UNIVERSIDAD DE GRANADA

Avances analíticos para la mejora de la información sobre calidad y estabilidad de alimentos vegetales con alto contenido graso

TESIS DOCTORAL

Programa de doctorado en Química
Universidad de Granada

Sandra Martín Torres

Dpto. de Química Analítica
Prof. Fermín Capitán García

Directores:

Luis Cuadros Rodríguez

Antonio González Casado

Editor: Universidad de Granada. Tesis Doctorales
Autor: Sandra Martín Torres
ISBN: 978-84-1117-500-5
URI: https://hdl.handle.net/10481/77145

AGRADECIMIENTO A LA FINANCIACIÓN

Esta Tesis Doctoral ha sido realizada gracias a los recursos facilitados por el grupo de investigación "Análisis en Alimentación y Medio Ambiente" (Cod. PAIDI: FQM-232) en las instalaciones del Departamento de Química Analítica de la Universidad de Granada.

Algunos de los resultados que se presentan son parte del proyecto de I+D+i / ayuda *RTC-2017-6170-2*, financiado/a por MCIN/ AEI/10.13039/501100011033/ y "FEDER Una manera de hacer Europa".

MINISTERIO DE CIENCIA E INNOVACIÓN

UNIÓN EUROPEA

FONDO EUROPEO DE DESARROLLO REGIONAL

"Una manera de hacer Europa"

AGENCIA ESTATAL DE INVESTIGACIÓN

Se agradece además la ayuda para la formación de profesorado universitario, en el marco del Plan Estatal de Investigación Científica y Técnica y de Innovación 2017-2020 [FPU19/02078], financiada por el Ministerio de Ciencia, Innovación y Universidades y disfrutada durante el período 2020-2022.

MINISTERIO DE UNIVERSIDADES

LISTADO DE ABREVIATURAS

1D	1 dimensión
2D	2 dimensiones
3iC	3 clases de entrada
ADA	American Dietetic Association
AESAN	Agencia Española de Seguridad Alimentaria y Nutrición
AOV	Aceite de Oliva Virgen
AOVE	Aceite de Oliva Virgen Extra
AV	Índice de anisidina
BPL	Buenas Prácticas de Laboratorio
BRC	British Retail Consortium
C18	Octadecilsilil
CAD	Detector de aerosol cargado
CE	Consejo Europeo
CN	Ciano
COI	Consejo Oleícola Internacional
DAD	Detector de fila de diodos
DOP	Denominación de Origen Protegida
DSC	Calorimetría diferencial de barrido
ECN	Número equivalente de carbono
EFSA	Autoridad Europea de Seguridad Alimentaria
ESR	Resonancia de espín electrónico
ETG	Especialidad Tradicional Garantizada
FCC	Codex de Productos Químicos para la Alimentación
FD	Detector de fluorescencia
FID	Detector de ionización de llama
FTIR	Infrarrojos por transformada de Fourier
GA	Algoritmo genético
GC	Cromatografía de gases
GC-FID	Cromatografía de gases - detector de ionización de llama

GC-MS	Cromatografía de gases - espectrometría de masas
HPLC	Cromatografía de líquidos de altas prestaciones
HPLC-DAD	Cromatografía de líquidos - detector de fila de diodos
IFS	International Food Standard
IFST	Instituto de Ciencia y Tecnología de los Alimentos
IGP	Indicación Geográfica Protegida
IR	Infrarrojo
ISO	Organización Internacional de Normalización
IUPAC	Unión Internacional de Química Pura y Aplicada
k-NN	*K*-vecino más cercano
LC	Cromatografía de líquidos
LV	Variable latente
LC-MS	Cromatografía de líquidos- espectrometría de masas
Min	Minutos
MIR	Infrarrojo medio
MS	Espectrometría de masas
NIR	Infrarrojo cercano
Nitro-PAH	Hidrocarburo aromático policíclico nitrado
nm	Nanómetro
NMR	Resonancia magnética nuclear
NP	Fase normal
(NP)HPLC	Cromatografía de líquidos de altas prestaciones en modalidad fase normal
OCU	Organización de Consumidores y Usuarios
OSI	Índice de estabilidad oxidativa
PAH	Hidrocarburo aromático policíclico
PC	Componente principal
PCA	Análisis de componentes principales
PET	Polietileno tereftalato
PLS-DA	Análisis discriminante de mínimos cuadrados parciales
PLSR	Regresión de mínimos cuadrados parciales

PV	Índice de peróxidos
RBF	Función de base radial
RI	Índice de refracción
RP	Fase invertida
(RP)HPLC	Cromatografía de líquidos de altas prestaciones en modalidad fase invertida
RT	Tiempo de retención
SIMCA	Modelado independiente por analogía de clases
sPLS-DA	Análisis discriminante de mínimos cuadrados parciales disperso
SRS	Puntuación de retención estándar
SVM	Máquinas de soporte vectorial
SVM-C	Máquinas de soporte vectorial de clasificación
TG	Triglicéridos
TLC	Cromatografía de capa fina
TPC	Compuestos polares totales
t_R	Tiempo de retención
UE	Unión Europea
UHPLC	Cromatografía líquida de ultra-alta eficiencia
USDA	Departamento de Agricultura de Estados Unidos
USP	Farmacopea de los Estados Unidos
UV	Ultravioleta
UV-Vis	Ultravioleta-visible
VIP	Importancia de la variable en la proyección

ÍNDICE

PROBLEMA, HIPÓTESIS Y JUSTIFICACIÓN: OBJETIVOS INICIALES

PROBLEMA, HIPOTESIS Y JUSTIFICACIÓN

La autentificación de los alimentos implica la confirmación de que las especificaciones que ofrecen son verdaderas. Puede incluir muchos aspectos como la identificación y/o la cuantificación de componentes característicos, adulterantes y/o contaminantes, y la verificación de los requisitos de calidad diferenciada, entre ellos el origen botánico u origen geográfico y el procedimiento de fabricación o transformación. En el Informe anual del 2018 de la Comisión Europea sobre fraude alimentario ('The EU Food Fraud Network and the System for Administrative Assistance – Food Fraud') se indica qué considera la UE como fraude alimentario: "El fraude alimentario se trata de acciones intencionadas tomadas por empresas o individuos con el propósito de engañar a compradores y obtener una ventaja indebida de los mismos, en violación de la legislación de la UE sobre la cadena agroalimentaria". En particular, el fraude alimentario consiste en la descripción errónea deliberada para engañar a los consumidores sobre la verdadera naturaleza y/u origen del producto o de alguno de sus ingredientes y suele ser práctica común en aquellos alimentos de alto valor añadido. Es por ello que es de interés avanzar en la información analítica que se proporciona a los consumidores, para que ésta tenga un buen soporte científico que defina su calidad y evitar así prácticas fraudulentas.

Centrándonos en los alimentos con alto contenido graso, en especial los aceites vegetales comestibles, estos constituyen matrices analíticas muy complejas. Además sus características físico-químicas y organolépticas no son estables, sino que están sujetas a cambios provocados por diferentes procesos de degradación. Una vez que el aceite está envasado, o filtrado y almacenado en bodega, y siempre que se conserve en acondiciones aceptables (no expuesto a fuentes de calor o luz intensa), la principal alteración se debe a procesos oxidativos, principalmente de autooxidación, denominados genéricamente de enranciamiento, en los que influyen el oxígeno, la temperatura y la luz. Es por eso que, aunque en el inicio no haya una clara intención de incurrir en fraude, es común encontrar en el mercado aceites etiquetados bajo una categoría comercial superior a la que realmente pertenecen.

La normativa actual obliga a los envasadores a señalar una fecha de consumo preferente en la etiqueta de los aceites comestibles que normalmente, por convenio de la mayoría de grandes envasadoras, se fija entre 12 y 24 meses. Después de ese periodo un aceite por ejemplo de oliva se puede consumir con plena garantía, ya que no es un producto perecedero, lo que sí ocurre es que sus atributos sensoriales empeoran con el tiempo. Se puede alterar el color

inicial, se merman los aromas e intensidad del frutado y por tanto el flavor, de modo que los aceites manifiestan defectos, dejan de cumplir con los requisitos de calidad y, como consecuencia, pierden la categoría de calidad, pudiendo llegar a ser considerados, en un caso extremo, como no aptos para el consumo humano. Por ello, no deberían de comercializarse en la categoría inicial.

Una alerta lanzada por la OCU (Organización de Consumidores y Usuarios) en el 2018 [**1**] señalaba que la mitad de los aceites analizados de diferentes marcas en productos de gran consumo, calificados como AOVE, (no aceites "gourmet" o "premium") no cumplían los requisitos de la norma del Consejo Oleícola Internacional (COI) que regula las características de los aceites para ser considerados como tales, sino que pertenecían a una categoría inferior, más barata. Según los resultados que la OCU reflejaba en su publicación, ningún aceite presentó problemas en parámetros como la calidad o el cuidado del fruto por lo que no estamos hablando de un problema de salud pública ni de seguridad alimentaria pero sí se trataría de un fraude en el etiquetado pues el envasador debe garantizar que el aceite mantiene sus características iniciales en todo el periodo de consumo que ofrece. Estos estudios se repiten cada año, con resultados similares. Además existen otras alertas como las lanzadas por la Agencia de Seguridad Alimentaria (AESAN) por la venta de aceite, etiquetado como "aceite de oliva" y "aceite de oliva virgen extra", pero que en realidad no es tal, sino un producto elaborado, envasado y distribuido fuera de las normas [**2**].

Se plantea por tanto la posibilidad de utilizar técnicas avanzadas de análisis químico combinadas con herramientas quimiométricas para autenticar, en términos de calidad, alimentos de alto contenido graso como son el aguacate o los aceites vegetales obtenidos del fruto de la aceituna o de distintas semillas vegetales. Además, se pretende establecer un índice de vida media como combinación o función de los datos analíticos de una muestra de aceite, que reflejará su estabilidad en orden a su utilización posterior, y que permitirá tener una estimación del periodo de vida útil del aceite considerado. Para ello se contará con un número suficiente de muestras representativas de la población; se realizarán medidas aisladas (modalidad univariable) y se empleará la metodología de "huella instrumental" cromatográfica (modalidad multivariable). Toda esta información se analizará mediante el empleo de métodos

1. OCU. OCU denuncia fraude en el etiquetado de 20 marcas de aceite de oliva virgen extra. https://www.ocu.org/organizacion/prensa/notas-de-prensa/2018/aceiteoliva251018, (accedido 06/06/2022).
2. AESAN. Alerta por elaboración, envasado y distribución clandestinas de aceite etiquetado como “de oliva virgen” y “de oliva virgen extra” procedente de España (ref. inf2022/0054). Ministerio de Consumo, Gobierno de España. https://www.aesan.gob.es/AECOSAN/web/seguridad_alimentaria/ampliacion/INF2022_54.htm, (accedido 20/06/2022).

quimiométricos de tratamiento de datos que permitan combinar los datos y huellas analíticas en una matriz. Así mismo, se conjugarán ensayos que permiten monitorizar la oxidación a tiempo real y la misma en condiciones aceleradas, por ejemplo en un proceso de fritura. Todo esto nos permitirá establecer la cinética de degradación del aceite y realizar una estimación de su vida útil para resolver los problemas propuestos en relación con la estabilidad de aceites vegetales comestibles.

OBJETIVOS

La tesis doctoral se centra en la autenticación, en términos de calidad y estabilidad, de alimentos vegetales con alto contenido graso. Para ello, se plantean dos objetivos generales[3]:

I. Emplear métodos analíticos cromatográficos aplicando la metodología de huella instrumental para autentificar, clasificar y diferenciar aceites vegetales en términos de calidad como función a su origen geográfico, botánico o estabilidad relativa.

II. Estudiar la estabilidad de diferentes aceites vegetales (destacando el aceite de oliva virgen extra) a lo largo de su periodo de consumo realizando medidas analíticas aisladas (por ejemplo, determinación de índice de refracción, anisidina, tocoferoles, etc.) y medidas de huella instrumental cromatográfica (por ejemplo intensidad de compuestos volátiles) y combinando éstas para desarrollar modelos cinéticos multivariables para la determinación de unos índices de estabilidad frente a la rancidez que permitan predecir su periodo de cumplimiento con los requisitos de calidad relativos a su categoría.

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

[3] *Por lógica en la secuenciación los objetivos deberían disponerse tras la contextualización, incluida en el Capítulo I. Sin embargo, por coherencia en la redacción de esta tesis doctoral, se ha decidido su inclusión tras la hipótesis planteada inicialmente.*

RESÚMEN

RESÚMEN

Esta Tesis Doctoral persigue desarrollar y aplicar herramientas analíticas para avanzar en la información científica proporcionada al consumidor en relación a la calidad de alimentos vegetales de alto contenido graso.

En primer lugar el *Capítulo I* permite poner en contexto sobre la información científica de la que se dispone para poder llevar a cabo la investigación que en esta Tesis Doctoral se presenta.

El *Capítulo II* muestra aplicaciones de la metodología de huellas instrumentales ('fingerprinting') para la autenticación de aguacates a partir de su extracto graso, en función de la variedad botánica y de su origen geográfico. Las huellas se adquirieron mediante cromatografía líquida y cromatografía de gases acopladas a diferentes detectores y empleando métodos sencillos y cortos tiempos de análisis. Éstas llevan implícita información sobre el perfil de composición de la fracción lipídica que está relacionada con características particulares de cada aguacate. Para poder obtener esta información, se aplicaron diversas herramientas quimiométricas de pre-procesamiento de datos y se establecieron modelos de clasificación multivariables. Para cada modelado se describieron y calcularon las métricas de clasificación adecuadas por lo que se presentan conclusiones sobre la información suministrada por cada tipo de huella, así como sobre la herramienta quimiométrica adecuada.

El *Capítulo III* presenta una propuesta de metodología para obtener señales cromatográficas independientes ("agnostizadas") del estado del sistema cromatográfico en el que se obtienen, que permitirá obtener mejoras en los sistemas de identificación cuando los análisis son llevados a cabo empleando equipos de espectrometría de masas acoplados a cromatografía, así como en la realización de modelos multivariables cuando las huellas cromatográficas son empleadas como señal de entrenamiento de un modelo de discriminación o clasificación.

En el *Capítulo IV* se presenta una propuesta innovadora para establecer un índice global de estabilidad calculado sobre la base de modelos cinéticos multivariables, que permita realizar una estimación fiable del periodo de vida útil del aceite considerado. Se realizaron ensayos de envejecimientos tanto en condiciones normales como en condiciones aceleradas, simulando el proceso de fritura. Se combinaron medidas analíticas aisladas, así como medidas de huellas instrumentales cromatográficas empleando para ello numerosas técnicas analíticas como: espectroscopía de absorción molecular, cromatografía de líquidos y cromatografía de gases. Todo esto permitió además la comparación, en términos de estabilidad, de diversos aceites vegetales comestibles.

Por último en el *Capítulo V* se realiza una discusión integrada de todos los resultados obtenidos en el desarrollo de esta Tesis.

