ESTUDIO ANALÍTICO DE LA FRACCIÓN TRANSESTERIFICADA DEL ACEITE DE OLIVA.

APLICACIÓN EN PROBLEMAS DE AUTENTIFICACIÓN DE ACEITE DE OLIVA

Tesis Doctoral

Programa oficial de Doctorado en Química

Ana María Jiménez Carvelo

Directores:

Dr. Luis Cuadros Rodríguez

Dr. Antonio González Casado

Para optar al título de:

Doctora en Química

Granada 2018

Editor: Universidad de Granada. Tesis Doctorales
Autor: Ana María Jiménez Carvelo
ISBN: 978-84-1306-093-4
URI: http://hdl.handle.net/10481/54772

Agradecimientos

Índice

ÍNDICE GENERAL

Resumen 23

Abstract 24

Presentación 27

Realización de la tesis doctoral 28

Estructura de la tesis doctoral 30

Contextualización 33

Aceite de oliva y sus métodos tradicionales de análisis 33

Metodología de huellas dactilares 38

Quimiometría 41

Capítulo de libro: *LC Fingerprinting approach for authenticating olive oil* 51

Artículo 1: *Emergent data mining/machine learning methods for the analytical evaluation of food quality and authenticity* 83

Comunicaciones a congresos derivadas de este capítulo 109

Problema 110

Hipótesis 111

Objetivos 112

Capitulo I. Huellas dactilares cromatográficas en fase normal 115

I.1. Presentación 115

I.2. Introducción 116

I.3. Instrumentación 121

I.4. Muestras 121

I.5. Condiciones experimentales 124

I.6. Pre-procesado de los datos 124

I.7. Métodos de clasificación multivariantes aplicados 124

Artículo 2: *One input-class and two input-class classifications for differentiating olive oil from other edible vegetable oils by use of the normal-phase liquid chromatography fingerprint of methyl-transesterified fraction* 129

Artículo 3: *A new analytical method for quantification of olive oil and palm oil in blends with other vegetable edible oils base on the chromatographic fingerprintints from methyl-transesterified fraction* 151

Articulo 4: *Classification of olive oil according to their cultivars based on second-order data using LC-DAD* 175

I.8. Discusión 195

Comunicaciones a congresos derivadas de este capítulo.................................... 196

Capitulo II. Huellas dactilares cromatográficas en fase invertida 199

II.1. Presentación 199

II.2. Introducción 200

II.3. Instrumentación 203

II.4. Muestras 204

II.5. Condiciones experimentales 204

II.6. Pre-procesado de los datos 204

II.7. Métodos de clasificación multivariantes aplicados 205

Artículo 5: *Fast-HPLC Fingerprinting to discriminate olive oil from other edible vegetable oils by multivariate classification methods* 209

II.8. Discusión 225

Comunicaciones a congresos derivadas de este capítulo.................................... 226

Capitulo III. Huellas dactilares espectroscópicas (espectrometría vibracional) 229

III.1. Presentación 229

III.2. Introducción 230

III.3. Instrumentación 233

III.4. Muestras 233

III.5. Condiciones experimentales 234

III.6. Pre-procesado de los datos 234

III.7. Métodos de clasificación multivariantes aplicados 234

Artículo 6: *Chemometric classification and quantification of olive oil in blends with any edible vegetable oils using FTIR-ATR and Raman* 239

spectroscopy ....

III.8. Discusión 271

Comunicaciones a congresos derivadas de este capítulo..................................... 272

Capitulo IV. Caracterización química de la fracción metil-transesterificada ... 275

IV.1. Presentación 275

IV.2. Introducción 276

IV.3. Instrumentación 280

IV.4. Muestras 282

IV.5. Condiciones experimentales 282

IV.6. Resultados y discusión 282

IV.7. Conclusiones 292

Estudios complementarios a la tesis doctoral 295

Participación en proyectos 299

Conclusiones finales y perspectivas futuras 303

Resumen/Abstract

RESUMEN

Esta Tesis Doctoral describe, desarrolla y aplica la metodología de huellas dactilares cromatográficas y espectroscópicas en el ámbito de la autentificación de aceite de oliva. Las aplicaciones se centran en la obtención de datos derivados de la medida instrumental de la fracción metil-transesterificada de muestras de aceite de oliva de diferentes categorías comerciales (virgen extra, virgen y oliva), aceite de orujo de oliva y muestras de otros quince tipos de aceites vegetales comestibles en los que se incluyen aceites de (por orden alfabético): avellana, cacahuete, canola, cártamo, colza, girasol, girasol alto oleico, lino, maíz, palma, semillas, sésamo, soja, trigo (germen), y uva (pepita).

El objetivo principal es el desarrollo de métodos analíticos rápidos que conlleven el empleo de diferentes técnicas analíticas como la cromatografía de líquidos acoplada a diferentes detectores o las espectroscopias vibracionales (infrarrojo y Raman), aplicando herramientas quimiométricas para extraer la información analítica relevante. A través de la aplicación conjunta de las técnicas analíticas y herramientas quimiométricas es posible aplicar diferentes procesos relacionados con la autentificación de aceite de oliva, como: discriminar entre aceite de oliva y otros aceites vegetales, detectar adulteraciones de aceite de oliva con otros aceites vegetales y cuantificar la proporción del mismo en mezclas con otros aceites vegetales.

Se aplican diversas técnicas analíticas, como la cromatografía de líquidos acoplada a un detector de aerosol en corona cargado (HPLC-CAD) y a un detector de fila de diodos (HPLC-DAD), la espectroscopia de infrarrojo cercano (FT-NIR) y medio (FT-MIR) y la espectroscopia Raman, para obtener diferentes huellas dactilares de la fracción metil-transesterificada de los aceites vegetales. Las matrices de datos correspondientes a estas huellas dactilares instrumentales serán tratadas aplicando una batería de técnicas y métodos propios de la quimiometría, como el análisis exploratorio de componentes principales (PCA), métodos de clasificación mediante análisis discriminante (DA) y de modelado de clases (CM), y métodos de calibración multivariable.

Además se aplica la cromatografía de gases acoplada a espectrometría de masas para caracterizar los componentes presentes en dicha fracción.

ABSTRACT

The present Doctoral Thesis describes, develops and applies the methodology of chromatographic and spectroscopic fingerprinting in the authentication olive oil field. The applications are based on the collecting of data as result of instrumental measuring of the methyl-transesterified fraction of olive oil samples in different market categories (extra virgin, virgin and olive), pomace olive oil and fifteen kinds of edible vegetable oils which include: hazelnut oil, peanut oil, canola oil, rapeseed oil, safflower oil, sunflower oil, sunflower high oleic oil, flax oil, corn oil, palm oil, seeds oil, sesame oil, soybean oil, wheat germ oil and grape seed oil.

The main aim is the development of fast analytical methods using different analytical techniques such as high- performance liquid chromatography coupled to different detectors and spectroscopy techniques as infrared (IR) and Raman and applying chemometric tools in order to extract the relevant information. If the analytical techniques and chemometric tools are implemented jointly it is possible to apply different processes for the authentication of olive oil so as to discriminate olive oil from other edible vegetable oils, to detect adulteration of olive oil with other edible vegetable oils and to quantify the proportion of olive oil in blends with other edible vegetable oils.

High performance liquid chromatography coupled to two detectors: (i) charged aerosol detector (HPLC-CAD) and (ii) diode array detector (HPLC-DAD), and Fouried transformed near (FT-NIR) and mid-infrared (FT-MIR), Raman spectroscopy are applied in order to collect different fingerprints from a methyl-transesterified fraction of the edible vegetable oils. The matrix of data corresponding to these fingerprints will be processed employing chemometric algorithms as principal component analysis (PCA), classification methods applying discriminant analysis (DA) and class modeling (CM) and multivariate calibration methods.

Finally, gas chromatography coupled to mass spectrometry detector is applied in order to characterize the compounds present in the methyl-transesterified fraction.

Presentación

PRESENTACIÓN

La tesis doctoral está regida por el Real Decreto 99/2011, de 28 de enero, por el que se regulan las enseñanzas oficiales de doctorado, modificado por el Real Decreto 534/2013 en lo relativo a la evaluación y defensa de la Tesis Doctoral [**1**].

El artículo 5 del Real Decreto establece las competencias que debe adquirir el doctorando:

a) *Compresión 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.*

Asimismo establece que una vez obtenido el título de Doctor, el cual debe proporcionar una alta capacitación profesional en ámbitos diversos, los doctores habrán adquirido, al menos, las siguientes capacidades y destrezas personales para:

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.*

---

[1] Real Decreto 99/2011, de 28 de enero, por el que se regulan las enseñanzas oficiales de doctorado (versión consolidada BOE-A-2011-2541), Boletín Oficial del Estado, (2016), 1- 17.

e) *Integrar conocimientos, enfrentarse a la complejidad y formular juicios con información limitada.*

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

Además, en al artículo 13 se declara 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 se enmarca dentro del Programa Oficial de Doctorado en Química, en la línea de investigación de "Metodologías de obtención de información analítica en sistemas reales", perteneciente a la Escuela de Doctorado de Ciencias, Tecnologías e Ingenierías de la Universidad de Granada.

Realización de la tesis doctoral

Esta tesis 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 del Dr. Luis Cuadros Rodríguez y del Dr. Antonio González Casado.

Además, durante el período de realización de la tesis se han llevado a cabo varias estancias de investigación en centros de investigación nacionales e internacionales.

- ❖ Estancia de tres meses de duración en el "Institute for Global Food Security" perteneciente a la Queen's University en Belfast (Irlanda del Norte). En esta estancia se aplicaron diferentes técnicas espectroscópicas que condujeron al capítulo III de esta Tesis Doctoral. Para la realización de esta estancia se contó con financiación del propio Grupo de Investigación "AnAMA". Curso académico 2015/2016.
- ❖ Estancia de dos meses de duración en el "Instituto Universitario de Plaguicidas y Aguas" (IUPA) de la Universitat Jaume I de Castellón (España). En esta estancia de aplicaron diferentes tecnologías de GC-MS para caracterizar la fracción metil-transesterificada del aceite de oliva, que condujeron al capítulo IV de esta Tesis Doctoral. Para la realización de esta estancia se recibió una ayuda por parte del Vicerrectorado de Investigación de la Universidad de Granada dentro del programa de Estancias Breves. Curso académico 2016/2017.
- ❖ Estancia de quince días de duración en el Departamento de Química Analítica i Química Orgánica de la Universitat "Rovira i Virgili" de Tarragona (España). Esta

estancia se realizó con el objetivo de aprender a desarrollar modelos de clasificación basados en fusión de datos de nivel bajo, medio y alto, obtenidos de señales analíticas de diferentes técnicas instrumentales: cromatografía de líquidos en sus dos modalidades de trabajo y técnicas espectroscópicas (FTIR y RAMAN). Para la realización de esta estancia se contó con financiación del propio Grupo de Investigación "AnAMA". Curso académico 2016/2017.

- Estancia de cuatro meses de duración en el Departamento de Química Analítica de la Facultad de Ciencias Bioquímicas y Farmacéuticas de la Universidad Nacional de Rosario en Rosario (Argentina). Esta estancia se realizó con el objetivo de aprender a trabajar con datos de orden 2 y a desarrollar modelos de regresión y clasificación. Las metodologías aprendidas en esta estancia condujeron a una publicación científica recogida en el capítulo I de esta Tesis Doctoral. Para la realización de esta estancia se recibió una beca por parte del Grupo Santander gestionada por la Universidad de Granada dentro del Programa Becas Iberoamérica. Santander Investigación. Curso académico 2017/2018.

Estructura de la tesis doctoral

Antes de dar comienzo a cada uno de los capítulos se desarrolla un apartado de "contextualización" donde se describe el estado del arte, con objeto de poner en contexto las metodologías aplicadas en la Tesis Doctoral, así como las revisiones bibliográficas que se realizaron, que dieron lugar a dos publicaciones científicas y a dos comunicaciones en congresos, y seguidamente se indican el problema, hipótesis y objetivos de la misma.

A continuación se expondrán cada uno de los capítulos. Cada capítulo comenzará con una breve presentación y una introducción acerca de las metodologías específicas aplicadas para poner en contexto el trabajo realizado. Así mismo se transcribirá una copia del manuscrito correspondiente a los artículos científicos, publicados o enviados a publicación, derivados del capítulo, una discusión de los resultados obtenidos destacando las relevancias principalmente encontradas, así como sus aportaciones para dar solución a las necesidades de la comunidad científica, y por último un listado de las comunicaciones a congresos.

Finalmente se presentan algunos estudios complementarios que se han desarrollado de forma paralela a la elaboración de esta a la tesis doctoral, las conclusiones finales de la misma y las perspectivas futuras.

Contextualización

*El **olivar** hace bien, aunque le hagan el mal*

Anónimo

CONTEXTUALIZACIÓN

Aceite de oliva y sus métodos tradicionales de análisis

Los aceites vegetales constituyen una de las principales fuentes de grasas en la dieta, siendo éste el caso de la dieta mediterránea donde el aceite de oliva es uno de los principales ingredientes. El aceite de oliva tal y como establece el Reglamento (CE) nº 1234/2007 [**1**] por el que se crea una organización común de mercados agrícolas y se establecen disposiciones específicas para determinados productos agrícolas, define el aceite de oliva virgen como:

> *Aceites obtenidos del fruto del olivo exclusivamente por medios mecánicos u otros procedimientos físicos aplicados en condiciones que excluyan toda alteración del producto, y que no se han sometido a ningún otro tratamiento que no sea su lavado, decantación, centrifugado o filtración, excluidos los aceites obtenidos con el uso de disolventes o de coadyuvantes de acción química o bioquímica, por un procedimiento de reesterificación o como resultado de cualquier mezcla con aceites de otros tipos.*

La fracción mayoritaria del aceite de oliva está compuesta por triglicéridos o triacilgliceroles (98-99%) que suponen los componentes principales del aceite de oliva. Entre ellos, destacan por su abundancia, el trioleatil glicerol o trioleína (OOO), el dioleatil-palmitil glicerol (OOP y OPO), el dioleatil-linoleatil glicerol (LOO y OLO) y el linoleatil-oleatil-palmitil glicerol (LOP, OLP, LPO). En lo que respecta a los ácidos grasos constituyentes de dichos triglicéridos, el ácido oleico (C18:1) es el que se encuentra en mayor proporción (70-75%) seguido obviamente de los ácidos palmítico (10-14%) y linoléico (7-9%) [**2**,**3**].

La fracción minoritaria constituye alrededor de un 2% del peso del aceite de oliva y está compuesta principalmente por esteroles, fosfolípidos, hidrocarburos, tocoferoles, polifenoles y colorantes (habitualmente denominados en bibliografía como pigmentos).

---

[1] Reglamento (CE) nº 1234/2007 del Consejo de 22 de octubre de 2007 por el que se crea una organización común de mercados agrícolas y se establecen disposiciones específicas para determinados productos agrícolas (Reglamento único para las OCM), Diario Oficial de la Unión Europea, L 299, 1-149.

[2] Boskou, D. (1998). Composición del aceite de oliva. En: Boskou D. (Ed). Química y tecnología del aceite de oliva. Editorial Mundi Prensa Libros, S.A.

[3] Dionisi, F., Hug B., Kamm, W. (2006). Autenticidad de grasas y aceites. En: Ducauze, C.J (Coord). Fraudes alimentarios. Legislación y metodología analítica. Editorial Acribia.

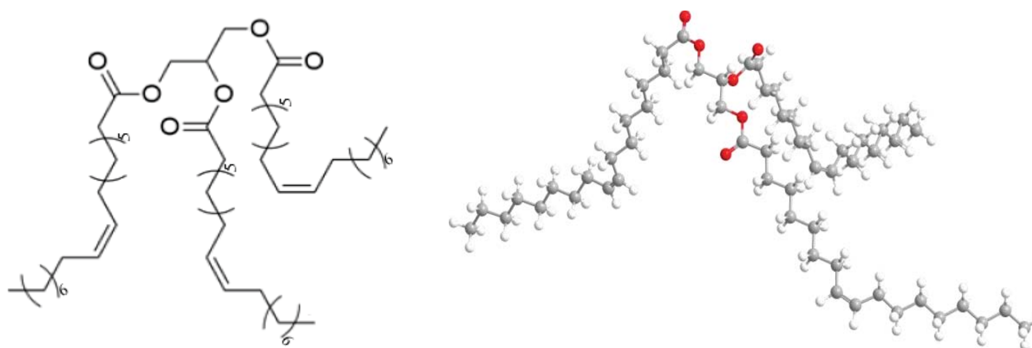


Figura 1. Estructura de la trioleína (OOO).

En la actualidad el Reglamento (CEE) nº 2568/91 relativo a las características de los aceites de oliva y de los aceites de orujo de oliva y sobre sus métodos de análisis [**4**], establece determinados valores de los parámetros de calidad y pureza para la comprobación de la conformidad de una muestra de aceite de oliva con la categoría declarada. En total existen alrededor de 50 parámetros que deben ser determinados para establecer la categoría del aceite, de los cuales el 80% son parámetros de pureza. La estimación de estos últimos parámetros se lleva a cabo para garantizar la autenticidad del aceite de oliva y detectar posibles adulteraciones del mismo con otros aceites vegetales. Por ejemplo valores altos en el contenido de las ceras indican una posible adulteración de aceite de oliva con aceite de orujo de oliva, ya que este último es obtenido de la pasta resultante de la molienda, rica en piel de aceituna, donde se encuentran la mayor proporción de estos compuestos.

Existen dos Reglamentos Europeos que regulan la información que se le debe dar al consumidor del alimento que se comercializa en relación con el tipo y contenido de aceites vegetales. El Reglamento (UE) nº 1169/2011 [**5**] sobre la información alimentaria facilitada al consumidor establece que cuando se declara que un alimento contenga mezcla de diferentes aceites vegetales es obligatorio declarar qué tipos de aceites vegetales han sido usados para la mezcla sin necesidad de especificar la proporción, y el Reglamento de Ejecución (UE) nº 29/2012 [**6**] sobre las normas de comercialización del aceite de oliva establece que cuando se declara que un alimento contiene aceite de oliva, se debe indicar obligatoriamente la proporción en la que se encuentra.

---

[4] Reglamento (CE) nº 2568/1991 de la Comisión de 11 de Julio de 1991 relativo a las características de los aceites de oliva y de los aceites de orujo de oliva y sobre sus métodos de análisis (versión consolidada 01991R2568), Diario Oficial de la Unión Europea (2016), 031.001, 1-128.

[5] Reglamento (UE) nº 1169/2011 del Parlamento Europeo y del Consejo de 25 de octubre de 2011 sobre la información alimentaria facilitada al consumidor, Diario Oficial de la Unión Europea, L 304, 18-20.

[6] Reglamento de Ejecución (UE) nº 29/2012 de la Comisión de 13 de Enero de 2012 sobre las normas de comercialización del aceite de oliva (versión consolidada 2012R0029), Diario Oficial de la Unión Europea (2016), 005.001, 1-15.

Además indica que:

Los Estados miembros podrán prohibir la producción en su territorio, para consumo interno, de las mezclas de aceite de oliva y otros aceites vegetales.

Si bien la legislación Europea permite la mezcla de aceite de oliva con otros aceites vegetales, en España existe una legislación específica que prohíbe las mezclas de aceite de oliva con otros aceites vegetales [**7**].

El aceite de oliva virgen es más caro que otros tipos de aceites vegetales comestibles debido a su proceso de obtención totalmente mecánico, y no siendo necesario someterlo a refino para el consumo humano como sucede con el resto de aceites vegetales. Además se ha demostrado que posee unas propiedades saludables excepcionales frente al colesterol, hipertensión, diabetes, así como acciones antiinflamatorias y antioxidantes, etc. [**8**,**9**,**10**,**11**]. Son por estos motivos por lo que es un buen candidato para ser adulterado con otros aceites vegetales para abaratar costes y tener un rendimiento económico más favorable y para obtener ciertas características de estabilidad y calidad nutricional de las mezclas resultantes.

Tradicionalmente para autentificar el aceite de oliva, se han aplicado técnicas separativas como la cromatografía de gases (GC) y la cromatografía de líquidos de altas prestaciones (HPLC), principalmente en su modalidad de fase invertida [**12**,**13**,**14**]. Los

---

[7] Real Decreto 308/1983, de 25 de enero, por el que se aprueba la Reglamentación Técnico-Sanitaria de aceites vegetales comestibles (versión consolidada BOE-A-1983-5543), Boletín Oficial del Estado, (2015), 1- 13.

[8] Gorzynik-Debicka, M., Przychodzen, P., Capello, F., Kuban-Jankowska, A., Gammazza, A.M., Knap, N., Wozniak, M., Gorska-Ponikowska, M. (2018). Potential health benefits of olive oil and plant polyphenols. International Journal of Molecular Science, *19*,547, 1-13.

[9] Wongwarawipat, T., Papageorgiou N., Bertisias, D., Siasos, G., Tousoulis, D. (2018). Olive oil-related anti-inflammatory effects on atherosclerosis potential clinical implications. Endocrome, Metabolic & Immune Disorders-Drug Targets, *18*, 51-62.

[10] Mariotti, M., Peri, C. (2014). The composition and nutrition properties of extra-virgin olive oil. En: Peri, C. (Ed). The Extra-Virgin Olive Oil Handbook. Editorial John Wiley & Sons.

[11] Fernández-Castillejo, S., Rubió, L., Hernández, A., Catalán, U., Pedret, A., Valls, R.M., Mosele, J.I, Covas, M.I., Remaley, A.T., Castañer, O., Motilva, M.J., Solá, R. (2017). Determinants of HDL cholesterol efflux capacity after virgin olive oil ingestion: interrelationships with fluidity of HDL monolayer. Molecular Nutrition and Food Research, *61*, 1700445 (1-9).

[12] Ben-Ayed R., Kamoun-Grati, N., Rebai, A. (2013). An Overview of the authentication of olive tree and oil. Comprehensive Reviews in Food Science and Food Safety, *12*, 218-226.

[13] Bajoud, A., Bandini, A., Fernández-Gutierrez, A., Carrasco-Pancorbo, A. (2018). Olive oil authentication: A comparative analysis of regulatory frameworks with special emphasis on quality and authenticity indices, and recent analytical techniques developed for their assessment. A review. Critical Reviews in Food Science and Nutrition, *58 (5)*, 832-857.

[14] Marikkar, J.M.M., Mirghani, M.E.E., Jaswir, I. (2016). Application of chromatographic and Infra-red spectroscopic techniques for detection of adulteration in food lipids: A review. Journal of Food Chemistry & Nanotechnology, *2*, 32-41.

datos experimentales recogidos por estas técnicas se han centrado principalmente en el estudio de los componentes mayoritarios, triglicéridos y/o ácidos grasos, y de algunos componentes minoritarios como esteroles, fenoles, tocoferoles, volátiles, etc. [**15**,**16**,**17**,**18**]. El perfil y/o la proporción de estos compuestos, dependerá de la variedad del fruto (picual, hojiblanca, arbequina, royal, frantoio, etc.) y de las condiciones agronómicas del cultivo, así como del proceso de extracción del aceite.

En la tabla 1 se recogen estudios sobre adulteraciones de aceite de oliva con otros aceites vegetales comestibles.

Tabla 1. Estudios sobre la fracción mayoritaria y minoritaria, para detectar adulteraciones de aceite de oliva con otros aceites vegetales.

Compuestos	Matriz	Técnica analítica	Ref.
Triglicéridos	AOVE-AA	HPLC-MALDI-ToF	[**19**]
	AOVE-AS; AOVE-AM; AOVE-AG; AOVE-AC	(DI)ESI-MS	[**20**]
	AOVE-AG	SFC-Q-ToF	[**21**]
	AOVE-AG; AOVE-AO	MALDI-ToF/MS	[**22**]

[15] Bosque-Sendra, J.M., Cuadros-Rodríguez, L., Ruiz-Samblás, C., De la Mata, P. (2012). Combining chromatography and chemometrics for the characterization and authentication of fats and oils triacyglyceerol compositional data- A review. Analytica Chimica Acta, *724,* 1-11.

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

[17] Tena, N., Aparicio-Ruiz, R., Koidis, A., García-González, D.L. (2018). Analytical tools in authenticity and traceability of olive oil. En: Montet, D. & Ray, R.C. (eds). Food traceability and authenticity: analytical techniques. Editorial: CRC Press, Taylor & Francis Group.

[18] Esteki, M., Simal-Gandara, J., Shahsavari, Z., Dashtaki, E., Heyden, Y.V. (2018). A review on the application of chromatographic methods, coupled to chemometrics, for food authentication. Food Control, *93*, 165-182.

[19] Calvano, C.D., Aresta, A., Zombonin, C.G. (2010). Detection of hazelnut oil in extra-virgin olive oil by analysis of polar components by micro-solid phase extraction based on hydrophilic liquid chrtomatography and MALDI-ToF mass spectrometry. Journal of Mass Spectrometry, *45,* 981-988.

[20] Alves, J.O., Sena, M.M., Augusti, R. (2014). Multivariate calibration applied to ESI mass spectrometry data: a tool to quantify adulteration in extra virgin olive oil with inexpensive edible oils. Analytical Methods, *6,* 7502-7509.

[21] Tu, A., Du, Z., Shuping, Q. (2016). Rapid profiling of triacyglycerols for identifying authenticity of edible oils using supercritical fluid chromatography-quadruple time-of-flight mass spectrometry combined with chemometric tools. Analytical Methods, *8,* 4226-4238.

[22] Jergovic, A.M., Persuric, Z., Saftic, L., Kraljevic Pavelic, S. (2017). Evaluation of MALDI-TOF/MS technology in Olive oil adulteration. Journal of American Oil Chemists' Society, *94,* 749-757.

Tabla 1. *Continuación*.

Compuestos	Matriz	Técnica analítica	Ref.
Ácidos grasos	AOVE-AS; AOVE-AG AOVE-AM; AOVE-AP; AOVE-AOO	GC-FID	[**23**]
	AOVE-AS; AOVE-AC; AOVE-AM	GC-MS	[**24**]
Tocoferoles	AOVE-AG	HPLC-FLD	[**25**]
Esteroles	AOVE-AG; AOVE-AS; AOVE-AC; AOVE-AA, AOVE-AC; AOVE-OP; AOVE-AM	GC-FID	[**26**]
Volátiles	AOVE-AG, AOVE-M, AOVE-AS	UV-IMS	[**27**]

EVOO: Aceite de oliva virgen extra; AA: Aceite de avellana; AS: Aceite de soja; AM: Aceite de maíz; AG: Aceite de girasol; AC: Aceite de canola; AO: Aceite de oliva (aceite de oliva refinado y aceite de oliva virgen); AP: Aceite de palma; AOO: Aceite de orujo de oliva; OP: Oleína de palma; AS: Aceite de semillas.

Para llevar a cabo estos análisis, el método habitualmente aplicado para el fraccionamiento del aceite de oliva con fines analíticos ha sido la saponificación (figura 2a). El problema de esta metodología es que es bastante tediosa y larga. En 1993 Bierderman *et al.* [**28**] propusieron una nueva estrategia de fraccionamiento que consistía en la transesterificación del aceite de oliva con metanol (figura 2b). Este nuevo método ofrece numerosas ventajas frente a la tradicional saponificación, ya que reduce el tiempo de trabajo, no se forman jabones, disminuye la cantidad de disolventes a utilizar y de muestra, mejora la extracción de compuestos minoritarios como los esteroles y es más rápido.

[23] Jabeur, H., Zribi, A., Bouaziz, M. (2016). Extra-Virgin olive oil and cheap vegetable oils: distinction and detection of adulteration as determined by GC and chemometrics. Food Analytical Methods, *9,* 712-723.

[24] Zhang, L., Yuan, Z., Li, P., Wang, X., Mao, J., Zhang, Q., Hu, C. (2017). Targeted multivariate adulteration detection based on fatty acid profiles and Monte Carlos one-class partial least squares. Chemometrics and intelligent laboratory system, *169,* 94-99.

[25] Bakre, S.M., Gadmale, D.K., Toche, R.B. (2015). Rapid determination of alpha tocopherol in olive oil adulterated with sunflower oil by reversed phase high-performance liquid chromatography. Journal of Food Science and Technology, *52(5),* 3093-3098.

[26] Srigley, C.T., Oles, C.J., Fardin Kia, A.R., Mossoba, M.M. (2016). Authenticity assessment of extra olive oil: evaluation of desmethylsterols and triterpene dialcohols. Journal of American Oil Chemists' Society, *93,* 171-181.

[27] Garrido-Delgado, R., Muñoz-Perez, E., Arce, L. (2018). Detection of adulteration in extra virgin olive oils by using UV-IMS and chemometric analysis. Food Control, *85,* 292-299.

[28] Biedermann, M., Grob, K., Mariani, C. (1993). Transesterification and on-line LC-GC for determining the sum of free and esterified sterols in edible oils and fats. Fett Wissenschaft Technologie (Fat Science and Technology), *95*, 127–133.

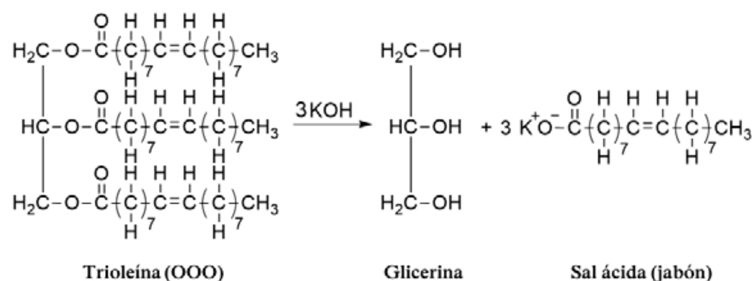


Figura 2a. **Saponificación** de la trioleína.

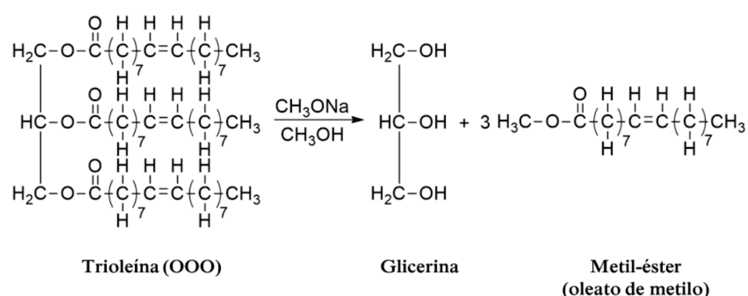


Figura 2b. **Transesterificación** de la trioleína.

Durante el transcurso de la transesterificación tienen lugar la rotura de las moléculas de todos los acil derivados del glicerol (triglicéridos y fosfolípidos) y otros compuestos de tipo éster (ceras, esteroles esterificados), produciéndose la formación de los esteres metílicos de los ácidos grasos y la liberación de la glicerina y, en su caso, de los esteroles. En este proceso se obtienen dos fracciones, la fracción acuosa que contiene los compuestos polares, y la fracción orgánica (fracción metil-transesterificada) donde se encuentran los esteres metílicos de los ácidos grasos, los esteroles y alcoholes terpénicos, mono y diglicéridos y otros compuestos. Es en esta última fracción donde Bierderman *et al.* identificaron metil esteroles, dimetil esteroles y alcoholes alifáticos.

Metodología de huellas dactilares

El aceite de oliva, al igual que otros aceites vegetales comestibles, después de la etapa de fraccionamiento, pueden ser analizados directamente por diferentes técnicas analíticas para obtener una **señal instrumental inespecífica** y aplicar la **metodología de "huellas dactilares"**.

En este contexto, la huella dactilar instrumental se puede definir como una señal inespecífica que se obtiene en poco tiempo, alrededor de 5 minutos, máximo 10 minutos y que contiene la suficiente información de un aceite como para poder diferenciarlo de otro aceite [**29**].

Esta metodología de huellas dactilares se basa en el mismo fundamento de identificación de personas por medio de la huella dactilar; brevemente consiste en

[29] Cuadros-Rodríguez, L., Ruiz-Samblás, C., Valverde-Som, L., Pérez-Castaño, E., González-Casado, A. (2016). Chromatographic fingerprinting: An innovative approach for food 'identitation' and food authentification. Analytical Chimica Acta, *909,* 9-23.

considerar en todo su conjunto el perfil de la señal instrumental inespecífica sin necesidad de identificar ni cuantificar ningún compuesto [**30**,**31**]. Por ejemplo, cuando se aplican técnicas cromatográficas para la obtención de "huellas dactilares cromatográficas" es común no tener picos bien resueltos y tener "jorobas", pero en consecuencia el tiempo de análisis cromatográfico se verá reducido considerablemente pero manteniendo toda la información de interés. Esta metodología se aleja de los esquemas convencionales de la cromatografía clásica donde prevalece la resolución de los picos del cromatograma con objeto de identificar cada compuesto o familia de compuestos. Por este motivo se debe recurrir a técnicas más avanzadas de tratamiento de datos para extraer la información útil, como son las técnicas quimiométricas.

Para la autentificación de aceite de oliva la metodología de huellas dactilares puede realizarse de tres formas o enfoques diferentes:

1. Un primer enfoque donde se obtiene la huella dactilar del aceite vegetal sin que éste haya sido sometido a ningún tratamiento. Se analiza en su forma y estado natural. Por ejemplo cuando se obtiene la huella dactilar cromatográfica de los triglicéridos del aceite de oliva.
2. Un segundo enfoque por el cual se obtiene la huella dactilar del aceite vegetal sometido previamente a una etapa de separación o fraccionamiento, para posteriormente aislar una familia concreta de compuestos. Por ejemplo cuando se obtiene la huella dactilar cromatográfica de los polifenoles o de los compuestos volátiles del aceite.
3. Un último enfoque en el que se obtiene la huella dactilar del aceite vegetal sometido previamente a una reacción química, de forma que se produce una alteración en la composición química del aceite y se forman nuevos compuestos químicos. Es similar a aplicar una reacción de formación de derivados o de 'derivatización'. Este último enfoque sería cuando se realiza una reacción de saponificación o de metil-transesterificación sobre el aceite vegetal.

Las técnicas analíticas más habituales de obtención de huellas dactilares son las cromatografías de líquidos, la cromatografía de gases, ambas acopladas a diferentes tipos de detectores, las técnicas espectroscópicas como las basadas en infrarrojo cercano (NIR), medio (FTIR) y Raman, y en resonancia magnética nuclear (RMN). Es importante remarcar que dependiendo de la configuración instrumental, los datos o huellas dactilares presentaran diferentes grados de complejidad en su procesado.

---

[30] Esslinger, S., Riedl, J., Fauhl-Hassek, C. (2014). Potential and limitations of non-targeted fingerprinting for authentication of food in official control. Food Research International, *60* 189-204.

[31] Ellis, D.I., Brewster, V.L., Dunn, W.B., Allwood, J.W., Golovanov, A.P., Goodacre, R. (2012). Fingerprinting food: current technologies for the detection of food adulteration and contamination, Chemical Society Reviews, *4*, 5706-5727.

En lenguaje matemático los conjuntos de números que representan una única información constituyen un tensor de datos, de forma que un escalar es un tensor de orden cero, un vector es un tensor de primer orden, una matriz es un tensor de orden dos, etc. Existe una nomenclatura para los diferentes tipos de datos obtenidos de los diferentes instrumentos analíticos: un tensor de **orden 0** es el obtenido por un equipo que genera un único dato numérico individual para una muestra, esto sería el equivalente a realizar por ejemplo una calibración univariante; un tensor de **orden 1** es un vector, es decir, que para una única muestra el instrumento ha generado múltiples medidas, esto sería el equivalente a obtener la huella dactilar aplicando NIR, FTIR, Raman, RMN, HPLC-UV-Vis y GC-FID; un tensor de **orden 2** es un matriz de datos, este tipo de datos son típicos de equipos instrumentales más complejos como HPLC-MS ó GC-MS, donde por cada muestra a un tiempo instrumental dado se tiene un espectro de masas determinado, o LC-DAD y LC-FLD donde por cada tiempo se tiene un espectro de absorción o de fluorescencia, respectivamente. No existe límite de órdenes de datos, pero es complicado encontrarse tensores de **orden 3** (o superior), que sería el equivalente a obtener un cubo de datos por cada muestra medida.

En la figura 3a se muestra una huella dactilar cromatográfica de un aceite de oliva obtenida por HPLC-CAD (orden 1), y en la figura 3b se muestra una huella dactilar cromatográfica de un aceite de oliva obtenida por HPLC-DAD (orden 2).

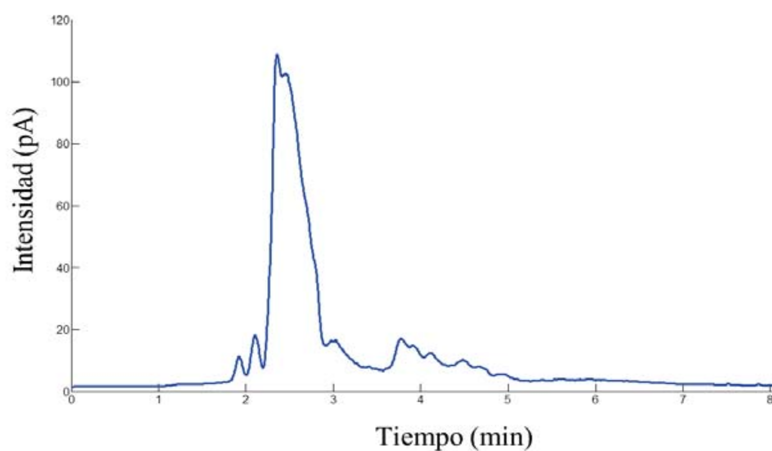


Figura 3a. **Huella dactilar cromatográfica** de la fracción metil-transesterificada de un AOVE, obtenida usando HPLC-CAD en la modalidad de fase normal.

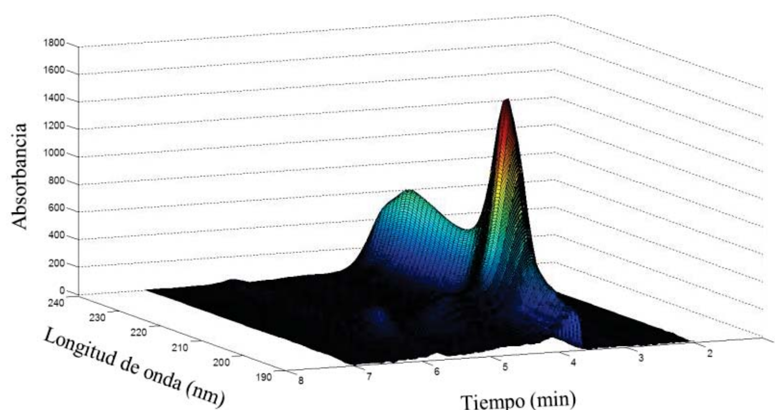


Figura 3b. **Huella dactilar cromatográfica** de la epoxidación de la fracción metil-transesterificada de un AOVE, obtenida usando HPLC-DAD en la modalidad de fase normal.

En general cuando se trabaja con más de un dato por muestra, es decir, tensores de orden 1 o superior se denomina análisis multivariante o análisis multivía. En este tipo de análisis cuando se tiene un conjunto de datos de muestras de orden 1, se les denominan **datos de dos vías**, ya que la unión de todos los vectores de un conjunto de muestras genera una matriz de datos. De este modo a los datos de orden 2 se les denomina **datos de tres vías** ya que generan un cubo, a los datos de orden 3 **datos de cuatro vías**, los cuales generan lo que se conoce como "matriz de cuatro vías", llamado así por sus denominación en inglés "four-way data array", y así sucesivamente [**32**]. En la figura 3 se muestra un esquema de la nomenclatura y el tipo de datos en análisis multivariante.

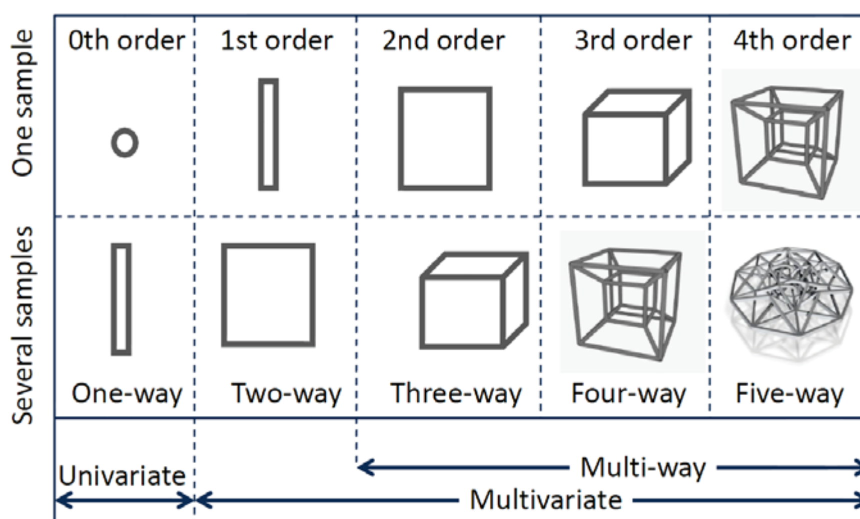


Figura 4. Nomenclatura del tipo de datos en análisis multivariante [**33**].

Quimiometría

Si se trabaja con la metodología de huellas dactilares instrumentales es necesario el uso de la quimiometría con objeto de extraer la información útil presente en la huella, aunque no de forma evidente. La quimiometría se define como *la ciencia que utiliza métodos matemáticos y estadísticos para extraer la mayor información de un grupo de señales analíticas complejas* [**34**,**35**,**36**].

---

[32] Escandar, G., Goicoechea, H.C., Muñoz de la Peña, A., Olivieri, A.C, (2014).Second- and higher-order data generation and calibration: A tutorial. Analytical Chimica Acta, *806*, 8-26.

[33] Olivieri, A.C, Escandar, G. (2014).Practical three-way calibration. Editorial Elsevier

[34] Hopke, P.K. (2003). The evolution of chemometrics. Analytical Chimica Acta, *500*, 365-377.

[35] Hibbert, D.B.; Minkkinen, P.; Faber, N.M.; Wise, B.M. (2009). IUPAC project: A glossary of concepts and terms in chemometrics. Analytical Chimica Acta, *642*, 3-5.

[36] Hibbert, D.B. (2016). Vocabulary of concepts and terms in chemometrics (IUPAC Recommendations. Pure and Applied Chemistry, *88*, 407-443.

Antes de aplicar cualquier algoritmo quimiométrico es necesario realizar una etapa de **pre-procesado de los datos** con objeto de linealizar las respuestas y eliminar fuentes extrañas de variabilidad para no entorpecer la estimación del modelo multivariante. Habitualmente las herramientas de pre-preocesado se aplican en función del tipo de datos con el que se trabaja. Por ejemplo cuando se obtienen huellas dactilares cromatográficas las etapas de pre-procesado o los pre-tratamientos más comunes son:

1. *Agrupación y superposición de todas las huellas dactilares:*

   En el caso de trabajar con datos de orden 1, consiste en generar la matriz de datos que contenga todos y cada uno de los vectores medidos de cada objeto de análisis. En este caso se tendrá una matriz de dimensiones (NxM), donde "N" (filas) se corresponde con el número de muestras analizadas y "M" (columnas) con el número de variables o de datos totales de cada vector. Para datos de orden 2, se genera un objeto tridimensional (cubo) o "arreglo de tres vías" de dimensiones (NxTxM), siguiendo con el ejemplo mostrado en la figura 3b, "N" corresponde al número de muestras, "T" corresponde a un modo instrumental (modo temporal) y "M" al otro modo instrumental (modo espectral).

2. *Corrección de la línea base:*

   Se aplica para eliminar desviaciones sistemáticas que pudiera darse en la línea base de las muestras analizadas debido a variaciones propias del equipo instrumental como la deriva entre el cromatograma de una muestra y otra, debido a factores de temperatura, presión y del propio detector. Es importante eliminar o miminizar lo máximo posible estas variaciones para que el modelo no se vea afectado.

3. *Reducción de variables (o decimación)*:

   Cuando se trabaja con una cantidad de datos/variables muy elevadas los recursos informáticos disponibles a veces son insuficientes llegando incluso en ocasiones a no ser capaces de elaborar los modelos y dar errores por la cantidad de datos que se manejan, principalmente cuando se trabaja con datos de orden 2. Por ese motivo en esta etapa en cromatografía se pueden seleccionar 1 de cada 2 variables (o en cualquier otra proporción no superior a 10) con lo que el número de variables de la huella se reduce a la mitad (o hasta la décima parte). Se debe tener la precaución de asegurar que, cuando se lleve a cabo esta decimación, el perfil de la huella se ha de mantener.

4. *Filtrado de ruido (suavizado):*

   Se centra en la eliminación de aquella componente de la señal que es independiente de la información analítica derivada de la muestra. Normalmente se refiere al ruido que acompaña a la señal analítica de interés y que no está relacionado con dicha señal analítica. Igual que en la etapa de decimación, es importante controlar el

suavizado para evitar la pérdida de información analítica. El algoritmo más utilizado para llevar a cabo el suavizado es el sugerido por Savitzky & Golay [**37**].

5. *Alineamiento de los picos cromatográficos:*

   En cromatografía esta etapa es probablemente la más importante, sobre todo en el caso de la cromatografía de líquidos donde los tiempos de retención suelen desplazarse más que en la cromatografía de gases que suele ser más reproducible. Existen diferentes algoritmos para alinear picos, los más habitualmente usados en cromatografía son COW [**38**,**39**] e ICOSHIFT [**40**]. En la tesis se ha aplicado este último para alinear los correspondientes picos. El algoritmo icoshift divide los cromatogramas en segmentos y éstos se van desplazando hacia los lados y alineándose con el pico de referencia que se encuentre en el segmento. Es importante que la selección de los intervalos que componen cada segmento sea correctamente seleccionada para evitar lo que se conoce en cromatografía como "artefactos".

6. *Centrado en la media:*

   Este pre-tratamiento sirve para corregir las diferencias entre las variable, y ello se logra variando los datos en torno a la media [**41**]. Se lleva a cabo calculando el valor medio de las variables de cada columna ($\bar{x}_i$) y restándolo a cada variable de la columna (x_{ij}):

   $$M_{ij} = x_{ij} - \bar{x}_j$$

   donde i es el número de fila y j el número de columna.

7. *Autoescalado:*

   Este pre-tratamiento se aplica normalmente cuando se tienen datos en diferentes escalas o adquiridos por diferentes equipos, por ejemplo cuando se aplica fusión de datos en donde es habitual fusionar datos medidos por diferentes equipos y con

---

[37] Savitzky, A., Golay, M.J.E. (1964). Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry, *36(8)*, 1627-1639.

[38] Vest Nielsena, N.P,Carstensenb, J.M., Smedsgaarda, I. (1998). Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimesed warping. Journal of Chromatography A, *805*, 17-35.

[39] Tomasi, G., Van der Berg, F., Andersson, C. (2004). Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. Journal of Chemometrics, *18*, 231-241.

[40] Tomasi, G., Savorani, F., Engelsen, SB. (2011). Icoshift: An effective tool for the alignment of chromatographic data. Journal of Chromatography A, *1248(43)*, 7832-7840.

[41] Bro, R., Smilde, A.K. (2003). Centering and scaling in component analysis. Journal of Chemometrics, *17*, 16-33.

objeto de que todos tengan el mismo peso en el modelo se les aplica un autoescalado. El autoescalado se realiza aplicando la siguiente ecuación:

$$M_{ij} = \frac{x_{ij} - \overline{x}_j}{s_j}$$

donde x_{ij} es la variable en cada columna, $\bar{x}_j$ es la media de las variables de cada columna y s_j es la desviación estándar de las variables de cada columna.

Cuando se trabaja con técnicas espectrométricas, como la espectroscopia de infrarrojo, habitualmente se aplican los pre-procesados 1 y 4 anteriormente citados, y además otro diferente denominado SNV [**42**], del inglés "*standard normal variate*", cuyo objetivo es eliminar la dispersión de los datos entre muestras. Este pre-tratamiento se realiza aplicando la misma ecuación matemática que para el autoescalado, pero difiere en la selección del elemento de cada vector que se utiliza al ser aplicado en la ecuación matemática. En el caso del autoescalado el elemento de cada vector, así como la media de las variables se calcula por columnas, mientras que en el caso del SNV se realiza por filas.

Una vez llevado a cabo el pre-procesado de los datos, se aplican las técnicas quimiométricas adecuadas (reconocimiento de patrones o pautas) para desarrollar los modelos multivariantes. Recientemente también se está utilizando la denominación de estas técnicas como métodos o herramientas de minería de datos ("data mining").

Los métodos de reconocimiento de pautas se dividen a su vez en supervisados y no supervisados. En el primero, el modelo se construye conociendo la clase a la que pertenece cada uno de los objetos o muestras, mientras que en la segunda el modelo no considera las clases y el reconocimiento de pautas permitirá explorar la presencia de agrupaciones naturales entre los objetos. A las técnicas quimiométricas de reconocimiento de pautas se les puede asignar niveles: (i) análisis exploratorio de datos, consisten principalmente en técnicas no supervisada de análisis de componentes principales (PCA) [**43**]; (ii) agrupamiento ("clustering"), es una técnica descriptiva, también no supervisada, que conforma agrupaciones por similitud entre un conjunto de objetos; (iii) análisis discriminante, son técnicas supervisadas donde el modelo de clasificación se genera estableciendo las fronteras de separación entre los objetos de cada clase de forma que cada objeto siempre va a pertenecer a una clase u otra y, por último, (iv) modelado de clases, también una técnica supervisada, que consiste en definir recintos cerrados en los que se incluyen los objetos de una misma clase. Estos recintos pueden ser independientes o solapar entre sí. Y por lo tanto puede darse el caso

[42] Barnes, R.J., Dhanoa, M.S., Lister, S.J. (1989). Standard normal variate transformation and De-trending of near-infrared diffue reflectante spectra. Applied Spectroscopy, *43(5)*, 772-777.

[43] Bro, R., Smilde, A.K. (2014). Principal component analysis. Analytical Methods, *6*, 2812-2831.

de que las regiones entre clases coincidan y haya algún objeto que puede ser clasificado en varias clases o en ninguna.

Además de las técnicas de reconocimiento de pautas, también hay técnicas de calibración multivariante que tratan de enmarcar los modelos dentro de métodos matemáticos de regresión lineal y no lineal, de manera que se usan métodos cuantitativos de predicción de propiedades específicas.

El **desarrollo de un modelo de clasificación** consta de dos etapas:

1. Entrenamiento del modelo. En esta etapa se seleccionan una cantidad de objetos patrones pertenecientes a una o varias clases conocidas. La selección de patrones se puede realizar de forma aleatoria o pueden aplicarse algoritmos de selección como el algoritmo de máxima disimilitud de Kennard-Stone [**44**], que permite seleccionar aquellas muestras que cubren todo el conjunto de muestras de manera uniforme, de forma que las muestras seleccionadas estén lo más "lejos" posible unas de otras.

2. Validación del modelo. En esta etapa se evalúa el poder clasificador del modelo, pudiéndose aplicar dos estrategias para ello:

    2.1. Validación interna cruzada. Se lleva a cabo sobre el conjunto de datos de entrenamiento y se aplica para diferentes propósitos: (i) Cuando el número de muestras disponible es relativamente pequeño y no se puede dividir en entrenamiento y validación externa; (ii) para evaluar el poder clasificador del modelo para los datos usados en entrenamiento y (iii) para seleccionar el número de componentes necesarios para establecer el modelo de clasificación. Existen diferentes tipos de validación interna cruzada, pero los dos más habituales son los denominados: "leave one out" y "venetian blinds". En todos ellos los datos de entrenamiento son divididos en segmentos (subcojuntos) y el modelo se construye utilizando algunos segmentos y dejando otros para validar el modelo, la forma de elección de dichos segmentos es en lo que difieren los diferentes métodos de validación interna cruzada.

    2.1. Validación externa. En esta etapa se usan patrones alternativos distintos a los usados en la etapa de entrenamiento y el modelo se aplica sobre estos patrones. Esta etapa a veces se ejecuta de forma errónea y es usual encontrar en bibliografía modelos de clasificación que han sido validados con los mismos objetos patrones que han sido utilizados en la etapa inicial de entrenamiento o de validación cruzada, con lo cual la fiabilidad del modelo no está asegurada.

---

[44] Kennard, R.W., Stone, L.A. (1969). Computer aided design of experiments. Technometrics, *11*, 137–148.

Para **evaluar la calidad de las clasificaciones** realizadas por el modelo se pueden calcular diferentes parámetros, aunque hay 4 parámetros fundamentales que se deberían expresar siempre que son: la **sensibilidad, la especificidad, el valor predictivo positivo y el valor predictivo negativo** y no únicamente los dos primeros como es habitual encontrarse en bibliografía. Y ya a partir de estos cuatro pueden calcularse el resto de parámetros como: relación de falsos positivos y relación de falsos negativos, índice de Youden, relación de verosimilitud para resultados positivos, relación de verosimilitud para resultados negativos, relación de diagnóstico, valor F, poder discriminante, eficiencia, área bajo la curva, coeficiente de Gini, G-media, coeficiente de correlación de Mattews, relaciones de aciertos por suerte, relaciones de errores por suerte, coeficiente Kappa [**45**]. Una vez comprobado o evaluado la calidad del modelo ya se lleva a cabo la aplicación del mismo sobre muestras cuya clase es desconocida y se desea conocer.

La **aplicación tradicional** de un modelo de clasificación ha sido llevar a cabo una clasificación binaria, donde el modelo es entrenado con **dos clases de entrada (2iC, 'two input-class')** utilizando patrones del objeto de interés y patrones alternativos y es validado, como se ha comentado anteriormente, con patrones del objeto de interés y patrones alternativos distintos a los usados en la etapa de entrenamiento. De acuerdo con esto tendríamos dos clases de entrada, la clase A como clase objetivo y la clase B como clase alternativa. Idealmente esto se traduciría, si se aplican métodos de análisis discriminante, en el que el modelo devolvería los objetos clasificados en una de las dos clases. Sin embargo, la situación real que ocurre habitualmente en métodos de modelado de clase es distinta. Como clases de entrada tenemos también la clase A y B, pero como clases de salida del modelo podemos tener 4 opciones: que nuestro objeto pertenezca a la clase A (clase objetivo), a la clase B, a ambas clases o que no se clasifique en ninguna clase.

También es posible llevar a cabo el mismo método de clasificación entrenando el modelo con una **sola clase de entrada (1iC, 'one input-class')**: la clase objetivo [**46**]. Por ejemplo, en problemas de autentificación de alimentos esta metodología presenta importante ventajas ya que no es necesario tener un amplio banco de muestras diferentes alimentos, sino que el modelo puede ser entrenado solo con objetos representativos del alimento de interés. La aplicación de esta metodología sólo es posible en métodos de modelado de clases ya que como se ha comentado anteriormente estos modelos se generan de forma independiente para cada clase de objetos. En análisis discriminante esto no es posible porque el modelo se genera delimitando fronteras entre

[45] Cuadros-Rodríguez, L., Pérez-Castaño, E., Ruiz-Samblás, C. (2016). Quality metrics in multivariate classification methods for qualitative analysis. Trends in Analytical Chemistry, *80*, 612–624.

[46] Rodionova, O.Y, Oliveri, P., Pomerantsev, A.L. (2016). Rigorous and compliant approaches to one-class classification. Chemometrics and Intelligent Laboratory Systems, *159*, 89–96.

las regiones características de cada clase, por lo que como mínimo hay que definir dos clases de entrada.

No obstante, en esta **tesis se ha propuesto** y se ha aplicado **una nueva metodología** de **desarrollo de modelos** utilizando métodos de análisis discriminante cuando sólo se tienen muestras de la clase objetivo o genuina. Para ello se define una clase fantasma ('dummy'), y el modelo es entrenado con la clase objetivo, la clase A y con la clase fantasma. Esta metodología se ha denominado como de **pseudo dos clases de entrada (*p*2iC, '*pseudo* two input-class')**. El motivo de aplicar y/o usar esta estrategia es poder aplicar métodos de análisis discriminante ya que son más sensibles y selectivos, cuando sólo se tiene la clase genuina. De esta forma se reduce entre otras cosas el tiempo y costo del análisis.

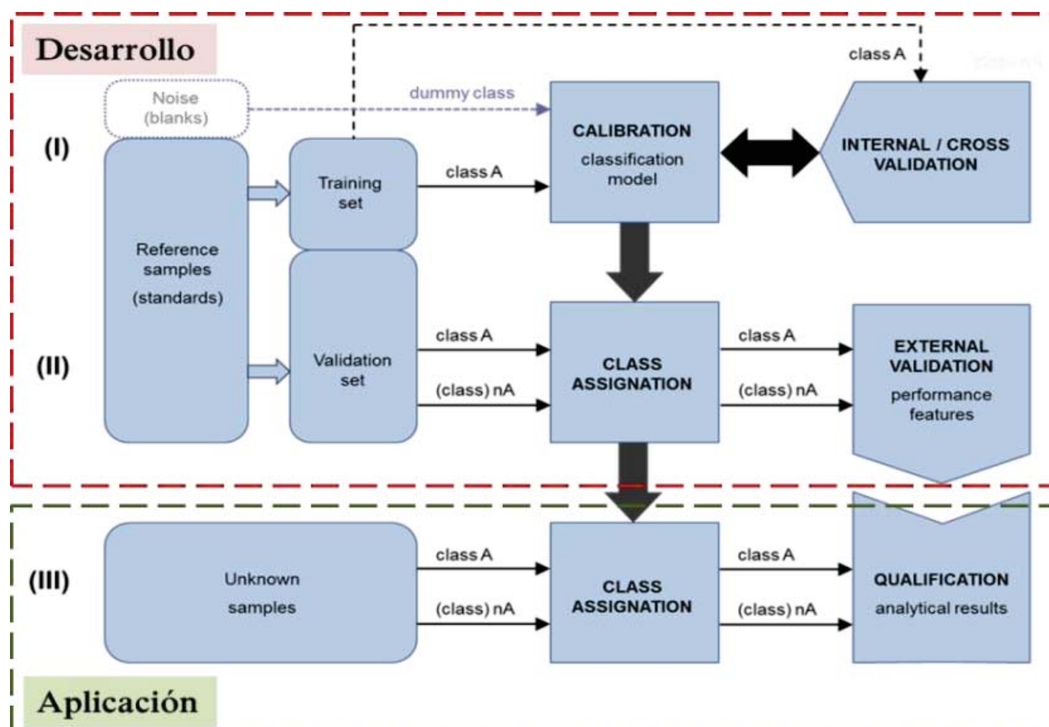


Figura 5. Esquema de desarrollo de un modelo de clasificación de "***pseudo* dos clases de entrada**" (*p*2iC) [**45**].

En esta Tesis Doctoral se aplica la metodología de las huellas dactilares, en conjunción con herramientas quimiometricas para autentificar el aceite de oliva utilizando para ello la fracción metil-transesterificada del mismo.

Además se han realizado dos revisiones bibliográficas relacionadas con lo anteriormente expuesto:

- Una revisión de la aplicación de las huellas dactilares cromatograficas aplicando la cromatografía de líquidos de alta eficiencia para la autentificación de aceite, así

como los diferentes modelos multivariantes aplicados con este fin pueden ser encontradas en el capítulo de libro publicado que se presenta a continuación.

- Una revisión bibliográfica sobre la aplicación de herramientas de minería de datos aplicadas en el campo de la alimentación.

Los correspondientes artículos científicos se presentan a continuación.

CAPÍTULO DE LIBRO

LC Fingerprinting approach for authenticating olive oil

(Capítulo de libro aceptado para su publicación el 09 de Septiembre 2018 por la editorial Nova Publisher en el libro: Authentication and Adulteration of Olive Oil)

Chapter

LC FINGERPRINTING APPROACH FOR AUTHENTICATING OLIVE OIL

Ana M. Jiménez-Carvelo✉, Antonio González-Casado, Estefanía Pérez-Castaño, Luis Cuadros-Rodríguez

Department of Analytical Chemistry, University of Granada,
Granada, Spain

Chapter

LC Fingerprinting Approach for Authenticating Olive Oil

Ana M. Jiménez-Carvelo**[*]*, ***Antonio González-Casado, Estefanía Pérez-Castaño and Luis Cuadros-Rodríguez***

Department of Analytical Chemistry, University of Granada, Granada, Spain

Abstract

This chapter aims to explain different approaches to authenticate food, followed by a brief clarification and differentiation of both fingerprinting and profiling terms. The liquid chromatographic (LC) fingerprinting describes a variety of analytical methods providing analytical signals related to the composition of foodstuffs in a non-selective way. The fingerprinting requires using chemometric tools to extract the most relevant and non-evident information from the chromatograms so that multivariate classification or regression models suitable to authenticate food can be developed. An overview of the applications of LC fingerprints in food using different chemometric tools has been included. The LC fingerprinting has been scarcely used in order to authenticate olive oil. However, the articles published up to now show that the LC fingerprinting is of great versatility comprising a powerful and efficient technique for authentication purposes which can be implemented in quality control laboratory so as to compete favourably with the spectroscopic techniques (i.e., NIR or NMR) hitherto used.

Keywords: olive oil, food authentication, liquid chromatography, fingerprinting, chemometrics

* Corresponding Author address: Phone: +34 958240797; fax: +34 958243328. Email: amariajc@ugr.es.

Abbreviations and acronyms

1iC, one input-class classification

2iC, two input-class classification

ACR, absolute centred residual

ANN, artificial neural network

biPLS-R, backward-interval partial least squares regression

CA, cluster analysis

CART, classification and regression trees

CASE, Computer-aided similarity evaluation system

CVA, canonical variate analysis

CAD, charged aerosol detector

DA, discriminant analysis

ECD, electrochemical detector

FA, factor analysis

FLD, fluorescence emission detector

FT-IR, Fourier transform-infrared spectrometry

FT-Raman, Fourier transform-Raman spectrometry

GA, genetic algorithm

HCA, hierarchical cluster analysis

HC-DA, hierarchical cluster-discriminant analysis

i-PLS, interval- partial least squares

IT, ionic trap mass analyser

kNN, k-nearest neighbours

LC, liquid chromatography

LDA, linear discriminant analysis

MIR, mid-infrared spectrometry

MLR, multiple linear regression

MS, mass spectrometry

NIR, near-infrared spectrometry

NMR, nuclear magnetic resonance spectrometry

OCPLS, one class partial least squares classification
OPLS-DA, orthogonal partial least squares-discriminant analysis
p2iC, pseudo two input-class classification
PMR, partial robust M-regression
PCA, principal component analysis
PCR, principal component regression
PLS-DA, partial least squares-discriminant analysis
PLS-R, partial least squares regression
QDA, quadratic discriminant analysis
Q-TOF, quadrupole-time of flight mass analyser
R2, determination coefficient
RBF, radial basis function
RDA, regularized discriminant analysis
RF-C, random forest classification
RF-R, random forest regression
SD, score distance
SESCF, similarity evaluation system for chromatographic fingerprinting
SIMCA, soft independent modelling of class analogy
SVM-C, support vector machine classification
SVM-R, support vector machine regression
UNEQ, unequal class modelling
UV-DAD, ultraviolet molecular absorption-diode array detector
UVD, ultraviolet molecular absorption detector
VIP, variable importance in projection.

Introduction

Food authentication involves the confirmation of the stated food specifications as true. This is a substantial requirement as authentication is bound to truthfulness, and food is considered authentic when it does not involve any fraud. There are different approaches to chemically authenticate the genuineness of a food and to detect fraud which are based on obtaining information on: (i) chemical composition, (ii) biomolecular markers and (iii) stable isotope ratios. The first approach is divided into three sub-types mainly depending on the scientific-technical basis applied and the type of analytical information used. These sub-types include: (i) chemical markers; (ii) component (compositional) profiles; and (iii) instrumental fingerprints.

The chemical markers are based on the determination of a number of chemical components characteristic of the authentic food ('positive' approach) or the adulterated food ('negative' approach), which are of interest for quality control purposes. The overall quality of food may be affected by many factors, including harvesting time, cultivation site, post-harvesting processing, adulterants or substitutes of raw materials, and procedures in extraction and preparation. For example, most of the reported studies on olive oil authentication are focused on applying the chemical markers strategy to distinguish olive oil from different commercial categories of olive oil [**1**] as well as to determine the optimal ripening stage of olive fruits [**2**]. The component profiling strategy is generally based on the quantification of determined compound groups which constitute a characteristic fraction of the material of the interest, and which can be expressed as contents (concentration) or composition (percentage). The instrumental fingerprints refer to the common peaks obtained through spectroscopic and chromatographic analysis (see Figure 1).

The fingerprinting methodology allows obtaining all the necessary information to characterise or authenticate a food [**3**]. This methodology is based on considering the entire signal achieved through an analytical technique (chromatography or spectroscopy) without having to identify or quantify individual compounds. The chromatograms or spectra are used as a whole analytical signal composed of complex bands, and the areas/heights of the individual peaks are not considered.

The terms fingerprinting and profiling are sometimes mistaken. Both terms come from the metabolomic terminology. Metabolomic strategies have been divided into two distinct approaches, *untargeted* (fingerprinting) and *targeted (profiling)* [**4, 5, 6**].

Fingerprinting (untargeted metabolomics) is the comprehensive analysis of all the measurable analytes in a sample, including unknown chemicals. Due to its comprehensive nature, it is necessary to use such advanced chemometric tools, as multivariate analysis to extract, reduce and process the extensive datasets. By contrast, profiling is the measurement of related metabolites in a group. In short, the difference between both terms arises from the type the analytical information involved.

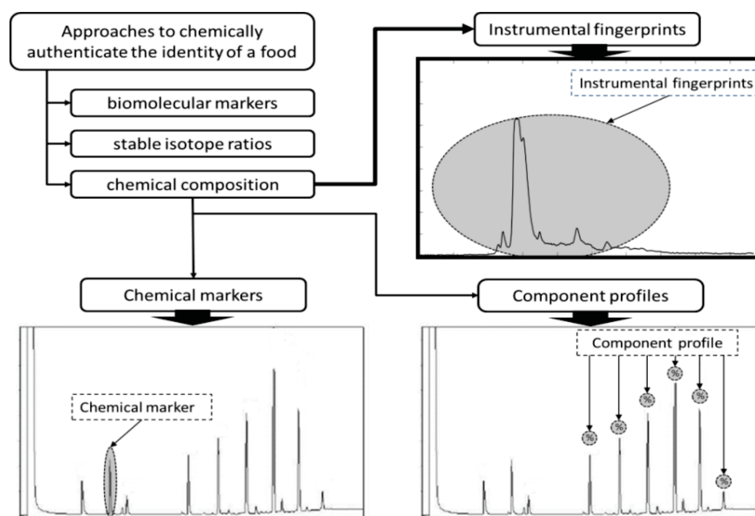


Figure 1. Approaches to determine the authenticity of a given food.

The fingerprinting methodology has become one of the most efficient and comprehensive methods to verify food identity [**7, 8, 9**]. The application of this methodological approach requires having an adequate number of samples which previously have been stated as authentic or genuine food in order to establish a representative data base of the genuine food population. Traditionally, vibrational spectrometry techniques as NIR, MIR, FT-IR and FT-Raman have been used for the determination of food authenticity, since these require no sample pre-treatment and only a few minutes for the analysis of each sample [**10, 11, 12, 13, 14, 15, 16, 17**]. The most recent advances apply NMR fingerprint for the characterisation of nutritional content, such as for the classification of red wines based on 1H and 13C NMR fingerprints [**18**], phospholipid fingerprints (31P NMR) to distinguish different types of milk [**19**] or classification of olive oils according to geographical origin [**20**], and virgin olive oil authentication [**21**]. However, these techniques possess low sensitivity. Recently, the different applications of NMR fingerprints to resolve authenticity problems in food have been reviewed [**22**]. Among the different analytical methods, the technology of LC fingerprinting has been widely used in many studies. It has

significantly contributed to ensure the authenticity and traceability of high-value foodstuffs since it has the features of high resolution, rapid analysis, high sensitivity, good stability, good reproducibility, with a wide choice of mobile phases and a wide variety of detectors. Research studies can be found in the literature applying LC fingerprinting to food science, i.e., to identify the botanical origin of honey based on the chromatographic fingerprint of the phenolic fraction [**23**], or to discriminate and classify red wines according to grape cultivar using LC coupled to quadrupole time-of-flight (Q-TOF) mass spectrometry (MS) [**24**]. Recently, various applications of chromatographic fingerprints to authenticate food have been reviewed [**25**].

Nevertheless, the use of chromatographic fingerprinting implies a methodological change in the way these techniques are used. Therefore, contrary to traditional chromatography, where the aim is to achieve a full chromatogram with well resolved peaks, here the main priority is to acquire a chromatogram in a short time interval (ideally, less than 5 min). Such a chromatogram should contain enough characteristic information to enable the verification or the authentication of the food under study. So as to compete with spectroscopic techniques it is important to obtain a chromatogram within such a short time interval.

Chemometrics

Chemometrics was introduced in the 1970s, and it is recently defined as the science of relating measurements made on a chemical system or process to the state of the system via application of mathematical or statistical methods, mostly of multivariate nature [**26**]. It is an interdisciplinary science that uses mathematical, statistical and logical methods to provide the largest information of a group of complex analytical signals. In this sense, the existence of interferences and the pre-treatment steps of samples do not pose a problem as chemometric methods can solve such problems [**27**].

