En este trabajo, la huella instrumental agnostizada obtenida del análisis de los compuestos orgánicos volátiles presentes en las avellanas por (HS)SPME-GC-MS se utilizó para evaluar las diferencias entre las muestras de la región del Piamonte (de variedad *Tonda Gentile Trilobata*) y otras muestras de avellanas de interés industrial de diferentes regiones de Italia y Turquía.

El análisis de componentes principales (PCA) reveló que las diferencias contenidas en la huella instrumental agnostizada se debían al envejecimiento, las condiciones de almacenamiento, el país de origen y las regiones de cultivo de cada una de las muestras analizadas. Los modelos de clasificación SIMCA, PLS-DA y SVM, aplicando diferentes tipos de pre-procesamiento, se desarrollaron para diferenciar entre las muestras de avellanas del Piamonte (clase "Gold") de las restantes muestras de avellanas (clase "Others"), obteniendo los mejores resultados en los modelos PLS-DA y SVM desarrollados.

VI.b. CONCLUSIONES DEL TRABAJO (EN ESPAÑOL)

En comparación con el uso de (bio)marcadores, la metodología de la *fingerprinting* ha demostrado su utilidad para la discriminación de la avellana en el marco de la evaluación de la calidad mediante el uso de información no cubierta por los (bio)marcadores, pero cifrada en la huella instrumental.

La aplicación del auto-escalado y centro medio como etapas independientes de preprocesamiento, combinadas con técnicas de reconocimiento de pautas no supervisadas, demostró que la huella instrumental agnostizada contenía suficiente información sobre las condiciones de envejecimiento y almacenamiento de las muestras, así como el país de origen de las mismas. Cuando se redujo el número de muestras de la matriz de huellas agnostizadas a sólo aquellas que fueron obtenidas en T0, se pudo observar cómo las muestras procedentes de Italia se sub-agrupaban según la región de origen. Estas diferencias fueron más evidentes cuando se aplicó el auto-escalado como opción de pre-procesado.

Las técnicas de reconocimiento de pautas supervisadas permitieron el desarrollo de modelos de clasificación atendiendo a la calidad de la avellana. Entre los diferentes modelos de clasificación desarrollados, el modelo PLS-DA con auto-escalado y SVM para ambas opciones de pre-procesamiento, mostraron los mejores resultados de clasificación.

Por último, el nuevo enfoque de la metodología de agnostización instrumental, en la que sólo se considera la variabilidad presente en todas las muestras, utilizando estas mismas como señal de referencia para el alineamiento, junto a técnicas quimiométricas con diferentes opciones de pre-procesado, puede considerarse como un enfoque válido en la resolución de problemas asociados a la metodología de *fingerprinting*. Sin embargo, este nuevo enfoque tan solo es válido para paliar la dependencia de la señal del tiempo de adquisición, pero no del instrumento utilizado para obtenerla. La falta de una referencia externa, pura y claramente identificada dificulta la posible transferencia de comportamiento entre diferentes instrumentos.

VI.c. ESTUDIO BIBLIOMÉTRICO

El trabajo recogido en este Capítulo recoge hace uso de diversas de procesamiento de datos y técnicas quimiométricas para evaluar la calidad de muestras de avellanas, desarrollando para ello modelos de clasificación.

Dado los objetivos finales del estudio científico como el enfoque con el que se ha abordado esta publicación hace que la categoría *Food Science and Technology* sea la más adecuada de la cual seleccionar una revista de reconocido prestigio donde publicar los resultados finales.

Dentro de esta categoría, la revista *Journal of Food Composition and Analysis* ocupa el puesto 30 de un total de 144 revistas que componen la categoría *Food Science and Technology*, por lo que esta revista se encuentra contenida en el primer cuartil (Q1).

Además, el *Journal of Food Composition and Analysis* está considerada como una revista multidisciplinar al pertenecer a una segunda categoría, en concreto la categoría *Chemistry, Applied* donde ocupa la posición 17 de un total de 74 revistas, nuevamente contenida en el primer cuartil (Q1) de la categoría.

Esta revista tiene por objetivo la publicación de trabajos originales centrados en la composición química de alimentos y los métodos empleados para su análisis. En base a esta información, el *Journal of Food Composition and Analysis* puede considerarse como una candidata idónea para la publicación de este trabajo. En fecha de entrega de esta Memoria, el trabajo contenido en este Capítulo ha sido enviado a la revista en cuestión para su evaluación o posible publicación.

VI.d. ARTÍCULO CIENTÍFICO

Evaluation of different chemometric tools combined with the GC-agnosticized fingerprint for the quality assessment of hazelnut

Fidel ORTEGA-GAVILÁN[✉] ¹, Simone SQUARA², Luis CUADROS-RODRÍGUEZ¹, M^a. Gracia BAGUR-GONZÁLEZ¹, Chiara CORDERO²

1 Department of Analytical Chemistry, Faculty of Science, University of Granada, C/ Fuentenueva s/n, E-18071, Granada (Spain)

2 Department of Drug Science and Technology, University of Turin, Via Pietro Giuria 9, I-10125, Turin, Italy

Highlights

- New instrument-agnosticizing methodology approach to obtain time-independent results
- Use of pattern recognition techniques to identify high quality hazelnuts
- Use of different pre-processing options applied in chemometric techniques

Keywords

Fingerprint methodology, hazelnut instrument-agnosticized fingerprint, hazelnut quality, pre-processing and chemometrics tools

Abstract

The hazelnut is a tree nut that is mainly used in confectionery industry to obtain sweets and chocolate pastes, being Turkey, Italy and USA the main producers of this foodstuff. Among all the known cultivars/origins, the Tonda Gentile Trilobata from Piedmont (Italy) is highly appreciated by confectionary industry due to its organoleptic properties and considered as “gold standard”.

Although the use of biomarkers is widely used in the evaluation of food quality, the fingerprinting methodology could provide more information related to hazelnut quality by

[✉] Corresponding author, e-mail: fog@ugr.es; Phone: +34 958240797.

making use of information non-visible in biomarkers but contained in the instrumental fingerprint.

In this work, the agnostised fingerprint obtained from the analysis of volatile organic compounds present in hazelnuts by (HS) SPME-GC-MS was used in order to evaluate the differences among samples from the Piedmont region (Tonda Gentile Trilobata cultivar) and other hazelnut samples of industrial interest from different regions of Italy and Turkey.

The PCA showed that the differences contained in the agnostised fingerprint were due to ageing, storage conditions, country of origin and growing regions of each sample. The SIMCA, PLS-DA and SVM models, applying different pre-processing types, were developed for the classification of hazelnut samples from Piedmont ("Gold") from the remaining hazelnut samples ("Others"), obtaining the best results for the PLS-DA and SVM models.

1. INTRODUCTION

The European hazelnut (*Corylus avellana* L.) is a tree nut that is mainly used in different processed food, primarily by the confectionery industry in the form of raw or roasted kernels and in addition to chocolate pastes. The total production in 2020 was estimated to be around 1 million tons (Food and Agriculture Organization of the United Nations, 2020), with Turkey being the most important sourcing country. Around the 62% of the global production comes from this country, where the main hazelnut production is located in the Black Sea region, followed by Italy ($\approx 6\%$, mainly located in the north western, central and southern regions) and USA ($\approx 4.5\%$, mainly produced in Oregon). From all the world cultivars, 18 different of them are produced in Turkey (Silvestri, Bacchetta, Bellincontro Cristofori, 2020) whereas more than 6 varieties are from Italy, among them the *Tonda Gentile delle Langhe* also known as *Tonda Gentile Trilobata* or *Nocciola Piemonte*, is considered by the food industry as a “Gold standard” due to its sensory properties and quality features.

For confectionery purposes, different hazelnuts cultivars/origins are usually blended together to obtain a consistent flavour profile. As different cultivars notably vary in both, morphological and hedonic characteristics, that lead to a wide flavours range, and considering the diversity in chemical composition and aroma precursors distribution (Cialì Rosso et al., 2020; Squara et al., 2022), not only the quality assessment *per se* but also the cultivars classification have a crucial repercussion in the supply chain costs. From these reasons, it is of primary importance the use of fast, accurate, informative, and objective quality control methods.

Human panels are nowadays employed to assess hazelnut quality based on International Standards (UNECE, U. S. DDP-04, 2010); testing is directed to evaluate conformity in morphological traits as calibre and shape, to exclude/count visible damage and, by sensory evaluation, detect flavour defects (e.g., stink bug damage, mouldy, rotten). These quality assessment procedures, costly and time-consuming, could take benefit of molecular resolution analytical tools that ultimately provide objective quality evaluation compatible with the growing market requests.

For quality assessment purposes, modern, objective and analytical approaches have been developed in the last years focusing on the identification of molecular (bio)markers, within the hazelnut metabolome. Caligiani et al. (2014) used nuclear magnetic resonance (NMR) spectroscopy discovering that the quali-quantitative profile of tryptophan, choline, and acetic acid were able to discriminate different cultivars as *Tonda Gentile Trilobata*,

Tonda Gentile Romana, and *Turkish* blend varieties. Cialiè Rosso et al. (2021a) used GC×GC-TOF MS to investigate primary metabolites signatures in raw and roasted *Tonda Gentile Trilobata* and *Anakliuri* cultivars farmed in Italy and Georgia. Authors identified proline, alanine, scyllo inositol, and citric acid as good discriminating (bio)markers for these cultivars.

Although the univariate approach, also considered as targeted approach, is simpler to understand and easier to apply on a routine basis, the fingerprinting methodology as an untargeted approach, has been shown to be a useful tool for quality assessment. Its best aiming arises from the possibility of extracting the comprehensive information set contained in the analytical signal that remains hidden when only target (bio)markers are used. This untargeted approach has been widely used in the quality food assessment as olive oil, cheese, beverages, honey among many others (Bagur González, Jiménez Carvelo, Ortega Gavilán & González Casado, 2021; Pérez Castaño, Sánchez Viñas, Gázquez Evangelista & Bagur González, 2018; Cevoli et. al., 201; Souard et. al, 2018; Chen, Jin, Fan & Wang, 2017).

On these bases, as the whole information contained in the instrumental fingerprint is usually represented as a data tensor, Chemometrics has been shown to be an indispensable tool in the extraction of hidden information that can be associated with parameters related to quality, variety, geographical origin among others, since it not only allows to know the natural grouping of the samples, but also it permits to establish discrimination/classification models suitable for quality control (Bagur González, 2021).

At the same time, in order to work with the most appropriate variables for each fingerprint, the pre-processing of them takes a crucial role when representative information on the variables under study wants to be obtained. All these techniques have been widely applied in different food matrices by different authors and from different types of instrumental techniques (Cuadros-Rodríguez, Ruiz-Samblás, Valverde-Som, Pérez Castaño & González-Casado, 2016; Kucharska Ambrożej & Karpinska, 2020).

Despite these advantages the instrumental fingerprint needs to be associated with specific compound families and with the instrumental technique from which it is obtained. In the case of hazelnut, the family of compounds that has generally been used applying the fingerprinting methodology is that of the volatile organic compounds (VOCs). Thus, in the last years, researchers have demonstrated that raw hazelnuts VOCs encrypts quality related signatures and those fingerprinting approaches have been able to detect and explain such information (Alasalvar et al., 2004; Squara et al., 2022).

For instance, Alasalvar et al. (2012) proved that the volatile fraction contained a specific signature for 18 native Turkish cultivars taking into account 39 compounds from several chemical classes. Cialiè Rosso et al. 2018 confirmed that by capturing the entire volatile fingerprint via high concentration capacity sampling coupled with high chromatographic resolution techniques, the natural clustering of commercial *Akçakoca*, *Ordu*, and *Tonda Gentile Romana* samples submitted to two different post-harvest drying methods was mainly driven by cultivar/geographical origin, while still being able to discriminate the different drying technology when selected VOCs were targeted. Origin-related discrimination capabilities with a fingerprinting procedure were also explored by other authors (Cordero et al. 2010, Nicolotti et al. 2013), who confirmed that cultivar differences are captured even after the roasting procedure, and that the volatile fingerprinting can be used to recognize different roasting degrees. Moreover, the volatile fraction of hazelnuts was also proven to be explanatory about spoilage and shelf-life evolutions when Stilo et al. (2021) implemented a fully automated untargeted-targeted fingerprinting approach based on GC×GC-MS to differentiate acceptable hazelnuts from samples showing perceivable off-odors.

In addition, one of the biggest challenges using fingerprinting methodology on a routine basis, is the analytical batch effect, that could include column bleed and/or the natural response fluctuation of the variables associated to analytical instrumentation, which are difficult to keep under control. In order to compensate for random errors and signal fluctuations, in this study, GC-MS fingerprints acquired over a one-year time span by a quality control procedure aiming at accurately quantify linear saturated aldehydes from raw kernels, was submitted to instrument agnostising methodology (Cuadros Rodríguez et al., 2021a & b). Results are critically evaluated and examined in light of the information potential encrypted in the VOCs fingerprint and related to samples cultivar/geographical origin, shelf life and storage conditions.

To make highly transferable the fingerprinting methodology, with possible ex-post applications on existing analytical batches, the study validates a novel approach of instrument agnostising methodology which adopts an internal standardization with selected samples traces/fingerprints instead of an external standard mixture solution as it is usually employed in the original approach. This new methodology requires a prior knowledge about the chemical composition of the sample, and it is mandatory that the selected compounds should be present in all samples used as external references, regardless their concentration.

2. MATERIALS AND METHODS

2.1. Chemicals and standard solutions

A mixture of α -thujone (95%, Merck Life Science Srl, Milan, Italy) and methyl 2-octinoate (99%, Merck Life Science Srl, Milan, Italy) of 100 mg/L for both standards in dibutyl phthalate (99%, Merck Life Science Srl, Milan, Italy), were used as internal standards (ISs).

Acetic acid, 1-butanol, hexanal, 3-methyl-4-heptanone, butyl butanoate, α -thujone, butyl benzoate and dibutyl phthalate were selected as reference compounds in the new approach of the instrument agnostizing methodology in gas chromatography. These eight compounds cover the complete operational range in which all organic volatile compounds in hazelnut samples appear.

2.2. Hazelnut samples

A total of 10 industrial batches of hazelnuts from harvest 2019 and 2020 were selected. Kernels, provided by Soremartec Srl (Alba-CN, Italy), were characterised by a homogeneous caliber (≈ 13 and 14 mm) and resulted conform to the industrial quality control procedure. Each batch of samples consists of hazelnuts from different cultivar/geographical origin, storage time and conditions as illustrated in Table 1.

Table 1. Description of different hazelnut batches used in this study.

Sample Code	Harvest	Country	Region
H21	2019	Turkey	Akçakoca
H23	2019	Turkey	Black Sea
H25	2020	Turkey	Akçakoca
H26	2020	Turkey	Black Sea
H27	2020	Italy	Piedmont
H28	2020	Italy	Lazio
H29	2020	Turkey	Akçakoca
H210	2020	Turkey	Akçakoca
H211	2020	Turkey	Black Sea
H212	2020	Turkey	Black Sea

Note. The absence of the codes H22 and H24 was due to their assignment to samples which were not covered in this study.

The different individual sample batches ($n = 10$) were obtained in accordance with the industrial sampling guidelines adopted in quality and safety controls applied to all input raw materials in the factory (Cialie Rosso et. Al., 2021b). Each sample was representative of an initial batch of 1 tonne, from which three aliquots of 20 kg were taken, subdivided into small portions for grinding.

Samples were stored for 1 year, analysed at checkpoints established at T0 (reception of samples before storage), T1 (6 months) and T2 (12 months). Over this period, two different storage conditions were used: (i) bad storage conditions (BSC) fixed at a temperature of 18 °C under air atmosphere and (ii) good storage conditions (GSC) fixed at a temperature of 5 °C in a modified atmosphere with 1% O₂ and 99% N₂. In both cases the relative humidity was kept fixed at 65%.

For each time of analysis under described storage conditions, only two of the three aliquots from each batch sample were analysed in triplicate, while the last aliquot was analysed only once ($n = 10 \times 7$).

2.3. Chromatography analysis

The chromatographic analysis was carried out using a GC-MS system Agilent 7890A GC unit, coupled to an Agilent 5977B MS detector provided by a single-quadrupole analyser and a High-Efficiency Source HES (Agilent, Little Falls, DE, USA) and equipped with a MPS-2 multipurpose auto-sampler (Gerstel GmbH, Germany). The analytical column was a capillary column 100% polydimethylsiloxane (DB 1; 10 m $\times$ 0.10 mm internal diameter $\times$ 0.40 μ m film thickness, maximum temperature 350 °C, Agilent Technologies, Santa Clara, CA, USA).