Por todo esto, se han realizado las siguientes contribuciones a congresos nacionales e internacionales:

1. *Autentificación de variedades de aguacate mediante cromatografía de líquidos (hplc) y métodos de clasificación multivariable.* XVI Reunión del Grupo Regional Andaluz de la Sociedad Española de Química Analítica (GRASEQA, 2018. Póster.
2. *Diferenciación de la variedad botánica de aguacates a partir de su huella dactilar cromatográfica (rp-hplc).* Primer congreso de Jóvenes Investigadores en Ciencias Agroalimentarias, 2018. Póster.
3. *Clasificación de aceites de oliva de acuerdo a su origen botánico*. XIX Simposium Científico-Técnico. Foro de la Industria Oleícola, Tecnología y Calidad. XIX Feria Internacional del Aceite de Oliva e Industrias Afines (EXPOLIVA), 2019. Póster.
4. *Classification of avocado varieties using liquid and gas chromatographic fingerprints data fusion.* X Colloquium Chemiometricum Mediterraneum , 2019. Oral Flash.
5. *Desarrollo de un modelo cinético mutivariable para la determinación del tiempo de vida útil de aceites vegetales comestibles.* XXII Reunión de la Sociedad Española de Química Analítica, 2019. Póster.
6. *Development of a multivariate kinetic model for the determination of the shelf life of virgin olive oils.* 17th Euro Fed Lipid Congress and Expo, 2019. Póster.
7. *'Coaching' educativo para trabajos de fin de grado (TFG) de 'Química Analítica'.* XVII Foro Internacional sobre la Evaluación de la Calidad de la Investigación y de la Educación Superior (FECIES), 2020. Póster.
8. *Desarrollo de un modelo multivariable de estabilidad para la determinación del tiempo de vida útil de aceites vegetales comestibles.* XX Feria Internacional del Aceite de Oliva e Industrias Afines (EXPOLIVA), 2022. Póster.
9. *Desarrollo de un modelo multivariable para la determinación del tiempo de vida útil de aceites vegetales refinados en condiciones acelerada*. XXIII Reunión de la Sociedad Española de Química Analítica, 2022. Oral.

Además, se han publicado los siguientes artículos científicos y capítulos de libros, que se recogen a lo largo de los diferentes capítulos de esta Tesis Doctoral:

I. **Capítulo de libro.** Jiménez-Carvelo, Ana M.; et al. 2021. Nontargeted fingerprinting approaches. In: C.M. Galanakis (Ed.), Food Traceability and Authentication. Academic Press / Elsevier, Cambridge, MA, pp. 163–193 ch. 6. DOI: 10.1016/B978-0-12-821104-5.00010-6

II. **Artículo científico.** Martín-Torres, Sandra; et al. 2019. Differentiation of avocados according to their botanical variety using liquid chromatographic fingerprinting and multivariate classification tree. Journal of the Science of Food and Agriculture 99, 4932-4941. DOI 10.1002/jsfa.9725

III. **Artículo científico.** Martín-Torres, Sandra; et al. 2020. Authentication of the geographical origin and the botanical variety of avocados using liquid chromatography fingerprinting and deep learning methods. Chemometrics and Laboratory Systems 199, 103960. DOI 10.1016/j.chemolab.2020.103960

IV. **Artículo científico.** Cuadros-Rodríguez, Luis; et al. 2021. Standardization of chromatographic signals – Part II: Expanding instrument-agnostic fingerprints to reverse phase liquid chromatography. Journal of Chromatography A 1641, 461973. DOI: 10.1016/j.chroma.2021.461973

V. **Capítulo de libro.** López-Ruíz, Rosalia; et al. 2022. Instrument-agnostizing methodology for liquid chromatography-mass spectrometry systems. In González Domínguez, Raúl (ed.) *Mass Spectrometry for Metabolomics-Methods in Molecular Biology,* Springer US, 1st Edition (in press). ISBN: 9781071626986

VI. **Artículo científco.** Martín-Torres, Sandra; et al. 2021. Applications of multivariate data analysis in shelf life studies of edible vegetal oils – A rewiew of the few past years. Food Packaging and Shelf Life 31, 100790. DOI: 10.1016/j.fpsl.2021.100790

VII. **Artículo científico.** Martín-Torres, Sandra; et al. 2022. Multivariate stability monitoring and shelf life models of deterioration of vegetable oils under real time ageing conditions – Part 1: Extra virgin olive oil as a main case of study. Food Packaging and Shelf Life, under review.

VIII. **Artículo científico.** Martín-Torres et al. 2022. Multivariate stability monitoring and shelf life models – Part 2: The matter of further edible vegetable oils. Food Packaging and Shelf Life, under review.

IX. **Artículo científico.** Martín-Torres et al. 2022. A comparison of the stability of refined edible vegetable oils under frying conditions: multivariate fingerprinting approach. Food Control, under review.

COMPETENCIAS ADQUIRIDAS

El Real Decreto 99/2011, del 28 de enero [**4**], es la norma que regula actualmente las enseñanzas oficiales de doctorado. Los estudios de doctorado tienen como objetivo la formación avanzada en la investigación y reúnen una serie de competencias básicas, así como aquellas otras que figuren en el Marco Español de Cualificaciones para la Educación Superior, y establece que los estudios de doctorado garantizarán la adquisición por el doctorando de las mimas. En concreto, el programa de doctorado en química recoge una serie de competencias Básicas, Personales y Específicas, descritas a continuación, las cuales el doctorando debe adquirir a lo largo de todo el periodo predoctoral.

Competencias básicas y generales

CB11 - Comprensión sistemática de un campo de estudio y dominio de las habilidades y métodos de investigación relacionados con dicho campo.

CB12 - Capacidad de concebir, diseñar o crear, poner en práctica y adoptar un proceso sustancial de investigación o creación.

CB13 - Capacidad para contribuir a la ampliación de las fronteras del conocimiento a través de una investigación original.

CB14 - Capacidad de realizar un análisis crítico y de evaluación y síntesis de ideas nuevas y complejas.

CB15 - 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.

CB16 - Capacidad de fomentar, en contextos académicos y profesionales, el avance científico, tecnológico, social, artístico o cultural dentro de una sociedad basada en el conocimiento.

Capacidades y destrezas personales

CA01 - Desenvolverse en contextos en los que hay poca información específica.

CA02 - Encontrar las preguntas claves que hay que responder para resolver un problema complejo.

CA03 - Diseñar, crear, desarrollar y emprender proyectos novedosos e innovadores en su ámbito de conocimiento.

CA04 - Trabajar tanto en equipo como de manera autónoma en un contexto internacional o multidisciplinar.

4. Real Decreto 99/2011, del 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.

CA05 - Integrar conocimientos, enfrentarse a la complejidad y formular juicios con información limitada.

CA06 - La crítica y defensa intelectual de soluciones.

Otras competencias

OC1 - Habilidad de encontrar fuentes de financiación para programas de investigación.

CE01 - Capacidad para idear, planificar y realizar de manera autónoma proyectos de investigación científica.

CE02 - La adquisición de las habilidades y destrezas para integrarse eficazmente en cualquier equipo de I+D+i en el ámbito de las Ciencias y Tecnologías Químicas

CE03 - Capacidad de asesorar, desde un punto de vista científico-técnico, a instituciones públicas y privadas en el ámbito de las Ciencias y Tecnologías Químicas

CE04 - Conocimiento de los avances científicos, de las últimas técnicas y de la instrumentación relacionados con su área química concreta de especialización

CE05 - Capacidad de actualización de su conocimiento científico y técnico de forma autónoma y continua

Cada una de estas competencias ha sido trabajada en mayor o menor grado durante la realización de esta Tesis Doctoral. Las acciones realizadas como la búsqueda bibliográfica de información de alto valor científico; participación en las actividades formativas (obligatorias y voluntarias) propuestas por el programa de doctorado; redacción y participación en proyectos de investigación; asistencia a congresos de carácter nacional e internacional; realización del trabajo experimental, lo que conllevó el manejo de equipos especializados, técnicas novedosas de análisis, en la mayoría de las ocasiones de forma no dirigida, la toma de decisiones, la organización del trabajo, el trabajo en equipo, la gestión de residuos generados y el registro de material y reactivos; tratamiento de datos y la discusión de resultados; el empleo de softwares de tratamiento de datos; la difusión de objetivos y resultados obtenidos en reuniones con expertos del sector y mediante la publicación de artículos científicos; impartición de docencias; así como la redacción y exposición de esta memoria final; entre otras muchas actividades más, me ha permitido completar una formación predoctoral integral y obtener las aptitudes necesarias para mi integración dentro de la comunidad científica.

CAPÍTULO I

CONTEXTUALIZACIÓN

I.1. Antecedentes. Aceites y alimentos con alto contenido graso

En la sociedad actual los consumidores demandan alimentos que además de cubrir sus necesidades energéticas, sean saludables y seguros. Dentro de este grupo de alimentos se encuentran aquellos ricos en lípidos polinsaturados. Los lípidos son moléculas presentes en todos los organismos vivos y, por tanto, están presentes en la mayoría de los alimentos, aunque en distintas concentraciones y con diferentes composiciones y propiedades [**1**]. Con respecto a sus propiedades físicas, los lípidos que son sólidos a temperatura ambiente son conocidos como grasas (como los originados en el mesocarpio del coco o de palma), mientras que los que son líquidos en las mismas condiciones se conocen como aceites (como los extraídos de la soja o las semillas de girasol).

De acuerdo a su naturaleza química, los aceites y grasas comerciales son mezclas de distintas moléculas orgánicas. En general, en la composición de aceites y alimentos con alto contenido graso se han diferenciado tradicionalmente dos fracciones: fracción mayoritaria, o saponificable, que representa generalmente más del 95% del peso total del alimento y constituida principalmente por triglicéridos, y en menor medida por diglicéridos, monoglicéridos, fosfolípidos, ceras y ácidos grasos libres; y la fracción minoritaria o insaponificable, que incluye una gran variedad de compuestos tales como esteroles y estanoles, alcoholes terpénicos, alcoholes alifáticos, tocoles y tocoferoles, clorofilas, carotenoides, vitaminas liposolubles o compuestos fenólicos, entre otros. Son por tanto muestras muy complejas y resulta difícil determinar de forma precisa la totalidad de los constituyentes minoritarios debido a su naturaleza diversa y a su baja concentración. Además, algunos de dichos constituyentes minoritarios sólo están presentes en el aceite crudo, y el procesado tecnológico, como la refinación, los elimina [**2**].

Por ello, aunque los aceites y grasas son principalmente fuente de lípidos en la dieta, también cubren las demandas del consumidor relacionadas con la salud, además de otras necesidades esenciales de la dieta. Debido principalmente a la disponibilidad y a las preferencias de los consumidores, los aceites y grasas comestibles son en su mayoría de origen vegetal [**3**].

Los aceites y, en general, los alimentos de origen vegetal con alto contenido graso son componentes esenciales en la dieta mediterránea y uno de los pilares de nuestra alimentación. Entre dichos alimentos destaca el aceite de

---

1. Sikorski, Z.E., Kolakowska, A. Chemical, Biological and Functional Aspects of Food Lipids, 2nd ed., CRC Press Taylor & Francis Group, New York, 2011.
2. Bosso, T. Olive oil Handbook, vol. I, Academic Pages, New York, 2017.
3. Fawzy Ramadan, M. Fruit Oils: Chemistry and Functionality, Springer, Switzerland, 2020.

oliva virgen (AOV), que se define, tal y como establece el Reglamento (CE) nº 1234/2007 [**4**] como '*aceite obtenido del fruto del olivo (Olea europea L.) únicamente por procedimientos mecánicos u otros procedimientos físicos en condiciones que no produzcan la alteración del aceite, que no hayan tenido más tratamiento que el lavado, la decantación, la centrifugación y el filtrado'*; encontrándose en la base de la pirámide de la Dieta Mediterránea y con efectos beneficiosos probados en la salud humana [**5**, **6**]. El término virgen garantiza que el aceite no ha sido mezclado con aceites de oliva de menor calidad o con otros aceites, ni ha sido sometido al proceso de refinado. El AOV de máxima calidad es el denominado aceite de oliva virgen extra (AOVE) que se caracteriza por la ausencia total de defectos fisicoquímicos y sensoriales y constituye una categoría comercial diferenciada del AOV.

España es el primer productor y exportador mundial de aceite de oliva virgen. El olivar y el aceite de oliva conforman un sector de gran relevancia económica, social y comercial en nuestro país y en los mercados internacionales, con una demanda en constante crecimiento, con una comercialización media en torno a 1.400.000 toneladas al año, representando casi el 70% de la producción de la UE y más del 40% de la producción mundial [**7**].

El aceite de oliva virgen extra es muy apreciado además de por sus propiedades sensoriales, por sus declaraciones de propiedades saludables debido, entre otras cosas, a su composición equilibrada de ácidos grasos (precursores de eicosanoides, ácidos linoleico y α-linolénico). Es la primera grasa reconocida oficialmente como saludable (además de su reconocimiento nutricional favorable) por varias autoridades, en particular por el Departamento de Agricultura de Estados Unidos (USDA) y en Europa, por la Autoridad Europea de Seguridad Alimentaria (EFSA), así como por el Consejo Europeo (CE).

Otro claro ejemplo de las propiedades saludables de algunos lípidos vegetales es el fruto del aguacate, único por su alto valor nutricional, debido a su alto contenido en ácidos grasos mayoritariamente insaturados. Es una rica fuente

4. Reglamento (CE) nº 1234/2007 de Consejo 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, (pp.1-149), 2007.
5. Consejo Oleícola Internacional. El aceite de oliva. http://www.internationaloliveoil.org/estaticos/view/83-designations-and-definitions-of-olive-oils, (accedido 28 marzo 2022).
6. Fenández Gutiérrez, A., Segura Carretero, A. El Aceite de Oliva Virgen: Tesoro de Andalucía, Fundación Unicaja, 2010.
7. Ministerio de Agricultura, Pesca y Alimentación – Gobierno de España. https://www.mapa.gob.es/estadistica/pags/anuario/2021Avance/AVANCE_ANUARIO/AvAE21.pdf (accedido 13 junio 2022).

de lípidos y es muy apreciado por ser una excelente fuente de energía y vitaminas. Según la 'American Dietetic Association' (ADA), el aguacate está clasificado como alimento funcional debido a sus probados efectos beneficiosos para la salud humana [**8**]. Al igual que el aceite de oliva, el aceite de aguacate se extrae de la pulpa del fruto, en lugar de la semilla, y puede consumirse crudo sin necesidad de refinarlo; así mismo puede clasificarse en virgen extra, virgen o puro, según los métodos y condiciones de extracción. El aceite de aguacate es cada vez más popular como fuente grasa con propiedades saludables. Actualmente existe una propuesta para la revisión de la Norma de aceites vegetales especificados en el CODEX (CODEX STAN 210-1999), y se está estudiando la inclusión del aceite de aguacate en dicha Norma [**9**].

Las propiedades químicas y físicas de las grasas y los aceites, así como su valor nutricional y estabilidad o resistencia a la degradación vienen determinadas en gran medida por la proporción de ácidos grasos que contienen, su longitud, posición de dobles enlaces y su posición dentro de la molécula de triacilglicerol, que varían para cada aceite y grasa individual; así como al contenido y composición de su fracción insaponificable. Son precisamente estos componentes minoritarios de la fracción insaponificable, presentes a muy bajas concentraciones (normalmente en el rango de los mg/kg) los que presentan y proporcionan a determinados aceites vegetales importantes propiedades biológicas (compuestos bioactivos). Entre dichos compuestos destacan el α-tocoferol (con actividad de vitamina E), el β-caroteno (que funciona como vitamina A), y los compuestos fenólicos (con alta actividad antioxidante lo que proporciona una mayor resistencia a la degradación por procesos oxidativos habituales) [**10**].