Different methods of multivariate analysis have been used to extract the relevant information from chromatographic fingerprints. Consequently, several areas of the chemometric tools can be distinguished:

- Similarity Analysis

The similarity between the different chromatographic fingerprints is measured by similarity indices. Such indices are defined as a quantity which describes the

equivalence of two objects characterised by multivariate data [26]. The use of similarity indices implies the decision to investigate how near two samples are to each other. Similarity analysis uses methods based on (i) distance, which employs several ways of measuring the distances (Euclidian, Manhattan, Chebyshev, Mahalanovis, etc.); and (ii) correlation. Pearson's correlation coefficient and determination coefficient are the most used.

Similarity evaluation of chromatographic fingerprints has been used to identify raspberry [**28**], for the authentication of honey from floral origin [**29**] and for the identification of Chinese tea from different plantations [**30**].

➭ Pattern Recognition

Different authors have defined this area [**31**]. The most recent and simple definition establishes that pattern recognition is the assignment of a label to an object characterised by data [**26**]. Pattern recognition is also an integral part in most machine intelligence systems built for decision making. Basically, the pattern recognition methodology attempts to find similarity between several groups of objects (qualitative approach).

The chemometric techniques under this field can be classified in different groups. Table 1 provides an overview of the major chemometric tools in food authentication.

Traditionally, these have been categorised as supervised and non-supervised methods, but more recently these are grouped as linear and non-linear classifiers, which are assigned levels.

Table 1. The most used chemometric pattern recognition methods.

	Level	Linear	Non-linear
Exploratory (non- supervised)	0	PCA, FA	
	1	CA	
Classification (supervised)	2	PLS-DA, LDA, OPLS-DA	kNN, SVM-C, ANN, CART, RF-C
	3	SIMCA, UNEQ	OCPLS

PCA: principal component analysis; FA: factor analysis; CA: cluster analysis; PLS-DA: partial least squares-discriminant analysis; OPLS-DA: orthogonal partial least squares-discriminant analysis; SIMCA: soft independent modelling of class analogy; UNEQ: unequal class modelling; kNN: k-nearest neighbour; SVM-C: support vector machine-classification; ANN: artificial neural network; CART: classification and regression tree; RF-C: random forest classification; OCPLS: one class- partial least squares.

- Level 0: Exploratory Analysis

It is based on proof if there is a natural grouping of a data set. The chemometric tool mostly employed is the principal component analysis (PCA), which is a particular method related to the factor analysis (FA). The principal components (PC) are calculated by successively capturing the greatest variance in the data set as a lineal combination between the original variables (chromatographic or spectroscopic intensities) of each object, which are described as:

$$X = T \times p^T + e$$

where X is the original data matrix, T is the score matrix, p^T is the transposed loading matrix and e is the residual or error matrix [**26, 32**].

Each PC comprises an axis in the data space, representing an underlying dimension that contributes to summarising or accounting for the original data set. PCA reduces the number of variables to evaluate which components contain essential information.

- Level 1: Cluster Analysis

It is a technique grouping similarity in accordance with a set of objects and generating groups of objects that are not similar to each other. It is a particular application of the similarity analysis.

The cluster analysis (CA) methods are divided into: (i) hierarchical methods, by which the number of clusters is not specified since their choice is a posteriori; and (ii) non-hierarchical methods, by which the number of conglomerates is previously defined [**33, 34**]. The result of performing cluster analysis is represented as a hierarchy tree named dendogram.

- Level 2: Classification

This level can be divided into purely statistical methods and machine learning (data mining) [**35**]. Although it is present in both sub-areas, the classification methods are generated through establishing the boundaries for the different categories defined by the training objects. The classical methods involve methods considered to be primarily statistics as discriminant analysis. The distinct classes of the objects are defined on the basis of several criteria. One of the most commonly used criteria is the so-called Baye's rule, which states that a sample should be assigned to the class to which it has the maximum probability to belong [**36**].

Classical methods, which involve methods that are considered to be primarily statistical such as the linear discriminant analysis (LDA) and sub-cases as quadratic discriminant analysis (QDA), canonical variate analysis (CVA), and regularised discriminant function (RDA) belong to this level. The same holds for partial least squares discriminant analysis (PLS-DA) or orthogonal partial least squares discriminant analysis (OPLS-DA). The latter tools (PLS-DA and OPLS-DA) are the most widely used for data treatment from fingerprints. PLS-DA involves building a PLS regression model to establish class limits and, then, performs a discriminant analysis (DA) to achieve classification. PLS-DA can be regarded as a linear two-class classifier. PLS-DA is represented as:

$$Xj = T \times p + e; \qquad Y_l = T \times q + f$$

where Xj is the matrix of the set of samples of j analytical measurement (spectrum or chromatogram), Y_l represents a numerical label for each sample according to its class, T is the score matrix, p and q are the loading matrices and e and f can be considered residuals [**37**].

One of the most widely-used supervised classification methods is the so-called k-nearest neighbour (kNN), which is a non-parametric method for objects based on the closest training examples in variable space [**26**].

At this level other more powerful chemometric techniques belonging to data mining can be found. Data mining is the process of extraction of implicit information from ‘big data,’ previously unknown and potentially useful from data. There is a wide variety of data mining algorithms. Traditionally, data mining methods have been used in the field of business, genetics and engineering. However, these methods are being recently employed in the food science field. Data mining methods can be extremely useful for solving problems related to food authentication. Research studies can be found in the literature where several data mining methods have been applied to a) detect adulteration of olive oil in blends with other vegetable edible oils [**38**], b) to detect adulteration of olive oil with camellia oil [**39**], and to c) classify the rice or hazelnut oils according to their geographical origin [**40, 41**]. In the food science field, the most representative applications are: (i) statistical models (e.g., support vector machines in classification (SVM-C)), (ii) neural network-based models (i.e., artificial neural network (ANN)), and (iii) decision trees and rules (i.e., classification and regression trees (CART), and random forests in classification (RF-C)).

SVM is based on building a hyperplane to discriminate the different classes of objects. The decision boundaries (hyperplanes) are determined, which maximises the separation of data in different classes [**26**]. ‘Support vectors’ are the points that form two lines parallel to the hyperplane, the distance between the points being the largest possible. The SVM classification is very efficient and is an extremely robust tool. Using the Kernel equations makes it possible to reduce the dimension of the working space [**42, 43**]. Although SVM is known as a non-linear classifier, it can also be used as a linear classifier. This is its main advantage with respect to PLS-DA.

ANN is computer network system of simple elements, interconnected in parallel and hierarchically organised, which process information via their dynamic state response to external inputs [**26**]. They attempt to interact with objects in the real world in the same way as the biological nervous system. ANN is based on a set of ‘nodes’ or ‘artificial neurons’ interconnected with each other in the form of a network [**44**].

Decision trees consist of nodes that form a ‘rooted tree.’ This means that it is a ‘directed tree’ with a node (root) lacking incoming edges. All other nodes have exactly one incoming edge. The classification of a sample begins in the node called ‘root’ of the tree. The instance is then evaluated at one node and it follows the branch appropriate to its outcome [**45**].

- Level 3: Class Modelling

It involves building a classification model in which each class in the training set is perceived independently. It defines a domain in space for each class of the model, so that it can be possible that the regions overlap and the new objects are classified in several classes. This area includes: (i) unequal class modelling (UNEQ), where each group is modelled with a multivariate normal distribution [46], and each new sample is assigned based on the Euclidean distance to each class; and (ii) soft independent modelling of class analogy (SIMCA), which performs a PCA on each class. Following this, an unknown is assigned to the particular class with which it has the lowest residual variance [26].

OCPLS (one class partial least squares classification) is a new chemometric tool developed by Xu et al. 2006 which is based on a special PLS regression. The objects are classified according to the OCPLS score distance (SD) and absolute centred residual (ACR) of predicted response. There are three variants of the function: (i) conventional ordinary linear OCPLS, (ii) nonlinear radial basis function (RBF) OCPLS, and (iii) partial robust M-regression (PMR) OCPLS [**47, 48**].

➷ Multivariate Calibration (Regression)

Multivariate regression establishes functional relationships between an analytical signal that provides a sample or set of samples, each of one including a property of these. Its matrix form is:

$$X \times b = Y + e$$

where X is a vector of the experimental data (analytical signal), b is the vector of the model coefficients, Y is the property to be predicted, and e is the vector of errors or residuals. Note that this expression is a reversal of the traditional form, x (concentration)/y (indication) of univariate linear regression [**26**].

Multivariate calibration is the application of multivariate regression in order to predict the value of the y-property of any sample by using the regression model. In these methods, it is assumed that there is a number of mixtures for which the amounts of each component are known and for which a number of properties has been measured. The chemometric tools regularly used in the food authentication field are partial least squares regression (PLS-R) [**49**], principal component regression (PCR) and, to a much

lesser extent, multiple linear regression (MLR) [**50**] which are linear methods. All the other aforementioned methods are lineal multivariate calibration methods, but some non-linear methods such as SVM regression (SVM-R), ANN CART, or RF regression (RF-R) can also be applied.

↳ Data Fusion

The use of two or more independent data sets in order to apply multivariate classification and/or calibration by collecting more and better relevant information deserve a special comment in this chapter. This was earlier stated by Forina et al. 2009 [**51**] "Better results are obtained by the synergism from different instruments, … the fusion, that is, the use of predictors from different instrumental sources or with different treatments, is one of the recent developments of chemometrics in Food Chemistry. Fusion is the consequence of laboratories that work on food problems and have several instruments, but many providing nonspecific information, such as fingerprints."

Data fusion is the process of synthesising data from different analytical techniques to obtain meaningful information that can be of greater value than single technique data. Various strategies can be used to fuse data: low-, mid- and high-level strategies [**52**]. In low-level data fusion, raw data from more than one source are directly concatenated. In mid-level data fusion, only a number of variables are fused. Previously, a reduction of variables is performed in order to select the variables. Several tools can be used to reduce the variables, for example interval-partial least squares (i-PLS) or variable importance in projection (VIP). In high-level data fusion, the classification results obtained from individual classification models are fused.

Currently, there are only a few studies in the literature which use data fusion for the authentication of olive oil [**53, 54**].

LC Fingerprinting

Sources of LC Fingerprints

In this section, an overview of the published studies on the applications of the fingerprinting methodology for food authentication purposes is presented. LC fingerprints are used for four mainly purposes:

- Similarity analysis, to evaluate the botanical or geographical origin of plant products. Similarity analysis is applied to calculate specific correlation parameters (correlation (r) and congruence (c) coefficients), and distance measurements. It is noteworthy the dissimilarity comparison of chromatographic fingerprints, which is a relative measure calculated between two objects that have the same nature or characteristics.

- Exploratory and clustering analysis, to discriminate genotypes or species. It differentiates geographical origin of different foods (i.e., rice, wine and sugarcane).

- Classification of foods regarding diverse criteria (i.e., authentication of geographical origin). In most of the studies carried out, classification is achieved using only the chromatographic fingerprint from a family compound fraction.

- Multivariate calibration, to predict the value of some properties, such as oxidative capacity; and, in some cases, to quantify the concentration of some food constituents in blends.

The number of published applications of LC fingerprinting is limited. Only 31 articles have been found using proper LC fingerprints, of which only 6 articles applied the methodology in order to authenticate olive oil. Table 2 summarises the main applications. The conventional applications, which are the most numerous, involve the use of classical chemometric tools as exploratory/clustering and classification (discrimination and class modelling) methods. Among the publications related to this issue, two studies should be higlighted because of their current interest. Perez-Castaño et al. 2015 applied two multivariate classification methods (PLS-DA and SIMCA) to differentiate the geographical origin of palm oil using the sterolic LC fingerprinting provided by two different detectors (ultraviolet and charged aerosol) coupled to LC

[55]. The results were compared by several proposed classification metrics in order to decide on the best method. Obisesan et al. 2017 [56] fused the fingerprints acquired in the former study to discriminate the geographical origin of edible palm oil. However, the main interest of this paper is to show how without a 'perfect' (well resolved) chromatogram but by applying the right chemometric tools it is possible to resolve a problem (in this case the authentication of geographical origin of palm oil).

Table 2. Reported applications of LC fingerprinting in food authetication

Detector	Food	Chemical	Purpose	Multivariate analysis	Ref.
UVD	Tofu-type soybean	Isoflavones (from methanolic extract)	Differentiation of Tofu-type soybeans	Similarity analysis: CASE	[57]
	Green tea	Aqueous extract	Evaluation of (dis)similarity	Similarity analysis: CASE; Heat maps; Euclidean and Mahalanobis distances	[58]
	Red wine	Phenols (from sample diluted with water)	Discrimination of Spanish wines from PDO, grape varieties, ageing periods and vintage	Exploratory analysis: PCA	[59]
	Edible vegetable oils	Sterols (from unsaponifiable fraction)	Discrimination of the olive oils from edible vegetable oils	Clustering: HCA Exploratory analysis: PCA Classification: PLS-DA	[60]
	Pomegranate peel	Phenols (from ethanolic extract)	Identitation of Chinese pomegranate from orchards	Similarity analysis: SESCF	[61]
	Palm oil	Methyl transesterified fraction	Differentiation of geographical origin	Exploratory analysis: PCA Classification: PLS-DA, SIMCA	[55]
	Beer	Polyphenols (from sample diluted with water)	Discriminate among ale and lager beers	Exploratory analysis: PCA Classification: PLS-DA, SIMCA	[62]

Table 2. *Continue*.

Detector	Food	Chemical	Purpose	Multivariate analysis	Ref.
UVD	Palm oil	Methyl transesterified fraction	Differentiation of geographical origin	Data fusion Classification: PLS-DA	[56]
UV-DAD	Extra virgin olive oil	Phenols (from hydroalcoholic extract)	Discrimination of the PDO Sabina olive oil	Classification: PLS-DA Data fusion Calibration: biPLS-GA	[53]
	Honeybush tea	Aqueous extract	Differentiation between wild-harvested and cultivated seedling plants	Exploratory analysis: PCA	[63]
			Evaluation of similarity	Similarity analysis: R2	
	Sugarcane	Flavonoids (from hydroalcoholic extract)	Discrimination of genotypes	Exploratory analysis: PCA	**[64]**
	Honey	Flavonoids (from methanolic extract)	Authentication and classification of floral origin	Similarity analysis: SESCF Exploratory analysis: PCA Classification: PLS-DA, SIMCA Calibration: PLS	[29]
	Skim milk powder	Foreign proteins (from free fat extract)	Detection of Adulteration	Exploratory analysis: PCA Classification: SIMCA	[65]
	Honey	Aqueous solution	Verification of the honey floral origin	Exploratory analysis: PCA Classification: kNN	[66]

Table 2. *Continue.*

Detector	Food	Chemical	Purpose	Multivariate analysis	Ref.
UV-DAD	Green tea	Phenols (from ethanolic extract)	Identitation of Chinese Ziyang green tea from different plantations	Similarity analysis: SESCF	[**30**]
	Tea	Metabolites ((from methanolic extract)	Cultivar origin identification	Similarity analysis: SESCF	[**67**]
	Corn	Peptides (from basic hydroalcoholic extract)	Distinguish the origins and varieties.	Similarity analysis: SESCF	[**68**]
	Fruit extract	Polyphenols (from acetone extract)	Authentication and classification of fruit type	Exploratory analysis: PCA Calibration: PLS	[**69**]
	Extra-virgin olive oil	Phenolic compounds compounds (from methanolic extract)	Verification of varietal origin	Data fusion Exploratory analysis: PCA Classification: kNN, PLS-DA, SIMCA	[**54**]
FLD	Red wine	Phenols (from sample diluted with water)	Discrimination of PDO Spanish wines, grape varieties, ageing periods and vintage	Exploratory analysis: PCA	[**59**]
	Extra-virgin olive oil	Phenolic compounds (from methanolic extract)	Verification of varietal origin	Data fusion Exploratory analysis: PCA Classification: kNN, PLS-DA, SIMCA	[**54**]

Table 2. *Continue.*

Detector	Food	Chemical	Purpose	Multivariate analysis	Ref.
ECD	Unifloral honeys	Phenolic (from methanolic extract)	Floral origin identification	Similarity analysis: SESCF	[**70**]
	Monofloral honeys	Phenolic acids (from methanolic extrac)	Floral origin identification	Similarity analysis: SESCF	[**71**]
CAD	Palm oil	Methyl transesterifed fraction	Differentiation of geographical origin	Exploratory analysis: PCA Classification: PLS-DA, SIMCA	[**55**]
	Olive oil from other edible vegetable oils	Methyl transesterified fraction	Discrimination of the olive oils from edible vegetable oils	Exploratory analysis: PCA Classification: kNN, PLS-DA, SVM-C, SIMCA	[**72**]
	Olive oil from other edible vegetable oils	Methyl transesterified fraction	Discrimination of the olive oils from edible vegetable oils	Exploratory analysis: PCA Classification: kNN, PLS-DA, SVM-C, SIMCA	[**73**]
	Olive and palm oil from other edible vegetable oils	Methyl transesterified fraction	Quantification of olive and palm oil in blends with other vegetable oils	Classification: SVM-C Calibration: PLS-R, SVR	[**74**]
	Palm oil	Methyl transesterified fraction	Differentiation of geographical origin	Data fusion Classification: PLS-DA	[**56**]
(IT)MS	Kiwi and pomelo	Catechins, phenols and flavonoids (from hydroalcoholic extract)	Classification according to the fruit species and subspecies	Exploratory analysis: PCA Clustering: HCA Classification: LDA	[**75**]

Table 2. *Continue.*

Detector	Food	Chemical	Purpose	Multivariate analysis	Ref.
(Q-TOF)MS	Red wine	Non-volatiles	Discrimination between grape varieties	Exploratory analysis: PCA	[**76**]
	Saffron	Metabolites (from basic ethanol extract)	Authenticity of saffron	Exploratory analysis: PCA Classification: PLS-DA, OPLS-DA	[**77**]
	Saffron	Metabolites (from hydroalcoholic extract)	Authentication of geographical origin	Exploratory analysis: PCA Classification: OPLS-DA	[**78**]
(Orbitrap)MS	Bread	Metabolites (from hydroalcoholic extract)	Differentiation of bread made with whole grain and refined wheat	Exploratory analysis: PCA Classification: SIMCA	[**79**]
	Rice	Metabolites (from methanolic extract)	Differentiation of geographical origin	Clustering: HCA Classification: HC-DA	[**80**]

Application of the LC Fingerprinting Methodology for the Authentication of Olive Oil

Olive oil is a product highly appreciated by consumers for its sensory properties as well as for its health promoting properties. In the food industry is common to incorporate olive oil as a valuable ingredient into numerous foodstuffs or to blend it with other edible oils and fats, with the aim of improving quality of such products or to lower costs and to have more favourable economic returns from the final product. Olive oil is mostly composed of triacylglycerols (or triglycerides), found in the saponifiable fraction and the unsaponifiable fraction which contains minor compounds such as tocopherols, carotenes, polyphenols, sterols, etc.

Most analytical methods used for the authentication of olive oil are based on the determination of compound-specific profiles of the major and minor fractions.

Composition of triacylglycerols, fatty acids, sterols and phenols to detect adulteration of olive oil with other vegetable oils has been determined by mainly using chromatographic methods. However, the scope of these methods is limited and not all of them are able to differentiate an olive oil from another vegetable oil. For example, some vegetable oils present a compound profile similar to that of olive oil. For this reason, it is necessary to apply chromatographic analysis with a different approach to the traditional one.

As described in previous sections, fingerprinting methodology in based on the acquisition of a non-specific analytical signal. For this compound, identification and quantitation steps are avoided. An important thing in this approach is to obtain a separation as fast as possible, though this would lead to a loss of chromatographic resolution in which the characteristic (and no-evident) information of the vegetable oil (in our case) is achieved. Thus, applying chemometric tools it is possible to extract the most relevant information and to differentiate the edible oils.

There are currently only a small number of studies having applied LC fingerprinting methodology to authenticate olive oil. There are two main approaches to apply LC fingerprinting: (i) to use the fingerprints from the intact material without having to change the matrix definition. For olive oil, this means that the oil is simply dissolved in a solvent and then analysed; and (ii) to carry out a previous fractionation step of the material to extract a specific family of compounds and to obtain a specific LC-fingerprint from this fraction. For example, for olive oil it is possible to obtain the fingerprint of the volatile, sterolic or phenolic fraction.

Several examples of these approaches have been published. De la Mata-Espinosa et al. 2011 developed a discrimination method of olive oil from other vegetable oils (canola, corn, flaxseed, grape seed, hazelnut, peanut, rapeseed, safflower, sesame, soybean and sunflower) employing fingerprints provided by HPLC-CAD. The analysis time was of 40 minutes and the PLS-DA model classified all the samples correctly except for the hazelnut oil [**81**]. This led to quantify olive oil in binary blends of other edible vegetable oils using triacyglycerol chromatographic fingerprints with a 10% of root mean square error of prediction (RMSEP) value [**82**]. Also, the LC fingerprinting methodology has been applied to assure the varietal origin of olive oils. Bajoub et al. 2016 used the entire oil chromatogram combined with PCA and PLS-DA to classify olive oil according olive

fruit variety. The olive varieties studied were Arbequina, Cornicabra, Frantoio, Hojiblanca and Picual, with several PLS-DA models being tested. Even with a small number of samples tested, the results showed that the phenolic fingerprint contains adequate information to group olive oil according to variety, mainly to discriminate Arbequina and Frantoio cultivar olive oils [83]. Further to this part of the work, the authors used data from liquid chromatography coupled to two detectors –diode array (DAD) and fluorescence (FLD)– in combination with low- and mid-level data fusion strategies to authenticate the varietal origin of olive oil. In this case, 140 extra-virgin olive oil samples from seven olive varieties were studied (Arbequina, Arbosana, Cornicabra, Frantoio, Picholine de Languedoc, Picholine Marocaine and Picual) and PCA, kNN, SIMCA and PLS-DA chemometric techniques were applied [54]. PLS-DA models combined with "mid level" data fusion provided the best results for the classification of the different olive oil varieties. The root mean square errors of prediction were found within the range 0.09–0.232.

Recently, a new approach for applying LC fingerprinting has been used in our research group. This new methodology is based on a chemical reaction of olive oil to form a derivative product. Traditionally, the most commonly used reaction has been the saponification reaction. Bagur-González et al. 2015 built a discrimination model to differentiate virgin olive oil from other edible oils using the LC fingerprinting of the sterol fraction. In a total of 51 vegetable oils analysed (virgin olive oil, sunflower oil, pomace oil, soybean plus corn oil and canola oil), from which 38 and 13 oils were employed respectively as calibration and evaluation purposes [**60**]. In 1993, Bierdermann et al. proposed a new strategy, which consisted of transesterifying olive oil with methanol [84]. This method offers many advantages in comparison to traditional saponification since it is less time consuming, reduces the volume of solvents needed, and improves the extraction yield of sterols. The cleavage of molecules occurring during transesterification leads to the formation of methyl esters from the fatty acids and the liberation of sterols. Two fractions are obtained within this process; (i) the water soluble fraction which contains the polar compounds, and (ii) the organic fraction (transesterified fraction) in which the methyl esters of the fatty acids, monoacylglycerols, diacylglycerols and other molecules can be found. In addition, methyl sterols, dimethyl sterols and linear alcohols were also identified in this latter fraction [**84**].

There are several studies dealing with the transesterification of olive oil and its further analysis by means of gas chromatography. There are, however, no previously published studies applying liquid chromatography on the transesterified fraction. To the best of our knowledge, there are no studies applying this methodology of oil transesterification to develop new methods for olive oil authentication. By employing this new strategy, our research group has been able to develop fast analytical methods to authenticate olive oil. LC fingerprinting from the methyl-transesterified fraction was acquired in both modalities (reverse phase and normal phase) using HPLC. In the normal-phase mode [**72**] the analysis lasted 8 minutes with 127 vegetable oil samples from different botanic species being analysed (66 samples of olive oil and 61 samples of other edible oils). The chemometric tools applied to build the classification models were kNN, PLS-DA, SVM-C, SIMCA and OCPLS. In nearly all cases, the classification rates from an external validation set were 100%. In the reversed-phase mode [**73**] a 3 cm short-column (usually used as a precolumn) was used to obtain the oil fingerprint from the methyl-transesterified fraction. The analysis time took only 4 min. The method provided a high classification performance: efficiency 0.98, correct classification rate 0.97, and Matthews's correlation coefficients 0.95. Three classification strategies were applied: one input-class classification (1iC), two input-class classification (2iC) and 1iC plus one 'dummy' class classification [**85**]. This latter strategy has been named by our group as 'pseudo' two input-class classification (p2iC) [**73**]. 1iC offers advantages in the food authentication field, because the classification model can be built from only a number of genuine foods (olive oil in our case), and it does not require other edible oils. It is important to highlight that p2iC has been devised, tested and developed by our group in order to allow for discrimination methods (i.e., PLS-DA or SVM-C) when only the target class is used for training. Discriminant methods require using two input-classes (2iC) in order to define the border among the different regions according to the classes of the samples studied. Nevertheless, it is possible to build the discrimination model by training the model with target class plus a fictitious class (i.e., dummy class). In both strategies (1iC and p2iC), the number of analyses to train the model is significantly lower.

Based on the studies referred above, it can be concluded that the methyl-transesterified fraction can be used to quantify olive oil in mixtures. We have developed a global and comprehensive analytical method to detect (classify) and quantify blends of olive oil

with most widely used edible vegetable oils [74]. The oils used in this work originate from different parts of the world. Although in the ‘real world’ the usual blends of olive oil with other seed oils are binary, a quality control laboratory does not know which individual seed oil or a mixture of seed oils were used in the adulteration process. For this reason, the proposed method attempted to cover the most habitual adulterant blends occurring.

Conclusion

Olive oil is a valuable foodstuff recognised by consumers for its nutritional and health promoting properties. This is the reason why it is subject to its adulteration with other cheaper vegetables oils, to faking its quality category or its botanical/geographical origin.

The effective use of fingerprinting methodology based on liquid chromatographic analytical signals for food authentication is still limited, and there are only a few analytical applications reported so far. In addition, the seeking of representative papers is complicated by the fact that scientists confuse the terminology and use indistinctly the terms chromatographic profiling and fingerprint.

The main goal of this chapter is to reveal the enormous potential of the LC-fingerprinting to solve problems related to olive oil authentication. For this reason, in order to be competitive with the most common spectrometric techniques, the chromatographic step should be rapid (less than 5 minutes) and simple (little or no sample preparation). The chromatographic signal is then treated as a whole by proper multivariate data analysis. For this purpose, a range of chemometric methods are available today in order to select and obtain the relevant information from the data. The application of the most appropriate chemometrics method to the same data set can make the difference between good and bad results. In addition, the application of data fusion strategies which remain largely unknown to food analysis may further improve the results.

References

[1] Sales, C., Cervera, M. I., Gil, R., Portolés, T., Pitarch, E., Beltran, J. (2017). Quality classification of Spanish olive oils by untargeted gas chromatography coupled to hybrid quadrupole-time of flight mass spectrometry with atmospheric pressure chemical ionization and metabolomics-based statistical approach. Food Chemistry, 216, 365-373.

[2] Magagna, F., Valverde Som, L., Ruíz Samblás, C., Cuadros Rodríguez, L., Reichenbach, S. E., Bichi, C., Cordero, C. (2016). Combined untargeted fingerprinting with comprehensive two-dimensional chromatography for volatiles and ripening indicators in olive oil. Analityca Chimica Acta, 936, 245-258.

[3] Ridl, J., Esslinger, S., Fauhl Hassek, A. (2015). Review of validation and reporting of non-targeted fingerprinting approaches for food authentication. Analityca Chimica Acta, 885, 17-32.

[4] Cubero-Leon, E., Peñalver, R., Maquet, A. (2014). Review on metabolomics for food authentication. Food Research International, 60, 95-107.

[5] Roberts, L. D., Souza, A. L., Gerszten, R., Clish, C. B. (2012). Targeted Metabolomics. Current Protocols in Molecular Biology, 30. 2, 1-24.

[6] Vinayaveklin, N., Saghtelian, A. (2010). Untargeted metabolomics. Current Protocols in Molecular Biology, 30. 1, 1-24.

[7] Esslinger, S., Rield, J., Fauhl-Hassek, C. (2014). Potential and limitations of non-targeted fingerprinting for authentication of food in official control. Food Research International, 60, 189-204.

[8] Ellis, D. I., Brewster, V. L., Dunn, W. B., Allwood, J. W., Golovanov, A. P., Goodacre, R. (2012). Fingerprinting food: current technologies for the detection of food adulteration and contamination. Chemical Society Reviews, 41, 5706-5727.

[9] Cifuentes, A. (2012). Food analysis: present, future, and foodomics. ISRN Analytical Chemistry, 1-16.

[10] Rohman, A., Che Man, Y. B. (2012). Application of Fourier transform infrared spectroscopy for authentication of functional food oils. Applied Spectroscopy Reviews, 47, 1-13.

[11] Madrile, L., Zeppa, G., Giovannozi, A. M., Rossi, A. M. (2016). Controlling protected designation of origin of wime by Raman spectroscopy. Food Chemistry, 211, 260-267.

[12] Zou, M. Q., Zhang, X. F., Qi, X. H., Ma, H. L., Dong, Y., Liu, C. W., Guo, X., Wang, H. (2009). Rapid authentication of olive oil adulteration by Raman spectrometry. Journal of Agricultural and Food Chemistry, 57, 6001-6006.

[13] López-Díez, E. C., Bianchi, G., Goodacre, R. (2003). Rapid quantitative assessment of the adulteration of virgin olive oils hazelnuts oils using Raman spectroscopy and chemometrics. Journal of Agricultural and Food Chemistry, 51, 6145-6150.

[14] Wu, L., Du, B., Heyden, Y. V., Chen, L., Zhao, L., Wang, M., Xue, X. (2017). Recent advancements in detecting sugar-based adulterants in honey - A challenge. Trends in Analytical Chemistry, 86, 25-38.

[15] Cozzolino, D. (2012). Recent trends on the use of infrared spectroscopy to trace and authenticate natural and agricultural food products. Applied Spectroscopy Reviews, 47, 518-530.

[16] Buyukgoz, G. G., Soforoglu, M., Akgul, N. B., Boyaci, I. H. (2016). Spectroscopic fingerprint of tea varities by Surface enhanced Raman spectroscopy. Journal of Food Science and Technology, 53, 1709-1716.

[17] Rohman, A., Che Man, Y. B., Yusof, F. M. (2014). The use FTIR spectroscopy and chemometrics for rapid authentication of extra virgin olive oil. Journal of the American Oil Chemist' Society, 91, 207-213.

[18] Geana, E. I., Popescu, R., Costinel, D., Dinca, O. R., Ionete, R. E., Stefanescu, I., Artem, V., Bala, C. (2016). Classification of red wines using suitable markers coupled with multivariate statistic analysis. Food Chemistry, 192, 1015-1024.

[19] Garcia, C., Lutz, N. W., Confort-Gouny, S., Cozzone, P. J., Armand. M., Bernad, M. (2012). Phospholipid fingerprints of milk from different mammalians determined by 31P NMR: Towards specific interest in human health. Food Chemistry, 135, 1777-1783.

[20] Longobardi, F., Ventrella, A., Napoli, C., Humpfer, E., Schütz, B., Schäfer, H., Kontominas, M. G., Sacco, A. (2012).Classification of olive oils according to geographical origin by using 1H NMR fingerprinting combined with multivariate analysis. Food Chemistry, 130, 177-183.

[21] Alonso Salces, R. M., Moreno Rojas, J. M., Holland, M. V., Reinero, F., Guillou, C., Héberger, K. (2010). Virgin olive oil authentication by multivariate analyses of 1H NMR fingerprints and δ 13C and δ 2H data. Journal of Agricultural and Food Chemistry, 58, 5586-5596.

[22] Ramakrishnan, V., Luthria, D. L. (2017). Recent applications of NMR in food and dietary studies. Journal of the Science of Food and Agriculture, 97, 33-42.

[23] Cavazza, A., Corradini, C., Musci, M., Salvadeo, P. (2012). High-performance liquid chromatographic phenolic compound fingerprint for authenticity assessment of honey. Journal of the Science of Food and Agriculture, 93, 1169-1175.

[24] Vaclevik, L., Lacina, O., Hajslova, J., Zweigenbaum, J. (2011). The use of high performance liquid chromatography-quadrupole time-of-flight mass spectrometry coupled to advanced data mining and chemometric tools for discrimination and classification of red wines according to their variety. Analityca Chimica Acta, 685, 45-51.

[25] Cuadros Rodríguez, L, Ruiz Samblás, C., Valverde Som, L., Pérez Castaño, E., González Casado, A. (2016). Chromatographic fingerprinting: An innovative approach for food 'identitation' and food authentication – A tutorial. Analityca Chimica Acta, 909, 9-23.

[26] Hibbert, D. B. (2016). Vocabulary of concepts and terms in chemometrics (IUPAC Recommendations 2016). Pure and Applied Chemistry, 88, 407-443.

[27] Hopke, P. K. (2003). The evolution of chemometrics. Analityca Chimica Acta, 500, 365-377.

[28] Chen, J., Lu, Y. H., Wei, D. Z., Zhou, X. L. (2009). Establishment of a fingerprint of raspberries by LC. Chromatographia, 70, 981-985.

[29] Zhou, J., Yao, L., Chen, L., Wu, L., Zhao, J. (2014). Floral classification of honey using liquid chromatography-diode array detection-tandem mass spectrometry and chemometric analysis. Food Chemistry, 145, 941-949.

[30] He, X., Li, J., Zhao, W., Liu, R., Zhang L., Kong, X. (2015). Chemical fingerprint analysis for quality control and identification of Ziyang green tea by HPLC. Food Chemistry, 171, 405-411.

[31] Dutt, V., Chaudhry, V., Khan, I. (2012). Pattern recognition: an overview. American Journal of Intelligent Systems, 2, 23-27.

[32] Bro, R., Smilde, A. (2014). Principal component analysis. Analytical Methods, 6, 2812–2831.

[33] Jain, A. K., Murty, M. N., Flynn, P. J. (1999). Data clustering: a review. ACM Computing. Surveys, 31(3), 264-323.

[34] Daszykowski, M., Walczak, B., Massart, D. L. Density-based clustering for exploration of analytical data. Anal. Bioanal. Chem., 2004, 380, 370-372.

[35] Brereton, R. G. (2015). Pattern recognition in chemometrics. Chemometrics and Intelligent Laboratory Systems, 149, 90-96.

[36] Bevilacqua, M., Bucci, R., Magrì, A. D., Magrì, A. L., Nescatelli, R., Marini, F. Classification and class-modelling, in: Marini, F. (Ed.), Chemometrics in food chemistry, Elsevier, Oxford (UK), 2013, pp. 171-233.

[37] Brereton, R. G., Lloyd, G. R. (2014). Partial least squares discriminant analysis taking the magic away. Journal of Chemometrics, 28, 213-225.

[38] Ruiz Samblás, C., Cadenas, J. M., Pelta, D. A., Cuadros Rodríguez, L. (2014). Application of data mining methods for classification and prediction of olive oil blends with other vegetable oils. Analytical and Bionalytical Chemistry, 406, 2591-2601.

[39] Liu, L., Hu, C., Liu, L., Zhang, S., Chen, K., He, D. (2017). Rapid detection and separation of olive oil and Camellia oil based on ion mobility spectrometry fingerprints and chemometrics models. European Journal of Lipid Science and Technology, 119, 1-9.

[40] Maione, C., Lemos Batista, B., Dobal Campiglia, A., Barbosa J, F., Barbosa, R. M. (2016). Classification of geographic origin of rice by data mining and inductively coupled plasma mass spectrometry. Computers and Electronics in Agriculture, 121, 101-107.

[41] Klockmann, S., Reiner, E., Bachmann, R., Hackl, T., Fischer, M. (2016). Food fingerprinting: metabolomic approaches for geographical origin discrimination of hazelnuts (corylus avellana) by UPLC-QTOF-MS. Journal of Agricultural and Food Chemistry, 64, 9253-9262.

[42] Luts, J., Ojeda, F., Van de Plas, R., De Moor, B., Van Huffel, S., Suykens, J. A. K. (2010). A tutorial on support vector machine-based methods for classification problems in chemometrics. Analityca Chimica Acta, 665, 129-145.

[43] Xu, Y., Zomer, S., Brereton, R. G. (2006). Support vector machine: A recent method for classification in chemometrics. Critical Reviews in Analytical Chemistry, 36, 177-188.

[44] Marini, F. (2010). Classification methods in chemometrics. Current Analytical Chemistry, 6, 72-79.

[45] Rokach, L., Maimon, O. Introduction to Decision Trees, in: Bunke, H., Wang, P. S. P. (Eds.), Data mining with decision trees. Theory and applications, World Scientific Publishing, Singapore, 2008, pp. 1-11.

[46] Mutihac, L., Mutihac, R. (2008). Mining in chemometric. Analityca Chimica Acta, 612, 1-18.

[47] Xu, L., Yan S. M., Cai, C. B., Yu, X. P. One-class partial least squares (OCPLS) classifier. Chemometr. Intell. Lab., 2013, 126, 1-5.

[48] Xu, L., Goodarzi, M., Shi, W., Cai, C. B., Jiang, J. H. (2014). A MATLAB toolbox for class modeling using one-class partial least squares (OCPLS) classifiers. Chemometrics and Intelligent Laboratory Systems, 139, 58-63.

[49] Wold, S., Sjöström, M., Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems, 58, 109-130.

[50] Westad, F., Bevilacqua, M., Marini, F. Regression, in: Marini, F. (Ed.), Chemometrics in food chemistry, Elsevier, Oxford (UK), 2013, pp. 127-168.

[51] Forina, M., Casale, M., Oliveri, P. Application of chemometrics to food chemistry, in: S. Brown, D., Tauler, R. Walczak, B. (Eds.), Comprehensive chemometrics: chemical and biochemical data analysis, Elsevier, Amsterdam (The Netherlands), 2009, pp. 75-126.

[52] Borràs, E., Ferré, J., Boqué, R., Mestres, M., Aceña, L., Busto, O. (2015). Data fusion methodologies for food and beverage authentication and quality assessment – A review. Analityca Chimica Acta, 891, 1-14.

[53] Nescatelli, R., Bonanni, R. C., Bucci, R., Magrì, A. L., Magrì, A. D., Marini, F. (2014). Geographical traceability of extra virgin olive oils from Sabina PDO by chromatographic fingerprinting of the phenolic fraction coupled to chemometrics. Chemometrics and Intelligent Laboratory Systems, 139, 175–180.

[54] Bajoud, A., Medina Rodríguez, S., Gómez Romero, S., Ajal, E. A., Bagur González, M. G., Fernandez Gutierrez, A., Carrasco Pancorbo, A. (2017).Assessing the varietal origin of extra-virgin olive oil using liquid chromatography fingerprints of phenolic compound, data fusion and chemometrics. Food Chemistry, 215, 245-255.

[55] 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.

[56] Obisesan Abidemi, K., Jiménez Carvelo, A. M., Cuadros Rodríguez, L., Ruisánchez, I., Callao, M. P. (2017). HPLC-UV and HPLC-CAD chromatographic data fusion for the authentication of the geographical origin of palm oil. Talanta, 170, 413-418.

[57] Shen, D., Wu, Q., Sciarappa, W. J., Simon, J. E. (2012). Chromatographic fingerprints and quantitative analysis of isoflavones in Tofu-type soybeans, Food Chemistry, 130, 1003-1009.

[58] Alaerts, G., Van Erps, J., Pieters, S., Dumarey, M., van Nederkassel, A. M., Goodarzi, M., Smeyers-Verbeke, J., Vander Heyden, Y. (2012). Similarity analyses of chromatographic fingerprints as tools for identification and quality control of green tea, Journal of Chromatography B, 910, 61–70.

[59] Serrano Lourido, D., Saurina, J., Hernández Cassou, S., Checa, A. (2012). Classification and characterisation of Spanish red wines according to their appellation of origin based on chromatographic profiles and chemometric data analysis. Food Chemistry, 135, 1425-1431.

[60] Bagur González, M. G., Pérez Castaño, E., Sánchez Viñas, M., Gázquez Evangelista, D. (2015). Using the liquid-chromatographic-fingerpint of sterols fraction to discriminate virgin olive from other edible oils. Journal of Chromatography A, 1380, 64-70.

[61] Li, J., He, X., Li, M., Zhao, W., Liu, L., Kong, X. (2015). Chemical fingerprint and quantitative analysis for quality control of polyphenols extracted from pomegranate peel by HPLC. Food Chemistry, 176, 7–11.

[62] Perez Rafols, C., Saurina, J. (2015). Liquid chromatographic fingerprints and profiles of polyphenolic compounds applied to the chemometric characterization and classification of beers. Analytical Methods, 7, 8733-8739.

[63] Schulze, A. E., De Beer, D., Manley, M., Joubert, E. (2014). Chemometric analysis of chromatographic fingerprints shows potential of Cyclopia Maculata (Andrews) kies for

production of standardized extracts with high xanthone content. Journal of Agricultural and Food Chemistry, 62, 10542-10551.

[64] Leme, G. M., Coutinho, I. D., Creste, S., Hojo, O., Carneiro, R. L., Bolzani, V., Cavalheiro, A. J. (2014). HPLC-DAD method to metabolic fingerprinting to the phenotyping of sugarcane genotypes. Analytical Methods, 6, 7781-7788.

[65] Jablonski, J. E., Moore, J. C., Harnly, J. M. (2014). Nontargeted detection of adulteration of skim milk powder with foreign proteins using UHPLC-UV. Journal of Agricultural and Food Chemistry, 62, 5198-5206.

[66] Kuś , P. M., van Ruth, S. (2015). Discrimination of Polish unifloral honeys using overall PTR-MS and HPLC fingerprints combined with chemometrics. LWT – Food Science and Technology, 62, 69-75.

[67] Ding, X. X., Ni, Y. N., Kokot, S. (2015). Analysis of different Flos Chrysanthemum tea samples with the use of two-dimensional chromatographic fingerprints, which were interpreted by different multivariate methods. Analytical Methods, 7, 961-969.

[68] Wang, C., He, H., Zhang, J. L., Li, X., Ma, Z. L. (2016). High performance liquid chromatography (HPLC) fingerprints and primary structure identification of corn peptides by HPLC-diode array detection and HPLC-electrospray ionization tandem mass spectrometry. Journal of Food and Drug Analysis, 24, 95-104.

[69] Pardo Mates, N., Vera, A., Barbosa, S., Hidalgo Serrano, M., Nunez, O., Saurina, J., Hernandez Cassou, S., Puignou, L. (2017). Characterization, classification and authentication of fruit-based extracts by means of HPLC-UV chromatographic fingerprints, polyphenolic profiles and chemometric methods. Food Chemistry, 221, 29-38.

[70] Zhao, J., Cheng, N., Xue, X. F., Wu, L. M., Zhu, X. H., Cao, W. (2015). Chromatographic ECD fingerprints combined with a chemometric method used for the identification of three light-coloured unifloral honeys. Analytical Methods, 7, 8393-8401.

[71] Zhao, J., Du, X. J., Cheng, N., Chen, L. Z., Xue, X. F., Zhao, J., Wu, L. M., Cao, W. (2016). Identification of monofloral honeys using HPLC-ECD and chemometrics. Food Chemistry, 194, 167-174.

[72] Jiménez Carvelo, A., Pérez Castaño, E., González Casado, A., Cuadros Rodríguez, L. (2017). One input-class and two input-class classifications for differentiating olive oil from other edible vegetable oils by use of the normal-phase liquid chrtomatography fingerprint of the methyl-transesterified fraction. Food Chemistry, 221, 1784-1791.

[73] Jiménez Carvelo, A., González Casado, A., Pérez Castaño, E., Cuadros Rodríguez, L. (2017). Fast-HPLC fingerprinting to discriminate olive oil from other edible vegetable oils by multivariate classifications methods. Journal of AOAC International, 100, 345-350.

[74] Jiménez Carvelo, A., González Casado, A., Cuadros Rodríguez, L. (2017). A new analytical method for quantification of olive and palm oil in blends with other vegetable edible oils based on the chromatographic fingerprints from the methyl-transesterified fraction. Talanta, 540-547.

[75] Sârbu, C., Nacu-Briciu, R. D., Kot-Wasik, A., Gorinstein, S., Wasik, A., Namieś nik, J. (2012). Classification and fingerprinting of kiwi and pomelo fruits by multivariate analysis of chromatographic and spectroscopic data. Food Chemistry, 130, 994-1002.

[76] Arbulu, M., Sampedro, M. C., Gómez-Caballero, A., Goicolea, A., Barrio, R. J. (2015). Untargeted metabolomic analysis using liquid chromatography quadrupole time-of-flight

mass spectrometry for non-volatile profiling of wines. Analityca Chimica Acta, 858, 32-41.
[77] Guijarro Diez, M., Nozal, L., Marina, M. L., Crego, A. L. (2015). Metabolomic fingerprinting of saffron by LC/MS: novel authenticity markers. Analytical and Bionalytical Chemistry, 407, 7197-7213.
[78] Rubert, J., Lacina, O., Zachariasova, M., Hajslova, J. (2016). Saffron authentication based on liquid chromatography high resolution tandem mass spectrometry and multivariate data analysis. Food Chemistry, 204, 201-209.
[79] Geng, P., Harnly, J. M., Chen, P. (2016). Differentiation of bread made with whole grain and refined wheat (T. aestivum) flour using LC/MS-based chromatographic fingerprinting and chemometric approaches. Journal of Food Composition and Analysis, 47, 92-100.
[80] Zhu, L., Yang, S. T., Li, G. K., Zhang, X. G., Yang, J., Lai, X. F., Yang, G. W. (2016). Simultaneous analysis of tocopherols, tocotrienols, phospholipids, gamma-oryzanols and beta-carotene in rice by ultra-high performance liquid chromatography coupled to a linear ion trap-orbitrap mass spectrometer. Analytical Methods, 8, 5628-5637.
[81] De la Mata Espinosa, P., Bosque Sendra, J. M., Bro, R., Cuadros Rodríguez, L. (2011). Discriminanting olive and non-olive oils using HPLC-CAD and chemometrics. Analytical and Bionalytical Chemistry, 399, 2083-2092.
[82] De la Mata Espinosa, P., Bosque Sendra, J. M., Bro, R., Cuadros Rodríguez, L. (2011). Olive oil quantification of edible vegetable oil blends using triacyglycerol chromatographic fingerprints and chemometrics tools. Talanta, 85, 177-182.
[83] Bajoud, A., Pacchiarotta, T., Hurtado-Fernández, E., Olmo García, L., García Villalba, R., Fernandez Gutierrez, A., Mayboroba, O. A., Carrasco Pancorbo, A. (2016). Comparing two metabolic profiling approaches (liquid chromatography and gas chromatography coupled to mass spectrometry) for extra-virgin olive oil phenolic compounds analysis: A botanical classification perspective. Journal of Chromatography A, 1428, 267-279.
[84] Biedermann, M., Grob, K., Mariani, C. (1993). Transesterified and on-line LC-GC for determining the sum of free and esterified sterols in edible oils and fats. European Journal of Lipid Science and Technology, 95, 127-133.
[85] Cuadros Rodríguez, L., Pérez Castaño, E., Ruiz Samblás, C. (2017). Quality performance metrics in multivariate classification methods for qualitative analysis. Trends in Analytical Chemistry, 80, 612-624.

ARTÍCULO DE REVISIÓN CIENTÍFICA

Emergent data mining/machine learning methods for the analytical evaluation of food quality and authenticity

(Artículo enviado en septiembre 2018 a la revista Analytical and Bioanalytical Chemistry)

Emergent data mining/machine learning methods for the analytical evaluation of food quality and authenticity

Ana M. JIMÉNEZ-CARVELO✉, ANTONIO GONZÁLEZ-CASADO, Luis CUADROS-RODRÍGUEZ

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

Abstract

The variety of information and data volume which is currently acquired by the modern analytical instrumentation, aimed to carry out a better authentication of foods, have been drastically increased in recent years. In order to deal with the large volume and complexity of the available trial data, several pattern recognition tools have been developed. The most widely used methods have been principal component analysis (PCA), partial least squares-discriminant analysis (PLS-DA), soft independent modelling by class analogy (SIMCA), k-nearest neighbour (kNN), parallel factor analysis (PARAFAC), or multivariate curve resolution-alternating least square (MCR-ALS). Nevertheless, there are other emergent data treatment methods, such as support vector machine (SVM), classification and regression tree (CART) or random forest (RF), which show a great potential and more advantages with regard to conventional ones. In this paper, the background of these methods is explained, and the reported studies in which these three methods have been applied in the food quality and authenticity field are reviewed and discussed. In addition, the employed technical terminology is clarified according to this particular working area.

Keywords

Data mining, random Forest, CART, decision tree, food chemistry

✉ Corresponding author: phone: +34 958240797; fax: +34 958243328; email: amariajc@ugr.es

1. Introduction

The assurance of the food authenticity has been the main concerns of many consumers, as well as manufacturers of high-quality products and official bodies and authorities in response to the need of consumer protection detecting of potential food frauds. Globalization and free trade agreements have fostered greater exchange and access to food globally and with them the problems associated with frauds such as adulteration, substitution, and falsification have increased. For this reason, multivariate analysis or pattern recognition techniques are a powerful tool to carry out quality control and food authentication [**1,2,3,4**].

The core goal of the multivariate pattern recognition methods is carried out the suitable data treatment to model and to characterise a set of objects or samples regarding a particular feature or behaviour. To this end, the significant and non-evident information is extracted allowing to establish relationships between the objects belonging to the working set, or between the set of objects and a particular feature, according to the similarity of their spectra, chromatograms, elementary analysis, images, etc. Likewise, these tools must allow the classification of new samples to a certain group or the reliable prediction of the value of a determined property in a quick and objective way [**5**].

Pattern recognition methods are divided into two main groups: unsupervised methods, where the main tools are: principal component analysis (PCA) and hierarchical cluster analysis (HCA), and supervised methods of analysis such as: k-nearest neighbour (kNN) [6], partial least squares-discriminant analysis (PLS-DA) [**7**] and soft independent modelling by class analogy (SIMCA) [**8**] among others. Figure 1 shows a straightforward flowchart of the conventional pattern recognition techniques.

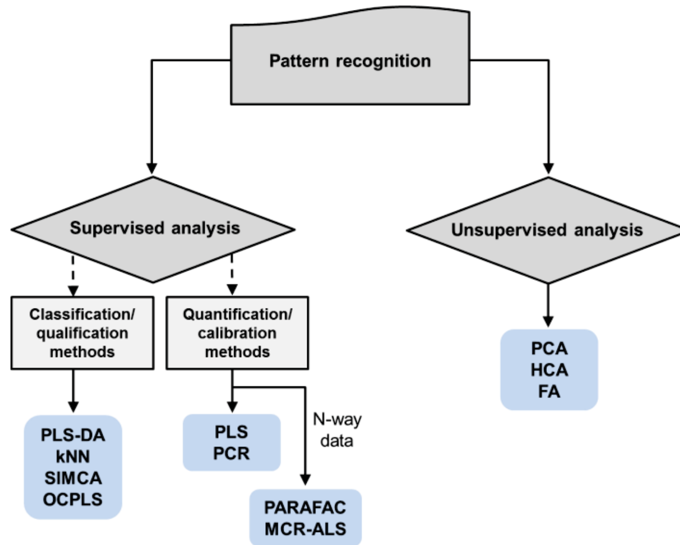


Figure 1. General overview of the traditional methods of pattern recognition.

The exploratory analysis is habitually used to scout the data structure and to evaluate if there are tendencies in the data set. The principal component analysis (PCA) is a valuable statistical tool whose goal is to maximize the information of the variance in the data and visually shows it in as few components, as possible. The main applications are: to provide information on the natural groupings of the objects and to reduce the number of variables necessary to represent the system, providing a new set of latent variables so-called principal components [**9**]. Nevertheless, sometimes PCA has been erroneously applied and in some works is used as a classification method to develop and validate classification models [**10**,**11**,**12**]; this is a grave mistake that unfortunately had been found to be still current. Cluster analysis is based on the intrinsic similarity between groups of objects. The results of the hierarchical clusters are presented in the form of a dendrogram where the objects are organized in rows according to their similarity [**13**].

The supervised methods of analysis are divided into two groups: (i) classification or qualification methods and (ii) calibration or quantitation methods. Multivariate classification/qualification methods have been defined as chemometric techniques designed to find mathematical models that can recognize each sample’s membership of its relevant class on the basis of a particular data set and they involve the application of different chemometric algorithms with two main statistical background relating to discrimination and class-modelling approaches [**14**]. There are a lot of classification techniques, the most common are: kNN, PLS-DA, SIMCA and unequal class modelling (UNEQ).

Multivariate calibration/quantitation methods are properly multivariate regression methods aimed to establish functional relationships between the analytical signal acquired from a set of samples each one and a characteristic feature or composition of these ones. The most widely used algorithm is partial least squares (PLS) regression [**15**].

The development of a pattern recognition supervised model draws up of two steps. In the first stage the model is built using a set of objects or samples in which the class or particular features are known (training set). The second stage is carried out to evaluate and to validate the previously established model; for this purpose, additional objects or samples fulfilling the same requirements are employed, which are not part of the original training set [**16**]. In these methods, it is assumed that there are a number of objects or samples which are acting as analytical standards because the outcome of interest (the qualitative class or the value of one or more quantitative features) are formerly known or they have already been measured.

Notice that the assignment of the objects or samples to a specific class could have a qualitative basis if kNN or SIMCA methods are applied or a quantitative basis if PLS-DA is used. Indeed, PLS-DA firstly performs a multivariate regression model in order to establish the discriminant function to separate the regions of the space for each class, and then the objects/samples are classified in one class or another [**17**].

In recent years, in the food field, the application of new pattern recognition algorithms are growing due to their advantages and potentiality to solve complex problems related to the food authenticity. The more significant one amount them are support vector machine (SVM), classification and regression trees (CART), and random forest (RF) which may be used to perform both classification and calibration models. Surprisingly, the application is still scarcer in food chemistry area, nevertheless in other areas as metabolomics is already widely used. Even some authors have stated their advantages in contrast to conventional techniques. For example in the case of PLS-DA vs SVM, it has been stated that *in comparison to PLS-DA, SVM is not influenced by the distribution of the different sample classes, on the contrary, it focuses on which side of the support vectors particular test samples fall* [**18**]. Similarly, in ecology area was reported the advantages of the RF algorithm [**19**].

In this paper, the employment of these emergent data mining/machine learning methods SVM, CART and RF applied in the food field, are reviewed and described. Examples

will be provided to demonstrate the potential of these techniques in this working area.

2. Terminology

The automation and computerization of the laboratories have resulted in numerous changes; one of them is the acquisition of a high amount of data. Nevertheless, the possession of a lot of data is often far from providing adequate answers, if it is not applied the right data processing tools. Collecting data is not synonymous with possessing information; the data must be treated and interpreted to convert them into useful information for the user or the analyst. The denomination of the kind of tools to process a high amount of data is different depending on the work area.

In the analytical chemistry area is where there is more variability of concepts. Some authors use: 'pattern recognition methods', 'multivariable analysis methods', and the concept most commonly employed is 'chemometric tools' to designate to the data treatment methods applied in the studies carried out. Chemometrics is defined as *an approach to analytical and measurement science that uses mathematical, statistical and other methods of formal logic to determine (often by indirect means) the properties of substances that otherwise would be very difficult to measure directly* [**20**].

In the engineering field for the processing of the signals or image is usual to label these kinds of techniques as tools of 'computational intelligence' or 'artificial intelligent'. The IUPAC defined the artificial intelligent as *the capability of a machine to perform human-like intelligence functions such as learning, adapting, reasoning and self-correction. The main areas of application are currently in expert systems, computer vision, natural language processing, robotics, and speech synthesis and recognition* [**21**]. Other authors define this term as *the interaction of several kinds of disciplines, such as computer science, cybernetics, information theory, psychology, linguistics, and neurophysiology. Artificial intelligence is a branch of computer science, involved in the research, design and application of intelligent compute*r [**22**]. In this area the artificial neural networks (ANN) are the most employed algorithm. It is based on a set of "nodes" or "artificial neurons" interconnected with each other in the form of network which try to simulate the network of neurons in the human brain [**23**]. ANN is not explained in this study due to its different application, although they can be included within the emergent machine learning methods [**24**,**25**,**26**].

Into the area of the health (medicine, pharmacy, biology and biotechnology) is used the term 'bioinformatics' which is defined as the *discipline encompassing the development and utilization of computational facilities to store, analyse, and interpret biological data* [**27**].

Besides these, there is a general term which concerns all these tools regardless of the area of work known as 'data mining'. This term appeared during the 1960s, but it was not until the eighties when its consolidation began together with the concept KDD (knowledge discovery in databases) [**28**,**29**]. Moreover, the concept 'machine learning' is also commonly used for the same purpose [**30**]. Both terms are often used interchangeably to refer to all these processing data techniques.

Aims of data mining can be descriptive, (i.e. uncovering similarities between the data set), or predictive (i.e. classify new data based on model built and validated previously). It is based on the collection, storage and treatment of a lot of data to take the most suitable decisions about a problem. Therefore, it is an interdisciplinary field with the overall objective of revealing relationships in the data from whatever source or origin. For this purpose are applied complex tools of treatment of data to detect and to found mainly hidden patterns, associations, and structures of the high amount of data or to select and to filter the useful information from big databases [**31**]. Machine learning concept is also known as: *the techniques involved in dealing with vast data in the most intelligent fashion (by developing algorithms) to derive actionable insights. In these techniques, we expect the algorithms to learn by them without being explicitly programmed* [**32**]. Consequently, data mining could be referred to the area in general and machine learning could be denoted for the algorithms.

Some authors classify the emergent data methods studied in this review in four machine learning classes: (i) information-based learning, (ii) similarity-based learning, (iii) probability-based learning, and (iv) error-based learning. DT (RF and CART) techniques are inside information-based learning class and SVM is within error-based learning [**33**]. This classification seems to us very successful, since the techniques are classified according to how they perform the building of the different regions for each classes of the classification model.

Over recent years, the field of the food chemistry is using the term data mining more frequently, leaving behind the concepts of 'pattern recognition techniques' or 'chemometric tools' to refer to the algorithms used to process the data. Even is being

taken up within ISO/IEC 17025 standard [**34**]. The inclusion of data mining methods to the area of food chemistry is owing to the need to perform the verification of the legal requirements of the different countries to confirm and to assure the authenticity of the food. Besides this, the consumers demand from the producer more information and knowledge about the foodstuff.

Likewise, the food sector is a highly competitive and global, thus food producers seek to consolidate in the national and international current emerging markets and to make a difference with their products. Product differentiation is 'the key' to take a leading position in the global market of the sector. For example a good strategy to take a prominent position over the competitors is to take advantage of the difference in chemical composition or organoleptic characteristics of the food. The current trend is developed quick analytical methods able to authenticate food products. All this has caused the development of analytical instrument more powerful and the application of the new methodologies to obtain more information about the objects/samples of study. For example the development of more advanced sensors which allow monitoring the food with a high level of detail with the consequence that a high amount of data is collected. Thus, alternative methods to the conventional techniques of processing of data are employed.

Traditionally, the data mining methods concept has been used in the food chemistry when techniques well-known as PCA, kNN, PLS-DA, SIMCA, or algorithms applied to second-order data as parallel factor analysis (PARAFAC) and multivariate curve resolution-alternating least squares (MCR-ALS) have been applied [**1,4,35,36**]. Nevertheless, as it has been reported in the introduction section the current trend in food chemistry is to use emerging methods of pre-processing data since they show advantages and greater potency than the conventional methods previously cited. Figure 2 shows a plot about the trend of publications in food chemistry applying the data mining methods in the recent years.

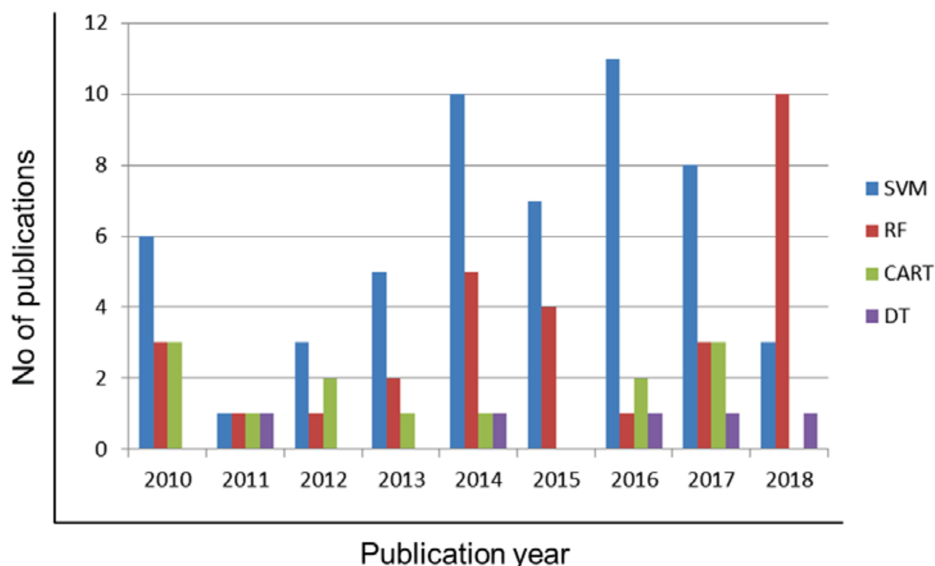


Figure 2. Tendency graph of the publication in the food chemistry area apply SVM, RF, DT and CART.

As can be seen in the graphic 2 the SVM techniques is the most commonly applied but in the two last year the employment of RF algorithm, whose application in food chemistry is scarcer, has increased significantly. Moreover, in the figure 3 is shown as the data mining term has increased its use in the food chemistry area in the recent years.

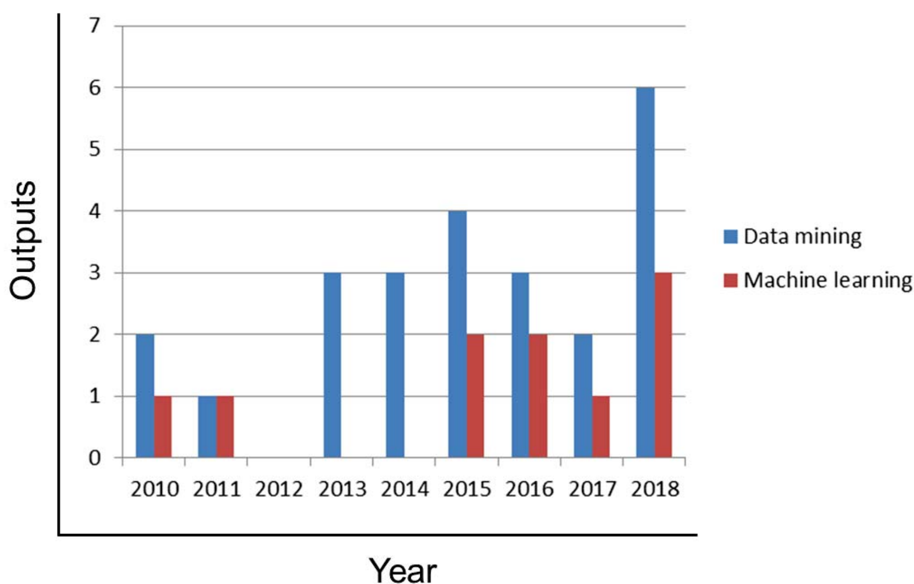


Figure 3. Tendency graph of the use of the data mining and machine learning concepts in the last 10 years.

3. Emergent tools

The most commonly algorithms applied for the processing data in food chemistry have been the conventional techniques cited in the previous sections with the main aim of performing the authentication of food. Authentication of food involves the verification of the genuineness of its declared specifications, which include several features such as the analytical determination of characteristic, detection of adulterants and / or contaminating components and the verification of differentiated quality requirements.

The studies published employ the conventional methods of analysis of data with the purpose of carrying out of the analysis of the similarity between signals for food identification, the classification of foods according to various criteria (i.e., botanical or animal species, geographical origin), detection of de adulterations, and the prediction of properties related to food quality, such as: antioxidant capacity [**37**,**38**,**39**,**40**].

Some authors declare that to perform a correct authentication of food is better to apply class modelling methods. They state that *to classify perfectively a new sample only is possible if this sample is a member of one of the predefines classes* [**41**] if discriminant analysis methods are employed in contrast to class modelling methods as SIMCA, which build a acceptance area around the target class, and for this reason if the new sample does not belong to the target class, it will be out of the acceptance area of the target class. No matter its class has not been predefined in the training stage of the classification model. Nevertheless, that is not all quite true since it is possible to establish confidence thresholds for each class from the training samples of the classification models using SVM or PLS-DA and in this way restrict the confidence area for each class and to avoid this inconvenience. Indeed, some authors have raised this strategy [**42**].

Moreover, different authors have collected the studies and the use of these techniques in different reviews [**43,44,45,46,47,48**]. Recently, a review paper has been published where is described an overview of the all the stages of the data analysis process carried out to take an appropriate decision to solve a problem. Nevertheless, in this work only conventional data processing techniques are mentioned, leaving out the new data mining methods such as RF and CART, although SVM is stated [**49**]. This demonstrates that the inclusion of data mining techniques within the field of analytical chemistry and specifically in the food chemistry is relatively recent.

Within the machine learning methods, SVM is the most applied method in food chemistry and from which there are more studies have been published (see figure 2) [**50**,**51**]. The goal of SVM is to find the best hyperplane in the space which splits the classes from the objects or samples applying a maximization method. The aim is maximizing the "margin", which is based on the sum of the distances from the hyperplane built to samples closer correctly classified in their corresponding classes, moreover SVM performs a penalizing the number of misclassified samples. The building of the hyperplane can be executed in the original space, thus the SVM model is linear or in a space of higher dimension using the kernel functions, therefore SVM model is non-linear [**52**]**.**

The main advantage of SVM over PLS-DA is the making of the separation regions of the different classes when these are not sufficiently evident. Nevertheless, the development of the classification methods using PLS-DA is easier and faster to carry out than SVM, since PLS-DA only performs a regression by partial least squares on the original data whereas SVM take into account the transformation of the data in a space of a higher dimension.

In the studies published in this area SVM is used for different purposes: (i) classification of the food according to their geographic origin, (ii) sensorial evaluation, (iii) detection of adulterations, (iv) quantification of compounds and (v) quality control.

DT predicts the label associated with an instance X by traveling from a root node of a tree to a leaf, where each label corresponding to one class. Each of those results creates additional nodes, which branch out into other possibilities. This gives it a structure similar to a tree. There are three kinds of nodes: probability nodes, decision nodes and terminal nodes. A probability node (root node) is represented by a circle and it shows the probabilities of certain results. A decision node is represented by a square and it shows a decision to be made, and a terminal node shows the final result of a decision path [**53**]. Thus, the classification of a new sample begins in the root node of the tree and it follows the branch appropriate to its outcome.

There are different DT techniques. In food chemistry the most common are CART and RF. Sometimes, these tools are called 'ensemble methods' because to obtain better predictive performance than could be obtained from any of the constituent models. The main idea behind is to combine several individual classifiers to obtain a classifier that outperforms every one of them [**54**]. RF is a combination of decision trees which is

constructed using different sets of randomly selected input (X) variables [**18**], in contrast to CART is a single tree which shows a lot of branch where the data set is split according to the selected decision, and the procedure is repeated as often as necessary. Figure 4 shows a plot about the differences between them [**55**].

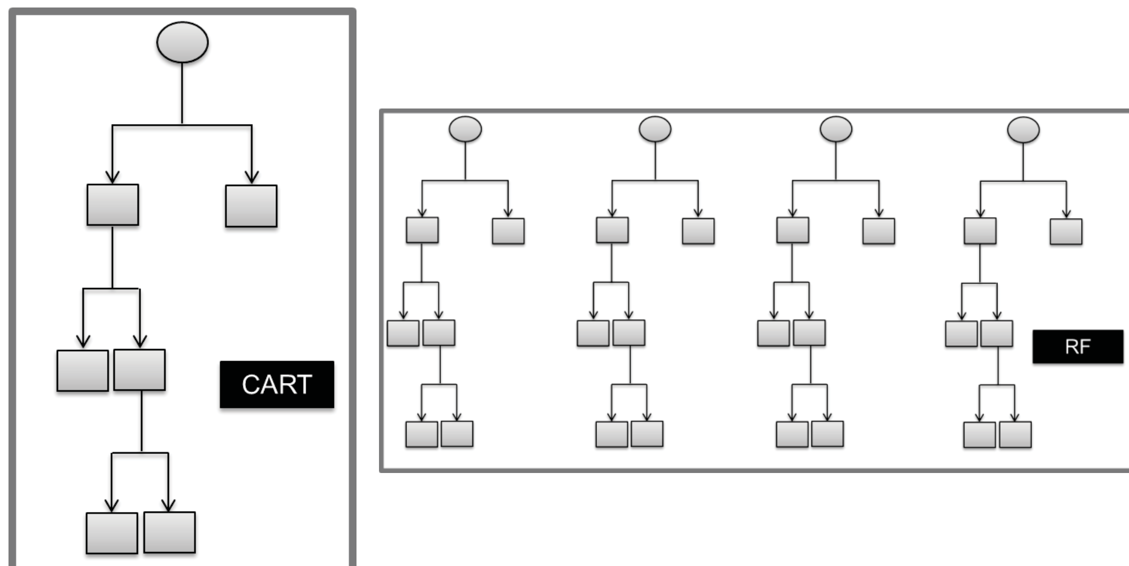


Figure 4. Diagram of classification and regression tree (CART) vs random forest (RF).