As a previous step to chromatographic separation, the VOCs were extracted by Solid phase microextraction (SPME), using a Divinylbenzene/Carboxen/Polydimethylsiloxane (DVB/CAR/PDMS) of 50/30 μ m fibre (Supelco Analytical, Bellefonte, PA, USA). Before VOCs extraction from hazelnut grain, the standard-in-fibre procedure was applied to pre load ISs on the SPME device. This procedure was performed on 5 μ L of ISs mixture placed in a 20 mL glass sealed vial and submitted to HS-SPME at the temperature of 50°C for 5 minutes. In addition, for VOCs extraction, 125 mg of hazelnut grain were placed in a 20 mL glass sealed vial. The samples were incubated in a shaker oven and submitted to HS-SPME at the temperature of 50°C for 30 minutes.

Fibre thermal desorption into the GC split/splitless was done on the injector port at 250 °C for 2 minutes of thermal desorption in pulsed, using a split ratio 1:10 and a pulse pressure at 295 kPa for another 2 minutes. Helium at a constant flow of 0.4 mL/min was used as carrier gas. The oven temperature program was set as follow: the column oven temperature of 50°C was held for 1 min, then programmed at 20°C/min to 260°C and finally programmed at 30°C/min up to 280°C, keeping this temperature fixed for 3 min before returning to the initial conditions.

The mass spectrometric condition was as follow: the ion source temperature was held isothermally at 230 °C while the quadrupole region was kept at 150°C and the ionization energy was 70 eV and the emission current 10 µA. The transfer line was kept at 280 °C through the analysis. The chromatograms were recorded in full-scan mode and the average spectra were acquired in the m/z range of 40 300 m/z with a scanning rate of 6,250 amu/s.

2.4. Data treatment

All data were acquired through chromatographic separation using the Mass Hunter software (Agilent Technologies, Wilmington, DE, United States). The individual chromatograms obtained were exported to netCDF format prior to data processing using OpenChrom® version 1.3 software (Lablicate & Scientific community, Hamburg, Germany), available free at www.lablicate.com/.

Once exported, the data were pre-processed using MATLAB, version 9.3 (Mathworks Inc., Natick, MA, USA) and an "ad hoc" home-programmed MATLAB script, called 'GC nostic' (version 0.1) which runs a number of algorithms included within the MATLAB Bioinformatic Toolbox. In addition, this script uses a specific chromatographic peak alignment algorithm based on the Kovat & Van Den Dool indices (d'Acampora Zellner et al., 2008; Cuadros Rodríguez et al., 2021), instead of applying a self-alignment as usual.

The script was applied in order to: (i) remove the lower and upper range values of each fingerprint that did not contain any useful information; (ii) filter and smooth using a Savitzky-Golay least-squares polynomial filter with a window size of 5 points and a second-order polynomial for noise removing; (iii) correct the baseline drifting; (iv) align time shifting and warping with respect to the external reference samples; and (v) normalise the intensities by using a suitable internal standard as reference. As a result, from the application of the script, a data matrix was obtained consisting of the 350

agnostized instrumental fingerprints from the chromatograms collected during the analysis of the hazelnut samples.

In addition, for supervised and unsupervised pattern recognition techniques, the software PLS_Toolbox, version 8.6.1 (Eigenvector Research Inc., Manson, WA, USA), which runs as an extension of MATLAB, was used.

3. RESULT AND DISCUSSION

Unsupervised pattern recognition technique, such as principal component analysis (PCA), was applied on the hazelnut agnostized fingerprint matrix obtained (350×4700) to evaluate the natural grouping of the analysed samples according to country of origin, and the storage conditions. Supervised Pattern Recognition Techniques such as soft Independent Modelling of Class Analogies (SIMCA), Partial Least Squares Discriminant Analysis (PLS-DA) or Support Vector Machine (SVM) were applied to create classification models in order to differentiate higher quality hazelnuts known as "Gold" from the remaining hazelnuts. These techniques were applied in the data matrix used in the PCA analysis, which was divided into two sub-matrices applying the Kennard-stones option integrated in PLS_Toolbox software as follows: a first data matrix (230×47000) used in the training of the model and a second data matrix (120×47000) used as an external validation set to test the performance of the model. Finally, a mean-center and an auto-scaling were applied independently on the data matrix before applying the different pattern recognition techniques. This strategy was adopted to evaluate which of these two pre-processing methods provided better results.

3.1. Principal Component Analysis

Auto-scaling

PCA choosing auto-scaling as pre-processing method was carried out by selecting the first 14 principal components (PCs) which explained 70.74% of the accumulated variance. Figure 1 shows the score plot of PC1 (16.90% of the variance) vs PC2 (12.26% of the variance), in which two well-defined groups are observed. Group I, with negative values for PC1 and positive values for PC2 (second quadrant), is constituted by hazelnut samples that were analysed at the beginning of the study considered as T0, while Group II with both positive and negative values for PC1 and PC2 (distributed between the first,

third and fourth quadrant) is constituted by hazelnuts analysed at T1 and T2. This behaviour suggests that regardless of storage conditions, significant compositional changes in the fraction of volatile organic compounds.

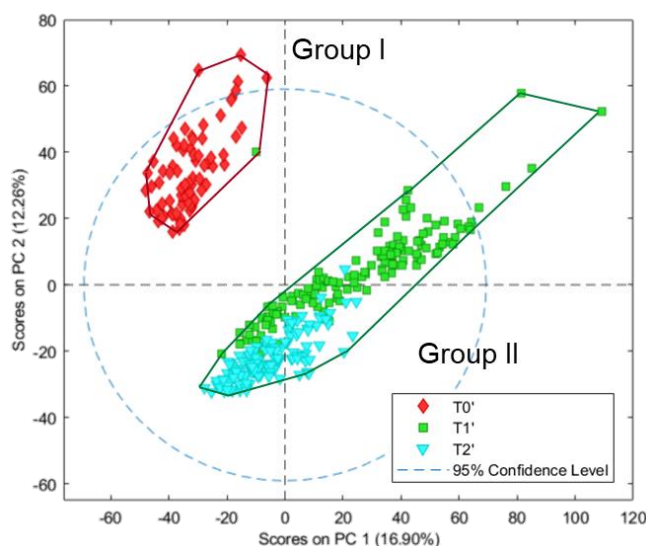


Figure 1. Score plot of PC1 vs PC2 from PCA using the agnostized fingerprint matrix (350 samples $\times$ 4700 variables) from hazelnut samples.

As components that explain a smaller percentage of variance are selected, new grouping trends can be observed. For example, although a priori PC3 (7.53 % of the variance) versus PC4 (7.25% of the variance) does not show any clear trend, ever since the different analysis times are overlapped (Figure 2a), Figure 2b reveals that samples of the same type of hazelnut are distributed along specific directions (indicated in the image with purple arrows) and are separated attending its preservation modes. However, due to the large amount of information contained in the data matrix, no logical separation according to good or bad storage conditions can be appreciated. Differences between sample preservation modes could be better appreciated by evaluating the differences of individual samples.

In order to evaluate this hypothesis, a new data matrix (35 $\times$ 4700) was created containing only the agnostized fingerprints of the hazelnut coded as Hz-3 for each of the analysis times and in each of the preservation modes. This analysis was carried out by selecting the first 6 PC, where the best separation among preservation modes is observed in the representation of PC1 vs PC5 (Figure 2c). In this figure it can be seen how samples stored in good conditions (stored in vacuum) take negative values for PC5 while hazelnuts

stored in bad conditions (stored in air atmosphere) take positive values for this component.

Besides the differences in terms of time of analysis or preservation methods, other separations such as cultivar/origin can be observed in the first PCA analysis carried out. In Figure 3 (a and b), where PC3 (7.53% of the variance) and PC10 (2.18% of the variance) are represented against PC4 (7.25% of the variance), it can be seen that, this one is the last PC which establishes differences between hazelnut samples according to origin country. Such trend is more evident in Figure 3b, where it is clearly observed how samples from different regions of Italy present positive values for PC 10 and negative values for PC4, with almost all samples from Italy being grouped in the second quadrant of the score plot. Differences between hazelnut samples according to origin's country can be better established by evaluating only one time of analysis.

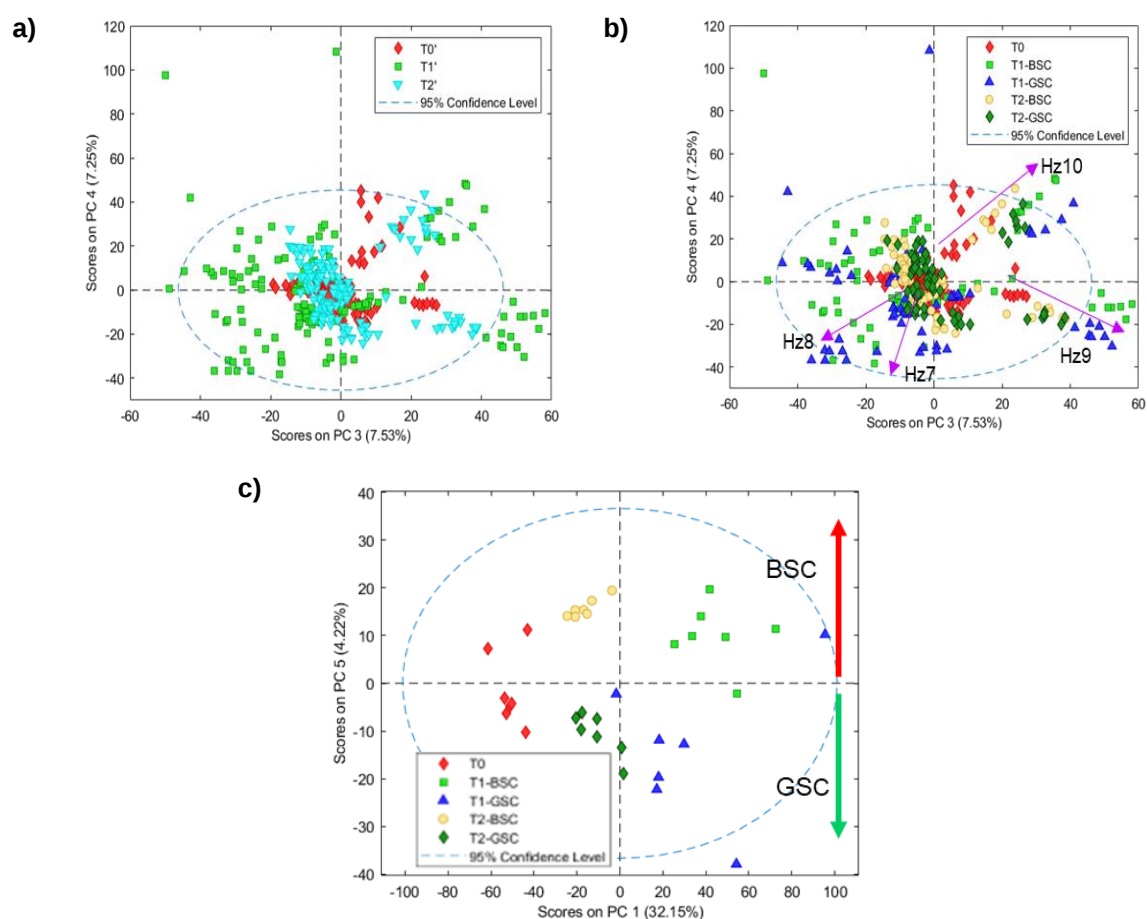


Figure 2. a) Score plot of PC3 vs PC4 from PCA using the agnostized fingerprint matrix (350 × 4700) from hazelnut samples, b) same score plot showing not only the analysis times, but also the preservation methods used in this study and c) score plot of PC1 vs PC5 from the PCA using only the agnostized fingerprint of the sample codified as Hz 3 (35 × 4700).

For this purpose, a new PCA was carried out on a data matrix consisting only of hazelnuts analysed at T0 (70×4700). The first 10 PCs, which explained 71.86% of the variance, were selected in this analysis. In Figure 3c, representing PC2 (13.03% of the variance) vs PC5 (5.76% of the variance), two groups can be clearly observed. Group I with both positive and negative values for PC 2 and PC 5 but scattered in the second, third and fourth quadrant of the graph, is composed by hazelnut blends from different regions of Turkey, while Group II with positive values for both PC2 and PC5 is composed by hazelnut samples from Italy. Furthermore, Group II can be subdivided into Group II-A, consisting of hazelnuts from the Piedmont region, and Group II-B, consisting of hazelnuts from the Lazio region.

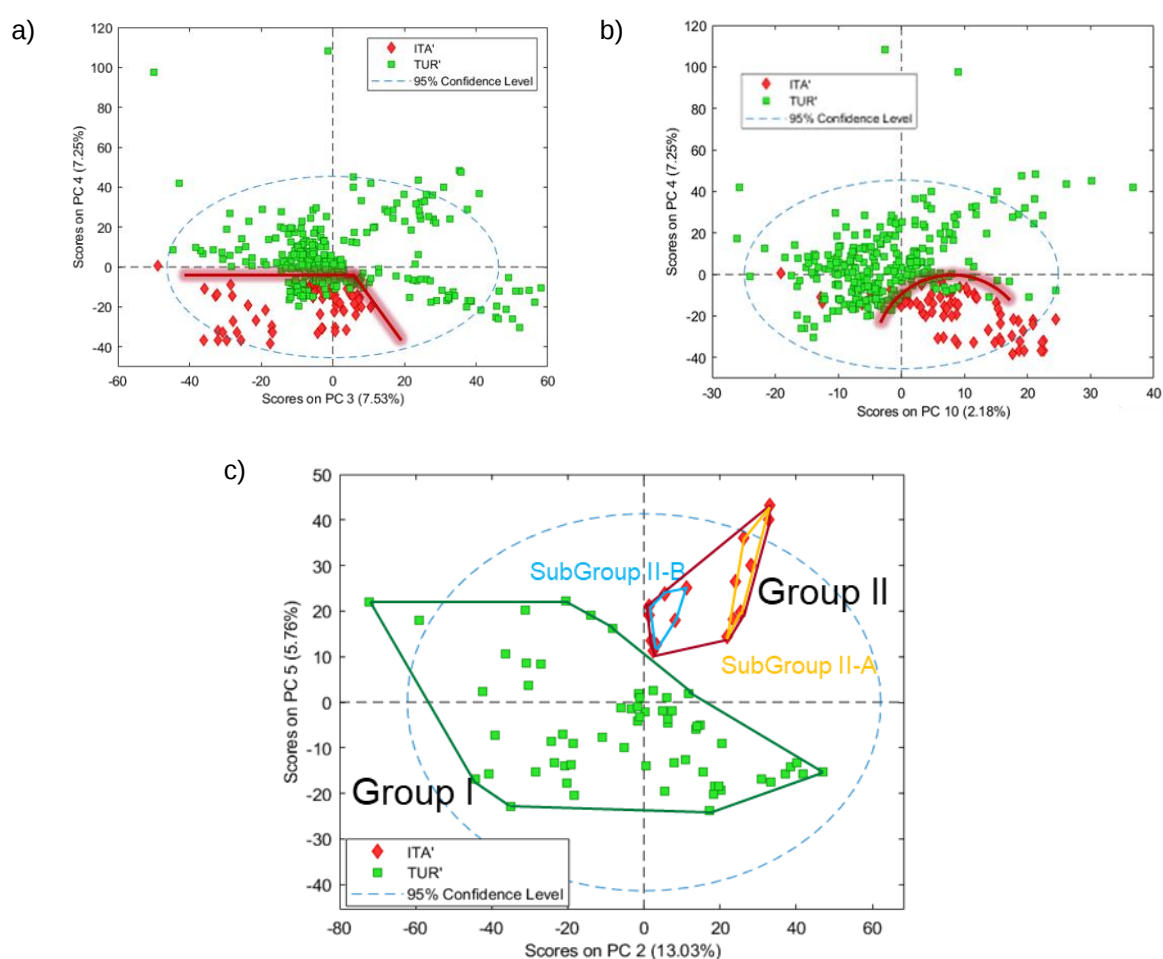


Figure 3. Score plot of: a) PC3 vs PC4 from PCA using the agnostized fingerprint matrix (350 × 4700) assigning as a label the origin country to each of the samples analysed, b) PC10 vs PC4 from the same PCA and c) PC2 vs PC5 from the PCA using only the agnostized fingerprint of the sample analysed at T0 (70 × 4700).