I.2. Calidad y autenticidad de alimentos

Una definición muy general pero útil de calidad alimentaria es la capacidad de satisfacer las expectativas del consumidor. Debido a que la calidad de los alimentos con alto contenido graso depende de la estabilidad de los compuestos polinsaturados, hace que realmente la calidad sea un estado

---

8. Krumreich, F.D., Borges, C.D., Mendonça, C.R.B., Jansen Alves, C., Zambiazi, R.C. Bioactive compounds and quality parameters of avocado oil obtained by different processes. Food Chem., 257 (2018), 376-381. https://doi.org/10.1016/j.foodchem.2018.03.048
9. Comisión del Codex Alimentarius, CRD1, Vigésima séptima reunión, 18 - 26 de octubre de 2021.
10. Aparicio, R., Harwood, J. Handbook of Olive Oil. Analysis and Properties, 2nd ed., Springer, New York, 2013. https://doi.org/10.1007/978-1-4614-7777-8

dinámico, que continuamente se va reduciendo, debido al paso del tiempo, después de su producción. Los productores y consumidores de productos alimentarios valoran cada día más la calidad y la seguridad alimentaria y es de interés avanzar en información analítica, bien soportada científicamente, para la caracterización de productos alimentarios de alto contenido graso en término de su calidad y seguridad. Además, el concepto de autenticidad, definido por algo original, verdadero, indudable, etc. aplicado a los alimentos, certifica que estos productos tienen un origen determinado en concordancia con las normas vigentes, quedando reflejada en las indicaciones adecuadas de la etiqueta de presentación.

La evaluación de la calidad de los alimentos se ha hecho históricamente en términos de salubridad y genuinidad. Sin embargo, en los últimos tiempos, factores como la identificación de la fuente botánica, del origen geográfico, las diversas y variadas contaminaciones que entran a través del medio ambiente, incluidos los residuos agroquímicos, el descubrimiento de nuevos peligros para la salud, la fabricación y comercialización de nuevas formulaciones y mezclas, los cambios químicos, bioquímicos y biológicos inducidos en la composición, los cambios químicos asociados con el procesamiento, incluidos los tratamientos térmicos, así como la aceptabilidad del consumidor, han llamado la atención tomando conciencia de las implicaciones de muchos de estos factores en el éxito de la operación de procesamiento de alimentos, el mantenimiento de la calidad y la uniformidad del producto y, sobre todo, la salubridad de los artículos alimentarios que llegan al consumidor [**11**]. Cualquier acto fraudulento, como la adulteración con ingredientes más baratos o la falsa declaración de origen (geográfico o botánico), reducen la calidad de los productos, engañan al consumidor y pueden incluso implicar un riesgo para la salud. Por ello, garantizar la autenticidad de los alimentos es una labor importante para todos los implicados en el comercio de alimentos: los consumidores, las autoridades de protección del consumidor y también los productores y comerciantes.

La comercialización de alimentos con alto contenido graso está frecuentemente regulada por normas de obligado cumplimiento especialmente en el ámbito de la seguridad alimentaria (residuos y contaminantes), pero también de calidad alimentaria, algunas de ellas se aplican internacionalmente, y otras solo en ciertos países, por lo que en ocasiones pueden actuar como barreras comerciales, dificultando la circulación de productos. Un claro ejemplo es el AOV que se rige por diferentes normas en función de dónde se comercialice: tres de las normas más importantes son las especificadas por la Unión Europea, el Consejo Oleícola Internacional y el Codex Alimentarius. Estas

11. Singhal, R.S., Kulkarni, P.R., Rege, D.V. Handbook of Indices of Food Quality and Authenticity, Woodhead Publishing Limited, England, 1997.

establecen los parámetros que se pueden utilizar para comprobar la calidad y la pureza de los aceites, así como los límites para cada parámetro aplicando una determinada metodología. Definen además las diferentes categorías comerciales de aceites de oliva en función de su calidad: virgen extra, virgen, corriente (solo en el Codex), y lampante (refinado y mezclas) [**12**].

Además de las normas de obligado complimiento existen una serie de estándares de calidad diferenciada, definida ésta como el conjunto de características peculiares y específicas de un alimento debidas al origen de las materias primas utilizadas y/o a los procedimientos de elaboración. Estos estándares están recogidos en el Reglamento (UE) 1151/2012 [**13**] y a su vez, en el Reglamento (UE) 2017/625 [**14**] que modifica el anterior en cuestiones relacionadas con el Control Oficial. Entre las dos regulaciones se garantiza el cumplimiento de unos requisitos de calidad diferenciada adicionales a los exigidos para el resto de productos convencionales. Son esquemas de carácter voluntario y los productos que cumplen con esos requisitos están inscritos en un registro de la UE y protegidos por derechos de propiedad intelectual. Los productos cuya calidad o características se deben al medio geográfico con sus factores naturales y humanos, y cuya producción, transformación y elaboración se realizan siempre en esa zona geográfica delimitada de la que toman el nombre se protegen bajo Denominación de Origen (DOP). Los productos con una Indicación Geográfica Protegida (IGP) poseen alguna cualidad determinada o reputación u otra característica que pueda atribuirse a un origen geográfico y cuya producción, transformación o elaboración se realiza en la zona geográfica delimitada de la que toma su nombre. Finalmente, las Especialidades Tradicionales Garantizadas (ETG) son los productos con rasgos específicos diferenciadores de otros alimentos de su misma categoría, y se producen a partir de materias primas tradicionales, o bien presentan una composición, modo de producción o transformación tradicional. Los alimentos que cumplen con estos estándares voluntarios de calidad proporcionan al consumidor una información clara sobre su origen aportando seguridad ante el consumidor y otorgando un valor añadido al producto.

---

12. Reglamento de ejecución (UE) Nº 1348/2013 de la comisión que modifica 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, 2013.
13. Reglamento (UE) 1151/2012 del Parlamento Europeo y del Consejo sobre los regímenes de calidad de los productos agrícolas y alimenticios, 2012.
14. Reglamento (UE) 2017/625 del Parlamento Europeo y del Consejo relativo a los controles y otras actividades oficiales realizados sobre los regímenes de calidad de los productos agrícolas y alimenticios para garantizar la aplicación de la legislación sobre alimentos y piensos, y de las normas sobre salud y bienestar de los animales, sanidad vegetal y productos fitosanitarios, y por el que se modifican los Reglamentos (CE y UE) previos (Reglamento sobre controles oficiales), 2017.

I.3. Estabilidad y vida media de aceites y alimentos con alto contenido graso.

La calidad de los productos alimenticios está directamente relacionada no sólo con las fases de producción y envasado, sino también con su estabilidad y vida útil. Conocer la estabilidad y vida útil de los alimentos es fundamental tanto para el consumidor como para la industria alimentaria con los objetivos de desarrollar nuevos productos alimentarios que ofrezcan la calidad esperada por el consumidor final en términos de contenido nutricional, apariencia, textura sabor y aroma, así como garantizar que los alimentos producidos y envasados cumplan con cualquier requerimiento legal, así como con cualquier declaración incluida en su etiqueta o envase.

De acuerdo a la legislación de la Unión Europea [**15**] la vida útil o vida media hace referencia a la "fecha de duración mínima" y exige que ésta se indique mediante una "fecha de consumo preferente" o "fecha de caducidad". La fecha de consumo preferente es la fecha hasta la que un producto alimenticio conserva sus propiedades específicas, por ejemplo, el sabor, aroma, aspecto y cualquier otra cualidad específica relacionada con el producto, siempre que el producto se haya almacenado adecuadamente y el envase esté sin abrir. La fecha de caducidad está relacionada con la seguridad y es la fecha en la que después un producto alimenticio puede no ser seguro.

Es por tanto que la "vida útil" de un alimento puede definirse como un período de tiempo finito tras la producción (en algunos casos tras la maduración o el envejecimiento) y el envasado durante el cual el producto alimentario conserva un nivel de calidad requerido en condiciones de almacenamiento bien definidas. Este nivel de calidad requerido permite que el producto sea aceptable para el consumo. Esta definición implica que para cualquier tipo de producto alimenticio debe haber un nivel de calidad definido que distinga los productos que todavía son aceptables para el consumo de los que ya no lo son. Este nivel de calidad se define como límite de aceptabilidad y el tiempo necesario para alcanzar el nivel de calidad correspondiente al límite de aceptabilidad se conoce como "vida útil primaria". Existe también lo que se denomina "vida útil secundaria", que se define como el período posterior a la apertura del envasado de un alimento durante el cual el producto mantiene el nivel de calidad aceptable. Implica que para cualquier tipo de producto

15. Reglamento (UE) Nº 1169/2011 del Parlamento Europeo y del Consejo sobre la información alimentaria facilitada al consumidor y por el que se modifican los Reglamentos (CE) Nº 1924/2006 y (CE) Nº 1925/2006 del Parlamento Europeo y del Consejo, y por el que se derogan la Directiva 87/250/CEE de la Comisión, la Directiva 90/496/CEE del Consejo, la Directiva 1999/10/CE de la Comisión, la Directiva 2000/13/CE del Parlamento Europeo y del Consejo, las Directivas 2002/67/CE, y 2008/5/CE de la Comisión, y el Reglamento (CE) Nº 608/2004 de la Comisión, 2011.

alimenticio debe haber un nivel de calidad definido que distinga los productos que todavía son aceptables para el consumo de los que ya no son aceptables.

En términos generales, cuando nos referimos a estabilidad de los alimentos esta puede incluir la estabilidad química (por ejemplo, la estabilidad oxidativa), la estabilidad física y la estabilidad microbiana. En la industria de los aceites y alimentos con alto contenido graso el concepto de estabilidad de una grasa o aceite generalmente se acepta como la vida útil de almacenamiento del producto hasta que la rancidez se hace evidente.

Hay tres vías de reacción principales responsables de la oxidación de lípidos en los alimentos: autooxidación, fotooxidación y oxidación enzimática. Los lípidos son susceptibles de sufrir procesos oxidativos en presencia de sistemas catalíticos como la luz, el calor, las enzimas, los metales, las metaloproteínas y los microorganismos, dando lugar al desarrollo de sabores extraños y a la pérdida de aminoácidos esenciales, vitaminas liposolubles y otros compuestos bioactivos. La degradación oxidativa, generalmente por autooxidación, también denominada enranciamiento, suele ser la principal preocupación ya que es la principal causante de los defectos sensoriales, aunque también pueden ocurrir simultáneamente otros tipos de deterioro y hacer que el problema sea más complejo. Aun así, la estabilidad oxidativa se refiere a la capacidad de los aceites/grasas y alimentos que contienen aceites/grasas para resistir el enranciamiento oxidativo (o deterioro) a lo largo de los periodos de procesamiento y almacenamiento.

Una mejor comprensión de los procesos de oxidación de los lípidos es especialmente útil para controlar estos procesos y desarrollar nuevos productos alimentarios que contengan aceites o grasas con una estabilidad oxidativa y una vida útil aceptables [**16**].

I.3.1. Mecanismos de oxidación lipídica

Los lípidos son los macroconstituyentes menos estables de los alimentos. La oxidación de los lípidos provoca pérdidas nutricionales y produce sabores, colores y compuestos indeseables que hacen que los alimentos sean menos aceptables o inaceptables para los consumidores.

Teniendo en cuenta que los triacilgliceroles (o triglicéridos), que son ésteres de glicerol con ácidos grasos, pueden constituir hasta el 99% de los lípidos de origen vegetal y animal, la oxidación de los lípidos se produce cuando los dobles enlaces de un ácido graso son atacados por oxígeno, hidrógeno y

16. Shahidi, F., Zhong, Y. Methods for Measuring Lipid Oxidation. In: F. Shahidi (Ed.), Bailey's Industrial Oil and Fat Products, 2nd ed., John Wiley & Sons, Nueva Jersey, 2020.

enzimas. Los mecanismos generales de los procesos oxidativos implican tres pasos de autooxidación: (1) iniciación, (2) propagación y (3) finalización. Cualquier discusión sobre la oxidación de lípidos debe comenzar considerando la reacción en cadena clásica de radicales libres. Los ácidos grasos insaturados (R-H) pierden un átomo de hidrógeno para formar un radical alquilo (R·) que reacciona rápidamente con el oxígeno molecular (O_2) para dar lugar a un radical peróxido (ROO·). El radical peróxido reacciona entonces con otro R-H para formar hidroperóxido (ROOH), un producto de oxidación primario. El hidroperóxido es insípido e inodoro y no tiene un impacto significativo en la calidad sensorial de los aceites/grasas. Sin embargo, es generalmente inestable y puede reaccionar con los iones hierro (II) (Fe^{2+}) para dar lugar a un radical alcoxilo lipídico (RO·), lo que conduce a la formación de productos de oxidación lipídica secundarios por β-escisión, como aldehídos (por ejemplo, hexanal y propanal), cetonas, alcoholes, ácidos orgánicos (por ejemplo, ácido hexanoico y propanoico), epóxidos e hidrocarburos. Estos productos de oxidación secundarios aportarían notas de sabor que afectan significativamente a la calidad sensorial de los aceites/grasas y de los alimentos que los contienen [**17**].

La fotooxidación se produce cuando un aceite o grasa se exponen a la luz en presencia de sensibilizadores como las clorofilas, feofitinas o riboflavinas que, activadas por la luz, reaccionan con el oxígeno para producir oxígeno singlete reactivo. De esta forma, el proceso de oxidación en presencia de luz ocurre muy rápido, sin período de inducción o iniciación. La tercera vía de inicio de los procesos oxidativos es la oxidación enzimática donde enzimas como las lipooxigenasas pueden incorporar el oxígeno a los ácidos grasos insaturados para iniciar de nuevo la fase de propagación. La Figura 1 muestra un esquema de autooxidación lipídica en tres etapas: iniciación, propagación y terminación.

Los aceites y alimentos a base de aceite/grasa son en su mayoría alimentos microbiológicamente estables, que almacenados en unas determinadas condiciones experimentan reacciones oxidativas y cuya vida útil depende de la compleja interacción entre factores intrínsecos (composición, propiedades físicas, antioxidantes) y factores extrínsecos (temperatura, material y condiciones de envasado, luz, humedad).

17. Logan, A., Nienaber, U. Xiangqing, S.P. Lipid Oxidation Challenges in Food Systems, AOCS Press, Illinois, 2013.

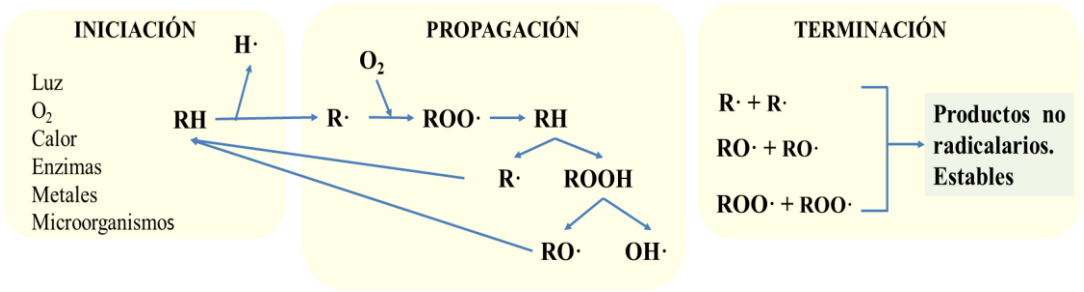


Figura 1. Esquema del proceso de oxidación lipídica representado en tres etapas, iniciación, propagación y terminación.