As mentioned in the introduction section RF has been employed scarce in food chemistry, however in the recent year its potential in this area has been shown with different studies, even new software has been developed to apply RF using spectroscopic techniques in the food chemistry area [**56**].

Table 1 shows a review of the main applications published since 2010 in food chemistry.

Table 1. Studies in food chemistry area where SVM, RF, CART and DT are applied.

Food	Purpose	Analytical technique	Tool	Ref.
Honey	Classification according to their floral and geographical origin	Rheometer	PCA, PLS, PCR, SVM	[**57**]
	Classification according to their	GC x GC-TOF	SIMCA, DPLS, LDA, **SVM**	[**58**]

	geographical origin			
	Classification according to their geographical origin	Electronic tongue	PCA, HCA, PLS, **SVM**	[**59**]
	Classification according to their phenolic composition	HPLC-UV	PLS-DA, **SVM**	[**60**]
	Classification according to their botanic origin	HPLC-IR, viscosimeter, HPLC-UV, spectrophotometer	**CART**	[**61**]
	Classification according to their botanic and geographical origin	ICP-MS	LDA, **CART**	[**62**]
Cocoa	Evaluation of sensory quality	Sensory tasting	PLS, **SVM**, MLR	[**63**]

Table 1. *Continue.*

Food	Purpose	Analytical technique	Tool	Ref.
Cocoa	Quantification of the total fat content	FT-NIR	PLS, **SVM**	[**64**]
	Classification according to their geographical origin	Electronic tongue	FDA, PCA, kNN, **SVM**	[**65**]
Olive oil	Detection and quantification of adulteration of extra virgin olive oil with other vegetable edible oils	Raman	PLS, **SVM**	[**66**]
	Discrimination of olive oil from other vegetable edible oils	HPLC-CAD	PCA, PLS-DA, OCPLS, kNN, SIMCA, **SVM**	[**67**]
	Discrimination of olive oil from other vegetable edible oils	HPLC-CAD	PCA, PLS-DA, OCPLS, kNN, SIMCA, **SVM**	[**68**]
	Discrimination of olive oil from other vegetable edible oils and quantification of the proportion of olive oil in blends with other vegetable oils.	FTIR, Raman	PCA, PLS-DA, OCPLS, kNN, SIMCA, **SVM**	[**69**]
	Quantification of olive oils in blends with other vegetable oils	HPLC-CAD	PLS, **SVM**	[**70**]
	Classification according to their geographical origin	HPLC-IR, GC-FID	PCA, **DT**	[**71**]

Table 1. *Continue.*

Food	Purpose	Analytical technique	Tool	Ref.
Olive oil	Classification according to their geographical origin	ICP-MS	PLS-DA, **SVM, RF**	[72]
Tofu	Study of the shelf-life	FTIR	PLS, **SVM**	[73]
Vegetable oil	Quantification of the fatty acid compounds	Raman	**SVM**	[74]
	Discrimination of vegetable oils according to their quality	GC-MS	PCA, **RF**	[75]
	Detection of the adulteration	GC x GC-TOF	PCA, HCA, **RF**	[76]
Andiroba Oil	Discrimination of andiroba oil adulterated and non-adulterated	FTIR-HATR	PLS-DA, **RF**	[77]
Licors	Quality control	HS-SPME-MS	PLS, **SVM**	[78]
Beef	Evaluation of sensory quality	Electronic nose	PCA, DFA, **SVM**	[79]
	Evaluation of sensory quality	GC-MS	LDA, SIMCA, PLS-DA, **SVM**	[80]
	Evaluation of sensory quality	Electronic nose	**SVM**	[81]
Coffee	Evaluation of the authenticity according to their trace element	ICP-MS	MLP, **SVM,** NB	[82]

Table 1. *Continue.*

Food	Purpose	Analytical technique	Tool	Ref.
Coffe	Classification according to their geographical origin	NIR and FTIR	**SVM**	[83]
Tea	Discrimination of Green and black tea	Voltammetric	PCA, **SVM**	[84]
	Classification between different kinds of tea	HPLC-UV	PCA, SVM, **RF**	[85]
	Discrimination of five varieties of green tea and quantification of polyphenols compounds	NIR, UV-Vis	PCA, PLS, **RF**	[86]
	Classification according to their geographical origin	ICP-AES	LDA, PLS-DA, **DT**	[87]
	Classification according to their botanic and geographical origin	UV-Vis	kNN, **CART**, SIMCA, PLS-DA, PCA-LDA	[88]
Ginseng	Classification according to their geographical origin	FT-MIR, NIR	RF	[89]
Persimmon	Classification according to their geographical origin	FT-NIR	HCA, PCA, **SVM**	[90]
Apple	Determination of pesticides	Raman	PLS, **SVM**	[91]
Cheese	PDO authenticity	SPME-MS	PCA, LDA, SIMCA, **SVM**	[92]

Table 1. *Continue.*

Food	Purpose	Analytical technique	Tool	Ref.
Cheese	Quality control	PTR-TOF-MS	**RF**	[93]
Jujube fruit	Classification according to their geographical origin	NIR	PCA, LDA, **SVM,** ANN	[94]
Wine	Classification according to their geographical origin	HPLC-DAD	**SVM**	[95]
	Evaluation of sensory quality	GC-MS	**RF**	[96]
	Classification according to their geographical origin	UV/Vis/NIR	LDA, SIMCA, **SVM**	[97]
	PDO authenticity	HPLC-DAD	LDA, SIMCA, **SVM**	[98]
Grape juice	Discrimination of organic grape juice from conventional grape juice	ICP-MS	**SVM, DT**	[99]
Sugar	Quality control	NIR	**SVM**	[100]
	Authenticity evaluation	ICP-MS	NB, **RF**	[101]
Rice	Discrimination of organic rice from conventional rice	ICP-MS	**SVM**	[102]
	Classification according to their geographical origin	ICP-MS	**SVM, RF**, ANN	[103]

Table 1. *Continue.*

Food	Purpose	Analytical technique	Tool	Ref.
Rice	Detection of adulterations	MS	**RF, SVM**, kNN	[**104**]
	Classification according to their geographical origin	Raman	SIMCA, PLS-DA, kNN, **SVM**	[**105**]
	Quality control	GC-MS	**RF**	[**106**]
	Quantification of ediphenphos	Raman	PCA, PLS, **RF**	[**107**]
Pepper	Determination of pesticides	Raman	**SVM**	[**108**]
Juices	Detection of adulteration of tomato juices	Electronic nose and tongue	CDA, **SVM**, PCR	[**109**]
	Detection of additives	Electronic nose	PLS, **SVM**, **RF**	[**110**]

Acronyms: CAD (charged aerosol detector), CART (classification and regression tree), CDA (canonical discriminant analysis), DAD (diode array detector), DFA (discriminant function analysis), DPLS (discriminant partial least squares), DT (decision tree), FID (flame ionization detector), FTIR-HATR (Fourier transform infrared spectroscopy- horizontal attenuated total reflectance), FT-NIR (Fourier transform-near infrared), GC (gas chromatography), HCA (hierarchical cluster analysis), HPLC (high performance liquid chromatography), HS (head space),ICP (inductively coupled plasma), IR (refractive Index), kNN (k-nearest neighbour), LDA (linear discriminant analysis), MLP (Multilayer Perceptron), MS (mass spectrometry), NB (naive Bayes), OCPLS (one class partial least squares), PCA (principal component analysis), PCR (principal component regression), PLS (partial least squares), PLS-DA (partial least squares-discriminant analysis), PTR (Proton-transfer-reaction), RF (random forest), SIMCA (soft independent modelling by class analogy), SPME (Solid-phase microextraction), SVM (support vector machine), TOF (time of fly), UV (ultra violet), Vis (visible).

Of all the works collected in table 1 where the conventional techniques and the machine learning methods have been applied, the authors would like to highlight the following studies due to their excellent results using SVM and DT.

Maione et al. [**103**] developed different classification methods to discriminate rice samples according to their geographical origin. For this purpose the authors applied RF, SVM and multilayer perceptron (MLP). The evaluation of these multivariate classification methods was carried out using the following performances metrics: accuracy, sensitivity, specificity and area under curve value (AUC) and in all the cases SVM and RF were better than MLP, achieving values above 95% in contrast to MLP which were below 90%. Likewise, Ni et al. [**87**] built several classification models to different green tea of different zones from China. In this case they compared the results employed linear discriminant analysis (LDA), PLS-DA and DT and they obtained an accuracy value of 100% when DT was used in contrast to PLS-DA and LDA whose results were 84.5% and 81.4% respectively. Finally, other study to highlight is carried out by Teye *et al.* [**65**] in which established classification models to distinguish cocoa beans samples using Fisher's discriminant analysis (FDA), kNN and SVM. The results revealed that SVM was better than kNN and FDA since 100% of the samples were correctly classified.

4. Final remarks

SVM, CART and RF constitute an emergent group of pattern recognition methods which are yielding promising results in the food quality and authenticity field. Considering only these three methods, SVM is by far the most used one and in most cases it is stated that SVM is performing better than other conventional methods more known, for example, PLS-DA.

In addition, CART and RF are alternative pattern recognition methods which are currently applied in the food field. In other related areas such as metabolomics, some authors have already highlighted the advantages of these machine learning methods as compared to conventional techniques. However, there are still very few reported studies in which CART and RF are applied in the food field even though their worth has been widely proved and outstanding results are been yielded.

References

[1] Zielinski A.A.F, Haminiuk C.W.I, Nunes C.A, Schnitzler E, van Ruth S.M, Granato D. Chemical composition, sensory properties, provenance, and bioactivity of fruit juices as assessed by chemometrics: a critical review and guideline. Compr Rev Food Sci F. 2014; 13:300-316.

[2] Bevilacqua M, Bro R, Marini F, Rinnan A, Rasmussen M, Skov T. Recent chemometrics advances for foodomics. Trends Anal Chem. 2017; 96:42-51.

[3] Brereton R.G, Jansen J, Lopes J, Marini F, Pomerantsev A, Rodionova O, Roger J.M, Walczack B, Tauler R. Chemometrics in analytical chemistry–part I: history experimental design and data analysis tools. Anal Bioanal Chem. 2017; 409:5891-5899.

[4] Callao M.P, Ruisánchez I. An overview of multivariate qualitative methods for fraud detection. Food Control. 2018; 86:283-293.

[5] Brereton R.G. Pattern recognition in chemometrics. Chemometr Intell Lab Syst. 2015;149:90-96.

[6] Steinbach M, Tan P.N. kNN: k-Nearst neighbors in X. Wu, V. Kumar, editors. The Top Ten Algorithms in Data Mining. Chapman & Hall / CRC Press, Boca Raton (FL); 2009, pp. 151-161.

[7] Ballabio D, Consonni V. Classification tools in chemistry. Part 1: linear models. PLS-DA. Anal Methods. 2013; 5:3790-3798.

[8] Oliveri P, Downey G. Multivariate class modelling for the verification of food-authenticity claims. Trends Anal Chem. 2012; 35:74-86.

[9] Bro R, Smilde A. Principal component analysis. Anal Methods. 2014; 6:2812–2831.

[10] Hakki Boyaci I, Tümay Temiz H, Selin Uysal R, Murat Velioglu H, Jafarzedh Yadegari R, Mahmoudi Rishkan M. A novel method for discrimination of beef and horsemeat using Raman spectroscopy. Food Chem. 2014; 148:37–41.

[11] Chung I.M, Kim J.K, Lee J.K, Kim S.H. Discrimination of geographical origin of rice (*oryza sativa L.*) by multielement analysis using inductively coupled plasma atomic emission spectrocopy and multivariate analysis. J Cereal Sci. 2015; 65:252–259.

[12] Sun X, Lin W, Li X, Shen Q, Luo H. Detection and quantification of extra virgin olive oil adulteration with edible oils by FT-IR spectroscopy and chemometric. Anal Methods 2015; 7:3939–3945.

[13] Drab K, Daszykowski. Clustering in Analytical Chemistry. J AOAC Int. 2014; 97:29–38.

[14] Cuadros Rodríguez L, Pérez Castaño E, Ruiz Samblas, C. Quality performance metrics in multivariate classification methods for qualitative analysis. Trends Anal. Chem. 2016; 80:612–624.

[15] Mehmood T, Ahmed B. The diversity in the applications of partial least squares:an overview. J Chemom. 2016; 30:4–17.

[16] Szymanska E, Gerretzen J, Engel J, Geurts B, Blanchet L, Buydens L.M.C.Chemometrics and qualitative analysis have a vibrant relantionship. Trends Anal Chem. 2015; 69:34–51.

[17] Brereton R.G, Lloyd G.R. Partial least squares discriminant analysis: taking the magic away. J Chemom. 2014; 28:213–225.

[18] Gromski P.S, Muhamadali H, Ellis D.I, Xu Y, Correa E, Turner M.L, Goodacre R. A tutorial review:metabolomics and partial least squares-discriminant analysis-a marriage of convenience of a shoutgun wedding. Anal Chim Acta 2015; 879:10–23.

[19] Cutler D.R, Edwars T.C, Beard K.H, Cutler A, Hess K.T, Gibson J, Lawler J.J. Random forest for classification in ecology. Ecology 2007; 88:2783-2792.

[20] Lavine B.K. Chemometrics. Anal Chem. 2000; 72:91–97.

[21] Kingston H.M, Kingston M.L. Nomenclature in laboratory robotics and automation (IUPAC Recommendations 1994). Pure Appl Chem. 1994; 66:609-630.

[22] Lu P, Chen S, Zheng Y. Artificial intelligence in civil engineering. Math Probl Eng. 2012; 1-22.

[23] Hibbert, D.B. Vocabulary of concepts and terms in chemometrics (IUPAC Recommendations 2016). Pure App Chem. 2016; 88:407-443.

[24] Yu P, Low M.Y, Zhou W. Design of experiments and regression modelling in flood flavaour and sensory analysis: a review. Trends Anal Chem. 2018; 71:202–215.

[25] Ropoli A.I, Panagou E.Z, Nychas J.E. Data mining derived from food analyses using non-invasive/non-destructive analytical techniques; determination of food authenticity, quality & safety in tandem with computer science disciplines. Trends Anal Chem. 2016; 50:11–25.

[26] Marini, F. Artificial neural networks in foodstuff analyses: trends and perspectives. A review. Anal Chim Acta 2009; 635:121–131.

[27] Duffus J.H, Nordberg M, Templenton D.M. Glossary of terms used in toxicology (IUPAC Recommendations 2007). Pure Appl Chem., 2007; 79:1153-1344.

[28] Mikut R, Resichl M. Data mining tools. Wires Data Min Knowl. 2011; 1:431-443.

[29] Han J, Kamber M, Pei J. Data mining: concepts and techniques. 3rd ed. Waltham: Morgan Kauffmann Publishers; 2012.

[30] Zheng W, Fu X, Ying Y. Spectroscopy-based food classification with extreme learning machine. Chemometr Intell Lab Syst. 2014; 139:42-47.

[31] Mitra S, Acharya T. Data mining: multimedia, soft computing and bioinformatics. New Jersey: John Wiley & Sons; 2003.

[32] Glossary of common machine learning, statistics and data science terms. https://www.analyticsvidhya.com/glossary-of-common-statistics-and-machine-learning-terms/ (accessed on September 10th).

[33] Kelleher J.D, Namee B.M, D'arcy A. Fundamentals of machine learning for predictive data analysis. Algorithms, worked examples, and case studies. Cambridge (MA): Massachusettes Institute of Technology; 2015.

[34] Ghernaout D, Aichouni M, Alghamin A. Overlapping ISO/IEC 17025:2017 into big data: a review and perpectives. Int J Sci Qual Anal. 2018; 4:83-92.

[35] Ropodi A.I, Panagou E.Z, Nychas G.J.E. Data mining derived from food analyses using non-invasive/non destructive analytical techniques; determination of food authenticity, quality & safety in tandem with computer science techniques. Trends Food Sci Technol. 2016; 50:11-25.

[36] Dai Q, Sun D.W, Xiong Z, Cheng J.H, Zeng X.A. Recent advances in data mining techniques and their applications in hyperspectral image processing for the food industry. Compr Rev Food Sci Food Saf. 2014; 13:891-905.

[37] Jiang L, Zheng H, Lu H. Applications of UV spectrometry and chemometric models for detecting olive oil-vegetable oil blends adulteration. J Food Sci Technol. 2015; 52:479-485.

[38] Cubero-Leon E, Peñalver R, Maquet A. Review on metabolomics for food authentication. Food Res Int. 2016; 60:95-107.

[39] Popescu R, Costinel D, Dinca O.R, Marinescu A, Stefanescu I, Ionete R.E. Discrimination of vegetable oils using NMR spectroscopy and chemometrics. Food Control. 2015; 48:84-90.

[40] Pisano P.L, Silva M.F, Olivieri A.C. Anthocyanins as market for the classification of Argentinean wines according to botanical and geographical origin. Chemometric modeling of liquid chromatography-mass spectrometry data. Food Chem. 2015; 175:174-180.

[41] Rodionova O.Y, Titova A.V, Pomerantesev A.L. Discriminant analysis in an innaproriate method of authentication. Trends Anal Chem. 2016; 78:17-22.

[42] Vitale R, Marini F, Ruckebush C. SIMCA modelling for ovelapping classes: fixed or optimised decision thershold. Anal Chem. (in press). DOI: 10.1021/acs.analchem.8b01270

[43] Bosque-Sendra J.M, Cuadros-Rodríguez L, Ruiz-Samblás C, De la Mata P. Combining chromatography and chemometrics for the characterization and authentication of fats and oils triacyglyceerol compositional data- A review. Anal Chim Acta 2012; 724:1-11.

[44] Berrueta L.A, Alonso-Salces R.M, Héberger K. Supervised pattern recognition in food analysis. J Chromatogr A. 2007; 1158:196-214.

[45] Oliveri P, Downey P. Multivariate class modeling for the verificaion of food-authenticity claims. Trends Anal Chem. 2012; 35:74-86.

[46] Khakimov B, Gürdeniz G, Engelsen S.B. Trends in the application of chemometrics to foodomics studies. Acta Alimentaria 2015; 44:4-31.

[47] Olivieri A.C. Recent advances in analytical calibration with multi-way data. Anal Methods 2012; 4:1876-1886.

[48] Ortiz M.C, Sarabia L. Quantitative determination in chromatographic analysis based on n-way calibration strategies. J Chromatogr A. 2007; 1158:94-110.

[49] Szymanska E. Modern data science for analytical chemical data. Anal Chim Acta 2018; 1028:1-10.

[50] Brereton R, Lloyd G.R. Support vector machine for classification and regression. Analyst 2010; 135:230-267.

[51] Luts J, Ojeda F, de Plas R.V, de Moor B, Van Huffel S, Suykens J.A.K. A tutorial on support vector machine-based methods for classifications problemas in chemometrics. Anal Chim Acta 2010; 665:129-145.

[52] Xu Y, Zomer S, Brereton R. Support vector machine: a recent method for classification in chemometrics. Crit Rev Anal Chem. 2006; 36:177-188.

[53] Witten I.H, Frank E. Data mining: practical machine learning tools and techniques. 2 nd ed. San Francisco (CA): Morgan Kaufmann Publishers; 2005.

[54] Ruiz-Samblás C, Cadenas J.M, Pelta D.A, Cuadros-Rodríguez L. Application of data mining methods for classification and prediction of olive oil blends with other vegetable oils. Anal Bional Chem. 2014; 406:2591-2601.

[55] Granitto P.M, Gasperi F, Biasioli F, Trainotti E, Furlanello C. Modern data mining tools in descriptive sensory analysis: a case study with a random forest approach. Food Qual Prefer. 2007; 18:681-689.

[56] Smith B.R, Baker M.J, Palmer D.S. PRFFECT: a versatile tool for spectroscopists. Chemometr Intell Lab Syst. 2018; 172:33-42.

[57] Wei Z, Wang J, Wang Y. Classification of monofloral honeys from different floral origins and geographical origins based on rheometer. J Food Eng. 2010; 96:469–479.

[58] Stanimirova I, Üstün B, Cajka T, Riddelova K, Hajslova J, Buydens L.M.C, Walczak B. Tracing the geographical origin of honeys based on volatile compounds profiles assessment using pattern recognition techniques. Food Chem. 2010; 118:171–176.

[59] El Alami El Hassani N, Tahri K, Llobet E, Bouchikhi B, Errachid A, Zine N, El Bari N. Emerging approach for analytical characterization and geographical classification of Moroccan and French honeys by means of a voltammetric electronic tongue. Food Chem. 2018; 243:36–42.

[60] Kemal Aloglu A, de B. Harrington P, Sahin S, Demir C, Gunes M.E. Chemical profiling of floral and chestnut honey using high-performance liquid chromatography-ultraviolet detection. J Food Compost Anal. 2017; 62:205–210.

[61] Popek S, Halagarda M, Jursa K. A new model to identify botanical origin of polish honeys based on the physicochemical parameters and chemometric analysis. LWT-Food Sci Technol. 2017; 77:482–487.

[62] Chudzinska M, Baralkiewicz D. Applicatiopn of ICP-MS method of determination of 15 elements in honey with chemometric approach for the verification of their authenticiy. Food Chem Toxicol. 2011; 49:2741–2749.

[63] Wood J.E, Allaway D, Boult E, Scott I.M. Operationally realistic validation for prediction of cocoa sensory qualities by high-throughput mass spectrometry. Anal Chem. 2010; 82:6048–6055.

[64] Teye E, Huang X. Novel prediction of total fat content in cocoa beans by FT-NIR spectroscopy based on effective spectral selection multivariate regression. Food Anal Methods. 2015; 8:945–953.

[65] Teye E, Huang X, Han F, Botchway F. Discrimination of cocoa beans according to geographical origin by electronic tongue and multivariate algorithms. Food Anal Methods. 2014; 7:360-365.

[66] Dong W, Zhang Y, Zhang B, Wang X. Quantitative analysis of adulteration of extra virgin olive oil using Raman spectroscopy improved by bayesin framework least squares support vector machines. Anal Methods. 2012; 4:2772–2777.

[67] Jiménez-Carvelo A.M, Pérez-Castaño E, González-Casado A, Cuadros-Rodríguez L. One input-class and two input-class classifications for differentiating olive oil from other edible vegetable oils by use of the normal phase liquid chrtomatography fingerprint of the methyl-transesterified fraction. Food Chem. 2017 ;221:1784–1791.

[68] Jiménez-Carvelo A.M, González-Casado A, Pérez-Castaño E, Cuadros-Rodríguez. L. Fast-HPLC fingerprinting to discriminate olive oil from other edible vegetable oils by multivariate classification methods. J AOAC int. 2017 ;100(2):345–350.

[69] Jiménez-Carvelo A.M, Osorio M.T, Koidis A, González-Casado A, Cuadros-Rodríguez L. Chemometric classification and quantification of olive oil in blends with any vegetable oils using FTIR-ATR and Raman. LWT-Food Sci Technol. 2017 ;86:174–184.

[70] Jiménez-Carvelo A.M, González-Casado A, Cuadros-Rodríguez L. A new analytical method for quantification of olive oil and palm oil in blends with other vegetable edible oils based on the chrtomatographic fingerprints from the methyl-transesterified fraction. Talanta. 2017 ;164:540–547.

[71] Nasibov E, Kantarci Savas S, Vahaplar A, Kinay A.O. A survey on geographic classification of virgin olive oil with using T-operators in fuzzy decision tree approach. Chemom Intell Lab Syst. 2016; 155:86–96.

[72] Sayago A, González-Dominguez R, Beltran R, Fernández-Recamales A. Combination of complementary data mining methods for geographical characterization of extra virgin olive oils based on mineral composition. Food Chem. 2018; 261:42–50.

[73] Xu L, Ye Z.H, Cui H.F, Yu X.P, Cai C.B, Yang H.W. Calibrating the shelf-life of chinese flavored dry tofu by FTIR spectroscopy and chemometrics: effects of data preprocessing and nonlinear transformation on multivariate calibration accuracy. Food Anal Methods. 2012; 5:1328–1334.

[74] Dong W, Zhang Y, Zhang B, Wang X. Rapid prediction of fatty acid composition of vegetable oil by Raman spectroscopy coupled with least saures support vector machines. J Raman Spectrosc. 2013; 44:1739–1745.

[75] Ai F.F, Bin J, Zhang Z.M, Huang J.H, Wang J.B, Liang Y.Z, Yu L. Application of random forests to select premium quality vegetable oils by their fatty acid composition. Food Chem. 2014; 143:472–478.

[76] Hu Wei, Zhang L, Li, P, Wang X, Zhang Q, Xu B, Sun X, Ma F, Ding X. Characterization of volatile compounds in four vegetable oils by headspace two-dimensional comprehensive chromatography time-of-flight mass spectrometry. Talanta. 2014; 129:629–635.

[77] Bachion de Santana F, Mazivila S.J, Gontijo L.C, Borges Neto W, Poppi R.J. Rapid discrimination between authentic and adulterated andiroba oil using FTIR-HATR spectroscopy and random forest. Food Anal Methods. 2018; 11(7):1927–1935.

[78] Cheng P, Fan W, Yan Xu. Quality grade discrimination of Chinese strong aroma type liquors using mass spectrometry and multivariate analysis. Food Res Int. 2013; 54:1753–1760.

[79] Papadopoulou O.S, Panagou E.Z, Mohareb F.R, Nychas G.J.E. Sensory and microbiological quality assessment of beef fillets using a portaable electronic nose in tandem with support vector machine analysis. Food Res Int. 2013; 50:241–249.

[80] Arredondo T, Oñate E, Santander R, Tomic G, Silva J.R, Sanchez E, Acevedo C.A. Application of neural networks and meta-learners to recognize beef from OTM cattle by using volatile organic compound. Food Bioprocess Technolo. 2014; 7:3217–3225.

[81] Mohareb F, Papadopoulou O.S, Panagou E, Nychas G.J.E. Esemble-based support vector machine classifiers as an efficient tool for quality assessment of beef fillets from electronic nose data. Anal Methods. 2016; 8:3711–3721.

[82] Barbosa R.M, Lemos Batista, B, Varrique R.M, Coelho V.A, Compiglia A.D, Barbosa Jr, F.The use of advanced chemometric techniques and trace element levels for controlling the authenticity of organic coffee. Food Res Int. 2014; 61:246–251.

[83] Bona E, Marquetti I, Varaschim Link J, Figueiredo Makimori G.Y, de Costa Arca V, Guimaraes Lemes A.L, Garcia Ferreira J.M, dos Santos Scholz M.B, Valderrama P,

Poppi R.J.Support vector machines in tandem with infrared spectroscopy for geographical classification of green arabica coffee. LWT-Food Sci Technol. 2017; 76:330–336.

[84] Liu N, Liang Y, Bin J, Zhang Z, Huasang J, Shu R, Yang K. Classification of green and black teas by PCA and SVM analysis of cyclic voltammetric signals from metallic oxide-modified electrode. Food Anal Methods. 2014; 7:472–480.

[85] Zheng L, Watson D.G, Johnston B.F, Clark R.L, Edrada-Ebel R.L, ELseheri, W. A chemometric study of chromatograms of tea extracts by correlation optimization warping in conjunction with PCA, support vector machines and random forest data modeling. Anal Chim Acta. 2009; 642:257–265.

[86] Wang X, Huang J, Fan W, Lu H. Identification of green tea varieties and fast quantification of total polyphenols by near-infrared spectroscopy and ultraviolet-visible spectroscopy with chemometric algorithms. Anal Methods. 2015; 7:787–792.

[87] Ni K, Wang J, Zhang Q, Yi X, Ma L, Shi Y, Ruan J. Multi-element composition and isotopic signatures for the geographical origin discrimination of green tea in China: a case study of Xihu Longjing. J Food Compost Anal. 2018; 67:104–109.

[88] Gonçalvez Dias Diniz P.H, Ferreira Barbosa M, Tavares de Melo Milanez K.D, Pistonesi M.F, Ugulino de Araújo M.C. Using UV-Vis spectroscopy for simultaneous geographical and varietal classification of tea infusions simulating a home-made tea cup. Food Chem. 2016; 192:374–379.

[89] Li Y, Zhang J.Y, Wang Y.Z. FT-MIR and NIR spectral data fusion: a synergetic strategy for the geographical traceability of Panax notoginseng. Anal Bioanal Chem. 2018; 410:91–103.

[90] Khanmohammadi M, Karami F, Mir-Marqués A, Garmarudi A.B, Garrigues S, de la Guardia M. Classification of persimmon fruit origin by near infrared spectrometry and least squares-support vector machines. J Food Eng. 2014; 142:17–22.

[91] Fan Y, Lai K, Rasco B.A, Huang Y. Determination of carbaryl pesticide in fuji apples using surface-enhanced Raman spectrocopy coupled with multivariate analysis. LWT-Food Sci Technol. 2015; 60:352–357.

[92] Majcher M.A, Kaczmarek A, Pawlik D, Pikul J, Jelén H.H. SPME-MS-based electronic nose as a tool for determination of authenticity of PDO Cheese, Oscypek. Food Anal Methods. 2015; 8:2211–2217.

[93] Fabris A, Biasioli F, Granito P.M, Aprea E, Cappelin L, Schuhfried E, Soukoulis C, Märk T.D, Gasperi F, Endrizzi I. PTR-TOF-MS and data mining methods for rapid characterisation of agro-industrial samples: influence of milk storage conditions on the volatile compounds profile of Trentingrana cheese.J Mass Spectrom. 2010; 45:1065–1074.

[94] Guo Y, Ni Y, Kokot S. Evaluation of chemical components and properties of the jujube fruit using near infrared spectroscopy and chemometrics. Spectrochim Acta A Mol Biomol Spectrosc. 2016; 153:79–86.

[95] da Costa n.L, Castro I.A, Barbosa R. Classification of cabernet sauvignon from two different countries in South America by chemical compounds and support vector machine. App Art Intell. 2016; 30(7):679–689.

[96] Vigneau E, Coureoux P, Symoneaux R, Guérin L, Villière A. Random forest: a machine learning methodology to highlight the volatile organic compounds involved in olfactory perception. Food Qual Prefer. 2018; 68:135–145.

[97] Martelo-Vidal M.J, Dominguez-Agis F, Vázquez M. Ultraviolet/visible/near-infrared spectral analysis and chemometric tools for the discrimination of wines between subzones

inside a controlled designation of origin: a case study of Rías Baixas. Aust J Grape Wine Res. 2013; 19:62–67.

[98] Martelo-Vidal M.J, Vázquez M. Polyphenolic profile of red wines for the discrimination of controlled designation of origin. Food Anal Methods. 2016; 9:332–341.

[99] Maione C, de Paula E.S, Gallimberti M, Batista B.L, Campiglia A.D, Barbosa Jr F, Barbosa R.M. Comparative study of data mining techniques for the authentication of organic grape juice based on ICP-MS analysis. Expert Syst Appl. 2016; 49:60–73.

[100] Ramírez-Morales I, Rivero D, Fernández-Blanco E, Pazos A. Optimization of NIR calibration models for multiple processes in the sugar industry. Chemometr Intell Lab Syst. 2016; 159:45–57.

[101] Barbosa R.M, Batista B.L, Bariao C.V, Varrique R.M, Coelho V.A, Campiglia A.D, Barbosa Jr, F.A simple and practical control of the authenticity of organic sugarcane samples based on the use of machine-learning algorithms and trace elements determination by inductively coupled plasma mass spectrometry. Food Chemistry. 2015; 184:154–159.

[102] Barbosa R.M, de Paula E.S, Paulelli A.C, Moore A.F, Oliveira Souza, J.M, Lemos Batista B, Campiglia A.D, Barbosa Jr, F. Recognition of organic rice samples based on trace elements and support vector machine. J Food Compost Anal. 2016; 45:95–100.

[103] Maione C, Lemos Batista, B, Campiglia A.D, Barbosa Jr F, Barbosa R.M. Classification of geographic origin of rice by data mining and inductively coupled plasma mass spectrometry. Comput Electron Agric. 2016; 121:101–107.

[104] Kyu Lim D, Phuoc Long N, Mo C, Dong Z, Cui L, Kim G, Kwon Won S. Combination of mass spectrometry-based targeted lipidomics and supervised machine learning algorithms in detecting adulterated admixtures of white rice. Food Res Int. 2017; 100:814–821.

[105] Feng X, Zhang Q, Cong P, Zhu Z. Preliminary study on classification of rice and detection of paraffin in the adulterated samples by Raman spectroscopy combined with multivariate analysis. Talanta. 2013; 115:548–555.

[106] Mahdavi V, Farimani M.M, Fathi F, Chassempour A. A targeted metabolomics approach toward undestanding metabolic variations in rice under pestice stree. Anal Biochem. 2015; 478:65–72.

[107] Weng S, Wang F, Dong R, Qiu M, Shao J, Huang L, Zhang D. Fast and quantitative analysis of ediphenpehos residue in ruce using surface-enhacenced raman spectroscopy. J Food Scie. 2018; 83(4):1179–1185.

[108] Li J.L, Sun D.W, Pu H, Jayas D.S. Determination of trace thiophanate-methyl and its metabolite carbendazim with teratogenic risk in red bell peper (*Capsicumannuum L.*) by surface-enhanced Raman imaging technique. Food Chem. 2017; 218:543–552.

[109] Hong X, Wang J. Detection of adulteration in cherry tomato juices based on electronic nose and tongue: comparasion of different data fusion approaches. J Food Eng. 2014;126:89–97.

[110] Qiu S, Wang J. The prediction of food additives in the fruit juice based on electronic nose with chemometrics. Food Chem. 2017; 230:208–214.

COMUNICACIONES A CONGRESOS

- A.M. Jiménez-Carvelo, L. Valverde-Som, C.Ruiz-Samblás, E. Pérez-Castaño, A. González-Casado, L. Cuadros-Rodríguez. **The use of a "dummy" class when performing (pseudo) two class multivariate classifications**. XVI Chemometrics in Analytical Chemistry. **Póster.**

- A. González-Casado, A.M. Jiménez-Carvelo, L. Cuadros-Rodríguez. **'Analytical-Chemometric' methodology for authentification of olive oil.** 9th Colloquium Chemiometricum Mediterraneum. **Oral.**

PROBLEMA

El aceite de oliva es un producto altamente apreciado por los consumidores y reconocido por sus propiedades saludables. En la industria alimentaria es común incorporar aceite de oliva como ingrediente de alimentos diversos con objeto de dotar a los productos que lo contienen de una mayor calidad. Como se ha comentado anteriormente en España se prohíbe la comercialización de mezclas de aceite de oliva con otros aceites vegetales y además si un alimento lo contiene como ingrediente es obligatorio indicar en la etiqueta la proporción del mismo.

Por estos motivos existe un gran interés en desarrollar métodos rápidos, sencillos y de aplicación general para controlar la presencia de aceite de oliva y poder detectar, en su caso, fraudes o adulteraciones del mismo. La respuesta técnica a todas estas situaciones se puede resumir en una única operación: autentificar el aceite de oliva.

Para ello, se van a desarrollar métodos analíticos mediante técnicas separativas como la cromatografía de líquidos en sus dos modalidades de trabajo y cromatografía de gases, acoplada a diferentes detectores y técnicas espectroscópicas como FTIR y Raman.

HIPÓTESIS

El uso de la fracción metil-transesterificada ha sido escaso en el ámbito analítico y no se conoce suficientemente la composición de dicha fracción, por lo que resulta aconsejable llevar a cabo una caracterización analítica de la misma mediante un estudio exhaustivo con diferentes técnicas cromatográficas acopladas a diversos detectores, siendo el basado en el uso de la espectrometría de masas en sus diversas modalidades, el que más potencialidad presente para este fin.

Además, la fracción transesterificada no ha sido hasta la fecha utilizada con fines de autentificación. Dada la complejidad química de dicha fracción, se recurre a la metodología de "huellas dactilares" y al empleo de métodos quimiométricas de tratamiento de datos para desarrollas modelos, a partir de los datos registrados diferentes equipos instrumentales, que permitan resolver los problemas propuesto en relación con la autenticidad en el aceite de oliva.

Cada especie vegetal presenta una huella dactilar instrumental (cromatográfica/espectroscópica) característica que depende de su composición química particular. La fracción transesterificada encierra la información necesaria para diferenciar los aceites vegetales, aunque ésta no es evidente. Por tanto, a partir del estudio de la huella dactilar cromatográfica/espectroscópica de la fracción transesterificada es posible crear modelos para diferenciar aceites vegetales, identificar la presencia de aceite de oliva y detectar fraudes alimentarios, y cuantificar la cantidad de dicho aceite en alimentos.

OBJETIVOS

La tesis doctoral se centra en la aplicación de la huella dactilar cromatográfica de la fracción metil-transesterificada del aceite de oliva para resolver problemas de autentificación del mismo.

Los objetivos generales de la tesis son:

- ⇨ Desarrollar métodos analíticos cromatográficos que sean rápidos aplicando la metodología de huellas dactilares para: (i) clasificar y diferenciar el origen botánico de los diferentes aceites vegetales comestibles; (ii) identificar y discriminar la presencia de otros aceites vegetales mezclados con aceite de oliva; y (iii) cuantificar la cantidad de aceite de oliva en mezclas con otros aceites vegetales o en alimentos que lo incorporen como ingrediente.

- ⇨ Caracterizar químicamente los distintos componentes que constituyen la fracción metil-transesterificada del aceite de oliva mediante el uso de la cromatografía de gases acoplada a diferentes espectrómetros de masas.

Los objetivos específicos se presentan al inicio de cada capítulo donde se explican los estudios llevados a cabo.

Capítulo I

"Solo puedes analizar los datos que tienes. Sé estratégico sobre qué reunir y cómo almacenarlo"

Marie Curie

CAPITULO I

'Huellas dactilares cromatográficas – fase normal'

I.1. Presentación

Este capítulo recoge los resultados obtenidos al aplicar las huellas dactilares cromatográficas de la fracción metil-transesterificada de los aceites vegetales obtenidas por cromatografía líquida en la modalidad de fase normal, y los métodos de clasificación multivariante desarrollados.

Los objetivos de este capítulo fueron:

- ⇨ Discriminar aceites de oliva de diferentes categorías de otros aceites vegetales comestibles.
- ⇨ Clasificar aceites de oliva virgen extra en función de su origen botánico.
- ⇨ Cuantificar la proporción de aceite de oliva y aceite de palma en mezclas con otros aceites vegetales.

Se aplicaron las tres estrategias de clasificación con diversos algoritmos adecuados a cada una de ellas, para el desarrollo de los modelos de clasificación con el objetivo de discriminar los aceites de oliva de diferentes categorías de otros aceites vegetales.

Este capítulo derivó en tres artículos publicados, cuyas referencias son:

1. *One input-class and two input-class classifications for differentiating olive oil from other edible vegetable oils by use of the normal-phase liquid chromatography fingerprint of the methyl-transesterified fraction* (Food Chemistry, 2017, *221,* 1784-1791).

2. *A new analytical method for quantification of olive and palm oil in blends with other vegetable edible oils based on the chromatographic fingerprints from the methyl-transesterified fraction* (Talanta, 2017, *164,* 540-547).

3. *Classification of olive oil according to their cultivars based on second-order data using LC-DAD* (*enviado en julio 2018 a Talanta*)

Seguidamente se realiza una breve introducción correspondiente a este capítulo.

I.2. Introducción

La cromatografía líquida de alta eficiencia o de altas prestaciones (HPLC) es una técnica separativa en la que los analitos que componen la fracción de interés de la muestra son arrastrados por flujo de una fase móvil líquida, a través de una fase estacionaria incluida en una columna, hasta llegar al detector. En función del material que conforma la fase estacionaria se seleccionan los disolventes que constituyen la fase móvil, lo que determina la modalidad de trabajo de la cromatografía líquida. Si se trabaja con una fase estacionaria polar, se usa una fase móvil apolar y se trabaja en la modalidad clásica de la cromatografía líquida que es la de fase normal. Por el contrario, para una fase estacionaria apolar se usa una fase móvil polar y la modalidad de cromatografía líquida es la de fase reversa o invertida. La fase normal, aun siendo la cronológicamente primera con la que se trabajó en cromatografía líquida, es la menos habitual.

Los estudios realizados en este capítulo se han llevado a cabo en la modalidad cromatográfica de fase normal. Esta modalidad de trabajo se caracteriza por separar los compuestos en base a su polaridad, dónde la fase estacionara de la columna presenta alta polaridad y las interacciones con los compuestos son específicas del grupo activo. Las fases estacionarias en esta modalidad están compuestas por siloxanos enlazados a grupos funcionales polares:

- ciano ($–(CH_2)_n–CN$)
- diol ($–(CH_2)_n–O–CH(OH)–CH_2OH$)
- amino ($–(CH_2)_n–NH_2$)
- dimetilamino ($–CH_2)_n–NH–(CH_2)_3$)

La fase móvil está constituida por disolventes no polares y se suele trabajar con éter etílico, cloroformo, diclorometano o n-hexano. Por lo tanto, el orden de elución de los compuestos va desde el menos polar que eluye primero al más polar que eluye más tarde.

La aplicación de ésta modalidad de cromatografía para el análisis de grasas y/o aceites ha sido reducida y son pocos los estudios publicados que pueden ser encontrados en bibliografía en lo últimos diez años. Se pueden encontrar algunos trabajos donde se ha

utilizado ésta modalidad para determinar triglicéridos en grasas lácteas [**1**] o para estudiar los productos de oxidación en aceites vegetales [**2**,**3**].

Habitualmente a un cromatógrafo de líquidos se le acoplan diferentes sistemas de medida, genéricamente denominados detectores. Los más habituales en el ámbito del análisis de grasas son:

- ⇨ Detector de absorción molecular ultravioleta-visible (UV-Vis). Devuelve el valor de absorbancia de los compuestos de una muestra a una longitud de onda determinada.
- ⇨ Detector de absorción molecular ultravioleta-visible de fila de diodos (DAD). Detecta los compuestos por su absorbancia también, pero difiere del UV-Vis en que puede hacer barridos en un rango de longitudes de onda y mostrar el espectro de absorbancia para cada tiempo de retención. Por consiguiente se pueden obtener datos de orden 2 (tres vías).
- ⇨ Detector de fluorescencia molecular (FLD): detecta los compuestos que emiten fluorescencia cuando son excitados a una determinada longitud de onda. Permite obtener espectros de excitación-emisión por lo que también es apto para adquirir datos de orden 3 (cuatro vías).
- ⇨ Detector de índice de refracción (RID): los compuestos son detectados por la refracción que provocan de la luz en una celda con respecto a un blanco. Su principal inconveniente es que no permite la elución en gradiente.
- ⇨ Detector de dispersión de luz en vapor (ELSD): el eluato procedente de la columna cromatográfica se mezcla con un corriente de gas y es nebulizado, la detección se produce cuando las gotas pequeñas de analito son atravesadas por un haz de luz y la dispersión de la luz es medida en un fotomultiplicador.
- ⇨ Espectrómetro de masas (MS): Puede hacer barridos en un rango de relaciones masa/carga (m/z) y mostrar el espectro de masas para cada tiempo de retención. Por consiguiente se pueden obtener datos de orden 2 (tres vías). Sirve para identificar compuestos desconocidos y permite elucidar la estructura química de las moléculas.

---

[1] Kalo, P., Kemppinen, A., Ollilainen, V. (2009). Determination of triagyglycerols in butterfat by normal-phase HPLC and electrospray-tandem mass spectrometry. Lipids, *44,* 169-195.

[2] Morales, A., Marmesat, S., Dobarganes, M.C., Márquez-Ruiz, G., Velasco, J. (2012). Evaporative light scattering detector in normal-phase high-performance liquid chromatography determination of FAME oxidation products. Journal of Chromatography A, *1254,* 62-70.

[3] Morales, A., Marmesat, S., Ruiz-Mendez, M.V., Márquez-Ruiz, G., Velasco, J. (2014). Formation of oxidation products in edible vegetable oils analyzed as FAME derivatives by HPLC-UV-ELSD. Food Research International, *62,* 1080-1086.

Los detectores acoplados más frecuentemente utilizados en estudios dirigidos a autentificar aceite de oliva han sido UV-Vis, FLD y MS. Algunos de los estudios publicados tienen por objeto: (i) reconocer su origen geográfico [**4**,**5**,**6**], (ii) diferenciar entre variedades de aceituna [**7**], y (iii) detectar adulteraciones del mismo con otros aceites vegetales y/o cuantificar la proporción de aceite de oliva en mezclas [**8**]. Además, como se ha comentado anteriormente, en todos los artículos citados en este párrafo, se ha aplicado la modalidad cromatográfica de fase reversa, no encontrándose hasta la fecha trabajos donde se aplique la modalidad de fase normal.

Además, de acoplar estos detectores a un cromatógrafo de líquidos, también se han llevado a cabo estudios de autenticidad de aceite oliva midiendo directamente con espectrofotómetros de absorción molecular UV-Vis y espectrofluorímetros [**9**,**10**,**11**,**12**].

Sin embargo, en esta tesis se ha utilizado otro detector bastante menos usual denominado detector de aerosol cargado ('charged aerosol detector', CAD), en el cual se genera un aerosol de partículas cargadas que son detectadas en un electrómetro y cuya

[4] Gil-Solsona, R., Raro, M., Sales, C., Lacalle, L., Diaz, R., Ibañez, M., Beltran, J., Sancho J.V., Hernández, F.J. (2016). Metabolomic approach for extra virgin olive oil origin discrimination making use of ultra-high performance liquid chromatography-quadrupole time of flight mass spectrometry. Food Control, *70 ,* 350-359.

[5] Bajoud, A., Carrasco-Pancorbo, A., Ajal, E.A., Ouazzani, N., Fernández-Gutierrez, A. (2015). Potential of LC-MS phenolic profiling combined with multivariate analysis as an approach for the determination of geographical origin of north Morrocan virgin olive oils. Food Chemistry, *166,* 292-300.

[6] Mohamed, M.B., Rocchetti, G., Montesano, D., Ali, S.B., Guasmi, F., Grati Kamoun, N., Lucini, L. (2018). Discrimination of Tunisian and Italian extra-virgin olive oils according to their phenolic and sterolic fingerprints. Food Research International, *106,* 920-927.

[7] Bajoud, A., Medina-Rodríguez, S., Gómez-Romero, M., Ajal, E.A., Bagur-González, M.G., Fernández-Gutierrez, A., Carrasco-Pancorbo, A. (2017). Assessing the varietal origin of extra-virgin olive oil using liquid chromatography fingerprints of phenolic compound, data fusion and chemometrics. Food Chemistry, *215 ,* 245-255.

[8] Carranco, N., Farrés-Cebrián, M., Saurina, J., Núñez, O. (2018). Authentication and quantitation of fraud in extra virgin olive oils based on HPLC-UV fingerprinting and multivariate calibration. Foods, *44 (7),* 1-15.

[9] Durán Merás, I., Domínguez Manzano, J., Airado Rodríguez, D., Muñoz de la Peña, A. (2018). Detection and quantification of extra virgin olive oil adulteration by means of autofluorescence excitation-emision profiles combined with multi-way classification. Talanta, *178,* 751-762.

[10] Aroca-Santos, R., Cancilla, J.C., Pérez-Pérez, A., Moral , A., Torrecilla, J.A. (2016). Quantifying binary and ternary mixture of monovarietal extra virgin olive oils with UV-vis absorption and chemometrics. Sensors and Actuators B: Chemical, *234,* 115-121.

[11] Melo Milanez, K.D.R., Araújo Nóbrega, C., Silva Nascimiento, D., Insausti, M., Ferández Band, B.S., Coelho Pontes, M.J. (2017). Multivariate modeling for detecting adulteration ox extra virgin olive oil with soybean oil using fluorescence and UV-Vis spectroscopies: A preliminary approach. LWT- Food Science and Technology, *85,* 9-15.

[12] Lia, F., Castellano, M., Zammi-Mangion, M. Farrugia, C. (2018). Application of fluorescence spectroscopy and chemometrics models for the detection of vegetable oil adulterants in Matese virgin olive oils. Journal of Food Science and Technology, (in press, https://doi.org/10.1007/s13197-018-3131-0).

carga total depende de la composición del eluato. En la figura 1 se muestra el diagrama de flujo del detector CAD.

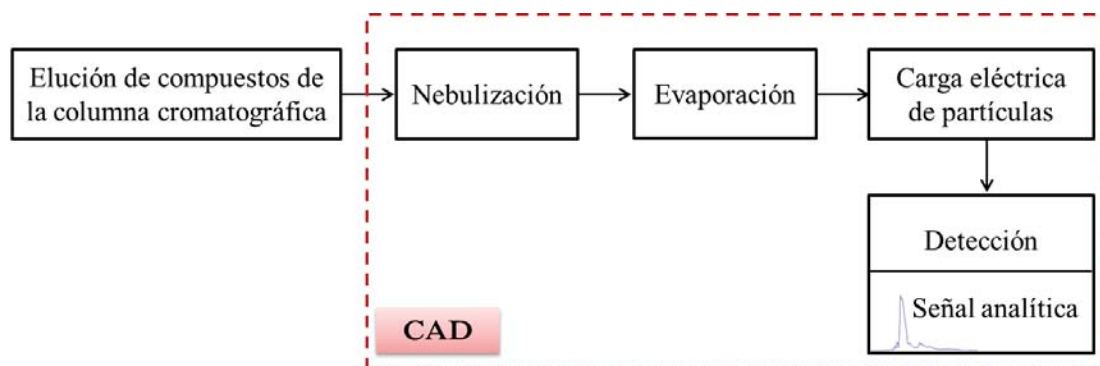


Figura 1. Diagrama de flujo del detector de aerosol cargado (CAD).

Seguidamente se describe de forma resumida el funcionamiento del CAD [**13**]:

1. La muestra procedente de la separación cromatográfica entra directamente a la cámara de nebulización, donde tiene lugar la formación de un aerosol de gotitas utilizando para ello un gas (N_2). Posteriormente, las gotitas de mayor tamaño son eliminadas por un conducto de drenaje y las de menor tamaño son "secadas" evaporando completamente el disolvente en un conducto de secado.

2. A las partículas procedentes del conducto de secado se les transfiere una carga positiva, previamente a su detección, utilizando para ello una fuente de iones de nitrógeno formados a través de una descarga en corona.

3. Las partículas cargadas son recogidas en un colector y posteriormente mediante un electrómetro se mide la cantidad de cargas.

4. La información es enviada al equipo informantico donde se observa la señal analítica de interés, la cual es directamente proporcional a la cantidad de analito en la muestra.

En la figura 2 se muestra a la izquierda un esquema más detallado de cada una de las etapas que tienen lugar dentro del detector y a la derecha el detector CAD utilizado en esta tesis.

[13] Gamache, P.H, Kaufman, S.L. (2017). Principles of charged aerosol detection. En: Gamache, P.H (ed). Charged aerosol detection for liquid chromatography and related separation techniques. Editorial: John Wiley & Sons, Inc.

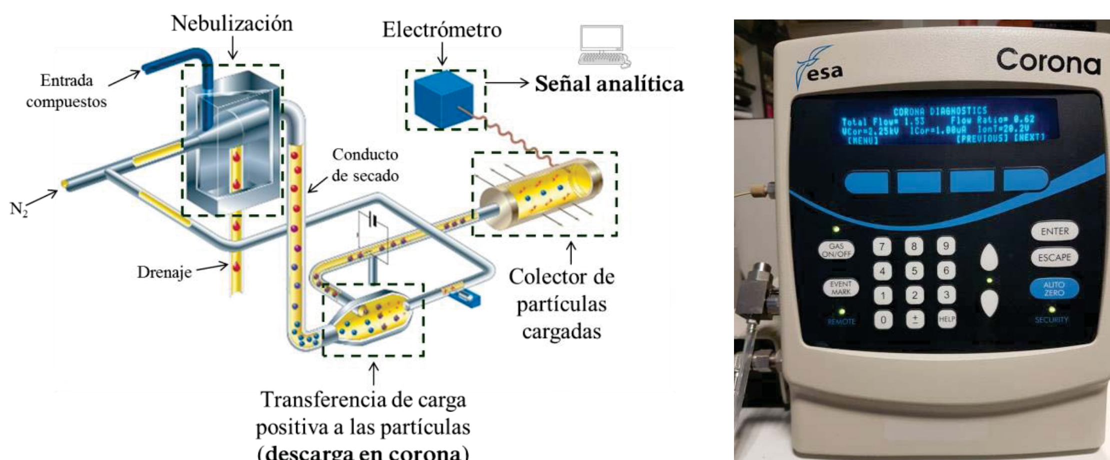


Figura 2. Esquema de funcionamiento del detector de aerosol cargado (CAD) (izquierda) [**14**] y detector CAD utilizado en esta tesis (derecha).

Las aplicaciones de este detector se centran principalmente en el ámbito farmacéutico: análisis de medicamentos, proteínas, e impurezas, entre otras [**15**,**16**,**17**,**18**]. Aunque también se pueden encontrar trabajos de su utilización para el análisis de lípidos, triglicéridos y ácidos grasos [**19**,**20**,**21**,**22**].

En el ámbito de la autenticidad de aceite de oliva los trabajos realizados utilizando este detector han sido los publicados por de la Mata *et al.*, cuyos estudios realizados en este grupo de investigación se centran en la utilización de la huella dactilar cromatográfica de los triglicéridos del aceite de oliva con el objetivo de: (i) discriminar entre aceite de

[14] P/N 70-6258, rev G, Corona CAD Detector. Operatig & Maintenance Manual. (2008). ESA Biosciences, Inc.

[15] Vehovec, T., Obreza, A. (2010). Review of operating principle and applications of the charged aerosol detector. Journal of Chromatography A, *1217,* 1549-1556.

[16] Almeling, S. Ilko, D., Holzgrabe, U. (2012). Charged aerosol detection in pharmaceutical analysis. Journal of Pharmaceutical and Biomedical Analysis, *69*, 50-63.

[17] Wang, R., Wang, X., Paulino, J., Alquier, L. (2013). Evaluation of charged aerosol detector for purity assessement of protein. Journal of Chromatography A, *1283,* 116-121.

[18] Wahl, O., Holzgrabe, U. (2015). Impurity profiling of ibandronate sodium by HPLC-CAD. Journal of Pharmaceutical and Biomedical Analysis, *114,* 254-264.

[19] Lísa, M., Fréderic, L., Holcapek, M., Sandra, P. (2007). Quantitation of triacyglycerols from plant oils using charged aerosol detection with gradient compensation. Journal of Chromatography A, *1176,* 135-142.

[20] Libong, D., Héron, S., Tchapla, A., Chamide, P. (2017). Lipid Analysis with the corona CAD. En: Gamache, P.H (ed). Charged aerosol detection for liquid chromatography and related separation techniques. Editorial: John Wiley & Sons, Inc.

[21] Ilko, D., Braun, A., Germershaus, O., Meinel, L., Holzgrabe, U. (2015). Fatty acid composition analysis in polysorbate 80 with high performance liquid chrtomatography coupled to charged aerosol detection. European Journal of Pharmaceutics and Biopharmaceutics, *94,* 569-574.

[22] Plante, M., Bailey, B., Acworth, I. (2009). The use of charged aerosol detection with HPLC for the measurement of lipids. En: Armstrong, D (ed). Lipidomics, Volume 1: Methods and protocols. Editorial: Humana Press.

oliva y otros aceites vegetales, (ii) cuantificar los triglicéridos presentes en el aceite de oliva, y (iii) cuantificar la proporción de aceite de oliva en mezclas binarias con otros aceites vegetales [**23**,**24**,**25**]. Aunque recientemente ha sido publicado un trabajo donde se utiliza este detector para el análisis de los triglicéridos del aceite de oliva [**26**].

A continuación se especifica la instrumentación analítica, las muestras, así como el pre-tratamiento de los datos y los métodos de clasificación utilizados en este capítulo.

I.3. Instrumentación

Para acometer los dos primeros objetivos específicos citados en la presentación de este capítulo se utilizó un cromatógrafo de líquidos Agilent 1100 provisto de una bomba cuaternaria, automuestrador y termostato de columna, acoplado a un detector de aerosol cargado en corona (figura 2, HPLC-CAD).

Para llevar a cabo el último objetivo específico se utilizó otro cromatografo de líquidos también un equipo Agilent 1100 de las mismas características que el anterior, pero acoplado a un detector de fila de diodos (HPLC-DAD).

En todos los estudios realizados se trabajó en la modalidad de fase normal, utilizando una columna LiChrospher 100 CN de 5 μm de tamaño de partícula y de longitud 250 mm y 4 mm de diámetro interno, como fase móvil se utilizó una mezcla de: n-hexano: isopropanol.

I.4. Muestras

Las muestras de aceites vegetales utilizadas en todos los estudios recogidos en esta tesis fueron adquiridas directamente en comercios de alimentación. Previo a su análisis cromatográfico se llevó a cabo la reacción de transesterificación metílica descrita por Bierderman et *al.* [**27**], la cual se modificó ligeramente con objeto de disminuir el tiempo de reacción. Además, en el estudio 3 se llevó a cabo la epoxidación de la

[23] de la Mata-Espinosa, P., Bosque-Sendra, J.M., Bro, R., Cuadros-Rodríguez, L. (2011). Olive oil quantification of edible vegetable oil blends using triacyglycerols chrtomatographic fingerprints and chemometric tools. Talanta, *85,* 177-182.

[24] de la Mata-Espinosa, P., Bosque-Sendra, J.M., Bro, R., Cuadros-Rodríguez, L. (2011). Discriminating olive oil and non-olive oils using HPLC-CAD and chemometrics. Analytical and Bioanalytical Chemistry, *399*, 2083-2092.

[25] de la Mata-Espinosa, P., Bosque-Sendra, J.M., Cuadros-Rodríguez, L. (2011). Quantification of triacyglycerols in olive oils using HPLC-CAD. Food Analytical Methods, *4,* 574-581.

[26] Lucci, P., Moret, S., Buchini, F., Ferlat, G., Conte, L. (2018). Improved analysis of olive oil triacyglycerols by UHPLC-charged aerosol detection. Journal of Food Composition and Analysis, *66,* 230-236.

[27] Biedermann, M., Grob, K., Mariani, C. (1993). Transesterification and on-line LC-GC for determining the sum of free and esterified sterols in edible oils and fats. Fett Wissenschaft Technologie (Fat Science and Technology), *95*, 127–133.

fracción metil-transesterificada con objeto de aumentar las diferencias entre las diferentes especies botánicas de aceite de oliva y poder clasificar acorde a éstas.

En la tabla 1 se detallan todas las muestras de aceites de oliva y de orujo de oliva y en la tabla 2 las muestras de aceites vegetales no oliva utilizadas en los estudios recogidos en este capítulo. También se realizaron mezclas de todos los aceites vegetales no oliva con aceites oliva, obteniendo mezclas de más de ocho aceites vegetales diferentes. La descripción de estas mezclas puede ser encontrada en el artículo "*a new analytical method for quantification of olive and palm oil in blends with other vegetable edible oils based on the chromatographic fingerprints from the methyl-transesterified fraction*" (Talanta, 2017,*164*, 540-547) de este mismo capítulo y en el artículo "*chemometric classification and quantification of olive oil in blends with any edible vegetable oils using FT-IR and Raman spectroscopy*" (LWT - Food Science and Technology, 2017, *86*, 174-184) recogido en el capítulo III.

Tabla 1. Muestras de aceite de oliva y orujo de oliva utilizada en los estudios.

Categoría	*Variedad de aceituna*	*Nº de muestras*	*Estudio**
ACEITES DE OLIVA			
Virgen extra	Arbequina	10	1, 2 y 3
	Picual	15	1, 2 y 3
	Hojiblanca	12	1, 2 y 3
	Cornicabra	8	1, 2 y 3
	Royal	4	1, 2 y 3
	Frantoio	10	1, 2 y 3
	Empeltre	1	1, 2 y 3
	Koroneiki	3	3
	Farga	1	3
	Shikitia	1	3
	Tosca	1	3
	Lucio	2	1, 2 y 3
	Loaime	3	3
	Arbosana	2	3
	Lechin	2	3
	Vidueña	1	3
	Manzanilla	6	1, 2 y 3
	Ocal	1	3
	Oliana	1	3
	Negrete	1	3
	Serrana	1	3
	Verdial	1	3

Tabla 1. *Continuación*.

Categoría	*Variedad de aceituna*	*Nº de muestras*	*Estudio**
ACEITES DE OLIVA			
	Alfafarenca	1	3
	Blanqueta	2	1, 2 y 3
	Picudo	4	1, 2 y 3
	No declarada	15	1 y 2
Virgen	No declarada	4	1 y 2
Oliva	No declarada	6	1y 2
ACEITES DE ORUJO DE OLIVA			
—	No declarada	6	1
	TOTAL DE ACEITES	**125**	

* **Estudio 1**: Discriminación de aceites de oliva de diferentes categorías de otros aceites vegetales comestibles; **Estudio 2:** Cuantificación de la proporción de aceite de oliva y palma en mezclas con otros aceites vegetales; **Estudio 3**: Clasificación de aceite de oliva virgen extra en función de su origen botánico.

Tabla 2. Muestras de no oliva utilizada en los estudios.

Tipo de aceite	*Nº de muestras*	*Estudio**
ACEITES DE NO OLIVA		
Cacahuete	5	1 y 2
Canola	4	1 y 2
Colza	4	1 y 2
Girasol	13	1 y 2
Lino	3	1 y 2
Maíz	5	1 y 2
Palma	7	1 y 2
Semillas	3	1 y 2
Soja	7	1 y 2
Trigo (germen)	1	1 y 2
Uva (pepita)	4	1 y 2
Sésamo	4	1 y 2
TOTAL DE ACEITES	**60**	

* **Estudio 1**: discriminación de aceites de oliva de diferentes categorías de otros aceites vegetales comestibles; **Estudio 2:** Cuantificación de la proporción de aceite de oliva y palma en mezclas con otros aceites vegetales.

I.5. Condiciones experimentales

Se desarrolló un método de análisis cromatográfico en la modalidad de fase normal que fue aplicado en los tres estudios que se recogen en este capítulo. Las condiciones de trabajo fueron:

- Fase móvil: mezcla n-hexano:isopropanol (96:4, v/v)
- Tipo de elución: isocrático
- Temperatura de la columna: 30 ºC
- Flujo: 1.2 mL/min
- Volumen de inyección: 20 µL
- Tiempo de análisis: 8 min

I.6. Pre-procesado de los datos

Para los dos primeros estudios: (i) discriminar aceites de oliva de diferentes categorías de otros aceites vegetales comestibles, y (ii) cuantificar la proporción de aceite de oliva y aceite de palma en mezclas con otros aceites vegetales, se aplicaron las mismas etapas de pre-tratamiento de los datos cuya descripción puede verse en los artículos publicados que se presentan al final de este capítulo. Todos los pre-tratamientos salvo el centrado en la media fueron realizados por una función programada y diseñada específicamente para el grupo de investigación, llamada "Medina" [**28**].

En el tercer estudio: 'clasificación de aceites de oliva virgen extra en función de su origen botánico' las huellas dactilares de la epoxidación de la fracción metil-transesterificada fueron datos de orden 2 adquiridos por detector de fila de diodos. En este caso no se realizó ningún tipo de pre-tratamiento de los datos, se comprobó que efectivamente los picos cromatográficos no se desplazaban en el tiempo y que el número de variables, aunque era elevado se podía operar con él sin necesidad de llevar a cabo una reducción previa de éstas.

I.7. Métodos de clasificación multivariante aplicados

En primer lugar se aplicó un análisis exploratorio (PCA) de los datos con objeto de observar si existían agrupaciones naturales de los objetos/muestras.

---

[28] Pérez-Castaño, E., Ruiz-Samblás, C., Medina-Rodríguez, S., Quirós-Rodríguez, V., Jiménez-Carvelo, A.M., Valverde-Som, L., González-Casado, A., Cuadros-Rodríguez. (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.

Para los modelos de clasificación que se desarrollaron en cada uno de los estudios aquí presentados se seleccionaron los conjuntos de entrenamiento y validación externa. En el conjunto de datos se reservó el 80% de las muestras para entrenar el modelo y el 20% para validar externamente el modelo. Se aplicaron los siguientes métodos quimiométricos: 'partial least squares-discriminant analysis' (PLS-DA), 'k-nearest neighbours' (kNN), 'support vector machine-classification' (SVM-C), 'soft independent modelling of class analogies' (SIMCA), 'radom forest' (RF) y 'multivariate curve resolution- alternating least square' (MCR-ALS).

La calidad de las clasificaciones realizadas por cada uno de los modelos fue evaluado calculando los siguientes parámetros: sensibilidad, especificidad, valor predictivo positivo, valor predictivo negativo, índice de Youden, relación de verosimilitud para resultados positivos, relación de verosimilitud para resultados negativos, valor F, poder discriminante, eficiencia, área bajo la curva, coeficiente de correlación de Matthews, y coeficiente Kappa. Todos estos parámetros se calcularon a partir de los resultados obtenidos sobre el conjunto de validación externa.

Para determinar la proporción de aceite de oliva y aceite de palma en mezclas con otros aceites vegetales se aplicaron los siguientes algoritmos de calibración: 'partial least squares' (PLS) y 'support vector-regression' (SVR).

Para evaluar los modelos se determinaron: el coeficiente de determinación, la media de los errores cuadráticos, la media de los errores absolutos y la mediana de los errores absolutos, sobre la predicción del conjunto de validación externa.

A continuación se presentan los tres artículos publicados, donde se describen de forma más detallada de los estudios realizados.

ARTÍCULO CIENTÍFICO

One input-class and two input-class classifications for differentiating olive oil from other edible vegetable oils by use of the normal-phase liquid chromatography fingerprint of the methy-transesterified fraction

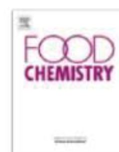
(Artículo publicado en 2017 en la revista: Food Chemistry, 221, 1784-1791)

Food Chemistry 221 (2017) 1784–1791

Contents lists available at ScienceDirect

Food Chemistry

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



Analytical Methods

One input-class and two input-class classifications for differentiating olive oil from other edible vegetable oils by use of the normal-phase liquid chromatography fingerprint of the methyl-transesterified fraction

Ana M. Jiménez-Carvelo *, Estefanía Pérez-Castaño, Antonio González-Casado, Luis Cuadros-Rodríguez

Department of Analytical Chemistry, University of Granada, c/ Fuentenueva, s.n., E-18071 Granada, Spain

One input-class and two input-class classifications for differentiating olive oil from other edible vegetable oils by use of the normal-phase liquid chromatography fingerprint of the methyl-transesterified fraction

Ana M. JIMÉNEZ-CARVELO [a,✉], Estefanía PÉREZ-CASTAÑO[a], Antonio GONZÁLEZ-CASADO [a], Luis CUADROS-RODRÍGUEZ [a]

[a] Department of Analytical Chemistry, University of Granada, c/ Fuentenueva, s.n. E-18071 Granada, Spain.

Abstract

A new method for differentiation of olive oil (independently of the quality category) from other vegetable oils (canola, safflower, corn, peanut, seeds, grapeseed, palm, linseed, sesame and soybean) has been developed. The analytical procedure for chromatographic fingerprinting of the methyl-transesterified fraction of each vegetable oil, using normal-phase liquid chromatography, is described and the chemometric strategies applied and discussed. Some chemometric methods, such as k-nearest neighbours (kNN), partial least squared-discriminant analysis (PLS-DA), support vector machine classification analysis (SVM-C), and soft independent modelling of class analogies (SIMCA), were applied to build classification models. Performance of the classification was evaluated and ranked using several classification quality metrics. The discriminant analysis, based on the use of one input-class, (plus a dummy class) was applied for the first time in this study.

Keywords

Olive oil authentication, methyl-transesterified fraction, chromatographic fingerprinting, one input-class and two input-class classification

✉ Corresponding author: phone: +34 958240797; fax: +34 958243328; email: amariajc@ugr.es

1. Introduction

Edible vegetable oils are important worldwide products, which are used as raw materials and/or ingredients in several foodstuffs. Although most vegetable oils are extracted from oilseeds, some are obtained directly from the fruit as a juice. This is the case for virgin olive oil, which is collected directly from olive fruits by mechanical procedures (grinding followed by centrifugation and/or decantation). Furthermore, in contrast to other vegetable oils, virgin olive oil is not refined for human consumption. Extra virgin olive oil is more expensive than other vegetable oils, owing to the specific process required for extraction [Jabeur, Zribi, Makni, Rebai, Abdelheidi, & Bouaziz, 2014]. For this reason, olive oils are susceptible to adulteration, with cheaper vegetable oils, to achieve an illicit profit. Unauthorized blends or adulteration of olive oil of any quality category with oils obtained from seeds is a particular problem in Spain, as well as other Mediterranean countries, which have specific legislation prohibiting the marketing of such blends. Therefore, it is desirable to develop rapid and simple methods to monitor the authenticity of olive oil.