Mean-center

All previous case studies performed with auto scaling as the pre-processing method prior to PCA were repeated by replacing the pre-processing method to mean-center. Although the variance percentage explained on the data matrix containing all hazelnut agnostized fingerprints (350×4700) was higher for a lower number of PCs (the first 6 PCs explain 88.04% of the variance), the results obtained did not show any clear grouping or trend compared to the auto-scaling pre-processing method. Only the case study, which evaluated the different storage conditions for the sample coded as Hz3 (35×4700) showed better results with the mean-center method compared with the auto-scaling method. In this case the first 6 PCs explain 97.64% of the total variance. Figure 4 shows the representation of PC1 (45.29% of the variance) versus PC 4 (4.67% of the variance). Three groups can be clearly observed in this figure: i) Group A with positive values of PC1 and close to zero in PC4 composed by all the repeats of the Hz3 sample analysed at the beginning of the study (T0); ii) Group B with negative values for both PC1 and PC4 composed by the multiple replicates of the same hazelnut analysed in different months (T1 and T2) with good storage conditions (GSC) and iii) Group C with negative values for PC1 and positive values for PC4 composed by the multiple replicates of the same hazelnut sample analysed in different months (T1 and T2) with bad storage conditions (BSC). Therefore, this behaviour reveals that regardless of the storage conditions, some changes appear in the VOC composition of the hazelnut samples. Nonetheless, it is also noticeable that according to this projection, the samples analysed at T0 show a greater similarity to those samples that were preserved at low temperatures under modified atmosphere (GSC).

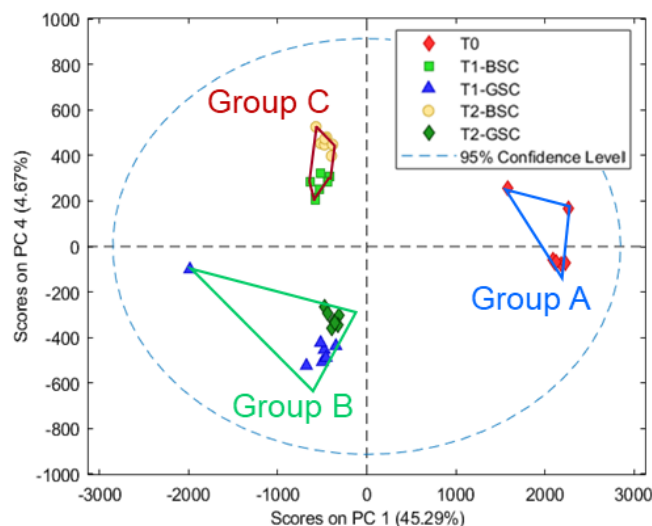


Figure 4. Score plot of PC1 vs PC4 from the PCA using only the agnostized fingerprint of the sample codified as Hz 3 (35×4700) with mean-center as pre-processing tool.

3.2. Supervised Pattern Recognition Analysis

Once the natural grouping of the samples had been evaluated, classification models were developed to differentiate hazelnuts of "*Tonda Gentile Trilobata*" variety, harvested in the Piedmont region and considered to be the hazelnut cultivar with the best hedonic properties, from the other samples analysed from other regions of Italy (Lazio region) and Turkey (Akçakoca and Black Sea region). Two input classes "Gold" - samples from the Piedmont region - and "Others" - samples from the rest of geographical locations (Lazio, Italy; Akçakoca and Black Sea regions, Turkey) - were assigned to the samples that constituted both the training matrix (composed by 24 samples of "Gold" class and 206 samples of "Others" class) and the external validation matrix (composed by 11 hazelnut samples of "Gold" class and 109 samples of "Others" class) described above.

Similar to the unsupervised pattern recognition techniques, all classification models were developed with two different pre-processing (auto-scaling and mean-center) applied independently.

Soft independent modelling of class analogies – SIMCA model

The SIMCA model, developed with auto-scaling as a pre-processing option, was established by selecting the first 8 PCs for "Gold" class and 15 PCs for "Others" class,

which explained respectively 81.98% and 74.68% of the variance for each of the modelled classes. Once the model was trained, the external prediction matrix was used to evaluate the potential of the model obtaining the Cooman's plot illustrated in Figure 5a. Among the eleven samples labelled as "Gold" in the external prediction set, only two samples were correctly classified as belonging to that class, while another two samples were wrongly classified as belonging to "Others" class. From the remaining seven samples, six of them were within the region of inconclusive samples (third quadrant) while only one was assigned to the "Not recognised" class (first quadrant). Regarding "Others" class practically all samples were correctly classified, with the exception of 5 samples which were assigned to "Not recognised" class.

The SIMCA model with mean-center pre-processing applied was established by selecting the first 4 PCs for the "Gold" class and 11 PCs for the "Others" class, which respectively explained 94.48% and 94.23% of the variance for each modelled class. This model shows similar results compared to the auto-scaling pre-processing option, thus two of the 11 samples labelled as "Gold" in the external prediction set were correctly classified as belonging to that class, while the remaining samples were assigned to the region of inconclusive samples. Furthermore, it is pinpointed that for "Others" class only a single sample was assigned to the class of unrecognised samples, while four other samples were included in the region of inconclusive samples.

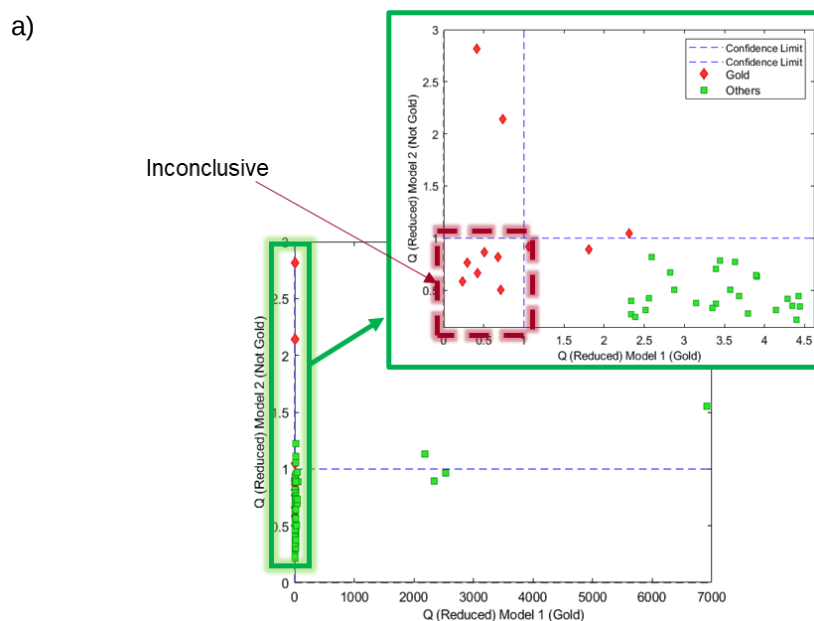


Figure 5. Cooman's plot obtained from the development of the SIMCA model for the different pre-processing options: a) auto-scaling and b) mean-center.

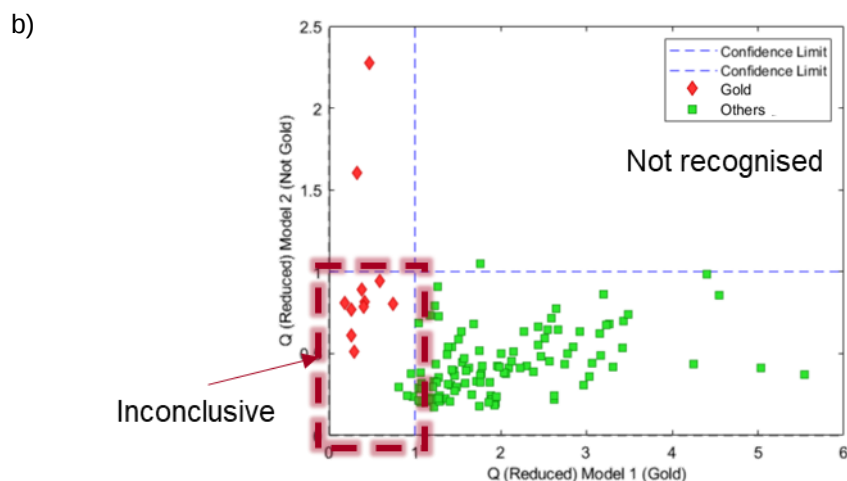


Figure 5. (cont.) Cooman's plot obtained from the development of the SIMCA model for the different pre-processing options: a) auto-scaling and b) mean-center.

For both pre-processing options, both samples included in the inconclusive region and misclassified samples for "Gold" class corresponded mainly to hazelnuts in poor storage conditions in the last month of analysis (T2). The SIMCA model considers that although these samples belong to "Gold" class, the agnostized fingerprint also has certain similarities with "Others" class, hence there is a certain probability of belonging to both modelled classes.

Partial least squares - discriminant analysis (PLS-DA)

The PLS-DA model obtained using auto-scaling as a pre-processing step prior to obtaining the model was developed by selecting the first 6 latent variables (LV), which explained 43.39% of the cumulative variance over the X-block and 90.77% of the cumulative variance over the Y-block. For the PLS-DA model, the threshold belonging to each class was set by "in lab-criteria" at the probability value of 0.5. Figure 6 shows the probability plot of samples belonging to the modelled classes. This picture shows how the samples belonging to the "Gold" class had probability values higher than the limit imposed for the development of the model, properly classifying all samples labelled as "Gold" for both datasets (training and external validation sets).

Moreover, the PLS-DA model developed using mean-center pre-processing step prior to model development was established using the first 5 LVs, which explained 82.81% of the

cumulative variance in the X-block and 75.08% of the cumulative variance in the Y-block. The results of the model classification are reported in Figure 6b. Once again, the membership threshold for the two modelled classes was set to a probability value of 0.5, as in the case of the auto-scaling. It can be seen how at least two samples labelled as "Gold" are misclassified, being within the "Others" class region. In this case the model results could be improved if the inter-class threshold was set around the value of 0.4. However, this option was rejected in order to apply the same criterion for classification in both approaches and thus make their results comparable.

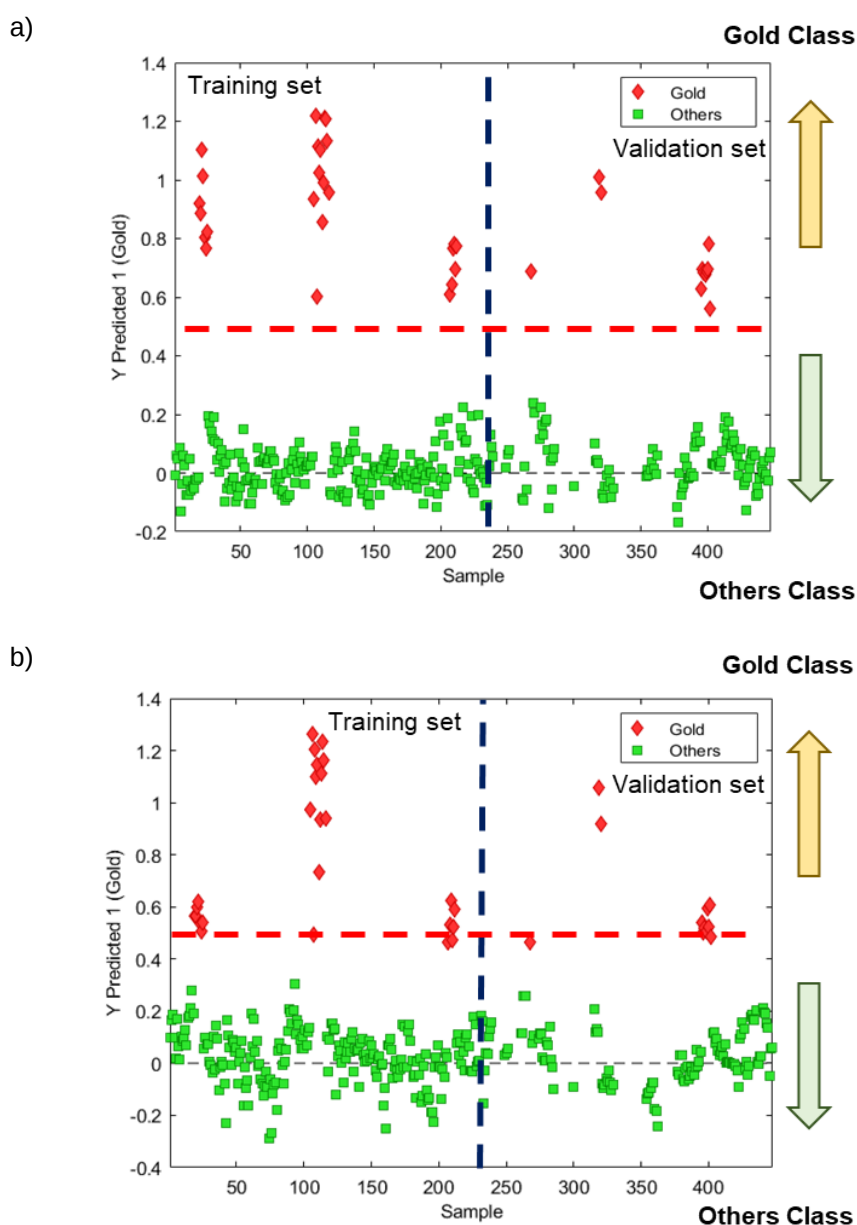


Figure 6. Prediction plot obtained from the development of the PLS-DA model for the different pre-processing options: a) auto-scaling and b) mean-center.

Support Vector Machine (SVM)

The SVM models for each of the pre-processing options used in this work were obtained without applying any prior comprehension during model development. In addition, the radial basis function (RBF) was applied as the Kernel algorithm, using the default values set by the PLS_Toolbox software for the gamma and cost parameters. In both cases the models showed similar results, and indicated successfully that the different samples belonged to the corresponding modelled classes. In Figure 7, it is noticeable that, for the model created by applying auto-scaling, there is a greater probability of belonging to the "Others" class in some samples labelled as "Gold". Nevertheless, the misclassification probability is negligible as these samples are not enough close to the threshold probability value which distinguishes the two modelled classes. This behaviour was not observed in the mean-center model.

The quality parameters of the three classification models for both pre-processed options, listed in Table 2, were calculated from the contingency tables provided in the supplementary material (Figure SM1-SM6). From the different classification models developed, the SVM, regardless of the pre-processing option selected, as well as the PLS-DA model with auto-scaling as pre-processing stage are the models that show the best values for sensitivity (SENS), specificity (SPEC), positive predicted values (PPV) and negative predicted values (NPV). In the case of PLS-DA with mean-center, the results of the SENS and FNR parameters were negatively affected by the value selected as the inter-class threshold.

Finally, from all of the models developed, the SIMCA for both pre-processed products showed the worst classification results in terms of belonging to the "Gold" class. This fact could be explained from the visual inspection of PCA because of the changes produced as a consequence over time have a greater effect on the VOC fraction compared to other factors such as the origin region, which has a negative impact on the prediction stage of the SIMCA model.

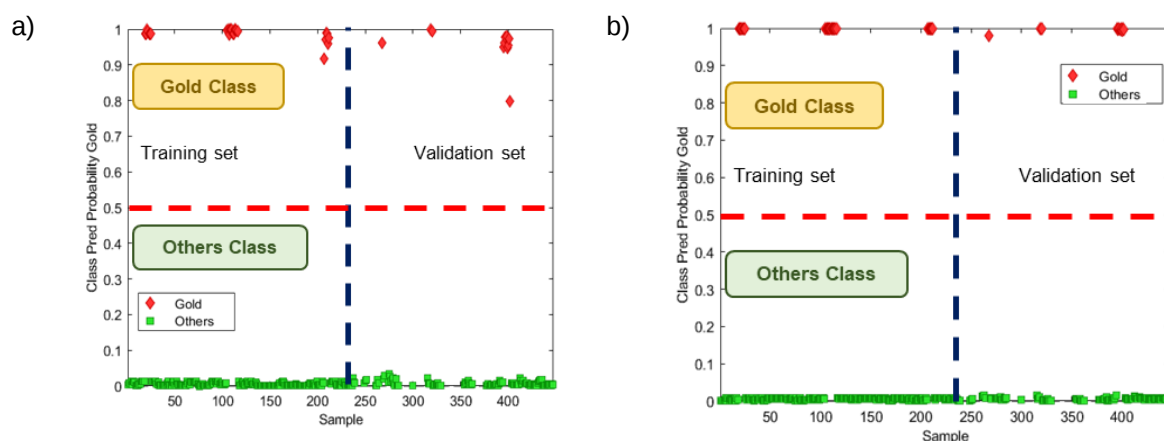


Figure 7. Probability prediction plot obtained from the development of the SVM model for the different pre-processing options: a) auto-scaling and b) mean-center.

4. CONCLUSION

Compared to the use of (bio)markers, fingerprint methodology has demonstrated its usefulness for hazelnut discrimination within the framework of quality assessment through the use of information not covered by (bio)markers, but encrypted in the instrumental fingerprint.

The implementation of both, auto-scaling and mean-center as independent pre-processing stages combined with unsupervised pattern recognition techniques, showed that the agnostized fingerprint contained enough information about the ageing and storage conditions of samples, as well as the country of origin of them. When the number of samples contained in the agnostized fingerprint matrix was reduced to only those obtained at T0, it could be observed that the samples from Italy were sub-grouped according to the region of origin, being these differences more evident when auto-scaling was applied.