I.3.2. Estudios de vida útil de aceites y alimentos con alto contenido graso.

En general, la monitorización de cualquier reacción se establece en el contexto de reacciones sencillas con productos fijos estables que se forman en altas concentraciones; por ejemplo:

$$A + B \rightarrow C + D$$

Esta situación difiere directamente de los mecanismos de oxidación de lípidos, que tiene múltiples vías posibles que cambian según las condiciones y el tiempo. En estas reacciones los productos son en realidad intermediarios (productos de oxidación primaria) que se degradan o transforman en otros compuestos, con rendimientos de los productos bastante bajos. Por ello, a fecha de hoy, monitorizar e interpretar la oxidación de lípidos se considera uno de los mayores desafíos analíticos tanto para la ciencia como la industria de los alimentos.

Una cuestión que persiste en el análisis de la oxidación de lípidos es decidir qué compuesto o compuestos químicos es o son los más adecuados para monitorizar – son los denominados marcadores oxidativos. La respuesta es difícil de encontrar y está fuertemente influenciada por el aspecto cuantitativo o cualitativo del estudio llevado a cabo en cuestión.

Llegados a este punto, debe quedar clara la diferencia existente entre: (1) un "estudio de estabilidad", dirigido a medir la cinética de agotamiento o el grado de deterioro con respecto a la calidad preestablecida, y (2) un "estudio de vida útil", destinado a realizar una estimación de un período finito de tiempo tras la producción (y en algunos casos, la maduración) y el envasado durante el cual el producto alimentario seguirá siendo seguro, conservará las características sensoriales, químicas, físicas y microbiológicas deseadas y cumplirá con cualquier declaración nutricional o de cualquier otra índole, cuando se almacena en las condiciones recomendadas. Por tanto, el objetivo de un estudio de vida útil es estimar un valor de tiempo derivado de la fusión de la

pérdida de calidad con el límite de aceptabilidad del consumidor. Para alcanzar este objetivo se requiere un enfoque multidisciplinario que explica la complejidad y el desafío que esta tarea supone. En otras palabras, los estudios de estabilidad se centran en las tasas de degradación mientras que los estudios de vida útil en los tiempos.

Aunque existen en literatura numerosos estudios que tratan la estabilidad de aceites, o alimentos de alto contenido graso, y utilizan el término vida útil, en su mayoría estos estudios no permiten la estimación de un tiempo de vida debido a la falta de una definición clara del límite de aceptabilidad o tiempo crítico a partir del cual el alimento deja de cumplir los estándares de calidad/seguridad esperados para su consumo [**18**,**19**,**20**]. La razón probablemente radique en que si bien, se pueden obtener datos precisos y confiables de la tasa de deterioro de la calidad mediante la aplicación de procedimientos de modelado adecuados, por otro lado, la identificación del límite de aceptabilidad todavía se lleva a cabo empíricamente y frustra cualquier intento por transformar la estimación de vida útil en una tarea con base científica.

Tal y como se refirió con anterioridad según el Reglamento (UE) 1169/2011, los productores de aceites y alimentos con alto contenido graso adquieren la responsabilidad de la datación de la fecha preferente de consumo en base a estudios de vida útil apropiados. Además, estándares como BRC Food Safety v8, IFS Food v7, ISO 22000:2018 han descrito requisitos aún más restrictivos para garantizar calidad y seguridad alimentaria, entre otros, estos estándares requieren que se adopten protocolos documentados para la definición de la vida útil, así como la implementación de un sistema de evaluación continua con pruebas y registros de la misma.

Un estudio de vida útil según el enfoque clásico se articula en tres etapas: una etapa preliminar, una etapa de análisis y obtención de datos experimentales, y el modelado final de los datos. La etapa preliminar implica la identificación del indicador o descriptor de deterioro oxidativo que se considera que es el parámetro analítico que tiene el impacto más crítico en la calidad del alimento,

---

18. Aktar, T., Adal, E. Determining the Arrhenius kinetics of avocado oil: oxidative stability under Rancimat test conditions. Foods, 8, 236 (2019), 1-13. https://doi.org/10.3390/foods8070236
19. Alonso Salces, R.M., Gallo, B., Collado, M.I., Sasía Arriba, A., Viacava, G.E., García González, D.L., Toschi, T.G., Servili, M., Berrueta, L.Á. ^{1}H-NMR fingerprinting and supervised pattern recognition to evaluate the stability of virgin olive oil during storage. Food Control, 123, 107831 (2021), 1-13. https://doi.org/10.1016/j.foodcont.2020.107831
20. Buratti, S., Malegori, C., Benedetti, S., Oliveri, P., Giovanelli, G. E-nose, e-tongue and e-eye for edible olive oil characterization and shelf life assessment: a powerful data fusion approach. Talanta, 182 (2019), 131-141. https://doi.org/10.1016/j.talanta.2018.01.096

así como su límite de aceptabilidad en unas determinadas condiciones de almacenamiento. Claramente éste es un enfoque univariable en el que se asume que el comportamiento de uno/varios descriptores oxidativos, y una correlación sencilla de estos, van a informar en todo momento del estado global del alimento. En esta etapa se suelen realizar pruebas preliminares de estabilidad para tener confianza en el comportamiento del producto y seleccionar el indicador crítico que mejor se adapte al uso. Dependiendo del principal evento de degradación responsable se han utilizado indicadores de deterioro de la calidad de los alimentos, químicos, físicos, microbiológicos o sensoriales. La búsqueda del indicador crítico está indisolublemente ligada a la identificación del límite de aceptabilidad, el parámetro más difícil de definir al desarrollar un estudio de vida útil. La elección del límite de aceptabilidad puede derivar de requisitos legales o voluntarios. Hay algunos ejemplos de límites de aceptabilidad obligatorios como en el caso de los aceites de oliva virgen [**21**]. En la mayoría de los casos, el límite de aceptabilidad es voluntario, y su elección es de los productores, quienes son libres de identificar este valor de acuerdo con su propia política y objetivos de calidad. En la actualidad, el límite de aceptabilidad se identifica con frecuencia empíricamente, pero son posibles enfoques más racionales para su identificación.

La segunda etapa consiste en la monitorización del alimento bajo condiciones simuladas de almacenamiento. Es la etapa donde se identifica lo que realmente sucede con el producto y se recopilan los datos experimentales midiendo los cambios del descriptor oxidativo a lo largo del tiempo bajo condiciones reales o aceleradas de envejecimiento. El enfoque más lento implica la monitorización bajo condiciones ambientales (es decir, temperatura, exposición a la luz, y humedad relativa) que imitan las condiciones reales de almacenamiento esperadas, elegidas cuidadosamente para simular de forma realista su almacenamiento y que deben mantenerse constantes hasta el final (estudios de vida útil en condiciones reales). Este enfoque permite la estimación directa de la vida útil del alimento, pero solo puede utilizarse cuando el deterioro de la calidad se produce relativamente rápido y cuando los cambios en las condiciones ambientales durante el almacenamiento se consideran insignificantes. Si por el contrario el deterior de la calidad del alimento avanza con bastante lentitud en condiciones reales de almacenamiento, es común intentar acelerar los estudios mediante el seguimiento bajo condiciones atmosféricas forzadas, capaces de acelerar la pérdida de calidad del alimento (estudios de vida útil bajo condiciones aceleradas). Este procedimiento permite

21. Reglamento de ejecución (UE) 2019/1604 de la Comisión por el que se modifica 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, 2019.

hacer una predicción de la vida útil extrapolando los resultados a las condiciones más suaves que suele experimentar el producto.

En la tercera y última etapa los datos que describen la evolución del indicador oxidativo en función del tiempo se someten a un modelado matemático según los principios cinéticos fundamentales para obtener una estimación o predicción de la vida útil. En general, los modelos intentan formular el comportamiento de los sistemas a partir del conocimiento de las propiedades de sus componentes. Se trata siempre de simplificaciones de la realidad, diseñadas para facilitar las predicciones y los cálculos. La modelización es un intento de aproximación a la realidad, siempre teniendo en cuenta las oportunidades y limitaciones que presentan. Esto se aplica, por supuesto, también a la modelización cinética de las reacciones en los alimentos.

I.3.3. Cinética de degradación

La cinética es el estudio de la velocidad de una reacción (ecuación 1.1), generalmente tomada como el cambio en la concentración de *C* a lo largo del tiempo y su dependencia respecto a la concentración de reactivos, temperatura, presencia de catalizadores y condiciones ambientales:

$$\mathrm{v} = \frac{\mathrm{dC}}{\mathrm{dt}} = \mathrm{k}\,\mathrm{C}^{\mathrm{n}} \qquad (1.1)$$

donde *k* es la constante de velocidad, *t* es el tiempo de almacenamiento y *n* es el orden de reacción. La velocidad de las reacciones químicas es un determinante importante de los cambios en la calidad de los alimentos y la vida útil. Por consiguiente, la tasa de cambios de un indicador de la estabilidad (*I*) en función del tiempo de almacenamiento (*t*) puede calcularse mediante la integración de la ecuación cinética general (ecuación 1.2):

$$\int_{\mathrm{I_o}}^{\mathrm{I}} \frac{\mathrm{dI}}{\mathrm{I^n}} = \int_0^{\mathrm{t}} \mathrm{k}\,\mathrm{dt} \qquad (1.2)$$

Dado que el orden de reacción no da ninguna indicación de los mecanismos reales de la reacción implicada en el agotamiento de la calidad de los alimentos, *k* se considera una constante "aparente" de velocidad.

La elección del orden de reacción de la ley de velocidad se realiza generalmente basándose en la selección del modelo que mejor se ajusta a los valores experimentales monitorizados para *I*. Los órdenes cero, primer orden y segundo orden son los más utilizados para describir el agotamiento de la calidad de los alimentos. En estos casos, la constante de velocidad se calcula mediante un análisis de regresión lineal de *I*, *ln I* o *1/I* en función del tiempo de almacenamiento. La elección del orden de reacción se apoya generalmente en un análisis estadístico basado en la comparación de los valores del coeficiente

de determinación de las regresiones (R^2), así como en un análisis de los residuos de la regresión [**22**,**23**].

Tras este punto, es necesario tener definido el límite de aceptabilidad y, una vez calculada *k*, se estima la vida útil bajo las condiciones de almacenamiento seleccionadas en el estudio integrando la siguiente ecuación (1.3) en función del tiempo:

$$t_{vu} = \frac{1}{k} \int_{I_o}^{I_{lim}} \frac{dI}{I^n} \qquad (1.3)$$

donde t_{vu} representa tiempo de vida útil, I_o representa el valor del descriptor oxidativo al inicio del estudio, justo después de su producción y envasado, mientras que I_{lim} representa el límite de aceptabilidad del mismo. En el esquema de la Figura 2 se muestra una representación del comportamiento del indicador *I* en función del tiempo considerando los órdenes de reacción más habituales, así como las ecuaciones para el cálculo de tiempo de vida útil resultantes tras la integración [**24**].

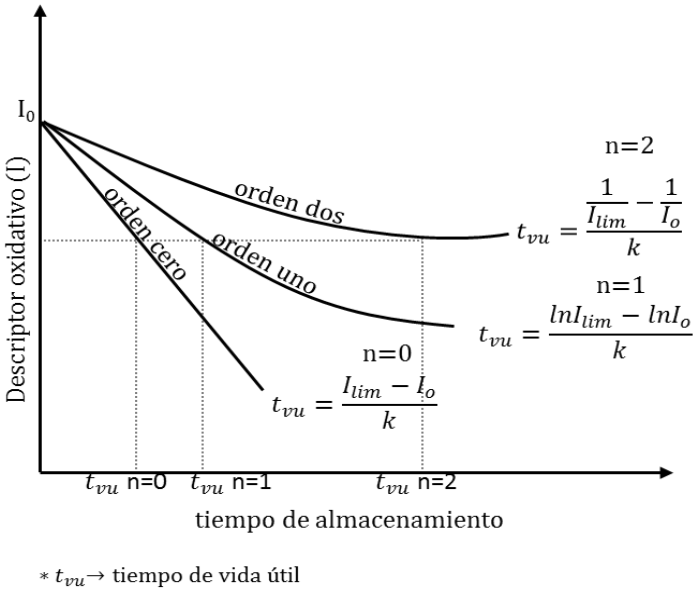


Figura 2. Comportamiento del indicador oxidativo I en función del tiempo para las reacciones que siguen una cinética de orden cero, uno y dos, así como las ecuaciones para la estimación de t_{vu} correspondientes.

Desgraciadamente, pueden observarse situaciones más complejas para los índices oxidativos que conducen a la inaplicabilidad del enfoque cinético clásico. Un ejemplo típico es la evolución de los descriptores que muestran una

22. Van Boekel, M.A.J.S. Kinetic Modelling of Reactions in Foods. CRC Press Taylor and Francis Group, New York, 2009.
23. Hu, M., Jacobsen, C. Oxidative Stability and Shelf Life of Foods Containing Oils and Fats. AOCS Press, United Kingdom, 2016.
24. Galanakis, C.M. Food Quality and Shelf Life. Academic Press, United Kingdom, 2019.

evolución sigmoidal, presentando un período de inducción (retardo) tras el cual, se produce un aumento progresivo hasta alcanzar un máximo, lo que indica que la reacción oxidativa se acerca a la terminación. En estos casos, los datos son modelados empleado modelos matemáticos totalmente empíricos, distintos de los clásicos, dando lugar a estimaciones fuertemente dependientes de las condiciones experimentales (tanto externas como intrínsecas (antioxidantes naturales, temperatura, composición inicial, luz, etc.)) [**25**,**26**,**27**,**28**].

En el caso de estudios acelerados, cuando la oxidación avanza lentamente en condiciones reales de almacenamiento, los datos son extrapolados para calcular la cinética en las condiciones que normalmente experimenta el producto. El requisito absolutamente necesario para predecir la vida útil de manera correcta es el conocimiento de la relación entre el factor de aceleración y la tasa de oxidación. Sólo si se cumple este requisito, la extrapolación a las condiciones reales podría utilizarse. La temperatura es, sin duda, el factor de aceleración más utilizado en estos estudios debido a su papel crítico a la hora de afectar a la cinética oxidativa, así como a la disponibilidad de una ecuación matemática teórica, la ecuación de Arrhenius (1.4), que describe la relación de las velocidades de reacción con la temperatura:

$$\mathrm{k} = \mathrm{A}\, \mathrm{e}^{-\frac{\mathrm{Ea}}{\mathrm{RT}}} \qquad (1.4)$$

donde *k* es la constante de velocidad de reacción, *R* es la constante molar de los gases (8,31 J/K mol), *T* es la temperatura absoluta (K), *Ea* es la energía de activación aparente (J/mol) y *A* es denominado factor pre-exponencial. Cuando se cumple, la ecuación de Arrhenius permite extrapolar la velocidad de reacción a una temperatura deseada midiendo la velocidad de oxidación al

25. Psomidou, E., Karakostas, K.X., Blekas, G., Tsimodou, M.Z., Boskou, D. Proposed parameters for monitoring quality of virgin olive oil (Koroneiki cv). Eur. J. Lipid Sci. Technol., 105, 8 (2003), 403-408. https://doi.org/10.1002/ejlt.200300780

26. Zanoni, B., Bertuccioli, M., Rovellini, P., Marotta, F., Mattei, A. A preliminary approach to predictive modelling of extra virgin olive oil stability. J. Sci. Food Agric, 85, 9 (2005), 1492-1498. https://doi.org/10.1002/jsfa.2135

27. Aragao, G.M.F., Corradini, M.G., Peleg, M. A phenomenological model of the peroxide value's rise and fall during lipid oxidation. J. Am. Oil Chem. Soc., 85 (2009), 1143-1153. https://doi.org/10.1007/s11746-008-1305-z

28. Aparicio Ruiz, R., Roca, M., Gandul Rojas, B. Mathematical model to predict the formation of pyropheophytin a in virgin olive oil during storage. J. Agric. Food Chem., 60, 28 (2012), 7040-7049, 2012. https://doi.org/10.1021/jf3010965

menos a tres temperaturas diferentes [**29**]. La ecuación de Arrhenius se ha utilizado en numerosos estudios para estimar la dependencia de la temperatura de la velocidad de degradación de aceites y alimentos de alto contenido graso recogiendo datos para varias condiciones de almacenamiento y determinando una constante de proporcionalidad entre las diferentes temperaturas que permite predecir la vida útil en condiciones ambientales de almacenamiento [**30**,**31**].