The analytical methodologies applied to authenticate the olive oil are, generally, based on the quantification of certain chemical markers, which constitute a characteristic fraction of the oils [Arvanitoyannis & Vlachos, 2007; Aparicio, Morales, Aparicio-Ruiz, Tena & García-González, 2013]. Thus, families of compound, such as fatty acids, triacylglycerols (or triglycerides) or sterols, have been proposed. Other chemical fractions, such as volatile compounds or phenols, have also been used but they are not stable enough to give reliable results.

Triglycerides represent 95-99% of the chemical composition of vegetable oils. The compositional characterization of these compounds, determined by gas chromatography (GC) or high-performance liquid chromatography (HPLC), has been proposed for the detection of other oils due to their specific compositional profiles [Aparicio & Aparicio-Ruiz, 2000; Ruiz-Samblás, Marini, Cuadros-Rodríguez, & González-Casado, 2012; Lerma-Garcia, Simó-Alfonso, Méndez, Lliberia & Herrero-Martínez, 2011]. The content of some fatty acids, such as linolenic and oleic acids, has also been used to detect blending in olive oils [Aparicio & Aparicio-Ruiz, 2000].

Fatty acids are quantified using GC, following derivatization to increase the volatility of the compounds, as necessary [Sanchéz de Medina, El Riachy, Priego-Capote & Luque de Castro, 2014; Fernandes, Fernandes, Simas, Barrera-Arellano, Eberlin, & Alberici,

2013]. Moreover, sterols are applied as markers of authenticity in vegetable oils. In order to characterize the compositional profile of these compounds, firstly, it is necessary to carry out saponification of the oil followed by isolation of free sterols by means preparative chromatography or solid phase extraction, and silanization. Then, GC analysis is performed [Gázquez-Evangelista, Pérez-Castaño, Sánchez-Viñas & Bagur-González, 2013]. Consequently, this methodology is difficult, tedious and time-consuming.

In 1993, Bierdermann et al. [Biedermann, Grob & Mariani, 1993] developed a new strategy that replaced the conventional saponification/ isolation process with a methyl-transesterification reaction. This approach, which inexplicably has been underused by the analytical community, requires less vegetable oil and facilitates extraction since soaps are not produced and the process is faster. The breakdown of molecules during transesterification leads to the formation of methyl esters from fatty acids and the liberation of sterols. Two fractions are obtained during this process: (1) the water soluble fraction, which contains the polar compounds, and (2) the organic fraction (transesterified fraction) in which fatty acid methyl esters, sterols, alcohol, monoglycerides, diglycerides and other molecules can be found. In the latter fraction, Bierdermann et al. [1993] identified methyl sterols, dimethyl sterols and linear alcohols. The methyl-transesterified fraction can be analysed by liquid chromatography to obtain a characteristic fingerprint of each vegetable oil, which might also be used to detect potential adulteration. The fingerprinting methodology is based on treating the entire or a part of the chromatogram as a whole, without identifying or quantifying each compound [Ellis et al., 2012; Cuadros-Rodríguez, Ruiz-Samblás, Valverde-Som, Pérez-Castaño, & González-Casado, 2016]. Effective implementation of fingerprinting requires the use of chemometric tools. Chromatograms are exported as data vectors and treated with pattern recognition methods to develop multivariate classification or regression models, which are suitable to differentiate among vegetable oils.

The chemometric methods fall in to two groups: supervised and non-supervised [Naes, Isaksson, Fearn & Davies, 2002; Marini, 2010]. In the first group, the category or class membership of each data vector is known and used to build the multivariate model. In contrast, the model from non-supervised methods does not consider this information [Correira & Ferreira, 2007]. Supervised classification methods are used to categorize objects (samples) in two or more classes according to a set of characteristic features of each class. Such features are extracted previously from information supplied for

standard objects and selected during the model-training step. In order to perform the classification process, two approaches could be applied: discriminant analysis methods and class-modelling methods [Bevilacqua, Nescatelli, Bucci, Magrì, Magrì & Marini, 2014]. A discriminant method works by finding the borders between groups of objects from different classes, while a class-modelling method defines a particular enclosed space for all the objects from the same class.

Sometimes class-modelling methods are described as 'one-class classifier' (e.g. SIMCA) where each class is modelled independently [Brereton, 2011] and as many model as classes are built. Classification is performed considering all the models simultaneously. In our opinion, however, this term should not be used as synonym for class-modelling since the two might be confused.

This study proposes a multivariate method to differentiate olive oil from other edible vegetable oils. For this, the methyl-transesterified fraction from each oil class (olive and non-olive) was analysed using normal-phase conventional high-performance liquid chromatography. The chromatograms (chromatographic fingerprints), acquired by means of a corona charged aerosol detector (CAD), were used as a source of analytical information to set up the classification models. Some common and well-established classification methods were applied, such as k-nearest neighbours (kNN), partial least squares discriminant analysis (PLS-DA), support vector machine classification (SVM-C) and soft independent modelling of class analogies (SIMCA). Two classification strategies were tried for each classification method according to the number of class used for model training: two input-class and one input-class classifications. In addition, the use of a 'dummy' class was proposed for applying discrimination methods with a one input-class strategy. The classification results from each method and strategy were compared and ranked on the basis of several classification performance metrics [Cuadros-Rodríguez, Pérez-Castaño & Ruiz-Samblás, 2016].

2. Materials and methods

2.1. Chemicals

All solvents used were HPLC grade. Isopropanol, n-hexane, methanol and tert-butyl methyl ether (TBME) were provided by the VWR International Eurolab, S.L.

(Barcelona, Spain). Sodium methoxide (MeONa), citric acid monohydrate, and anhydride sodium sulphate were purchased from Merck (Darmstadt, Germany). The nitrogen (99.9999 %) used was provided by Air Liquid (Madrid, Spain).

2.2. Chromatography

The analyses were carried out with an Agilent 1100 series liquid chromatograph (Santa Clara, USA) equipped with a column thermostat (Eppendorf CH30), a quaternary pump and degasser auto sampler. Detection was performed with a corona charged aerosol detector (CAD) (ESA Bioscienses Inc., Chemlsford, MA, USA). Agilent ChemStation software (rev. B.02.01-SR1) for LC systems was used to collect and process data.

The HPLC analysis was carried out on a (250 × 4 mm i.d, 5 μm) column Lichrospher® 100 CN. The column temperature was set at 30 ºC during the entire operation. The composition of the mobile phase was n-hexane/isopropanol (96:4, v/v) at a flow rate of 1.2 mL min^{-1}. The injection volume was 20 µL and the run time was only 8 min.

2.3. Samples

A total of 127 vegetable oil samples of different types were analysed. The samples were obtained directly from local providers. More specifically, 66 samples were different categories of marketed olive oil (virgin extra, virgin, refined+virgin, and pomace+virgin), and the other 61 were canola, safflower, corn, peanut, sunflower, (no-specified) seed, grapeseed, palm, linseed, sesame, and soybean oils. Table 1 summarizes the different vegetable oils and the number of samples analysed for each.

2.4. Sample preparation

Previous to the chromatographic analysis, a transesterification reaction was applied. A modification of the procedure described by Biedermann et al [Biedermann, Grob & Mariani, 1993] was used. For this, 0.1 g of oil was weighed into a centrifuge tube. 1 mL of extracting agent (MeONa at 10 % in methanol in TBME, 4:6 (v/v)) was added and mixed with the oil. The mixture was stirred for 20 s and then allowed to stand for 20 min. This step was repeated twice. Then, 1 mL of water and 8 mL of hexane was added, and the mixture centrifuged for 3 min at 1,500 g. The aqueous phase was removed with

a Pasteur pipette and 1 mL of 1 % citric acid in water added to the residual. Again, the aqueous phase was eliminated before 2 g of anhydrous sodium sulphate added and the mixture allowed to stand for 20 min. The methyl-transesterified organic fraction was passed through a polytetrafluoroethylene (PTFE) membrane syringe filter (0.22 µm) and the solution stored at –20ºC until analysis. For the chromatographic analysis, 200 µL of transesterified solution was added to a 2 mL HPLC vial before 450 µL of n-hexane was added and 20 µL injected.

Table 1. Class and types of vegetable oils analysed.

Class	Category/type	Nº samples
Olive oil (66 samples)	Virgin extra	50
	Virgin	4
	"Refined" [a]	6
	"Pomace" [b]	6
Non-olive oil (61 samples)	Canola	4
	Safflower	4
	Corn	5
	Peanut	5
	Sunflower [c]	13
	Seeds	6
	Grapeseed	4
	Palm	7
	Linseed	3
	Sesame	3
	Soybean	7

[a] A marketed blend of refined and virgin olive oil (5-10 %).

[b] A marketed blend of pomace and virgin olive oil (5-10 %).

[c] Two samples of high-oleic sunflower oils are included.

2.5. Chemometrics

The raw data files from each chromatogram were obtained in a CSV file and exported to MATLAB (version R2013a). In this way, a data vector composed of 839 variables defined each chromatogram. The data pre-processing was done with a home-

programmed MATLAB function, "Medina" (version 10) [Pérez Castaño et al., 2015]. This function implemented several algorithms from the MATLAB Bioinformatics Toolbox™ and 'icoshift' (*interval correlation optimized shifting*) algorithm [Tomasi, Savorani & Engelsen, 2011] to align the peaks of the chromatograms. The steps for pre-processing the data were: (1) raw chromatograms data grouping and overlay; (2) selection of interval of interest in chromatograms; (3) filtered of the raw chromatograms data to eliminate noise of signal analytical; (4) correction of the baseline using the 'msbackadj' function (included in the Bioinformatics Toolbox™); (5) alignment of the peaks with the function 'icoshift'; and finally (6) mean centring of the data set.

The original dataset was divided in two groups: (1) the training set, which was made up of 84 oil samples (44 olive oil, 40 non-olive oil), and (2) the validation (or test) set composed of the remaining oil samples (25 olive oil, 18 non-olive oil). Selection was carried out ensuring that a sample from each class of oil was allocated to one vegetable oil group or the other. Within each group, the samples were selected randomly.

Classification of the vegetable oils was achieved using multivariate chemometric pattern recognition in the PLS_Toolbox (version 7.5.2, Eigenvector Research, Wenatchee, WA).

Principal Component Analysis (PCA)

The main aim of PCA is to reduce the number of variables to evaluate which components contain essential information. Each principal component (PC) is a lineal combination between original variables (chromatographic intensities) of each object, which are described as: $X=T\times P^T$ where X is the original data matrix, T is the score matrix and P is the transposed loading matrix [Bro & Smilde, 2014].

k-Nearest Neighbours (kNN)

kNN is a based-similarity classification method that uses distance measures between objects. The classification is carried out as follows: first, a multidimensional hyperspace is defined with the training set and, then, the prediction is performed. The assigned class of each new object will be one where the number of k-neighbours is largest [Correira & Ferreira, 2007; Alsberg, Goodacre, Rowland & Kell, 1997] and k is an odd integer that

could be selected previously. Each sample is classified based on the most represented classes of the k-nearest samples.

Partial Least Squares Regression-Discriminant Analysis (PLS-DA)

PLS-DA is a latent variable-based method that builds a PLS regression model on latent variables (LV) to establish limits of the class and, then, carries out a discriminant analysis (DA) to classify the samples [Bevilacqua, Nescatelli, Bucci, Magrì, Magrì & Marini, 2014; Ballabio & Consonni, 2013]. In order to develop the best PLS model, it is necessary to optimize the number of LVs to be used in advance.

Support Vector Machine Classification (SVM-C)

SVM is a based-machine learning method. As with PLS-DA, SVM-C works by carrying out a SVM regression model for building hyperplanes in a multidimensional space that separates the different classes of objects [Xu, Zomer, Brereton, 2006; Luts, Ojeda, Van de Plas, De Moor, Huffel & Suykends, 2010]. SVM can be optimized with 'nu' and 'C' parameters. The former optimizes a model with an adjustable parameter Nu [0 → 1], which indicates the upper boundary for the number of misclassifications allowed, and the latter optimizes a model with an adjustable cost function C [0 → ∞], which indicates how strongly misclassifications should be penalized [SVM Function Settings, Eigenvector Documentation wiki. URL http://wiki.eigenvector.com/index.php?title=SVM_Function_Settings. Accessed 29.06.15].

Soft Independent Modelling of Class Analogies (SIMCA)

This chemometric technique performs as many principal component (PC) models as input-classes in study and, then, the classification is carried out from the distance of the object to the centre of each principal component score space [Bevilacqua, Nescatelli, Bucci, Magrì, Magrì & Marini, 2014]. The assignment of each unknown sample to a particular class is based on the nearest distance to the corresponding regions established by the PC model.

Two input-class (2iC) and one input-class (1iC) classification

Usually a two-class classification method (or more properly, two output-class classification) requires using two input-classes, the target class and the non-target class (in this paper, olive and non-olive classes). The term 'output' is related to the classes to which objects or samples will be assigned as result of the classification while the term 'input' refers to the class that is used to train the classification model [Cuadros-Rodríguez, Pérez-Castaño, & Ruiz-Samblás, 2016]. It is also possible to perform the same classification method by training the model with a single input-class, *i.e.* the target class.

Working with one input-class classification has significant advantages. For example, in food authentication, the model can be built with data from only genuine foods (target class) and it is not necessary to have other foods (non-target class) to train the model. Consequently, the necessary experimental work is halved. When this model is applied on unknown foods, only those recognized by the model will be declared as "true" whereas the remaining food will be refused and they are candidate to be considered as "false". The greater the training set of genuine representative samples, the better the quality classification performance. Obviously, this strategy can be applied to differentiate olive oils from other edible vegetable oils.

This is a very easy task when a class-modelling method is applied because each class is modelled independently. This approach has been used already with SIMCA [López, Trullos, Callao & Ruisanchez, 2014]. However, the discriminant methods, such as PLS-DA or SVM-C, usually require two input-classes to define the discrimination model. Although some proposals have been reported as one-class PLS (OCPLS) [Xu, Yan, Cai & Yu, 2013], in fact, this is a class-modelling method. To resolve this drawback, a fictitious class or 'dummy' class could be used as a substitute for the second class (the non-target class). The dummy class should be defined from inactive objects that do not have analytical information of interest for the target class, e.g. analytical blank.

In this study, both 2iC and 1iC strategies were applied to devise a classification model for differentiating olive oil from non-olive oil. When the 1iC was applied, a dummy class was from the dataset provided using 30 chromatograms for the solvent blank.

3. Results and discussion

A chromatogram was recorded for each vegetable oil sample. Figure 1 shows the superposed chromatograms for all vegetable oil samples. Two regions could be easily differentiated: (1) region A shows a major peak, which was essentially composed of methyl esters of fatty chains derived from triglycerides, phospholipids, waxes, esterified sterols and free fatty acids, and (2) region B that was composed of several minor peaks and contained information about the families of free sterols and terpenic alcohols.

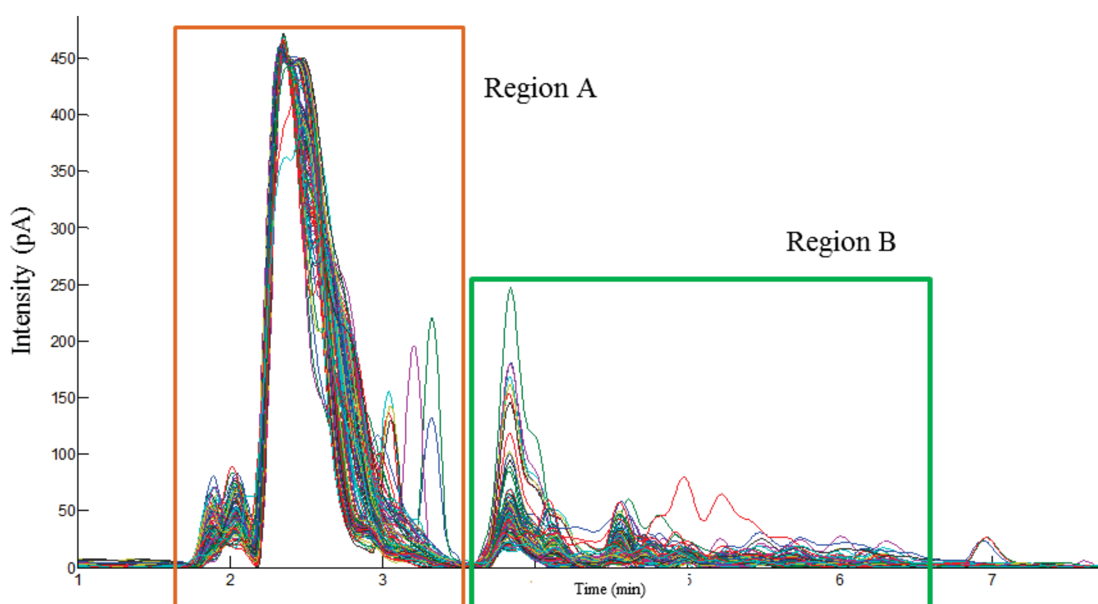


Figure 1. Superposed chromatograms of the 127 vegetable oil samples showing the two characteristic regions (see text for additional explanations). The chromatograms have been previously pre-processed with the exception of the mean centring step.

Exploratory Analysis

A principal component analysis (PCA) was carried out considering the dataset composed of the whole chromatogram from each vegetable oil sample. Four PCs were enough to explain 87.16% of the variance. Figure 2a shows the biplot for scores on the PC2-PC1 plane. PC1 and PC2 explained 56.2% and 17.3% of the variance, respectively. Three groups of vegetable oils could be distinguished easily, which corresponded with olive oil (centre left), palm oil (top left) and other vegetable oils (right).

Two additional PCA were carried out, one for each of the regions of the chromatograms to check if both regions grouped the oil samples in the same way. Figure 2b and 2c show the biplot for scores on the PC2-PC1 plane, corresponding to the data subset from regions A and B, respectively.

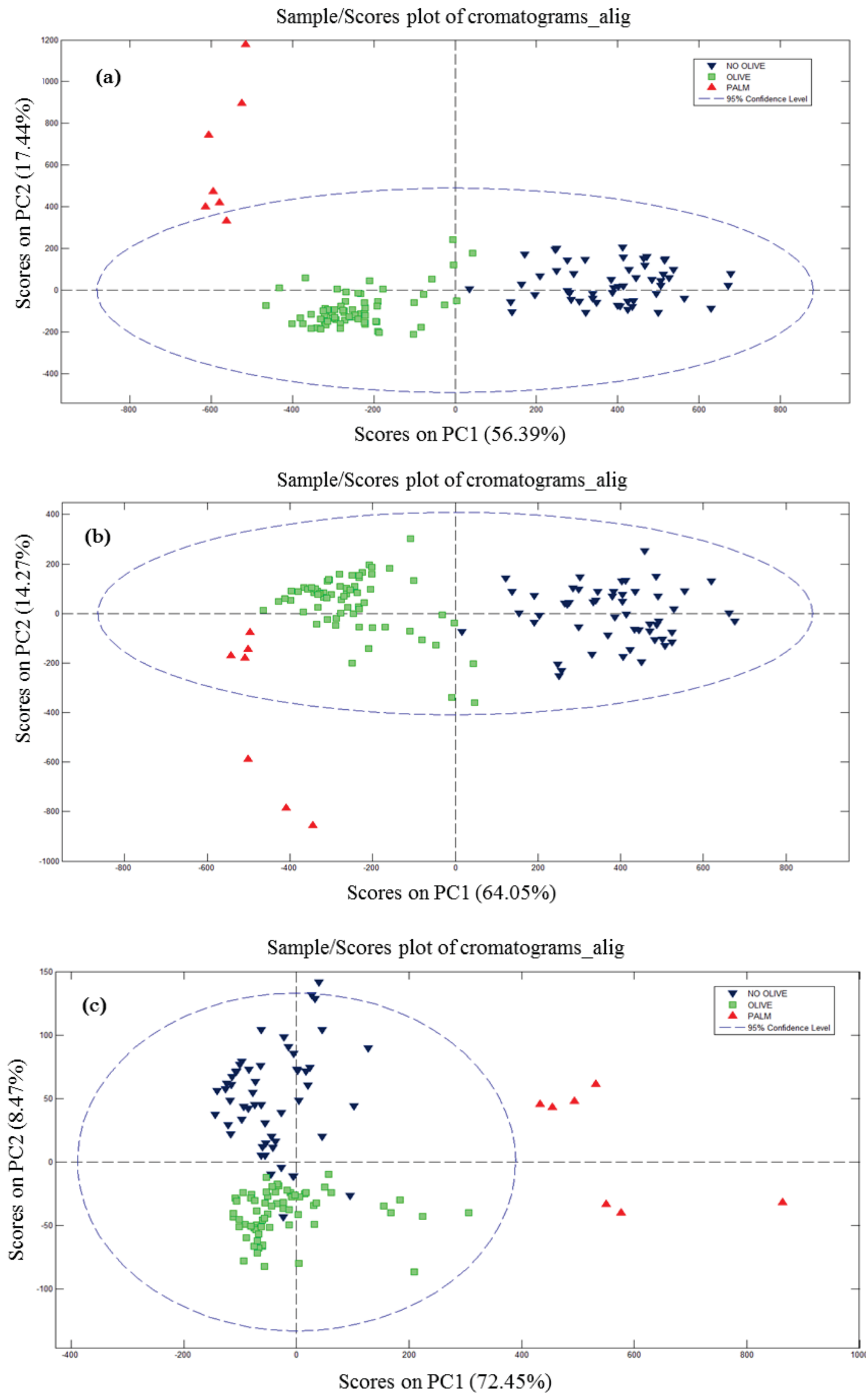


Figure 2. PCA scores biplot obtained from the fingerprint data of the methyl-transesterified fraction of the 127 vegetable oil samples: **(a)** PC2-PC1 plane of the whole chromatogram; **(b)** PC2-PC1 plane from region A; **(c)** PC2-PC1 plane from region B.

The three scores biplots allowed differentiation in similar ways to the three sample groups and, in principle, there was no conclusive reason –from a chemometric point of view– to select one dataset or the others. However, looking the chromatographic retention time, region A was preferred to minimize the analysis time.

Two input-class (2iC) classification

In order to differentiate olive oils from other vegetable oils, a two input-class (2iC) classification strategy was applied where the target class was 'olive oil' and the alternative class was, generally, denoted as 'non-olive oil'. Four well-established classification methods were tried: kNN, PLS-DA, SVM-C and SIMCA.

To differentiate the two vegetable oils classes, k=3 was enough to decide the neighbour distance in the kNN model. The olive class was defined by a class predicted probability value equal to 1, while the non-olive class was defined by a probability of 0. Classification of the samples contained in the validation set was carried out directly by the software. All of olive oil samples were well classified (probability=1) and the non-olive oil samples were also classified correctly (probability=0), with exception of palm oil samples, which had an assigned probability of 0.5; in this case, we also classified these samples as non-olive oil.

The PLS-DA model was built using four LVs, with 92.91% of the variance explained. Each class was characterized by a predicted value around 1 for olive oil and 0 for non-olive oil. The classification threshold established by the software from the corresponding probability curves was a predicted value of 0.6 for the olive oil class.

The SVM-C model was optimized with 'C-svc' and 'nu-svc' parameters, and the results obtained in both cases were similar. As in the kNN method, the olive class was assigned to samples with a predicted probability value equal to 1 and the non-olive class was defined by samples with a probability of 0. The software also carried out the class assignment for the validation samples. Both olive and non-olive oil samples were classified correctly.

Figure 3 (a) and (b) show the classification plots obtained from both 2iC PLS-DA and 2iC SVM-C methods.

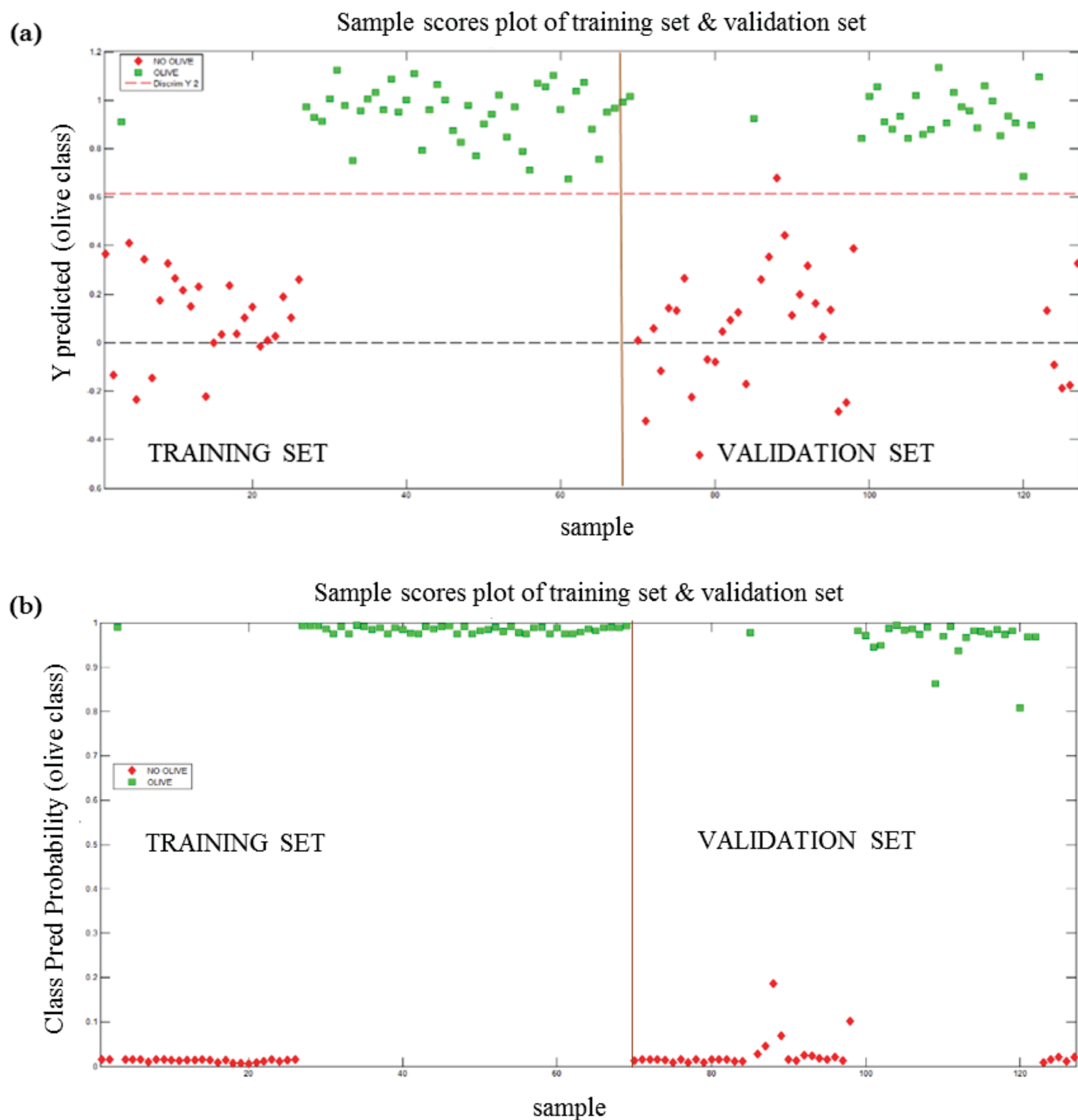


Figure 3. Classification plots on the 2iC classification strategy: **(a)** PLS-DA; **(b)** SVM-C.

The application of SIMCA implies building of two PC models. The number of PCs chosen for each model was four for 'olive oil' and five for 'non-olive oil'. The software carried out classification of the validation dataset based on the Q-residual values for each olive oil sample. Samples with a normalized Q-residual (95% confidence) value less than $\sqrt{2}$ were classified as olive oil.

Table 2 shows the different quality performance features for the 2iC classification method, calculated according to the olive oil samples classification. These show that, in this classification scenario, 2iC kNN and SVM-C were faultless and, in contrast, 2iC SIMCA performed poorly.

Table 2. Values of the quality performance features of the different 2iC classification methods.

Performance features	kNN	PLS-DA	SVM-C	SIMCA
Sensibility (or Recall)	1.00	1.00	1.00	0.48
Specificity	1.00	0.94	1.00	1.00
Positive predictive value (Precision)	1.00	0.96	1.00	1.00
Negative predictive value	1.00	1.00	1.00	0.58
Youden index	1.00	0.94	1.00	0.48
Positive likelihood rate	–	18.00	–	–
Negative likelihood rate	0.00	0.00	0.00	0.52
F-measure	1.00	0.98	1.00	0.65
Discriminant power	–	–	–	–
Efficiency (or Accuracy)	1.00	0.98	1.00	0.70
AUC (Correctly classified rate)	1.00	0.97	1.00	0.74
Matthews correlation coefficient	1.00	0.95	1.00	0.53
Kappa coefficient	1.00	0.95	1.00	0.44

The hyphen "–" is signifying that the performance feature cannot be determined

One input-class (1iC) classification

Since the aim of this study was differentiation of olive oil from other vegetable oils, the classification model could be trained using objects from the olive oil class. In this way, the objects recognized by the model should be assigned as olive oil whereas the remainder, regardless of their botanical origin, should be classified as non-olive oils. The same classification methods, kNN, PLS-DA, SVM-C and SIMCA, were applied. For each, a confidence interval-based classification criterion was established because the default classification threshold defined by the software was not applicable.

The kNN model conformed with k=3, but did not generate good results and all the non-olive oil samples were misclassified because they were considered to be "nearest neighbours" to the target class (olive oil). Thus, the 1iC strategy was not applicable for the kNN method.

Two strategies were applied for 1iC PLS. In a first step, a PLS-DA classification with dummy class was performed using the PLS_Toolbox. Next, a one-class PLS without

dummy class (OCPLS) was performed using software provided by Xu [Xu, Yan, Cai, & Yu, 2013].

A conventional PLS-DA was built with only two LVs explaining 99.74% of the variance. A confidence interval was established centred on 1, which was the value assigned for the olive oil class. The width of the interval was calculated as plus/minus 2.33-times the standard deviation (s) from the predicted values for the olive oil samples in the training set. The expression 2.33×s is an "ad-hoc" application, recommended by the EC for estimating the decision limit (DL), formally termed as CCα, concerning the performance of analytical methods in the case of substances for which no permitted limit has been established [EU Commission Decision, 2002]. This decision limit defines the limit at and above which it can be concluded with an error probability of α that a sample is non-compliant. Strictly speaking, the correct expression would be: DL = $1.645\times\sqrt{2}$ s, where 1.645 is the critical value for the standardized normal distribution (α = 1%) and s the whiting-batch standard distribution of the difference between the predicted values of both the target and the non-target samples, which are considered equal, and consequently: $s = \sqrt{s^2(targ) + s^2(non\text{-}targ)} = \sqrt{2}\,s(targ)$. The coefficient 2.33 is the result of multiplying 1.645 $\times\sqrt{2}$ (or 1.414). The confidence band is calculated from an estimated standard deviation of 0.026.

Figures 4(a) and 4(b) show the classification plots obtained from the 1iC PLS-DA method.

Most olive oil samples were included within the confidence interval while the no-olive oils were not. However, samples in the non-olive oil class were separate into two subclasses on both sides of the interval. The seed oils were located in the upper region whereas the palm oils were in the lower region. This surprising outcome implies the classification scenario is suitable for implementing a three output-class classification (olive oil, palm oil, and generically seed oil) from a one input-class strategy, making it possible to distinguish palm oil from a classification model trained only with olive oils. Currently, the authors are working to develop and apply this approach.

OCPLS was built with seven LVs. For classification purpose, the regions pre-established by the software were used. The results are showed in Table 3.

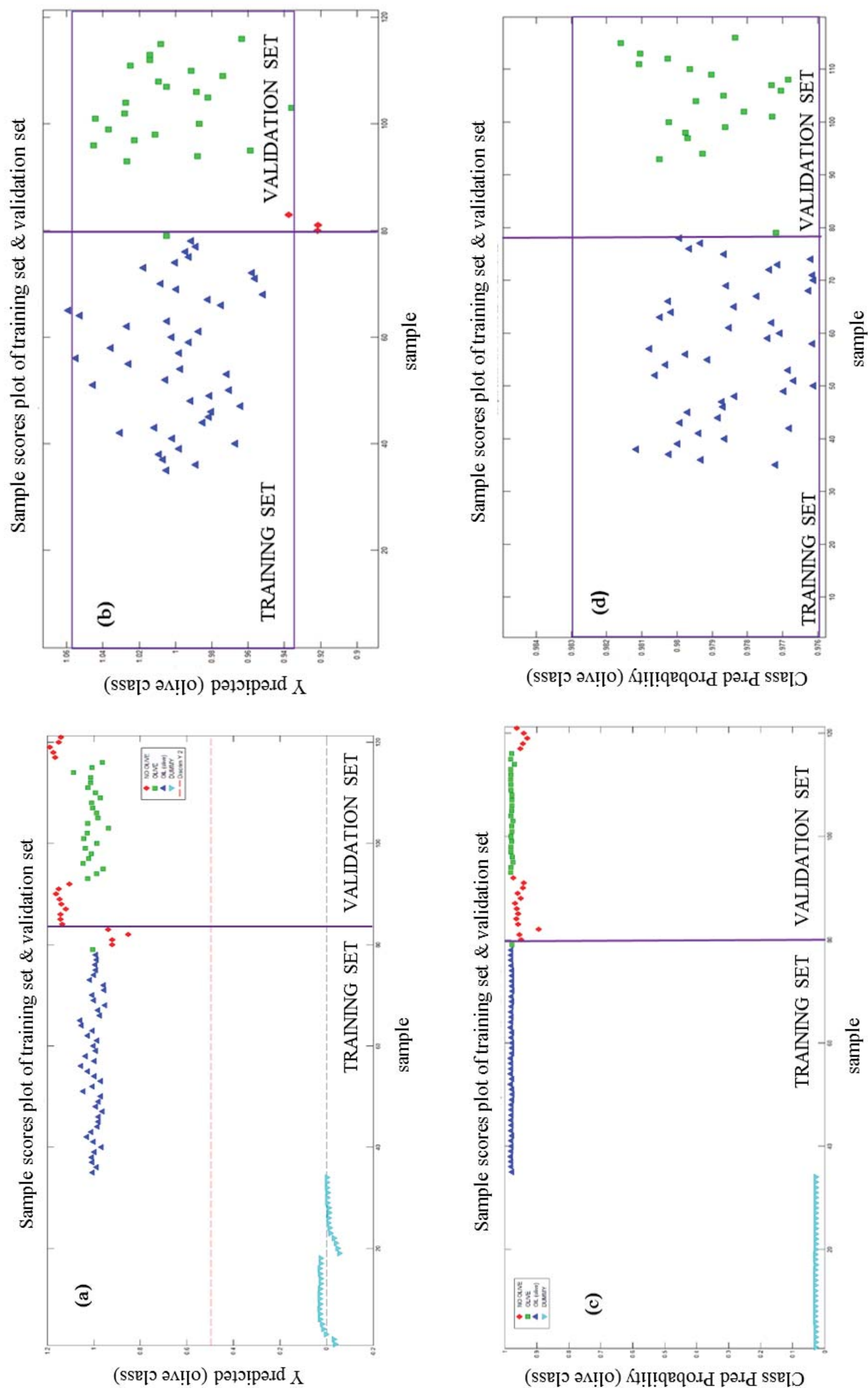


Figure 4. Classification plots on the 1iC classification strategy: **(a)** and **(b)** PLS-DA full plot and zoomed plot, respectively; **(c)** and **(d)** SVM-C full plot and zoomed plot, respectively. In addition, the confidence bands are superposed on **(b)** and **(d)** plots.

SVM-C classification was carried out by optimization of 'C-svc' and 'nu-svc' parameters, and the results obtained in both cases were similar. All the oil samples were assigned to a predicted probability close to 1 and always distant from 0, which was assigned to the dummy class. Specifically, the probability value was ca. 0.98 for the olive oil class and less but always greater than 0.92 for the non-olive class. The confidence interval was determined by means of a probability interval centred on the average olive oil class probability calculated from the training set. The width of the interval was also calculated as plus/minus 2.33 times the standard deviation from the predicted class probability. The estimated value of the probability standard deviation was 0.0015. Figures 4 (c) and (d) show the classification plots obtained from the 1iC SVM-C method.

Finally, the SIMCA method was also applied. Since SIMCA is a class-modelling method, two options were applied: i) a double PCA model using both the olive oil and dummy classes; and ii) a single model from the olive oil class. In both cases, five PCs were used to build the olive oil model. In both cases, a sample oil was classified as olive oil when the normalized Q-residual (95% confidence) value was less than $\sqrt{2}$.

Table 3. Values of the quality performance features of the different 1iC classification methods.

Performance features	With a dummy class				Without dummy class	
	kNN	**PLS-DA**	**SVM-C**	**SIMCA**	**OCPLS**	**SIMCA**
Sensibility (or Recall)	1.00	0.96	0.88	0.88	0.80	0.80
Specificity	0.00	1.00	1.00	0.83	0.89	1.00
Positive predictive value (Precision)	0.58	1.00	1.00	0.88	0.91	1.00
Negative predictive value	–	0.95	0.86	0.83	0.76	0.78
Youden index	0.00	0.96	0.88	0.71	0.69	0.80
Positive likelihood rate	1.00	–	–	5.28	7.20	–
Negative likelihood rate	–	0.04	0.12	0.14	0.23	0.20
F-measure	0.74	0.98	0.94	0.88	0.85	0.89
Discriminant power	–	–	–	0.86	0.83	–
Efficiency (or Accuracy)	0.58	0.98	0.93	0.86	0.84	0.88
AUC (Correctly classified rate)	0.50	0.98	0.94	0.86	0.84	0.90
Matthews correlation coefficient	–	0.95	0.87	0.71	0.68	0.79
Kappa coefficient	0.00	0.95	0.86	0.71	0.67	0.77

The hyphen "–" is signifying that the performance feature cannot be determined

Table 3 shows the quality performance features of the different 1iC classification methods. In contrast with the 2iC classification method, the 1iC PLS-DA provided the best classification performance and 1iC SIMCA (without dummy class) was, again, the worst.

4. Conclusions

In this study, several classification methods were applied and the application strategy has been discussed. Four well-established classification methods were used, namely kNN, PLS-DA, SVM-C and SIMCA. Each was applied using two classification strategies designated as two input-class (2iC) and one input-class (1iC) classifications. This is the first time a dummy class has been used to perform discriminant analysis methods with a single input-class. This new approach does not require having and analysing samples from the non-target class (non-olive vegetable oil) in order to train the classification model. In order to assess and rank the different classification methods and strategies, several quality classification metrics were calculated. kNN and SVM-C, on the one hand, and PLS-DA, on the other, proved to be the best when 2iC or 1iC classification strategies were applied, respectively. Furthermore, the proposed analytical method consumed less time in sample treatment (transesterification reaction, 60 min) and chromatographic elution (8 min) than previous methods (saponification, 120 min) and chromatographic analysis (40 min).

References

Alsberg, B.K., Goodacre, R., Rowland, J.J., & Kell, D.B. (1997). Classification of pyrolysis mass spectra by fuzzy multivariate rule induction-comparison with regression, K-nearest neighbour, neural and decision-tree methods. *Analytica Chimica Acta, 348,* 389–407.

Aparicio, R., & Aparicio Ruíz, R. (2000). Authentication of vegetable oils by chromatographic techniques. *Journal of Chromatography A*, *881*, 93–104.

Aparicio, R., Morales, M.T., Aparicio-Ruiz, R., Tena N., & García-González D.L. (2013). Authenticity of olive oil: Mapping and comparing official methods and promising alternatives. *Food Research International, 54*, 2025–2038.

Arvanitoyannis, I.S., & Vlachos, A. (2007). Implementation of physicochemical and sensory analysis in conjunction with multivariate analysis towards assessing olive oil authentication/adulteration. *Critical Reviews in Food Science and Nutrition, 47*,441–498.

Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: linear models. PLS-DA. *Analytical Methods, 5,* 3790–3798.

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.

Biedermann, M., Grob, K., & Mariani, C. (1993). Transesterification and on-line LC-GC for determining the sum of free and esterified sterols in edible oils and fats. *Fett Wissenschaft Technologie - Fat Science Technology*, *95*(4), 127–133.

Brereton, R.G. (2011). One-class classifiers. *Journal of Chemometrics, 25,* 225–246.

Bro, R., & Smilde, A. (2014). Principal component analysis. *Analytical Methods, 6,* 2812–2831.

Correira, M.M., & Ferreira, P.R.M. (2007). Non-supervised pattern recognition methods: Exploring chemometrical procedures for evaluating analytical data. *Quimica Nova, 30,* 481–487.

Cuadros Rodríguez, L., Pérez Castaño, E., & Ruiz Samblás, C. (2016). Quality performance metrics in multivariate classification methods for qualitative analysis. *Trends in Analytical Chemistry, 80*, 612–624.

Cuadros Rodríguez, L., Ruiz Samblás, C., Valverde Som, L., Pérez Castaño, E., & González Casado, A. (2016). Chromatographic fingerprinting: An innovative approach for food 'identitation' and food authentication - A tutorial. *Analytica Chimica Acta, 909*, 9–23.

Ellis, D.I., Brewster, V.L., Dunn, W.B., Allwood, J.W., Golovanov, A.P., & Goodacre R. (2012). Fingerprinting food: current technologies for the detection of food adulteration and contamination. *Chemical Society Reviews*, *41*, 5706-5727.

EU Commission Decission (2002/657/EC) implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results, Official Journal of the European Communities, L 221/8-36.

Fernandes, A.M.A.P., Fernandes, G.D., Simas, R.C., Barrera-Arellano, D., Eberlin, M.N., & Alberici, R.M. (2013). Quantitation of triacylglycerols in vegetable oils and fats by easy ambient sonic-spray ionization mass spectrometry. *Analytical Methods, 5,* 6969-6975.

Gázquez Evangelista, D., Pérez Castaño, E., Sánchez Viñas, M., & Bagur González, M.G. (2013). Using offline HPLC-GC-FID 4-Desmethylsterols Concentration Profiles, Combined with Chemometric Tools, to Discriminate Different Vegetable Oils. *Food Analytical Method*, *7*, 912–925.

Jabeur, H., Zribi, A., Makni, J., Rebai, A., Abdelheidi, R., & Bouaziz, M. (2014). Detection of chemlali extra-virgin olive oil adulteration mixed with soybean oil, corn oil, and sunflower oil by using GC and HPLC. *Journal of Agricultural and Food Chemistry, 62,* 4893–4904.

Lerma García, M.J., Simó Alfonso, E.F., Méndez, A., Lliberia, J.L., & Herrero Martínez, J.M. (2011). Classification of extra virgin olive oils according to their genetic variety using linear discriminant analysis of sterol profiles established by ultra-performance liquid chromatography with mass spectrometry detection. *Food Research International, 44,* 103–108.

López, M.I., Trullos, E., Callao, M.P., & Ruisánchez, I. (2014). Multivariate screening in food adulteration: untargeted versus targeted modelling. *Food Chemistry, 147,* 177-181.

Luts, J., Ojeda, F., Van de Plas, R., De Moor, B., Van Huffel, S., & Suykens, J.A.K.(2010). A tutorial on support vector machine-based methods for classification problems in chemometrics. *Analytica Chimica Acta, 665,*129–145.

Marini, F. (2010). Classification Methods in Chemometrics. *Current Analytical Chemistry*, *6*, 72–79.

Naes, T., Isaksson, T., Fearn T., & Davies, T. (2002). A user-friendly guide to multivarate calibration and classification, NIR Publicationes, Chichester.

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.

Ruiz Samblás, C., Marini, F., Cuadros Rodríguez, L., & González Casado, A. (2012). Quantification of blending of olive oils and edible vegetable oils by triacylglycerol fingerprint gas chromatography and chemometric tools. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences, 910*, 71–77.

Sánchez de Medina, V., El Riachy, M., Priego-Capote, F., & Luque de Castro, M.D.(2014). Composition of fatty acids in virgin olive oils from cross breeding segregating populations by gas chromatography separation with flame ionization detection. *Journal of the Science of Food and Agriculture, In press*, doi:10.1002/jsfa.7030.

SVM Function Settings, Eigenvector Documentation Wiki [accessed 09 November 2015], http://wiki.eigenvector.com/index.php?title=SVM_Function_Settings

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.

Xu, L., Yan, S-M., Cai, C-B., Yu, X-P. (2013). One-class partial least squares (OCPLS) classifier. *Chemometrics and Intelligent Laboratory Systems, 126*, 1-5.

Xu, Y., Zomer, S., & Brereton, R.G. (2006). Support Vector Machines: A recent method for classification in chemometrics. *Critical Reviews in Analytical Chemistry, 36,* 177–188.

ARTÍCULO CIENTÍFICO

A new analytical method for quantification of olive and palm oil in blends with other vegetable edible oils based on the chromatographic fingerprints from the methyl-transesterified fraction

(Artículo publicado en 2017 en la revista: Talanta, 164, 540-547)

Talanta 164 (2017) 540–547

Contents lists available at ScienceDirect

Talanta

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

A new analytical method for quantification of olive and palm oil in blends with other vegetable edible oils based on the chromatographic fingerprints from the methyl-transesterified fraction

Ana M. Jiménez-Carvelo*, Antonio González-Casado, Luis Cuadros-Rodríguez

Department of Analytical Chemistry, University of Granada, c/Fuentenueva, s.n, Granada, E-18071 Spain

A new analytical method for quantification of olive and palm oil in blends with other vegetable edible oils based on the chromatographic fingerprints from the methyl-transesterified fraction

Ana M. JIMÉNEZ-CARVELO ✉, Antonio GONZÁLEZ-CASADO , Luis CUADROS-RODRÍGUEZ

Department of Analytical Chemistry, University of Granada, c/ Fuentenueva, s.n. E-18071 Granada, Spain.

Abstract

A new analytical method for the quantification of olive oil and palm oil in blends with other vegetable edible oils (canola, safflower, corn, peanut, seeds, grapeseed, linseed, sesame and soybean) using normal phase liquid chromatography, and applying chemometric tools was developed. The procedure for obtaining of chromatographic fingerprint from the methyl-transesterified fraction from each blend is described. The multivariate quantification methods used were Partial Least Square-Regression (PLS-R) and Support Vector Regression (SVR). The quantification results were evaluated by several parameters as the Root Mean Square Error of Validation (RMSEV), Mean Absolute Error of Validation (MAEV) and Median Absolute Error of Validation (MdAEV). It has to be highlighted that the new proposed analytical method, the chromatographic analysis takes only eight minutes and the results obtained showed the potential of this method and allowed quantification of mixtures of olive oil and palm oil with other vegetable oils.

Keywords

Olive oil, palm oil, blends, vegetable oil, fingerprint, liquid chromatography

✉ Corresponding author: phone: +34 958240797; fax: +34 958243328; email: amariajc@ugr.es

1. Introduction

A common practise in the foodstuffs industry in the elaboration of foods is the use of mixtures of different vegetable edible oils. The food products can be produced with these blends (oils used for frying) or contained as ingredient. This action happens mainly due to two reasons: (i) to reduce cost and to have a better product performance, or (ii) to obtain certain characteristics of stability or nutritional quality of the blends.

To harmonize such practices in Europe for the foodstuffs industry, there are two European Regulations which describe the compulsory characteristics and the provision of food information to consumers. A specific Regulation about foods containing olive oil [**1**], and a general Regulation about foods containing blends of vegetable edible oil different to olive oil [**2**]. Concerning to olive oil, European Regulation nº 29/2012 establishes that *The percentage of olive oil and certain indications specific to products consisting exclusively of a blend of vegetable oils should therefore be clearly shown on the labelling and … the blend concerned must bear the following trade description: 'Blend of vegetable oils (or the specific names of the vegetable oils concerned) and olive oil', directly followed by the percentage of olive oil in the blend. By contrast, European Regulation nº 1169/2011 only indicates that May be grouped together in the list of ingredients under the designation 'vegetable oils' followed immediately by a list of indications of specific vegetable origin …* Therefore, in case of blends of vegetable edible oils (non-olive oil) is not obligated to declare the percentage for each oil.

The verification of compliance of these legal regulations requires rapid and simple analytical methods, which can be used as routine analysis methods. These methods should identify the type and botanic species of each vegetable oil used as ingredient in food or in mixtures of different vegetable edible oils. Furthermore, if the vegetable oil used is olive oil, the method must be able to determine the proportion of olive oil in relation to the total mass of the food.

Extra virgin olive oil (EVOO) is produced from the pressing of the fruits of the olive tree (Olea europea) by mechanical procedures (grinding followed by centrifugation and/or decantation). The olive oil is composed by two fractions: (i) the majority or saponifiable fraction, mainly constituted by triglycerides and free fatty acids; and (ii) the unsaponifiable fraction chemically composed by minor components as sterols, phenols, tocopherols etc. [**3**]. EVOO has excellent sensory and a high content in oleic acid and phenols compounds, for this reason many health claims attributed to it. It is

associated to the reduction of the content of cholesterol in blood, cardiovascular risks and of illnesses as the obesity and the diabetes [**4**]. Thus, it is a product highly appreciated by consumers and is candidate to be adulterated with other vegetable oils to increase their profitability [**5,6**].

Currently, most studies to date about quantification of olive oil in blends with other vegetable oils are based on the study of chromatographic fingerprinting olive oil major fraction (triglycerides and fatty acids) applying chemometric tools as partial least squares regression (PLS-R) and partial component regression (PCR). Mainly, the gas chromatography and the spectroscopy techniques as near infrared region (NIR) and Fourier transform-infrared region (FTIR) are the analytical techniques used [**7,8,9,10,11,12**]. Recently, it has been reported a work about quantification of mixtures of monovarietal extra virgin olive oils using the visible ultraviolet fingerprint and applying artificial neural networks (ANN) [**13**].

There is also interest in detecting and quantifying adulterations of palm oil with other vegetable oils due to social and environmental impacts of this oil industry [**14**]. Palm oil is obtained from the fruit of the palm (Elais Guineensis), originally from Gulf of Guinea and in West Africa, where is not suitable for storage and it must be refined [**15**]. Currently, it is cultivated in different tropical places of the world; the main plantations are in Indonesia and Malaysia [**16**]. From palm plant several types of oil are produced: (i) the orange-red crude palm oil, obtained from the flesh of the palm. This oil is known as 'palm oil'; and (ii) the yellow crude palm kernel oil obtained from seed palm fruit. Palm oil is composed by 50% saturated fatty acids, the rest are monounsaturated and polyunsaturated fatty acids and contains few phytosterols [**17**].

Palm oil is the most consumed oil worldwide. This is due to its low production costs and its multiple uses [**18**]. The growth of trade in this product has produced the maximum planting of palms at the expense of the devastation of protected forests, with the consequent threat to the life of the animal and plant species that live there. This illegal human activity is causing serious environmental problems in countries such as Malaysia. For this reason in 2004 *Roundtable on Sustainable Palm Oil (RSPO) was founded* [**19,20**]. *This is a non-profit organization whose aim is to join all sectors* of the palm oil industry and ensure that the production of this oil respects the environment and *not threaten the biodiversity.* There are also other organizations such as the Federación Nacional de Cultivadores de Aceite de Palma (Fedepalma) (National Federation of Palm

Oil Growers) created in 1962 in Colombia, formed by farmers of palm oil in the country and whose aims is to support the development of the palm in Colombia and work sustainability of it [**21**].

Approximately 1% of this oil is used as initial material for obtaining biodiesel, 19% in chemical oil industry (paints, detergents, cosmetics etc.) and 80% is mainly intended for the food industry for different purposes. Palm oil is used for frying due to it resists high temperatures compared with other vegetable oils. It is used as ingredient in breads and pastries and as substitute for some fat due to its low price compared to others [**22**].

Actually, there are few studies about quantification of palm oil in blends with other vegetable edible oils; the analyses performed so far mainly applying spectroscopic techniques such as FTIR and chemometric tools for data processing [**23**,**24**].

In the context of food authenticity it is crucial to verify that the information shown on the label of a food containing olive oil or vegetable oil blends are correct. Aims of this study are to develop multivariate methods of quantification of olive oil and palm oil in blends with other vegetable edible oils. For that, it was obtained a chromatographic fingerprinting from methyl-transesterified fraction of mixtures of olive oil and palm oil with other vegetable oils by liquid chromatography. Different chemometric tools have been applied for the quantification of olive oil and palm oil as Partial Least Squares Regression (PLS-R) and Support Vector Regression (SVM-R). In both cases, the results of quantification models have been evaluated by the Root Mean Square Error of Validation (RMSEV), Mean Absolute Error of Validation (MAEV) and Median Absolute Error of Validation (MdAEV).

2. Materials and methods

2.1. Chemicals

All solvents used were HPLC grade. Isopropanol, n-hexane, methanol and tert-butyl methyl ether (TBME) were provided by the VWR International Eurolab, S.L. (Barcelona, Spain).

Other reagents, sodium methoxide (MeONa), citric acid monohydrate, and anhydride sodium sulphate were purchased from Merck (Darmstadt, Germany). The nitrogen (99.9999 %) used was provided by Air Liquid (Madrid, Spain).

2.2. Samples

The olive oil samples, to build the blends were twenty-five, including three categories: extra virgin (EVOO), virgin (VOO) and olive oil (OO, blend of virgin and refined) of the different olive fruit varieties. In addition, fifty-two edible oils samples of eight vegetable types, obtained each one from different trademark, were used: soya oils (8), sunflower oils (11), canola oils (10), corn oils (5), seeds oils (5), peanut oils (5), sesame oils (4) and grapeseed oils (4).

For the calibration set, six pure samples of olive oil and six mixtures of different vegetable edible oils were also analysed and twenty-four blends samples were prepared by mixing several olive oils (EVOO, VO and OO) with diverse vegetable edible oils in different percentage units (g of olive oil by 100 g of blend oil), obtaining four different levels of concentration from 20 to 80% (w/w). For each specific vegetable type, all the available trademarks were used in mixing. For example, the blend number 19 was composed by four EVOO of different olive fruit varieties ('hojiblanca' (2), 'manzanilla' and 'picual') and five vegetable edible oil (soya (2), canola, sesame and grapeseed oils). Likewise, for validation set twenty-four blends samples were prepared. Table 1a and table 1b show the blends of oils samples for the calibration set and validation set respectively.

The palm oil samples from different geographic origins: Africa (AF), Asia (AS) and America (AM), to build the blends were seventeen. The fifty-two vegetable edible oils were the same that were used to make blends with olive oil.

Table 1. Percentage of the olive oil and other vegetable edible oil in the oil blend samples.

	Nº	Composition	Nº	Composition
(a)	*Calibration set*			
	1	100% MDVO*	19	60% EVOO + 40% MDVO
	2	100% MDVO	20	36% EVOO + 12% OO +12% VOO + 40% MDVO
	3	100% MDVO	21	40% EVOO + 10% VOO + 10% OO + 40% MDVO
	4	100% MDVO	22	60% EVOO + 40% MDVO
	5	100% MDVO	23	36% EVOO + 12% OO +12% VOO + 40% MDVO
	6	100% MDVO	24	40% EVOO + 10% VOO + 10% OO + 40% MDVO
	7	20% EVOO+ 80% MDVO	25	40% EVOO + 20% VOO + 20% OO + 20% MDVO
	8	15% EVOO + 5% OO+ 80% MDVO	26	80% EVOO + 20% MDVO
	9	15% EVOO + 5% VOO+ 80% MDVO	27	60% EVOO + 20% OO + 20% MDVO
	10	20% EVOO+ 80% MDVO	28	40% EVOO + 20% VOO + 20% OO + 20% MDVO
	11	15% EVOO + 5% OO+ 80% MDVO	29	80% EVOO + 20% MDVO
	12	15% EVOO + 5% VOO+ 80% MDVO	30	60% EVOO + 20% OO + 20% MDVO
	13	30% EVOO + 10% OO+ 60% MDVO	31	100% EVOO
	14	30% EVOO + 10% OO+ 60% MDVO	32	100% EVOO
	15	30% EVOO + 10% VOO+ 60% MDVO	33	100% VOO
	16	30% EVOO + 10% OO+ 60% MDVO	34	100% EVOO
	17	30% EVOO + 10% OO+ 60% MDVO	35	100% VOO
	18	30% EVOO + 10% VOO+ 60% MDVO	36	100% EVOO
(b)	*Validation set*			
	1	68% EVOO + 32% MDVO*	13	68% EVOO + 32% MDVO*
	2	17.50% VOO + 82.50% MDVO	14	17.50% VOO + 82.50% MDVO
	3	93% VOO + 7% MDVO	15	93% VOO + 7% MDVO
	4	44% EVOO + 56% MDVO	16	44% EVOO + 56% MDVO
	5	5% EVOO + 95% MDVO	17	5% EVOO + 95% MDVO
	6	68% EVOO + 32% MDVO	18	68% EVOO + 32% MDVO
	7	70% VOO + 30% MDVO	19	70% VOO + 30% MDVO
	8	31% EVOO + 69 MDVO	20	31% EVOO + 69 MDVO
	9	52% EVOO + 48% MDVO	21	52% EVOO + 48% MDVO
	10	25% EVOO + 75% MDVO	22	25% EVOO + 75% MDVO
	11	90% EVOO + 10% MDVO	23	90% EVOO + 10% MDVO
	12	40% EVOO + 60% MDVO	24	40% EVOO + 60% MDVO

** MDVO: Mixture of different vegetable edible oils (non-palm and non-olive oils)*

For the calibration set, a total of six pure samples palm oil from different origins were analysed: two from Africa, two from America and two from Asia. Moreover, the same six blends of different vegetable edible oils that for quantification olive oil were used. Twenty-four blends were also prepared by mixing of palm oils (AF, AS and AM) with various vegetable edible oils in different percentage units (g of palm oil by 100 g of blend oil), obtaining four different levels of concentration from 20 to 80% (w/w). In this case, for validation set only eighteen blends were prepared. Table 2a and table 2b show the composition of each blend which were used as calibration and validation set respectively. The calibration chart, for the calibration set of olive oil blends is shown in Figure 1.

Table 2a. Percentage of the palm oil and other vegetable edible oil in the oil blend samples.

	Nº	Composition	Nº	Composition
(a)	*Calibration set*			
	1	100% MDVO*	19	30% AF + 15% AM + 15% AS + 40% MDVO
	2	100% MDVO	20	60% AS + 40% MDVO
	3	100% MDVO	21	30% AS + 30% AF + 40% MDVO
	4	100% MDVO	22	30% AF + 15% AM + 15% AS + 40% MDVO
	5	100% MDVO	23	60% AS + 40% MDVO
	6	100% MDVO	24	30% AS + 30% AF + 40% MDVO
	7	10% AF + 10% AS + 80% MDVO	25	60% AF + 20% AM + 20% MDVO
	8	10% AF + 10% AS + 80% MDVO	26	54% AS + 26% AF + 20% MDVO
	9	5% AF + 10% AS + 5% AM + 80% MDVO	27	60% AS + 20% AF + 20% MDVO
	10	10% AF + 10% AS + 80% MDVO	28	60% AF + 20% AM + 20% MDVO
	11	10% AF + 10% AS + 80% MDVO	29	54% AS + 26% AF + 20% MDVO
	12	5% AF + 10% AS + 5% AM + 80% MDVO	30	60% AS + 20% AF + 20% MDVO
	13	20% AF + 20% AS + 60% MDVO	31	100% AF
	14	30% AS + 10% AF + 60% MDVO	32	100% AF
	15	30% AS + 10% AF + 60% MDVO	33	100% AS
	16	20% AF + 20% AS + 60% MDVO	34	100% AS
	17	30% AS + 10% AF + 60% MDVO	35	100% AM
	18	30% AS + 10% AF + 60% MDVO	36	100% AM

** MDVO: Mixture of different vegetable edible oils (non-palm and non-olive oils)*

Table 2b. Percentage of the palm oil and other vegetable edible oil in the oil blend samples.

(b)	*Validation set*		
1	43% AM + 57% MDVO	10	43% AM + 57% MDVO
2	18% AF + 82% MDVO	11	18% AF + 82% MDVO
3	25% AF + 75% MDVO	12	25% AF + 75% MDVO
4	34% AS + 66% MDVO	13	34% AS + 66% MDVO
5	56% AS + 44% MDVO	14	56% AS + 44% MDVO
6	53% AS + 47% MDVO	15	53% AS + 47% MDVO
7	49% AS + 51% MDVO	16	49% AS + 51% MDVO
8	38% AS + 62% MDVO	17	38% AS + 62% MDVO
9	30% AS + 70% MDVO	18	30% AS + 70% MDVO

** MDVO: Mixture of different vegetable edible oils (non-palm and non-olive oils)*

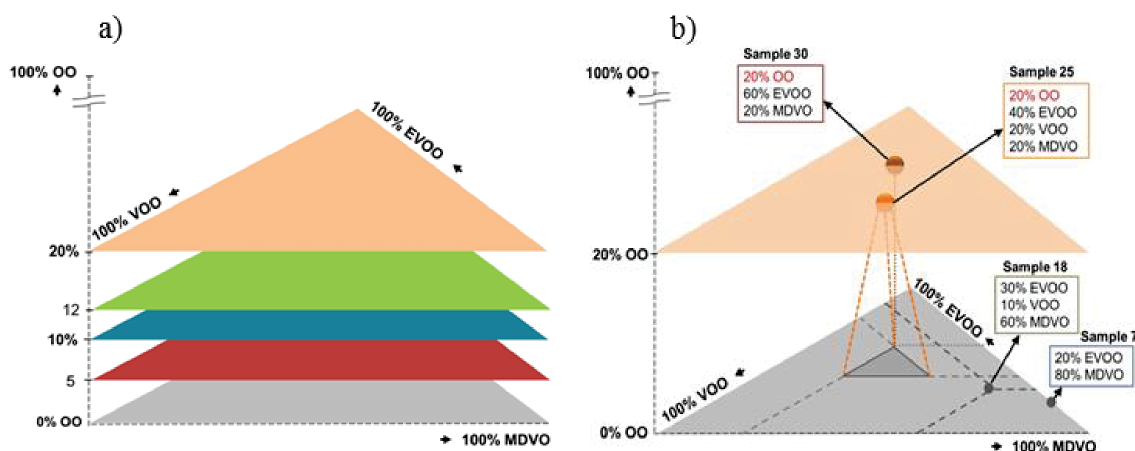


Figure 1. (a) Calibration chart of the training set for blends of olive oil. (b) Example of four blends of olive oil with other vegetable edible oils.

All the vegetable oils samples were directly obtained from local providers and were stored in dark bottles, at 6ºC until their analysis.

2.3. Sample preparation

Previous to the chromatographic analysis, a transesterification reaction was applied to the pure vegetable oil samples and all the different oil blends prepared. A modification of the original procedure described by Biedermann et al was applied [25]. For this, 0.1 g of oil was weighed into a centrifuge tube. 1 mL of extracting agent (MeONa at 10 % in

methanol in TBME, 4:6 (v/v)) was added and mixed with the oil. The mixture was stirred during 20 s, and then allowed to stand during 20 min. This step was repeated twice. Then 1 mL of water and 8 mL of hexane was added, and then the mixture was centrifuged 3 min at approximately 1,500g of acceleration. The aqueous phase was removed with a Pasteur pipette followed by addition of 1 mL of 1 % citric acid in water. Again, the aqueous phase was eliminated and 2 g of sodium sulphate anhydrous were added, and the mixture was allowed to stand during 20 min.

The methyl-transesterified organic fraction was filtered with a syringe filter of polytetrafluoroethylene (PTFE) membrane 0.22 µm and the subsequent solution was stored at –25ºC until analysis. For the chromatographic analysis, 350 µL of transesterified solution was added to a 2 mL HPLC vial.

2.4. Chromatographic conditions

The analyses were carried out with an Agilent 1100 series liquid chromatograph (Santa Clara, USA) equipped with a column thermostat (Eppendorf CH30), a quaternary pump and degasser auto sampler. Detection was performed with a corona charged aerosol detector (CAD) (ESA Bioscenses Inc., Chemlsford, MA, USA). Agilent ChemStation software (rev. B.02.01-SR1) for LC systems was used.

The chromatography fingerprint from methyl transesterified fraction was obtained by HPLC using a (250 × 4 mm i.d, 5 µm) column Lichrospher® 100 CN. The column temperature was set at 30 ºC during the entire operation. The composition of the mobile phase was n-hexane/isopropanol (96:4, v/v). The injection volume was 20 µL, moreover, also 5 µL were injected for mixtures of palm oil. The run time was 8 min.

2.5. Chemometrics

The raw data files from each chromatogram were obtained in a CSV file, and then exported to MATLAB format (version R2013a). The original analytical data were divided in different groups to perform the statistical analysis. In both quantifications (olive oil and palm oil), the calibration data set was made up of 36 samples. The validation set for olive oil quantification was composed of 24 samples and for palm oil quantification was composed of 18 samples.

All chemometric treatments were carried out by using the PLS Toolbox (Eigenvector Research Inc., Wenatchee, WA), for MATLAB software (Mathworks Inc., Natick, MA, USA).

Pre-processing data

The data pre-processing was done with a home-programmed MATLAB function named "Medina" (version 10) [**26**]. This function implements several algorithms from Matlab Bioinformatics Toolbox™, and 'icoshift' (interval correlation optimized shifting) algorithm to align the peaks of the chromatograms [**27**].

Two types of pre-processing of data were employed. First a pre-processing for calibration set was applied and then a second pre-processing for the validation set. The pre-processing steps for calibration set were: (1) raw chromatograms data grouping and overlay; (2) selection of interval of interest in chromatograms; (3) filtered of the raw chromatograms data to eliminate noise of signal analytical; (4) correction of the baseline using the 'msbackadj' function (included in the Bioinformatics ToolboxTM); (5) alignment of the peaks with the function 'icoshift'; (6) obtaining of the average chromatogram of all blends, and finally (7) mean centring of the data set. The following steps for validation set data were: (1) raw chromatograms data grouping and overlay; (2) selection of interval of interest in chromatograms; (3) filtered of the raw chromatograms data to eliminate noise of signal analytical; (4) correction of the baseline using 'msbackdj' function; (5) alignment of the peaks concerning the average chromatogram of the blends of the calibration set with the function 'icoshift'.

Previous screening: multivariate classification

A method of quantification should demonstrate first that the sample of interest is adulterated prior to quantification, using a method of classification previously validated based on the same vector of data.

Surprisingly, this methodology is not applied to study about quantification of oils, being a basic operation of good analytical practice. In the figure 2 is shown the correct process to carry out the quantification.

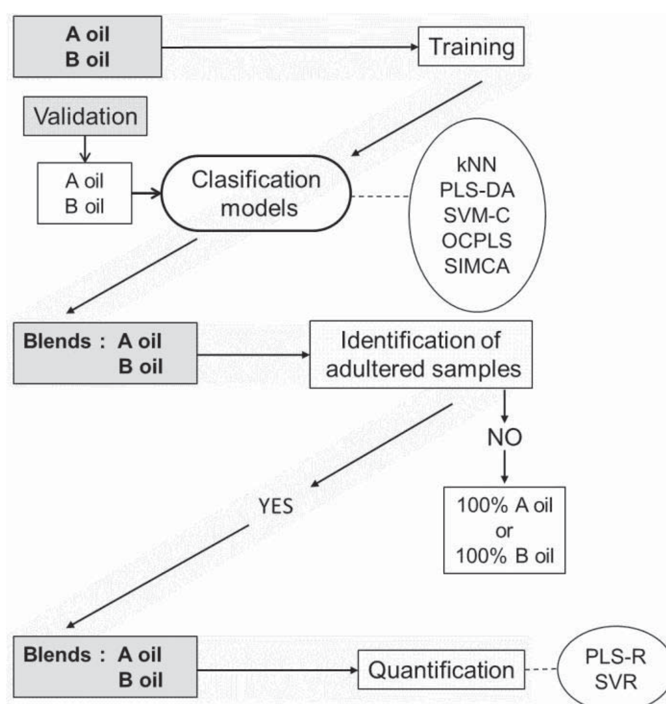


Figure 2. Flow chart describing the global process: training, classification and quantification.

In a previous paper published by authors, different models of classification of olive oil from other vegetable edible oils were developed using several chemometric methods as k-nearest neighbours (kNN) [**28**], partial least squared-discriminant analysis (PLS-DA) [**29**], one-class partial least squares (OCPLS) [**30**], support vector machine classification (SVM-C) [**31**], and soft independent modelling of class analogies (SIMCA) [**32**]. All the models were validated and were able to distinguish between olive oil from other vegetable edible oils. Similarly, models were established and validated for mixture of palm oil.

In accordance with the cited above, firstly a previous screening was performed about the blends of the validation set of olive oil and palm oil respectively, to ascertain if classification models were able to detect 24 mixtures of olive oil and 18 mixtures of palm oil with other vegetable oils. All the chemometric models were applied in both validation-set.

Once concluded all the classification models the calibration models for quantification of blends of olive oil and palm oil were developed.

Quantification: multivariate calibration

Partial least square-regression (PLS-R) and support vector regression (SVR) are supervised chemometric methods to build models of quantitative prediction. The main difference between them is that PLS-R is a linear chemometric method and SVR builds a hyperplane (or set of hyperplanes) for perform the prediction.