Supervised pattern recognition techniques allowed the development of hazelnut quality classification models. Among the different classification models developed, the PLS-DA model with auto-scaling and SVM for both pre-processing options, showed the best classification results for higher quality hazelnut samples, referred as "Gold" class in this paper.

Finally, the new approach of instrument-agnostizing methodology, where only the variability present in all samples is considered, using as alignment reference signal these same samples, in conjunction with chemometric techniques with different pre-processing options, can be considered as a valid approach in problem resolution associated with

fingerprint methodology. However, this new approach overcomes the dependence of the signal on the time of acquisition, but not on the instrument used to acquire it. The lack of a clearly identified external and pure reference makes difficult the possible behavioural transfer between different instruments.

Table2. Summary of classification performance metrics obtained for the different model and pre-processing option applied in this study.

	SIMCA		PLS-DA		SVM	
	Auto-scaling	Mean center	Auto-scaling	Mean center	Auto-scaling	Mean center
	Class 1 (Gold)					
Occurrence (OCCUR)	0.09	0.09	0.09	0.09	0.09	0.09
Inconclusive rate (IR)	0.01	0.00	–	–	–	–
Sensitivity (SENS)	0.18	0.18	1.00	0.82	1.00	1.00
Specificity (SPEC)	0.95	0.95	1.00	1.00	1.00	1.00
False positive rate (FPR)	0.05	0.05	0.00	0.00	0.00	0.00
False negative rate (FNR)	0.82	0.82	0.00	0.18	0.00	0.00
Positive predictive value (precision) (PPV)	1.00	1.00	1.00	1.00	1.00	1.00
Negative predictive value (NPV)	0.98	1.00	1.00	0.98	1.00	1.00
Youden index (YOUDE)	0.14	0.14	1.00	0.82	1.00	1.00
Positive likelihood rate (LR (+))	4.00	4.00	–	–	–	–
Negative likelihood rate (LR (-))	0.86	0.86	0.00	0.18	0.00	0.00
Classification odds ratio (COR)	4.67	4.67	–	–	–	–
F-measure (F)	0.31	0.31	1.00	0.90	1.00	1.00
Discriminant power (DP)	0.37	0.37	–	–	–	–
Efficiency (or accuracy) (EFFIC)	0.88	0.88	1.00	0.98	1.00	1.00
Misclassification rate (MR)	0.12	0.12	0.00	0.02	0.00	0.00
AUC (correctly classified rate)	0.57	0.57	1.00	0.91	1.00	1.00
Gini coefficient (Gini)	0.14	0.14	1.00	0.82	1.00	1.00
G-mean (GM)	0.42	0.42	1.00	0.90	1.00	1.00
Matthew's correlation coefficient	0.41	0.42	1.00	0.90	1.00	1.00
Chance agreement rate (CAR)	0.81	0.79	0.83	0.85	0.83	0.83
Chance error rate (CER)	0.17	0.17	0.17	0.17	0.17	0.17
Kappa coefficient (KAPPA)	0.41	0.45	1.00	0.89	1.00	1.00
PROB (1/1)	0.29	0.29	1.00	1.00	1.00	1.00
PROB (2/2)	0.92	0.92	1.00	0.98	1.00	1.00
PROB (1/2)	0.08	0.08	0.00	0.02	0.00	0.00
PROB (2/1)	0.71	0.71	0.00	0.00	0.00	0.00

Note. The hyphen "–" refers to metrics that cannot be determined since a division by zero is involved.

SUPPLEMENTARY MATERIAL

		11	110	121
Assigned class	Not recognized (Nr)	6	0	6
	Inconclusive (Inc)	1	5	6
	Others	2	105	107
	Gold	2	0	2
		Gold	Others	T
		Actual		

Figure SM1. Contingency values for the SIMCA classification model using auto-scaling as pre-processing option.

		11	110	121
Assigned class	Nr	9	4	13
	Inc	0	1	1
	Others	0	105	105
	Gold	2	0	2
		Gold	Others	T
		Actual		

Figure SM2. Contingency values for the SIMCA classification model using mean-center as pre-processing option.

		11	110	121
Assigned class	Inc	0	0	0
	Others	0	110	110
	Gold	11	0	11
		Gold	Others	T
		Actual		

Figure SM3. Contingency values for the PLS-DA classification model using auto-scaling as pre-processing option.

Assigned class		11	110	121
	Inc	0	0	0
	Others	2	110	112
	Gold	9	0	9
		Gold	Others	T
		Actual		

Figure SM4. Contingency values for the PLS-DA classification model using mean-center as pre-processing option.

Assigned class		11	110	121
	Inc	0	0	0
	Others	0	110	110
	Gold	11	0	11
		Gold	Others	T
		Actual		

Figure SM5. Contingency values for the SVM classification model using auto-scaling as pre-processing option.

Assigned class		11	110	121
	Inc	0	0	0
	Others	0	110	110
	Gold	11	0	11
		Gold	Others	T
		Actual		

Figure SM6. Contingency values for the PLS-DA classification model using mean-center as pre-processing option.

REFERENCES

Food and Agriculture Organization of the United Nations, FAOSTAT - Production quantities of hazelnuts (2020). accessed on May 2nd, 2022.

<https://www.fao.org/faostat/en/#data/OCL/visualize>

Silvestri, C., Bacchetta, L., Bellincontro, A. & Cristofori, V. (2020) Advances in cultivar choice, hazelnut orchard management, and nut storage to enhance product quality and safety: an overview. *Journal of the Science of Food and Agriculture*, 101(1), 27-43.

<https://doi.org/10.1002/jsfa.10557>

Cialiè Rosso, M., Mazzucotelli, M., Bicchi, C., Charron, M., Manini, F., Menta, R., Fontana, M., Reichenbach, S.E., Cordero, C. (2020) 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. *Journal of Chromatography A*, 1614, 460739.

<https://doi.org/10.1016/j.chroma.2019.460739>

Squara, S., Stilo, F., Cialiè Rosso, M., Liberto, E., Spigolon, N., Genova, G., Castello, G., Bicchi, C., Cordero, C. (2022). *Corylus avellana* L. Aroma Blueprint: Potent Odorants Signatures in the Volatilome of High Quality Hazelnuts. *Frontiers in Plant Science*, 13, 840028

<https://doi.org/10.3389/fpls.2022.840028>

UNECE, U. S. (2010). DDP-04 Concerning the Marketing and Commercial Quality Control of Hazelnut Kernels. United Nations New York, Geneva.

Caligiani, A., Coisson, J. D., Travaglia, F., Acquotti, D., Palla, G., Palla, L., & Arlorio, M. (2014). Application of ¹H NMR for the characterisation and authentication of "Tonda Gentile Trilobata" hazelnuts from Piedmont (Italy). *Food Chemistry*, 148, 77-85.

<https://doi.org/10.1016/j.foodchem.2013.10.001>

Cialiè Rosso, M., Stilo, F., Bicchi, C., Charron, M., Rosso, G., Menta, R., ... & Cordero, C. (2021a). Combined untargeted and targeted fingerprinting by comprehensive two-dimensional gas chromatography to track compositional changes on hazelnut primary metabolome during roasting. *Applied Sciences*, 11, 525.

<https://doi.org/10.3390/app11020525>

Bagur González, M.G., Jiménez Carvelo, A.M., Ortega Gavilán, F. & González Casado A. (2021) Chromatographic methods. In M. Galanakis (Ed) Food authentication and traceability, (pp 65–99), London, United Kingdom, Elsevier.

<https://doi.org/10.1016/B978-0-12-821104-5.00012-X>

Pérez-Castaño, E., Sánchez-Viñas, M., Gázquez-Evangelista, D., & Bagur-González, M. G. (2018). Quantification of extra virgin olive oil in dressing and edible oil blends using the representative TMS-4, 4'-desmethylsterols gas-chromatographic-normalized fingerprint, *Food chemistry*, 239, 1192-1199.

<https://doi.org/10.1016/j.foodchem.2017.07.069>

Cevoli, C., Cerretani, L., Gori, A., Caboni, M. F., Toschi, T. G., & Fabbri, A. (2011). Classification of Pecorino cheeses using electronic nose combined with artificial neural network and comparison with GC–MS analysis of volatile compounds, *Food Chemistry*, 129, 1315-1319.

<https://doi.org/10.1016/j.foodchem.2011.05.126>

Souard, F., Delporte, C., Stoffelen, P., Thévenot, E. A., Noret, N., Dauvergne, B., ... & Stevigny, C. (2018). Metabolomics fingerprint of coffee species determined by untargeted-profiling study using LC-HRMS, *Food chemistry*, 245, 603-612.

<https://doi.org/10.1016/j.foodchem.2017.10.022>

Chen, H., Jin, L., Fan, C., & Wang, W. (2017). Non-targeted volatile profiles for the classification of the botanical origin of Chinese honey by solid-phase microextraction and gas chromatography–mass spectrometry combined with chemometrics, *Journal of separation science*, 40, 4377-4384.

<https://doi.org/10.1002/jssc.201700733>

Cuadros-Rodríguez, L., Ruiz-Samblás, C., Valverde-Som, L., Pérez-Castaño, E., & González-Casado, A. (2016). Chromatographic fingerprinting: An innovative approach for food identification and food authentication—A tutorial. *Analytica Chimica Acta*, 909, 9-23.

<https://doi.org/10.1016/j.aca.2015.12.042>

Kucharska-Ambrożej, K., & Karpinska, J. (2020). The application of spectroscopic techniques in combination with chemometrics for detection adulteration of some herbs and spices. *Microchemical Journal*, 153, 104278.

<https://doi.org/10.1016/j.microc.2019.104278>

Alasalvar, C., Odabasi, A. Z., Demir, N., Balaban, M. Ö., Shahidi, F., & Cadwallader, K. R. (2004). Volatiles and flavor of five Turkish hazelnut varieties as evaluated by descriptive sensory analysis, electronic nose, and dynamic headspace analysis/gas chromatography-mass spectrometry. *Journal of food science*, 69, SNQ99-SNQ106.

<https://doi.org/10.1111/j.1365-2621.2004.tb13382.x>

Alasalvar, C., Pelvan, E., Bahar, B., Korel, F., & Ölmez, H. (2012). Flavour of natural and

roasted Turkish hazelnut varieties (*Corylus avellana* L.) by descriptive sensory analysis, electronic nose and chemometrics. *International journal of food science & technology*, 47, 122-131.

<https://doi.org/10.1111/j.1365-2621.2011.02817.x>

Cialie Rosso, M. (2018). Advanced analytical approaches for “omic” investigations of high-quality food matrices of vegetable origin. Published online 2019. accessed on May 2nd, 2022.

https://dott-sfb.campusnet.unito.it/att/XXXII/Abstract_CIALIE_ROSSO.pdf

Cordero, C., Liberto, E., Bicchi, C., Rubiolo, P., Schieberle, P., Reichenbach, S. E., & Tao, Q. (2010). 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 origins. *Journal of Chromatography A*, 1217, 5848-5858

<https://doi.org/10.1016/j.chroma.2010.07.006>

Nicolotti, L., Cordero, C., Cagliero, C., Liberto, E., Sgorbini, B., Rubiolo, P., & Bicchi, C. (2013). Quantitative fingerprinting by headspace-Two-dimensional comprehensive gas chromatography-mass spectrometry of solid matrices: Some challenging aspects of the exhaustive assessment of food volatiles. *Analytica Chimica Acta*, 798, 115-125.

<https://doi.org/10.1016/j.aca.2013.08.052>

Stilo, F., Liberto, E., Spigolon, N., Genova, G., Rosso, G., Fontana, M., ... & Cordero, C. (2021). 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*, 340, 128135.

<https://doi.org/10.1016/j.foodchem.2020.128135>

Cuadros-Rodríguez, L., Ortega-Gavilán, F., Martín-Torres, S., Medina-Rodríguez, S., Jimenez-Carvelo, A. M., González-Casado, A., & Bagur-González, M. G. (2021). Standardization of chromatographic signals – Part I: Towards obtaining instrument-agnostic fingerprints in gas chromatography. *Journal of Chromatography A*, 1641, 461983.

<https://doi.org/10.1016/j.chroma.2021.461983>

Cuadros-Rodríguez, L., Martín-Torres, S., Ortega-Gavilán, F., Jiménez-Carvelo, A. M., López-Ruiz, R., Garrido-Frenich, A., ... & González-Casado, A. (2021). Standardization of chromatographic signals – Part II: Expanding instrument-agnostic fingerprints to reverse

phase liquid chromatography. *Journal of Chromatography A*, 1641, 461973.

<https://doi.org/10.1016/j.chroma.2021.461973>

Cialiè Rosso, M., Stilo, F., Mascrez, S., Bicchi, C., Purcaro, G., & Cordero, C. (2021b). Shelf-Life Evolution of the Fatty Acid Fingerprint in High-Quality Hazelnuts (*Corylus avellana* L.) Harvested in Different Geographical Regions. *Foods*, 10(3), 685.

<https://doi.org/10.3390/foods10030685>

d'Acampora Zellner, B., Bicchi, C., Dugo, P., Rubiolo, P., Dugo, G., & Mondello, L. (2008). Linear retention indices in gas chromatographic analysis: a review. *Flavour and Fragrance Journal*, 23, 297-314.

<https://doi.org/10.1002/ffj.1887>

Capítulo VII

Discusión integrada

VII.1. INTRODUCCIÓN

El creciente interés por el control de la calidad y autenticidad de los alimentos, así como el continuo desarrollo de métodos analíticos que aseguren dichos parámetros se ha puesto de manifiesto de forma constante a lo largo de esta Tesis. De manera específica, dos de los alimentos que suscitan mayor interés en algunos países de la UE son el aceite de oliva y la avellana.

El aceite de oliva es uno de los principales productos exportados en España, siendo dicho país uno de los principales productores a nivel mundial. Por otro lado, la avellana, altamente apreciada en Italia, es una de las principales materias primas en la industria de la repostería, y es frecuente su uso en la elaboración de chocolates, praliné y cremas untables de cacao. Ambos se caracterizan por ser alimentos bastante complejos desde el punto de vista composicional, por lo que su calidad depende de manera simultánea de múltiples parámetros que pueden ser determinados de manera instrumental o sensorial.

De los diferentes tipos de alimentos estudiados en esta Tesis, el aceite de oliva virgen o virgen extra es el único de ellos en el que la UE obliga a los organismos de control oficial a que realicen un análisis sensorial para verificar su denominación de venta durante la comercialización. Ante esta obligatoriedad, y partiendo de la premisa descrita en la introducción de esta Tesis en la que se considera al panel de cata como un instrumento analítico multisensor, cobra mayor sentido el disponer de MRC específicamente diseñados para su uso en análisis sensorial que ayuden en el entrenamiento y control del panel.

Cuando estos materiales están destinados al uso exclusivo en el control de parámetros sensoriales, es recomendable hacer un uso combinado del análisis sensorial e instrumental durante el proceso de certificación. De los diferentes estudios necesarios que se deben realizar para alcanzar la certificación, el estudio de estabilidad es con diferencia el que presenta mayor dificultad cuando la metodología de trabajo a seguir se basa en el uso de huellas instrumentales (*instrumental fingerprints*).

El estudio de la estabilidad necesita del análisis de unidades individuales representativas de una misma muestra en diferentes periodos de tiempo donde se pueden observar cambios significativos en diferentes meses que tienen un origen muy diverso:

- i) al tratarse de muestras de origen biológico, los cambios pueden producirse como consecuencia de los propios procesos de oxidación y descomposición que tienen lugar de forma natural en la muestra independientemente de que esta se conserve en condiciones óptimas, o
- ii) debido a la dependencia de la señal analítica que se adquiere con el estado en el que se encuentra el instrumento en el momento en que se registra dicha señal.

No obstante, dichas causas no son exclusivas de la monitorización de la estabilidad de un MR, ya que es bien conocido la relación que existe, sobre todo en los métodos cromatográficos, entre el registro de la señal y las condiciones en las que se encuentra el equipo en cada momento. Por otro lado, cualquier huella instrumental obtenida a partir de muestras biológicas, puede considerarse como una "señal viva" y por tanto cambiante, bien por evolución propia del material biológico, o bien por la influencia de matrices nuevas que modifiquen su composición natural (caso de muestras compuestas).

Prevenir o eliminar estas diferencias tiene también cierta dificultad, ya que, en el caso de los cambios producidos en la muestra biológica, la mejor solución es establecer protocolos de almacenamiento adecuados que minimicen cualquier reacción de oxidación o degradación térmica de la muestra, así como la de mantener una trazabilidad estricta en la que se tenga control de la muestra con el fin de evitar cualquier mezcla fraudulenta o contaminación accidental. En el caso de la dependencia de la señal del estado del instrumento, el desarrollar una solución como la planteada en esta Tesis, mediante la agnostización de cada huella instrumental, permitiría independizarla de los parámetros experimentales, manteniendo toda la información encriptada en la misma. A su vez esto permitiría una comparación entre huellas (huellas agnostizadas) más fácil, ya que en cierto modo se minimizarían los problemas anteriormente descritos.