Aunque a priori parece una técnica sencilla y adecuada para establecer el agotamiento de la calidad, en la práctica, aparecen numerosas desviaciones de la ecuación de Arrhenius que podrían causar sobreestimación en la predicción de vida útil. La extrapolación de las tasas de reacción a las temperaturas habituales de almacenamiento a partir de datos acelerados debe realizarse sólo dentro del rango de temperaturas que se ha demostrado experimentalmente que se ajusta al modelo de Arrhenius. En otras palabras, la metodología de Arrhenius requiere una adaptación a las circunstancias específicas del producto considerado [**32**,**33**].

I.4. Técnicas y metodologías analíticas

El análisis de lípidos comprende una amplia gama de enfoques, técnicas y métodos que sirven para fines totalmente diferentes condicionados, en gran parte, por las exigencias comerciales. El progreso en la utilización de grasas y aceites para la producción de productos alimenticios útiles, depende de un conocimiento profundo de las características de las materias primas, de los cambios efectuados por cada proceso y de los requisitos del producto alimenticio preparado individualmente. Los análisis físicos, químicos y de

---

29. Van Boekel, M.A.J.S. Kinetic modelling of food quality: a critical review. Compr. Rev. Food Sci. Food Saf., 7, 1 (2008), 144-158. https://doi.org/10.1111/j.1541-4337.2007.00036.x
30. Rodríguez, G., Villanueva, E., Glorio, P., Baquerizo, M. Estabilidad oxidativa y estimación de la vida útil del aceite de sacha inchi (*Plukenetia volubilis L.*), Sci. Agropecu., 6, 3 (2015), 155-163. https://doi.org/10.17268/sci.agropecu.2015.03.02
31. Conte, L., Milani, A., Calligaris, S., Rovellini, P., Lucci, P., Nicoli, M.C. Temperature dependence of oxidation kinetics of extra virgin olive oil (EVOO) and shelf-life prediction. Foods, 9, 295 (2020), 1-13. https://doi.org/10.3390/foods9030295
32. Calligaris, S., Manzocco, L., Conte, L.S., Nicoli, M.C. Application of a modified Arrhenius equation for the evaluation of oxidation rate of sunflower oil at subzero temperatures. J. Food Sci., 69 (2006), 361-366. https://doi.org/10.1111/j.1365-2621.2004.tb09896.x
33. Syed, A. Oxidative stability and shelf life of vegetable oils. In: M. Hu, C. Jacobsen (Eds.), Oxidative Stability and Shelf Life of Foods Containing Oils and Fats. Academic Press, United Kingdom, (187-207), 2016.

rendimiento son las herramientas de las que disponen los procesadores de grasas y aceites para la compra de materias primas, el desarrollo de nuevos productos y la evaluación de los productos elaborados por medio del conocimiento y control de su composición, sus propiedades estructurales y funcionales, las reacciones esperadas; y controlar así el etiquetado mediante la aplicación de la investigación científica. Por ejemplo, los listados nutricionales de ácidos grasos saturados, trans-, poliinsaturados, colesterol, vitaminas y otras características del producto exigidos en el etiquetado de los productos alimenticios requieren un análisis exacto para identificar los valores originales y garantizar el cumplimiento de los requisitos legales. Análisis de propiedades generales de la muestra como el contenido de humedad, grado de acidez o estado de deterioro oxidativo (OSI) son prácticas comunes. También son importantes los análisis para detectar trazas de componentes, como plaguicidas u oligoelementos, para garantizar el cumplimiento de la normativa gubernamental relativa a estos compuestos. Además, para que los productos se comercialicen a nivel internacional, los procedimientos de análisis deben ser robustos y ampliamente reconocidos. Organizaciones como las que se enumeran a continuación persiguen la estandarización de métodos, para lo que desarrollan procedimientos analíticos, que aunque ofrecen orientaciones similares, no son idénticas [**34**]:

- *AOAC International. Association of Analytical Communities*
- *AOCS. The American Oil Chemists' Society*
- *BSI. The British Standards Institution*
- *ISO. The International Organisation for Standardisation*
- *IUPAC. The International Union of Pure and Applied Chemists*

Previo a la realización de cualquier medición, es necesario obtener una muestra representativa y transportarla sin que sufra alteraciones que modifiquen el parámetro a determinar. Existen procedimientos estándar para todas estas etapas [**35**,**36**]. Hay que prestar atención a la temperatura de almacenamiento, la naturaleza del contenedor, la inhibición de la actividad enzimática y la posible adición de antioxidantes. Si no se consideran todas estas cuestiones, incluso el análisis más robusto dará lugar a resultados y conclusiones poco fiables.

---

34. Gunstone, F.D. Chemistry Oils Fats. Blackwell Publishing, United Kingdom, 2004.
35. UNE EN-ISO 661:2006. Aceites y grasas de origen animal y vegetal. Preparación de la muestra para análisis.
36. UNE EN- ISO 5555:2002.Aceites y grasas de origen animal y vegetal. Toma de muestras.

La diversidad de productos y aplicaciones, además de los diversos procesos industriales a los que alimentos con alto contenido graso pueden ser sometidos (destinados a consumo en crudo, refinamiento, hidrogenación, etc.) hacen que sea necesario establecer requisitos analíticos individuales en términos de propiedades químicas, físicas, de consistencia, nutricionales, de rendimiento y de otro tipo, que deben medirse y mantenerse para garantizar que los productos funcionen como se han diseñado. Además, aunque se dispone de avanzados métodos de evaluación analítica, solo unos pocos se utilizan en la industria alimentaria, ya que cada uno tiene sus propias limitaciones. Algunos pueden ser punteros en investigación, pero éstos pueden no ser adecuados para el análisis de rutina en la industria alimentaria por diferentes motivos: complejidad de uso, limitaciones instrumentales en los laboratorios de rutina, coste económico del ensayo, etc. Por ejemplo, los métodos ESR (resonancia de espín electrónico), NMR (resonancia magnética nuclear), DSC (calorimetría diferencial de barrido), quimioluminiscencia y fluorescencia se utilizan ampliamente en investigación académica, pero no como análisis de rutina en las industrias de alimentos.

Por ello, en este capítulo nos centraremos en las técnicas y metodologías analíticas más empleadas actualmente en el control de la autenticidad, calidad y estabilidad de aceites y alimentos vegetales con alto contenido graso.

I.4.1. Técnicas espectrométricas

La espectroscopia se ha destacado como una potencial técnica analítica para la caracterización (y detección de prácticas fraudulentas) de aceites y grasas por su reproducibilidad y naturaleza no destructiva. Además, no es necesaria la preparación previa de la muestra ni el uso de reactivos tóxicos, lo que abarata su coste. En este sentido, la espectroscopia en el UV-VIS se basa en la medida de absorbancias de los compuestos en el rango de 195 a 780 nm. Estas absorbancias son debidas a las ganancias energéticas de los electrones de valencia de las moléculas. Así, se realiza la determinación de extinción específica (valores k de los aceites de oliva) a 232 y 268 nm, o la medida de la absorbancia a 410, 450 y 470 nm para la determinación de los carotenoides y a 660 nm para los compuestos ricos en clorofila.

Por otro lado, la espectrofluorimetría emplea el mismo rango de longitudes de onda que en espectroscopía UV-Vis pero en este caso para medir la excitación de las moléculas. Las moléculas excitadas emiten radiación electromagnética, denominada fluorescencia. A partir de la medida de la intensidad de esa fluorescencia (ya sea de excitación o de emisión) podemos poder obtener información de los compuestos. Como ejemplo podemos destacar: la florescencia de tocoferoles a 300-390 nm, el estudio de productos de oxidación

de los ácidos grasos a 445-475 nm, determinación de vitamina E fluorescente a 525 nm y la determinación de productos de degradación de clorofila a y b (681 nm). La espectrofluorimetría se puede emplear como medida directa [**37**,**38**], o bien como sistema de detección tras una separación cromatográfica [**39**], para evaluar tanto autenticidad como calidad/estabilidad de alimentos con alto contenido graso.

La espectroscopia vibracional que agrupa a su vez las espectroscopías del infrarrojo cercano (NIR), infrarrojo medio (MIR) y Raman. La espectroscopía NIR utiliza la región desde el extremo de las longitudes más altas del visible (750 nm) hasta los 2500 nm del espectro electromagnético (12500-4000 cm^{-1}) y se basa en las vibraciones moleculares de sobretono y combinación. La espectroscopia MIR (2500-25000 nm, o 4000-400 cm^{-1}) está vinculada a la espectroscopia vibracional y en los efectos rotacionales-vibracionales asociados. Por último, la espectroscopia Raman se basa en la dispersión inelástica de una luz monocromática, que interactúa con las vibraciones moleculares, los fotones u otras excitaciones de la muestra, dando lugar a un desplazamiento de las frecuencias de energía. El desplazamiento de las frecuencias energéticas proporciona información sobre los modos vibracionales de la muestra [**40**]. En la actualidad, la espectroscopia NIR, Raman y, principalmente, MIR se han empleado para desarrollar métodos rápidos y sencillos para detectar la adulteración y para evaluar la autenticidad y los parámetros de calidad intrínsecos de grasas y aceites comestibles.

---

37. Domenici, V., Ancora, D., Cifelli, M., Serani, A., Veracini, C.A., Zandomeneghi, M. Extraction of pigment information from near-UV Vis absorption spectra of extra virgin olive oils. J. Agric. Food Chem., 62 (2014), 9317-9325. https://doi.org/10.1021/jf503818k

38. Kyriakidis, N.B., Skarkalis, P. Fluorescence spectra measurement of olive oil and other vegetable oils. J. AOAC Int., 83, 6 (2000), 1435-1439. https://doi.org/10.1093/jaoac/83.6.1435

39. Lucci, P., Bertoz, V., Pacetti, D., Moret, S., Lanfranco C. Effect of the refining process on total hydroxytyrosol, tyrosol, and tocopherol contents of olive oil. Foods, 9 (3), 292 (2020), 1-11. https://doi.org/10.3390/foods9030292

40. Nunes, C.A. Vibrational spectroscopy and chemometrics to assess authenticity, adulteration and intrinsic quality parameters of edible oils and fats. Int. Food Res. J., 60 (2014), 255-261. https://doi.org/10.1016/j.foodres.2013.08.041

Estos métodos se apoyan generalmente en un enfoque multivariable y requieren la aplicación de técnicas quimiométricas [**41**,**42 43**].

I.4.2. Técnicas cromatográficas

La mayoría de los métodos analíticos para la determinación e identificación de lípidos se basan actualmente en la potente capacidad de separación de las técnicas cromatográficas. La cromatografía es una técnica muy útil para separar, purificar e identificar los diferentes componentes de una mezcla, empleando para ello una columna de separación cromatográfica [**44, 45**]. El concepto de separar los componentes de la muestra en una columna fue desarrollado por primera vez en 1903 por Mikhail Tswett, quien introdujo el término cromatografía en 1906, por su logro de separar pigmentos o colores [**46**]. El desarrollo hasta la actualidad de la técnica ha dado lugar a sofisticados equipos que permiten la separación de una amplia variedad de mezclas, pero el principio en el que se basa sigue siendo el mismo.

En una separación cromatográfica, un pequeño volumen de muestra es introducido a la entrada de una columna. En el interior de la columna se encuentra la fase estacionaria. Una fase móvil transporta los componentes de la muestra a lo largo de la columna, donde debido a las diferentes interacciones que se producen entre los componentes de la muestra y las dos fases, estos migran a través del sistema a diferentes velocidades y eluyen de la columna con diferentes tiempos de retención. Al final de la columna, un dispositivo de

---

41. Li, Q., Chen, J., Huyan, Z., Kou, Y., Xu, L., Yu, X., Gao J.M. Application of Fourier transform infrared spectroscopy for the quality and safety analysis of fats and oils: a review. Crit. Rev. Food Sci. Nutr., 59, 22 (2019), 3597-3611. https://doi.org/10.1080/10408398.2018.1500441
42. Wang, K., Li, Z., Li, J., Lin, H. Raman spectroscopic techniques for nondestructive analysis of agri-foods: a state-of-the-art review. Trends Food Sci. Technol., 118 (2021), 490-504. https://doi.org/10.1016/j.tifs.2021.10.010
43. Su, W.H., Sun, D.W. Mid-infrared (MIR) spectroscopy for quality analysis of liquid foods. Food Eng. Rev., 11 (2019), 142-158. https://doi.org/10.1007/s12393-019-09191-2
44. Esteki, M., Simal Gandara, J., Shahsavari, Z., Zandbaaf, S., Dashtaki, E., Vander Heyden, Y. A review on the application of chromatographic methods, coupled to chemometrics, for food authentication. Food Control, 93 (2018), 165-182. https://doi.org/10.1016/j.foodcont.2018.06.015
45. 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 chromatography fingerprint of the methyl transesterified fraction. Food Chem., 221 (2017), 1784-1791. https://doi.org/10.1016/j.foodcont.2018.06.015
46. Lundanes, E., Reubsaet, L., Greibrokk, T. Chromatography. Basis Principles, Sample Preparations and Related Methods. Wiley-VCH, Germany, 2014.

medida (detector) proporciona una señal para todos los componentes que eluyen o para un número limitado de ellos dependiendo de si se trata de un detector universal o un detector selectivo. Los sistemas de medida más ampliamente utilizados en los análisis cromatográficos de aceites son, con diferencia: detectores espectrofotométricos que miden absorbancia en la zona UV-visible, como el detector de absorción UV-Vis mediante fila de diodos (DAD) [**47**]; detectores que miden fluorescencia (FD) [**48**]; y detectores universales como el detector de ionización de llama (FID) [**49**] o de espectrometría de masas (MS) [**50**].