PLS-R calculates a single partial least squares regression model using the given number of components 'ncomp' to predict a dependent variable 'y' from a set of independent variables 'x' [**33**]. The standard PLS-R model is defined as [**34**]:

$$X=TP^{T}+E$$

$$Y= Tq^{T}+f$$

Where X is the independent variable (X-block) data and Y is the dependent variable (Y-block) data.

This chemometric tool is based on the search for the relationship between the matrix X, known as 'vector of independent variables', and another matrix Y, known as 'matrix of predictor variables'. In this study, the measurement data file of chromatographic fingerprinting is the 'matrix of predictor variables' and the concentration is the 'vector of independent variables'. The scores are obtained by decomposing the data matrices into a sum of rank one-component matrices.

SVR this is a non-linear regression method which can be considered a hybrid of Multiple Linear Regression (MLR) and Locally Weighted Regression (LWR). The calibration step selects calibration samples (called "support vectors") which are deemed the most critical to defining and optimization of an error function (regression). Depending on the definition of this error function, two types of SVR models: (i) Epsilon-SVR (ε-SVR) optimizes a model using the adjustable parameters epsilon (upper tolerance on prediction errors) and C (cost of prediction errors larger than epsilon), and (ii) Nu-SVR optimizes a model using the adjustable parameter Nu (0 → 1] which indicates a lower bound on the number of support vectors to use given as a fraction of total calibration samples [**35**].

3. Results and discussion

Previous screening multivariate classification of olive oil blend samples

Figure 3a shows the chromatograms of the whole calibration data set and figure 3b the chromatograms of validation set, both after of the pre-processing. As described in the previous section, identification and quantification of the blends were carried out in two successive stages: at first, a classification screening was used to separate pure olive oils from blends and then calibration models were built.

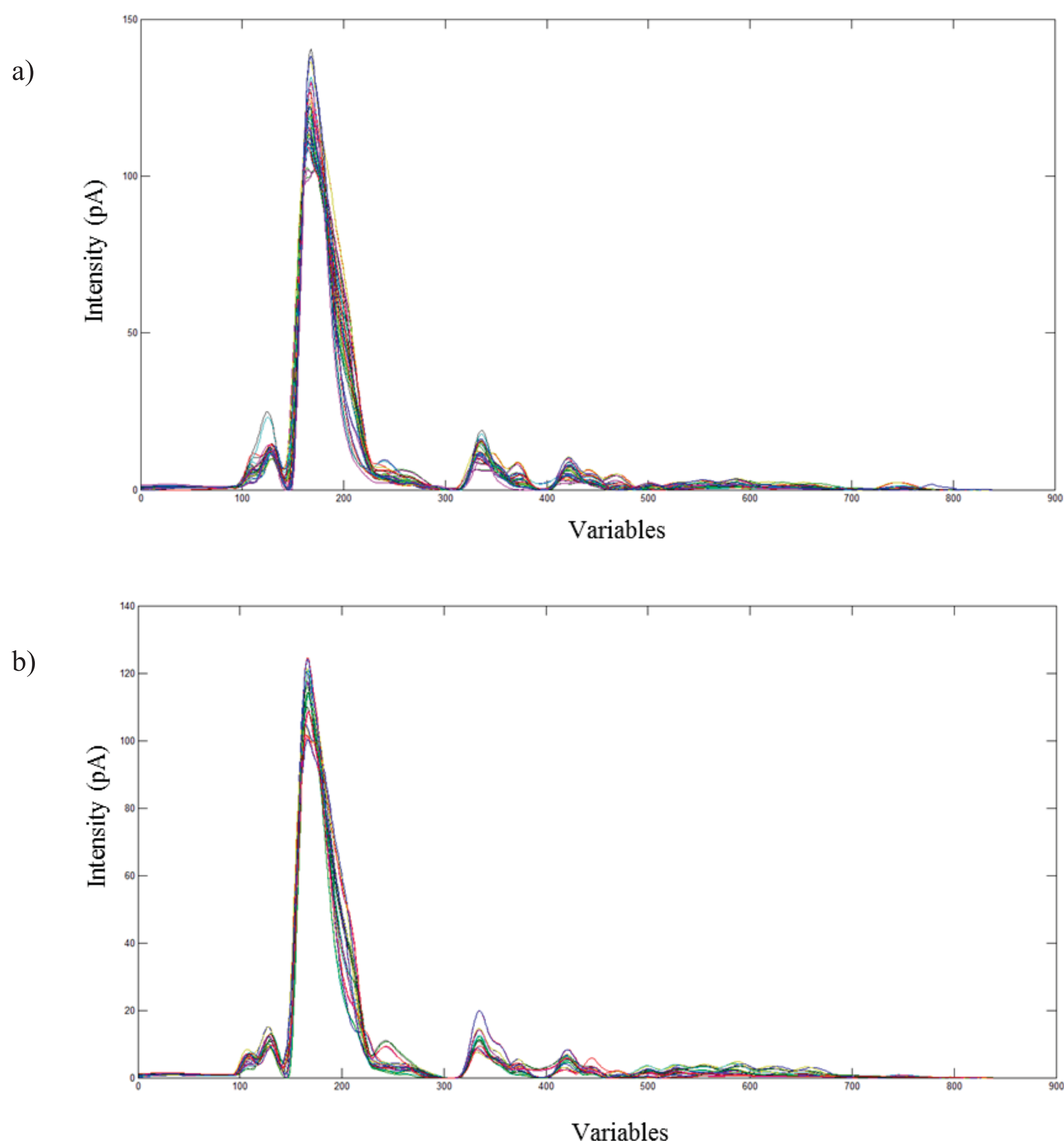


Figure 3. Superposed chromatograms of the vegetable oils blends samples of olive oil: (a) calibration and (b) validation set.

In the figure 4 is shown the SVM-C model for the blends of olive oil. As can be seen, the model distinguishes between samples of pure olive oil from the blends. Once identified adulterated samples, it was carried out the quantification.

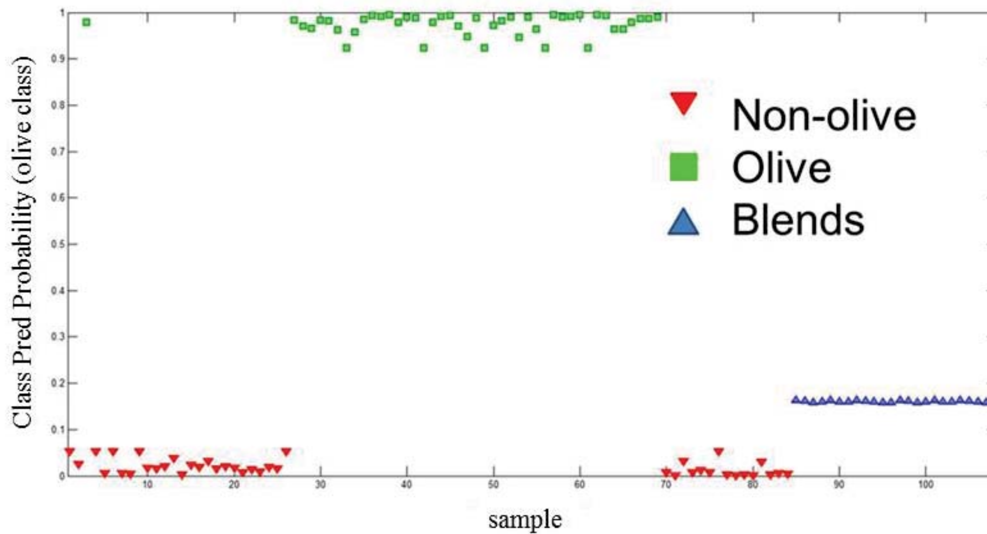


Figure 4. Classification plot for the blends of olive oil.

Quantification of olive oil

For calibration models PLS-R and SVR in all its varieties were used. The reliability of the models was established on the basis two criteria: (i) the suitability of the model by means of the determination coefficient (R^2), and (ii) the errors of quantification (validation errors) were evaluated with the Root Mean Square Error of Validation (RMSEV), Mean Absolute Error of Validation (MAEV) and Median Absolute Error of Validation (MdAEV) [**36**].

$$\mathrm{RMSEV} = \sqrt{\frac{\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}{n}} \; ; \quad \mathrm{MAEV} = \mathrm{mean}\,(|\hat{y}_t - y_t|) \; ;$$

$$\mathrm{MdAEV} = \mathrm{median}\,(|\hat{y}_t - y_t|)$$

Where 'n' is the number of different validation blends and $'\hat{y}'$ the predicted values of the blend for time 't' of a regression's dependent variable 'y_t'.

The results obtained in terms of R^2, RMSEV, MAEV and MdAEV are presented in the table 3.

Table 3. Quality metrics of PLS-R and SVR models for the blends of olive oil.

	PLS-R	ε-SVR			nu-SVR		
		None	PLS	PCA	None	PLS	PCA
R^2	0.99	1.00	0.99	0.99	1.00	0.99	0.99
RMSEV*	5.7	6.1	8.4	13.5	6.4	8.6	10.4
MAEV*	5.0	4.6	5.6	8.6	4.5	5.7	9.0
MdAEV*	4.9	5.1	7.1	9.9	5.3	7.1	8.7

* Units: g of olive oil / 100 g blend oil

Firstly, the PLS-R calibration model was built with five latent variables. The evaluation of the method linearity was performed in accordance with proportional relationship between responses versus blend concentrations. The determination coefficient value R^2 was 0.99 and venetian blinds were used to choose the optimal model complexity. In order to validate the model 24 test samples of blends of olive oil with other vegetable edible oils in different proportions were process by PLS-R. Then, for each sample, the PLS model predicted of concentration. The process is displayed in the figure 5. Table 3 shows an error, averaged from the three validation errors, about 5.2 percentage units (g of olive oil by 100 g of blend oil).

Next, SVR model in all its varieties (ε-SVR and nu-SVR) was applied. All the models were tested both without and with X-block compression by PCA and PLS. In the models with X-block compression the R^2 estimated was 0.99 and 1.00 for SVR models without reducing variables of X-block. In the validation step, in both cases (ε-SVR and nu-SVR), without X-block compression and with PLS X-block compression given similar quantification results. If no compression data is applied, the averaged errors are in the order of 5 percentage units. When PCA was applied to the data in order to reduce the number of X-block variables, ε-SVR model gave a RMSEV and MdAEV values higher than nu-SVR. RMSEV, MAEV and MdAEV values from both models, with X-block compression by PCA, are beyond 10 percentage units.

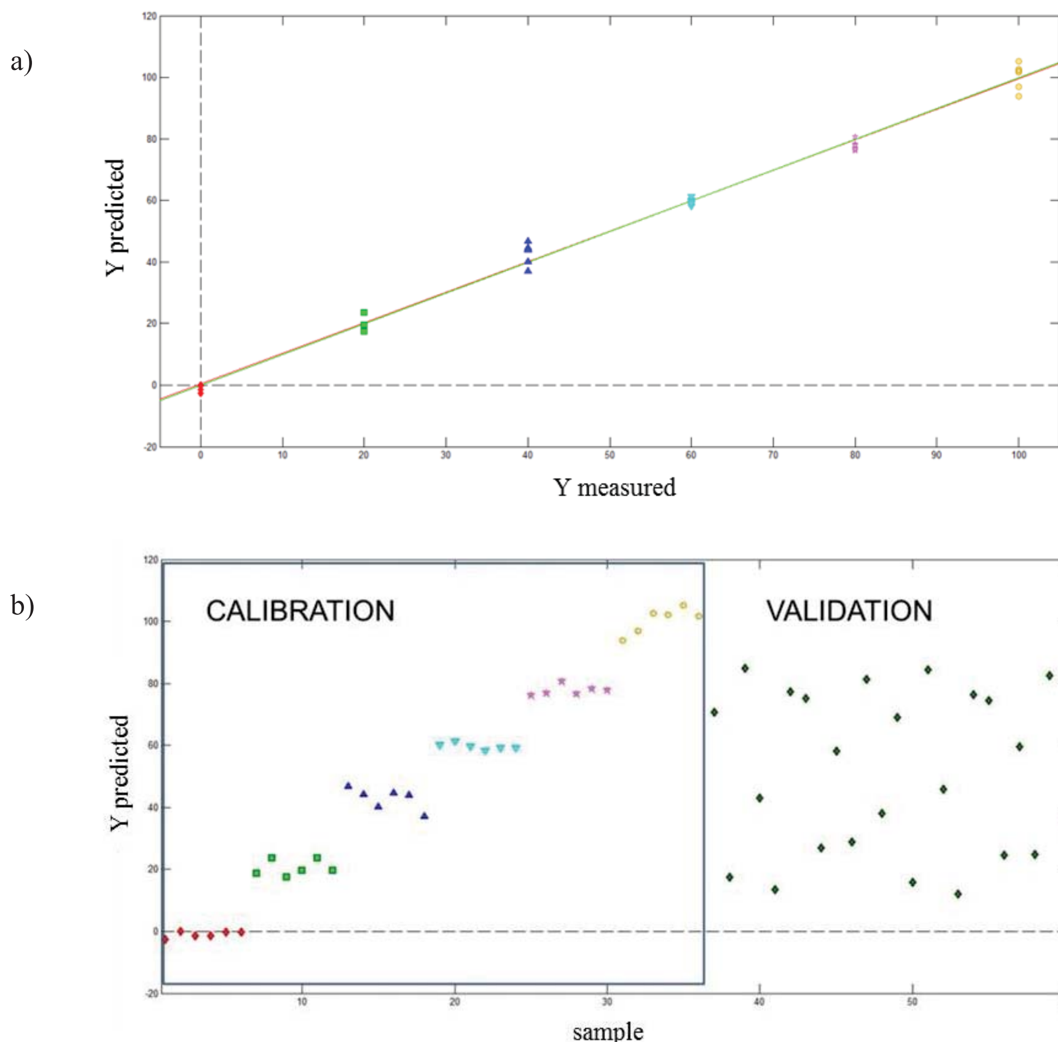


Figure 5. (a) The calibration model of PLS for training set of blends of olive oil, (b) distribution of samples in the calibration set (left) and the validation set (right) of blends olive oil.

Previous screening multivariate classification of palm oil blend samples

Figure 6 shows chromatographic fingerprints from methyl transesterified of mixtures of palm oils with other vegetable edible oils of the calibration set. Two regions can be easily differentiated: the region 1 shows a major peak which is essentially composed by the methyl esters of the fatty chains; and the region 2 which is composed by several minor peaks and contains basically information on the families of free sterols and terpenic alcohols.

As in the case of the blends of olive oil, first, a classification model to identify the adulterated samples of palm oil about the validation set composed of 18 blends was applied. Next, the quantification model was carried out.

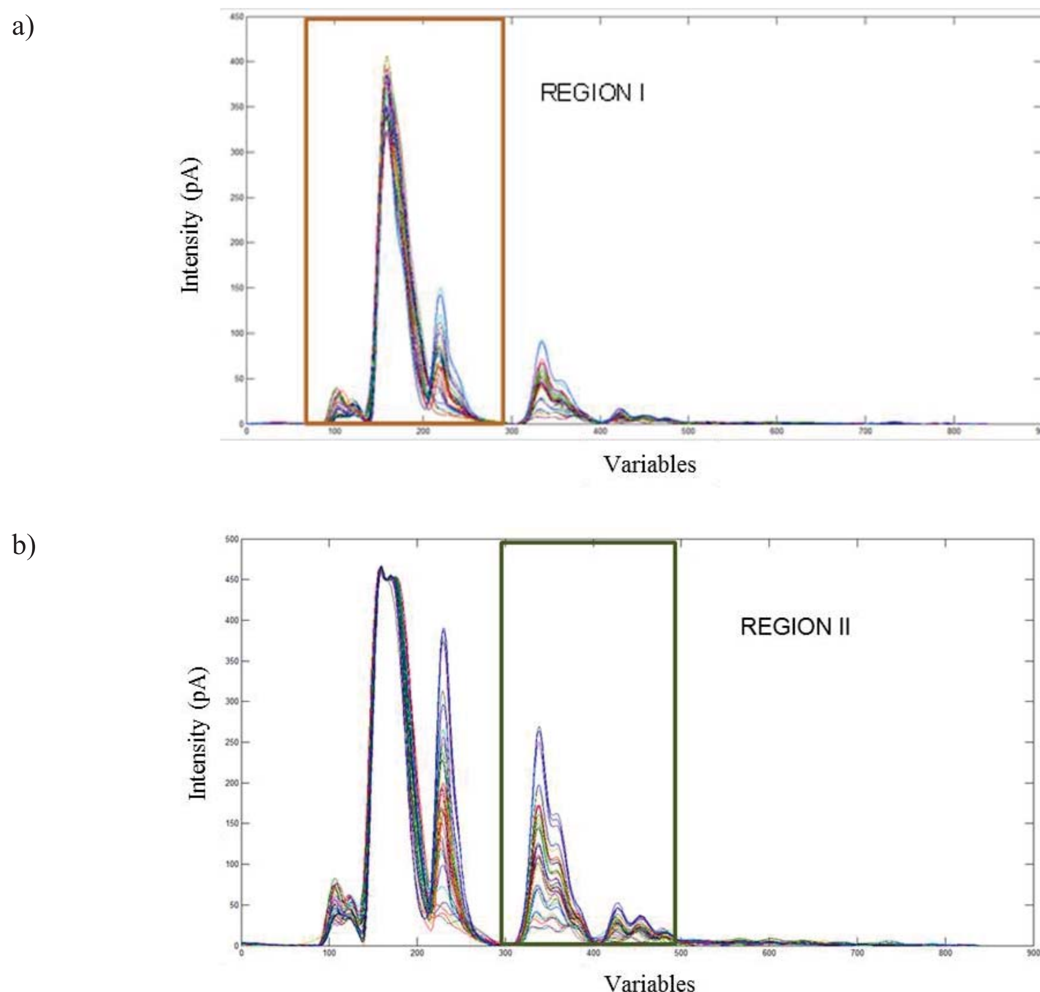


Figure 6. Superposed chromatograms of the vegetable oils blends samples of palm oil of calibration set showing the two characteristic regions: (a) region I and (b) region II.

Quantification of palm oil

The criteria to evaluate of quantification models were the same that in the quantification of olive oil. The results obtained are shown in the table 4.

Models were developed for the region 1 and region 2 to see if the quantification improves. Due to the distinct intensity of the peaks from the region 1 and region 2, different injection volumes were required in order to obtain a signal sufficiently high which did not saturate the detector. For this, 5 µL and 20 µL were selected for collect the chromatographic peaks corresponding to regions 1 and 2.

Firstly, the PLS-R models on regions I and II were built with 6 latent variables. The model on region 1 shows both best fitting and lower validation errors. In addition, ε-SVR and nu-SVR models were applied without and with reducing X-block data. In conclusion, quantification results on region 2 were worse than on region 1. PLS-R and ε-SVR with X-block PLS compression on region I were the best models. In these cases, R^2 were 0.99 and the validation error values were about 5.5-9.2 percentage units.

Table 4. Quality metrics of PLS-R and SVR models for the blends of palm oil.

	PLS-R	ε-SVR			nu-SVR		
		None	PLS	PCA	None	PLS	PCA
(a) Region I							
R^2	0.99	0.99	0.99	0.92	0.99	0.99	0.92
RMSEV*	9.0	11.0	9.2	17.4	11.4	9.1	17.7
MAEV*	6.3	9.4	5.5	17.2	9.8	5.2	17.7
MdAEV*	7.1	10.5	7.8	16.5	11.0	7.5	16.8
(b) Region II							
R^2	1.0	0.99	0.99	0.92	0.99	0.99	0.92
RMSEV*	13.1	17.9	11.3	14.9	18.5	11.3	12.4
MAEV*	11.6	16.0	7.7	9.3	17.0	7.7	6.5
MdAEV*	12.3	15.2	9.7	12.5	15.7	9.7	9.5

* Units: g of olive oil / 100 g blend oil

4. Conclusion

It has been proved that the chromatographic fingerprint from methyl-transesterified provides the information needed to quantify the amount of olive oil and palm oil in blends with other vegetable edible oils. In basis to these chromatograms, multivariate models to quantify of olive oil and palm oil applying multivariate regression methods have been developed.

In this study, a previous screening classification step has been proposed. Two well-established calibration methods have been then applied: PLS-R and SVR in all its varieties. The models of quantification of olive oil give better results (minor errors in

validation) than when are applied about palm oil. PLS-R model is the best model in quantification of olive oil and also in quantification of palm oil when the region 1 of the chromatogram is used.

Both proposed methods could be labelled as a "global method" of detection and quantification of olive oil or palm oil, respectively. They can detect adulterations of olive and palm oil with other vegetable edible oils, and to quantify the composition in the blends, regardless of the nature or the number of the vegetable oil(s) used as adulterant(s). The fact that is not necessary previously to identify the vegetable oil or vegetable oils mixed with the olive or palm oil is a great advantage for the (official or private) laboratories which carry out quality control of olive or palm oil.

Reference

[1] Regulation (UE) Nº 29/2012 on marketing standards for olive oil (consolidated version), 2012R0029, Official Journal of the European Union. (2016) 005.001–1-15.

[2] Regulation (UE) Nº 1169/2011 on the provision of food information to consumers, amending Regulations (EC) No 1924/2006 and (EC) No 1925/2006 of the European Parliament and of the Council, and repealing Commission Directive 87/250/EEC, Council Directive 90/496/EEC, Commission Directive 1999/10/EC, Directive 2000/13/EC of the European Parliament and of the Council, Commission Directives 2002/67/EC and 2008/5/EC and Commission Regulation (EC) No 608/2004, Official Journal of the European Union (2011), L 304/18.

[3] M.C. Ramírez-Tortosa, S. Granados, J.L. Quiles, Chemical composition, Types and Characteristics of Olive Oil in: J.L. Quiles, M.C. Ramírez-Tortosa, P. Yaqoob (Eds), Olive Oil and Health, Cab International, Wallingford (UK), 2006, pp. 45-62.

[4] F. Visioli, P. Bogani, C. Galli, Healthful Properties of olive oil Minor components in D. Boskou (Ed), Chemistry and Technology, AOCS Press,Champaign (IL), 2006, pp. 173-190.

[5] A. Tay, R.K. Singh, S.S. Krishnan, J.P. Gore, Authentication of Olive Oil Adulterated with Vegetable Oils Using Fourier Transform Infrared Spectroscopy, LWT - Food Science and Technology. 35 (2002) 99-103

[6] X. Sun, W. Lin, X. Li, Q. Shen, H. Luo, Detection and quantification of extra virgin olive oil adulteration with edible oils by FT-IR spectroscopy and chemometrics, Analytical methods. 7 (2015) 3939-3945.

[7] M. Monfreda, L. Gobbi, A. Grippa, Blends of olive oil and seeds oils: Characterisation and olive oil quantification using fatty acids composition and chemometric tools, Part II, Food Chemistry. 145 (2014) 584-592.

[8] A. Tay, R.K. Singh, S.S Krishnan, J.P. Gore, Authentification of olive oil adulterated with vegetable oils using fourier transform infrared spectroscopy, Food Science & Technology. 35 (2002) 99-103.

[9] A. Rohman, Y.B. Che Man, Determination of extra virgin olive oil in quaternary mixture using FTIR spectroscopy and multivariate calibration, Spectroscopy. 26 (2011) 203-211.

[10] X. Sun, W. Lin, Q. Shen, H. Luo, Detection and quantification of extra virgin olive oil adulteration with edible oils by FT-IR spectroscopy and chemometrics, Analytical Methods. 7 (2015), 3939-3945.

[11] H. Jabeur, A. Zribi, M. Bouaziz, Extra-virgin olive oil and cheap vegetable oils: distinction and detection of adulteration as determined by GC and chemometrics, Food Analytical Methods. 9 (2016) 712-723.

[12] C. Ruiz-Samblás, F. Marini, L. Cuadros-Rodríguez, A. González-Casado, Quantification of blending of olive oils and edible vegetable oils by triacyglycerol fingerprint gas chromatography and chemometric tools, Journal of Chromatography B. 910 (2012) 71-77.

[13] R. Aroca-Santos, J.C. Cancilla, A. Pérez-Pérez, A. Moral, J.S. Torrecilla, Quantifying binary and ternary mixtures of monovarietal extra virgin olive oils with UV-vis absorption and chemometrics, Sensors and Actuators B: Chemical. 234 (2016) 115-121.

[14] C. Ruiz-Samblás, C. Arrebola-Pascual, A. Tres, S. Van Ruth, L. Cuadros-Rodríguez, Authentication of geographical origin of palm oil by chromatographic fingerprinting of triacyglycerols and partial least square-discriminant analysis, Talanta. 116 (2013) 788-793.

[15] R.H.V. Corley, P.B. Tinker, The origin and development of the oil palm industry in: R.H.V. Corley, P.B. Tinker (Eds), The Oil Palm, Blackwell Science, Oxford (UK), 2003, pp. 1-26.

[16] G. Van Duijin, Traceability of the palm oil supply chain, Lipid Technology. 25 (2013) 15-18.

[17] P. Toy Gee, Analytical characteristics of crude and refines palm oil and fractions, European Journal of Lipid Science and Technology. 109 (2007) 373-379.

[18] B.M. Siddique, A. Ahmand, M.H. Ibrahim, S. Hena, M. Rafatullah, M.O. A.K, Physico-chemical properties of blends of palm olein with other vegetable oils, Grasas y aceites. 61 (2010) 423-429.

[19] RSPO, Roundtable on Sustainable Palm Oil [accesed July 18 2016], http://www.rspo.org/about.

[20] M. Leegwater, G. Van Duijin, Traceability of RSPO-certified Sustainable Palm Oil in: O.M. Lai, C.P. Tan, C.C. Akoh (Eds), Palm Oil, production, processing, characterization and uses, AOCS Press, Urbana (IL), 2012, pp. 713-736.

[21] Fedepalma, Federación Nacional de Cultivadores de Palma de Aceite (Spanish) [accesed July 18 2016] http://web.fedepalma.org.

[22] K.W. Kai Seng, M.N. Shamsudin, A.A Abdel Hammed, The Economics of Malaysian Palm oil production, marketing and utilization in: O.M. Lai, C.P. Tan, C.C. Akoh (Eds), Palm Oil, production, processing, characterization and uses, AOCS Press, Urbana (IL), 2012, pp. 211- 235.

[23] A. Rohman, Y.B. Che Man, Fourier transform infrared (FTIR) spectroscopy for analysis of extra virgin olive oil adulterated with palm oil, Food Research International. 43 (2010) 886-892.

[24] A. Rohman, Y.B. Che Man, Palm oil analysis in adulterated sesame oil using chromatography and FTIR spectroscopy, European Journal Lipid Science Technology. 113 (2011) 522-527.

[25] Biedermann M., Grob K., Mariani C. Transesterification and on-line LC-GC for determining the sum of free and esterified sterols in edible oils and fats. Eur J Lipid Sci Tech. 1993; 95:127–133.

[26] E. Pérez Castaño, C. Ruiz Samblás, S. Medina Rodríguez, V. Quirós Rodríguez, A.M. Jiménez Carvelo, L. Valverde Som, A. González Casado, L. Cuadros Rodríguez, Comparison of different analytical classification scenarios: application for the geographical origin of edible palm oil by sterolic (NP) HPLC fingerprinting, Analytical Methods. 7 (2015) 4192–4201.

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

[28] M. Steinbach, P.N. Tan, kNN: k-Nearst Neighbors in X. Wu, V. Kumar (Eds), The top ten algorithms in data mining, Chapman & Hall/CRC, Boca Raton (FL) , 2009, pp. 151-161 .

[29] D. Ballabio, V. Consonni, Classification tools in chemistry. Part 1: linear models. PLS-DA, Analytical Methods. 5 (2013) 3790-3798.

[30] L.Xu, S.M. Yan, C.B. Cai, X.P. Yu, One-class partial least squares (OCPLS) classifier, Chemometrics and Intelligent Laboratory System. 126 (2013) 1-5.

[31] J. Luts, F. Ojeda, R.V. de Plas, B. De Moor, S.V. Huffel, J.A.K. Suykens, A tutorial on support vector machine-based methods for classifications problems in chemometrics, Analytical Chimica Acta. 665 (2010) 129-145.

[32] P. Oliveri, G. Downey, Multivariate class modelling for the verification of food-authenticity claims, Trends in Analytical Chemistry. 35 (2012) 74-86.

[33] PLS-R Function Settings, Eigenvector Documentation Wiki [accessed July 21 2016], http://wiki.eigenvector.com/index.php?title=Pls

[34] F. Westad, M. Bevilacqua, F. Marini, Regression in: Federico Marini (Ed), Chemometrics in food chemistry, Elsevier, Oxford (UK), 2013, pp. 127-168.

[35] SVM Function Settings, Eigenvector Documentation Wiki [accessed July 25 2016], http://wiki.eigenvector.com/index.php?title= SVM_Function_Settings

[36] R.J. Hyndman, A.B Koehler, Another look at measures of forecast accuracy, International Journal of Foresting, 22 (2006) 679-688.

ARTÍCULO CIENTÍFICO

Classification of olive oils according to their cultivars based on second-order data using LC-DAD

(Artículo enviado en Julio 2018 a la revista: Talanta)

Elsevier Editorial System(tm) for Talanta
Manuscript Draft

Manuscript Number: TAL-D-18-02491

Title: Classification of olive oils according to their cultivars based on second-order data using LC-DAD

Article Type: Research Paper

Keywords: Olive oil authentication; Liquid chromatography; Three-way data classification method; Multivariate curve resolution; Random forest

Corresponding Author: Miss ANA MARIA JIMENEZ-CARVELO,

Corresponding Author's Institution: University of Granada

First Author: ANA MARIA JIMENEZ-CARVELO

Order of Authors: ANA MARIA JIMENEZ-CARVELO; Carlos M.CRUZ; Alejandro C.Olivieri; Antonio González-Casado; Luis Cuadros-Rodríguez

Classification of olive oils according to their cultivars based on second-order data using LC-DAD

Ana M. JIMÉNEZ-CARVELO✉[1], Carlos M. CRUZ[2], Alejandro C. OLIVIERI[3], ANTONIO GONZÁLEZ-CASADO[1], Luis CUADROS-RODRÍGUEZ[1]

[1] Department of Analytical Chemistry, [2] Department of Organic Chemistry, Faculty of Sciences, University of Granada, C/ Fuentenueva s/n, E-18071, Granada, Spain

[3] Instituto de Química Rosario (IQUIR–CONICET), Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531, S2002LRK, Rosario, Argentina

Abstract

Second-order data, acquired using liquid chromatography coupled to a diode array detector were used to classify extra virgin olive oils samples according to their cultivars. The chromatographic fingerprints from the epoxidised fraction were obtained using normal-phase liquid chromatography. To reduce the data matrices two strategies were employed: (1) multivariate curve resolution-alternating least squares (MCR-ALS) and (2) a new strategy proposed in this work based on the fusion of the mean data profiles in both spectral and time domains. Several conventional chemometric tools were then applied to both raw and reduced data: principal component analysis (PCA), partial least-squares-discriminant analysis (PLS-DA), soft independent modelling of class analogies (SIMCA) and n-way partial least-squares-discriminant analysis (NPLS-DA). Furthermore, an emergent multivariate classification method known as random forest (RF) has been first applied to second-order data. It was shown that RF is more efficient than conventional tools. Indeed, the obtained sensibility, specificity and accuracy are 1.00, 0.92 and 0.95 respectively; these performance metrics are significantly better than the values found for the other methods.

Keywords

Olive oil authentication; Liquid chromatography; Three-way data classification method; Multivariate curve resolution; Random forest

✉ Corresponding author: phone: +34 958240797; fax: +34 958243328; email: amariajc@ugr.es

1. Introduction

Extra-virgin olive oil (EVOO) is a food which contains valuable bioactive compounds as tocopherols and tocotrienols (vitamin E), β-carotenes, sterols, or phenols, which confer cardioprotective, antioxidant and anti-inflammatory properties over the health of consumers [**1,2**]. Furthermore, it is mainly composed by triacylglycerols (more than 90%) having a high proportion of monounsaturated fatty acids, especially oleic acid (around 70%). Its chemical composition could vary depending on many factors such as cultivar, agronomic conditions, extraction process, and ripeness, among others [**3**]. EVOO have thus characteristic organoleptic properties [**4**] due to the presence of many different flavouring organic compounds.

This essential food represents a treasure into the Mediterranean diet, giving unique flavour and aroma to the dishes where it is employed. In the last few years, EVOO is gaining ground in high-quality cuisine, due to its broad spectrum in terms of organoleptic properties which allow choosing each cultivar for a specific flavour. Around 1700 olive varieties are being cultivated nowadays according to the World Catalogue of Olive Varieties of the International Olive Council (IOC). Nevertheless, only a handful of them are mostly used to produce olive oil [**5**].

In the last years, the main producers of olive oil have shown a special interest in the marketing of monovarietal olive oils as a way to improve the competitiveness, and to try to deal with the effects of the globalization process in the olive oil sector. The aim is to market high-quality olive oil with specific organoleptic characteristics, which reflect the effect of the cultivar and the geographical origin where it has been grown. A good strategy to take a prominent position over the competitors is to take advantage of the difference in chemical composition, organoleptic characteristics or the kind of cultivar of each EVOO, bearing a recognised quality-differentiated food seal as the 'Protected Designation Origin' (PDO) or 'Protected Geographical Indication' (PGI) according to European regulations [**6**], and also labelling the oil as monovarietal EVOO, i.e., an extra-virgin olive oil obtained from a single kind of olive fruit botanical variety. These credentials, enforced within the EU and being gradually expanded internationally via bilateral agreements between the EU and non-EU countries, add value to the final product and bring exclusivity to the consumer.

The 'arbequina' cultivar is a commonly botanical variety in Spain since the XVII century. The monovarietal olive oil obtained from arbequina olive fruits shows special

organoleptic properties in comparison with other olive varieties, characterized for their freshly and fruity aroma and for showing a slight pungency or even none. These particular organoleptic properties make this olive oil an appreciated product for a wide spectrum of consumers. In Spain there are some PDO concerns to Arbequina cultivars as 'Estepa' (South of Spain) [**7**], 'Les Garrigues' [**8**] or 'Siurana' [**9**] (North of Spain).

In this sense, proper analytical methods which enable to distinguish quickly and reliably cultivar olive oils are currently demanded. There are some works reporting the classification of EVOO according to its cultivar using spectroscopic techniques [**10**], liquid chromatography [**11,12,13,14**] or gas chromatography [**15,16,17**]. Nevertheless, all these works are based on the quantification of specific compounds or on the study of the profile of a family of components such as chlorophylls, sterols, fatty acids and phenolic compounds.

On the other hand, it is possible to develop a global method for the classification of EVOO according to its cultivar by applying the chromatographic fingerprinting methodology which combines second-order data with chemometric tools. Conventionally, second-order data have been used for the quantification of compounds due to what is known as 'the second-order advantage', i.e., 'the analytes can be quantitated in the presence of uncalibrated interfering substances'. Therefore, only small sets of pure compounds are required for building the calibration model, instead of large calibration sets containing all possible interfering substances. The main algorithms employed to process these data are: (i) parallel factor analysis (PARAFAC) [**18**], (ii) multivariate curve resolution-alternating least squares (MCR-ALS) [**19**] and (iii) unfolded or multidimensional partial least-squares with residual bilinearization (UPLS-RBL or NPLS-RBL) [**20**].

Nevertheless, the application of this kind of data to build multivariate classification models for authentication of olive oils has not been extensively explored. The literature reports some studies applying PARAFAC together with unfolded principal component analysis (UPCA) to discriminate between commercial samples of virgin and pure olive oils [**21**], to detect adulterations in EVOO samples from the PDO [**22**], or PARAFAC with unfolded partial least-squares-discriminant analysis (UPLS-DA) to detect adulteration of olive oils with other vegetable oils and to quantify the proportion in binary blends [**23**].In all these studies, fluorescence spectroscopy was mainly employed. As far as we know, no studies have been reported where these algorithms are combined

with chromatographic data and traditional supervised pattern recognition methods such as partial least-squares discriminant analysis (PLS-DA) and soft independent modelling of class analogies (SIMCA), or with recently introduced classification methods such as random forest (RF). Only few applications are known in the food field with second-order data to authenticate the cultivar of extra-virgin olive oils.

The aim of this study is to discriminate between arbequina extra-virgin olive oil from extra-virgin olive oils from other cultivars, using three-way data to develop multivariate classification methods. For this purpose, we have developed a quick analytical method using high performance liquid chromatography coupled to a UV absorption diode array detector (HPLC-DAD). The second-order data were processed with PLS-DA, SIMCA and RF, in their original format or by first reducing them using MCR and a newly proposed approach. In addition, a set of quality metrics: (i) sensitivity, (ii) specificity, (iii) positive (or precision) and negative predictive values, (iv) Youden index, (v) positive and negative likelihood ratios, (vi) classification odds ratio; (vii) F-measure (or F-score), (viii) discriminant power, (ix) efficiency (or accuracy), (x) AUC (area under the receiver operating curve), (xi) G-mean; (xii) Matthews correlation coefficient and (xiii) Kappa coefficient, were used to assess the performance of the classifications.

2. Materials and methods

2.1. Chemicals and reagents

HPLC-grade solvents (n-hexane, isopropanol, methanol and tert-butyl methyl ether (TBME)) were purchased from VWR International Eurolab, S.L. (Barcelona, Spain).

Other reagents, sodium methoxide (MeONa), citric acid monohydrate, and anhydride sodium sulphate were provided by Merck (Darmstadt, Germany), sodium sulphate anhydrous was provided by Panreac, S.L (Barcelona, Spain) and 3-chloroperbenzoic acid was purchased from Sigma-Aldrich (Missouri, USA).

2.2. Samples

Sixty-four single-variety extra virgin olive oil samples (EVOO) of different regions from Spain and olive fruit varieties were analysed. The samples were obtained directly from local providers. More specifically, 20 samples were from 'arbequina' fruit variety and 44 samples were from different fruit varieties which include: 'picual', 'hojiblanca', 'cornicabra', 'frantoio', 'koroneiki', 'picudo', 'royal', 'loaime', 'lechin', 'lucio', 'arbosana'

and 'manzanilla'. Table 1 summarizes the different EVOO and the number of samples analysed.

Table 1. Classes and olive fruit varieties of extra virgin olive oil analysed.

Class	Fruit varieties	Nº samples
'Arbequina' (20 samples)	'arbequina'	20
'Non-arbequina' (44 samples)	'picual'	10
	'hojiblanca'	4
	'cornicabra'	5
	'frantoio'	3
	'koroneiki'	3
	'picudo'	4
	'royal'	3
	'loaime'	3
	'lechin'	1
	'lucio'	3
	'arbosana'	2
	'manzanilla'	3
	Total	**64**

2.3. Sample preparation

First a transesterification reaction was applied to the EVOO samples. This reaction is a modification of the original procedure described by Bierdemann et al. [24]. For further information regarding to this modification see references [25,26]. Then, the methyl-transesterification fraction of the EVOO samples was epoxidised as follows: 1000 µL of the transesterified fraction were added to a 10 mL tube and mixture with 1000 µL of a solution of 5% (m/v) 3-chloroperbenzoic acid in TBME. The tube was stirred for 20 s and then allowed to stand for 10 min. Next, 4 mL n-hexane and 1 mL 20% sodium sulphate anhydrous in water were added, and the mixture was shaken. The aqueous phase was removed with a Pasteur pipet and finally the organic fraction was filtered using a syringe filter of polytetrafluoroethylene (PTFE) membrane with a 0.22 µm pore diameter. The solution was stored in cold until analysis.

For chromatographic analysis, 200 μL of the stored solutions was transferred to a 2 mL HPLC vial. The epoxidisation step was carried out to enhance the difference between arbequina EVOOs and the ones from other cultivars.

2.4. Instrumentation

The chromatographic analysis was carried out with an Agilent 1100 series liquid chromatography (Santa Clara, CA) equipped with a G1316A column thermostat, G1311A quaternary pump, a G1379A degasser and a G1313A autosampler. Detection was performed with a G1315B diode-array detector (DAD). Agilent ChemStation software (rev.A.09.03 [1417]) for HPLC systems was used.

2.5. Chromatographic analysis

The chromatographic fingerprint from the epoxidised fraction was obtained by HPLC-DAD using a column Lichrospher® 100 CN (250×4 mm, i.d, 4 μm) provided by Merck (Darmstadt, Germany). During the analysis the column temperature was constant at 30 ºC. Isocratic chromatographic conditions were employed using a mixture of n-hexane/isopropanol (96:4, v/v) as mobile phase at a flow rate of 1.2 mL min−1. The injection volume was 20 μL and the run time was only 8 min. The DAD collected spectra every 2 s in the range 190-400 nm, each 1 nm.

2.6. Chemometrics

The raw data files from each chromatogram were exported in 'comma separated value' (CSV) format, and then converted to MATLAB format (version R2013b). The dimension of the matrix for each sample was of 1343×211 where 1343 is the number of rows corresponding to the number of elution times and 211 is the number of absorbance spectra recorded. It is important to notice that the chromatographic fingerprints from the epoxidised fraction were reproducible from sample to sample due to the short chromatographic run time (3-4 min); for this reason, it was not necessary to apply any alignment procedure.

The original dataset was randomly split into a training set, which was composed of 44 EVOO samples (14 EVOO samples from arbequina cultivar and 30 from non-arbequina cultivar) and an external validation set was made up with 20 EVOO samples (6 EVOO samples from arbequina cultivar and 14 from non-arbequina cultivar).

MCR-ALS and NPLS-DA were applied using the interface MVC2 MATLAB toolbox, freely available on the internet [**27**]. Conventional multivariate chemometrics pattern recognition such PCA, SIMCA and PLS-DA, were employed using PLS_Toolbox ver 8.5.1 (Eigenvector Research Inc., Wenatchee, WA). RF was employed using perClass ver 4.7 (Delft, Netherlands). All the interface graphics, MVC2 toolbox, PLS Toolbox and perClass were designed for MATLAB software (Mathworks Inc., Natick, MA, USA).

3. Results and discussion

A two-way data array was recorded for each EVOO sample. Figure 1 illustrates a chromatographic-spectral landscape for an EVOO sample from 'cornicabra' cultivar.

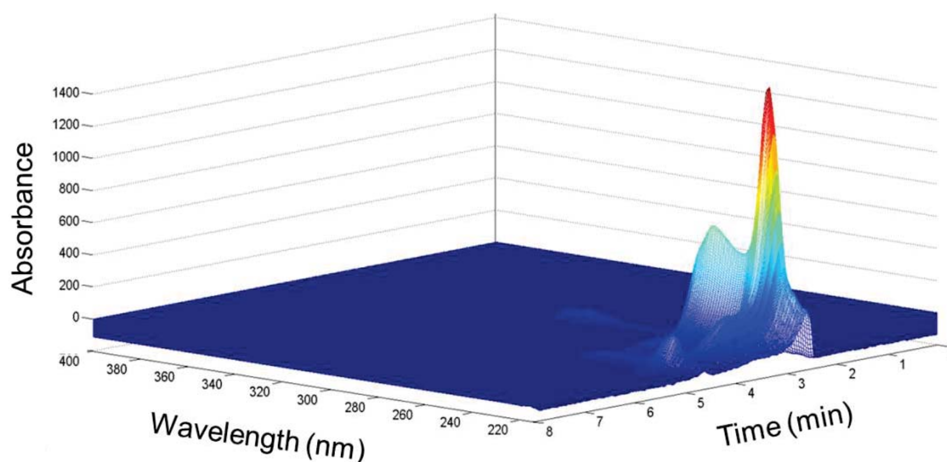


Figure 1. Time-wavelength chromatographic landscape of an extra virgin olive oil from 'cornicabra' cultivar.

Variable reduction

Two strategies of variable reduction were employed: (i) strategy 1, named "decomposition and vector fusion" (DVF) and (ii) strategy 2, using MCR-ALS for the resolution into individual components. Figure 2 shows a flow chart of the two strategies performed.

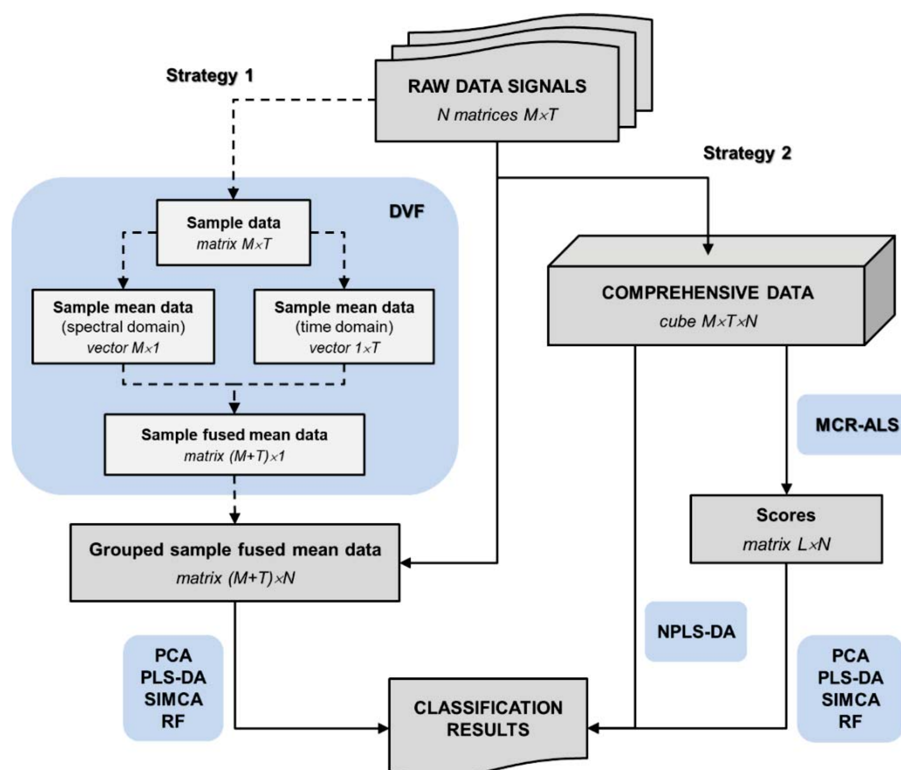


Figure 2. Flow chart showing the strategies applied for the treatment of the two-way data (N = number of objects (EVOO samples); M = number of variables in the spectral domain (wavelengths of the UV absorption spectrum); T = number of variables in the time domain (retention times of the chromatogram); L = number of latent variables (principal components)).

(i) Strategy 1 for variable reduction: DVF

For each sample, the corresponding mean vectors in both time and spectral domains were obtained. In this way, two individual vectors per sample were computed, a mean vector of size 1343×1 (time domain) and another mean vector of size 211×1 (spectral domain). These two vectors were then fused, so that the resulting fused vector was composed of 1544 variables. Finally, the fused vectors for all samples were grouped in a single matrix of dimension 64×1544 (64 samples and 1544 variables). Figure 3 displays the mean vectors in the time and spectral domain for an EVOO sample from 'cornicabra' cultivar, respectively. Figure 4 shows the overlay of the fused mean vectors from the 64 EVOO samples.

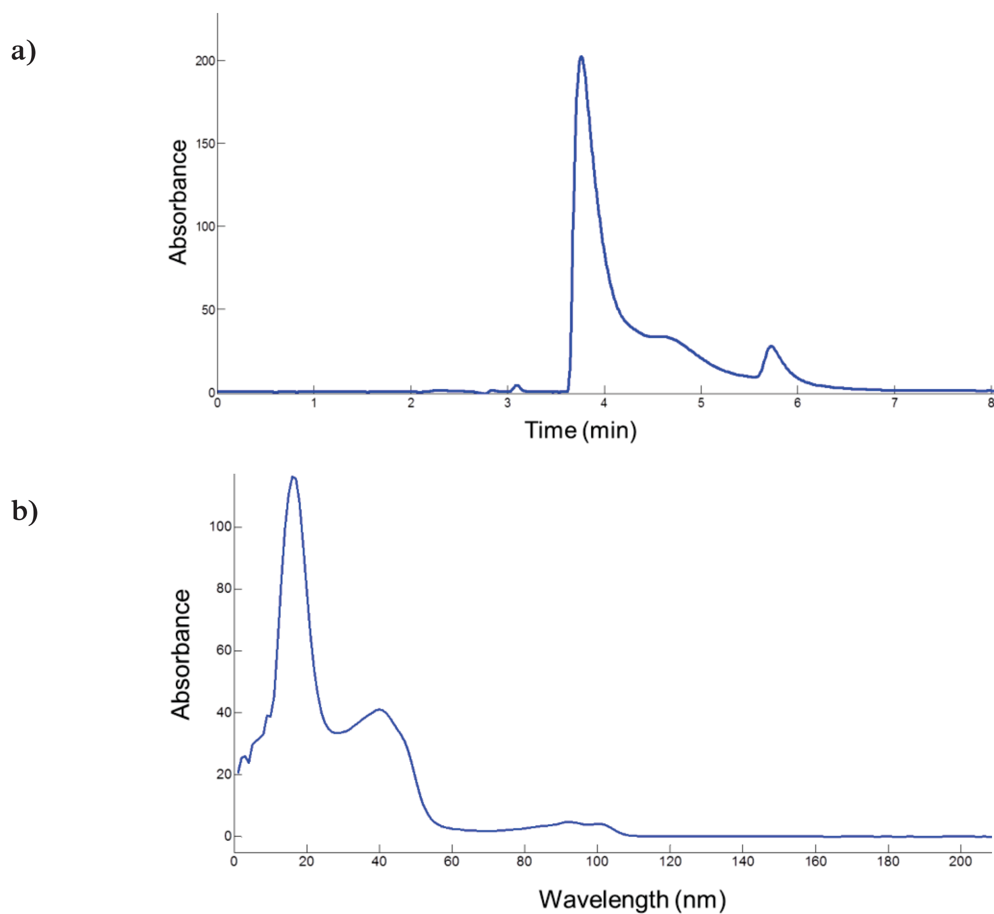


Figure 3. Plot of the mean vectors for an EVOO sample from 'cornicabra' cultivar: (a) mean vector in the time domain and (b) mean vector in the spectral domain.

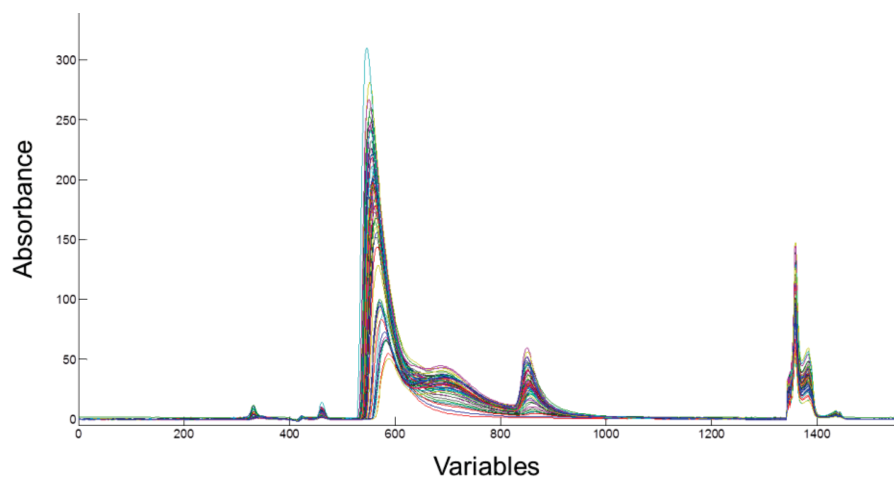


Figure 4. Overlay of fused mean vectors of all EVOO samples.

(ii) Strategy 2 for variable reduction: MCR-ALS

The successful application of this algorithm requires that enough selectivity exists in the spectral domain. If the samples show similar spectra, they cannot be resolved into individual components using MCR-ALS. In these cases, if the chromatograms are reproducible, matrix augmentation can be performed along the spectral domain before MCR-ALS is applied [**28**].

MCR-ALS was applied to the row-wise augmented matrix (i.e., along the spectral domain). The number of components was estimated using principal component analysis of the augmented data matrix under a series of constraints: non-negativity in both domain (time and spectral) and none unimodality. According to the PCA results, 8 components were selected, which explained 99.94% of the data variance. After MCR-ALS decomposition, the chromatographic fingerprint information was arranged into a matrix of dimension 64×8 (64 samples and 8 components), which was subsequently processed with PCA, PLS-DA, SIMCA and RF for classification purposes.

Exploratory analysis

Two PCA models were built using each of the matrices computed by strategies 1 and 2, to test if there was some natural grouping in the data set. Both PCA models were built with four principal components (PCs) (98.96% and 98.08% of explained variance for each strategy, respectively). They grouped the samples in a similar way.

Figure 5 shows the scores score-score plot on the PC4 vs PC1 plane. PC4 and PC1 explained 4.08% and 68.18% of the variance, respectively. Two groups of EVOO cultivars are distinguished: the positive region of PC4 mainly groups the EVOOs of the arbequina cultivar, while the positive region of PC1 clusters the EVOOs of the other cultivars.

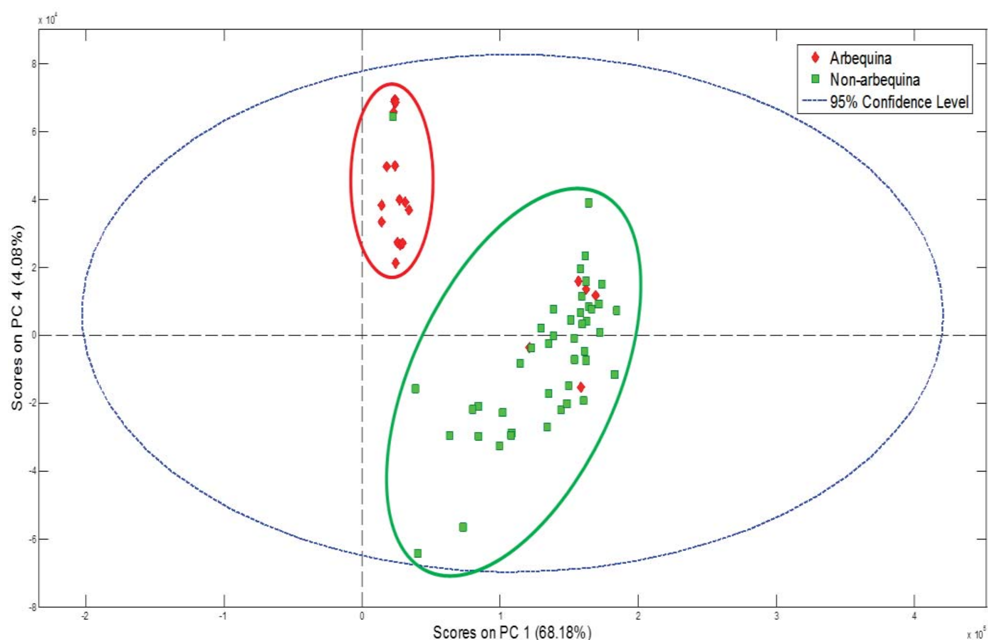


Figure 5. PC4 vs PC1 plot from the matrix obtained from application of MCR of the epoxidised fraction of the 64 EVOO samples from different cultivars.

Conventional multivariate classification methods

As mentioned in the Introduction, multivariate classification methods using second-order data for the authentication of the cultivar kind of EVOOs are scarce, and the most commonly applied chemometric in these cases is linear discriminant analysis (LDA).

In the present report, three classification models were developed: (i) two well-known classification methods (PLS-DA and SIMCA) using the resulting matrices obtained from the application of both strategies 1 and 2 (see above), and (ii) NPLS-DA over the raw three-way data array . The main aim was to test whether there was significant difference between the classification methods usually applied to first-order (PLS-DA and SIMCA) and second-order data (NPLS-DA) when the chromatographic fingerprinting methodology is applied.

For classification purposes with PLS-DA, the class "arbequina" was indexed with the value 1 and the class "non-arbequina" with the value 0. The classification threshold was established by the software around of the value 0.6 for the arbequina class.

The classification of the samples with SIMCA was carried out from both Q-reduced (Q) and Hotelling T^2-reduced values. The classification region for the arbequina class was established according to Q and T^2 values equal to 1, meaning that a sample must take values lower than 1 to be classified in the arbequina class.

Table 2 shows the specifications of the PLS-DA and SIMCA models.

Table 2. Characteristics of the PLS-DA and SIMCA models

	PLS-DA		SIMCA			
	LVs	% var	PCs 'Arb-Class'	% var	PCs 'nArb-Class'	% var
(a) Strategy 1						
	4	98.83	4	99.48	5	99.60
(b) Strategy 2						
	5	97.29	6	99.92	6	99.87

The raw three-way data array was analysed using NPLS-DA . The estimated number of latent variables (LVs) was 15 according to the leave-one-out cross-validation method. As in the PLS-DA model, the "arbequina" class was denoted using the number 1 and the "non-arbequina" class using the number 0.

The prediction results from both strategies were the same. Table 3 shows the results of PLS-DA, SIMCA and NPLS-DA models and Table 4 presents the classification quality metrics calculated from the prediction results on the external validation set.

In Table 3, the wrongly classified samples have been highlighted. As can be seen, PLS-DA and NPLS-DA classification models are more efficient than SIMCA. Furthermore, the former two models misclassified the same samples, which might indicate that the labels of the containers of these EVOOs samples are incorrect, i.e. that these samples are not monovarietal EVOOs, but blends of extra virgin olive oils from different olive fruit varieties (or 'coupages').

Table 3. Prediction results of arbequina and non-arbequina classification from the external validation set using PLS-DA, SIMCA and NPLS-DA.

Class	Sample number	Class Ref	PLSDA	SIMCA	NPLS-DA
			Clas Pred	**Clas Pred**	**Clas Pred**
Arbequina (Arb)	28	1	0	0	0
	30	1	0	0	0
	67	1	1	1	1
	69	1	1	1	1
	70	1	1	1	1
	74	1	1	0	1
Non-arbequina (nArb)	1	0	0	0	0
	10	0	0	0	0
	14	0	0	0	0
	16	0	0	0	0
	19	0	0	0	0
	34	0	0	0	0
	37	0	0	0	0
	39	0	0	1	0
	48	0	0	0	0
	51	0	0	0	0
	59	0	0	0	0
	61	0	0	0	0
	73	0	1	1	1

Ref: reference; Pred: predicted

Table 4. Values of the quality metrics from the conventional multivariate classification methods.

Performance features	PLS-DA	SIMCA	NPLS-DA
Sensibility (Recall)	0.67	0.50	0.67
Specificity	0.92	0.85	0.92
Positive predictive value (Precision)	0.80	0.60	0.80
Negative predictive value	0.86	0.79	0.86
Youden index	0.59	0.35	0.59
Positive likelihood ratio	8.67	3.25	8.67
Negative likelihood ratio	0.36	0.59	0.36
Classification odds ratio	24.00	5.50	24.00
F-measure	0.73	0.55	0.73
Discriminant power	0.76	0.41	0.76
Efficiency (or Accuracy)	0.84	0.74	0.84
AUC (Correctly classified rate)	0.79	0.67	0.79
G-mean	0.78	0.65	0.78
Matthews correlation coefficient	0.62	0.36	0.62
Kappa coefficient	0.62	0.36	0.62

Emergent multivariate classification methods

Random forest (RF) was first employed to process the second-order data. This algorithm is a combination of several prediction trees, which then selects the best split at each node among a random selection of predictor variables. RF shows significant advantages about other more applied classification methods and it is able to build a more robust classification model than other conventional algorithms. Moreover, RF readily handles larger numbers of predictors and the cross-validation is unnecessary because it generates an internal unbiased estimate of the generalization error (test error) as the forest building progresses. The potential of RF for modelling linear and nonlinear multivariate calibration allows to be used for feature selection too, with two different objectives: (i) to find the subset of features with the minimum possible generalization error, or (ii) to select the smallest possible subset with a given discrimination capability [**29**].

Both classification models using the reduced data sets by strategies 1 and 2 achieved the

same results. In both cases, 20 trees were combined to perform the prediction of the classes of the EVOO samples. Table 5 shows the obtained classification contingency table on the external validation data set, and table 6 displays the prediction results and the different classification quality metrics for the RF models, respectively.

Table 5. Contingency charts for the RF classification models**.**

		Decision of the classifier		
		Arb class	**nArb class**	*Total*
True class	**Arb class**	6	0	*6*
	nArb class	2	12	*14*
	Total	*8*	*12*	*20*

Table 6. Values of the classification quality metrics from the RF models.

Performance features	RF (strategy 1)	RF (strategy 2)
Sensibility (Recall)	1.00	1.00
Specificity	0.92	0.92
Positive predictive value (Precision)	0.76	0.76
Negative predictive value	1.00	1.00
Youden index	0.92	0.92
Positive likelihood ratio	13.00	13.00
Negative likelihood ratio	0.00	0.00
Classification odds ratio	–	–
F-measure	0.92	0.92
Discriminant power	–	–
Efficiency (or Accuracy)	0.95	0.95
AUC (Correctly classified rate)	0.96	0.96
G-mean	0.96	0.96
Matthews correlation coefficient	0.89	0.89
Kappa coefficient	0.88	0.88

The hyphen "–" indicates that the performance feature cannot be determined

The RF results are significantly better than the obtained ones from the previously applied conventional classification methods. The sensibility, specificity and efficiency from PLS-DA and NPLS-DA were 0.67, 0.92 and 0.84, respectively, while the same performances featured by the RF model were 1.00, 0.92 and 0.95, respectively. This suggests that the analysis of second-order data with to a powerful algorithm such as RF is a promising methodology to authenticate cultivars of EVOO samples.

4. Conclusions

The potential of second-order (or) fingerprint data obtained using LC-DAD to identify and discriminate extra-virgin olive oils from 'arbequina' botanical variety in respect of other varieties of milled olive fruits has been proved. A binary classification method has been developed and properly validated by applying three multivariate classification algorithms, including two widely-recognised methods (partial least-squares-discriminant analysis, PLS-DA, and soft independent modelling of class analogies, SIMCA) and a third one (random forest, RF) which is much less known and has been first used on second-order data. Surprisingly RF has shown itself to be the more efficient one in validation, yielding values of sensibility, specificity and accuracy of 1.00, 0.92 and 0.95, respectively, which are significantly better than the values found for the other methods.

Before building multivariate classification models, the raw three-way data matrices have been reduced by applying two strategies: (1) multivariate curve resolution-alternating least squares (MCR-ALS), and (2) a new strategy named "decomposition and vector fusion" (DVF) which has been proposed in this work and based on the fusion of the mean vector obtained from the signal profiles in both spectral and time domains. No differences on the performance classification are found when both strategies are applied.

References

[1] M.I. Covas, Olive oil and the cardiovascular system, Pharmacol. Res. 55 (2007) 175-186. https://doi.org/10.1016/j.phrs.2007.01.010

[2] R.W. Owen, A. Giacosa, W.E. Hull, R. Haubner, B. Spiegelhalder, H. Bartsch, The antioxidant/anticancer potential of phenolic compounds isolated from olive oil, Eur. J. Cancer 36 (2000) 1235-1247. https://doi.org/10.1016/S0959-8049(00)00103-9

[3] R. Aparicio, G. Luna, Characterisation of monovarietal virgin olive oils, Eur. J. Lipid Sci. Technol. 104 (2002) 614-627. https://doi.org/10.1002/1438-9312(200210)104:9/10<614::AID-EJLT614>3.0.CO;2-L

[4] C. Campestre, G. Angelini, C. Gasbarri, F. Angerosa, The compounds responsible for the sensory profile in monovarietal virgin olive oils, Molecules 22 (2017) 1833. https://doi.org/10.3390/molecules22111833

[5] G. Bertolini, G. Prevost, C. Messeri, G. Carignani, Olive germplasm: cultivars and world-wide collections, FAO, Rome, 1998.

[6] Regulation (EU) No 1151/2012 of the European Parliament and of the Council of 21 November 2012 on quality schemes for agricultural products and foodstuffs. Official Journal of the European Union. (2012) L343/1-29.

[7] Commission Implementing Regulation (EU) Nº 2017/597 of 15 March 2017 approving non-minor amendments to the specification for a name entered in the register of protected designations of origin and protected geographical indications (Estepa (PDO)). Official Journal of the European Union. (2017) L81/15.

[8] Commission Regulation (EU) Nº 1902/2004 of 29 October 2004 amending the specification of a name appearing in the Annex to Regulation (EC) No 1107/96 on the registration of geographical indications and designations of origin (Les Garrigues). Official Journal of the European Union. (2004) L328/1-73.

[9] Commission Regulation (EU) Nº 2156/2005 of 23 December 2005 amending the specification of a protected designation of origin listed in the Annex to Regulation (EC) No 1107/96 (Siurana) (PDO). Official Journal of the European Union. (2005) L342/1-54.

[10] N. Sinelli, M. Casale, V. Di Egidio, P. Oliveri, D. Bassi, D. Tura, E. Casiraghi, Varietal discrimination of extra virgin olive oils by near and mid infrared spectroscopy, Food Res. Int. 43 (2010) 2126-2131. https://doi.org/10.1016/j.foodres.2010.07.019

[11] A. Cichelli, G.P. Pertesana, High-performance liquid chromatography analysis of chlorophylls, pheophytins and carotenoids in virgin olive oils: chemometric approach to variety classification, J. Chromatogr. A 1046 (2004) 141-146. https://doi.org/10.1016/j.chroma.2004.06.093

[12] A. Bajoub, S. Medina-Rodríguez, M. Gómez-Romero, E.A. Ajal, M.G. Bagur-González, A. Fernández-Gutiérrez, A. Carrasco-Pancorbo, Assesing the varietal origin of extra-virgin olive oil using liquid chromatography fingerprints of phenolic compound, data fusion and chemometrics, Food Chem. 215 (2017) 245-255. https://doi.org/10.1016/j.foodchem.2016.07.140

[13] N.P. Kalogiouri, R. Aalizadeh, N.S. Thomaidis, Application of an advanced and wide scope non-target screeninworkflow with LC-ESI-QTOF-MS and chemometrics for the classification of the Greek olive oil varieties, Food Chem. 256 (2018) 53-61. https://doi.org/10.1016/j.foodchem.2018.02.101

[14] M.B. Mohamed, F. Gusami, S.B. Ali, F. Radhouani, J. Faghim, T. Triki, N.G. Kammoun, C. Baffi, L. Lucini, C. Benincasa, The LC-MS/MS characterization of phenolic

compounds in leaves allows classifying olive cultivars grown in South Tunisia, Biochem. Syst. Ecol. 78 (2018) 84-90. https://doi.org/10.1016/j.bse.2018.04.005

[15] M.R. Alves, S.C. Cunha, J.S. Amaral, J.A. Pereira, M.B Oliveira, Classification of PDO olive oils on the basis of their sterol composition by multivariate analysis, Anal. Chim. Acta 549 (2005) 166-178. https://doi.org/10.1016/j.aca.2005.06.033

[16] G. Gurdeniz, B. Ozen, F. Tokatli, Classificaton of Turkish olive oils with respect to cultivar, geographic origin and harvest year, using fatty acid profile and mid-IR spectroscopy, Eur. Food Res. Technol. 227 (2008) 1275-1281. https://doi.org/10.1007/s00217-008-0845-7

[17] A. Kritioti, G. Menexes, C. Drouza, Chemometric characterization of virgin olive oils of the two major Cypriot cultivars based on their fatty acid composition, Food Res. Int. 103 (2018) 426-437. https://doi.org/10.1016/j.foodres.2017.10.064

[18] R. Bro, Parafac. Tutorial and applications, Chemom. Intell. Lab. Syst. 38 (1997) 149-171. https://doi.org/10.1016/S0169-7439(97)00032-4

[19] R. Tauler, Multivariate curve resolution applied to second order data, Chemom. Intell. Lab. Syst. 30 (1995) 133-146. https://doi.org/10.1016/0169-7439(95)00047-X

[20] A.C. Olivieri, On a versatile second-order multivariate calibration method based on partial least-squares and residual bilinearization: second-order advantage and precision properties, J. Chemometrics. 19 (2005) 253-265. https://doi.org/10.1002/cem.927

[21] F. Guimet, J. Ferré, R. Boqué, F.X. Rius, Application of unfold principal component analysis and parallel factor analysis to the exploratory analysis of olive oils by means of excitacion-emission matrix fluorescence spectroscopy, Anal. Chim. Acta. 515 (2004) 75-85. https://doi.org/10.1016/j.aca.2004.01.008

[22] F. Guimet, J. Ferré, R. Boqué, Rapid detection of olive-pomace oil adulteration in extra virgin olive oils from the protected denomination of origin "Siurana" using excitation-emision fluorescence spectroscopy and three-way methods of analysis, Anal. Chim. Acta. 544 (2005) 143-152. https://doi.org/10.1016/j.aca.2005.02.013

[23] I. Durán Merás, J. Domínguez Manzano, D. Airado Rodríguez, A. Muñoz de la Peña, Detection and quantification of extra virgin olive oil adulteration by means of autofluorescence excitacion-emission profiles combined with multi-way classification, Talanta. 178 (2018) 751-762. https://doi.org/10.1016/j.talanta.2017.09.095

[24] M. Bierdermann, K. Grob, Mariani, Transesterification and on-line LC-GC for determining the sum of free and esterified in edible oils and fats, Eur. J. Lipid Sci. Technol. 95 (1993) 127-133. https://doi.org/10.1002/lipi.19930950403

[25] A.M. Jiménez-Carvelo, E. Pérez-Castaño, A. González-Casado, L. Cuadros-Rodríguez, One input-class and two input-class classifications for differentiating olive oil from other edible vegetable oils by use of the normal-phase liquid chrtomatography fingerprint of the methyl-transesterified fraction, Food Chem. 221 (2017) 1784-1791. https://doi.org/10.1016/j.foodchem.2016.10.103

[26] A.M. Jiménez-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. 100 (2) (2017) 345-350. https://doi.org/10.5740/jaoacint.16-0411

[27] A.C. Olivieri, H.L. Wu, R.Q. Yu, MVC2: a matlab graphical interface toolbox for second-order multivariate calibration, Chemom. Intell. Lab. Syst. 96 (2009) 246-251. https://doi.org/10.1016/j.chemolab.2009.02.005

[28] P. L. Pisano, M.F. Silva, A.C. Olivieri, Exploration of liquid chromatographic-diode array for Argentinean wines by extended multivariate curve resolution, Chemom. Intell. Lab. Syst. 132 (2014) 1-7. https://doi.org/10.1016/j.chemolab.2013.12.010

[29] C. Ruiz-Samblas, J.M. Cadenas, D.A. Pelta, L. Cuadros-Rodríguez, Application of data mining methods for classification and prediction of olive oil blends with other vegetable oils, Anal. Bioanal. Chem. 406 (2014) 2591-2601. https://doi.org/10.1007/s00216-014-7677-z

I.8. Discusión

A la vista de los resultados obtenidos, podemos concluir que la huella dactilar cromatográfica de la fracción metil-transesterificada del aceite de oliva contiene la suficiente información para poder diferenciarlo de otros aceites vegetales, detectar adulteraciones y determinar la proporción del mismo en mezclas con otros aceites vegetales, reduciendo de esta forma el tiempo y el costo del análisis de la cromatografía convencional.