En el caso de las huellas instrumentales obtenidas mediante cromatografía, una forma de independizar la señal de las condiciones experimentales podría consistir en estudiar el comportamiento de compuestos con un comportamiento cromatográfico similar a los analizados (patrones externos) frente a las condiciones experimentales nominales estándar utilizadas en el equipo. El estudio de dicho comportamiento permitiría su transferencia al resto de señales mediante la sustitución de los valores de

tiempos de retención por nuevos valores empíricos independientes de las particularidades de cada método/instrumento/ análisis.

Esta metodología, descrita en esta Tesis, es denominada como agnostización de la señal instrumental. La denominación está basada en aplicar en el ámbito de la química analítica un significado o connotación libre del término "agnosticismo", que ya ha sido propuesto por otros autores, y cuyo uso no es original en esta Memoria.

La agnostización presentaría una gran aplicabilidad en el campo de la química analítica, ya que aparte de la monitorización de la estabilidad de MR, sería de gran utilidad en problemas analíticos basados en mezclas no autorizadas o incluso en la creación de bases de datos de alimentos con el fin de establecer modelos de clasificación y cuantificación entre otras muchas aplicaciones que contribuyeran a la mejora de la calidad alimentaria.

Estas aplicaciones potenciales se han ido desarrollado en los diferentes estudios que dotan de contenido a esta Tesis y que se han materializado en diversas publicaciones científicas. En cada uno de ellos se ha propuesto una adaptación concreta atendiendo a las particularidades de cada caso de estudio (tipo de cromatografía –GC o LC–, modalidad cromatográfica –baja o alta temperatura en GC o fase normal o invertida en caso de LC–).

VII.2. PLANTEAMIENTO DE LA PROBLEMÁTICA EN LA PRIMERA CAMPAÑA DE CERTIFICACIÓN DE MATERIALES DE REFERENCIA

En este apartado y como complemento al trabajo publicado recogido en el Capítulo II se comentan parte de los resultados no publicados. En su totalidad, los resultados forman parte del convenio establecido entre la Consejería de Agricultura, Pesca y Desarrollo Sostenible de la Junta de Andalucía con la Oficina de Transferencia de Resultados de Investigación de la Universidad de Granada (OTRI-UGR) titulado "Elaboración, certificación y distribución de nueve Materiales de Referencia Certificados (MRC) para análisis organoléptico de aceite de oliva –SENSOLEO-MRC 2017– (OTRI 3898-2017). A su vez, han servido como punto de partida para desarrollar el resto de estudios incluidos en esta Tesis Doctoral.

Una de las etapas más críticas en el proceso de producción de MR es el envasado. El motivo se debe a la necesidad por parte del productor de que todos los laboratorios

que vayan a utilizar dicho material, reciben exactamente un producto similar. Por ello es necesario que durante el proceso de certificación se asegure la homogeneidad del candidato a MR.

Este estudio podría realizarse siguiendo la metodología aplicada en este trabajo, en el que la homogeneidad se establece por comparación de las señales obtenidas mediante GC con detector de ionización de llama analizando diferentes unidades seleccionadas de forma sistemática, de tal forma que se cubra todo el rango de unidades envasadas para cada uno de los candidatos a MR seleccionados en el proyecto.

La comparación de dos señales requiere previamente de un ajuste o pre-procesado con la finalidad eliminar cualquier fuente de variabilidad que provenga del instrumento, de modo que todas las señales contengan el mismo número de elementos, se presenten en la misma escala de intensidades y se haya eliminado cualquier fuente de ruido o de deriva, dando como resultado un nuevo conjunto de datos denominado huella instrumental.

En el pre-procesado de la señal cromatográfica se incluye una etapa de alineamiento, la cual es necesaria debido a las pequeñas desviaciones que se produce en los tiempos de retención como consecuencia de la degradación térmica de la fase estacionaria contenida en la columna capilar. Un hecho destacable es que, durante los análisis realizados en el estudio de homogeneidad para un mismo material, las desviaciones observadas en los valores de tiempo de retención de la señal son mínimas, siendo solamente apreciables en los últimos minutos del análisis. En la Figura 20 se muestran los 12 cromatogramas obtenidos para el estudio de homogeneidad de uno de los candidatos a MR.

En este caso, al tener todos los cromatogramas perfiles similares (puesto que provienen de un mismo material que ha sido analizado en condiciones de repetibilidad), los algoritmos estadísticos de alineamiento, como el algoritmo de desplazamiento optimizado por corrección de intervalos (interval correlation optimized shifting – Icoshift) o el algoritmo de deformación optimizada por correlación (correlation optimized warplings – COW), son capaces de corregir las deformaciones producidas en el eje de tiempos sin ningún tipo de dificultad. De hecho, en el trabajo recogido en el Capítulo II de esta Tesis, se muestra como tras aplicar el pre-procesado, las deformaciones ocurridas durante los últimos minutos del análisis son corregidas.

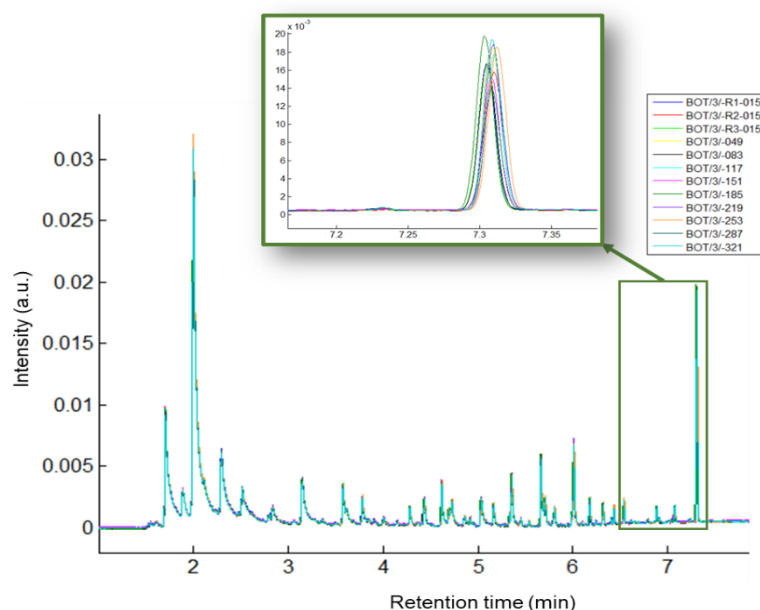


Figura 20. Cromatogramas obtenidos en el estudio de homogeneidad para la muestra BOT/3/ antes de aplicar el pre-procesado.

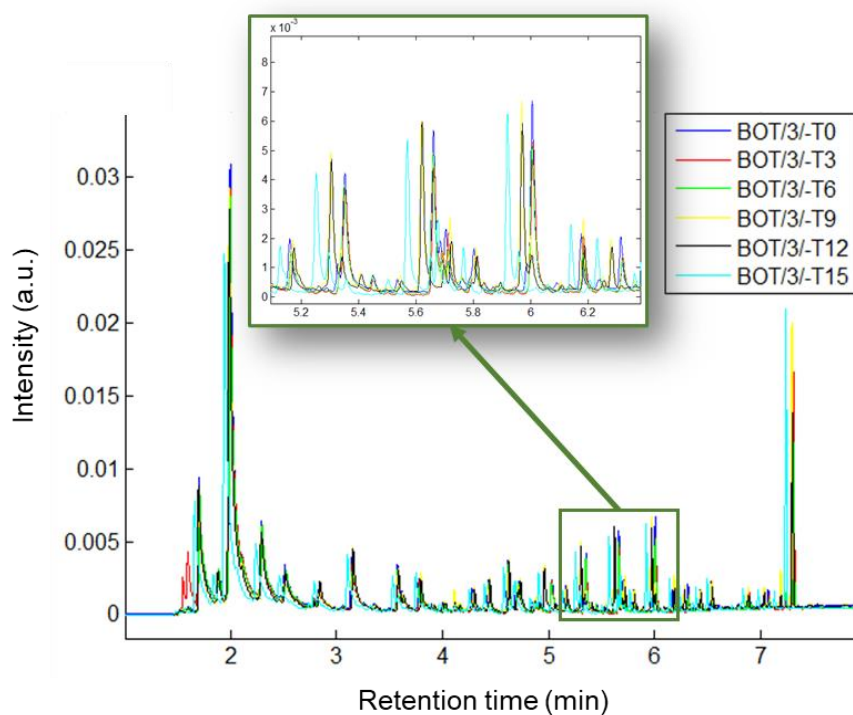
Una vez certificados los materiales de referencia se dio comienzo a la monitorización de la estabilidad para cada uno de ellos, de acuerdo al procedimiento descrito en la sección 3.2 del Capítulo IV de la Tesis. Desde el punto de vista del tratamiento de la señal, es en este estudio donde aparecieron mayores dificultades en la obtención de resultados comparables, ya que se desarrolla en condiciones de reproducibilidad. De hecho, a lo largo de este estudio se pudo observar que para cada MR los cromatogramas presentaban diferencias significativas causadas por:

- i) la evolución natural de la muestra biológica a pesar de encontrarse almacenadas en oscuridad y a bajas temperaturas.
- ii) la deformación de la señal como consecuencia de la degradación térmica de la fase estacionaria tras un uso continuado de la columna capilar.

En la Figura 21 se muestra una captura de pantalla de los cromatogramas obtenidos para cada una de las unidades analizadas de un mismo tipo de MR en los diferentes meses en los que se llevó a cabo la monitorización de su estabilidad. En esta figura se observa con claridad los efectos descritos previamente a lo largo de los diferentes meses seleccionados para su análisis, siendo más evidentes en la ampliación de la figura. De todas estas señales, aquellas correspondientes con la recepción de las muestras (T0) y el último mes de análisis en el que concluyó el estudio (T15) son las

que muestran mayores diferencias entre sí, al haberse evidenciado la coexistencia de ambos efectos (Figura 21 b).

a)



b)

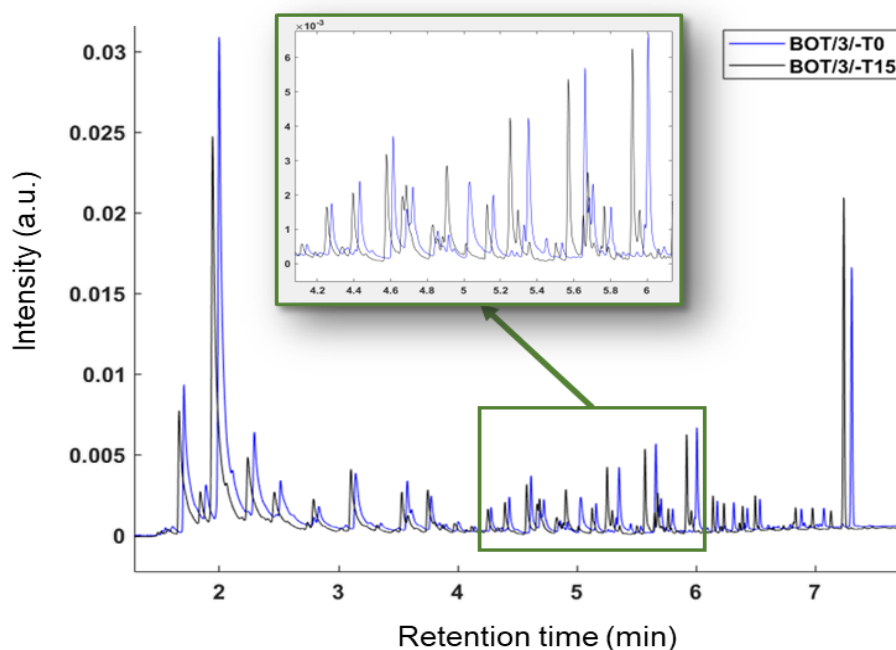


Figura 21. a) Cromatogramas relativos al material codificado como BOT/3/ en los diferentes meses de análisis, b) Cromatogramas relativos a los meses de análisis T0 y T15 para el mismo material

Aunque parecía que estos desplazamientos se podían corregir fácilmente aplicando los algoritmos típicos de alineamiento de señales, no siempre se obtuvieron los

resultados esperados detectándose que, dichos algoritmos comenzaban a dar fallos cuando:

- i) las señales proceden de diferentes tipos de muestra o
- ii) cuando las señales proceden del mismo tipo de muestra, pero han sido obtenidas en diferentes equipos o con meses de diferencia como serían los casos de estudio planteados.

Estos hechos obligarían a readaptar continuamente los parámetros de alineamiento, haciendo que fueran específicos para cada conjunto de señales y que dificultaran la posibilidad de reproducir los resultados en otros laboratorios o a otros tiempos.

VII.3. DESARROLLO Y PRIMERAS APLICACIONES DE LA AGNOSTICACIÓN INSTRUMENTAL EN CROMATOGRAFÍA DE GASES

Gracias al apoyo económico obtenido por parte del Fondo Europeo de Desarrollo Rural (FEADER), la Asociación Europea para la Innovación en materia de Productividad y Sostenibilidad Agrícola (EIP-AGRI) y a la Consejería de Agricultura Ganadería, Pesca y Desarrollo Sostenible de la Junta de Andalucía, por la financiación del proyecto denominado Sistema de Referencia y Armonización de Paneles de Cata de Aceite de Oliva Virgen – Proyecto INTERPANEL (GOP3-JA-16-0012), junto con el apoyo técnico y la financiación parcial recibida por parte de la Organización Interprofesional del Aceite de Oliva Español, se pudo **estudiar, desarrollar y aplicar un procedimiento íntegro que intentase minimizar la dependencia de la señal instrumental del estado del instrumento en el momento de su adquisición**. Fue la financiación de estos proyectos de investigación los que hicieron posibles las publicaciones recogidas en los Capítulos III y IV de esta Memoria, los cuales fueron realizados paralelamente en el tiempo.

Tal y como se indica en el Capítulo III, los primeros antecedentes en la "corrección de las desviaciones" producidas en los tiempos de retención de compuestos presentes en una muestra fueron facilitados por Kovats y Van Den Dool. Estos autores proponían el uso de Índices de Retención (I_R) obtenidos a partir de series homólogas de determinados compuestos químicos para obtener un sistema de referencia universal que les permitiese identificar compuestos orgánicos desconocidos. Dichos I_R se

obtuvieron por la interpolación del tiempo de retención del compuesto de interés con los tiempos de retención de los estándares eluidos antes y después de dicho componente.

Si la interpolación se diese, no solo entre el valor correspondiente con el tiempo de retención relacionado con el máximo de intensidad registrado para cada compuesto, sino con todos los valores que conforman la señal de la muestra, el resultado sería una huella instrumental basado en el sistema de referencia constituido por la serie de compuestos e independiente del tiempo de retención. El "nuevo" cromatograma así obtenido abandona el eje de tiempos de retención y se dibuja sobre un nuevo dominio caracterizado por los índices de retención que está relacionado con las propiedades químicas de la serie homologa de compuestos.

Desde la publicación de los trabajos de Kovats y Van Den Dool se han descrito múltiples familias de compuestos que podrían ser empleados para poder establecer el nuevo dominio del eje x. En todos los casos se indica que la selección de la serie homologa debe realizarse en función de las condiciones del propio análisis, ya que no todas las familias de compuestos presentan la misma estabilidad térmica en todos los rangos de temperatura. En el trabajo recogido en el Capítulo III se muestran dos casos de estudio en los que se hace uso de series homologas diferentes dependiendo de si la separación cromatográfica se realiza a baja (temperaturas del orden de 250-270 °C) o alta temperatura (temperaturas sobre 370-400 °C).

VII.3.1. Agnosticismo aplicado separaciones cromatográficas de baja temperatura

De acuerdo con lo descrito en el Capítulo III, para la agnostización de la huella instrumental, la mezcla de patrones externos debe ser analizada con un método similar al empleado en el análisis de la muestra, y además los compuestos que constituyen dicha mezcla deben abarcar el intervalo de tiempo en el que aparecen los compuestos de interés. En este caso se probaron dos disoluciones comerciales de n-alcenos preparados en n-hexano, la primera constituida por la mezcla C6-C20 y la segunda constituida por la mezcla C6-C40, cuyos cromatogramas se encuentran recogidos en la Figura 22. Las condiciones de análisis de la mezcla de patrones externos quedan recogidas en la Tabla 8.