Además, se pueden diferenciar las tres modalidades cromatográficas de mayor interés que son la cromatografía en capa fina (TLC), la cromatografía de líquidos de altas prestaciones (HPLC) y la cromatografía de gases (GC). En TLC la fase estacionaria, generalmente de sílice, está colocada sobre una superficie plana (placa de vidrio, metal o plástico). La mezcla que se va a separar se coloca al borde de la placa. La fase móvil o disolvente asciende por la placa mediante la atracción capilar y transporta los distintos componentes de la mezcla a diferentes alturas, en función de la polaridad, u otros mecanismos, y así se consigue la separación. La técnica de cromatografía en capa fina ha sido utilizada en métodos oficiales tanto para aplicar un fraccionamiento previo, p.e., de esteroles y estanoles en aceites [**51**], o para la determinación de compuestos como los ácidos trans-octadecanoicos, productos de oxidación o fosfolípidos de aceites, grasas y derivados [**52**]. Además, la TLC ha sido

---

47. Tasioula Margari, M., Tsabolatidou, E. Extraction, separation, and identification of phenolic compounds in virgin olive oil by HPLC-DAD and HPLC-MS. Antioxidants, 4, 3 (2015), 548-562. https://doi.or/ 10.3390/antiox4030548
48. Huyan, Z., Ding, S., Liu, X., Yu, X. Authentication and adulteration detection of peanut oils of three flavor types using synchronous fluorescence spectroscopy. Anal. Methods, 10 (2018), 3207-3214. https://doi.org/10.1039/C8AY00837J
49. Siano, F., Vasca, E. GC-FID analysis to evaluate the possible adulteration of extra virginolive oil with different vegetable oils. J. Chem. Educ., 97, 11 (2020), 4108-4116. https://doi.org/10.1021/acs.jchemed.0c00278
50. Alberdi Cedeño, J., Ibargoitia, M.L., Guillén, M.D. Monitoring of minor compounds in corn oil oxidation by direct immersion-solid phase microextraction-gas chromatography/mass spectrometry. New oil oxidation markers. Food Chem., 290 (2019), 286-294. https://doi.org/10.1016/j.foodchem.2019.04.001
51. International Olive Council. Determination of the sterol composition and content and alcoholic compounds by capillary gas chromatography. COI/T.20/ Doc. No 26/Rev. 4, 2018.
52. IUPAC 2.642. Standard Methods for the Analysis of Oils, Fats and Derivatives, 7th ed, Oxford Blackwell Scientific Publication, London, 1992.

aplicada, de manera cualitativa y cuantitativa, con fines de autenticidad de aceites comestibles [**53**].

HPLC consiste en una separación empleando mezclas de disolvente como fase móvil y micropartículas de sílice o sílice ligadas a grupos alquilos (los más habituales C18) como fase estacionaria. Podemos diferenciar a su vez HPLC de fase normal que ocurre cuando la fase estacionaria es polar y por tanto la fase móvil apolar, y HPLC de fase invertida (normalmente nombrada como fase inversa) referida a columnas no polares y mezclas de disolventes polares. La diversidad de aplicaciones de HPLC tanto en fase normal como, mayoritariamente, en fase invertida en el análisis de aceites vegetales es incontable. Proporciona uno de los mejores métodos para analizar mezclas de triglicéridos, componentes mayoritarios de los aceites, por lo que se pueden encontrar numerosas aplicaciones en términos de autenticidad [**54**] y detección de adulteraciones [**55**]. Además, esta técnica se emplea para la determinación de compuestos que influyen directamente en la estabilidad de aceites y alimentos con alto contenido graso como determinación del contenido en antioxidantes [**56**], compuestos fenólicos [**57**], tocoferoles y tocotrienoles [**58**,**59**], compuestos polares resultantes de la hidrólisis, oxidación y

---

53. Sherma, J., Rabel, F. A review of thin layer chromatography methods for determination of authenticity of foods and dietary supplements. J. Liq. Chromatogr. Relat. Technol., 41, 10 (2018), 645-657. https://doi.org/10.1080/10826076.2018.1505637
54. Obisesan, K.A, Jiménez Carvelo, A.M., Cuadros Rodríguez, L., Ruisánchez, I., Callao, M.P. HPLC-UV and HPLC-CAD chromatographic data fusion for authentication of geographical origin of palm oil. Talanta, 170 (2017), 413-418. https://doi.org/10.1016/j.talanta.2017.04.035
55. Carranco, N., Farrés Cebrián, M., Saurina, J., Núñez, O. Authentication and quantitation of fraud in extra virgin olive oils based on HPLC-UV fingerprinting and multivariate calibration. Foods, 7, 4 (2018), 44. https://doi.org/10.3390/foods7040044
56. Kim, J.M., Choi, S.H., Shin, G.H., Lee, J.H., Kang, S.R., Lee, K.Y., Lim, H.S., Kang, T.S., Lee, O.H. Method validation and measurement uncertainty for the simultaneous determination of synthetic phenolic antioxidants in edible oils commonly consumed in Korea. Food Chem., 213 (2016), 19-25. http://dx.doi.org/10.1016/j.foodchem.2016.06.053
57. International Olive Council. Determination of biophenols in olive oils by HPLC. COI/T.20/Doc. Nº 29, 2009.
58. Spanish Association for Standardisation and Certification (AENOR). Animal and vegetable fats and oils. Determination of tocopherol and tocotrienol contents by high-performance liquid chromatography. UNE-EN ISO 9936, 1-23, 2016.
59. Tanno, R., Kato, S., Shimizu, N., Ito, J., Sato, S., Ogura, Y., Sakaino, M., Sano, T., Eitsuka, T., Kuwahara, S., Miyazawa, T., Nakagawa, K. Analysis of oxidation products of α-tocopherol in extra virgin olive oil using liquid chromatography-tandem mass spectrometry. Food Chem., 306, 125582 (2020), 1-9. https://doi.org/10.1016/j.foodchem.2019.125582

polimerización de triglicéridos [**60**] o colorantes naturales tales como las clorofilas [**61**, **62**].

GC es una técnica eficaz de separación en la que una pequeña cantidad de muestra se volatiliza y se inyecta en una corriente de gas inerte (suele ser nitrógeno, helio o hidrógeno, en orden de mayor poder de resolución) a alta temperatura. Esta corriente de gas atraviesa una columna cromatográfica (normalmente un tubo capilar de sílice fundida en cuya pared interior suele estar recubierta la fase estacionaria, polar o apolar) que separa los componentes de la mezcla. La columna se calienta a una serie de temperaturas, limitadas únicamente por la estabilidad térmica de la fase estacionaria y de los analitos. La elución depende de la temperatura, así que cuanto más baja sea la temperatura, se obtienen normalmente mejores separaciones, por lo que es necesario llegar a un compromiso entre el tiempo de elución y la eficacia de la separación modificando las temperaturas (gradiente). Los componentes separados, emergen de la columna a intervalos discretos de tiempo y son registrados por un sistema de detección adecuado. En el campo del análisis de alimentos de alto contenido graso, se emplea principalmente para separar y cuantificar los ácidos grasos en forma de sus ésteres metílicos [**63**], para caracterizar el perfil de triglicéridos [**64**] y para identificar compuestos volátiles [**65**]. Esto hace que se puedan encontrar numerosas aplicaciones de cromatografía de gases acopladas a diversos

---

60. Khor, Y.P., Hew, K.S., Abas, F., Lai, O.M., Cheong, M.Z., Nehdi, I.A., Sbihi, H.M., Gewik, M.M., Tan, C.P. Oxidation and polymerization of triacylglycerols: in-depth investigations towards the impact of heating profiles. Foods, 8, 475 (2019), 1-15. https://doi.org/10.3390/foods8100475 1-15
61. Spanish Association for Standardisation and Certification (AENOR). Vegetable fats and oils. Determination of the degradation products of chlorophylls a and a' (pheophytins a, a' and pyropheophytins). UNE-EN ISO 29841, 1-15, 2009.
62. Delpino Rius, A., Cosovanu, D., Eras, J., Vilaró, F., Balcells, M., Canela Garayoa, R. A fast and reliable ultrahigh-performance liquid chromatography method to assess the fate of chlorophylls in teas and processed vegetable foodstuff. J. Chromatogr. A, 1568 (2018), 69-79. https://doi.org/10.1016/j.chroma.2018.07.016
63. IUPAC 2.304. Capillary column gas-liquid chromatography of fatty acid methyl esters, 1992.
64. Ruiz Samblás, C., González Casado, A., Cuadros Rodríguez, L. Triacylglycerols determination by high-temperature gas chromatography in the analysis of vegetable oils and foods: a review of the past 10 years. Crit. Rev. Food Sci. Nutr., 55, 11 (2015), 1618-1631. https://doi.org/10.1080/10408398.2012.713045
65. Mota, M.F.S., Waktola, H.D., Nolvachai, Y., Marriott, P.J. Gas chromatography – mass spectrometry for characterisation, assessment of quality and authentication of seed and vegetable oils. Trends in Anal. Chem., 138, 116238 (2021), 1-14. https://doi.org/10.1016/j.trac.2021.116238

detectores con fines de verificación de la autenticidad [**66**] y detección de adulteraciones [**67**], así como en estudios de estabilidad y vida útil [**68**].

Se puede concluir que las técnicas de cromatografía descritas (TLC, HPLC y GC) pueden utilizarse para la separación, purificación y/o análisis de muestras lipídicas complejas. Estas técnicas proporcionan una gran cantidad de datos que ofrecen mucha información al analista. Por lo tanto, el correcto tratamiento quimiométrico de los datos mediante técnicas multivariable de minería de datos como el análisis exploratorio no supervisado, los métodos supervisados de clasificación, o métodos de cuantificación son herramientas útiles para la autentificación de diferentes aceites comestibles.

I.5. Minería de datos (quimiometría): huellas instrumentales

El avance en técnicas analíticas cada vez más sofisticas que proporcionan una gran cantidad de datos hace necesario el uso de herramientas adecuadas de análisis y tratamiento de datos para poder extraer e interpretar la información recogida en los mismos. La minería de datos, por definición, abarca el estudio de los sistemas de bases de datos, que se vuelve crucial al trabajar con grandes conjuntos de datos ('big data'). Un análisis adecuado de los datos, mediante técnicas apropiadas puede mejorar la toma de decisiones al elucidar patrones y tendencias en grandes cantidades de datos complejos.

En una primera aproximación, podemos distinguir entre dos enfoques para abordar una estrategia analítica: el enfoque dirigido o 'targeted' y el enfoque no dirigido o 'untargeted'. En el primer enfoque, se controlan marcadores moleculares específicos y seleccionados previamente y/o durante el análisis. Esto requiere, generalmente, complejos tratamientos de muestra previos y métodos muy selectivos que proporcionan información limitada sobre especies químicas individuales a la vez: como ejemplo, en un estudio de autenticidad

---

66. Vera, D.N., Jiménez Carvelo, A.M., Cuadros Rodríguez, L., Ruisánchez, I., Callao, M.P. Authentication of the geographical origin of extra-virgin olive oil of the Arbequina cultivar by chromatographic fingerprinting and chemometrics. Talanta, 203 (2019), 194-202. https://doi.org/10.1016/j.talanta.2019.05064
67. Tian, L., Zeng, Y., Zheng, X., Chiu, Y., Liu, T. Detection of peanut oil adulteration mixed with rapeseed oil using gas chromatography and gas chromatography–ion mobility spectrometry. Food Anal. Methods, 12 (2019), 2282-2292. https://doi.org/10.1007/s12161-019-01571-y
68. Oueslati, I., Manaï, H., Madrigal Martínez, M., Martínez Cañas, M.A., Sánchez Casas, J., Zarrouk, M., Flamini, G. Volatile molecular markers of VOO thermo-oxidation: effect of heating processes, macronutrients composition, and olive ripeness on the new emitted aldehydic compounds. Int. Food Res. J., 106 (2018), 654-665. https://doi.org/10.1016/j.foodres.2018.01.029

una adulteración sólo puede detectarse si el adulterante se conoce antes de iniciar el estudio y es buscado explícitamente por el analista. También se conoce como análisis dirigido el análisis de perfiles de componentes o 'profiling' donde el análisis se dirige a múltiples marcadores secundarios conocidos, normalmente de una misma familia química. En general, el perfil suele ser más rico en información y proporciona un mayor poder de clasificación que la búsqueda de un único marcador [**69**].

En un segundo enfoque, el análisis no dirigido, también denominado 'fingerprinting' se basa en un enfoque holístico que pretende proporcionar información sobre todo el producto alimentario, normalmente como huella instrumental analítica ('fingerprint'), ofreciendo una imagen simplificada y global del alimento analizado. Por lo tanto, podría definirse como una metodología analítica destinada a obtener información específica vinculada a la composición química de un determinado material sobre su identidad o sobre una cualidad distintiva cualitativa o cuantitativa a partir de una señal instrumental inespecífica (huella) que contiene la información de interés de forma no evidente y no explícita y que necesita ser extraída mediante la aplicación de métodos quimiométricos específicos de minería de datos [**70**]. El punto fuerte de la huella instrumental es su capacidad para detectar múltiples cambios pequeños en el producto alimentario y de extraer estos cambios como información valiosa a través de los métodos multivariable adecuados.

Mientras que en un enfoque dirigido podemos encontrar procedimientos y métodos estandarizados, se considera que un análisis no dirigido es un enfoque más rápido y de alto rendimiento, aunque adolece de una falta de procedimientos estandarizados. El primer método normalizado que empieza a recoger esta terminología ha sido publicado recientemente. Se trata de la Especificación Técnica ISO/TS 22115 para la separación de lípidos en aceites y grasas mediante cromatografía de gases en cuyo título aparece literalmente el término “método de huellas dactilares” [**71**].

En general, el uso de la metodología de huellas instrumentales suele basarse en la construcción de una base de datos adecuada con muestras de referencia

---

69. Ballin, N.Z., Laursen, K.H. To target or not to target? Definitions and nomenclature for targeted versus non-targeted analytical food authentication. Trends Food Sci. Technol., 86 (2019), 537543. https://doi.org/10.1016/j.tifs.2018.09.025
70. Cuadros Rodríguez, L., Ortega Gavilán, F., Martín Torres, S., Arroyo Cerezo, A., Jiménez Carvelo, A.M. Chromatographic fingerprinting and food identity/quality:potentials and challenges. J. Agric. Food Chem., 69 (2021), 14428-14434. https://doi.org/10.1021/acs.jafc.1c05584
71. ISO/TS 22115. Animal and vegetable fats and oils —Separation of lipid classes by capillary gas chromatography (fingerprint method), 2021.

auténticas con la que comparar la huella analítica de la muestra obtenida y la realización de modelos cuantitativos o cualitativos de discriminación de las mismas según diversos fines. También se pueden utilizar los índices de similitud [**72**] para comparar las muestras con una referencia cuando no se dispone de esta base de datos.

Por todo lo expuesto anteriormente, numerosas técnicas son usadas en el enfoque no dirigido de huella dactilar que proporcionan una gran cantidad de datos de gran complejidad. El tratamiento correcto de estos datos es de suma importancia. En este sentido, se presenta la **Publicación I** donde se expone una visión general del enfoque no dirigido, que amplía el conocimiento tanto de las técnicas, como de las herramientas multivariables de tratamiento de datos utilizados en la metodología de huellas dactilares, así como una declaración de sus aplicaciones más recientes en los campos de calidad y autenticidad de los alimentos. Una descripción más detallada de las técnicas de minería de datos más habituales se proporciona en el Capítulo II.

72. Ortega Gavilán, F., Valverde Som, L., Rodríguez García, P., Cuadros Rodríguez, L., Bagur González, M.G. Homogeneity assessment of reference materials for sensory analysis of liquid foodstuffs. The virgin olive oil as case study. Food Chem., 322, 126743 (2020), 1-12. https://doi.org/10.1016/j.foodchem.2020.126743

PUBLICACIÓN I.