Han sido aplicadas tres estrategias de desarrollo de métodos de clasificación multivariante y además una de ellas (*pseudo* dos clases de entrada o *p*2iC) ha sido propuesta y aplicada por primera vez en el trabajo de esta tesis doctoral, obteniéndose buenos resultados en las clasificaciones.

Todos los estudios realizados son de gran importancia en el sector del aceite de oliva, ya que el método analítico desarrollado aplicando la metodología de huellas dactilares cromatográficos en conjunción con herramientas quimiométricas, permite **autentificar el aceite de oliva** en un **único análisis**, sin necesidad de determinar alrededor de los 50 parámetros tal y como recoge la legislación actual para poder detectar fraudes en el aceite de oliva. Este tipo de análisis permitiría llevar a cabo un control de calidad del aceite de oliva en las empresas del sector de forma rápida y sencilla.

Además, el uso de la huella dactilar cromatográfica de la fracción metil-transesterificada se ha permitido el desarrollo de un **método global** que podría ser aplicado en laboratorios de rutina para la detección de mezclas de aceite de oliva con más de 10 aceites vegetales no oliva diferentes en diferentes proporciones.

COMUNICACIONES A CONGRESOS

- A.M. Jiménez Carvelo, E. Pérez-Castaño, L. Valverde Som, A. González Casado, L. Cuadros Rodríguez. **Diferenciación de aceites vegetales a partir de la huella cromatográfica (HPLC-CAD) de la fracción esterólica.** XIV Reunión del grupo regional andaluz de la sociedad española de química analítica (GRASEQA). **Póster.**

- A.M. Jiménez Carvelo, L. Cuadros Rodríguez. **Diferenciación de aceites vegetales a partir de la huella cromatográfica (HPLC-CAD) de la fracción esterólica.** Jornadas <<Avances en calidad y tecnología alimentaria>>. **Póster.**

- A.M. Jiménez Carvelo, L. Valverde Som, A. González Casado, L. Cuadros Rodríguez. **Determinación de la proporción de aceite de oliva en mezclas con otros aceites vegetales.** XVII Feria internacional del aceite de oliva e industrias afines (EXPOLIVA). **Póster.**

- A.M. Jiménez Carvelo, E. Pérez Castaño, A. González Casado, L. Cuadros Rodríguez. **Identificación de la presencia y determinación de la proporción de aceite de palma en mezclas con otros aceites vegetales. Aplicación de la estrategia de "huellas dactilares" cromatográficas.** XX Reunión de la sociedad española de química analítica (SEQA). **Póster.**

- A.M. Jiménez Carvelo, E. Pérez Castaño, A. González Casado, L. Cuadros Rodríguez. **Aplicación de la estrategia de clasificación multivariante de "pseudo" dos clases de entrada (una clase diana + una clase fantasma) en la autentificación de aceite de oliva.** XV Reunión del grupo regional andaluz de la sociedad española de química analítica (GRASEQA). **Oral.**

- A.M. Jiménez Carvelo, Carlos M. Cruz, A. González Casado, L. Cuadros Rodríguez, T. Koidis. **Classification of olive oils according to their cultivar.** 16th Euro Fed Lipid Congress. **Póster.**

Capítulo II

"Haz las cosas lo más simple que puedas, pero no te limites a lo simple"

Albert Einsten

CAPITULO II

'Huellas dactilares cromatográficas – fase invertida'

II.1. Presentación

Este capítulo recoge los resultados obtenidos al aplicar la metodología de huellas dactilares cromatográficas en la modalidad de fase invertida y los métodos de clasificación multivariante desarrollados con el **objetivo** de discriminar aceites de oliva de diferentes categorías de otros aceites vegetales comestibles.

Al igual que en el capítulo I, se aplicaron tres estrategias de clasificación aplicando métodos de análisis discriminante (PLS-DA, SVM y kNN) y métodos de modelado de clases (SIMCA y OCPLS).

Este capítulo derivó en un artículo publicado, cuya referencia es:

1. *FAST-HPLC FINGERPRINTING TO DISCRIMINATE OLIVE OIL FROM OTHER EDIBLE VEGETABLE OILS BY MULTIVARIATE CLASSIFICATION METHODS* (Journal of AOAC International, 2017, *100 (2),* 345-350).

El citado artículo fue consecuencia de una invitación personalizada del editor de la revista a uno de los directores de esta tesis doctoral para un número especial dedicado a la aplicación de la metodología de huellas dactilares espectroscópicas o cromatográficas, en conjunción con herramientas quimiométricas, para el desarrollo de métodos de clasificación enfocados a problemas relacionados con materiales agrícolas y productos alimenticios.

Seguidamente se realiza una breve introducción correspondiente a este capítulo.

II.2. Introducción

La cromatografía líquida de alta eficiencia o de altas prestaciones (HPLC) en la modalidad de trabajo de fase invertida (más comúnmente denominada fase reversa) es la más ampliamente utilizada cuando se aplica esta técnica instrumental. En esta modalidad la fase estacionaria es de carácter apolar y la fase móvil de carácter polar.

Las fases estacionares en esta modalidad están formadas por sílice modificada a la cual se unen grupos orgánicos frecuentemente octilos (C8) o octadecilos (C18). Estas modificaciones generan una superficie hidrofóbica que retienen preferentemente a los compuestos apolares. El grado de hidrofobicidad vendrá regido por la longitud de la cadena orgánica [**1**].

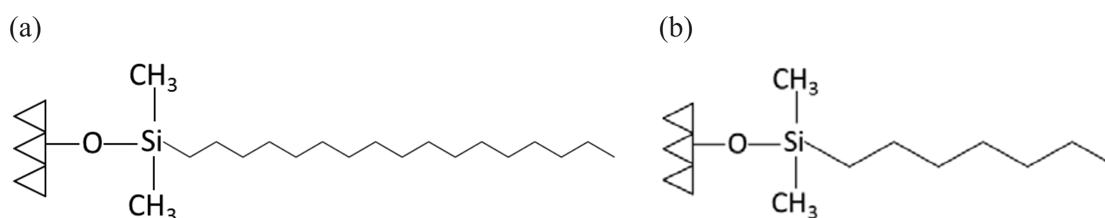


Figura 1. Fases estacionarias más frecuentes en las columnas en fase reversa: (a) C18 y (b) C8.

También son comunes las fases estacionarias de sílice modificada con uniones a grupos fenilos (Ph) o cadenas de hidrocarburos más cortas cómo etil (C2) o hexil (C6).

La fase móvil puede estar constituida por mezclas de diferentes disolventes, entre los que se pueden destacar acetonitrilo, metanol, acetona y agua por ser los más empleados. La capacidad de elución de la fase móvil se regula variando la proporción de agua y disolvente orgánico, de forma que cuanto mayor sea la proporción de agua existirá menor elución de los compuestos apolares, y por tanto, mayores tiempos de retención para éstos. Cuando en la fase móvil el agua se encuentra en una alta proporción, los compuestos cargados se disolverán en ésta lo que impide la interacción con la fase estacionaria. Para contrarrestar este proceso se añade un modificador en el agua, el cual anula las cargas, favoreciendo así las interacciones analito-fase estacionaria.

Para el análisis de grasas y/o aceites se ha aplicado principalmente esta modalidad de cromatografía, en la cual se utiliza mayoritariamente columnas de tipo C18 convencionales. Para las separaciones de los ácidos grasos libres y los ésteres metílicos de los ácidos grasos (FAMEs) el orden de elución de los compuestos es el mismo y se produce de menor a mayor número de carbonos, y dentro de compuestos con el mismo número de carbonos eluyen primero los que presentan mayor número de insaturaciones

[1] Pesek, J.J., Matyska, M.T. (2010). Reverse-phase Chromatography. En: Cazes, J. (Ed). Enclyclopedia of Chromatography. Editorial Taylor and Francis Group.

[**2**]. En el caso del análisis de triglicéridos el orden de elución se produce en función del valor del parámetro ECN (número equivalente de carbonos), de forma que eluyen de menor a mayor valor de ECN [**3**]. El parámetro ECN se define cómo:

$$ECN = CN - 2n$$

dónde CN es el número total de átomos de carbonos en las cadenas alquílicas (no se consideran los carbonos del esqueleto glicérico), y n es la suma de enlaces dobles presentes en dichas cadenas grasas.

Para la autentificación de aceite de oliva tradicionalmente se han aplicado métodos de análisis con una duración de más de diez minutos. En la tabla 1 se muestran algunos ejemplos de aplicación de la cromatografía líquida en la modalidad de fase reversa para autentificar el aceite de oliva.

Tabla 1. Estudios de autentificación de aceite de oliva aplicando cromatografía líquida en la modalidad de fase reversa.

Detector*	Objetivo	Columna	Tiempo (min)	[Ref]
FLD	Detectar adulteraciones	C18 (250 mm × 4.6 mm, 5 µm)	12	[**4**]
	Detectar adulteraciones	C18 (250 mm × 4.6 mm, 5 µm)	30	[**5**]
DAD	Clasificar acorde al origen geográfico	C18 (150 mm × 4.6 mm, 1.8 µm)	23	[**6**]

FLD: detector de fluorescencia; **DAD**: detector de fila de diodos.

[2] Lin, J.T., Mckeon, T.A. (2005). Reverse-phase HPLC separation of free fatty acids and their methyl esters. En: Lin, J.T., Mckeon, T.A. (Eds). HPLC of acyl lipids. Editorial HNB Publishing.

[3] Nikolova-Damyanova, B. (2005). Silver-ion HPLC of fatty acids and triacylglycerols. En: Lin, J.T., Mckeon, T.A. (Eds). HPLC of acyl lipids. Editorial HNB Publishing.

[4] Bakre, S.M., Gadmale, D.K., Toche, R.B., Gaikwad, V.B. (2015). Rapid determination of alpha tocopherol in olive oil adulterated with sunflower oil by reverse phase high-performance liquid chromatography. Journal of Food Science and Technology, *52 (5)*, 3093-3098.

[5] Chen, H., Angiuli, M., Ferrari, C., Tombari, E., Salvetti, G., Bramanti, E. (2011). Tocopherol speciation as first screening for the assessment of extra virgin olive oil quality by reverse-phase high-performance liquid chromatography/fluorescence detector. Food Chemistry, *125*, 1423-1429.

[6] Bajoud, A., Carrasco-Pancorbo, A., Ajal E.A., Ouazzani, N., Fernández-Gutiérrez, A. (2015). Potential of LC-MS phenolic profiling combined with multivariate analysis as an approach for the determination of the geographical origin of north Moroccan virgin olive oils. Food Chemistry, *166*, 292-300.

Tabla 1. *Continuación.*

Detector*	Objetivo	Columna	Tiempo (min)	[Ref]
DAD	Caracterizar los aceites de una denominación de origen protegida	C18 (250 mm × 4.6mm, 5 μm)	45	[**7**]
FLD+DAD	Clasificar acorde a la variedad de aceituna	C18 (150 mm × 4.6 mm, 1.8 μm)	23	[**8**]
RID	Detectar adulteraciones	C18 (250 mm × 4.6 mm, 5 μm)	34	[**9**]
	Clasificar acorde al origen geográfico	C18 (250 mm × 4 mm, 4 μm)	55	[**10**]
MS	Clasificar acorde a la variedad de aceituna	C18 (150 mm × 4.6 mm, 1.8 μm)	23	[**11**]
	Identificar marcadores de variedad de aceituna en aceites de oliva virgen extra	C18 (100 mm × 2.1 mm, 1.8 μm)	18	[**12**]

FLD: detector de fluorescencia; **DAD**: detector de fila de diodos; **RID:** detector de índice de refracción; **MS:** detector de espectrometría de masas.

[7] Nescatelli, R., Bonanni, R.C., Bucci, R., Magrì, A.L., Magrì, A.D., Marini, F. (2014). Geographical traceability of extra virgin olive oils from Sabina PDO by chromatographic fingerprinting of the phenolic fraction coupled to chemometrics. Chemometrics and Intelligent Laboratory Systems, *139*, 175-180.

[8] Bajoud, A., Medina-Rodríguez, S., Gómez-Romero M., Ajal E.A., Bagur-González M.G., Fernández-Gutiérrez, A., Carrasco-Pancorbo, A. (2017). Assessing the varietal origin of extra-virgin olive oil using liquid chromatography fingerprints of phenolic compound, data fusion and chemometrics. Food Chemistry, *215,* 245-255.

[9] Jabeur, H., Zribi, A., Makni, J., Rebai, A., Abdelhedi, R., Bouaziz, M. (2014). Detection of chemlali extra-virgin olive oil adulteration mixed with soybean oil, corn oil, and sunflower oil by using GC and HPLC. Journal of Agricultural and Food Chemistry, *62,* 4893-4904.

[10] Olliver, D., Artaud, J., Pinatel, C., Durbec, J.P., Guérère, M. (2003). Triacyglycerol and fatty acid compositions of French virgin olive oils. Characterization by chemometrics. Journal of Agricultural and Food Chemistry, *51,* 5723-5731.

[11] Bajoud, A., Pacchiarotta, T., Hurtado-Fernández, E., Olmo-García, L., García-Villalba, R., Fernández-Gutiérrez, A., Mayboroda, O.A., Carrasco-Pancorbo, A. (2016). Comparing two metabolic profiling approaches (liquid chromatography and gas chromatography coupled to mass spectrometry) for extra-virgin olive oil phenolic compounds analysis: A botanical classification perspective. Journal of Chromatography A, *1428,* 267-279.

[12] Gil-Solsona, R., Raro, M., Sales, C., Lacalle, L., Díaz, R., Ibañez, M., Beltran, J., Vicente Sancho, J., Hernández F.J. (2016). Metabolomic approach for extra virgin olive oil origin discrimination making use of ultra-high performance liquid chromatography-quadrupole time-of-flight mass spectrometry. Food Control, *70,* 350-359.

Tabla 1. *Continuación.*

Detector*	Objetivo	Columna	Tiempo (min)	[Ref]
MS	Clasificar acorde a la variedad de aceituna	C18 (100 mm × 2.1 mm, 2.2 µm)	20	[**13**]
	Clasificar acorde a la variedad de aceituna	C8 (50 mm × 4.6 mm, 5 µm)	20	[**14**]
	Identificar marcadores responsables de las características sensoriales en aceites de oliva virgen extra	C18 (100 mm × 2.1 mm, 2.2 µm)	11	[**15**]

MS: detector de espectrometría de masas.

En este capítulo la innovación ha sido modificar la longitud de las columnas tradicionales empleando únicamente una columna corta C18 de tan sólo 3 cm de longitud, habitualmente utilizada como precolumna, con lo que se logra acortar el tiempo de análisis a 4 minutos.

A continuación se especifica la instrumentación analítica, las muestras, así como el pre-tratamiento de los datos y los métodos de clasificación utilizados en este capítulo.

II.3. Instrumentación

Para llevar a cabo este estudio se utilizó el mismo equipo de cromatografía de líquidos Agilent 1100 acoplado a un detector CAD descrito en el capítulo 1.

Como columna cromatográfica se utilizó una pre-columna KromaPhase C18 de 5 µm de tamaño de partícula y de longitud 3 cm y 4 mm de diámetro interno, como fase móvil se utilizó una mezcla de acetonitrilo, metanol y agua con 0.05% de ácido fórmico.

[13] Kalogiouri, N.P., Aalizadeh, R., Thomaidis, N.S. (2018). Application of an advance and wide scope non-target screening workflow with LC-ESI-QTOF-MS and chemometrics for the classification of the Greek olive oil varieties. Food Chemistry, *256,* 53-61.

[14] Mohamed, M.B., Guasmi, F., Ben Ali, S., Radhouani, F., Faghim, J., Triki, T., Kammoun, N.G., Baffi, C., Lucini, L., Benincasa, C. (2018). The LC-MS/MS characterization of phenolic compounds in leaves allows classifying olive cultivars grown in South Tunisia. Biochemical Systematics and Ecology, *78,* 84-90.

[15] Kalagiouri, N.P., Alygizakis, N.A., Aalizadeh, R., Thomaidis, N.S. (2016). Olive oil authenticity studies by target and nontarget LC-QTOF-MS combined with advanced chemometric techniques. Analytical and Bioanalytical Chemistry, *408,* 7955-7970.

II.4. Muestras

Se utilizaron las mismas muestras que aparecen en la tabla 1 del capítulo I. Cincuenta aceites de oliva virgen extra, cuatro aceites de oliva virgen, cinco aceites de oliva refinado, cinco aceites de orujo de oliva y cincuenta y cinco aceites de no oliva, de entre los cuales se encontraban: aceite de canola, aceite de maíz, aceite de cacahuete, aceite de girasol, aceite de semillas, aceite de uva (pepita), aceite de palma, aceite de sésamo y aceite de soja.

Como fue declarado en el capítulo I, todas las muestras fueron sometidas a la reacción de transesterificación metílica previa a su análisis cromatográfico [**16**].

II.5. Condiciones experimentales

Se desarrolló un método de análisis cromatográfico en la modalidad de fase reversa o invertida. Las condiciones de trabajo fueron:

- Fase móvil: mezcla acetonitrilo, metanol y agua (80:5:15, v/v/v; 0.05% ácido fórmico)
- Tipo de elución: isocrático
- Temperatura de la columna: 30 ºC
- Flujo: 1.2 mL/min
- Volumen de inyección: 20 µL
- Tiempo de análisis: 4 min

II.6. Pre-procesado de los datos

En este estudio se llevaron a cabo dos etapas de pre-procesado de los datos, que pueden verse en mayor detalle en el artículo que se presenta el final de este capítulo. En una primera etapa se realizó el pre-procesado de los blancos de disolvente (n-hexano) y se les aplicó las mismas etapas de pre-tratamiento que a las muestras de aceites, y se obtuvo el cromatograma promedio de dichos datos. En la segunda etapa se realizó el pre-procesado de los datos de los aceites vegetales, los cuales fueron llevados a cabo con la función "Medina" mencionada en el capítulo I la cual fue ligeramente modificada de forma que a cada cromatograma de aceite vegetal se le sustrajera el cromatograma promedio de los blancos de disolvente.

[16] Biedermann, M., Grob, K., Mariani, C. (1993). Transesterification and on-line LC-GC for determining the sum of free and esterified sterols in edible oils and fats. Fett Wissenschaft Technologie (Fat Science and Technology), *95*, 127–133.

II.7. Métodos de clasificación multivariante aplicados

En primer lugar se realizó un análisis de componentes principales (PCA) de los datos para evaluar las agrupaciones naturales de las muestras, dónde se puedo observar que los aceites de palma eran muy diferentes al resto de aceites vegetales.

Seguidamente se dividió el conjunto de datos en entrenamiento y validación externa y se aplicaron los siguientes algoritmos quimiométricos: 'partial least squares-discriminant analysis' (PLS-DA), 'k-nearest neighbours' (kNN), 'support vector machine-classification' (SVM-C), 'soft independent modelling of class analogies' (SIMCA) y 'one-class partial least-squares' (OCPLS).

La calidad de las clasificaciones realizadas por cada uno de los modelos fue evaluado calculando los siguientes parámetros: sensibilidad, especificidad, valor predictivo positivo, valor predictivo negativo, índice de Youden, relación de verosimilitud para resultados positivos, relación de verosimilitud para resultados negativos, valor F, poder discriminante, eficiencia, área bajo la curva, coeficiente de correlación de Matthews, y coeficiente Kappa. Todos estos parámetros se calcularon a partir de los resultados obtenidos sobre el conjunto de validación externa.

A continuación se presenta el artículo publicado, donde se describe de forma más detallada del estudio realizado.

ARTÍCULO CIENTÍFICO

Fast-HPLC fingerprinting to discriminate olive oil from other edible vegetable oils by multivariate classification methods

(Artículo publicado en 2017, en la revista: Journal of AOAC International,100 (2), 345-350)

JIMÉNEZ-CARVELO ET AL.: JOURNAL OF AOAC INTERNATIONAL VOL. 100, NO. 2, 2017 345

SPECIAL GUEST EDITOR SECTION

Fast-HPLC Fingerprinting to Discriminate Olive Oil from Other Edible Vegetable Oils by Multivariate Classification Methods

ANA M. JIMÉNEZ-CARVELO, ANTONIO GONZÁLEZ-CASADO, ESTEFANÍA PÉREZ-CASTAÑO, and LUIS CUADROS-RODRÍGUEZ
University of Granada, Department of Analytical Chemistry, c/ Fuentenueva, s.n. E-18071 Granada, Spain

Fast-hplc fingerprinting to discriminate olive oil from other edible vegetable oils by multivariate classification methods

Ana M. JIMÉNEZ-CARVELO ✉, Antonio GONZÁLEZ-CASADO , Estefanía PÉREZ-CASTAÑO, Luis CUADROS-RODRÍGUEZ

Department of Analytical Chemistry, University of Granada, c/ Fuentenueva, s.n. E-18071 Granada, Spain.

Abstract

A new analytical method for the differentiation of olive oil from other vegetable oils using reverse phase liquid chromatography, and applying chemometric techniques was developed. A 3 cm short column was used to obtain the chromatography fingerprinting of the methyl-transesterified fraction for each vegetable oil. The multivariate classification methods used were k-nearest neighbours (kNN), partial least squared-discriminant analysis (PLS-DA), one-class partial least squares (OCPLS), support vector machine classification (SVM-C), and soft independent modelling of class analogies (SIMCA). The discrimination of olive oil from other vegetable edible oils was evaluated by several classification quality metrics. Several strategies for the classification of the olive oil were used: (i) one input-class, two input-class and pseudo two input-class. It has to be highlighted that in the new proposed analytical method, the chromatographic analysis takes only four minutes and has been firstly applied in this study for olive oil authenticating by means discriminant analysis.

Keywords

Reverse-phase, liquid chromatography, olive oil, fingerprinting, fast-chromatography, HPLC-CAD, chemometric, performance features

✉ Corresponding author: phone: +34 958240797; fax: +34 958243328; email: amariajc@ugr.es

1. Introduction

The vegetable oils constitute one of the main sources of fats in the diet, being the olive oil one of the oils more appreciated by the consumers thanks to their valuable nutritional properties which makes it different from others edible oils. The International Olive Council (IOC), based in Madrid (Spain), is the only intergovernmental organisation in the world in the field of olive oil. It enables meeting between government representatives and experts to discuss problems and concerns, draw up quality standards and it advices governments on the decisions they should take. The olive oil is regulated in the European Union (EU) by two legal provisions: (i) *the Commission Regulation (CEE) No. 2568/91 on the characteristics of olive oil and olive-residue oil* (actually named olive-pomace oil) *and on the relevant methods of analysis* which has undergone 28 amends since its publication in 1991, the last one dated in October 2015 (more than one per year); and (ii) *the Commission Implementing Regulation (UE) No. 29/2012 on marketing standards for olive oil*, amended also in December 2013. The United States Department of Agricultural (USDA) published in October 2010 the *United States Standards for grades of olive oil and olive-pomace-oil* (2010), followed in May 2012 of the *Grading Manual for Olive Oil and Olive-Pomace Oil*. The manual defines the standardized process for inspection, testing and authentication and provides a path for authentication and USDA certification through voluntary laboratory testing. Other countries, which have their own similar regulations, are the People's Republic of China, in force since March 2009, and Australia even though the first edition of the standard has not appeared yet.

The verification of the requirements specified in these regulations requires, if it is possible, rapid and straightforward analytical methods suitable to characterize the olive oil and to detect potential frauds due to adulterations of olive oil with other cheaper vegetable edible oils. This way reducing the cost and achieve an illicit profit.

These analytical methods are upgraded in a regular way considering the recommendations propose by IOC. The major analytical methodologies applied to authenticate olive oil are focused on the characterization of specific analytical profiles from both major components (saponifiable fraction) and minor components (unsaponifiable fraction). The composition of triglycerides [**1**], fatty acids [**2**] or phytosterols [**3**] is determined by chromatographic methods in order to detect if an olive oil has been adulterated with other vegetable edible oils [**4**].

The main problem is that the scope of these methods is limited and not all adulterations of olive oil with other vegetable oils are easily detectable. This is because some vegetable edible oils have a compounds profile similar to olive oil. For instance, there is not a well-established analytical method to detect the adulteration of olive oil with hazelnut oil since this oil has both triacylglycerol and fatty acid compositional profiles similar to olive oil. The IOC is claiming for many years in order to that method is developed.

Therefore, it is advisable to develop reliable, rapid and simple methods to control the authenticity of olive oil. An attractive and innovative analytical strategy, which uses the chromatogram as a whole, is being recently applied. This strategy is known as 'chromatographic fingerprinting' [**5**], starts from the premise that the chemical information necessary to properly characterise and authenticate a particular foodstuff is contained into the chromatographic profile. This critical and useful information is non-evident (it is hidden) but it could be extracted by applying appropriate chemometrics tools.

In the last years, the authentication of olive oil based on instrumental fingerprinting has increased, being the analytical spectroscopy as near infrared (NIR), Fourier-transform infrared (FTIR), Raman and nuclear magnetic resonance (NMR) those most used ones. This is because the sample does not suffer any alteration, is not destroyed, and these are quick [**6**,**7**]. In a similar way, when the chromatographic fingerprinting is applied, a poor resolved chromatogram is used as a whole analytical signal composed of complex bands (as a spectrum) and the areas/heights of the individual peaks are not considered. Recently, a comprehensive review about chromatographic fingerprint strategy including the ground of the methodology and the analytical applications to date has been reported [**5**].

In a previous paper, the HPLC fingerprints of the sterols fraction obtained through a saponification reaction has been used for discriminating virgin olive oils from other vegetable edible oils [**8**]. The sample preparation methodology applied to isolate these compounds was long and tedious. For this reason a new strategy has been developed, based on the method proposed by Biedermann et al. [**9**], which substitutes the traditional saponification reaction of the vegetable oils for a methyl transesterification [**10**]. The resulting transesterified fraction gathers the chemical information needed for differentiate each vegetable oil. Consequently, a nonspecific 'chromatographic

fingerprinting' of each vegetable oil is acquired where this fraction is analysed using liquid chromatographic. Each plant species presents a characteristic chromatographic fingerprinting, related to its specific chemical composition, which could be applied to discriminate among different vegetable oils. Consequently, it is not necessary to identify each one of the chemical components.

The aims of this study are: (i) to propose a quick new analytical chromatographic method and (ii) to probe the applicability of a new classification strategy designated by us as "*pseudo* two input-class classification", to differentiate olive oil from other vegetable edible oils. The methyl-transesterified fraction of each vegetable oil has been analysed by conventional reverse phase high-performance liquid chromatography using a 'short-column', habitually used as precolumn, instead of a typical chromatography columns. In this way, the solvent consumption and the chromatographic run time (4 minutes) are drastically reduced. The chromatographic signal has been used as analytical information to set up the classification models.

Three classification strategies were applied: (i) one input-class (1iC) classification, two input-class (2iC) classification and one input-class plus one 'dummy' class (or *pseudo* two input-class (*p*2iC) classification). Some common models of classification have been used, as k-nearest neighbours (kNN), partial least squares-discriminant analysis (PLS-DA), one-class partial least squares (OCPLS), support vector machine-classification (SVM-C) and soft independent modelling of class analogies (SIMCA). In addition, several quality metrics: (i) sensitivity, (ii) specificity, (iii) positive (or precision) and negative predictive values, (iv) Youden index, (v) positive and negative likelihood ratios, (vi) F- measure (or F- score), (vii) discriminant power, (viii) efficiency (or accuracy), (ix) AUC (area under the receiver operating curve), (x) Matthews correlation coefficient and (xi) Kappa coefficient, were used in order to assess the performance of the classifications. Extensive information on both the specific features of the classification strategies and the meaning of the classification quality metrics can be found in a recent tutorial [**11**].

2. Materials and methods

2.1 Chemicals

All solvents used were HPLC grade. Methanol was provided by Panreac (Barcelona, Spain), deionized water was obtained using a Milli-Q-system (Millipore, Bedford, MA, USA) and acetonitrile was provided by Merck (Darmstadt, Germany).

Other reagents, as sodium methoxide (MeONa), citric acid monohydrate, anhydride sodium sulphate were purchased from Merck (Darmstadt, Germany), formic acid and tert-buthyl methyl ether (TBME), were purchased from VWR International Eurolab, S.L (Barcelona Spain). The nitrogen (99.9999 %) used was provided by Air Liquid (Madrid, Spain).

2.2 Chromatography

The analysis was performed with an Agilent 1100 series liquid chromatograph (Santa Clara, USA) equipped with a column thermostat (Eppendorf CH30), a quaternary pump and degasser auto sampler. Detection was performed with a corona charged aerosol detector (CAD) (ESA Bioscenses Inc., Chemlsford, MA, USA). Agilent ChemStation software (rev. B.02.01-SR1) for LC systems was used.

The chromatography fingerprint from the methyl-transesterified fraction was obtained by HPLC using a KromaPhase C18 precolumn (30 x 4.0 mm i.d, 5 µm) from Scharlau (Barcelona, Spain). The temperature of the column was set at 30 ºC during the entire operation. The mobile phase was constituted by acetonitrile/methanol/water (0.05 % formic acid) (80:5:15, v/v) at a flow rate of 1.2 mL min-1 and a run time of 4 min. The volume of injection was 20 µL.

2.3 Samples

A total of 119 edible vegetable oil samples of different trademark and types were analysed. The samples were purchased directly in local markets. The following oils were analysed: 64 olive oils of different categories (virgin extra (50), virgin (4), refined+virgin (5), and pomace+virgin (5)), and the other 55 samples were: canola oil

(7), corn oil (5), peanut oil (5), sunflower oil (13), (no-specified) seeds oil (5), grapeseed oil (4), palm oil (7), sesame oil (3), and soybean oil (6).

2.4 Sample preparation

Previous to the chromatographic analysis, a transesterification reaction was applied to the sample oil. A modification of the original procedure described by Biedermann et al was applied [**9**]. For this, 0.1 g of oil was weighed into a centrifuge tube. 1 mL of extracting agent (MeONa at 10 % in methanol in TBME, 4:6 (v/v)) was added and mixed with the oil. The mixture was stirred during 20 s, and then allowed to stand during 20 min. This step was repeated twice. Then 1 mL of water and 8 mL of hexane was added, and then the mixture was centrifuged 3 min at 1,500g. The aqueous phase was removed with a Pasteur pipette followed by addition of 1 mL of 1 % citric acid in water. Again, the aqueous phase was eliminated and 2 g of sodium sulphate anhydrous were added, and the mixture was allowed to stand during 20 min.

The methyl-transesterified organic fraction was filtered with a syringe filter of polytetrafluoroethylene (PTFE) membrane 0.22 μm and the subsequent solution was stored at –25ºC until analysis. For the chromatographic analysis, 350 μL of transesterified solution was added to a 2 mL HPLC vial.

n-hexane was used as analytical blank in order to obtain the representative chromatograms of the 'dummy' class.

2.5 Chemometrics

The raw data files from each chromatogram were obtained in a CSV file, and then exported to MATLAB format (version R2013a). The data pre-processing was done with a home-programmed MATLAB function named "Medina" (version 13) [**10**]. This function implements several algorithms from Matlab Bioinformatics Toolbox™, and 'icoshift' (interval correlation optimized shifting) algorithm to align the peaks of the chromatograms [**12**]. Two types of pre-processing of data were applied. First a pre-processing for analytical blank was applied and then a second pre-processing on the raw dataset from the oils. First pre-processing had as mean aims to overlap and align the analytical blanks chromatograms and to obtain the average of all samples of analytical blank chromatogram analysed. The following steps for the first pre-processing the data

were: (1) raw chromatograms analytical blank data grouping and overlay; (2) selection of interval of interest in chromatograms; (3) filtered of the raw chromatograms analytical blank data to eliminate noise of signal analytical; (4) correction of the baseline using the 'msbackadj' function (included in the Bioinformatics ToolboxTM); (5) alignment of the peaks with the function 'icoshift'; and finally (6) obtaining of the pooled chromatogram. The following steps for the second pre-processing the data were: (1) raw chromatograms data grouping and overlay; (2) selection of interval of interest in chromatograms; (3) filtered of the raw chromatograms data to eliminate noise of signal analytical; (4) correction of the baseline using 'msbackdj' function; (5) alignment of the peaks with the function 'icoshift'; (6) averaging of the entire chromatograms of analytical blank was subtracted; and finally (7) mean centring of the data set.

Three classification strategies have been applied: (i) one input-class (1iC) classification; (ii) two input-class (2iC) classification; and (iii) one input-class plus one 'dummy' class classification (or *pseudo* two input-class (*p*2iC) classification). Usually a 2iC classification strategy has been only used to carry out the differentiation of olive oils from other vegetable edible oils applying discriminate analysis as PLS-DA or SVM-C. This is because discriminant methods require using two classes in order to define the border among the different regions of the discriminant model. However, it is also possible to perform the discrimination models training the classification model with a single class. It offers some advantages for the food authentication considering that the model could be built only from sufficient number of genuine foods (authenticated olive oils, independently of the quality category) and it is not necessary that other vegetable oils are advisable. In this way, the number of previous analyses to train the model is significant lesser. The *p*2iC methodology can be considered as a 2iC discriminant method, which used a single effective class plus a fictitious class (or 'dummy' class). In all cases, the target class was 'olive oil' while the non-target class was 'non-olive oil' for 2iC and 'dummy class' for *p*2iC.

For each strategy applied (2iC, *p*2iC and 1iC) the original vector data set were divided in different groups. The selection was carried out in a random way. For 2iC, the training set which is made up of 80 oil samples (42 olive oils and 38 non-olive-oils), and the validation set composed by the remaining oil samples (24 olive oils and 17 non-olive oils). For *p*2iC, the training set was made up of 41 olive oil samples and 62 analytical blanks, and the validation set composed by 41 oil samples (24 olive oils and 17

non-olive oils). For 1iC, the training set was composed by 41 olive oils samples and the validation set by 41 oil samples (24 olive oils and 17 non-olive oils).

Once it was done, the classification of the vegetable oils was performed. PLS_Toolbox (version 7.95, Eigenvector Research, Wenatchee, WA) was applied for exploratory analysis and classification methods: principal component analysis (PCA) [**13**], k-nearest neighbors (kNN), partial least squares discriminant analysis (PLS-DA) [**14**], soft independent modelling of class analogies (SIMCA) [**15**], and support vector machine-classification (SVM-C) [**16**]. There are two commonly used versions of SVM classification, 'C-SVC' and 'nu-SVC'. C-SVC optimizes a model using the adjustable parameter C (0→inf) to apply a penalty to the optimization for data points which is not correctly separated by the classifying hyperplane. In Nu-SVC the C penalty parameter is replaced by a nu [0 →1] parameter which applies a slightly different penalty [**17**].

In addition, one-class partial least squares (OCPLS) was performed applying the three variants of the function: (i) conventional ordinary linear OCPLS, (ii) nonlinear radial basis function (RBF) OCPLS, and (iii) partial robust M-regression (PMR) OCPLS, using the software provided by Xu et al. [**18**]. All the options offered by the software were tested.

3. Results and discussion

A chromatogram was recorded for each vegetable oil sample and for each analytical blank. The chromatograms of the analytical blank were obtained of a single vial which was analysed several times. Figure 1a shows the superposed chromatograms of all vegetable oil samples, figure 1b shows the superposed chromatograms of analytical blanks, and figure 1c shows the average chromatogram of the analytical blanks.

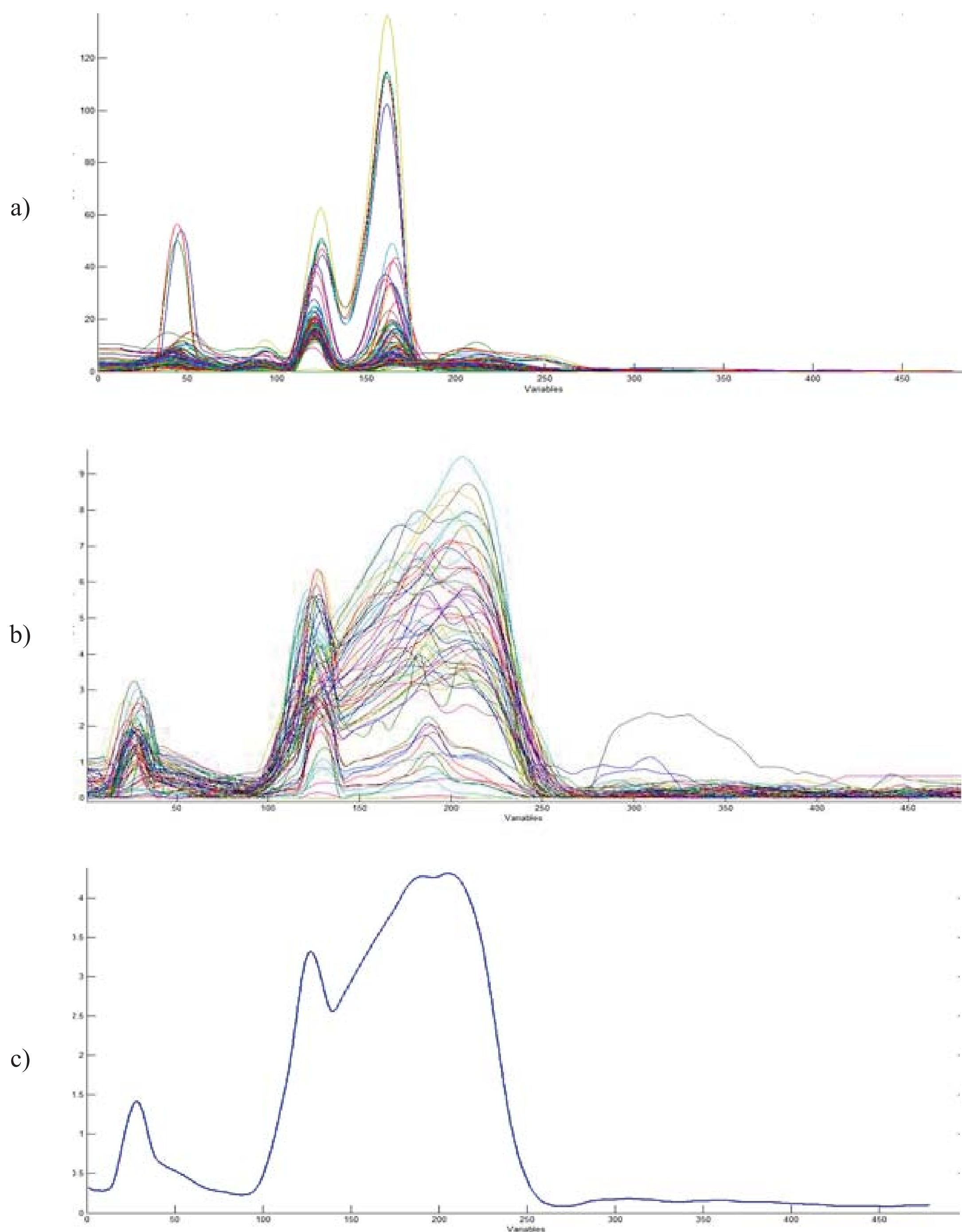


Figure 1. (a) Superposed chromatograms of the 119 vegetable oil samples; (b) superposed chromatograms of the 62 analytical blank samples (dummy class); and (c) average chromatogram of the analytical blank samples. The chromatograms have been previously pre-processed with the exception of the mean centring step.

Exploratory analysis

A principal component analysis (PCA) was previously carried out for exploring if there were natural groupings of different oils. Four PCs were enough to explain the 98.2% of the variance. Figure 2 shows the scores on the PC2-PC1 plane. The explained variance for PC1 and PC2 was 88.0% and 6.8% respectively. As it can be observed, the first principal component groups the palm oils (bottom right), and the second component separates olive oils from the other vegetable oils.

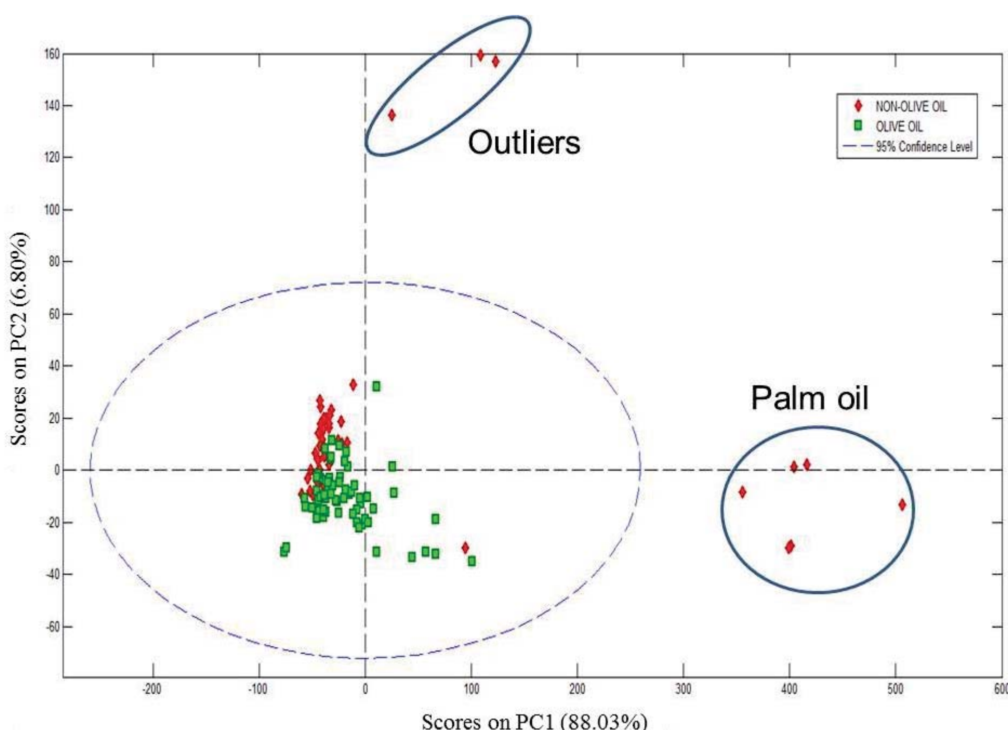


Figure 2. PC2-PC1 score biplot obtained from the fingerprint data (the whole chromatogram) of the methyl-transesterified fraction of the 119 vegetable oil samples.

Class-modelling methods

In order to differentiate olive oils from other vegetable oils, two class-modelling classification methods were performed: SIMCA and OCPLS.

The three classification strategies: 2iC, *p*2i and 1iC were tested when SIMCA is applied. For all models, the differentiation was carried out by the software PLS_TOOLBOX from the Q-residual values from each sample of olive oil. Samples with a normalised Q-residual value less than $\sqrt{2}$ were classified as olive oil. Firstly a 2iC SIMCA classification was performed. For each class a PCA model was building, 5 PCs and 3 PCs were chosen for olive oil and non-olive oil classes respectively.

Secondly, a *p*2iC SIMCA was carried out. In this case, 6 PCs and 3 PCs were chosen for olive oil class and 'dummy' classes respectively. At last, a 1iC SIMCA was performed with 5 PCs for olive oil model.

Next, OCPLS model (a 1iC classification strategy) was performed. Conventional OCPLS was built with 4 latent variables (LVs), RBF OCPLS with 7 LVs and PMR OCPLS with 3 LVs. For classification purposes, the regions pre-established by the software were used.

In the case of SIMCA the 2iC strategy gave better results than *p*2iC and 1iC. The values of efficiency, AUC and Matthews's correlation coefficient were high values about 0.8. For OCPLS, the ordinary linear classified better than NRB and PMR functions. In addition, the metrics values for OCPLS (ordinary linear classified) were lower than 2iC SIMCA model. The results for each class modelling methods are showed in Table 1.

Table 1. Values of the quality performance metrics from the different class-modelling classification methods.

Performance features	SIMCA			OCPLS		
	2iC	*p*2iC	1iC	Ordinary linear	NRB	PMR
				1iC	1iC	1iC
Sensibility (Recall)	0.88	0.04	0.04	0.88	0.67	0.67
Specificity	0.76	1.00	1.00	0.59	0.59	0.65
Positive predictive value (Precision)	0.84	1.00	1.00	0.75	0.70	0.73
Negative predictive value	0.81	0.43	0.43	0.77	0.56	0.58
Youden index	0.64	0.04	0.04	0.46	0.25	0.31
Positive likelihood rate	3.72	–	–	2.13	1.62	1.89
Negative likelihood rate	0.16	0.96	0.96	0.21	0.57	0.52
F-measure	0.86	0.08	0.08	0.81	0.68	0.70
Discriminant power	0.75	–	–	0.55	0.25	0.31
Efficiency (Accuracy)	0.83	0.44	0.44	0.76	0.63	0.66
AUC (Correctly classified rate)	0.82	0.52	0.52	0.73	0.63	0.66
Matthews correlation coefficient	0.65	0.13	0.13	0.49	0.25	0.31
Kappa coefficient	0.65	0.03	0.03	0.48	0.25	0.31

The hyphen "–" is signifying that the performance feature cannot be determined

Discriminant analysis methods

Three well-established classification methods were tried: kNN, PLS-DA, and SVM-C. In all cases, only 2iC and *p*2iC strategies were tested because, as it has been stated, 1iC strategy is not feasible for discrimination analysis.

The *p*2iC strategy is not applicable for perform the kNN model because the objects from the class non-olive are always considered by the classification model as 'nearest neighbours' to the objects from the target class (olive oil) that from the 'dummy' class. It is obvious because a non-olive oil chromatogram is always more like an olive oil

chromatogram than an analytical blank chromatogram. Then, the 2iC strategy was only applied. The model was performed with k=3, the olive class was defined by a class predicted probability equal to 1, while the non-olive class was defined by a probability of 0. All samples of olive oil and other vegetable edible oils were well classified, with exception of the palm oil samples, which were classified as olive oil with a class predicted probability value of 1.

Firstly, PLS-DA model using 2iC methodology was performed. 4 LVs were selected to build the model explaining a 98.2% of variance. The olive class was assigned with value 1 and non-olive oil class with value 0. In second step a PLS-DA model using *p*2iC methodology was built with 5 LVs which explaining a 95.0% of variance. The discrimination of olive oil from other vegetable edible oil was carried out establishing a confidence interval from the estimated standard deviation (s) of the predicted values for the predicted values of the olive oil samples of the training set. The interval was the arithmetic mean of the predicted values multiplied by 2.33×s. This expression is similar to that which has been recommended for calculating the analytical decision limit (LD) for substances no permitted in foodstuffs [**19**].

SVM-C classification was carried out by optimizing of "nu" and "c" parameters. In addition, all the models were tested both without and with variable reduction by PCA and PLS. This variable reduction is named X-block compression by the software [**17**]. The 2iC and *p*2iC strategies were applied. The results obtained for 2iC (nu)SVM-C, were similar, independently of the type of X-block compression (none, PCA and PLS). Nevertheless, for 2iC (c)SVM-C, the results were better when an X-block compression with PLS was applied. For 2iC, the samples were always directly classified by the software. Afterward, *p*2iC strategy was applied using also the different options of X-block compression. The discrimination of the samples was also classified calculated a confidence interval as plus/minus 2.33 times the standard deviation from the predicted class probability. In both cases, X-block compression with PLS given serious classification errors and any oil was well classified; for this reason, the quality performance metrics were not calculated.

In short, the 2iC strategy gave good results for all the discriminant methods. However, the *p*2iC strategy yielded only good results when PLS-DA method was applied. In the Table 2 shows the quality performance metrics for the different methods.

Table 2. Values of the quality performance metrics from the different discriminant analysis classification methods.

Performance features	kNN	PLS-DA		(nu)SVM-C					(c)SVM-C				
				None		PCA		PLS	None		PCA		PLS
	2iC	2iC	*p*2iC	2iC	*p*2iC	2iC	*p*2iC	2iC	2iC	*p*2iC	2iC	*p*2iC	2iC
Sensibility (Recall)	1.00	0.96	0.58	1.00	0.75	1.00	0.67	1.00	0.96	0.63	1.00	0.67	1.00
Specificity	0.76	0.71	0.24	0.76	0.41	0.65	0.24	0.76	0.76	0.41	0.88	0.29	0.94
Positive predictive value (Precision)	0.86	0.82	0.52	0.86	0.64	0.80	0.55	0.86	0.85	0.60	0.92	0.57	0.96
Negative predictive value	1.00	0.92	0.29	1.00	0.54	1.00	0.33	1.00	0.93	0.44	1.00	0.38	1.00
Youden index	0.76	0.66	-0.18	0.76	0.16	0.65	-0.10	0.76	0.72	0.04	0.88	-0.04	0.94
Positive likelihood rate	4.25	3.26	0.76	4.25	1.28	2.83	0.87	4.25	4.07	1.06	8.50	0.94	17.00
Negative likelihood rate	0.00	0.06	1.77	0.00	0.61	0.00	1.42	0.00	0.05	0.91	0.00	1.13	0.00
F-measure	0.92	0.88	0.55	0.92	0.69	0.89	0.60	0.92	0.90	0.61	0.96	0.62	0.98
Discriminant power	–	0.96	-0.20	–	0.18	–	-0.12	–	1.03	0.04	–	-0.04	–
Efficiency (Accuracy)	0.90	0.85	0.44	0.90	0.61	0.85	0.49	0.90	0.88	0.54	0.95	0.51	0.98
AUC (Correctly classified rate)	0.88	0.83	0.41	0.88	0.58	0.82	0.45	0.88	0.86	0.52	0.94	0.48	0.97
Matthews correlation coefficient	0.81	0.70	-0.19	0.81	0.17	0.72	-0.11	0.81	0.75	0.04	0.90	-0.04	0.95
Kappa coefficient	0.79	0.69	-0.19	0.79	0.17	0.68	-0.10	0.79	0.74	0.04	0.90	-0.04	0.95

The hyphen "–" is signifying that the performance feature cannot be determined

4. Conclusion

The use of a short column for obtaining chromatographic fingerprinting from methyl-transesterified fraction and the application of chemometric classification methods, to solve problems the olive oil authentication has gave good results.

In this study, several classification strategies and methods have been applied, and the results have been discussed. Class-modelling (SIMCA and OCPLS) and discriminant analysis (kNN, PLS-DA and SVM-C) methods have been tested. Each one has been applied using three classification strategies, designated as two input-class (2iC) classification, *pseudo* two input-class (*p*2iC) classification and one input-class (1iC) classification. The main advantage of the *p*2iC strategy, proposed in this paper, it is can use the discriminant analysis methods as the class modelling methods using only the samples of the target class to train the model.

To assess the different classification scenarios (methods and strategies), several quality classification metrics have been calculated. The discriminant analysis methods have proven to be better classifiers than class-modelling methods. It should be emphasized that the *p*2iC strategy yield better results in discriminant analysis methods than class-modelling methods.

Finally, it should be remarked that the approach of working with a single effective input-class is especially advantageous for two reasons: (i) it only requires experimental data from that class and the number of necessary analysis might be reduced; (ii) it is not possible collect a sufficient number of representative samples of all the other non-olive oils.

5. References

[1] Bosque Sendra, J.M., Cuadros Rodríguez, L., Ruíz Samblás, C. & de la Mata, A.P. (2012) Combining chromatography and chemometrics for the characterization and authentication of fats and oils from triaglycerol compositional data- A review. *Analytical Chimica Acta*, **724**, 1–11.

[2] Monfreda, M., Gobbi, L. & Grippa, A. (2012) Blends of olive oil and sunflower oil: Characterisation and olive oil quantification using fatty acid composition and chemometric tools. *Food Chemistry* **134**:2283–2290.

[3] Gázquez Evangelista, D., Pérez Castaño, E., Sánchez Viñas, M. & Bagur González, M.G. (2013) Using offline HPLC-GC-FID 4-desmethylsterols concentration profiles, combined with chemometric tools, to discriminate different vegetable oils. *Food Analytical Method*, **7**, 912–925.

[4] Aparicio, R. & Aparicio-Ruiz, R. (2000) Authentication of vegetable oils by chromatographic techniques. *Journal of Chromatograph*y A, **881**:93–104.

[5] Cuadros Rodríguez, L., Ruiz Samblás, C., Valverde Som, L., Pérez Castaño, E. & González Casado, A. (2016) Chromatographic fingerprinting: an innovative approach for food 'identitation' and food authentication - A tutorial. *Analytica Chimica Acta*, **909**, 9–23.

[6] Osorio M.T., Haughey S.A., Elliott C.T. & Koidis, A. (2015) Identification of vegetable oil botanical speciation in refined vegetable oil blends using an innovative combination of chromatographic and spectroscopic techniques. *Food Chemistry*, **189**, 67–73.

[7] Tay A., Singh .RK., Krishnan S.S. & Gore J.P. (2002) Authentication of Olive Oil Adulterated with Vegetable Oils Using Fourier Transform Infrared Spectroscopy. *LWT - Food Science and Technology*, **35**, 99–103.

[8] Bagur Gonzalez, M.G., Perez 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.

[9] Biedermann M., Grob K. & Mariani C. (1993) Transesterification and on-line LC-GC for determining the sum of free and esterified sterols in edible oils and fats. *Fat Science Technology*, **95**,127–133.

[10] 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.

[11] Cuadros Rodríguez, L., Pérez Castaño, E. & Ruiz Samblás, C. (2016) Quality performance metrics in multivariate classification methods for qualitative analysis. *Trends in Analytical Chemistry*, **80**, 612–624.

[12] 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.

[13] Ballabio, D. & Consonni, V. (2013) Classification tools in chemistry. Part 1: linear models. PLS-DA. *Analytical Methods*, **5**, 3790–3798.

[14] Bro R. & Smilde A.K (2014) Principal component analysis. *Analytical Methods*, **6**, 2812–2831.

[15] Wold, S. & Sjostrom, M. (1977) SIMCA: A method for analyzing chemical data in terms of similarity and analogy, in *Chemometrics Theory and Application*, Kowalski, B.R. (Ed), ACS Symposium Series 52, Washington D.C., pp. 243–282.

[16] Luts J., Ojeda F., Van de Plas R., De Moor B., Van Huffel S. & Suykens J.A.K. (2010) A tutorial on support vector machine-based methods for classifictions problems in chemometrics. *Analytical Chimica Acta*, **665**, 129–145.

[17] SVM Function Settings, Eigenvector Documentation wiki. URL: http://wiki.eigenvector.com/index.php?title=SVM_Function_Settings. Accessed 29.05.16.

[18] Xu L., Yan S.M., Cai C.B. & Yu X.P. (2013) One-class partial least squares (OCPLS) classifier. *Chemometrics and Intelligent Laboratory Systems*, **126**, 1–5.

[19] EU Commission Decision (2002/657/EC) implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results, *Official Journal of the European Communities*, **L 221**, 8-36.

II.8. Discusión

En vista de los resultados obtenidos se puede concluir que el empleo de toda la huella dactilar cromatográfica de la fracción metil-transesterificada del aceite de oliva es válida para autentificar el aceite de oliva.

Además, el empleo de una pre-columna de 3 cm en lugar de una columna cromatográfica convencional ha hecho posible reducir el tiempo de análisis a tan sólo 4 minutos, lo cual conduce a una reducción considerable en el costo del análisis, no sólo por el recorte del tiempo de análisis sino porque el coste de una pre-columna con respecto a una columna cromatográfica convencional es notablemente menor.

Los resultados de este estudio y las mejoras con respecto a los análisis convencionales de cromatografía líquida en la modalidad de fase invertida, pueden ser de gran utilidad en el sector del aceite de oliva ya que podrán emplear este método para autentificar el aceite de oliva, evitando realizar una determinación de todos los parámetros actuales definidos en el Reglamento (CE) Nº 2568/91 relativo a las características de los aceites de oliva y de los aceite de orujo de oliva y sobre sus métodos de análisis.

Por otro lado, el método analítico desarrollado puede ser aplicado en cualquier laboratorio de rutina que disponga únicamente de un cromatógrafo de líquidos, el cual habitualmente se emplea en la modalidad de fase reversa. De esta manera para autentificar el aceite de oliva no se tendrá que cambiar de modalidad de trabajo del equipo instrumental.

COMUNICACIONES A CONGRESOS

- A.M. Jiménez Carvelo, A. González Casado, L. Cuadros Rodríguez. **Uso de cromatografía líquida "rápida" (fast-HPLC) para la discriminación de aceite de oliva de otros aceites vegetales aplicando herramientas quimiométricas.** XVIII Feria internacional del aceite de oliva e industrias afines (EXPOLIVA). **Oral.**

- A.M. Jiménez Carvelo, A. González Casado, L. Cuadros Rodríguez. **Aplicación conjunta de LC "rápida" y quimiometría para la autentificación de aceite de oliva.** IX Congreso argentino de química analítica. **Oral.**

Capítulo III

"Investigar es ver lo que todo el mundo ha visto, y pensar lo que nadie más ha pensado"

Albert Szent-Györgyi

CAPITULO III

'Huellas dactilares espectroscópicas– espectrometría vibracional'

III.1. Presentación

Este capítulo recoge los resultados obtenidos al aplicar la metodología de huellas dactilares espectroscópica sutilizando espectrometría FT-NIR, FTIR-ATR y Raman y los métodos de clasificación multivariante desarrollados con los **objetivos** de: (i) discriminar aceites de oliva de diferentes categorías de otros aceites vegetales comestibles, y (ii) cuantificar la proporción de aceite de oliva en mezclas con otros aceites vegetales.

Al igual que en el capítulo I y II se aplicaron tres estrategias de clasificación dependiendo del número de clases utilizadas en el desarrollo de los modelos, y diversos métodos de análisis discriminante (PLS-DA, SVM y kNN) y métodos de modelado de clases (SIMCA y OCPLS). Además se cuantificó la proporción de aceite de oliva en mezclas con otros aceites vegetales aplicando dos técnicas de calibración multivariante (PLS y SVR).

Este capítulo derivó en un artículo publicado, cuya referencia es:

1. *Chemometric classification and quantification of olive oil in blends with any edible vegetable oils using FTIR-ATIR and Raman spectroscopy* (LWT-Food Science and Technology, 2017, *86,* 174-184).

Seguidamente se realiza una breve introducción correspondiente a este capítulo.

III.2. Introducción

Dentro del ámbito de la espectrometría vibracional, una de las técnicas espectroscópicas más utilizadas es la espectroscopia de infrarrojo. Ésta proporciona información acerca de los procesos de absorción de las moléculas que se encuentran en la muestra cuando la radiación electromagnética interacciona con la materia. La región de infrarrojo del espectro electromagnético se encuentra comprendida entre 12800-10 cm^{-1}dónde: el infrarrojo cercano (NIR) está comprendido entre 12800-400 cm^{-1}, el infrarrojo medio (MIR) entre 4000-400 cm^{-1} y el infrarrojo lejano (FIR) entre 400-10 cm^{-1}.

El estudio realizado en este capítulo se ha llevado a cabo en la región del infrarrojo medio utilizando un equipo que incorporaba el método de transformada de Fourier (FT-IR). En FT-IR en función del tipo de medida que se realice existen diferentes técnicas de muestreo [**1**,**2**,**3**]:

- Transmisión: se utilizan principalmente para sustancia sólidas. Se requiere realizar una pastilla en presencia de KBr, la cual se colocará en el portamuestras para su posterior análisis y registro del espectro infrarrojo.

- Reflectancia total atenuada (ATR): es un tipo de reflexión que se produce cuando la radiación electromagnética entra en contacto con un cristal de alto índice de refracción, de forma que se genera una onda evanescente sobre la superficie del cristal dónde se encuentra depositada la muestra y posteriormente se registra su espectro de infrarrojo.

- Reflectancia difusa (DRIFTS): en este caso cuando la radiación electromagnética incide sobre la materia, la energía es reflejada en todas las direcciones.

- Reflectancia especular: en este caso cuando la luz electromagnética incide sobre la muestra y ésta la refleja como si fuese un espejo.

En función del tipo de análisis que se desee realizar se pueden encontrar y/o ensamblar al equipo de FTIR el accesorio necesario que lleve a cabo la técnica de muestreo deseada. El estudio que es presentado en este capítulo se realizó un equipo ATR-FTIR que se muestra en la figura 1.

---

[1] Smith, B.C. (2011). Preparing samples properly. En: Smith, B.C. (Ed). Fundamentals of Fourier transform infrared spectroscopy. Editorial: Taylor & Francis Group.

[2] Stuart, B.H. (2004). Experimental methods. En: Stuart, B.H. (Ed). Infrared spectroscopy:fundamentals and applications. Editorial: John Wiley & Sons.

[3] ATR Specac: Reflectancia total atenuada. Technical note. Teknokroma.

Figura 1. Equipo ATR-FTIR utilizado para llevar acabo los análisis de este capítulo. A la izquierda de la imagen se muestra el accesorio ATR.

Existe otro tipo de espectrometría vibracional, la espectroscopia Raman que se diferencia del infrarrojo en el fundamento físico, ya que en ésta lo que se mide es la dispersión de la radiación electromagnética cuando interacciona con la materia al contrario del IR que mide la absorción.

La dispersión Raman se mide en el rango comprendido entre 250-2900 cm^{-1} y el análisis consiste en medir la intensidad y frecuencia de fotones que se dispersan en el material cuando es irradiado con luz monocromática de alta intensidad, habitualmente se utiliza un láser para ello. La principal ventaja con respecto a la técnica de infrarrojo es que no requiere contacto con la muestra para llevar a cabo la medida, sino que la muestra se encuentra en viales o en algún tipo de compartimento, además no se presenta la interferencia del agua, sin embargo presenta el inconveniente de si la muestra medida presenta fluorescencia ésta puede interferir en la intensidad de las bandas Raman [**4**,**5**].

En la figura 2 se muestra el equipo utilizado Raman empleado en el estudio presentado en este capítulo.

Figura 2. Equipo Raman utilizado para llevar acabo los análisis de este capítulo.

[4] Pérez Pueyo, R. (2005). Procesado y optimización de espectros RAMAN mediante técnicas de lógica difusa: aplicación a la identificación de materiales pictóricos. Tesis Doctoral. ISBN: 8468909696 (D.L: B.15685-2005).

[5] McCreerey, R.L. (2005).Introduction and scope.En: Winefordner, J.D. (Ed).Raman Spectroscopy for Chemical Analysis. Editorial: John Wiley & Sons.

La principal ventaja de ambas técnicas es que no son destructivas y que no se requiere a priori ningún tipo de tratamiento previo de la muestra. El empleo de éstas para la autentificación de en alimentos es muy extendido [**6**], en el caso de autentificación de aceite de oliva se pueden encontrar trabajos para: (i) discriminar aceite de oliva en función de su origen geográfico o variedad botánica [**7**,**8**,**9**]; (ii) detectar adulteraciones de aceite de oliva con otros aceites vegetales y cuantificar la proporción del mismo en mezclas [**10**,**11**,**12**,**13**]; y (iii) autentificar la calidad del fruto y del aceite de oliva [**14**,**15**].

La principal desventaja de la mayoría de los estudios realizados con estas técnicas radica en que se aplican habitualmente en mezclas binarias cuando se usan para la detección de la adulteración de aceite de oliva con otros aceites vegetales y la posterior determinación de la proporción del mismo en dichas mezclas. Ello implica que sólo se considera el aceite de oliva adulterado con un único aceite vegetal, con lo cual los métodos dejan de ser globales y sólo podrían ser usados cuando la adulteración sea con ese tipo de aceite vegetal. Eso quiere decir que sería necesario desarrollar un método, y un modelo de cuantificación, por cada tipo de aceite vegetal utilizado en la mezcla con aceite de oliva.

---

[6] Esteki, M., Shahsavari, Z., Simal-Gandara, J. (2018). Use of spectroscopic methods in combination with linear discriminant analysis for authentication of food products. Food Control, *91,* 100-112.

[7] Abdallah, M., Vergara-Barberán, M., Lerma-García, M.J., Herrero-Martínez, J.M., Simó-Alfonso, E.F., Guerfel, M. (2016). Cultivar discrimination and prediction of mixtures of Tunisian extra virgin olive oils by FTIR.European Journal of Lipid Science and Technology, *118*, 1236-1242.

[8] Sánchez-López, E., Sánchez-Rodríguez, M.I., Marinas, A., Urbano, F.J., Caridad, J.M., Moalem, M. (2016). Chemometric study of Andalusian extra virgin olive oils Raman spectra: Qualitative and quantitative information. Talanta, *156-157,* 180-190.

[9] Sinelli, N., Casale, M., Di Egidio, V., Oliveri, P., Bassi, D., Tura, D., Casiraghi, E. (2010). Varietal discrimination of extra virgin olive oils by near and mid infrared spectroscopy. Food Research International, *43,* 2126-2131.

[10] De Luca, M., Restuccia, D., Clodoveo, M.L., Puoci, F., Ragno, G. (2016). Chemometric analysis for discrimination of extra virgin olive oil from whole and stoned olive paste. Food Chemistry, *202*, 432-437.

[11] Sun, X., Lin, W., Li, X., Shen, Q., Luo, H. (2015).Detection and quantification of extra virgin olive oil adulteration with edible oils by FT-IR spectroscopy and chemometrics. Analytical Methods, *7,* 3939-3945.

[12] De la Mata, P., Domínguez-Vidal, A., Bosque-Sendra, J.M., Ruiz-Medina, A., Cuadros-Rodríguez, L., Ayora-Cañada, M.J. (2012). Olive oil assessment in edible oil blends by means of ATR-FTIR and chemometrics. Food Control, *23,* 449-455.

[13] Tiryaki, G.Y., Ayvaz, H. (2017).Quantification of soybean oil adulteration in extra virgin olive oil using portable Raman spectroscopy.Food Measure, *11,* 523-529.

[14] Gouvinhas, I., De Almeida, J.M.M.M., Carvalho, T., Machado, N., Barros, A. I.R.N.A. (2015).Discrimination and characterization of extra virgin olive oils from three cultivars in different maturation stages using Fourier transform infrared spectroscopy in tandem with chemometrics.Food Chemistry, *174*, 226-232.

[15] Guzmán, E., Baeten, V., Fernández Pierna, J.A., García-Mesa, J.A. (2012). A portable Raman sensor for the rapid discrimination of olives according to fruit quality.Talanta, *93,* 94-98.

Pero en la realidad, a priori, no se va a conocer qué aceite va a ser usado para adulterar el aceite de oliva por lo tanto es necesario desarrollar un **método global** que abarque la mayoría de las posibilidades. En este capítulo se desarrolla un método en el cual se tiene en cuenta la adulteración de aceite de oliva con más de un aceite vegetal obteniéndose mezclas de 10 aceites vegetales.

A continuación se especifica la instrumentación analítica, las muestras, así como el pre-tratamiento de los datos y los métodos de clasificación utilizados en este capítulo.

III.3. Instrumentación

Para llevar a cabo este estudio se utilizaron los siguientes equipos:

- ⇨ FT-IR (Nicolet iS5) equipado con ATR (iD7) con un divisor de haz de KBr/Ge y detector de sulfato de triglicinadeuterado de recuperación rápida (DTGS).
- ⇨ FT-NIR (Antaris II) equipado con una fibra de reflectancia difusa y un detector de arseniuro de indio y galio (InGaAs).
- ⇨ Raman (IDRAMAN Reader) con una emisión de 785 nm láser, cuya potencia era de 23.4 mW, para llevar a cabo la excitación. Para la detección se utilizó un detector de carga acoplada (CCD) potenciado, de matriz de 2048 elementos, con refrigeración termoeléctrica (TEC) a 10ºC.