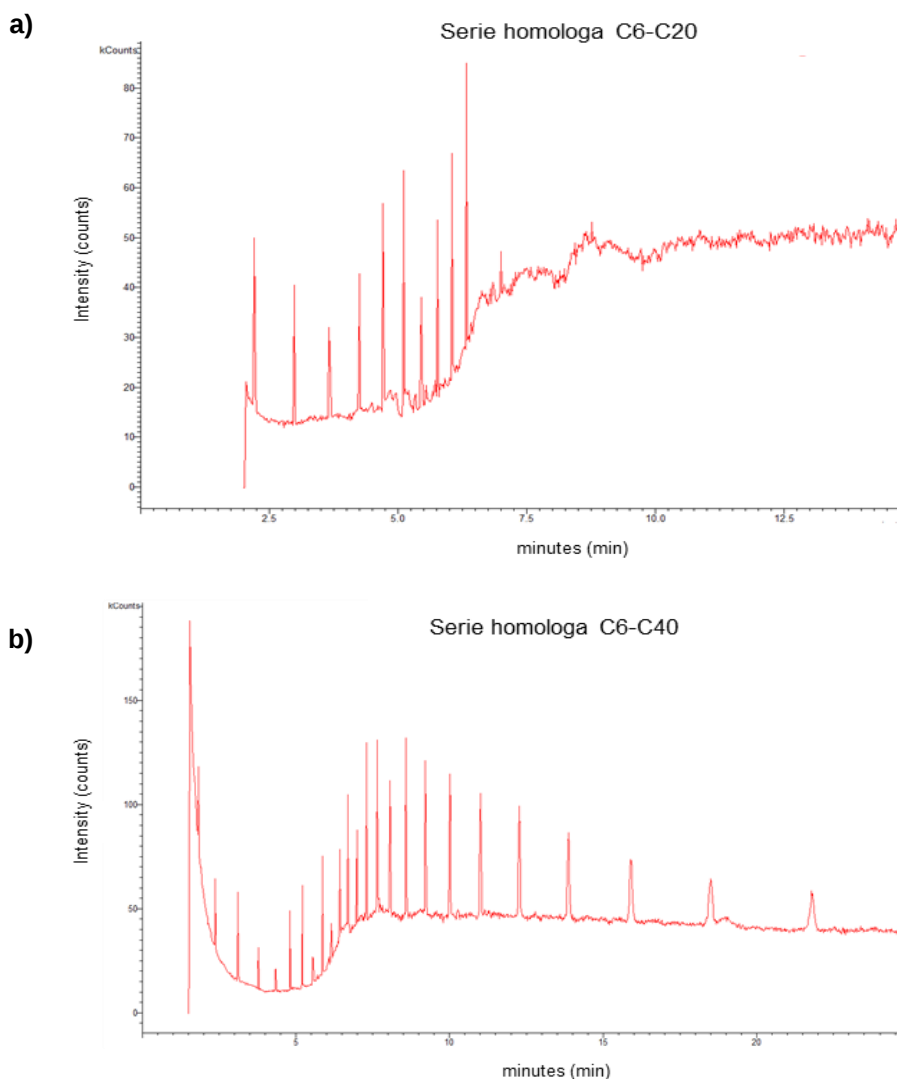


Figura 22. Cromatogramas obtenidos en el análisis de las dos series homologas de hidrocarburos empleadas a) C6-C20 y b) C6-C40

En el caso de la primera mezcla de *n*-alcanos, la separación se realiza en los primeros 6.3 min del análisis, tiempo insuficiente para cubrir todo el intervalo deseado de las muestras de aceite empleadas en el primer caso de estudio.

La segunda mezcla comercial requiere de casi 80 minutos para extraer de la columna capilar el *n*-tetracontano (C40), excediendo el intervalo en que aparecen los compuestos volátiles de las muestras de aceite. Hay que señalar que en la Figura 22b tan solo se muestran los 25 primeros minutos de análisis donde se aprecian claramente hasta el *n*-tetratriacontano (C34).

Tabla 8. Método cromatográfico empleado en la separación de la mezcla de patrones externos de n-alcanos.

Jeringa	
Volumen	10 µL
Columna capilar	
Nombre comercial	HP-INNOWAX
Fase estacionaria	100% polietilenglicol
Dimensiones	30 m × 0.25 mm × 0.25 µm
Rampa de temperatura programada	
Volumen de inyección	2 µL
Temp. Inyector	250 °C
Rampa Temp.	80 °C (2.5 min), 240 °C (40 °C/min), 240 °C (70 min)
Tiempo de análisis	78.5 min
Condiciones del espectrómetro de masas	
Retraso de disolvente	2 min
Temp. línea de transferencia	250 °C
Temp. Fuente de Ionización	250°C
Energía electrón de ionización	70 eV
Corriente de Emisión	10 µA
Velocidad de scan	1.2 s
Modo de adquisición	Full-scan, intervalo 10-400 m/z

Sobre la base a estos resultados, surgieron al menos dos posibilidades:

- i) trabajar con una serie homologa constituida por un numero intermedio de n-alcanos por ejemplo una mezcla C6-C30, o
- ii) emplear la segunda serie homologa (C6-C40) y desechar el exceso de información de los últimos 60 min de análisis los cuales no aportan ningún tipo de información o ventaja en el proceso de transferencia del comportamiento de la serie homologa a la muestra de interés.

Para los estudios recogidos en los Capítulos III y IV, los cuales fueron realizados de forma simultánea, se optó por la segunda opción. La principal razón se debió a la falta de disponibilidad de una disolución comercial C6-C30 en el momento en que se llevaron a cabo los diferentes análisis.

VII.3.2. Agnosticismo aplicado separaciones cromatográficas de alta temperatura

Llegados a este punto, la agnostización instrumental solo contemplaba el análisis a baja temperatura, pero no todos los compuestos químicos, volátiles o no volátiles, consiguen separarse en dicho intervalo. Por ejemplo, el análisis de triglicéridos se realiza en GC a temperaturas superiores a 300 grados en columnas capilares de alta polaridad y con una fase estacionaria altamente termoestable, para lograr no solo volatilizar este tipo de compuestos de elevado peso molecular, sino también lograr su separación tendiendo tanto al grado de insaturación como al número de carbonos.

El principal problema que tendría la aplicación de la metodología propuesta haciendo uso de la serie de *n*-alcanos sería la alta volatilidad que presenta a las temperaturas de trabajo lo que dificulta su separación cromatográfica. Dado que el método cromatográfico para la determinación de triglicéridos transcurre a alta temperatura sin producirse ningún tipo de degradación, se propuso como serie candidata a utilizar para la agnostización de la señal, una mezcla constituida por triglicéridos, sin importar que estos compuestos seleccionados a utilizar como patrones externos pudieran estar o no incluidos de forma natural en las muestras analizadas.

Para ello, en el segundo caso de estudio, se preparó en el propio laboratorio una mezcla de triglicéridos constituidos por tres cadenas iguales de ácidos grasos saturados (trimiristina, tripalmitina, triestearina y triaraquidina) para la obtención de la señal de referencia externa con la que poder aplicar la metodología de agnostización instrumental. Comparada con la serie homóloga de hidrocarburos, el número necesario para cubrir el intervalo de tiempos en que aparecen los triglicéridos en muestras de aceite de oliva es considerablemente menor. Como posible alternativa a ampliar el número de compuestos, cabría la posibilidad de incorporar a esta mezcla triglicéridos en los que los tres ácidos grasos, idénticos entre sí, incorporaran una o más insaturaciones (por ejemplo, trioleína). En los experimentos realizados en el laboratorio para los primeros ensayos se hizo uso de la triestearina y la trioleína, ambas analizadas por separado y haciendo uso de una rampa de temperatura más suave a la descrita en la sección 3.2 del Capítulo III y recogida en la Tabla 9. Se pudo comprobar que ambos compuestos estaban suficientemente separados como para ser identificados.

Sin embargo, dada la necesidad de modificar el gradiente de temperatura para disminuir el tiempo de análisis se optó por utilizar solo uno de ellos (en este caso la

triestearina). El motivo de esta elección se debió a que un aumento de la temperatura provocaba una disminución de la resolución entre los picos de ambos patrones, hecho que dificultaría la interpolación de aquellos triglicéridos contenidos en las muestras de aceite estudiadas que actuarían con un comportamiento similar a los patrones analizados en el sistema cromatográfico, dificultando a su vez la obtención de huellas instrumentales agnostizadas.

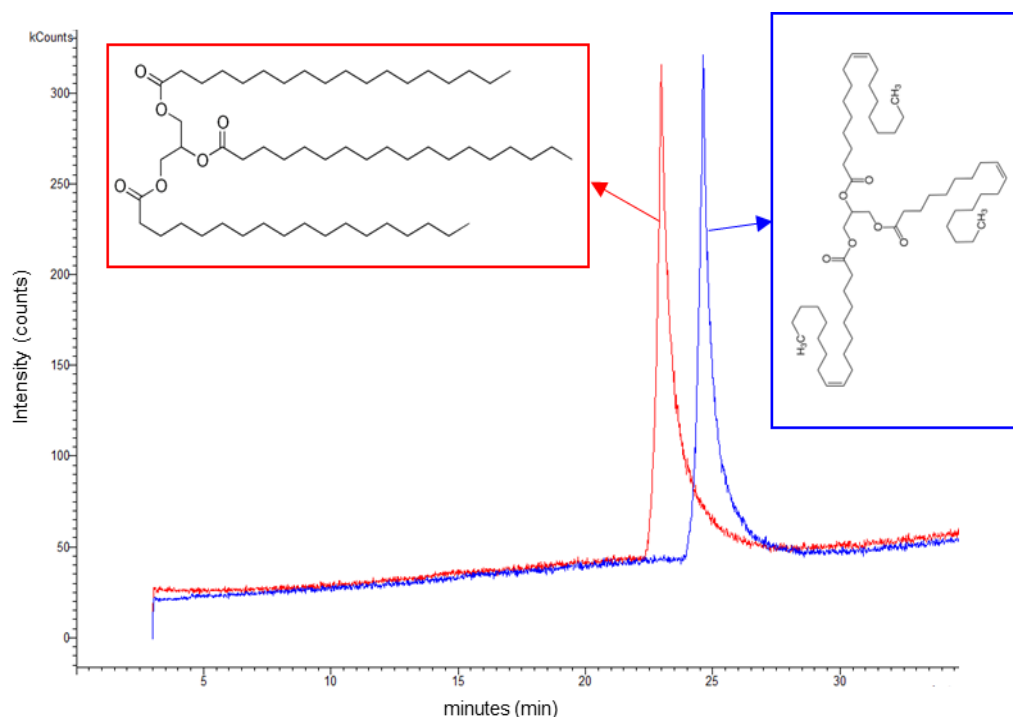


Figura 23. Cromatogramas obtenidos durante el análisis de la triestearina (marcada en rojo) y la trioleína (marcada en azul).

En la Figura 23 se muestra la separación de ambos compuestos aplicando las condiciones descritas en la Tabla 9.

Tabla 9. Método cromatográfico empleado en la separación de la mezcla de patrones externos de n-alcános.

Jeringa	
Volumen	10 µL
Columna capilar	
Nombre comercial	Rtx-65 TG
Fase estacionaria	Crossbond®65% diphenyl; 35% dimethyl polysiloxane
Dimensiones	30 m × 0.32 mm × 0.1 µm
Rampa de temperatura programada	
Volumen de inyección	2 µL
Temp. Inyector	370 °C
Rampa Temp.	315, 350 °C (1 °C/min)
Tiempo de análisis	35 min
Condiciones del espectrómetro de masas	
Retraso de disolvente	2 min
Temp. línea de transferencia	350 °C
Temp. Fuente de Ionización	250°C
Energía electrón de ionización	70 eV
Corriente de Emisión	10 µA
Velocidad de scan	1.2 s
Modo de adquisición	Full-scan, intervalo 200-1000 m/z

VII.4. NUEVAS APLICACIONES EN CROMATOGRAFÍA DE LÍQUIDOS EN FASE NORMAL

Una vez establecida y aplicada la metodología de agnostización instrumental en GC para dos casos de estudio en dos fracciones diferentes del aceite de oliva bien conocidas dentro del Grupo de Investigación FQM-232, se planteó la posibilidad de adaptar esta nueva metodología a LC en la modalidad de fase normal complementando las evidencias ya encontradas por miembros del Grupo referidas a la potencialidad de la agnostización instrumental en la modalidad de fase invertida. Para este estudio se contó únicamente con la financiación de los fondos propios del Grupo de Investigación. Las muestras empleadas en este estudio fueron elaboradas en el propio laboratorio para simular una situación en donde los aceites sufren mezclas no autorizadas en el transporte.

Al igual que sucede en GC, los valores de posición representados por el tiempo de retención están estrechamente relacionados con el estado del instrumento en el momento del análisis. Tanto es así que al comparar la señal obtenida del análisis de una misma muestra en diferentes equipos cromatográficos y en diferentes momentos, se observan desplazamientos en el tiempo de retención. En LC este efecto se ve además acentuado por las interacciones que se producen durante la separación de compuestos, incluso para un único instrumento a lo largo de un día de análisis. Mientras que en GC los analitos se separan únicamente por su interacción con la fase estacionaria a medida que son arrastrados por un gas portador inerte, en LC el analito no solo interactúa con la fase estacionaria, sino que además lo hace con el disolvente o mezclas de disolventes empleados como fase móvil, que a su vez interactúa con la propia fase estacionaria.

Comparada con la GC, donde existe un amplio abanico de posibilidades de series de compuestos homólogos que surgen a raíz de las publicaciones de Kovats y Van Den Dool, en LC este tipo de mezclas son bastante reducidas y solamente aplicables a determinados tipos de fases móviles y estacionarias. Además, todas las propuestas a día de hoy han sido para la modalidad de fase invertida.

El primer paso para la aplicación de la metodología en fase normal sería establecer una mezcla de compuestos químicos que cubrieran en la medida de lo posible el intervalo de tiempos en los que aparecen los compuestos de interés en el análisis de una muestra determinada. Estos compuestos pueden pertenecer a una misma familia, por ejemplo, isómeros de posición de un mismo compuesto al tener polaridades diferentes entre sí, o mezclas de varias familias de compuestos que, pese a no tener ningún tipo de relación o semejanza estructural entre sí, las condiciones de análisis hacen que su separación abarque el intervalo de trabajo. Esta primera etapa requiere de bastantes pruebas basadas en ensayo-error y consumen bastante tiempo.

Por ello, el primer paso para aplicar esta metodología sería establecer una huella instrumental de referencia obtenida del análisis de una mezcla de compuestos analizados como patrones externos. Esta etapa se trata a su vez del paso más complicado, ya que la mezcla de patrones externos se debe diseñar evaluando posibles compuestos de los que se sospecha que, bien por los grupos funcionales de los que dispone o bien por su estructura general, se puede presuponer que aparezcan a un tiempo de retención determinado y próximo al tiempo de retención de alguno de los compuestos presentes en la muestra.

El caso de los compuestos empleados en el Capítulo IV requirió de múltiples pruebas de ensayo-error donde se evaluaron familias muy diversas hasta dar con una mezcla que ocupase gran parte del intervalo de trabajo. El método empleado en el análisis de dicha mezcla se describe en la Tabla 10, mientras que la Figura 24 muestra gráficamente el gradiente de elución utilizado.

Tabla 10. Descripción del método de análisis empleado en la separación de la mezcla de compuestos externos.

Fase estacionaria	Sílica (Si 60)		
Fase móvil A	Hexano / 2-propanol (99:1)	Flujo	0.8 mL/min
Fase móvil B	2-Propanol	Presión de trabajo	21-22 bares
Tiempo	Fase A %	Fase B %	
0	100	0	
15	90	10	
17	90	10	
20	100	0	

Post-time 1 min

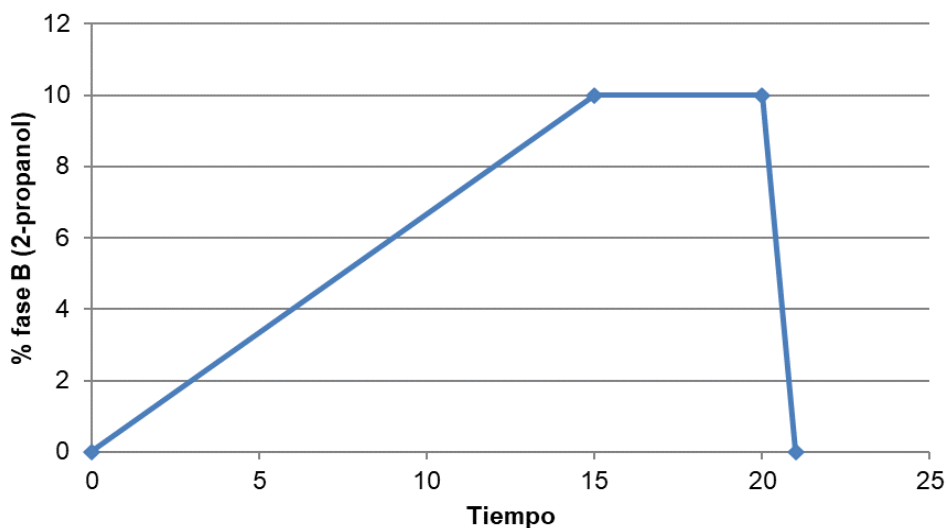


Figura 24. Representación gráfica del gradiente de elución aplicado en la separación de los patrones externos utilizados para la agnostización instrumental en (NP)HPLC-DAD.