CAPÍTULO DE LIBRO

Non-targeted fingerprinting approaches

En el libro: *Food Authentication and Traceability, Charis M. Galanakis (ed.), Academic Press, 2021.*
https://doi.org/10.1016/C2019-0-03808-3


6

Nontargeted fingerprinting approaches

Ana M. Jiménez-Carvelo, Sandra Martin-Torres,
Luis Cuadros-Rodríguez, Antonio González-Casado

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

Chapter 6. Non-targeted fingerprinting approaches

Ana M. JIMÉNEZ CARVELO✉, Sandra MARTIN TORRES, Luis CUADROS RODRÍGUEZ, Antonio GONZÁLEZ CASADO

Department of Analytical Chemistry, Faculty of Sciences, University of Granada, C/ Fuentenueva, s/n, E-18071 Granada (Spain); ✉ *amariajc@ugr.es*

Abstract

This chapter provides an overview of the different analytical techniques which are capable of generating analytical signals suitable to be employed for applying non-targeted fingerprinting as well as the level of information given for each of them, depending on the specificity of the signal (high, middle or level informative signal). Moreover, a brief explanation of the basis for developing multivariate qualitative and quantitative methods is also included. Likewise, the conventional validation process of these methods is described and a new approach for establishing the validation requirements of the qualitative screening methods is discussed. Similarly, the most recently applications of the fingerprinting methodology for food quality and food authentication are declared. Finally, advantages, weakness and future prospects in this field are stated.

Key Words

Food authentication, fingerprinting, chemometrics, screening methods, multivariate method validation.

Contents

Note: *Abbreviations and acronyms contained in this chapter are listed and explained in the section 8.*

1. Introduction

In the current world scenario of food consumption, there has been a growing interest among consumers to buy high quality healthy foods. The main reason is related to the increased awareness about the right food which has to ensure the health of the consumer. Therefore, in this scenario, authentication and traceability play a very important role in controlling the quality and safety of food. Food safety is an essential aspect for the food industry, so there are increasing controls on products throughout the entire supply chain, "from farm to table". Hence, the need is twofold, on the one hand, to verify that a product complies with the descriptions on the label to detect adulteration and fraud and, on the other hand, track the product through each of the manufacturing stages. All this is done to respond to consumer demands, the demands of export and import markets, and government regulations.

Food authenticity can be defined by the legal recognition of distinctive characteristics of a genuine food product, being its uniqueness, quality (and safety), and traceability guaranteed by a recognition regulatory framework. Usually, this recognition is a bureaucratic and rigorous process concerning a large number of entities, including the scientific community, governmental agencies, and bodies assessing quality and food stakeholders. Food fraud should be considered as a highly dynamic purposive act of substitution, addition, tampering or misrepresentation of food, food ingredients or food packaging, carried out for the purpose of economic gain in which fraudsters aim to escape the regulatory and industry controls. Food authenticity is not a new concept, but it seems to be surging. The engagement with fraudsters and detection of adulterated and contaminated food is continuous and becoming increasingly more sophisticated, with the leading food categories of reported food fraud including adulteration and mislabelling of milk and dairy products, meat, seafood, wines, spirits, edible oils, honey, fruit juices, coffee and tea, organic food and products, and clouding agents.

It is at this point that Analytical Chemistry plays a fundamental role, since analysts have to be responsible of develop methodologies that satisfy all these demands. Developing increasingly effective methodologies for the detection of food adulteration, the confirmation of both food traceability and authenticity and verifying their compliance with labelling is a challenge for analytical chemists worldwide. Different chemical approaches have been carried out to address this problem markers, targeted (profiling) and non-targeted (fingerprinting) methodologies. This chapter is aim to describe the use of methodologies based on non-targeted fingerprints, in which the signal can be obtained by means of different fast and non-specific analytical techniques to be subsequently complemented by the corresponding chemometric tools.

Research to date has tended to focus on analytical methods to detect food fraud. A number of excellent reviews on analytical techniques on both targeted (*i.e*. when chemical species of interest is known) and non-targeted methods (*e.g.* screening) have been recently published [**1**,**2**,**3**]. Non-targeted analysis, for food fraud, is comprised of using an analytical technique that affords a sufficiently resolved instrumental signal of the representative authentic sample. While the instrumental signal may vary depending on the technique, it will be possible to compare the signal profile of the sample in question to a library of signal profiles gathered previously that represent historic material that has been shown to be fit for purpose. Several analytical techniques, such as spectrometric, electrochemical and chromatographic approaches have been employed for monitoring food authentication including many important and complex analytical platforms such as hyphenated and multisensory techniques. Then, appropriate multivariate data analysis (chemometrics) is used to determine the similarity between the new and historic samples [**4**].

Every analytical instrument or method can be classified according to the type of data it provides. Using existing terminology from mathematics, we can define a tensor as a mathematical object that can hold data. Measuring equipment that generates a single datum per sample is a zero-order instrument because a single number is a zero-order tensor. In other words, the data for a single chemical sample consist of a scalar; indeed, a scalar is a zeroth-order tensor. Examples of zero-order instruments are ion selective electrodes and single-filter photometers. Instruments capable of generating multiple measurements at one time or for one sample are considered first-order instruments; all types of spectrometers, chromatographs, and even arrays of zero-order sensors are first-order instrument. Hence a single index suffices to organise the data and the measurements can be put into an ordered array referred to as a vector of data (or first-order tensor).

Second-order instruments can generate a matrix (second-order tensor) of data per sample; these instruments are used in the hyphenated techniques such as EEM fluorescence, GC-MS, LC-DAD, MS/MS, either 2D chromatographies or spectrometries, and in sophisticated sensors. Data generated by these instruments can be viewed equivalently, for example for GC-MS, as a set of chromatograms, each measured at a different nominal mass, or as a set of mass spectra that change over time as determined by the chromatography.

There is no limit to the maximum order of data that can be generated. Excitation-emission time decay fluorescence spectroscopy produces a third-order tensor of data per sample. The more instrument complexity, the more advantage obtained from the data. The most obvious advantage is increased analyte selectivity. For example, when measuring a scalar, bias in a sample due to a varying interference cannot be detected. However, it is often the case that

large quantities of data do not provide as much additional information as they might. This problem arises not because the additional data do not contain additional information, but because the data are in a form that does not make the information readily available. Substantial effort is required to identify the underlying structure of large data sets ('big data'). Data having more than three variables become less comprehensible as the number of variables increases. The evolution and application of multivariate and megavariate statistical methods (chemometrics, machine learning) have solve many authentication problems in a wide variety of food products [**5**,**6**].

2. Non-targeted analytical methods - fingerprinting

Traditional strategies for the food authenticity and fraud control have relied on the determination of the amount of a marker compound or a set of marker compounds and the comparison of the obtained values for the test material with those established for the genuine material. This approach has been classically defined as targeted analysis. These targeted methods detect a single chemical species at a time and, therefore, often provide only limited information: an adulteration can be detected only if the adulterant is known beforehand and explicitly searched for by the analyst. Profiling strategies had been also afforded. A profiling analysis targets multiple secondary markers. The exact number of targets to constitute a profile is not scientifically described or internationally harmonised. In general, profiling is often richer in information and provides increased classification power compared to a single or dual-marker based targeted method.

In recent decade, the fingerprinting strategies [**7**] are gaining more and more popularity thanks to advancements in the analytical instruments that are able to generate enormous amount of data at once and the application of chemometric techniques. Instrumental fingerprint refers to the characteristic unspecific instrumental signal (spectrum, voltamperogram, thermogram, chromatogram, electropherogram or image) from the analysed sample which can be related to its properties, complex chemical composition and thus to its authenticity in the same way as a human fingerprint is specific of a certain person and unequivocally identify him/her. Instrumental fingerprint is provided and recorded by an analytical instrument that requires a mathematical treatment, normally a chemometric approach including projection, clustering, modelling techniques, etc. with the goal of characterizing the food. The term thus recalls a comprehensive description of a test material that is carried out in a non-selective (non-targeted) way [**8**].

A collection of papers that gather applications of instrumental fingerprints to food authentication or food adulteration were published for the last ten years.

Two different approaches have been noticed from the point of view of the involved foods. Vertical revisions, aimed at collecting the different applications of a specific analytical technique, and horizontal revisions devoted to gather generic issues as authenticity, adulteration or food fraud in different foodstuffs and beverages. But it may be noted that the information contained in both revision approaches is quite similar and the quoted papers are frequently the same ones. A detailed description of the studies reported following a vertical approach is presented in section 6, while brief comments of the available literature following the horizontal approach are already included in the following lines.

In 2019, Medina *et al.* has reviewed the application of food fingerprints to monitor food authenticity and safety [**9**]. The potentiality of this methodology in official food control has also been discussed by Esslinger *et al.* [**10**]. Same year, the use of fingerprints using a metabolomic approach was discussed by Cubero-Leon *et al.* [**11**]. In addition, several issues about the application of fingerprints to the detection of adulteration and food fraud have been also reviewed [**12**,**13**,**14**]. Even the application of fingerprinting allowing detection of food adulteration and fraud in China has been also revised [**15**]. As a curiosity, a study in which the fingerprinting methodology was applied to evaluate food quality changes deserves be highlighted [**16**].

In addition, a number of handbooks and guidebooks were published in the last 10 years in which applications of the fingerprint methodology to food analysis appear have been considered. The most significant ones are collected and briefly commented on Table 1.

Table 1. Some significant handbooks/guidebooks focussed on food authentication.

Title (Year)	Description	Ref
New analytical approaches for verifying the origin of food (2013)	This handbook is derived from the results of the European TRACE project (2005-2009) where the instrumental fingerprinting strategy was first proposed for food authentication purposes. Parts II and III deserve to be highlighted. In Part II, chapter 6 is devoted to vibrational spectroscopy in studies for food origin authentication (it includes fingerprinting-based applications of NIR, MIR and Raman), and chapter 7 is focused on the description of the different chemometrics tools. Part III shows a vertical revision of different food commodities.	[**17**]
Food protected designation of origin – Methodologies and applications (2013)	Protected designation of origin (PDO) is a quality seal created by the European Union (EU) to provide added value to their regional foods. The aim of the book is to present the analytical approaches designed for discrimination of PDO products. It is split into three parts, although the most interesting parts are part II and part III. Part II consists of 10 chapters, which describe different analytical and chemometric techniques to be employed for the control of PDO food where fingerprinting plays a major role. As the previous book, part III is composed of 11 chapters, each one dedicated to different food commodity applications.	[**18**]
Advances in food analysis research (2015)	Although none of its chapters deals specifically with fingerprinting, chapters 1, 5 and 10 describe some applications of this methodology.	[**19**]
Advances in food authenticity testing (2016)	It is divided into three parts. The first one makes a vertical revision of the different analytical techniques used in food analysis; in most of these chapters, papers applying fingerprinting are reported. Chapter 3 include an interesting study of food authentication using UV-Vis spectrometry alone or combined with NIR and MIR spectrometries. A horizontal revision on food authenticity and adulteration are showed in the second and third part of the book.	[**20**]
Food authentication – Management, analysis and regulation (2017)	It describes the application of different analytical techniques on food authentication. From Chapter 4 to Chapter 10, the analytical techniques applied for food authentication using mainly fingerprinting are discussed. At the end of chapter 17, the different chemometric tools that may be used for authentication purposes are described.	[**21**]

Food integrity handbook – A guide to food authenticity issues and analytical solutions (2018)	It is divided into different chapters each dedicated to a food commodity: animal products, fish and fish products, plant products, beverages and related products, fats and oils, food additives. At the end of each chapter there is a table summary of applications of the different analytical techniques used for the authentication. Most of these applications employ fingerprint methodology.	[**22**]
Modern techniques for food authentication (2018)	This book is made up of 18 chapters focused on the most common analytical techniques used to authenticate foodstuffs, among which spectrometric (MIR, NIR, Raman, UV-visible, fluorescence, HSI) and chromatographic (GC LC) techniques can be highlighted for their suitability to apply fingerprinting. In addition, chapter 17 is exclusively focused on describing some chemometric tools.	[**23**]
Food traceability and authenticity – Analytical techniques (2018)	It is mainly aimed at describe different tools used to verify and ensure the food traceability. Only in three chapters (6, 13 and mainly 16) develop contents related to food authentication, in which applications of instrumental fingerprinting are considered.	[**24**]
Fingerprinting techniques in food authentication and traceability (2019)	It compiles several examples showing applications using the most common analytical techniques. Despite the tittle, fingerprinting is hardly considered.	[**25**]
Evaluation technologies for food quality (2019)	This book is divided into 5 parts describing 5 sorts of technologies for food quality evaluation. The second part focused on applications of chemical analytical techniques, and only in the chapters devoted to spectrometric techniques (UV-visible, NIR, Raman, HIS and NMR) are fingerprinting-based applications revealed.	[**26**]

3. Analytical techniques for fingerprinting

Fingerprints can be generated through many analytical techniques. Herewith, fingerprinting signals can be classified in three broad categories according to the kind of information they handed: high, middle and low informative. The first ones provided useful and readily available information from the whole sample whereas the last ones need a separation (maybe also derivatisation) step, previous to the signal measuring, to produce discriminatory information. Figure 1 shows a classification proposal of common analytical techniques which are consistent with the fingerprinting methodology.

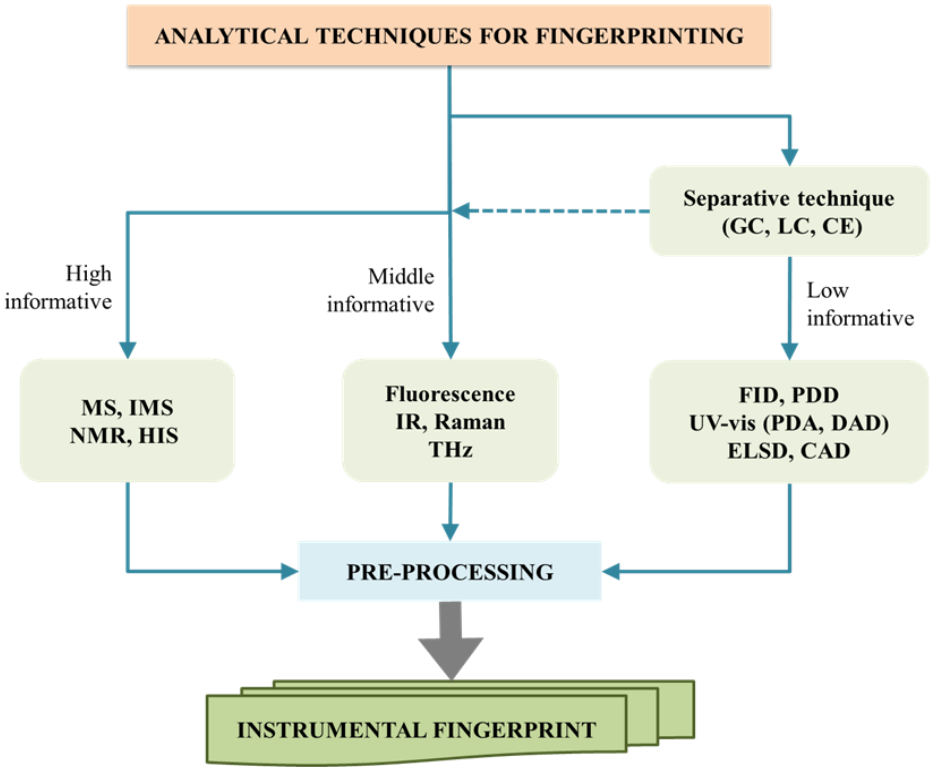


Figure 1. Overview of common analytical techniques consistent with the fingerprinting methodology according to the specificity of the analythical signal.