III.4. Muestras

Se analizaron sesenta y nueve muestras de aceite de oliva de las cuales cincuenta y dos eran aceite de oliva virgen extra, cuatro aceites de oliva virgen, cinco de aceite de oliva y seis de orujo de oliva. De igual forma también se analizaron setenta y nueve muestras de aceites vegetales no oliva, entre los que se encontraban aceites de: avellana, cacahuete, canola, cártamo, germen de trigo, girasol, maíz, palma, pepita de uva, semillas, sésamo y soja.

También se analizaron un total de veintisiete mezclas de aceite de oliva con otros aceites vegetales, además cada mezcla implicaba de 5 a 10 aceites diferentes. En el artículo que se presenta al final de este capítulo aparece detallada la composición y el número de aceites utilizados en cada una de las mezclas preparadas.

Como se declaró en los capítulos anteriores, las muestras fueron sometidas a una reacción de transesterificación metílica previa al análisis por ambas técnicas, de acuerdo a la metodología inicialmente propuesta por Biedermann*et al.* [**16**].

---

[16] Biedermann, M.,Grob, K.,Mariani, C. (1993).Transesterification and on-line LC-GC for determining the sum of free and esterified sterols in edible oils and fats. Fett Wissenschaft Technologie (Fat Science and Technology), *95*, 127-133.

III.5. Condiciones experimentales

- Rango de medida: (i) 4000-550 cm^{-1} (ATR-FTIR), (ii) 4000-10000 cm^{-1} (FT-NIR) y (iii) 200-3200 cm^{-1} (Raman).
- Número de scans: 32 cuando se utilizó ATR-FTIR y FT-NIR y 3 scans cada 10 segundos en el caso de Raman.
- Número de análisis por muestra: por triplicadoen todos los equipos instrumentales.

III.6. Pre-procesado de los datos

En todos los casos se trabajó con el espectro promedio resultado de las medidas por triplicado de cada una de las muestras analizadas.

En el caso de las técnicas de infrarrojo el equipo instrumental no generaba automáticamente el promedio de las medidas realizadas por triplicado de cada muestra, por lo que en una primera etapa se realizó el agrupamiento de cada una de las tres medidas por muestra y se obtuvo el espectro promedio, tanto para los análisis llevados a cabo por FT-NIR como para los análisis por ATR-FTIR. En el caso de las medidas realizadas por Raman,esta etapa no fue necesaria ya que el propio instrumento generaba el espectro promedio y éste era exportado para su posterior tratamiento quimiométrico.

Una vez obtenidos los espectros promedios de cada una de las muestras, se agruparon y se superpusieron para generar las matrices de datos utilizadas para los diferentes modelos quimiométricos. El resto de etapas de pre-procesado de los datos como el suavizado y selección de las variables de interés para el desarrollo de los modelos de clasificación y cuantificación multivariable, pueden ser leídas en el artículo publicado que se presenta al final de este capítulo.

III.7. Métodos de clasificación multivariante aplicados

En primer lugar se desarrollaron tres modelos PCA, uno por cada una de las técnicas analíticas empleadas para observar tendencias y/o agrupaciones de cada los objetos analizados.

Seguidamente se aplicó el algoritmo de 'Kennard-Stone' (KS) [**17**] para seleccionar las muestras a utilizar en el entrenamiento de los modelos de clasificación. Una vez establecidos los sub-conjuntos para entrenamiento y validación externa de los modelos, se desarrollaron los modelos de clasificación multivariante aplicando técnicas de análisis discriminante: 'partial least squares-discriminant analysis' (PLS-DA), 'support vector machine-classification' (SVM-C), y 'k-nearest neighbours' (kNN), y técnicas de

---

[17] Kennard, R.W., Stone, L.A. (1969). Computer aided design of experiments. Technometrics, *11*, 137–148.

modelado de clases: 'soft independent modelling of class analogies' (SIMCA) y 'one class partial leasts quares' (OCPLS).

La calidad de las clasificaciones realizadas por cada uno de los modelos fue evaluado calculando los siguientes parámetros: sensibilidad, especificidad, valor predictivo positivo, valor predictivo negativo, índice de Youden, relación de verosimilitud para resultados positivos, relación de verosimilitud para resultados negativos, valor F, poder discriminante, eficiencia, área bajo la curva, coeficiente de correlación de Matthews, y coeficiente Kappa. Todos estos parámetros se calcularon a partir de los resultados obtenidos sobre el conjunto de validación externa.

Además para llevar a cabo la cuantificación de la proporción de aceite de oliva en mezclas con otros aceites vegetales se emplearon 'partial least squares' (PLS) y 'support vector-regression'' (SVR).

Para evaluar la bondad del ajuste del modelo se utilizó el coeficiente de determinación (R^2), mientras que para el desempeño de los modelos se determinaron: la media de los errores cuadráticos (RMSE), la media de los errores absolutos (MAE) y la mediana de los errores absolutos (MdAE), sobre la predicción del conjunto de validación externa [**18**].

En el caso de FT-NIR no se pudieron obtener señales de las muestras y no se pudo llevar a cabo las clasificaciones ni las cuantificaciones debido a la baja selectividad de dicha técnica instrumental.

A continuación se presenta el artículo publicado, donde se describen de forma más detallada de los estudios realizados.

[18] ASTM E1655-05 (2012). Standard practices for infrared multivariate quantitative analysis. ASTM International, West Conshohocken (PA).

ARTÍCULO CIENTÍFICO

Chemometric classification and quantification of olive oil in blends with any edible vegetable oils using FTIR-ATR and Raman spectroscopy

(Artículo publicado en 2017, en la revista: LWT-Food Science and Technology, *86*, 174-184)

LWT - Food Science and Technology 86 (2017) 174–184

Contents lists available at ScienceDirect

LWT - Food Science and Technology

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

Chemometric classification and quantification of olive oil in blends with any edible vegetable oils using FTIR-ATR and Raman spectroscopy

Ana M. Jiménez-Carvelo [a,*], María Teresa Osorio [b], Anastasios Koidis [b], Antonio González-Casado [a], Luis Cuadros-Rodríguez [a]

[a] Department of Analytical Chemistry, University of Granada, c/ Fuentenueva, s.n., E-18071 Granada, Spain
[b] Institute for Global Food Security, Queen's University, 18-30 Malone Road, Belfast BT9 5BN, Northern Ireland, UK

Chemometric classification and quantification of olive oil in blends with any edible vegetable oils using ftir-atr and raman spectroscopy

Ana M. JIMÉNEZ-CARVELO[1] ✉, María Teresa OSORIO[2], Anastasios KOIDIS[2], Antonio GONZÁLEZ-CASADO [1] , Luis CUADROS-RODRÍGUEZ [1]

[1] Department of Analytical Chemistry, University of Granada, c/ Fuentenueva, s.n. E-18071 Granada, Spain.

[2] Institute for Global Food Security, Queen's University, 18-30 Malone Road, Belfast BT9 5BN, Northern Ireland, UK

Abstract

Samples of olive oils (n=67) from different qualities and samples of other vegetable edible oils (including soybean, sunflower, rapeseed, corn oil etc; n=79) were used in this study as pure oils. Previous to spectroscopy analysis, a transesterification step was applied to the pure vegetable oil samples and all the different oil blends were then prepared to create in-house blended samples. Spectral acquisition was performed with typical parameters to collect the FTIR and Raman fingerprints. For the olive/non-olive classification model, three classification strategies have been applied: (i) one input-class (1iC) classification; (ii) two input-class (2iC) classification; and (iii) one input-class plus one 'dummy' class classification (or pseudo two input-class (p2iC) classification). The multivariate classification methods used were k-nearest neighbours (kNN), partial least squared-discriminant analysis (PLS-DA), one-class partial least squares (OCPLS), support vector machine classification (SVM-C), and soft independent modelling of class analogies (SIMCA). The multivariate quantification method used was partial least square-regression (PLS-R). FTIR fingerprints showed excellent classification ability to distinguish pure olive from non-olive oil. When PLS-DA or SVM-C techniques are

applied, 100% of olive oil samples and 92% of other vegetable edible oils are correctly classified. In general FTIR fingerprints were more discriminative than Raman's in both classification and regression scenarios.

Keywords

vegetable oils, discrimination, fingerprinting, pattern recognition, spectroscopic techniques

✉ Corresponding author: phone: +34 958240797; fax: +34 958243328; email: amariajc@ugr.es

1. Introduction

As a natural product that is produced using 'only mechanical means' from olive drupes, olive oil is protected by various regulations and institutions such as the EU Regulations (Regulation UE, 2016; Regulation UE, 2011; Commission Regulation EEC, 2016) and Codex Alimentarius (Codex Stan, 2015). Due to its increasing popularity, it has always been the target for fraudulent practises such as substitution fraud with cheaper oils (blends). To prevent that, authenticity of olive oil is described adequately in the legislation. The top two qualities of olive oil that exist are the extra-virgin and the virgin olive oil and both of them must comply to certain well defined physical, chemical and sensorial parameters. There are several standard methods that are used to determine these parameters. For example, with the use of chromatographic techniques detection of several major and minor constituents of olive oil (fatty acids, tocopherols, carotenoids etc.) is achieved. Nowadays rapid and novel methods are continuously developed (such as those based on spectroscopy), as alternatives to the standard methods offering speed, efficiency (less resources required) and accuracy in authenticity testing.

Actually, studies about authentication of olive oil using spectroscopic techniques are based on the application of chemometric tools to develop multivariate models that are able to differentiate pure olive oils from adulterated olive oil with other vegetable edible oil. Then, the proportion of olive oil in these blends is quantified; therefore, although blends of olive oil with other vegetable oils are allowed by the legislation, there is a restriction of labelling them as "olive oils" if the olive oil in the blend does not exceed 50% (Regulation UE, 2016). Consequently, a proper method of control must be established. (Sun, Lin, Li, Shen and Luo (2015) reported: (i) a principal component analysis (PCA) model to discriminate extra virgin olive oil from binary blends of olive oil with camellia oil, soybean oil, sunflower oil and corn oil; and (ii) a quantification

model using partial least squares (PLS) to quantify the olive oil in binary blends. López-Díez and Goodacre (2003) described a PCA model to differentiate pure extra virgin olive oil from adulterated olive oil with hazelnut oil, and a PLS model to quantify the amount of olive oil in the mixtures. Similar studies to the above mentioned ones are shown in Table 1. This table shows five papers using FTIR to detect adulteration of olive oil with other vegetable oil in blends binary, only Gurdeniz and Ozen (2009) develop a model to quantify olive oil in ternary blends. For Raman spectroscopy five works are reported, as in FTIR all the authors detect and quantify olive oil in blends binary, except Rohman and Che Man (2012) which quantifies olive oil in quaternary blends.

The main disadvantage of the reported models to authenticate olive oil using spectroscopic techniques, such as FTIR and Raman, is the low number of different botanical species used to build the blends of olive oil with other edible vegetable oils. Most authors employ a small set of oils to elaborate the blends, and sometimes using a single olive oil or a limited number of vegetable edible oil (non-olive oil) in the different mixtures prepared. For example, Tay, Singh, Krishnan and Gore (2002) reported a method to authenticate olive oil using only thirty two olive oil and seven vegetable edible oils (non-olive oil) to build the different blends (Tay et al., 2002). Thus, the resulting models cannot be considered as global methods to detect adulteration of olive oil (independently of the cultivars) with any edible vegetable oil. Moreover, some authors erroneously apply PCA as discriminant analysis technique to develop and validate classification models of olive oil (Sun et al.,2015). PCA is an unsupervised data analysis technique used to explore the variability in the dataset and to evaluate if there are different groups of samples when the dimensionality of the data decreases. This exercise should not be used for classification purposes. In the literature

there is only one published study where it is developed a classification model to distinguish pure olive oil from other pure vegetable oil using FTIR or Raman spectroscopy. De la Mata et al. (2012) reported a partial least squares discriminant analysis (PLS-DA) aiming to distinguishing between olive oil and binary mixture of non-olive samples applying ATR-FTIR.

The aims of this study are: (i) discrimination of pure olive oil/non-pure olive oil, (ii) detection of adulterated olive oil and (iii) quantification of olive oil in blends (from binary to heptenary mixtures) with other vegetable edible oils using a number of chemometric techniques. For this purpose, we have developed a global and comprehensive analytical method to differentiate, detect and quantify olive oil in blends with any edible oils. The number of oils used in this work is wide, and spread worldwide. Although, in the "real world" the usual blends of olive oil with other seed oil are binary, a quality control laboratory does not know which was and/or how many were the seed oils used in adulteration, if any. For this reason, the proposed method aims at covering binary and higher-order blends which could be found.

Table 1. Chemometric methods using FTIR or Raman for the authentication of olive oil found in the literature.

Nº	Analytical technique	Amount and types of edible oils	Blends	Aims	Chemo-metrics	Results (Quality Features)	Ref.
1	FTIR (4000-650 cm^{-1})	EVOO[t] (40), CAM[a] (5), SOY[b] (5), SUN[c] (5) and COR[d](5) oils	EVOO-CAM, EVOO- SOY, EVOO-SUN, EVOO-COR	Classification model of EVOO and binary blends of EVOO with edible oil Quantification of EVOO in binary blends	PCA and PLS	R^2 : 0.98 - 0.99 RMSE: 1.9 % (EVOO-SUN); 9.5% (EVOO-CA); 1.72 % (EVOO-SOY); 2.2% (EVOO-COR)	(Sun,Lin, Li, Shen & Luo, 2015)
2	FTIR (1200-900 and 2949-2885 cm^{-1})	EVOO (1), GSO[e] (1), RBO[f] (1), WO[g] (1) oils	GSO-WO, EVOO-RBO, EVOO-RBO-GSO-WO	Quantification of EVOO in quaternary mixture	PLS	R^2 : 0.99 ; RMSE : 3.7%	(Rohman & Che Man, 2011)
3	FTIR (1207-1018, 1517-1222 and 3050-2927 cm^{-1}) GC	EVOO, CAN[h], COR, GSO, SOY, SES[i], SUN and WO oils	EVOO-SES	Classification model of EVOO and other pure edible oil based on their fatty acids profiles. Quantification of EVOO in blends of EVOO-SES	PLS and PCR	R^2 : 1.00 ; RMSE : 7.0% (PLS) R^2 : 0.997; RMSE: 1.1% (PCR)	(Rohman & Che Man, 2012)

[a] Camellia; [b] Soybean; [c] Sunflower; [d] Corn; [e] Grapeseed; [f] Rice bran; [g] Walnut; [h] Canola; [i] Sesame; [j] Hazelnut; [k] Corn germ; [l] Rape seed; [m] Garlic; [n] Bean with Omega 3; [o] Safflower; [p] Wheat germ; [r] flaxseed; [s]cottonseed, [t] Extra virgin olive oil, [u] Pomace olive oil

RMSE: Root mean square error; MAE: Mean absolute error

Table 1. *Continue.*

Nº	Analytical technique	Amount and types of edible oils	Blends	Aims	Chemo-metrics	Results (Quality Features)	Ref.
4	FTIR (4000-1000 cm^{-1})	EVOO (6), HAZ[j] (6), SUN (6), COR (3), COG[k] (2) and SOY (6) oils	EVOO-HAZ, EVOO-SUN, EVOO-CORN, EVOO-SOY	Classification of vegetable oils using LDA Determination of EVOO adulteration	LDA and MLR	R^2 : 0.91; MAE: 2.0 (EVOO-HAZ) R^2 : 0.99; MAE: 1.7 (EVOO-SUN) R^2 : 0.99; MAE: 1.5 (EVOO-CORN) R^2 : 0.98; MAE: 1.9 (EVOO-SOY)	(Lerma-García, Ramis-Ramos, Herrero-Martínez & Simó-Alfonso, 2010)
5	FTIR (3080-2800 cm^{-1})	EVOO (25), COR, SUN, RPS and COT[s] oils	EVOO-SUN-COR, EVOO-COT, EVOO-RPS	Classification model of EVOO and adulterated EVOO Quantification of EVOO in the mixtures	SIMCA and PLS	R^2 : 0.99; RMSE: 10.4% (EVOO-SUN-COR) R^2 : 0.95; RMSE: 14.0% (EVOO-COT) R^2 : 0.93; RMSE: 13.2% (EVOO-RPS)	(Gurdeniz & Ozen, 2009)
6	Raman (1000-3000 cm^{-1})	EVOO (31) and HAZ (10) oils	EVOO-HAZ	Quantification of EVOO in blends with hazelnut oil	PCA and PLS	$R^2 = 0.98$ RMSE 10.94%	(López-Diez , Bianchi & Goodacre, 2003)
7	Raman (800-1800 and 2850-3020 cm^{-1})	EVOO (18), RPS[l], SES, GAR[m], BOM[n], SUN (3), WO, SAF[o] (2), SOY, WGE[p], and FLA[r] oils	EVOO-SUN#1, EVOO-SUN#2, EVOO-SUN#3	Discrimination model of EVOO and adulterated EVOO. Estimation of the SUN oil content in EVOO	PCA and PLS	R^2 : 0.99; RMSE(cross-valid): 9.81% R^2 : 0.99; RMSE(cross-valid): 9.88% R^2 : 0.98; RMSE (cross-valid): 9.71%	(El-Abassy, Donfack & Materny, 2009)
8	Raman (1000-1800 cm^{-1})	EVOO (6), POO[u] (1), SOY (3), SUN (3), RPS (2) and COR (2) oils	EVOO-SOY, EVOO-SUN, EVOO-RPS, EVOO-COR	Model to detect adulterated EVOO	PCA	Intensity ratio	(Zhou et al., 2009)
9	Raman (800-1800 cm^{-1})	EVOO (5), SOY (3), corn (3) and SUN (3) oils	EVOO-SUN, EVOO-SOY, EVOO-COR	Quantification of EVOO in binary blends	Bay-LS-SVM, LS-SVM and PLS	R^2 : 0.99; RMSE: 5.1% (Bay-LS-SVM) R^2 : 0.99; RMSE: 6.9% (LS-SVM) R^2 : 0.99; RMSE: 8.4% (PLS)	(Dong, Zhang, Zhang & Wang, 2012)
10	Raman (1000-1800 cm^{-1})	EVOO and SOY oil	EVOO-SOY	Quantification of SOY adulteration in EVOO	PLS	$R^2 = 0.99$; RMSE: 1.3%	(Tiryaki & Ayvaz, 2016)

[a] Camellia; [b] Soybean; [c] Sunflower; [d] Corn; [e] Grapeseed; [f] Rice bran; [g] Walnut; [h] Canola; [I] Sesame; [j] Hazelnut; [k] Corn germ; [l] Rape seed; [m] Garlic; [n] Bean with Omega 3; [o] Safflower; [p] Wheat germ; [r] flaxseed; [s]cottonseed, [t] Extra virgin olive oil, [u] Pomace olive oil

RMSE: Root mean square error; MAE: Mean absolute error

2. Materials and methods

2.1. Chemicals

Isopropanol, *n*-hexane, methanol and tert-butyl methyl ether (TBME) were purchased from VWR International Eurolab, S.L. (Barcelona, Spain) and all of them were of HPLC grade. Other reagents, such as sodium methoxide, citric acid monohydrate, and anhydrous sodium sulphate were purchased from Merck (Darmstadt, Germany). The nitrogen (99.9999 %) used was provided by Air Liquid (Madrid, Spain).

2.2. Instrumentation

FT-IR spectra were obtained on a NICOLET iS5 spectrometer (Thermo Scientific, Waltham, Massachusetts, USA) equipped with a DTGS detector and KBr beam splitter. Spectra were obtained in the range of 4000 cm^{-1} to 550 cm^{-1} with a resolution of 2 cm^{-1} using a monolithic diamond attenuated total reflectance (ATR iD7) accessory. All the spectra were recorded at room temperature with 32 scans.

Raman measurements were carried out using IDRAMAN Reader (Ocean Optics, Oxford, UK) with 785 nm emission of a laser (23.4 mW at sample) for excitation. The laser was focused on the sample contained in 2 mL vial. For signal detection, a 2048-element NIR-enhanced CCD array with thermoelectric cooling to 10 °C was employed. An averaged spectrum for each sample was recorded in the range of 200 to 3200 cm^{-1}, using an integration time of 10 s each 3 scans.

NIR spectra were obtained using Antaris II (Thermo Electron Corporation, Waltham, Massachusetts, USA) FT-NIR analyzer, equipped with a diffuse reflection fibre optic and InGaAs detector. All the spectra, in the range of 4000 to 10000 cm^{-1}, were recorded at room temperature with 32 scans.

In all cases, each sample was analysed in triplicate.

2.3. Samples

Pure vegetable edible oils used to the classification models

67 samples of olive oils and 79 samples of other vegetable edible oils were used in this study. The samples of olive oils were constituted by 52 extra virgin olive oils (EVOO) samples, including 41 samples from 10 different monovarietals ("Arbequina", "Hojiblanca", "Picual", "Royal", "Manzanilla", "Cornicabra", "Empeltre", "Frantoio", "Verdial" and "Blanqueta") and 26 samples of varietal mixtures, 4 virgin olive oil samples (VOO), 5 olive oils, blend of virgin and refined (OO) and 6 pomace olive oil samples (POO). Vegetable edible oil samples (non-olive oils) consisted of 8 hazelnut oils, 5 peanut oils, 10 canola oils, 2 safflower oils, 12 sunflower oils, 2 flax oils, 5 corn oils, 9 palm oils, 8 seeds oils (marketing mixture of unidentified seeds), 4 sesame oils, 8 soybean oils, 1 wheat oil and 4 grapeseed oils. In addition, a speciality olive oil extracted from previously dehydrated olive fruits was also added in this group. All samples were collected from marketed edible oils, purchased in food stores and sourced from respective partners from multiple geographical locations.

Blends of olive oil with other vegetable edible oils

To build the blends were used 27 olive oil samples, of which 22 EVOO (including 16 monovarietal oils), 3 VOO and 2 OO. In addition, 52 edible oils samples of 8 botanical origins, obtained each one from different suppliers, were used: 8 soybean oils, 11 sunflower oils, 10 rapeseed (canola) oils, 5 corn oils, 5 seeds oils (commercial blends of unknown seed oils), 5 peanut oils, 4 sesame oils and 4 grapeseed oils. Table 2 shows details on the composition of the different blends.

All the oil samples were stored at 4 ºC until the sample preparation in order to provide realistic testing conditions.

Table 2. Percentage and composition of the olive oil and other vegetable edible oil in the oil blend samples.

Nº		Composition
(a) Calibration set		
1	100% MDVO[a]	20% seed#1 oil, 20% peanut#1 oil, 20% sunflower#1 oil, 20% canola#1 oil and 20% corn#1 oil
2	100% MDVO	20% soybean#1 oil, 20% soybean#2 oil, 20% sunflower#2 oil, 20% canola#2 oil and 20% grapeseed#1 oil
3	100% MDVO	20% seed#2 oil, 20% sesame#1 oil, 20% peanut#2 oil, 20% corn#2 oil and 20% grapeseed#2 oil
4	20% EVOO[b] + 80% MDVO	10% EVOO#1, 5% EVOO#2, 5% EVOO#3, 13% soybean#3 oil, 13% canola#3 oil, 13% corn#1 oil, 13% seed#3 oil, 13% grapeseed#3 oil and 13% peanut#2 oil
5	15% EVOO + 5% OO[c] + 80% MDVO	5% OO#4, 5% EVOO#5, 5% EVOO#6 , 5% EVOO#7, 8% sunflower#3 oil, 8% sunflower#4 oil, 8% canola#1 oil, 8% canola#4 oil, 16% corn#3 oil, 16% sesame#2 oil and 16% peanut#3 oil
6	15% EVOO + 5%VOO[d] + 80% MDVO	5% EVOO#8, 5% EVOO#9, 5% EVOO#11, 5% VOO#10, 13% sunflower#5 oil, 13% sunflower#2, 26% corn#2 oil and 26% grapeseed#2 oil
7	30% EVOO + 10% OO + 60% MDVO	10% OO#12, 10% EVOO#5, 10% EVOO#7, 10% EVOO#14, 15% soybean#4 oil, 15% canola#5 oil, 15% seed#3 oil and 15% peanut#4 oil
8	30% EVOO + 10% OO + 60% MDVO	10% EVOO#6, 10% EVOO#13, 10% EVOO#8, 10% OO#4, 15% sunflower#6 oil, 15% canola#5 oil, 15% corn#4 oil and 15% grapeseed#2 oil
9	30% EVOO + 10% VOO + 60% MDVO	10% EVOO#1, 10% EVOO#3, 10% EVOO#2, 10% VOO#15, 15% sunflower#7 oil, 15% corn#1 oil, 15% sesame#1 oil and 15% peanut#4 oil
10	60% EVOO + 40% MDVO	15% EVOO#6, 15% EVOO#7, 15% EVOO#13, 15% EVOO#14, 5% soybean#1 oil, 5% soybean#5 oil, 10% canola#6 oil, 10% sesame#3 oil and 10% grapeseed#4 oil

Table 2. *Continue.*

Nº		Composition
11	36% EVOO + 12% OO + 12% VOO + 40% MDVO	12% EVOO#6, 12% EVOO#2, 12% EVOO#5,12% OO#12, 12% VOO#15, 8% canola#7 oil, 8% corn#5 oil, 8% seed#4 oil, 8% grapeseed#8 oil and 8% peanut#3 oil
12	40% EVOO + 10% VOO + 10% OO + 40% MDVO	10% EVOO#9, 10% EVOO#11, 10% EVOO#1, 10% EVOO#8, 10% VOO#10, 10% OO#12, 7% sunflower#8 oil, 6.6% canola#8 oil, 6.6% corn#2 oil, 6.6% sesame#2 oil, 6.6% seed#2 oil and 6.6% peanut#5 oil
13	40% EVOO + 20% VOO + 20% OO + 20% MDVO	20% EVOO#5, 20% EVOO#2, 20% VOO#15, 20% OO#12, 5% sunflower#9 oil, 5% corn#3 oil, 5% seed#1 oil and 5% grapeseed#3 oil
14	80% EVOO + 20% MDVO	30% EVOO#6, 25% EVOO#7, 25% EVOO#9, 5% seed#5 oil, 5% peanut#1 oil, 5% canola#9 oil and 5% canola#2 oil
15	60% EVOO + 20% OO + 20% MDVO	20% EVOO#11, 20% EVOO#1, 20%EVOO#13, 20% OO#12, 5% soybean#6 oil, 5% corn#1 oil, 5% sesame#4 oil and 5% grapeseed#1 oil
16	100% EVOO	100% EVOO#16
17	100% EVOO	100% EVOO#17
18	100% VOO	100% VOO#18
	(b) Validation set	
1	68% EVOO + 32% MDVO*	68% EVOO#6, 25% corn#5 oil, 3% peanut#3 oil and 4% grapeseed#4 oil
2	17.50% VOO + 82.50% MDVO	17.50% VOO#15, 17% sunflower#8, 11% soybean#4 oil, 28% canola#6 oil, 26% peanut#1 oil and 0.5% seed#5 oil
3	93% VOO + 7% MDVO	93% VOO#19, 2% corn#3 oil and 5% sesame#4 oil
4	44% EVOO + 56% MDVO	44% EVOO#20, 13% peanut#3 oil, 8% canola#5 oil and 35% canola#4 oil
5	5% EVOO + 95% MDVO	5% EVOO#27, 40% canola#9 oil, 23% soybean#2 oil, 7% grapeseed#2 oil, 15% canola#4 oil and 10% sunflower#3 oil

Table 2. *Continue.*

Nº		Composition
6	68% EVOO + 32% MDVO	68% EVOO#21, 10% sesame#4 oil, 7% soybean#7 oil and 15% seed#3 oil
7	70% VOO + 30% MDVO	70% VOO#10, 1% sunflower#2 oil, 9% sesame#1 oil, 17% corn#1 oil and 3% sunflower#1
8	31% EVOO + 69 MDVO	31% EVOO#22, 24% sunflower#3 oil, 13% sesame#4 oil, 20% soybean#7 oil, 2% peanut#5 oil and 10% grapeseed#2
9	52% EVOO + 48% MDVO	52% EVOO#23, 28% canola#7, 13% soybean#6 oil, 5% grapeseed#1 oil and 2% sesame#4 oil
10	25% EVOO + 75% MDVO	25% EVOO#24, 25% corn#1 oil, 25% sunflower#2 oil and 25% peanut#3 oil
11	90% EVOO + 10% MDVO	90% EVOO#25, 5% canola#2 oil and 5% soybean#2 oil
12	40% EVOO + 60% MDVO	40% EVOO#26, 30% peanut#1 oil and 30% canola#6

[a] *MDVO: Mixture of different vegetable edible oils (non-olive oils)*
[b] *EVOO: Extra virgin olive oil*
[c] *OO: Olive oil*
[d] *VOO: Virgin olive oil*

2.4. Sample preparation

Previous to the spectrometric analysis, a transesterification reaction was applied to the pure vegetable oil samples and all the different oil blends prepared. This reaction was carried out using 0.1 g/mL sodium methoxide in a methanol/TBME mixture, 4:6 (mL:mL), and then the extraction was performed with *n*-hexane. In this alkaline medium, the free fatty acids presents in the oil are not methylated {(Li & Watkins, 2001)}. A modification of the original procedure described by Biedermann *et al.* was applied (Bierdermann *et al.*, 1993). A detailed description of the procedure followed is described elsewhere (Jímenez-Carvelo, Pérez-Castaño, González-Casado & Cuadros-

Rodríguez, 2017). The subsequent solution was stored at –25º C until analysis with less than 5% headspace under nitrogen.

2.5. Chemometrics

The FTIR and FT-NIR raw data files were exported to MATLAB (Mathworks, Massachusetts, USA, version R2013a). In order to reduce the variability associated to the intensity and derived from baseline, or other sources such as scattering effects, source or detector variations, or other general instrumental sensitivity effects, *standard* normal variate (SNV) and smoothing applying the Savitzky-Golay algorithm (second order polynomial filter with a 9-point window and first derivative) were used. Different chemometric tools have been applied for classification, including k-nearest neighbours (kNN), partial least squares discriminant analysis (PLS-DA), support vector machine-classification (SVM-C), one-class partial least squares (OCPLS) and soft independent modelling of class analogies (SIMCA). The classification results from each method have been evaluated on the basis of several quality metrics, such as: (i) sensitivity, (ii) specificity, (iii) positive (or precision) and negative predictive values, (iv) efficiency (or accuracy), (v) AUC (area under the receiver operating curve), (vi) Matthews correlation coefficient and (vii) Kappa coefficient. The meaning and way to calculate these metrics was recently reviewed (Cuadros-Rodríguez, Pérez-Castaño & Ruiz-Samblás, 2016).

Partial least squares regression (PLS-R) has been applied for quantification. Root Mean Square Error of Validation (RMSEV), Mean Absolute Error of Validation (MAEV) and Median Absolute Error of Validation (MdAEV) were used for accuracy assessment of the quantification methods.

Olive/non-olive classification models

Three classification strategies have been applied: (i) one input-class (1iC) classification; (ii) two input-class (2iC) classification; and (iii) one input-class plus one 'dummy' class classification (or *pseudo* two input-class (*p*2iC) classification).

The main difference between the strategies pursued is the number of class used to build the classification model. More detailed information can be found at the references Jímenez-Carvelo, Pérez-Castaño, González-Casado and Cuadros-Rodríguez (2017) and Jímenez-Carvelo, González-Casado, Pérez-Castaño and Cuadros-Rodríguez (2017).

For each strategy applied (2iC, *p*2iC and 1iC) the original vector data set of pure vegetable oil was divided into different groups to perform the classification model. The selection was carried out using the Kennard-Stone (KS) algorithm (Kennard & Stone, 1969)]. For 2iC, the training set was made up of 98 samples (44 olive oils and 54 non-olive oils), and the remaining oil samples (23 olive oils and 25 non-olive oils) composed the validation set. For *p*2iC, the training set which was made up of 61 samples (44 olive oils and 17 analytical blanks), and the validation set composed by 102 samples (23 olive oils and 79 non-olive oils). For 1iC, the training set was composed by 44 olive oils samples and the validation set by 102 samples (23 olive oils and 79 non-olive oils).

Once it was done, the classification models were developed. PLS_Toolbox (version 8.02, Eigenvector Research, Wenatchee, WA) for MATLAB environment was applied for reducing of variables and classification methods: principal component analysis (PCA) (Bro, 2014), k-nearest neighbours (kNN) (Steinbach & Tan, 2009), partial least squares-discriminant analysis (PLS-DA) (Ballabio & Consonni, 2013) soft independent modelling of class analogies (SIMCA) (Bylesjö et al., 2006), and support vector machine-classification (SVM-C) (Luts et al., 2010). Moreover, one-class partial least squares classification (OCPLS) (Xu, Yan, Cai & Yu, 2013) was performed applying the

three variants of the function: (i) conventional ordinary linear OCPLS, (ii) nonlinear radial basis function (RBF) OCPLS, and (iii) partial robust M-regression (PMR) OCPLS, using the software provided by Xu *et al.*, (2013). All the options offered by the software were tested.

Adulterated olive oils detection models

Classification models to detect adulterations of olive oil with other vegetable edible oils were developed in order to apply a screening classification method previous to carry out the quantification. In this case, it was only applied the 2iC strategy. The training set was composed by 56 samples (44 pure olive oils and 12 adulterated olive oils) and validation set was made up of 35 samples (23 pure olive oils and 12 adulterated olive oils). As in the olive/non-olive classification models, PCA was used to reduce the variables, kNN, PLS-DA, SIMCA, SVM-C and OCPLS techniques were applied to developed the different classification models.

Olive oil quantification model

The original analytical data were divided in different groups to perform the statistical analysis. The calibration data set was made up of 18 samples whose adulteration levels were from 20 to 80 (g olive oil/100 g blend oil). The validation set for olive oil quantification was composed of 12 samples. The composition of the different samples is shown in Table 2. Partial least squares-regression (PLS-R) was used to build the model of quantitative prediction using standard parameters.

3. Results and discussion

Figure 1 shows a spectral fingerprint of the transesterified fraction of an EVOO sample with the three different spectroscopic techniques used. The FTIR and Raman fingerprints of EVOO show strong molecular vibrations and good variability between samples, the NIR fingerprints appear non selective. Therefore, classification and quantification models in FT-NIR were not developed due to the low specificity of the spectra of the transesterified fraction from the different vegetable edible oil samples.

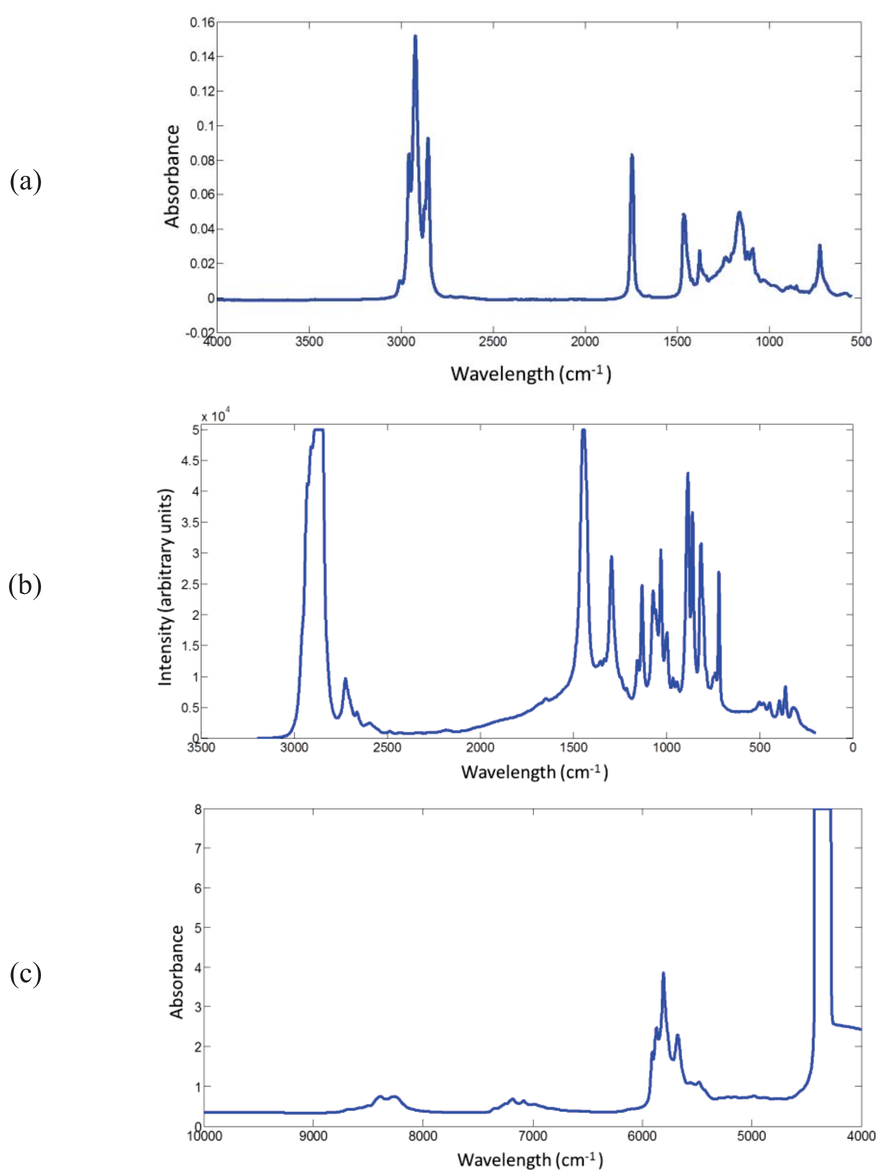


Figure 1. Examples of vibrational spectra of extra virgin olive oil (EVOO) acquired from: (a) FTIR (b) Raman and (c) FT-NIR.

3.1. Selection of variables

In order to reduce the number of variables and visualise the data a PCA model was obtained using FTIR and Raman fingerprints. In both spectroscopic techniques the selection of variable was performed examining the PCA loading plot. For that purpose, the regions of the spectra where the intensity of the loading was high were selected. Although the initial region of the Raman spectrum (2900-2800 cm^{-1}) shows a high value of the loadings, it was not finally selected since it did not improve the performance of the classification and quantification models.

The PCA model from FTIR data was developed with four principal components (PCs) which explain 98.87% of the variance. Figure 2 shows both the plot for FTIR spectrum and PCA loading plot with the three regions selected. The frequencies of the regions 1, 2 and 3 were 3100-2700 cm^{-1}, 1800-1600 cm^{-1} and 1205-1080 cm^{-1} respectively.

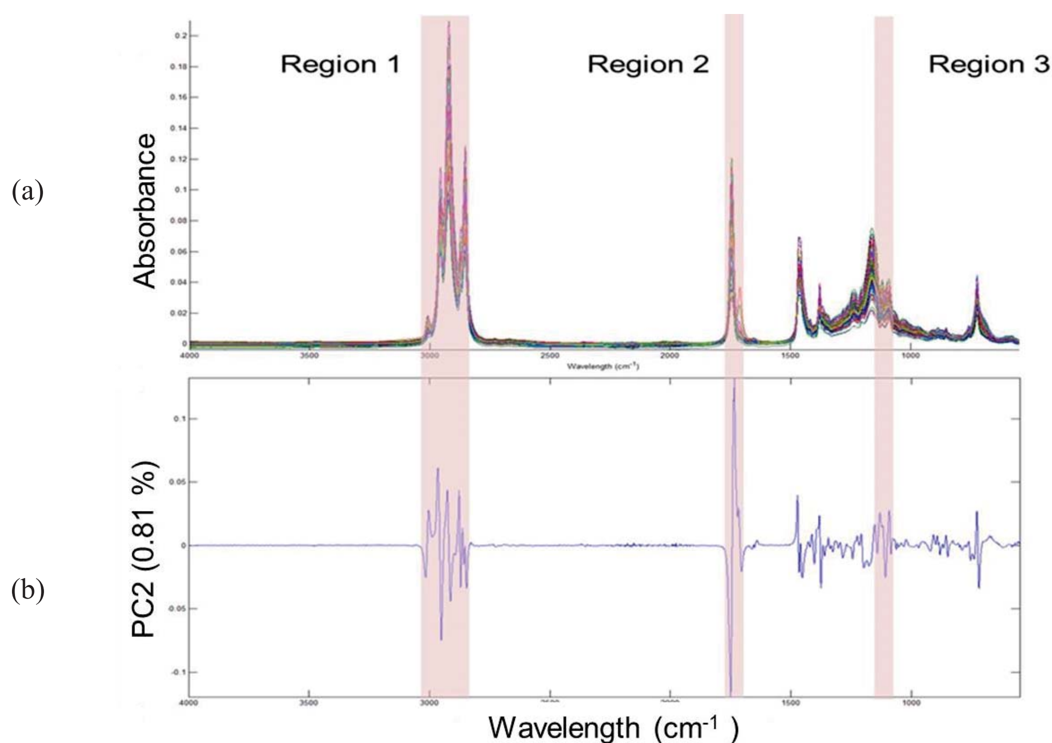


Figure 2. (a) Superposed FTIR spectra and (b) loading plot of the 146 vegetable oil samples showing the three regions selected.

The Raman spectra of all 146 samples were recorded. For the PCA model, four PCs were enough to explain 99.84% of the variance. Only the 950-650 cm^{-1} range, as shown in Figure 3, was chosen for analysis.

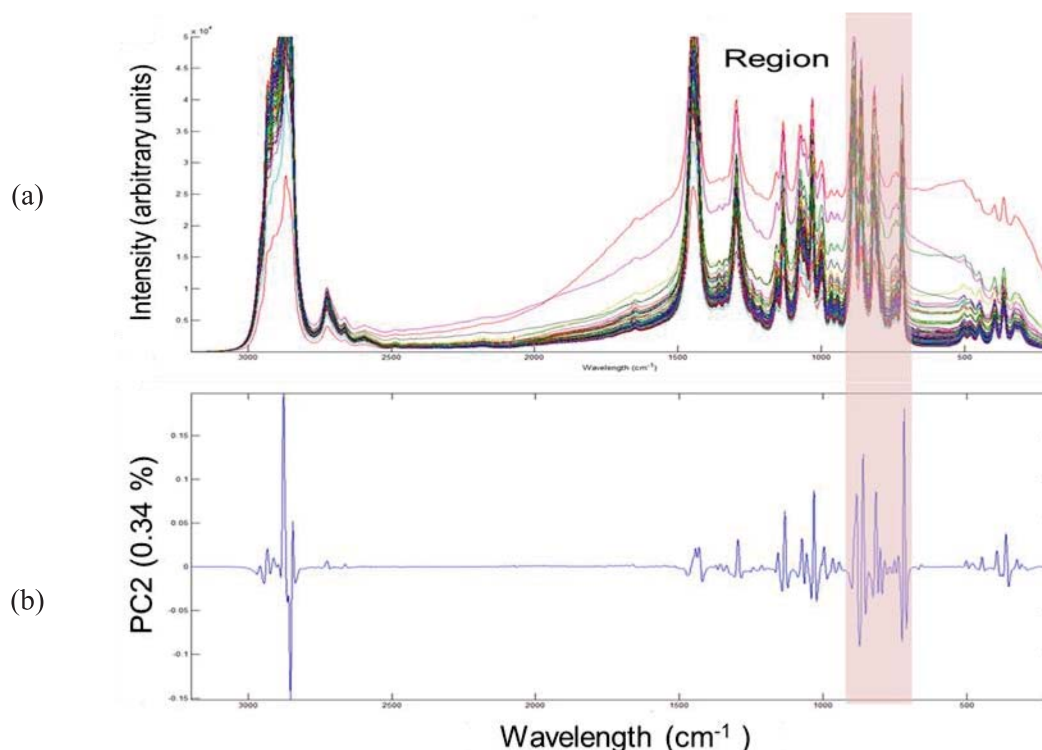


Figure 3. Plot of (a) superposed Raman spectra and (b) loading plot of all the vegetable edible oil samples showing the region selected.

3.2. Olive/non-olive classification models

FTIR

In order to differentiate pure olive oils from other pure vegetable edible oils, different models were tested using the three regions selected; however, the best performance statistics were obtained for the models generated using the region 2.

The two-input class (2iC) strategy was used to develop the model applying the chemometric methods: kNN, PLS-DA, SVM-C and SIMCA. One-input class (1iC)

strategy was applied when OCPLS and SIMCA models were performed and lastly, *pseudo* two-input class (*p*2iC) strategy was only applied to SIMCA model.

The target class was "olive oil" and the non-target class was "non-olive". In kNN, PLS-DA and SVM-C the olive class was assigned to samples with a predicted probability value equal to 1 and the non-olive class was defined by samples with a probability of 0. K=3 was enough to decide the neighbour distance in the kNN model. Classification of the samples of the validation set was performed directly by the software. There were only five samples misclassified, two olive oils samples and three non-olive oils samples (canola, peanut and hazelnut oils).

The PLS-DA model was built using six latent variables (LV), with 75.68% of the variance explained. Only one sample was not well classified corresponding to non-olive oil (canola oil). The efficiency, area under the receiver operating curve (AUC) and Matthews's correlation coefficient were 0.98, 0.98 and 0.96 respectively. Figure 4 shows the classification plot obtained from the PLS-DA method.

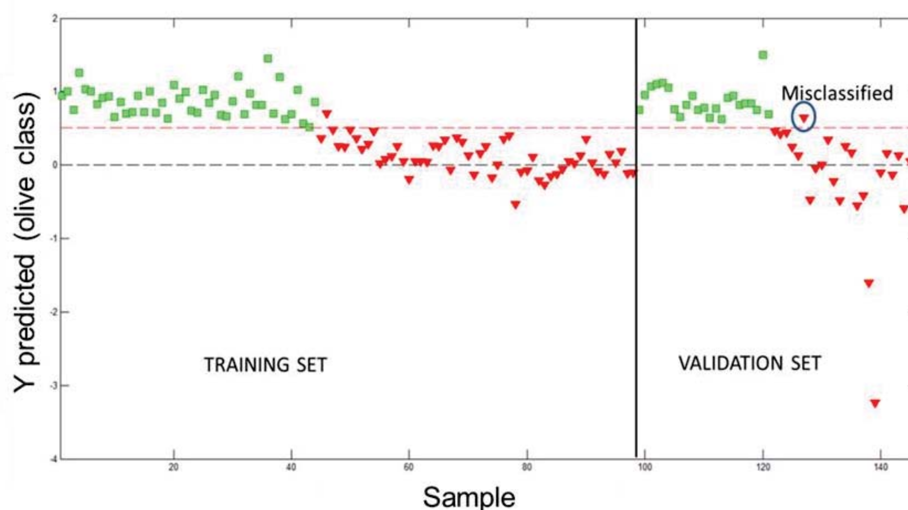


Figure 4. Classification plot from FTIR applying two input-class (2iC) classification strategy on PLS-DA. The green squares (■) and the red triangles (▼) represent the olive and non-olive class, respectively.

The SVM-C model was developed optimizing the "C" and "nu" operational parameters. There are two commonly used versions of SVM classification, 'C-SVC' and 'nu-SVC'. “C” represents the penalty associated with errors. “Nu” is an alternative parameter for specifying the penalty associated with errors. It indicates a lower bound on the number of support vectors to use, given as a fraction of total calibration samples, and an upper bound on the fraction of training samples which are errors (misclassified) (SVM Function Settings, Eigenvector Documentation wiki. URL http://wiki.eigenvector.com/index.php?title=Svmda. Accessed 13.06.17). The results obtained in all the cases were similar. Moreover, all the models were tested with and without variable reduction using PCA and PLS. This variable reduction is named X-block compression by PLS_Toolbox software. The best results were obtained when an X-block compression with PLS was applied. The samples were directly classified by the software.

The 2iC, *p*2iC and 1iC strategies were tested to generate the SIMCA models. Classification of samples was performed using means of the normalised (also called, reduced) statistics values of residual-Q_r and Hotelling-T^2_r (Marini, 2010). The samples with values lower than 1 for both statistics were classified as olive oil. Firstly, a 2iC SIMCA classification was carried out. PCA model was built with 4 PCs and 5 PCs for olive and non-olive oil classes, respectively. Secondly, a *p*2iC SIMCA model was developed. In this case, 4 PCs and 3 PCs were chosen for olive oil and 'dummy' classes respectively. At last, a 1iC SIMCA was performed with a 4 PCs for olive oil model.

An OCPLS class-model was also developed. Conventional OCPLS was built with 4 LVs, RBF OCPLS with 5 LVs and PMR OCPLS with 6 LVs. The regions for the samples classification were pre-established by the software.

To sum up, the 2iC strategy gave good results for all the discriminant methods. PLS-DA and SVM with reduction of variable using PLS were the best models; yielding the same classification results. The sensibility and specificity of all models were 1.00 and 0.96 respectively. In contrast, to SIMCA model lead to better classification results when 1iC strategy was used. The results for each model are shown in Table 3.

Table 3. Values of the quality performance metrics from the different FTIR olive/non-olive classification methods.

Performance features Region 2 (1680-1800 cm^{-1})	kNN	PLS-DA	(nu)SVM-C			(c)SVM-C			SIMCA			OCPLS		
			None	PCA	PLS	None	PCA	PLS				Ordinary linear	RBF	PRM
	2iC	2iC	2iC	2iC	2iC	2iC	2iC	2iC	2iC	*p*2iC	1iC	1iC	1iC	1iC
Sensibility (Recall)	0.91	1.00	1.00	0.83	1.00	1.00	0.87	1.00	0.91	0.87	0.96	0.74	0.87	0.57
Specificity	0.88	0.96	0.88	0.88	0.96	0.92	0.84	0.96	0.32	0.76	0.71	0.80	0.20	0.94
Positive predictive value (Precision)	0.88	0.96	0.88	0.86	0.96	0.92	0.83	0.96	0.55	0.51	0.49	0.52	0.24	0.72
Negative predictive value	0.92	1.00	1.00	0.85	1.00	1.00	0.88	1.00	0.80	0.95	0.98	0.91	0.84	0.88
Efficiency (Accuracy)	0.90	0.98	0.94	0.85	0.98	0.96	0.85	0.98	0.60	0.78	0.76	0.78	0.35	0.85
AUC (Correctly classified rate)	0.90	0.98	0.94	0.85	0.98	0.96	0.85	0.98	0.62	0.81	0.83	0.77	0.54	0.75
Matthews correlation coefficient	0.79	0.96	0.88	0.71	0.96	0.92	0.71	0.96	0.29	0.54	0.56	0.48	0.08	0.55
Kappa coefficient	0.79	0.96	0.88	0.71	0.96	0.92	0.71	0.96	0.23	0.50	0.50	0.46	0.04	0.54

Raman

In a similar way to FTIR, the 2iC strategy was applied with all the chemometric methods, *p*2iC strategy only with SIMCA and 1iC strategy with SIMCA and OCPLS. The classification criteria were the same as for FTIR with the different chemometric methods.

kNN classification model was built with k=3. Seven oil samples were misclassified (4 olive oils and 3 non-olive oils). The values of the quality performance metrics were similar with those obtained from FTIR models. Four LVs explaining 99.99% of the variance were enough to develop the PLS-DA model. This model was less efficient than PLS-DA model from FTIR.

As in the previous case of FT-IR spectra, SVM-C classification models were developed and tested with and without X-Block compression (reduction of variables). The results of all the models were the same excepting the (nu)-SVM-C model with reducing the variable by PLS. In this model all the samples were classified in both classes (olive and non-olive oil classes) and the values of the quality performance metrics were not satisfactory.

The SIMCA classification model was built with 3 PCs and 4 PCs for olive and non-olive oil classes applying the 2iC strategy. On the contrary to FTIR model, this SIMCA model classified better the samples of the validation set. 3 PCs for each class were enough to develop the *p*2iC SIMCA model. This model classified all the oil samples in the class of non-olive oils. At last, the 1iC SIMCA model was built with 4 PCs for olive oil model.

OCPLS classification models were developed. In this case partial robust M-regression (PMR) OCPLS was the best model.

As in the case of FT-IR, the discriminant analysis methods gave good classification results; PLS-DA model was the best model. In contrast with the results for FTIR, SIMCA provided better results when the 2iC strategy was applied. Table 4 shows the results for each model.

Table 4. Values of the quality performance metrics from the different Raman olive/non-olive classification methods.

Performance features (650-950 cm^{-1})	kNN	PLS-DA	(nu)SVM-C			(c)SVM-C			SIMCA			OCPLS		
			None	PCA	PLS	None	PCA	PLS	2iC	*p*2iC	1iC	Ordinary linear	RBF	PRM
	2iC	2iC	2iC	2iC	2iC	2iC	2iC	2iC				1iC	1iC	1iC
Sensibility (Recall)	0.83	0.88	0.83	0.83	1.00	0.83	0.83	0.83	0.67	0.00	0.50	0.42	0.46	0.33
Specificity	0.88	0.88	0.88	0.88	0.00	0.88	0.88	0.88	0.88	1.00	0.94	0.29	0.23	0.95
Positive predictive value (Precision)	0.87	0.88	0.87	0.87	0.49	0.87	0.87	0.87	0.84	-	0.71	0.15	0.15	0.67
Negative predictive value	0.85	0.88	0.85	0.85	-	0.85	0.85	0.85	0.73	0.76	0.86	0.62	0.58	0.82
Efficiency (Accuracy)	0.86	0.88	0.86	0.86	0.49	0.86	0.86	0.86	0.78	0.76	0.83	0.32	0.28	0.80
AUC (Correctly classified rate)	0.86	0.88	0.86	0.86	0.50	0.86	0.86	0.86	0.77	0.50	0.72	0.36	0.34	0.64
Matthews correlation coefficient	0.71	0.76	0.71	0.71	-	0.71	0.71	0.71	0.56	-	0.50	-0.25	-0.29	0.37
Kappa coefficient	0.71	0.76	0.71	0.71	0.00	0.71	0.71	0.71	0.55	0.00	0.48	0.24	-0.19	0.34

The hyphen "-" is signifying that the performance feature cannot be determined

3.3. Adulterated olive oils detection models

Discriminant analysis and class-modelling methods were used for the discrimination of pure EVOO and EVOO adulterated with several vegetable edible oils. The chemometric techniques used and the criteria for classification were the same that to olive/non-olive classification models. Table 5 and 6 show the classification results of the different models tested from FTIR and Raman techniques. From FTIR, the best results were obtained when PLS-DA was applied. On the contrary, from Raman, the best models were obtained when SVM-C (optimizing with 'nu' operational parameter) without and with X-Block compression by PLS was used. Only four EVOO adulterated samples were misclassified.

Table 5. Values of the quality performance metrics from the different FTIR adulterated olive oils detection models.

Performance features Region 2 (1680-1800 cm-1)	kNN	PLS-DA	SIMCA	(nu)SVM-C			(c)SVM-C		
				None	PCA	PLS	None	PCA	PLS
	2iC	2iC	2iC	2iC	2iC	2iC	2iC	2iC	2iC
Sensibility (Recall)	0.83	0.70	0.70	1.00	1.00	1.00	1.00	1.00	0.87
Specificity	0.50	0.92	0.58	0.00	0.00	0.00	0.00	0.00	0.67
Positive predictive value (Precision)	0.76	0.94	0.76	0.66	0.66	0.66	0.66	0.66	0.83
Negative predictive value	0.60	0.61	0.50	-	-	-	-	-	0.73
Efficiency (Accuracy)	0.71	0.77	0.66	0.66	0.66	0.66	0.66	0.66	0.80
AUC (Correctly classified rate)	0.66	0.81	0.64	0.50	0.50	0.50	0.50	0.50	0.77
Matthews correlation coefficient	0.34	0.58	0.27	-	-	-	-	-	0.55
Kappa coefficient	-0.11	0.55	0.27	0.00	0.00	0.00	0.00	0.00	0.55

The hyphen "-" is signifying that the performance feature cannot be determined

Table 6. Values of the quality performance metrics from the different Raman adulterated olive oils detection models.

Performance features (650-950 cm-1)	kNN	PLS-DA	SIMCA	(nu)SVM-C			(c)SVM-C		
				None	PCA	PLS	None	PCA	PLS
	2iC	2iC	2iC	2iC	2iC	2iC	2iC	2iC	2iC
Sensibility (Recall)	0.83	0.92	0.46	1.00	1.00	1.00	0.88	0.71	0.83
Specificity	0.08	0.67	0.33	0.67	0.00	0.67	0.67	0.00	0.67
Positive predictive value (Precision)	0.65	0.85	0.58	0.86	0.67	0.86	0.84	0.59	0.83
Negative predictive value	0.20	0.80	0.24	1.00	-	1.00	0.73	0.00	0.67
Efficiency (Accuracy)	0.58	0.83	0.42	0.89	0.67	0.89	0.81	0.47	0.78
AUC (Correctly classified rate)	0.46	0.79	0.40	0.83	0.50	0.83	0.77	0.35	0.75
Matthews correlation coefficient	-0.11	0.61	-0.20	0.76	-	0.76	0.55	-0.35	0.50
Kappa coefficient	-0.10	0.61	-0.19	0.73	0.00	0.73	0.55	-0.33	0.50

The hyphen "-" is signifying that the performance feature cannot be determined

3.4. Olive oil quantification model

Quantitative analysis of blends of olive oil with other vegetable edible oils was performed building a specific PLS-R model from FTIR and Raman fingerprints on the regions previously selected (see section 3.1). In order to achieve more realistic conditions of the composition of olive oil, the proportion of olive oil in the blends of the

training and validation set is different, in contrast to some research work about quantification of olive oil using spectroscopic techniques in which the composition is similar in both set.

The reliability of the different models was established on the basis of: (i) the determination coefficient (R^2) and (ii) the errors of quantification (validation errors) were evaluated with the Root Mean Square Error of Validation (RMSEV), Mean Absolute Error of Validation (MAEV) and Median Absolute Error of Validation (MdAEV) (Hyndman & Koehler, 2006; ASTM E1655-05, 2012). The results obtained (g EVOO/100 g blend) in terms of R^2, RMSEV, MAEV and MdAEV were 0.86, 17.6, 14.6 and 16.0 respectively from FTIR and 0.93, 34.2, 27.8 and 29.6 respectively from Raman. Figure 5 shows the concentration values obtained from the PLS model vs. the actual concentration of any vegetable edible oil in olive oil samples using FTIR-ATR.

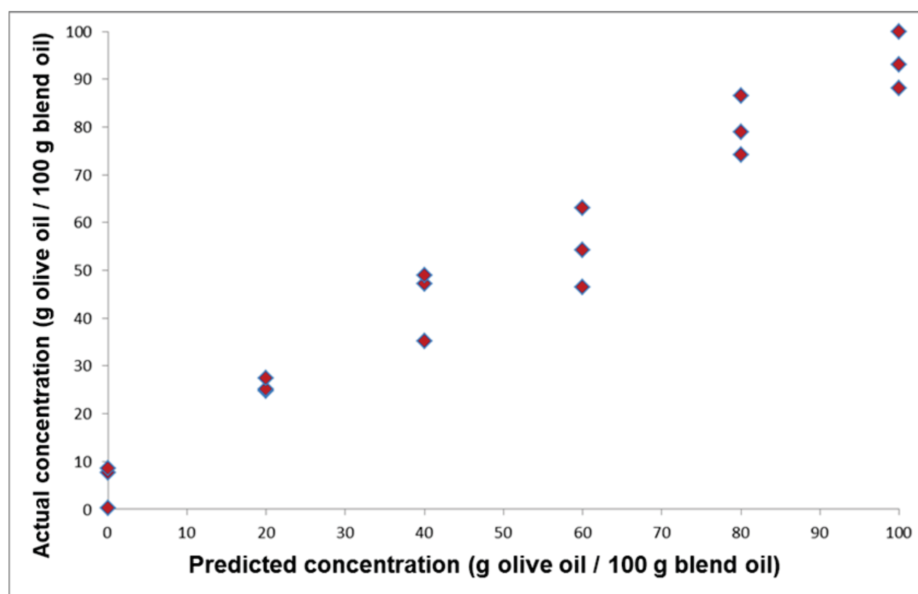


Figure 5. Concentration values for adulteration obtained from the PLS model vs. the actual concentration of olive oil using FTIR-ATR.

Although the R^2 obtained from FTIR is not sufficiently good, the validation errors (about 15-17%) are better than the validation errors obtained from Raman (about 28-34%).

4. Conclusion

Methyl-transesterified provides the information needed to authenticate of olive oil. The method developed could be named "global method" of detection, discrimination and quantification of olive oil in blends with other vegetable edible oils. Moreover, due to a transesterification step prior to spectroscopic analysis the problem of the low selectivity of these techniques has been resolved. Using FTIR and applying PLS-DA is performed without the need of any resource intensive chromatographic analysis. Discriminant analysis classified well the 100% olive oils samples and in addition, the proportion of olive oil in blends with other vegetable edible oils has been successively quantified using PLS-R.

Abbreviations and acronyms

1iC, one input-class classification

2iC, two input-class classification

ATR, attenuated total reflectance

AUC, area under the receiver operating curve

Bay-LS-SVM, Bayesian-least squares-support vector machine

BOM, bean with omega

CAM, camellia oil

CAN, canola oil

COG, corn germ oil

COR, corn oil

COT, cottonseed oil

EVOO, extra virgin olive oil

FLA, flaxseed oil

FT-IR, Fourier transform-infrared spectroscopy

FT-NIR. Fourier transform-near infrared spectroscopy

GAR, garlic oil

GSO, Grapeseed oil

HAZ, hazelnut oil

kNN, k-nearest neighbors

KS, Kennard-Stone

LDA, linear discriminant analysis

LS-SVM, least squares-support vector machine

LV, latent variables

MAE, Mean absolute error

MAEV, mean absolute error of validation

MdAEV, median absolute error of validation

MDVO, mixture of different vegetable edible oils (non-olive oil)

MLR, multiple linear regression

OCPLS, one class partial least squares classification

OO, olive oil

*p*2iC, *pseudo* two input-class classification

PC, principal component

PCA, principal component analysis

PCR, principal component regression

PLS-DA, partial least squares-discriminant analysis

PLS-R, partial least squares regression

PLS-R, partial least squares regression

PMR, partial robust M-regression

POO, pomace olive oil

R^2, determination coefficient

RBF, radial basis function

RBO, rice bran oil

RMSE, Root mean square error

RMSEV, root mean square error of validation

RPS, rapeseed oil

SAF, safflower oil

SES, sesame oil

SIMCA, soft independent modelling of class analogy

SNV, standard normal variate

SOY, soybean oil

SUN, sunflower oil

SVM-C, support vector machine classification

TBME, tert-butyl methyl ether

VOO, virgin olive oil

WGE, wheat germ oil

WO, walnut oil

References

Regulation (UE) Nº 29/2012 on marketing standards for olive oil (consolidated version 2012R0029), Official Journal of the European Union. (2016) 005.001, 1-15.

ASTM E1655-05(2012). (2012). Standard practices for infrared multivariate quantitative analysis. ASTM International, West Conshohocken (PA).

Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: linear models. PLS-DA. *Analytical Method, 5*, 3790–3798.

Biedermann, M., Grob, K., & Mariani, C. (1993). Transesterification and on-line LC-GC for determining the sum of free and esterified sterols in edible oils and fats. *European Journal of Lipid Science and Technology, 95*, 127–133.

Bro, R., & Smilde, A. (2014). Principal component analysis. *Analytical Method, 6*, 2812–2831.

Bylesjö, M., Rantalainen, M., Cloarec, O., Nicholson, J.K., Holmes, E., & Trygg, J. (2006). OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *Journal of Chemometrics, 20*, 341–351.

CODEX STAN 33-1981. Revision 2015. Standard for olive oil oils and olive pomace oils. Codex Alimentarius, FAO/WHO.

Commission Regulation (EEC) Nº 2568/1991 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 (2016), 031.001, 1-127.

Cuadros-Rodíguez, L., Pérez-Castaño, E., & Ruiz-Samblás, C. (2016). Quality performance metrics in multivariate classification methods for qualitative analysis. *Trends in Analytical Chemistry, 80*, 612-624.

De la Mata, P., Dominguez-Vidal, A., Bosque-Sendra, J.M., Ruiz-Medina, A., Cuadros-Rodríguez, L., & Ayora-Cañada, M. (2012). Olive oil assessment in edible oil blends by means of ATR-FTIR and chemometrics. *Food Control, 23,* 449–455.

Dong, W., Zhang, Y., Zhang, B., & Wang X. (2012). Quantitative analysis of

adulteration of extra virgin olive oil using Raman spectroscopy improved by Bayesian framework least squares support vector machine. *Analytical Method, 4*, 2772–2777.

El-Abassy, R.M., Donfack, P., & Materny A. (2009). Visible Raman spectroscopy for the discrimination of olive oils from different vegetable oils and the detection of adulteration. *Journal of Raman Spectroscopy, 40*, 1284–1289.

Gurdeniz, G., & Ozen B. (2009). Detection of adulteration of extra virgin olive oil by chemometric analysis of mid-infrared spectral data. *Analytical Method, 116*, 519–525.

Hyndman, R.J, & Koehler, A.B. (2006). Another look at measures of forecast accuracy. *International Journal of Forescasting*, *22,* 679–688.

Jiménez-Carvelo, A.M., González-Casado, A., Pérez-Castaño, E., & Cuadros-Rodríguez L. (2017). Fast-HPLC fingerprinting to discriminate olive oil from other edible vegetable oils by multivariate classification methods. *Journal of AOAC International, 100(2*), 1-6.

Jiménez-Carvelo, A.M., Pérez-Castaño, E., González-Casado, A., & Cuadros-Rodríguez, L. (2017). One input-class and two input-class classifications for differentiating olive oil from other edible vegetable oils by use of the normal-phase liquid chromatography fingerprint of the methyl-transesterified fraction. *Food Chemistry, 221*,1784–1791.

Kennard, R.W., & Stone, L.A. (1969). Computer aided design of experiments. *Technometrics, 11*, 137–148.

Lerma-García, M.J., Ramis-Ramos, G., Herrero-Martínez, J.M., & Simó-Alfonso, E.F. (2010). Authentication of extra virgin olive oils by Fourier-transform infrared spectroscopy. *Food Chemistry, 118*, 78–83.

Li, Y. & Watkins, B.A. (2001). Unit 1.2: Analysis of Fatty Acids in Food Lipids, in Whitaker, J. (Ed.), *Current Protocols in Food Analytical Chemistry*. John Wiley and Sons, Inc., New York, D1.2.1-D1.2.15.

López-Díez, E.C., Bianchi, G., & Goodacre, R. (2003). Rapid and quantitative assessment of the adulteration of virgin olive oils with hazelnut oils using Raman spectroscopy and chemometrics. *Journal of Agricultural and Food*

Chemistry, 51, 6145–6150.

Luts, J., Ojeda, F., De Plas, R.V., De Moor, B., Huffel, S.V., & Suykens, J.A.K. (2010). A tutorial on support vector machine-based methods for classifications problems in chemometrics. *Analytical Chimica Acta, 665*,129–145.

Marini, F. (2010). Classification methods in chemometrics. *Current Analytical Chemistry, 6,* 72-79.

Regulation (UE) Nº 1169/2011 on the provision of food information to consumers, amending Regulations (EC) No 1924/2006 and (EC) No 1925/2006 of the European Parliament and of the Council, and repealing Commission Directive 87/250/EEC, Council Directive 90/496/EEC, Commission Directive 1999/10/EC, Directive 2000/13/EC of the European Parliament and of the Council, Commission Directives 2002/67/EC and 2008/5/EC and Commission Regulation (EC) No 608/2004, Official Journal of the European Union (2011), L 304/18.

Rohman, A., Che, & Man, Y.B. (2011). Determination of extra virgin olive oil in quaternary mixture using FTIR spectroscopy and multivariate calibration. *Spectroscopy, 26*, 203–211.

Rohman, A., & Che Man, Y.B. (2012). Authentication of extra virgin olive oil from sesame oil using FTIR spectroscopy and gas chromatography. *International Journal of Food Properties*, 15, 1309–1318.

Steinbach, M., & Tan, P.N. (2009). kNN: k-Nearst neighbors. In: Wu X., Kumar V, (Eds), *The top ten algorithms in data mining* (pp. 151–161). Boca Raton (FL): Chapman & Hall//CRC.

Sun, X., Lin, W., Li, X., Shen, Q., & Luo, H. (2015). Detection and quantification of extra virgin olive oil adulteration with edible oils by FT-IR spectroscopy and chemometrics. *Analytical Methods*, *7*, 3939–3945.

SVM Function Settings, Eigenvector Documentation Wiki [accessed 13 June 2017], http://wiki.eigenvector.com/index.php?title=Svmda.

Tay, A., Singh, R.K., Krishnan, S.S., & Gore, J.P. (2002). Authentication of olive oil adulterated with vegetable oils using Fourier transform infrared

spectroscopy. *LWT – Food Science and Technology, 35,* 99-103.

Tiryaki, G.Y., & Ayvaz, H. (2017). Quantification of soybean oil adulteration in extra virgin olive oil using portable Raman spectroscopy. *Journal of Food Measurement and Characterization, 11*, 523-529.

Xu, L., Yan, S.M., Cai, C.B., & Yu, X.P. (2013). One-class partial least squares (OCPLS) classifier*. Chemometrics and Iintelligent Laboratory System, 126*, 1–5.