Una lista de los compuestos que se presentaron como posibles candidatos antes de seleccionar la composición final de la mezcla de patrones se recoge en la Tabla 11.

Tabla 11. Compuestos químicos evaluados como posibles patrones externos para la aplicación de la metodología de agnostización instrumental en fase normal.

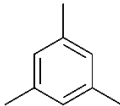
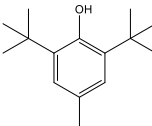
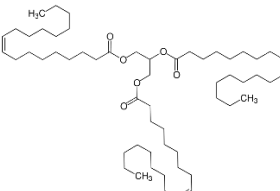
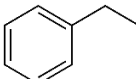
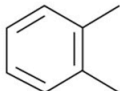
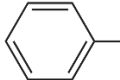
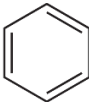
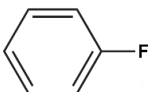
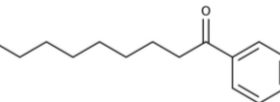
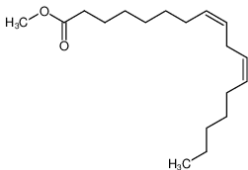
Compuesto	Formula	Estructura	Tiempo (min)
1,3,5- Trimetilbenceno	C_9H_{12}		3.16
Butilhidroxitolueno (BHT)	$C_{15}H_{24}O$		3.18
Trioleína	$C_{57}H_{114}O_6$		3.19
Etilbenceno	C_8H_{10}		3.19
o-Xileno	C_8H_{10}		3.20
Tolueno	C_7H_8		3.21
Benceno	C_6H_6		3.23
Fluorobenceno	C_6H_5F		3.28
Nonanofenona	$C_{15}H_{22}O$		3.42
Linoleato de metilo	$C_{19}H_{34}O_2$		3.42

Tabla 11. (cont.) Compuestos químicos evaluados como posibles patrones externos para la aplicación de la metodología de agnostización instrumental en fase normal.

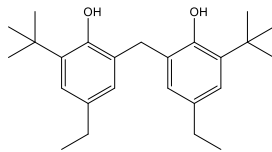
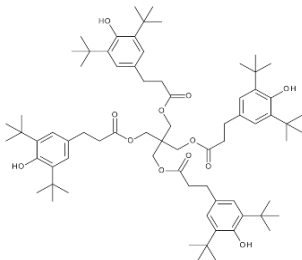
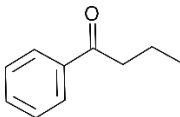
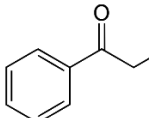
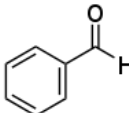
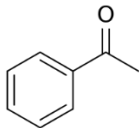
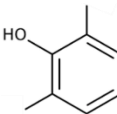
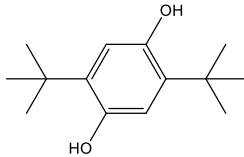
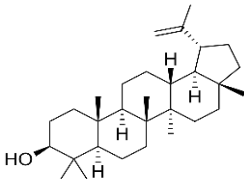
Compuesto	Formula	Estructura	Tiempo (min)
Irganox 425	$C_{25}H_{36}O_2$		3.51
Irganox 1010	$C_{73}H_{108}O_{12}$		3.63
Butirofenona	$C_{10}H_{12}O$		3.68
Propiofenona	$C_9H_{10}O$		3.90
Benzaldehído	C_7H_6O		4.21
Acetofenona	C_8H_8O		4.78
2,6-Dimetilfenol	$C_8H_{10}O$		5.16
2,5-Di- <i>t</i> -Butil hidroquinona	$C_{14}H_{22}O_2$		5.38
Lupeol	$C_{30}H_{50}O$		6.48

Tabla 11. (cont.) Compuestos químicos evaluados como posibles patrones externos para la aplicación de la metodología de agnostización instrumental en fase normal.

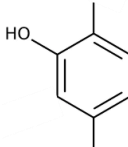
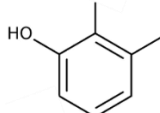
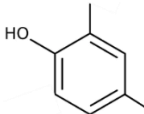
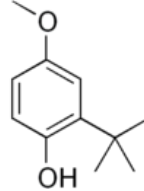
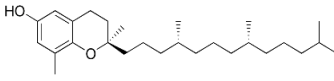
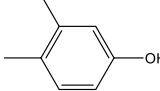
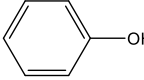
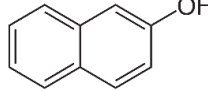
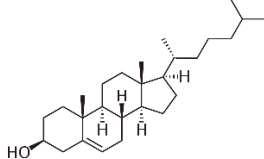
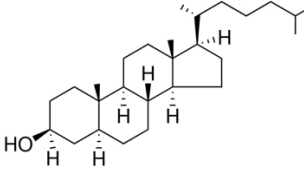
Compuesto	Formula	Estructura	Tiempo (min)
2,5-Dimetil fenol	$C_8H_{10}O$		6.44
2,3-Dimetil fenol	$C_8H_{10}O$		6.50
2,4-dimetilfenol	$C_8H_{10}O$		6.58
Butilhidroxianisol (BHA)	$C_{11}H_{16}O_2$		6.91
δ-Tocoferol	$C_{27}H_{46}O_2$		7.05
3,4-dimetilfenol	$C_8H_{10}O$		7.40
Fenol	C_6H_6O		7.69
2-Naftol	$C_{10}H_8O$		7.96
Colesterol	$C_{27}H_{46}O_2$		8.41
Colestanol	$C_{27}H_{48}O$		8.52

Tabla 11. (cont.) Compuestos químicos evaluados como posibles patrones externos para la aplicación de la metodología de agnostización instrumental en fase normal.

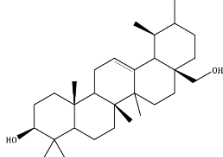
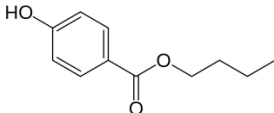
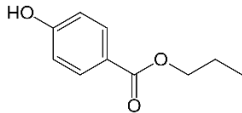
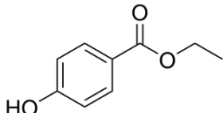
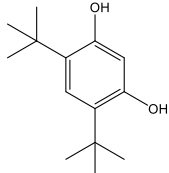
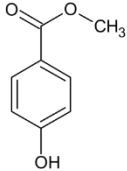
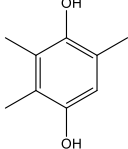
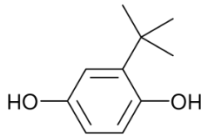
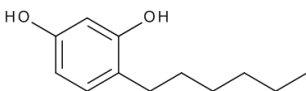
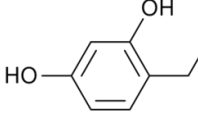
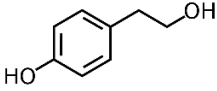
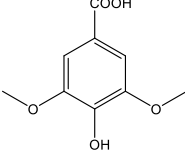
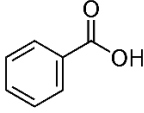
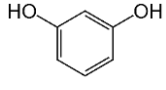
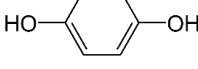
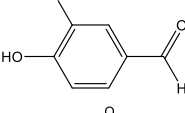
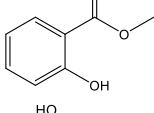
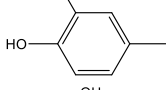
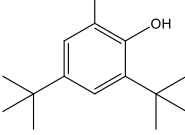
Compuesto	Formula	Estructura	Tiempo (min)
Uvaol	C₃₀H₅₀O₂		9.32
<i>p</i> -Hidroxibenzoato de butilo	C ₁₁ H ₁₄ O ₃		9.63
<i>p</i> -Hidroxibenzoato de propilo	C ₁₀ H ₁₂ O ₃		9.74
<i>p</i> -Hidroxibenzoato de etilo	C ₉ H ₁₀ O ₃		10.00
4,6-di- <i>t</i> -Butilresorcinol	C ₁₄ H ₂₂ O ₂		10.25
<i>p</i> -Hidroxibenzoato de metilo	C ₈ H ₈ O ₃		10.41
Trimetil hidroquinona	C ₉ H ₁₂ O ₂		10.89
<i>t</i> -Butil hidroquinona	C ₁₀ H ₁₄ O ₂		12.38
4-Hexilresorcinol	C ₁₂ H ₁₈ O ₂		12.43

Tabla 11. (cont.) Compuestos químicos evaluados como posibles patrones externos para la aplicación de la metodología de agnostización instrumental en fase normal

Compuesto	Compuesto	Compuesto	Compuesto
4-Etilresorcinol	C₈H₉O₂		12.84
Tirosol	C ₈ H ₁₀ O ₂		-
Ácido siríngico	C ₉ H ₁₀ O ₅		-
Ácido benzoico	C ₇ H ₆ O ₂		-
Resorcinol	C ₆ H ₆ O ₂		-
Hidroquinona	C ₆ H ₆ O ₂		-
Vanilina	C ₈ H ₈ O ₃		-
Metil-3-metoxisalicilato	C ₈ H ₈ O ₃		-
4-Metil catecol	C ₇ H ₉ O ₂		-
3,5-di- <i>t</i> -Butilcatecol	C ₁₄ H ₂₂ O ₂		-

Nota I. En negrita se resaltan los compuestos finalmente seleccionados.

Nota II. El guion "-" indica aquellos compuestos que quedaron completamente retenidos en la columna

La Figura 25 muestra un mapa de calor representando el espectro-cromatograma (señal 2D) obtenido para una de las muestras de aceites estudiados.

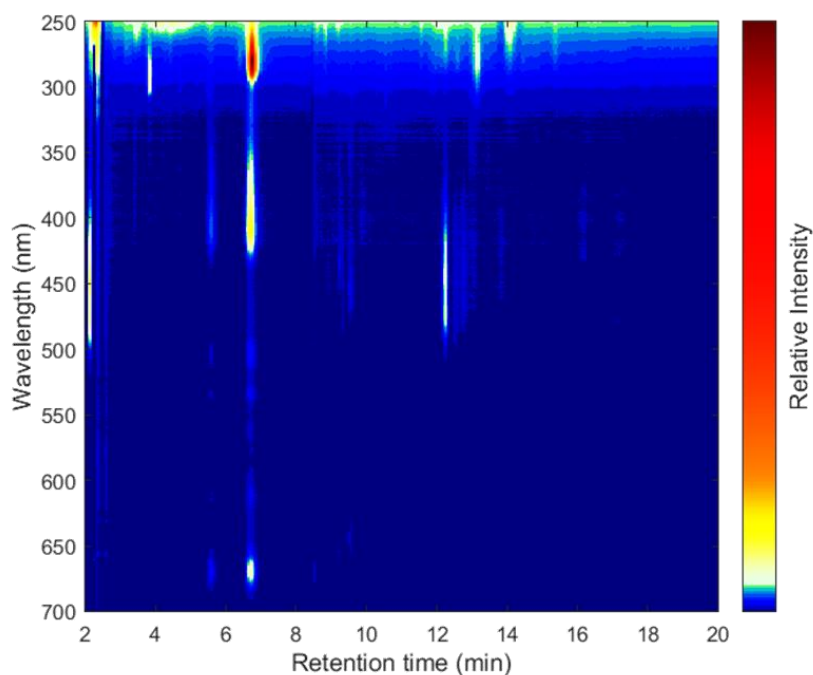


Figura 25. Espectro-cromatograma original obtenido tras el análisis de una muestra de aceite analizado mediante (NP)HPLC-DAD.

Si se tiene en cuenta que el último compuesto seleccionado como patrón externo en el estudio fue el 4-etilresorcinol (ver Tabla 11), que presenta un tiempo de retención aproximado de 12.8 min, se puede observar que la separación cromatográfica de los últimos compuestos se produce en torno al minuto 16 y por tanto la mezcla empleada para poder agnostizar la señal no abarcaba por completo el intervalo de trabajo utilizado. Esta falta de referencias en los últimos minutos se debió a la imposibilidad de encontrar compuestos estables y suficientemente puros con un precio asequible que en las condiciones de análisis pudieran aparecer en torno a esos valores de tiempo de retención.

Como el resto del intervalo de trabajo si se encontraba bien definido, se consideró la posibilidad de poder extrapolar la información a aquellos intervalos de trabajo no cubiertos por la mezcla de patrones, de tal manera que bajo esta premisa la huella instrumental 2D obtenida podría considerarse agnostizada (véase Figura 26). Esta aproximación se puede considerar válida ya que la extrapolación se realizó en un intervalo lo suficientemente estrecho como para que el posible error producido no fuese demasiado excesivo.

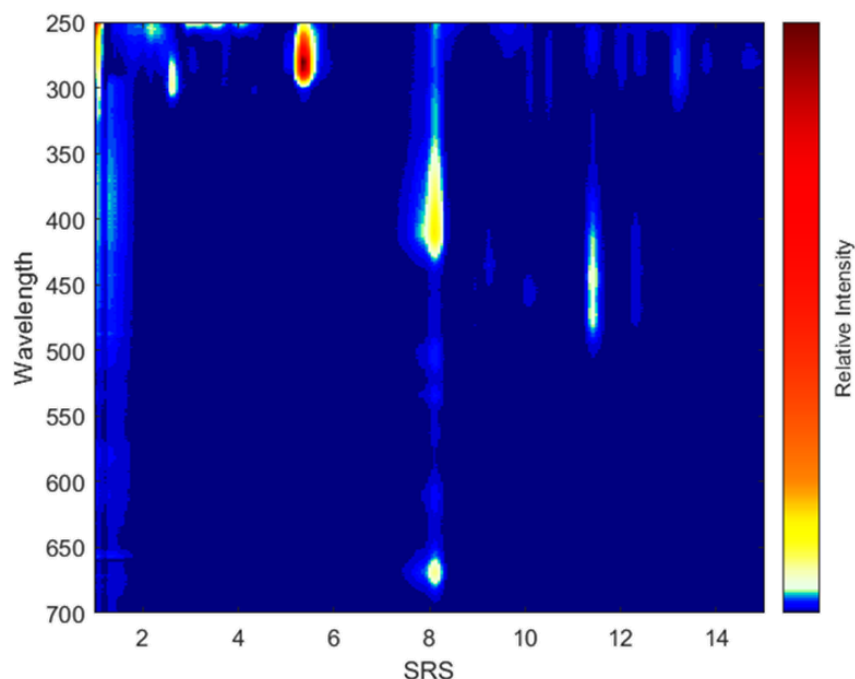


Figura 26. Huella instrumental 2D agnostizada obtenido tras el análisis de una muestra de aceite analizado mediante (NP)HPLC-DAD.

Una vez obtenida la matriz de huellas instrumentales agnostizadas para el análisis de los compuestos responsables del color, se procedió a la utilización de técnicas quimiométricas con la finalidad de evaluar posibles situaciones reales de mezclas fraudulentas lo que completó el trabajo experimental que se recoge en el Capítulo V y que como se ha comentado anteriormente, a fecha de entrega de esta Memoria, se encuentra en fase de revisión.

Otro hecho remarcable del mismo es el enfoque dado a la interpretación de los resultados obtenidos al aplicar el modelo de clasificación SIMCA. Mientras que la metodología seguida en el desarrollo de un modelo de clasificación de este tipo se basa en buscar una correcta pertenecía a las clases modeladas para considerar que el modelo es adecuado, en este caso se consideró que los resultados eran realmente buenos si como resultado final en el conjunto de validación, el modelo conducía a una mala clasificación de las muestras incluidas en el mismo, ya que el aumento del número de muestras no clasificadas de forma concluyente implicaría que los cambios producidos en la huella instrumental agnostizada se deberían exclusivamente a la

presencia de un aceite vegetal diferente al definido para cada clase tratándose por tanto de una mezcla fraudulenta.

VII.5. EXPANDIENDO LA METODOLOGÍA A NUEVAS MATRICES ALIMENTARIAS

Finalmente, gracias a la Ayuda de Movilidad Internacional 2020 para Estudiantes de Programas de Doctorado de la Universidad de Granada con la cual se pudo financiar el periodo de estancia internacional en la Università degli Studi di Torino (Italia), se procedió a la evaluación de la agnostización instrumental utilizando una nueva matriz alimentaria como es la avellana. En concreto, el Grupo de Investigación al que pertenece la Profesora Chiara Cordero (supervisora durante el periodo de estancia) se centra en el análisis de la calidad y seguridad de alimentos entre los que se incluyen las avellanas gracias al acuerdo de colaboración existente entre dicho grupo de investigación con el Grupo Ferrero, (Soremartec Italia S.R.L.).

En la nueva aproximación se planteó como hipótesis de partida el independizar la señal obtenida en el análisis de las muestras reales (avellanas) utilizando las propias huellas instrumentales de las mismas como referencias externas, es decir la utilización de la propia matriz como patrón externo para la agnostización.

Para conseguir tal independencia, se consideró que el hecho de disponer de suficiente información acerca de la composición de todo un conjunto de muestras analizadas, es decir que se dispone del conocimiento adecuado de los compuestos químicos se encuentran presentes en todas las muestras, en este caso de avellanas, independientemente de su variedad, país de origen o año de cosecha, dichos compuestos pueden seleccionarse como referencia externa y así evaluar su comportamiento a lo largo de sucesivos análisis. De este modo, la propuesta sería similar a una hipotética adición de la mezcla de patrones externos sobre las muestras analizadas, pasando a actuar dichos patrones externos como patrones internos.

Para alcanzar la agnostización de la huella se consideró necesario depurar la señal suministrada por el equipo con una doble finalidad:

- i) maximizar las intensidades de los compuestos representativos a todas las muestras y seleccionados como mezcla de patrones externos.

- ii) Minimizar la señal del resto de compuestos para evitar posibles interferencias durante la transferencia del comportamiento de dichos compuestos al resto de señales.

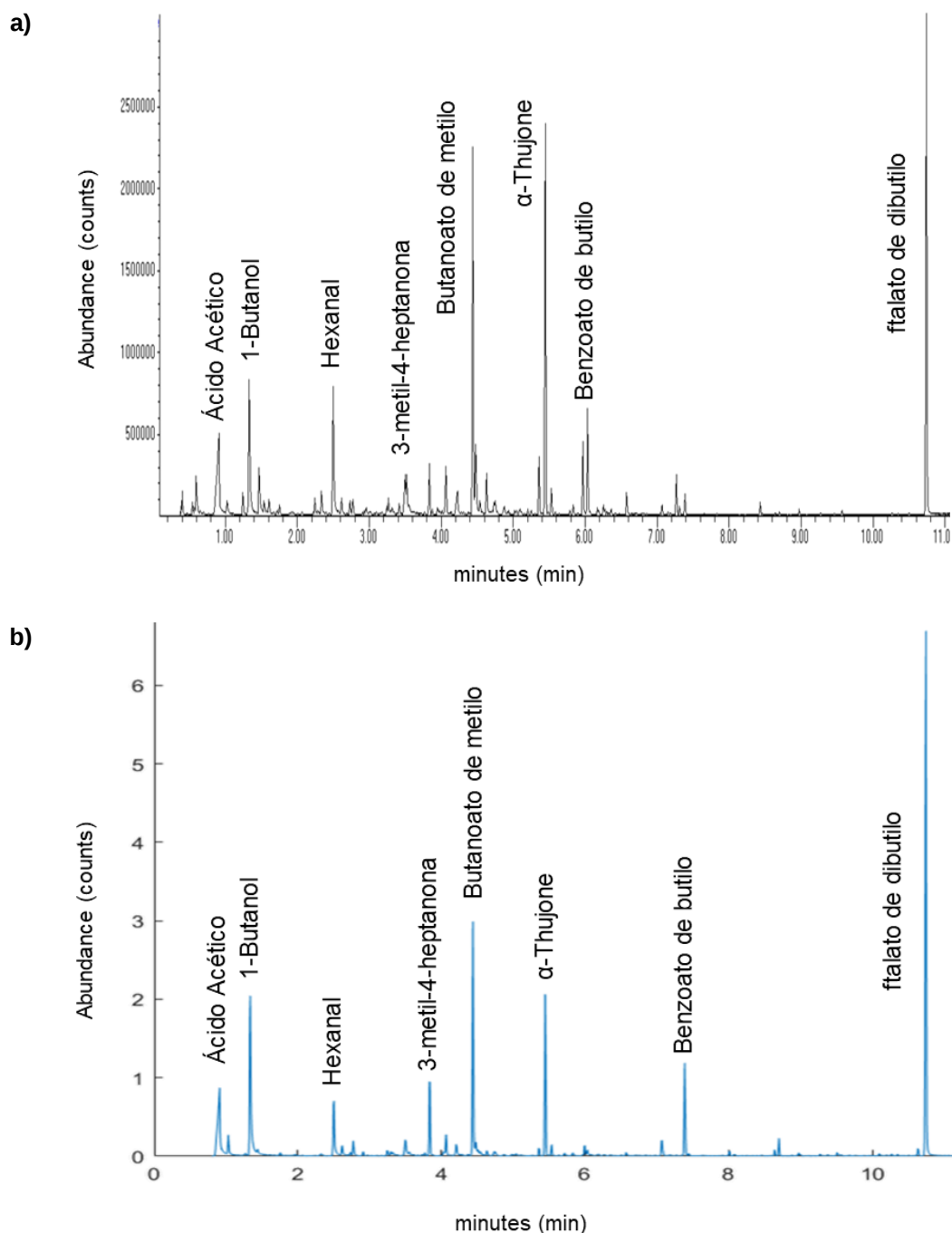


Figura 27. (cont.) a) Cromatograma obtenido tras el análisis de una muestra de avellana mediante (HS)-SPME-GC-MS antes del procesado/depuración de la señal; b) señal obtenida a partir del cromatograma a) tras su procesado y depuración. En ambas figuras se muestran los compuestos seleccionados como referencias externas para el proceso de agostización.

La Figura 27 muestra la misma señal (cromatograma), correspondiente al perfil de compuestos volátiles de una muestra de avellana, antes y después de su depuración.

A su vez, la Figura 28 muestra dos huellas instrumentales agnostizadas siguiendo la metodología propuesta.

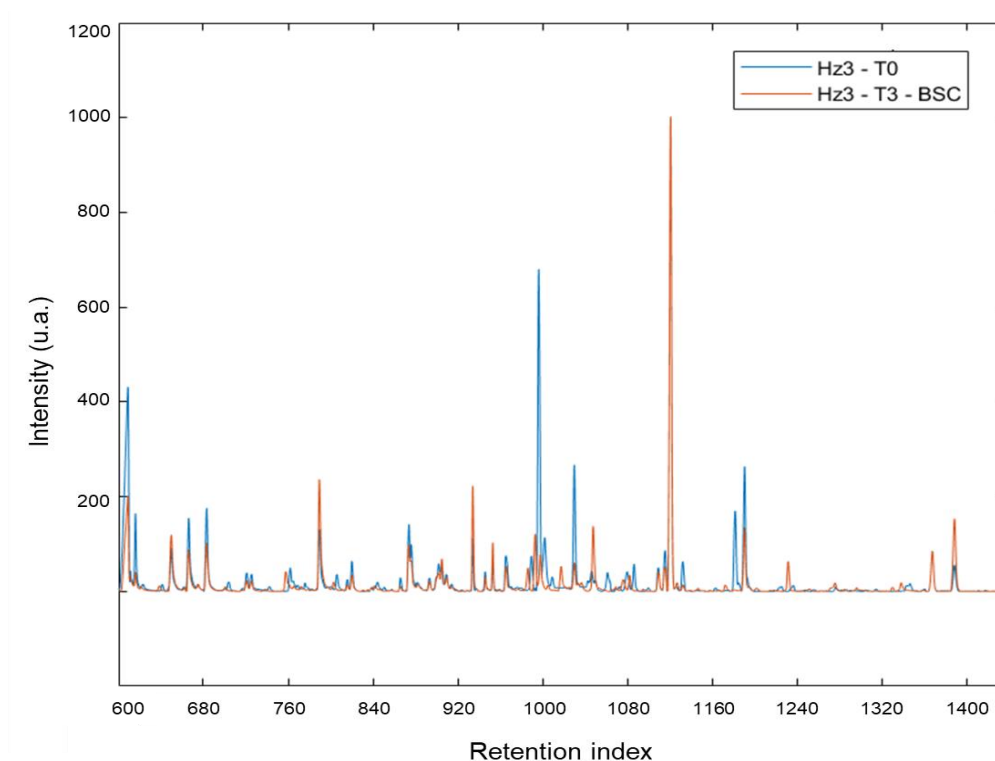


Figura 28. Huella dactilar agnostizada para la misma muestra de avellana analizadas en diferentes momentos (tiempo 0 y tiempo 2) siguiendo la nueva aproximación de la metodología.

El principal inconveniente de esta aproximación reside en la necesidad de disponer de información concreta acerca de la composición de las muestras a analizar. Esto limitaría su uso a aquellos equipos cromatográficos que estén acoplados a un espectrómetro de masas donde, con ayuda de las bases de datos incorporada en el software del equipo, se pueda proceder a la correcta identificación de todos los compuestos y a partir de la información obtenida, seleccionar aquellos que se vayan a utilizar como referencia.

Otra limitación importante encontrada radica en el hecho de que esta aproximación solo sirve para garantizar la no dependencia de la señal del momento de su adquisición, pero no del instrumento tal y como se plantea en uno de los objetivos originales de la propuesta. Esta limitación se puede justificar atendiendo a la falta de una referencia externa aislada de la muestra que permita establecer una relación inequívoca entre uno o varios instrumentos de forma rápida y sencilla.

Aunque la utilización de la señal producida por parte de los constituyentes de una muestra como conjunto de patrones para agnostizar la huella instrumental, ha permitido solucionar problemas asociados con la dependencia de la señal en el momento de la adquisición, la falta de una referencia externa y pura claramente identificada (serie externa de patrones) dificulta la posible transferencia del comportamiento entre diferentes instrumentos. Por tanto, para conseguir ambos fines y siempre que sea posible, es recomendable el uso de una mezcla de patrones analizados como referencia externa independientemente de si estos son o no una serie homologa de compuestos.

Desde el punto de vista quimiométrico, trabajar con huellas cromatográficas agnostizadas y al menos independientes del momento de adquisición de las señales instrumentales que se utilizan como datos originales, por ejemplo, permite contribuir a la mejora del ciclo de vida del modelo clasificatorio. Dicho de otro modo, en etapas sucesivas del ciclo las muestras predichas correctamente en una etapa anterior pueden utilizarse para reentrenar el modelo de clasificación mejorando los parámetros de calidad del mismo.

El uso de la agnostización instrumental, así aplicada y relacionada con la problemática asociada a la calidad alimentaria, permite extraer conclusiones acerca del país de procedencia, la región de cultivo o el estado de conservación del alimento estudiado, hecho que ha quedado demostrado en el caso de las muestras de avellanas utilizadas durante el desarrollo de este estudio.

Conclusiones

- Se establece una metodología novedosa y precisa para la evaluación de la homogeneidad de muestras de aceite de oliva mediante el uso de huellas instrumentales obtenidas mediante GC-FID correspondientes a la fracción de compuestos orgánicos volátiles (VOC), previamente aislados por microextracción en fase sólida.
- La estrategia cromatográfica complementa el uso del análisis sensorial cuando la finalidad del estudio se encuentra orientada a la evaluación de la homogeneidad, facilitando la certificación de MR de aceites de oliva virgen ya que estos resultan ser herramientas clave para el entrenamiento y armonización de paneles sensoriales.
- Esta estrategia general se basa en la similitud de señales permitiendo revelar la falta de homogeneidad entre unidades muestreadas del MR candidato. Los gráficos de control obtenidos a partir de los índices de similitud permiten comprobar, de forma sencilla, cuándo se superan los criterios de aceptación de los valores límite, de forma similar a lo dispuesto en el proceso de certificación.
- El uso de técnicas de reconocimiento de pautas no supervisados (como HCA y PCA) ha permitido confirmar la agrupación natural de las unidades pertenecientes a cada uno de los materiales de referencia e identificar posibles marcadores químicos.
- Se ha propuesto una metodología innovadora denominada "**agnosticismo instrumental**" que establece una mejora significativa en la obtención de señales cromatográficas armonizadas y comparables con el fin de conseguir nuevas huellas instrumentales o señales analíticas estandarizadas, sin importar el instrumento utilizado o la naturaleza de los datos brutos adquiridos.
- El uso de un conjunto de patrones adecuados, tanto internos como externos, permite transferir la influencia del instrumento, principalmente relacionada con disminuciones de intensidades de señal, o desplazamientos como consecuencia de la degradación de la fase estacionaria, al resto de huellas cromatográficas, independientemente de si estos se han obtenido en la misma tanda o en diversas tandas analizadas en días diferentes.

- El agnosticismo instrumental se presenta como una metodología prometedora para ser utilizada en el caso de señales analíticas que requieran un proceso de separación previo a su obtención. Esta metodología permite obtener señales reproducibles en las que, por ejemplo, un cambio en el perfil de la señal cromatográfica no suponga un problema a la hora de alinear huellas.
- El uso combinado del análisis sensorial e instrumental se presenta como una metodología rigurosa para la certificación de la intensidad organoléptica de atributos sensoriales, ya que las técnicas instrumentales han permitido la evaluación de la homogeneidad y estabilidad mediante la evaluación de la similitud entre huellas cromatográficas agnostizadas.
- Se ha propuesto un conjunto de filtros estadísticos de calidad desarrollados para el tratamiento de datos generados en diferentes sesiones de cata. Su evaluación en una situación real ha permitido obtener resultados estadísticamente fiables de forma rápida y sencilla. El conjunto de filtros propuestos también puede aplicarse en ensayos de armonización de paneles de cata.
- El método analítico desarrollado para el análisis directo de muestras de aceite de oliva por (NP)HPLC-DAD ha permitido la detección simultánea de diferentes familias de compuestos (algunos de ellos relacionados con el color de los aceites de oliva) cuya composición cambia cuando se produce una mezcla con cualquier otro aceite vegetal, lo que permite, no sólo una reducción considerable del tiempo de análisis y del procesamiento de datos necesarios para obtener un resultado fiable, sino también una reducción del coste del análisis.
- El uso de la agnostización instrumental minimiza la dependencia de la señal con respecto del sistema cromatográfico. Cabe destacar que el procedimiento desarrollado no solo se ha utilizado por primera vez con un conjunto de estándares químicos externos fiables, sino que además ha permitido evaluar la correcta aplicación de esta metodología en cromatografía líquida en fase normal.
- La aplicación de herramientas quimiométricas mediante métodos de reconocimiento de pautas supervisados y no supervisados a la huella bidimensional agnostizadas de mezclas de aceites ha revelado que las mayores diferencias se deben a la tonalidad presente en las mismas. El modelo de clasificación SIMCA desarrollado ha sido capaz de poner de manifiesto las

mezclas en aquellas muestras que originalmente tenían una fuerte tonalidad verdosa, pero no detecta las mezclas producidas cuando la tonalidad del AOVE original comercializado es diferente.

- El perfil de similitud desarrollado y aplicado en el análisis de mezclas de aceite de diferente calidad ha demostrado ser una herramienta sencilla para evitar los problemas de clasificación errónea, además de ser útil para evidenciar la existencia de mezclas independientemente de la tonalidad de la muestra.
- Como alternativa a los modelos de clasificación, el perfil de similitud, que no requiere una reducción de variables haciendo uso de toda la información especto-cromatográfica obtenida en el análisis, permite una reducción significativa del tiempo requerido para el procesamiento de datos y la toma de decisiones, facilitando la toma de decisiones.
- El análisis directo de muestras de aceite de oliva por (NP)HPLC-DAD combinado con herramientas quimiométricas y/o perfiles de similitud puede considerarse como una herramienta útil para la prevención/detección de posibles mezclas fraudulentas asegurando la trazabilidad de las muestras de aceite.
- En el caso de la avellana, la agnostización de huellas instrumentales se muestra como una herramienta útil para la discriminación de dicho alimento en el marco de la evaluación de la calidad ya que contempla el uso de información cifrada en la huella instrumental pero no cubierta por (bio)marcadores.
- La combinación de una etapa de pre-procesado (siendo la mejor opción en el caso de estudio el auto-escalado) de la huella con técnicas de reconocimiento de pautas no supervisadas, ha demostrado que, para el caso de las avellanas, la huella instrumental agnostizada contiene suficiente información sobre las condiciones de envejecimiento y almacenamiento de las muestras, así como el país de origen de las mismas.
- Las técnicas de reconocimiento de pautas supervisadas han permitido el desarrollo de modelos de clasificación atendiendo a la calidad de la avellana.
- Finalmente, el nuevo enfoque de agnostización instrumental en la que sólo se considera la variabilidad presente en todas las muestras, utilizando estas

mismas como señal de referencia para el alineamiento, junto a técnicas quimiométricas con diferentes opciones de pre-procesado, puede considerarse como un enfoque válido en la resolución de problemas asociados a la metodología de huellas instrumentales.