Zou, M.Q., Zhang, X.F., Qi, X.H., Ma, H.L., Dong, Y., Liu, C.W., Guo, X., & Wang H. (2009). Rapid authentication of olive oil adulteration by Raman spectrometry. *Journal of Agricultural and Food Chemistry*, *57*, 6001–6006.

III.8. Discusión

Los resultados obtenidos permiten concluir que la aplicación de las técnicas de espectrometría vibracional ATR-FTIR y Raman sobre la fracción metil-transesterificada del aceite de oliva dan óptimos resultados para autentificar el aceite de oliva cuando está mezclado con otros aceites vegetales.

Aunque no se haya "cumplido" con la premisa o ventaja principal de las técnicas espectroscópicas, la cual es poder medir la muestra sin realizar un pre-tratamiento previo de la muestra, se ha demostrado que realizando una reacción de transesterificación metílica sobre el aceite ha compensado la falta de selectividad que presentan éstas técnicas frente a las técnicas cromatográficas. Asimismo presenta una principal ventaja al sector olivarero, ya que es posible detectar adulteraciones de aceite de oliva con más de un aceite vegetal aplicando un procedimiento relativamente sencillo y con una técnica analítica ya totalmente implantada en los laboratorios de control de la calidad.

Del mismo modo, gracias a llevar a cabo una etapa de pre-tratamiento de muestra previo al análisis espectroscópico, se ha podido desarrollar un **método global** de detección y cuantificación de aceite de oliva en mezclas con otros aceites vegetales empleando diferentes tipos de aceites en una misma mezcla. A diferencia de la mayoría de los estudios encontrados en literatura en los cuales aunque obtengan buenas puntuaciones en los modelos de clasificación y cuantificación presentan un principal inconveniente y es que **no se encuentran modelos** que impliquen más de tres aceites diferentes en una misma mezcla, lo cual presenta una dificultad debido a qué cuando una muestra de aceite llegue a un laboratorio de análisis a priori no se conoce con qué tipo de aceite ha podido ser adulterado (sí es que se da el caso de dicha adulteración) con lo cual la fiabilidad del modelo no está asegurada.

COMUNICACIONES A CONGRESOS

- A.M. Jiménez Carvelo, A. González Casado, L. Cuadros Rodríguez. **Aplicación de FTIR-ATR y RAMAN para identificar y cuantificar aceite de oliva en mezclas de aceites vegetales.** XVI Reunión del grupo regional andaluz de la sociedad española de química analítica (GRASEQA). **Oral Flash.**

Capítulo IV

“Cada solución da pie a una nueva pregunta”

David Hume

CAPITULO IV

'Caracterización química de la fracción metil-transesterificada'

IV.1. Presentación

Este capítulo recoge el análisis de la fracción metil-transesterificada de aceites de oliva virgen extra monovarietales mediante cromatografía de gases acoplada a espectrometría de masas de alta resolución (GC-(APCI/QTOF)MS), con el objetivo de identificar los compuestos presentes en la misma.

Aunque la idea original era la de aplicar cromatografía de líquidos acoplada con espectrometría de masas de alta resolución, la naturaleza química del sistema material a estudiar aconsejó el uso de la cromatografía de gases.

La parte experimental de este trabajo se llevó a cabo durante una estancia predoctoral de investigación en el Instituto Universitario de Plaguicidas y Aguas (IUPA) de la Universidad "Jaume I" de Castellón, centro de investigación de excelencia en el campo de la espectrometría de masas, bajo la supervisión del Dr. Joaquín Beltrán Arandes y en el seno del grupo de investigación "Química Analítica en Salud Pública y Medio Ambiente".

IV.2. Introducción

La caracterización de los compuestos o familia de compuestos presentes en una matriz puede llevarse a cabo aplicando dos enfoques o análisis, conocidos como análisis no dirigido ('non-targeted' o 'untargeted') y análisis dirigido ('targeted').

El análisis no dirigido se aplica para detectar e identificar compuestos desconocidos en la matriz de estudio. Este enfoque suele ir acompañado del empleo de herramientas estadísticas y/o de reconocimiento de pautas con el objeto de poder determinar marcadores específicos de cada muestra analizada. Este análisis se caracteriza por llevarse a cabo sin disponer necesariamente de patrones o materiales de referencia, contar con un número elevado de compuestos y presentar un elevado grado de dificultad para elucidar la estructura molecular de éstos. Los equipos instrumentales más empleados para llevar a cabo este tipo análisis son los espectrómetros de masas de alta resolución híbridos de cuadrupolo con analizador de tiempo de vuelo (QTOF) o los espectrómetros de masas de alta resolución con analizador tipo Orbitrap™. El modo de trabajo empleado en ambos casos es análisis mediante "barrido completo" ('full-scan') de fragmentos moleculares de masas, es decir se registra el espectro de masas completo a cada tiempo de retención. Este tipo de técnicas permiten una alta resolución, sensibilidad y selectividad y están siendo ampliamente empleadas en estudios de autentificación de alimentos [**1**,**2**,**3**]. Este campo está adquiriendo mayor relevancia en los últimos años y mayor atención por la comunidad científica así como por las autoridades para llevar a cabo el control oficial de productos alimentarios [**4**,**5**].

De acuerdo con esto, se ha aplicado este tipo de enfoque en el análisis de aceite de oliva para clasificar el aceite de oliva de origen griego en función del tipo de aceituna [**6**], para discriminar el aceite de oliva en función de su origen geográfico [**7**,**8**,**9**] o para

---

[1] Cubero-Leon, E. Peñalver, R., Maquet, A. (2014). Review on metabolomics for food authentication. Food Research International, *60*, 95-107.

[2] Riedl, J., Esslinger, S., Fauhl-Hassek, C. (2015). Review of validation and reporting of non-targeted fingerprinting approaches for food authentification. Analytica Chimica Acta, *885*, 17-32.

[3] Danezis, G.P., Tsagkaris, A.R., Brusic, V., Georgiou, C.A. (2016). Food authentication: state of the art and prospect. Current opinion in Food Science, *10*, 22-31.

[4] Esslinger, S., Riedl, C., Fauhl-Hassek, C. (2014). Potential and limitations of non-targeted fingerprinting for authentication of food in official control. Food Research International, *60*, 189-204.

[5] Kunzelmann, M., Winter, M., Aberg, M., Hellenäs, K.E., Rosén, J. (2018). Non-targeted analysis unexpected food contaminants using LC-HRMS. Analytical and Bioanalytical Chemistry, *410*, 5593–5602.

[6] Kalogiouri, N.P., Aalizadeh, R., Thomaidis, N.S. (2018). Application of an advanced and wide scope non-target screening workflow with LC-ESI-QTOF-MS and chemometrics for the classification of the Greek olive oil varieties. Food Chemistry, *256*, 53-61.

[7] Melucci, D., Bendini, A., Tesini, F., Barbieri, S., Zappi, A., Vichi, S., Conte, L., Gallina Toschi, E. (2016). Rapid direct analysis to discriminate geographic origin of extra virgin olive oils by flash gas chromatography electronic nose and chemometrics. Food Chemistry, *204*, 263-273.

determinar compuestos orgánicos volátiles responsables de ciertos atributos sensorial del aceite de oliva [**10**].

El análisis dirigido, a diferencia del anteriormente citado, se aplica además para cuantificar un grupo determinado de compuestos definidos previamente. En este tipo de análisis ya sí se trabaja con patrones o materiales de referencia, esto implica que se cuenta con un conocimiento *a priori* de los compuestos presentes en la matriz de estudio. Como consecuencia, se reduce el error de identificación de un determinado compuesto, por lo que este tipo de análisis suele ser más sensible que el análisis no dirigido y los métodos de pre-tratamiento de muestran suelen ser más específicos para aislar los compuestos de interés. Sin embargo, es en este pre-tratamiento donde se encuentra la principal desventaja del análisis dirigido, ya que un método determinado para identificar un compuesto o familia de compuestos puede que no sea igual de efectivo para otro grupo de compuestos.

Este enfoque ha sido el tradicionalmente usado para el control analítico de la seguridad alimentaria y la autentificación química de alimentos, como por ejemplo para determinar residuos de antibióticos en la miel [**11**], identificar y cuantificar adulterantes en mozzarellas [**12**], discriminar aceites de oliva en función de la variedad de aceituna mediante un análisis de los compuestos fenólicos [**13**], o para cuantificar pesticidas en diferentes matrices de frutas y verduras [**14**].

---

[8] Gerhardt, N., Birkenmeir, M., Sanders, D., Rohn, S., Weller, P. (2017). Resolution-optimized headspace gas chromatography-ion mobility spectrometry (HS-GC-IMS) for non-targeted olive oil profiling. Analytical and Bioanalytical Chemistry, *409*, 3933-3942.

[9] Gil Solsona, R., Raro, M., Sales, C., Lacalle, L., Díaz, R., Ibañez, M., Beltran, J., Sancho, S.V., Hernández, F.J. (2016). Metabolomic approach for extra virgin olive oil origin discrimination making use of ultra-high performance liquid chromatography-quadrupole time-of-flight mass spectrometry. Food Control, *70,* 350-359.

[10] Vichi, S., Cortés-Francisco, N., Romero, A., Caixach, J. (2014). Determination of volatile thiols in virgin olive oil by derivatisation and LC-HRMS, and relation with sensory attributte. Food Chemistry, *149*, 313-318.

[11] Tölgyesi, A., Barta, E., Sohn, M., Sharma, V.K. (2018). Determination of antimicrobial residues in honey by liquid chromatography tandem mass spectrometry. Food Analytical Methods, *11,* 2043-2055.

[12] Dal Bosco, C., Panero, S., Navarra, M.A., Tomai, P., Curini, R., Gentili, A. (2018).Screening and assessment of low-molecular-weight biomarkers of milk from cow and water buffalo: an alternative approach for the rapid identification of adulterated water buffalo mozzarellas. Journal of Agricultural and Food Chemistry, *66,* 5410-5417.

[13] Monasterio, R.P., Olmo García, L., Bajoub, A., Fernández Gutierrez, A., Carrasco Pancorbo, A. (2017). Phenolic compounds profiling of virgin olive oils from different varieties cultivated in Mendoza, Argentina, by using liquid chromatography-mass spectrometry. Journal of Agricultural and Food Chemistry, *65,* 8184-8195.

[14] Cervera, M.I., Portolés, T., López, F.J., Beltrán, J., Hernández, F. (2014). Screening and quantification of pesticide residue in fruits and vegetables making use of gas chromatography-quadrupole time-of-flight mass spectrometry with atmospheric pressure chemical ionization. Analytical and Bioanalytical Chemistry, *406,* 6843-6855.

Ambos enfoques trabajan de forma complementaria y en una primera etapa es común aplicar un análisis no dirigido con objeto de llevar a cabo un screening de los posibles compuestos presentes en el objeto de estudio para posteriormente llevar a cabo un análisis dirigido para bien cuantificar la cantidad de los mismos o para usar los marcadores seleccionados en la etapa no dirigida y llevar a cabo una autentificación con compuestos de estructura conocida. En este sentido se han aplicado ambas metodologías para determinar pesticidas en frutas y verduras [**15**] o para clasificar aceites de oliva como aceites de oliva virgen extra o aceites de oliva con defectos [**16**].

Las técnicas instrumentales empleadas para llevar a cabo este tipo de análisis han sido la cromatografía de líquidos de (ultra) alta eficacia y la cromatografía de gases, acopladas ambas a espectrometría de masas. Ésta última se emplea para la detección, identificación y elucidación de la estructura de compuestos basándose en la fragmentación de moléculas en iones derivados; una vez obtenidos dichos iones, éstos se separan de acuerdo a su relación masa- carga (m/z) y posteriormente generan sendas señales que son detectadas y cuantificadas en intensidad. A continuación se muestra el esquema general de los elementos que componen un espectrómetro de masas: fuente de ionización, analizador de iones y detector o sistema de medida.

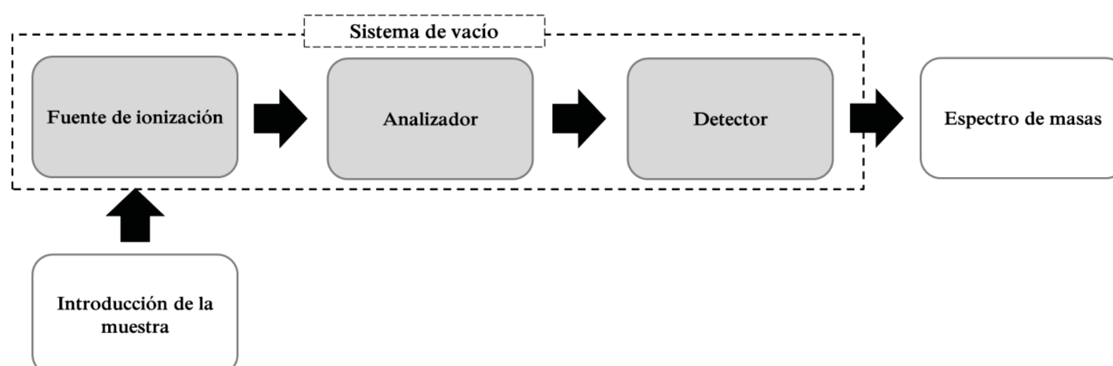


Figura 1. Diagrama de flujo simplificado de un espectrómetro de masas convencional.

En la fuente de ionización es donde tiene lugar la transformación (fragmentación) de las moléculas en especies moleculares iónicas en fase gaseosa, con la correspondiente pérdida de electrones o hidrogenoiones. Existen diferentes tipos de fuentes de ionización, entre las más comúnmente usadas se encuentran la ionización por impacto de electrones (EI) , ionización por electroespray (ESI), ionización química a presión atmosférica (APCI), ionización por desorción láser asistida por una matriz (MALDI) e ionización por análisis directo en tiempo real (DART). La elección de un sistema de

[15] Cervera, M.I., Portolés, T., Pitarch, E., Beltrán, J., Hernández, F. (2012). Application of gas chromatography time-of-flight mass spectrometry for target and non-target analysis of pesticide residues in fruits and vegetables. Journal of Chromatography A, *1244*, 168-177.

[16] Kalogiouri, N.P., Alygizakis, N.A., Aalizadeh, R., Thomaidis, N.S. (2016). Olive oil authenticity studies by target and nontarget LC-QTOF-MS combined with advanced chemometric techniques. Analytical and BioanalyticalChemistry, *408,* 7955-7970.

ionización u otro dependerá del tipo de muestra a analizar y de la clase de información que se desea obtener [**17**].

En el analizador de iones es dónde se logra la separación física de los diferentes fragmentos iónicos en función de su relación masa/carga (m/z), y posiblemente sea el componente más importante o de mayor relevancia de un espectrómetro de masas, ya que la resolución, la precisión en la medida y el rango de masas que se pueden obtener dependen en gran medida de su diseño. La resolución de un espectrómetro de masas es la capacidad de éste para discernir entre dos iones de masa similar. Por todo ello, los analizadores de iones se pueden clasificar atendiendo al grado de resolución, teniendo así espectrómetros de masas de baja resolución y de alta resolución. Los espectrómetros de baja resolución más ampliamente utilizados son el simple cuadrupolo (Q) y la trampa de iones (IT). En el caso de la espectrometría de masas de alta resolución lo más habituales son el analizador de triple cuadrupolo (QQQ), cuadrupolo con analizador de tiempo de vuelo (QTOF) y analizador Orbitrap™, y analizador de sector magnético, aunque éste último está bastante menos implantado.

Por último la información sobre el flujo de iones es detectada y presentada en forma de un espectro de masas utilizando detectores del tipo de fotomultiplicadores o copa de Faraday.

Los modos de adquisición del espectro de masas son [**18**]: (i) modo de barrido ('scan', o más comúnmente conocido como 'full-scan') que es el modo utilizado cuando se aplica el análisis no dirigido y consiste en realizar un barrido completo entre un rango de m/z seleccionado previamente con el objetivo de tener una información global del objeto de estudio; (ii) modo mediante selección de iones ('SIM', 'selected ion monitoring') que consiste en seleccionar unos iones concretos en un intervalo de tiempo concreto, este modo es el que se emplea cuando se quieren cuantificar compuestos conocidos en la muestra; y (iii) espectros de masas en tándem, o 'MS/MS', que implica dos etapas, primero una etapa de aislamiento de un ion fragmento previamente seleccionado, conocido como ion precursor, y una segunda donde se produce la una nueva fragmentación de dicho ion en iones más pequeños, que son los que llegan al detector.

A continuación se muestra el estudio experimental llevado a cabo para el análisis de la fracción metil-transesterificada de aceites de oliva virgen extra monovarietales, llevado a cabo con un equipo de cromatografía de gases acoplado a un espectro de masas con analizador QTOF con objeto de identificar los compuestos presentes en dicha fracción.

---

[17] Martín Gómez, M.C., Ballesteros González, M. (2010). Espectrometría de masas y análisis de biomarcadores. Monografía XXX: Biomarcadores: analítica, diagnóstico y terapéutica. Real Academia de Farmacia, http://dx.doi.org/ES/monoranf.v0i0.1066 .

[18] de Hoffman, E., Stroobant V. (2007). Mass Spectrometry - Principles and Applications. 3rd ed. John Wiley & Sons, Chichester, UK.

IV.3. Instrumentación

Para llevar a cabo los análisis se utilizó un cromatógrafo de gases Agilent 7890A acoplado a un espectrómetro de masas con ionización química a presión atmosférica (APCI) como sistema de ionización y un analizador hibrido de simple cuadrupolo con tiempo de vuelo como analizador de masas (GC-(APCI/QTOF)MS).

La separación cromatográfica se llevó a cabo utilizando una columna de 5% difenil-95% dimetilpolixilosano (DB-5MS) de 0.25 µm de tamaño de partícula, 30 m de longitud y 0.250 mm de diámetro interno. Como gas portador se utilizó helio.

IV.4. Muestras

Todas las muestras analizadas fueron aceites de oliva virgen extra monovarietales, la mayoría adquiridos en comercios en diferentes puntos de España y otros donados por laboratorios de análisis de aceites o empresas productoras o envasadoras de aceite de oliva virgen extra. En la tabla 1 se indican las diferentes variedades, así como el número de muestras utilizadas en este estudio

Como se ha comentado en los anteriores capítulos, todas las muestras de aceite de oliva fueron sometidas a una reacción de transesterificación metílica antes de su análisis cromatográfico.

Tabla 1. Muestras de aceite de oliva analizadas.

Categoría	*Variedad de aceituna*	*Nº de muestras*
Virgen extra	Arbequina	5
	Picual	8
	Hojiblanca	4
	Cornicabra	5
	Royal	3
	Frantoio	3
	Koroneiki	3
	Farga	1
	Shikitia	1
	Tosca	1
	Lucio	2
	Loaime	3
	Arbosana	2
	Lechin	1
	Vidueña	1
	Manzanilla	3

Tabla 1. *Continuación.*

Categoría	*Variedad de aceituna*	*Nº de muestras*
	Ocal	1
	Oliana	1
	Negrete	1
	Serrana	1
	Verdial	1
	Alfafarenca	1
	Blanqueta	1
	Picudo	4
	No declarada	2
	TOTAL DE ACEITES	59

Además de las muestras mostradas en la tabla 1 también se analizaron patrones de compuestos comúnmente presentes en el aceite de oliva y que a priori cabría esperar encontrar (véase tabla 2). Se analizaron todos los patrones en hexano a una concentración de 1 mg/L.

Tabla 2. Patrones analizados.

Familia	*Compuesto*	*Formula molecular*
Ácido graso	Ácido palmítico	$C_{16}H_{32}O_2$
	Ácido esteárico	$C_{18}H_{36}O_2$
	Ácido oléico	$C_{18}H_{34}O_2$
Éster metílico de ácido graso	Miristato de metilo	$C_{15}H_{32}O_2$
	Palmitoleato de metilo	$C_{17}H_{32}O_2$
	Palmitato de metilo	$C_{17}H_{34}O_2$
	Estereato de metilo	$C_{19}H_{38}O_2$
	Linolenato de metilo	$C_{19}H_{32}O_2$
	Linoleato de metilo	$C_{19}H_{34}O_2$
	Oleato de metilo	$C_{19}H_{36}O_2$
	Araquidato de metilo	$C_{21}H_{42}O_2$
4-Desmetilesteroles	β-sitosterol	$C_{29}H_{50}O$
	Estigmasterol	$C_{29}H_{48}O$
Triglicéridos	Trioleina*	$C_{57}H_{104}O_6$

*El patrón de trioleína no fue analizado puro sino que se analizó la trioleina sometida a la misma reacción transesterificación metílica que los aceites de oliva.

IV.5. Condiciones experimentales

Las condiciones del método analítico utilizado para la obtención de los cromatogramas fueron las que se describen a continuación.

Condiciones del cromatógrafo de gases:

- Gas portador: Helio a 1.2 mL/min
- Modo de inyección: pulsos sin división ('pulsed splitless') (50 psi por pulso)
- Volumen de inyección: 2 µL
- Rampa de temperatura:

Temp. inicial (ºC)	Temp. final (ºC)	Rampa (ºC/min)	Tiempo (min)
60	60	-	1.0
60	180	30	4.0
180	320	10	14.0
320	320	-	6.0
		Tiempo total de análisis	25.0

Condiciones del espectrómetro de masas:

- Sistema de ionización: ionización química a presión atmosférica (APCI), 5 eV
- Temperatura de la línea de transferencia: 300 ºC
- Temperatura de la fuente de ionización: 150 ºC
- Modo de adquisición: barrido ('full scan')
- Rango de masas: 50-650 m/z

IV.6. Resultados y discusión

➡ Análisis de patrones

En primer lugar se llevó a cabo el análisis de la trioleína transesterificada, ya que es el triglicérido mayoritario en los aceites de oliva, con el objetivo de comprobar que productos de reacción se obtenían al llevar a cabo la transesterificación metílica. En la figura 2 se muestra el cromatograma obtenido dónde pueden verse los picos característicos de los productos de reacción, así como las estructuras químicas de cada uno de los compuestos identificados.

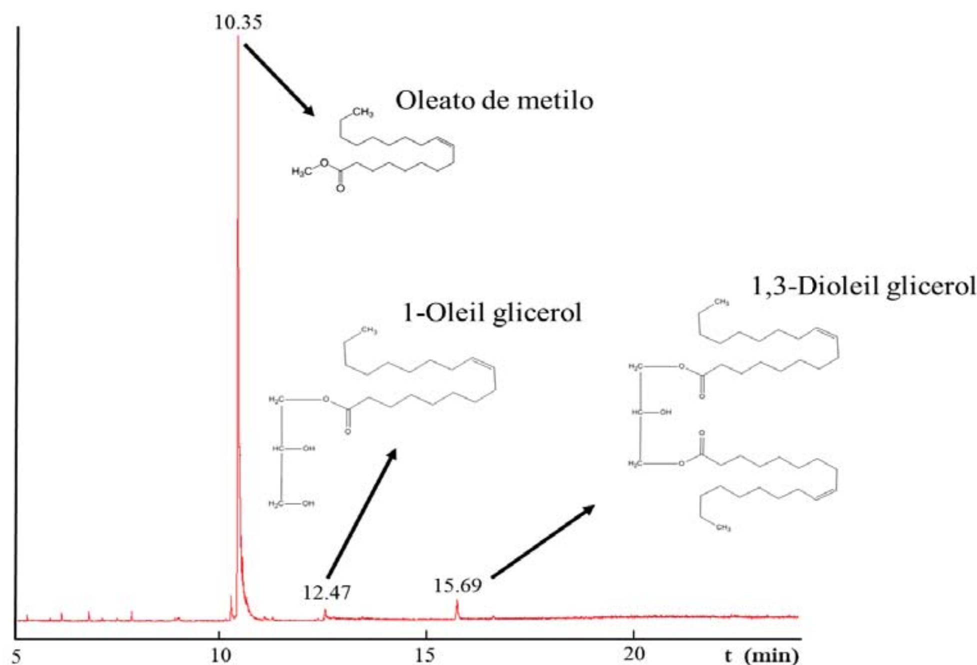


Figura 2. Señal cromatográfica (TIC) de la trioleína transesterificada.

El pico obtenido a 10.35 min corresponde al oleato de metilo que se encuentra en mayor proporción y cuya identificación se corroboró cromatografiando un patrón de oleato de metilo, y comparando que ambos presentaban el mismo tiempo de retención y espectros de masas similares (véase figura 3).

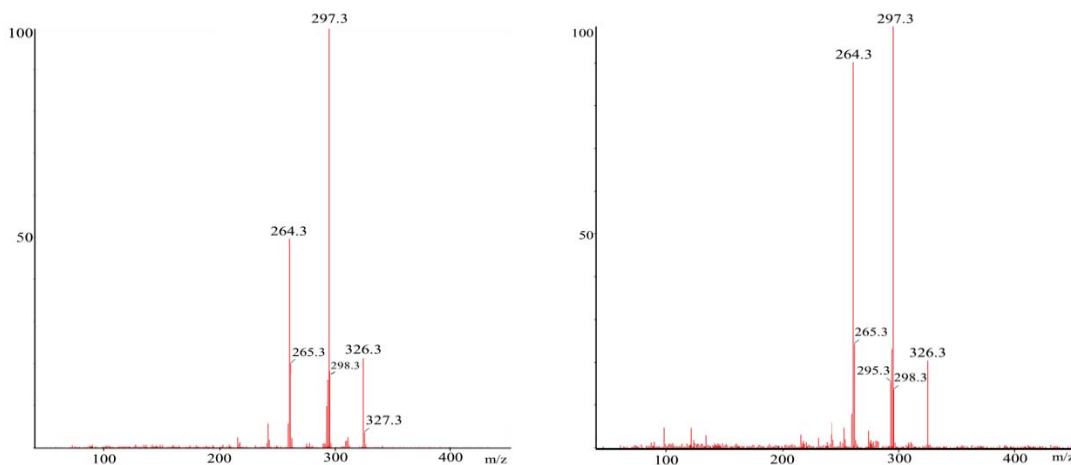


Figura 3. Espectros de masas correspondientes al oleato de metilo. A la izquierda el oleato de metilo producto de la transesterificación de la trioleína y a la derecha el patrón puro de oleato de metilo.

El pico obtenido a 12.47 min corresponde al compuesto 1-oleil glicerol, cuya identidad fue elucidada por nosotros analizando el espectro de masas correspondiente. En dicho espectro no se observó el ión molecular (356.2916 m/z) del compuesto 1-oleil glicerol, ni tampoco el $[M-17]^+$ (339.2916 m/z) resultante de la pérdida de una molécula de agua y ganancia de un átomo de hidrógeno. Pero en cambio sí se observaron lo fragmentos

282 m/z, 283 m/z y 264 m/z característicos del ácido oleico, por lo que se concluyó que el compuesto 1-oleil glicerol se fragmenta liberando el residuo de ácido oleico de su estructura. Además esto fue corroborado cuando se analizó el patrón de ácido oleico puro (véase figura 4).

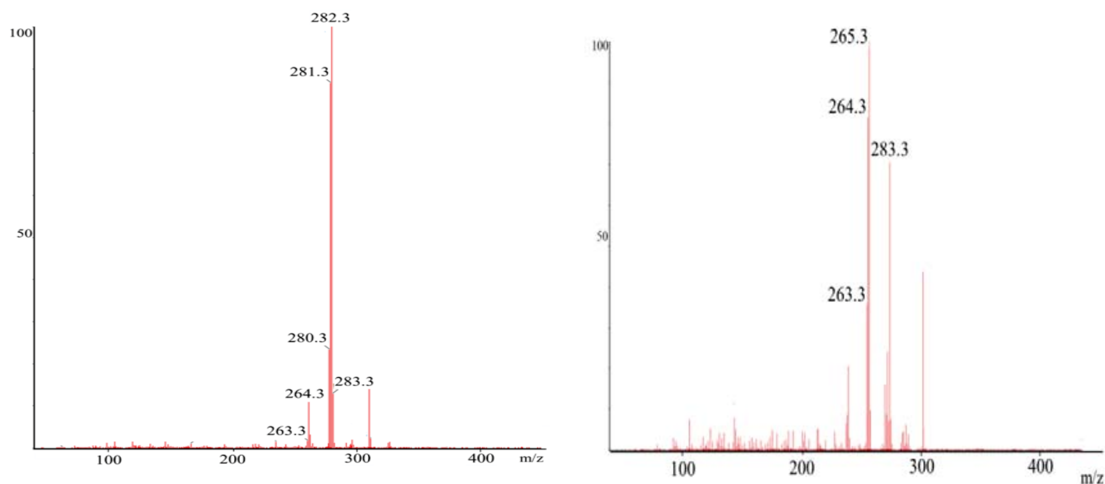


Figura 4. A la izquierda el espectro de masas correspondiente al 1-oleil glicerol, y a la derecha el espectro de masas correspondiente al patrón de ácido oleico.

Por último el pico correspondiente al minuto 15.69 se le asignó el diglicérido 1,3-dioleil glicerol, cuya identidad fue también elucidada por nosotros analizando el espectro de masas correspondiente. Al igual que en el caso anterior, no se observó el ión molecular (620.5361 m/z), ni el $[M-17]^+$ (603.5361 m/z). Los fragmentos más abundantes resultante de la ionización química de este compuesto fueron 336 m/z, 337 m/z, 338 m/z y 339 m/z, patrón isotópico resultante de la pérdida de una molécula de ácido oleico del 1,3-dioleil glicerol, además de observarse en menor proporción los fragmentos 282 m/z y 264 m/z característicos como se ha comentado anteriormente del ácido oleico (véase figura 5).

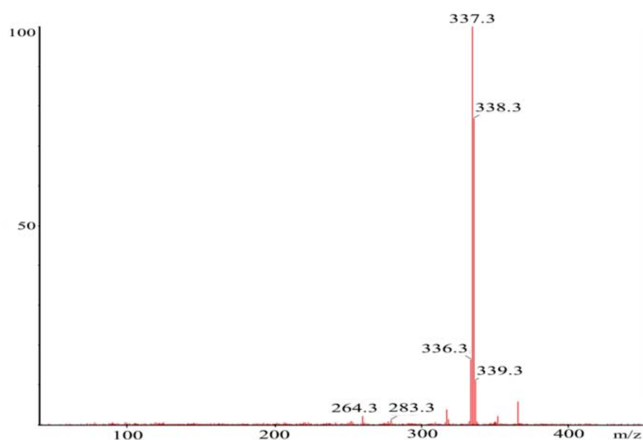


Figura 5. Espectro de masas correspondientes al 1,3-dioleil glicerol.

De esta forma, se seleccionaron los siguientes fragmentos para la trioleína transesterificada:

Tabla 3. Tiempo de retención (tr) y señales m/z de los compuestos resultantes al llevar a cabo la reacción de transesterificación metílica de la trioleína.

Compuesto		*m/z*	*tr (min)*
trioleina transesterificada	oleato de metilo	297.29, 264.25	10.35
	1-oleil glicerol	282.28, 264.25	12.47
	1,3-dioleil glicerol	337.34, 283.27, 264.25	15.69

De igual forma se llevó a cabo el análisis del resto de patrones recogidos en la tabla 2, y se seleccionaron los fragmentos característicos de los mismos (véase tabla 4).

Tabla 4. Tiempo de retención (tr) y señales m/z de los patrones.

Compuesto	*m/z*	*tr (min)*
Miristato de metilo	243.25	7.40
Palmitoleato de metilo	269.26	8.77
Palmitato de metilo	271.28	8.91
Ácido palmítico	257.27	9.42
Linoleato de metilo	294.27	10.31
Oleato de metilo	297.30, 264.26	10.36
Linolenato de metilo	292.26	10.36
Estereato de metilo	298.31	10.56
Ácido oléico	283.28, 265.28	10.85
Ácido esteárico	285.30	11.05
Araquidato de metilo	326.34	12.27
β-sitosterol	414.38, 397.38	19.57
Estigmasterol	412.41, 395.41	19.72

➡ Análisis de muestras de aceite de oliva virgen extra

Seguidamente se realizó la asignación de compuestos para los picos cromatográficos de las muestras analizando los espectros de masas correspondientes. Todas las muestras presentaban tres picos comunes a los tiempos de retención de 11.18, 12.81 y 16.14 min. En la figura 6 se representan tres aceites de oliva virgen extra dónde puede verse los tres picos citados.

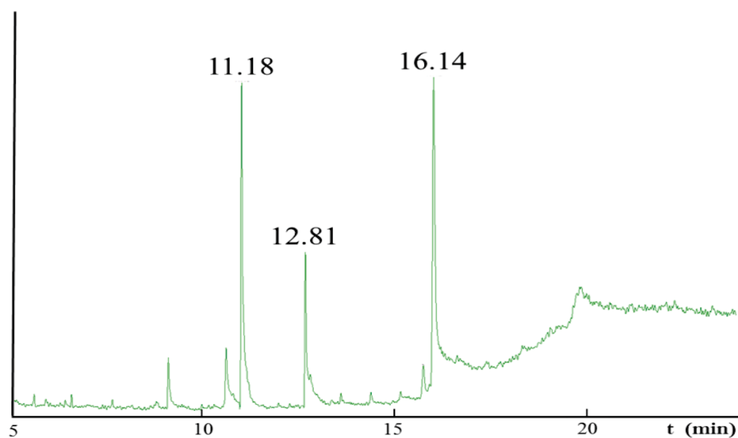


Figura 6a. Señal cromatográfica (TIC) de una muestra de aceite de oliva virgen extra de variedad **arbequina**.

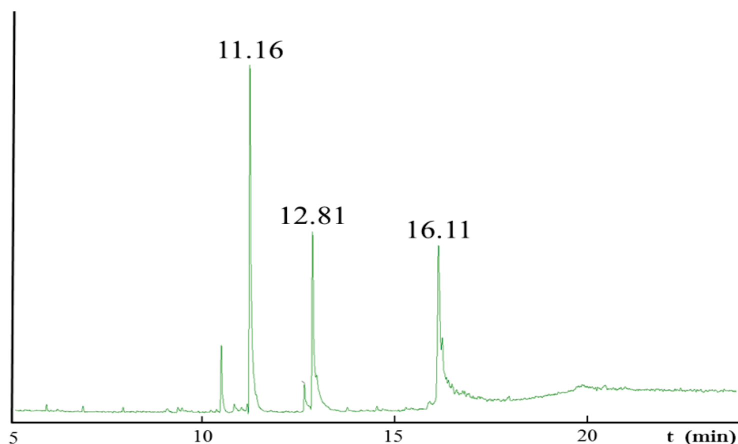


Figura 6b. Señal cromatográfica (TIC) de una muestra de aceite de oliva virgen extra de variedad **picual**.

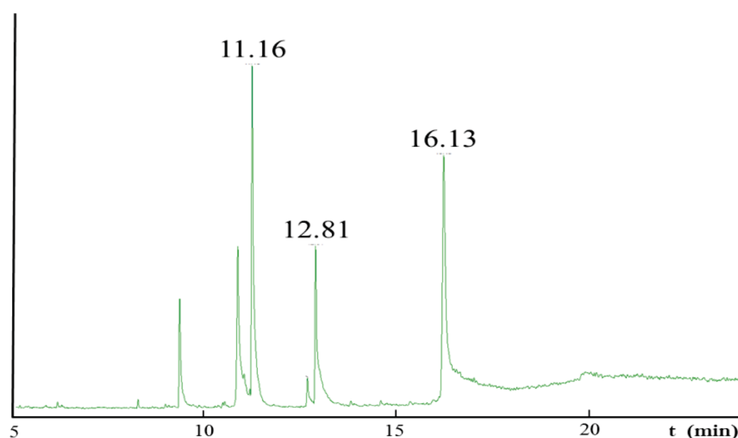


Figura 6c. Señal cromatográfica (TIC) de una muestra de aceite de oliva virgen extra de variedad **cornicabra**.

El pico cromatográfico a minuto 11.18 fue asignado al monoglicérido que presenta en su composición un residuo de ácido palmítico y es denominado 1-palmitoil glicerol. Se concluyó que era este compuesto ya que, aunque el ión molecular (330.276 m/z) no se observó, sí se localizó el ion $[M-17]^+$ (312.276 m/z) correspondiente a la pérdida de una molécula de agua y ganancia de un átomo de hidrógeno. Además de los fragmentos 258 m/z, 257 m/z y 256 m/z que representa el patrón isotópico del ácido palmítico, que fue igualmente comprobado con el espectro de masas del patrón de ácido palmítico. Igualmente el tiempo de retención al que aparece el pico también nos lleva a la misma conclusión ya que aparece antes que el pico del 1-oleil glicerol (comentado anteriormente) cuya masa molecular y estructura molecular es mayor que el del

1-palmitoil glicerol, y tal y como sucede en cromatografía de gases los compuestos son retenidos en función del número de carbonos y de las insaturaciones de los compuestos.

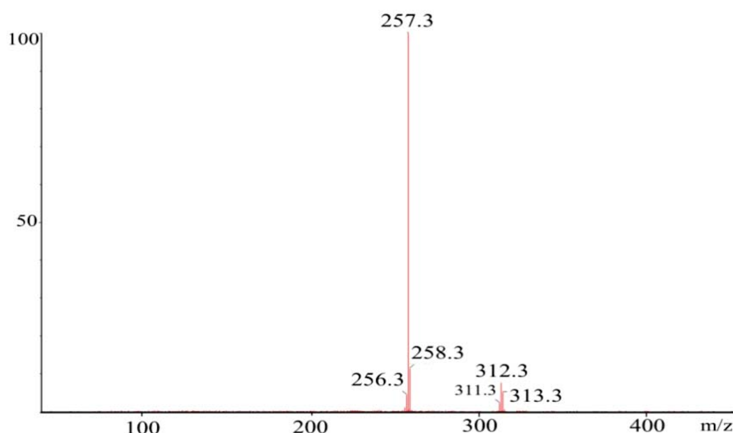


Figura 7. Espectro de masas correspondiente al **1-palmitoil glicerol** de un aceite de oliva virgen extra de variedad **picual**.

El siguiente pico cromatográfico significativo aparece en el minuto 12.81, y fue asignado al 1-estearoil glicerol, otro monoglicerido que en este caso presenta un residuo de ácido esteárico en su composición. De igual forma que en el caso anterior tampoco se observó el ión molecular (358.3072 m/z), pero sí el fragmento $[M-17]^+$ (341.3072 m/z) correspondiente a la pérdida de una molécula de agua y ganancia de un átomo de hidrógeno. Del mismo modo que en los picos elucidados anteriormente, se observaron los fragmentos 284 m/z, 285 m/z y 286 m/z característicos del ácido esteárico y asimismo se verificó con el espectro de masas del patrón de ácido esteárico.

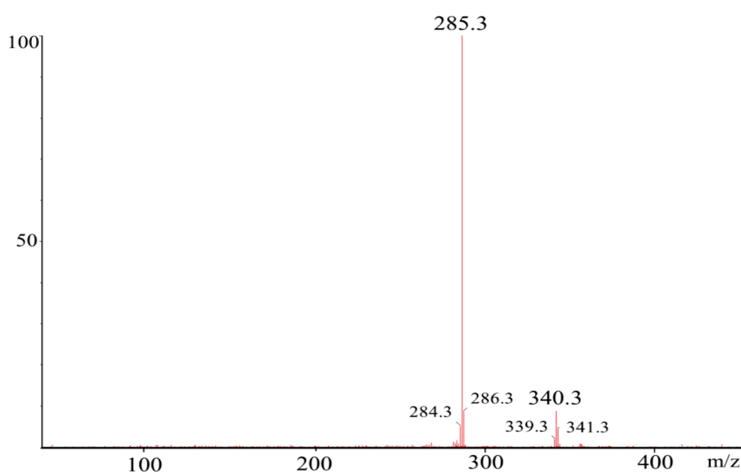


Figura 8. Espectro de masas correspondiente al **1-estearil glicerol** de un aceite de oliva virgen extra de variedad **picual**.

El último pico correspondiente al minuto 16.14 no se consiguió elucidar la estructura del compuesto al que era debido. Analizando el espectro de masas se observó cómo fragmentos mayoritarios el ión 410 m/z y el 426 m/z.

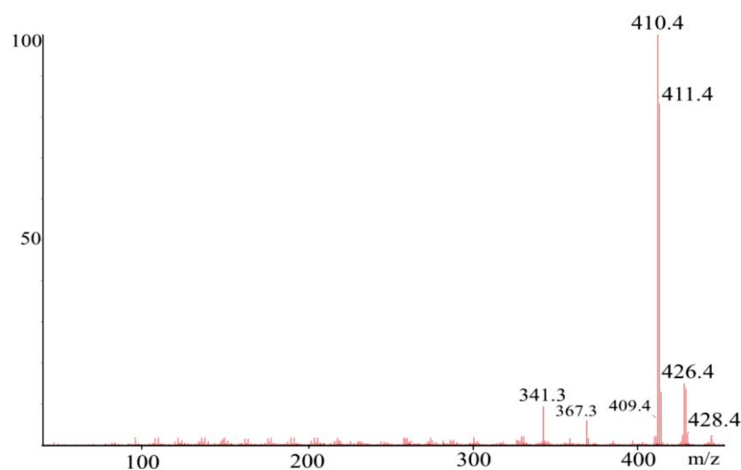


Figura 9. Espectro de masas correspondiente al pico a 16.11 min. de un aceite de oliva virgen extra de variedad picual.

Por el tiempo de retención se pensó que podría ser algún diglicérido, tal y como sucedía con la trioleína transesterificada cuyo pico a 15.67 min correspondiente con el 1-3-dioleil glicerol. Sin embargo, analizando los triglicéridos más abundantes en el aceite de oliva y considerando los correspondientes diglicéridos, no se concluyó que fuese alguno de los más comunes. Esto nos llevó a pensar las hipótesis de que podría ser algún tipo de subproducto generado en la reacción de transesterificación metílica, o el resultado de que diferentes diglicéridos se localizaran en un mismo pico, de modo que cuando se produzca la fragmentación de los iones en la fuente de ionización los fragmentos se reagrupen y den lugar a nuevas especies químicas.

Evaluando los otros picos de los cromatogramas de las diferentes muestras de aceite de oliva, se lograron identificar la mayoría de los picos: t_R 9.02 min, palmitato de metilo; t_R 9.29 min, ácido palmítico libre; t_R 10.46 min, oleato de metilo; t_R 10.77 min, ácido oleico libre; t_R 12.60 min, 1-oleil glicerol. La mayoría de las muestras presentaban todos estos picos.

A continuación se muestran algunos cromatogramas y sus correspondientes espectros de masas de muestras aceites de oliva virgen extra dónde pueden observarse dichos picos.

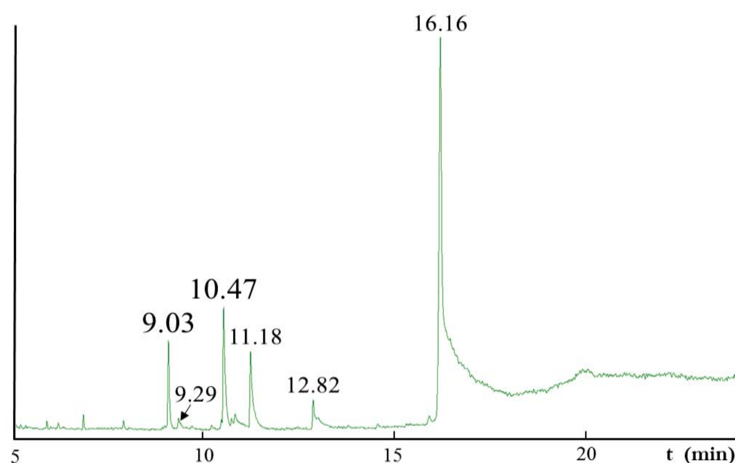


Figura 10. Señal cromatográfica (TIC) de un aceite de oliva virgen extra de la variedad **alfafarenca**.

Los picos cuyos tiempos de retención fueron 9.03 y 10.47 min fueron asignados a los ésteres palmitato de metilo y oleato de metílico respectivamente, ya que se compararon con el tiempo de retención y con el espectro de masas de los patrones de los mismos compuestos.

El pico en el minuto 9.29 fue asignado al ácido palmítico libre por comparación en tiempos de retención y en espectros de masas con el patrón del mismo (véase figura 11). El hecho de encontrar ácido palmítico libre en el aceite de oliva virgen extra u otro tipo de ácido libre como el ácido oleico, que también fue identificado en el minuto 10.78 en otras muestras de aceite de oliva, es debido a que en los aceites virgen extra existe una pequeña proporción de éstos ácidos libres que son los que proporcionan la acidez del aceite. Pero además cuando se lleva a cabo la reacción de transesterificación metílica en este estudio, se aplica una etapa de lavado con agua y junto con los restos de metóxido sódico que se encuentran en el medio, se podría producir una hidrólisis de los ésteres metílicos de los ácidos grasos dando lugar a los correspondientes ácidos libres.

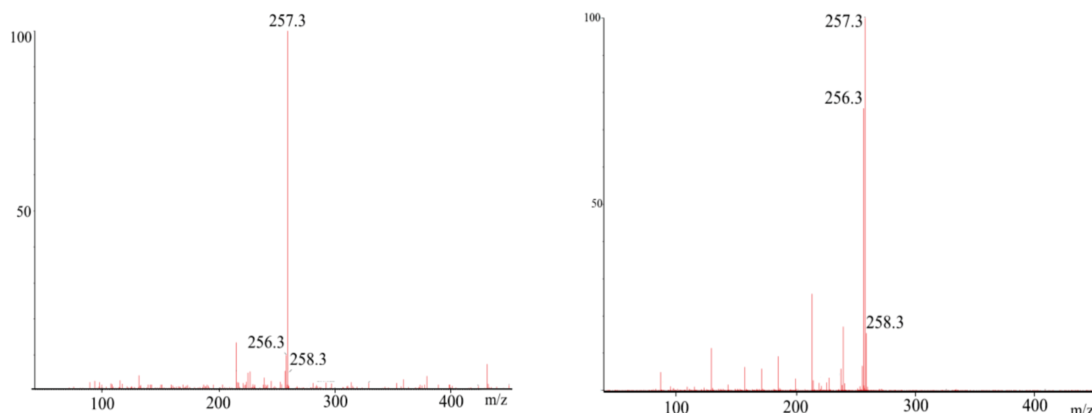


Figura 11. A la izquierda el espectro de masas correspondiente al ácido palmítico del aceite, y a la derecha el espectro de masas correspondiente al patrón de ácido palmítico.

En la siguiente muestra, correspondiente a un aceite de oliva virgen extra de variedad cornicabra, se pueden observar dos nuevos picos cromatográficos a 10.79 y 12.60 min, los cuales fueron asignados al ácido oleico y al 1-oleil glicerol, respectivamente (véase figura 12).

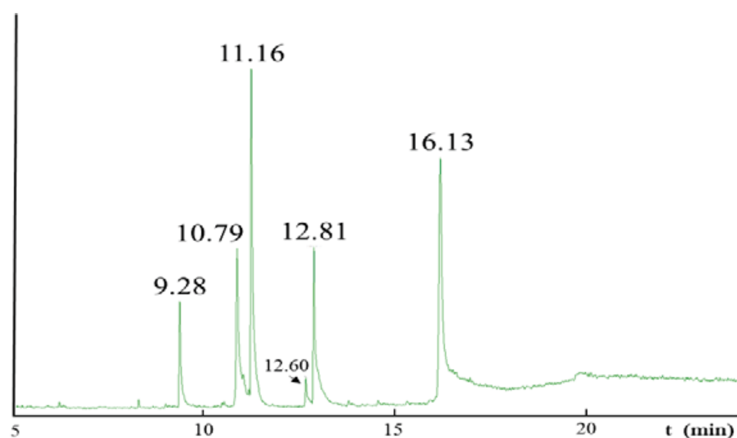


Figura 12. Señal cromatográfica (TIC) de un aceite de oliva virgen extra de la variedad **cornicabra**.

La identificación de ambos pico fue confirmada, al igual que en todos los casos anteriores, por comparación de los tiempos de retención y de sus espectros de masas correspondientes con el patrón de ácido oleico y con el espectro de masas asignado al 1-oleil glicerol a partir de la trioleína transesterificada. En la figura 13 se muestra el espectro de masas del pico a tiempo 12.60 min de la muestra de aceite de oliva, de la variedad cornicabra, y el espectro de masas de la trioleína transesterificada identificada en el minuto 12.47.

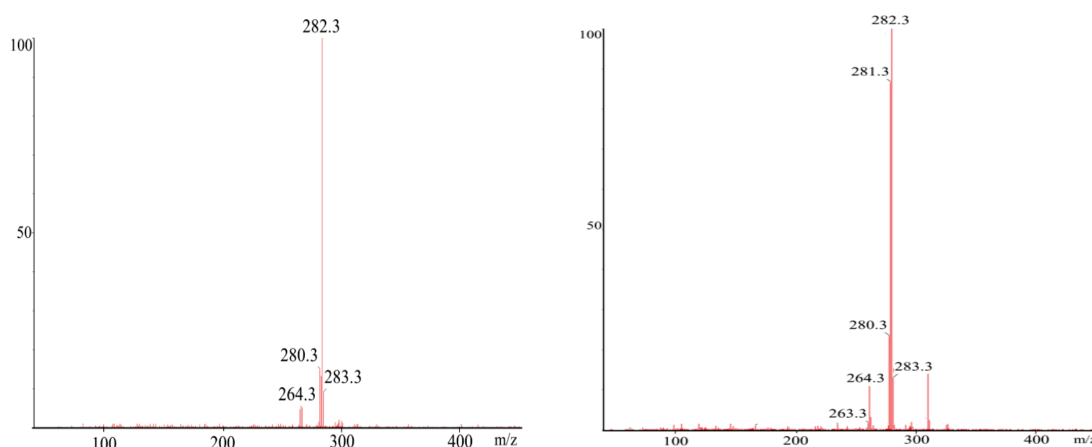


Figura 13. A la izquierda el espectro de masas correspondiente al 1-oleil glicerol del aceite de variedad cornicabra, y a la derecha el espectro de masas correspondiente 1-oleil glicerol de la trioleína transesterificada.

Aunque este estudio estuvo centrado principalmente en la búsqueda de compuestos mayoritarios del aceite de oliva y efectivamente en prácticamente todas las muestras los picos encontrados eran los comentados y descritos anteriormente, hubo una muestra que presentaba un pico a 13.05 min con elevada intensidad y el cual no estaba presente en ninguna de las restantes 58 muestras de aceite estudiadas, por ese motivo se llevó a cabo el estudio del mismo, y cuya estructura fue elucidada por nosotros y comprobada con la librería de espectros NIST, que aunque en esta librería los espectros registrados son cuando se utiliza la fuente ionización de impacto electrónico (EI) cuyo poder de fragmentación es mayor que la fuente APCI utilizada en este estudio. Los espectros de

masas eran prácticamente iguales, y se concluyó que correspondía con el compuesto 2,2'-metilen bis(4-metil-6-*tert*butilfenol), denominado comercialmente como Vulkanox BKF. Este compuesto suele ser utilizado como aditivo antioxidante y en la fabricación de ciertos envases de plástico destinados a contener aceite. La aparición de este compuesto en esta muestra puede ser debida o bien porque el compuesto haya migrado del envase original al aceite de oliva o porque haya sido añadido de forma ilícita al aceite para aumentar su estabilidad, aunque nos inclinamos por la primera opción como la más probable.

En la figura 14 se muestra la señal cromatográfica (TIC) donde pueden ver los tres picos comunes en todos los aceites y además el pico correspondiente al minuto 13.05.

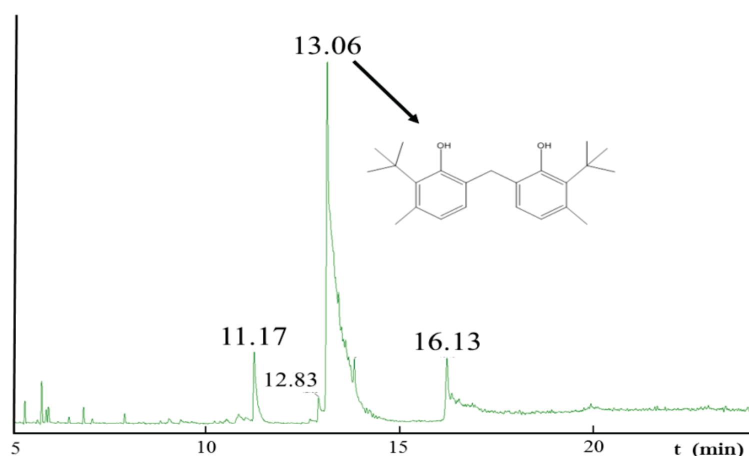


Figura 14. Señal cromatográfica (TIC) de un aceite de oliva virgen extra de la variedad **lechín**.

En la figura 15a puede verse el espectro de masas de la muestra de aceite de variedad lechín donde se observan los fragmentos mayoritarios 340 m/z, 284 m/z y 177 m/z. Y si lo comparamos con el espectro que aparece en la base de datos de la librería de espectros NIST mostrado en la figura 15b, podemos observar que aparecen los mismos fragmentos mayoritarios y que el patrón isotópico de ellos es prácticamente el mismo.

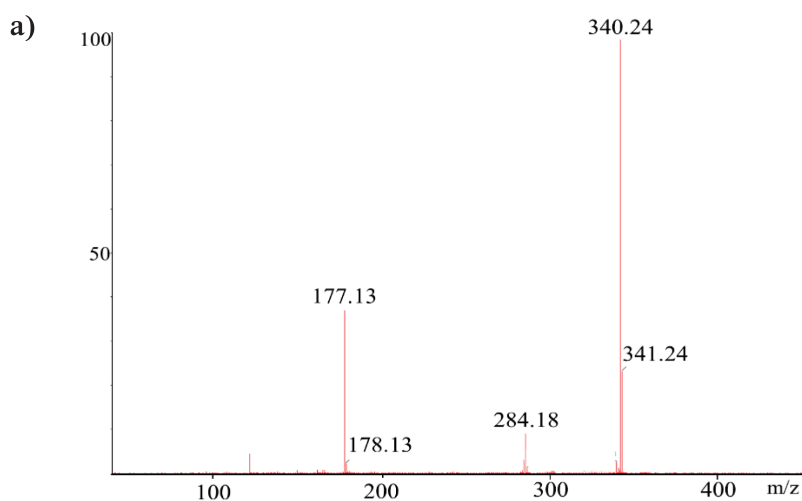
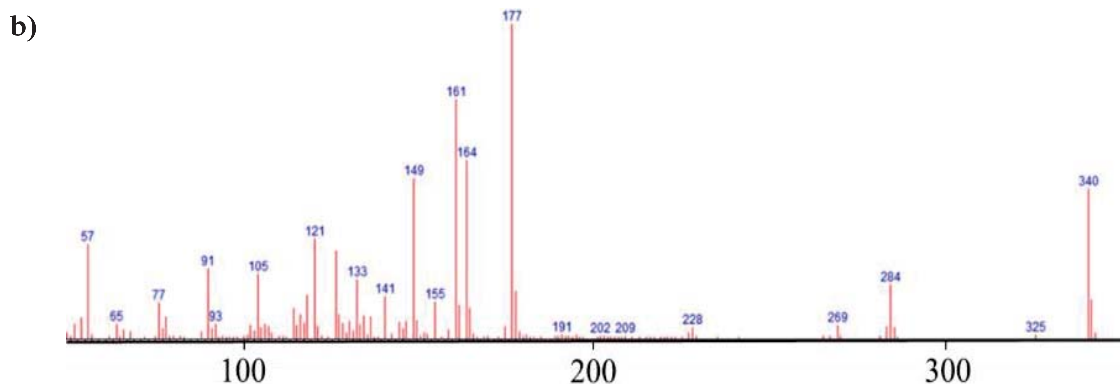


Figura 15. a) Espectro de masas correspondiente al pico eluido a 13.06 min de la muestra de aceite de oliva de variedad lechín. **b)** Espectro de masas de la base de datos NIST correspondiente al 2,2'-metilen bis(4-metil-6-*tert*butilfenol)



IV.7. Conclusiones

Los resultados obtenidos permiten corroborar la hipótesis inicial de los compuestos que cabría esperar en la fracción metil-transesterificada de los aceites de oliva virgen extra.

Se han identificado los compuestos principales y mayoritarios como el ácido oleico y palmítico en sus estructuras de ácidos libres y esterificados formando los ésteres metílicos correspondientes, entre otros. Además de caracterizar y elucidar la estructura química de algunos monoglicéridos resultantes de llevar a cabo la reacción de transesterificación metílica de los diferentes aceites de oliva.

Asimismo queda reflejada la potencialidad de este tipo de técnicas instrumentales cuando se aplica el enfoque 'no dirigido' para identificar los compuestos presentes en una matriz. Aunque no se haya podido desarrollar un método de cromatografía líquida acoplada a espectrometría de masas con objeto de poder relacionar de forma óptima las "jorobas" obtenidas en las cromatogramas adquiridos y utilizados en capítulos anteriores de esta tesis para desarrollar los diferentes modelos de clasificación, con los compuestos responsables de las mismas.

Quedaría pendiente aplicar el enfoque 'dirigido' con objeto de poder cuantificar los compuestos identificados y comprobar las proporciones y/o cantidades de los mismos presentes en las diferentes variedades de aceite de oliva virgen extra. Y del mismo modo estudiar la fracción transesterificada de aceites de no oliva con objeto de observar los compuestos presentes en la misma, aunque este trabajo ya sería complementario al objetivo de esta tesis doctoral.

Estudios complementarios a la tesis doctoral

ESTUDIOS COMPLEMENTARIOS A LA TESIS DOCTORAL

De forma paralela a la realización de la tesis doctoral se llevaron otros estudios relacionados con la autentificación de alimentos, algunos de ellos en colaboración con otros centros nacionales e internacionales de investigación, para: (i) clasificar el aceite de palma acorde a su origen geográfico, (ii) detectar adulteraciones de aceite de oliva de Argentina con otros aceites vegetales y autentificar su origen geográfico, y (iii) desarrollar y validar un panel de jamón.

- ❖ Pérez-Castaño, E., Ruiz-Samblas, C., Medina-Rodríguez, S., Quirós-Rodríguez, V., Jiménez-Carvelo, A.M., Valverde-Som, L., González-Casado, A., Cuadros-Rodríguez, L. (2015). Comparision of different analytical classification scenarios. Application for geographical origin of edible palm oil by sterolic (NP)HPLC fingerprinting. Analytical Methods, *7*, 4192-4201.
- ❖ Abidemi Obsisesan, K., Jiménez-Carvelo, A.M., Cuadros-Rodríguez, L., Ruisánchez, I., Callao, P. (2017). HPLC-UV and HPLC-CAD chromatographic data fusion for authentication of geographical origin of palm oil. Talanta, *170*, 413-418.
- ❖ Jiménez-Carvelo, A.M., Lozano, V., Olivieri, A. (2019). Comparative chemometric analysis of fluorescence and near infrared spectroscopies for authenticity confirmation and geographical origin of Argentinean extra virgin olive oils. Food Control, *96*, 22-28.
- ❖ González-Casado, A., Jiménez-Carvelo, A.M., Cuadros-Rodríguez, L. Sensory quality control of dry-cured ham: a comprehensive methodology for sensory panel qualification and method validation. *Enviado a la revista Meat Science (Mayo 2018).*
- ❖ Oriol, S., Jiménez-Carvelo, A.M., Callao, M.P, Ruisánchez, I. Authentication of the geographical origin of Arbequina extra-virgin olive oil using Raman spectroscopy and chemometrics. *Enviado a la revista Food Chemistry (Septiembre 2018).*

Participación en proyectos, convenios y contratos de investigación

PARTICIPACIÓN EN PROYECTOS, CONVENIOS Y CONTRATOS DE INVESTIGACIÓN

Durante el periodo de realización de la tesis doctoral, se ha participado o colaborado en diversos proyectos, convenios y contratos de investigación relacionados directa o indirectamente con los estudios experimentales recogidos en la tesis doctoral.

- Estudio de las diferentes tecnologías analíticas actuales aptas para complementar o constituir una alternativa futura al método comunitario de análisis sensorial denominado «panel test» (Proyecto INSTRUMENSORIAL-VOLÁTILES). Organización Interprofesional del Aceite de Oliva Español, (Ref.: OTRI-UGR 3501-2014).

- Análisis de los datos experimentales recopilados durante el desarrollo del proyecto: "Estudio de la evolución de los ésteres alquílicos en el aceite oliva virgen extra". Organización Interprofesional del Aceite de Oliva Español (Ref.: OTRI-UGR 3501-02-2015).

- Certificación y distribución de nueve materiales de referencia certificados (MRC) para análisis organoléptico de aceite de oliva (SensOLEO-MRC 2015). Consejería de Agricultura, Pesca y Desarrollo Rural, Junta de Andalucía (Ref.: OTRI-UGR 3637-2015).

- Desarrollo de un sistema de gestión de la calidad para el funcionamiento de un panel analítico de cata científica para el análisis sensorial de los alimentos a ser certificados, que reúna los requisitos establecidos en la norma UNE-EN ISO/IEC 17025. Fundación Qualytech Alimentación (Ref.: OTRI 3680-2015).

- Certificación y distribución de cinco materiales de referencia certificados (CMR) para análisis físico-químicos de aceite de oliva (InterOLEO-MRC 2016). Consejería de Agricultura, Pescar y Desarrollo Rural, Junta de Andalucía (Ref.: OTRI-UGR 3755-2016).

- Asesoramiento técnico entre la Organización Interprofesional del Aceite de Oliva Español y el profesor de la Universidad de Granada D. Luis Cuadros Rodríguez para la creación del Grupo Operativo Supra-autonómico SENSOLIVE-OIL. Organización Interprofesional del Aceite de Oliva Español y Ministerio de Agricultura Pesca, Alimentación y Medioambiente (Ref.: OTRI-UGR 3881-2017).

- Elaboración, certificación y distribución de nueve materiales de referencia certificados (MRC) para análisis organoléptico de aceite de oliva (SensOLEO-MRC 2017). Consejería de Agricultura, Pesca y Desarrollo Rural, Junta de Andalucía (Ref.: OTRI-UGR 3898-2017).

Conclusiones finales y perspectivas futuras

CONCLUSIONES FINALES

En esta sección se recogen las conclusiones derivadas de los estudios presentados en esta Tesis Doctoral.

I. Se han llevado a cabo dos revisiones bibliográficas: (i) sobre la aplicabilidad de las huellas dactilares cromatográficas usando la cromatografía de líquidos para autenticar el aceite oliva, así como la aplicación de técnicas de reconocimiento de pautas para este fin, la cual derivó en una publicación como capítulo de libro; y (ii) sobre el uso de nuevas técnicas emergentes de minería de datos para el tratamiento de datos en el ámbito de la alimentación, cuya potencialidad está siendo demostrada en los últimos estudios publicados en este campo, y resultó en un artículo científico de revisión, en formato "review".

II. Se han desarrollado y optimizado dos métodos analíticos de cromatografía de líquidos, en sus dos modalidades de trabajo (fase normal y fase invertida), acoplada a un detector de aerosol cargado en corona (HPLC-CAD) para obtener la huella dactilar cromatográfica de la fracción metil-transesterificada de aceites de oliva de diferentes categorías comerciales, aceites vegetables comestibles distintos al aceite de oliva y mezclas de aceite de oliva con otros aceites vegetales. Paralelamente, los cromatogramas obtenidos se han convertido en un vector da datos para su posterior tratamiento quimiométrico.

II.1. Se han aplicado dos estrategias ampliamente conocidas de desarrollo de modelos de clasificación multivariable, denominadas: (i) estrategia de dos clases de entrada y (ii) estrategia de una sola clase de entrada. Además se ha desarrollado y aplicado por primera vez en esta tesis, una nueva estrategia de desarrollo de modelos de clasificación denominada por nosotros '*pseudo* dos clases de entrada' (por sus siglas en inglés, "*pseudo* two-input class", *p*2iC) para poder emplear métodos de análisis discriminante cuando sólo se tengan muestras de la clase de interés, ya que éstos son más selectivos.

II.2. Se ha sustituido el uso de una columna de cromatografía convencional por un pre-columna para disminuir los tiempos de desarrollo cromatográfico y así obtener los cromatogramas de forma más rápida, lo cual permitió obtener huellas dactilares en 4 minutos, obteniéndose muy buenos resultados en los modelos de clasificación.

III. Se ha desarrollado un método global de discriminación de aceite de oliva de otros aceites vegetales y de cuantificación del mismo en mezclas con otros aceites vegetales comestibles aplicando técnicas vibraciones como son la técnica ATR-FTIR y Raman, permitiendo la cuantificación de aceite de oliva en mezclas compuestas por más de 10

aceites vegetales distintos. Aunque se haya llevado a cabo un pre-tratamiento previo al análisis de las muestras, esto ha mejorado notablemente la selectividad de éstas técnicas para resolver problemas relativos a la autentificación de aceite de oliva.

IV. Se ha caracterizado la fracción metil-transesterificada del aceite de oliva, identificando los principales ácidos grasos como son el ácido oleico y palmítico, los ésteres metílicos derivados de los mismos, y los residuos de triglicéridos, cuya derivatización no fue completa, en forma de monoglicéridos resultantes de llevar a cabo la reacción de transesterificación metílica sobre los aceites de oliva.

V. Se ha llevado a cabo la comparación entre las técnicas vibracionales y las técnicas cromatográficas y se ha podido concluir que:

V.1. Aunque las técnicas vibracionales presentan menor dificultad de uso y la medida sea rápida, el análisis no está automatizado, por lo que el analista debe permanecer delante del equipo y llevar a cabo las medidas personalmente, lo cual alarga el tiempo real de análisis. A diferencia de las técnicas cromatográficas, en las cuales el analista programa una secuencia de análisis y no tiene que estar delante del equipo mientras ésta se lleva a cabo, permitiendo trabajar de forma paralela.

V.2. La aplicación de la metodología de huellas dactilares presenta mayor potencialidad cuando se emplean técnicas cromatográficas en lugar de técnicas vibracionales. Esto se debe principalmente al fundamento de ambas técnicas analíticas. Las técnicas vibracionales reportan información sobre los grupos funcionales presentes en la muestra, por lo que si las muestras de estudio presentan los mismos o similares grupos funcionales, la selectividad de estas técnicas disminuye notablemente, haciendo que la metodología de huellas no sea del todo efectiva. Mientras que en el caso de las técnicas cromatográficas la información mostrada es relativa a los compuestos o familia de compuestos y por tanto una baja resolución del cromatograma ya presenta excelentes resultados, y si éste no fuera el caso, siempre es posible mejorar la separación cromatográfica para hacer más evidente ésta información.

VI. Se ha realizado también una comparación entre las diferentes modalidades de trabajo de la cromatografía de líquidos y se ha podido concluir que:

VI.1. En el estudio de la fracción metil-transesterificada la modalidad de trabajo de cromatografía líquida en fase normal es más selectiva y no da lugar a ningún error de clasificación. Esto se debe principalmente a la naturaleza química de los compuestos presenten en dicha fracción ya que son completamente solubles en disolventes orgánicos como n-hexano, utilizado para llevar a cabo la extracción

de los mismos y utilizado en el método cromatográfico desarrollado. Por tanto esta modalidad de trabajo era la óptima para estudiar esta fracción.

VI.2. Aunque la modalidad de fase invertida o reversa sea menos selectiva que la modalidad de fase normal también ha dado buenos resultados en la autentificación de aceite de oliva, evitando así el uso de disolventes orgánicos como el n-hexano.

PERSPECTIVAS FUTURAS

El trabajo presentado cumple con los objetivos establecidos inicialmente. Sin embargo tras el desarrollo de la tesis y de los resultados obtenidos se abren nuevas aportaciones que permiten realizar una seria de investigaciones futuras.

Teniendo en cuenta el trabajo realizado y presentado en el capítulo I sobre diferenciación de aceites de oliva de acuerdo a su variedad botánica, se podrían realizar diferentes estudios para aumentar las diferencias entre las fracciones de los aceites de oliva, estudiando la posibilidad de introducir otro tipo de grupo funcional diferente al epóxido aquí presentado.

Así mismo llevar a cabo la caracterización de la fracción metil-transesterificada del aceite de oliva con cromatografía de líquidos acoplada a espectrometría de masas con objeto de correlacionar las 'jorobas' formadas por conjuntos de picos cromatográficos no resueltos encontradas en las huellas dactilares obtenidas y utilizadas en los diferentes modelos multivariantes desarrollados.

VALORACIÓN GLOBAL DE LA TESIS DOCTORAL

Como resumen de todo el período doctoral se han adquirido diversas competencias, cómo son el manejo de diferentes plataformas analíticas, especialmente la cromatografía de líquidos, y también, aunque en menor medida, de cromatografía de gases acoplada a espectrometría de masas y de técnicas vibracionales como el infrarrojo y la espectroscopía Raman. Además de adquirir un alto nivel en el uso de métodos de reconocimiento de pautas sobre datos de diferentes órdenes, aunque principalmente de orden 1, y de fusión de datos.

Aunque en la tesis doctoral no se aporta ningún estudio dónde se aplique la metodología de fusión de datos ya que los modelos desarrollados y validados eran muy buenos y no fue necesario aplicar la fusión. Así mismo ha permitido adquirir la capacidad de

divulgación científica mediante la presentación oral y en cartel de los trabajos que conforman esta Tesis Doctoral en congresos nacionales e internacionales.