(The meaning of the abbreviations and acronyms are shown in section 8 of this chapter).

MS, IMS, NMR and HIS are useful analytical techniques for multi-component simultaneous analysis without spending time and effort on investigating separation conditions. They provided high-specific information and accordingly the corresponding analytical signals acquired by direct measurement from the food sample could be used as fingerprints.

Analytical techniques used as measuring devices in separative techniques as FID, PDD, PDA, DAD, ELSD or CAD detectors are classified as low informative

techniques. They respond in a similar way to all (or many all) compounds or families of compounds in a sample, they are also usually considered as universal detectors. This lack of analytical selectivity makes it impossible for them to obtain useful fingerprints as the information on the chemical composition would be confused. This lack of informative competence is compensated by a preceding separation technique that allows obtaining signals related to profiles of chemical compounds that are measured consecutively and not simultaneously.

Fluorescence, IR, Raman and THz molecular spectrometric signals show intermediate characteristic between the two previous categories and these techniques could be classified as middle informative techniques. However, in occasions on which sophisticated separation conditions must be applied, such as when characterising trace components in complex food matrices, both middle and hight informative techniques could be also coupled to chromatographic or electrophoretic separative instruments.

A more detailed description of the analytical techniques capable of providing informative fingerprints for authentication purposes is collected in the following subsections. In addition, a comprehensive review on the applicability of these techniques on different food matrices will be given on section 6.

3.1. Mass spectrometric fingerprinting

One of the most powerful analytical techniques in food authentication field is MS which operates by the formation of positively or negatively charged ions and subsequent detection and measurement of the mass to charge (m/z) ratio. MS produces therefore a mass spectrum representing the mass fingerprint of the sample components. The analysis is carried out in four stages: sample introduction, molecular fragmentation and ions formation in the ionisation source, ions separation based on their m/z ratio in a mass analyser and finally ions detection. A number of different MS fingerprinted set-ups are possible based on the ionisation technique and the mass analyser used. In this regard, soft-ionisation methods may be preferred for authentication purposes because an easier pattern mass spectrum is then obtained from each chemical component.

Depending on the system characteristics and the goal of the analysis, the sample introduction could involve a previous separation step (chromatographic or electrophoretic separation) prior to MS analysis or direct introduction to the mass spectrometer which includes DIMS, MALDI, PTR-MS, SRI-MS and SIFT-MS [**12**].

In recent years a new mod of mass spectrometry is appearing, known as AMS [**27**] that is characterized by the fact that the ambient ionisation step occurs outside the instrument and only the ions produced are injected; there are three types depending on the ionisation method: ASAP, DART and DESI.

3.2. Ion mobility spectrometric fingerprinting

IMS is a gas-electrophoretic separation technique in which ionised compounds are separated in a neutral gas phase at atmospheric or near to atmospheric pressure. Therefore, separation takes places under an electric field and is the result of the difference in mobility of ions in the drift cell.

Despite IMS fundamentals have been developed since the beginning of the 20th century, it has not been until the recent commercialisation of hyphenated IM-MS instruments when this technique has really caught the attention of researchers from multiple fields, including food science. There is a wide variety of stand-alone IMS and IM-MS instruments on the market based on different technologies with different advantages. They can be classified in time-dispersive, space-dispersive and trapping (*i.e.*, ion confinement and release) technologies. Despite the fact that IMS can be used alone as analytical tool, it can also be coupled to other analytical separation techniques such as GC, LC or CE.

IMS has mainly been applied to targeted or semi-targeted analysis. However, it is expected that it will be used in non-targeted approaches, such as food fingerprinting [**28**].

3.3. Nuclear magnetic resonance spectrometric fingerprinting

NMR spectrometry is one of the major techniques used for the non-targeted analysis of food for its authentication [**12**]. NMR is a technique which takes advantage of the energy absorption by atomic nuclei with non-zero spins (mainly ^{1}H, ^{13}C, ^{19}F and ^{31}P) and in the presence of a magnetic field. The energy absorptions of the atomic nuclei are affected by the nuclei of surrounding molecules, which cause small local modifications to the external magnetic field. NMR can therefore provide detailed information about the molecular structure of a food sample (identify functional groups) and determine the physical and chemical properties of atoms within a molecule, given that the observed interactions of an individual atomic nucleus are dependent on the atoms surrounding it.

NMR fingerprinting aims at establishing a global approach and the tested system is submitted to the NMR experiment as a whole. A simple quantification of the major compounds, which are characterised by one or several signals in

the NMR spectrum, can be performed. This type of analysis is particularly attractive for several reasons: it is non-destructive, non-selective and cost effective; requires little or no sample pre-treatment; uses small amounts of organic solvents or reagents; and typically takes only a few minutes per sample [**29**,**30**].

3.4. Molecular absorption spectrometric fingerprinting

Molecular UV-Vis fingerprints are probably the less informative one and thus it is not suitable for authentication for most of the foods. It has been only exploited for authentication of coloured liquid foodstuffs as juices and wine. Usually molecular UV-Vis absorption spectrophotometers are used as a detector coupled to liquid chromatographic system.

Vibrational spectrometries, as IR or Raman are rapid and not destructive techniques. IR signals are associated with molecular vibrations, specifically with overtones and combinations of fundamental vibrations, mainly of the light atoms CH, OH, and NH. NIR, FTIR (or MIR) and Raman fingerprints have been successfully used to control the authenticity and quality of food [**36**].

The far-infrared or terahertz region refers to a very small gap between the microwave and infrared regions of the electromagnetic spectrum. By comparison with the relatively well-developed and widespread previous spectrometries, research and application of the THz waves are still in their infancy. THz spectrometry can be used to study the vibrational activities of molecules. THz wave can penetrate many commonly used nonpolar dielectric materials and indicate vibrational modes of many biomolecules, which make it an extremely attractive tool for agri-food products inspection. Torsional and rotational modes of molecules can also be observed in the low frequency THz region. In addition, compared with the commonly used UV-Vis and IR spectra, the wavelength of the THz spectral band is longer and thus will not be easily affected by scattering [**31**].

3.5. Molecular fluorescent spectrometric fingerprinting

Molecular fluorescence spectrometry, because of its rapid analysis, relatively inexpensive in addition to its ease of use, sensitivity and selectivity, instrumental versatility, speed of analysis and its non-destructive character, is a powerful analytical tool for using in food authenticity fields. Comparing with the other spectrometric analytical techniques (e.g., IR or UV molecular adsorption), molecular fluorescence spectrometry is the greater sensitivity and can be used to analyse materials including endogenous fluorescent compounds even at very low concentration levels. Molecular fluorescent spectrometry can be used alone

(measured on the solution or in front-face mode) or it can be coupled with separation techniques. Unfortunately, conventional fluorescence spectrometry offers a weak selectivity in the analysis of complex samples because of the intrinsic broad nature of a molecular fluorescence spectrum. It gives information about the fluorophores naturally occurring on the sample and also about the full environment of those fluorophores in the sample matrix. Thus, molecular fluorescence spectrometry is not suitable for the analysis of complex multi-component samples without prior separation, normally by chromatography, due to severe overlaps of spectrum bands. EEM fluorescence spectrometry, in which a total fluorescence spectrum is obtained by systematically varying the excitation and emission wavelengths and collecting the resulting data matrix (consisting of excitation plotted against emission wavelengths), may improve the selectivity of analysis in a moderate way. Due to the additional mode, the capability for resolution of overlapped fluorescence spectra is improved. The 2D character of EEM fluorescence spectra implies a three-way nature for a set of samples which need to be analysed by adequate chemometric methods. At present, multiway methods coupled with EEM fluorescence spectrometry are mainly used to quantitative analysis, and only a few literatures were applied to classification research [**32**,**33**].

3.6. Hyperspectral imaging fingerprinting

As an emerging novel analytical spectrometric technology, hyperspectral imaging has been introduced for the identification and quantification of chemical constituents as well as the simultaneous determination of their location or spatial distribution in the sample. This technology has recently been applied as a powerful process analytical tool for rapid, non-contact, and non-destructive inspection of a wide variety of food and agricultural products [**34**]. HSI captures a stack of images, forming a three-dimensional structure of multivariate image data (hypercube), consisting of a spectrum for each pixel in the image. Almost all of the hyperspectral methods use a pixel-based approach, which yields a large number of predominantly NIR spectra per sample, but discards the spatial coherence present in the analysis. This technology combines the advantages of image analysis and spectrometry to analyse heterogeneous structures and obtain chemical information in the same image but also it offers many advantages over conventional methods making hyperspectral imaging to stand out over them. It is a non-contact, non-destructive analytical methodology that enables multi-component information to be obtained from a sample beyond the ability to identify the spatial distribution of multiple chemical and physical components in a sample [**35**].

3.7. Chromatographic fingerprinting

Chromatographic techniques, as GC and LC, are mainly employed to separate the different components of a complex mixture. Then, depending on the coupled measuring system (or detector), each individual constituent or constituent family will generate specific informative signals which can be assumed of as a food sample fingerprint of one or more classes of compounds occurring in the sample suitable for qualitative/quantitative analysis. In chromatographic fingerprinting the identification of each single peak is not necessary [**7**]. Strictly speaking, a chromatographic fingerprinting is the results of two steps: (a) a total, or partial components separation, and (ii) a measurement related to each single or co-eluted component. Therefore, the profile and the intensity of the fingerprint signal depend on both separation and measuring performances.

Regarding the detection of the compounds separated in the GC systems, detectors had been generally classified as universal (measure nearly all components in a mixture) or selective (detect components with a specific structural feature in their molecules). Examples of the most commonly used universal detectors include the FID and PDD. With no doubt the most used measuring system in GC systems in the latest authenticity studies reported is the MS instrument.

The most employed detection systems coupled to LC techniques are the UV-Vis light absorbance detector, particularly PDA or DAD, fluorescence detector, ELSD and CAD. As in the CG case, the MS instruments already constitute the more promising measuring system for food authentication. Currently, hyphenated techniques combining separation techniques (e.g. GC, LC, CE) with spectrometric techniques (e.g. MS, NMR) are receiving special attention given their ability to solve complex analytical problems.

3.8. Capillary electrophoretic fingerprinting

CE is based on the separation of charged molecules based on their electrophoretic mobility in dissolution. An electric field is applied to the ends of a capillary column and the ions migrate through it. The signal generated when ions are detected as they are eluted from the column is referred to as a capillary electropherogram.

CE technique shows a notable potential for food fraud detection based on product fingerprint, usually also coupled to MS instrument. CE is capable of rapid, low cost and high-resolution analysis with low consumption of mobile phase. Main drawbacks of the technique are both poor reproducibility (compared to chromatography) and sensitivity.

4. Multivariate methods/models

Once instrumental fingerprint is acquired, regardless the employed analytical technique, multivariate methods are needed to obtain relevant information about the sample. The given definition of instrumental fingerprint implies that the whole signal obtained from the corresponding measuring device must be a complete and available description of the tested sample. Therefore, a pre-processing of the instrumental response is necessary to remove all the excess of information that masks signal of interest prior to develop of the multivariate method. Raw spectral data need to be submitted to mathematical transformation in order to minimise the source of spectral variability not related to the research subject as well as to maximise the comprehensive interpretation of chemical information from chemical data, to gather deep knowledge about the system. The main employed pre-processed techniques are [**36**]:

- Baseline correction: It is applied to eliminate systematic deviations that could occur in the baseline of the analysed sample due to variations on the instrumental equipment.
- Variable reduction (sometimes referred to as decimation): When working with a very high mass of data/variables, the available chemometric techniques are sometimes insufficient for their properly treatment. Care should be taken to ensure that, when the reduction is carried out, the profile of the fingerprint must be kept.
- Noise filtering (smoothing): It focuses on the elimination of that component of the instrumental response independent of the analytical information derived from the sample. As in the reduction stage, it is important to control the smoothing to avoid the loss of critical analytical information.
- Mean centring: It is carried out by calculating the mean value of the variables in each column and subtracting each variable from the column.
- Autoscaling: This pre-treatment is normally applied when there are data in different scales in order to ensure that all have the same weight in the data-modelling.
- Alignment of the signal peaks, standard normal variate, multiplicative scatter correction, orthogonal signal correction, selection of interest region and correction of death pixel pre-processes are also common employed depending on the technique of analysis.

A multivariate model is a powerful statistical tool that uses multiple variables to forecast or understand the behaviour of an objects/samples population. The development of the multivariate model can be carried out with two different goals which give rise to [**37**,**38**]:

- Explanatory models (also known how descriptive models). The aim is to describe and/or explain the sources of the phenomenon in study. These kinds of models are widely used in the food quality field, for example an explanatory model of degradation such as oxidation. Knowing that the exposure to some environmental conditions leads to higher risk of having loss of quality is valuable information since if the exposure decreases this will result or will be expect an increase of shelf-life.
- Predictive models. The purpose is to know the value of a qualitative o quantitative feature from future observations, *i.e.*, it is extrapolating into the unknown using the known relationships between the objects. Unlike previous methods these are more widely used in food quality/authentication field, since the goal is to forecast if a sample whose class is unknown belong to particular class or to predict some value of a particular property.

In this sense it is possible to say that explanatory model is all about 'what can be done about it?', whereas predictive model is all about 'what is likely to happen?'

The development of these models is performed applying pattern recognition methods. The terminology employed to name these methods depends on the area of study. In analytical chemistry field is common to use 'pattern recognition methods' or 'multivariate analysis methods', although more recently the terms 'data mining methods', 'machine learning methods' and 'deep learning methods' have been incorporated. These methods are powerful tools to conduct food quality control and food authentication [**39**].

In general terms the development of the multivariate method involves two steps [**40**]: (i) training stage in which the model is trained using objects whose class is known; and (ii) validation stage in which the model is evaluated to assure the appropriate performance of it. The objects selection to be used at one or the other step of the process can be carried out randomly or applying a proper subset splitting method such as CADEX algorithm (so-called Kennard-Stone algorithm), which allow to pick those samples representative of the all the objects population [**41**]. In addition, it is important to highlight that depending on the model built (explanatory or predictive) the validation stage is achieved in a different way and this will condition the application of them. In this context, some authors confuse the kinds of validation required contingent upon the method purpose, so when they state that the method has been validated and can be used in routine laboratory is an incorrect statement [**42**]. Although it is true that sometimes the model validation is influenced by the number of objects/samples.

As stated before, in food quality and authentication field is more common to build predictive models, so in this case for the model evaluation should be

carried out an external validation. In this, additional samples that fulfil the same requirements but were not part of the original training set are employed. A cross-validation, however, should be applied for explanatory models; this validation is performed on the data from training set and it is employed for different reasons: (i) when there is a limited available quantity of objects and it is not possible to split in two dataset (training and external validation), (ii) to evaluate the classificatory power of the model with the training data, and (iii) to select the components number necessary to establish the predictive model. There are several kinds of the cross-validation, although the most common are called 'leave one out' and 'venetian blinds'. In all of them, the data from training set are divided into subset (segments) and the model is developed using some segments and leaving others to evaluate the model, how are the segments chosen is the main difference between the types of cross-validation [**41**,**42**].

Pattern recognition methods are divided into two groups: (i) unsupervised methods which are employed to check if there is natural grouping between the objects/samples, in these methods the class to which the samples belong is not considered and (ii) supervised methods, in these is known the class to which the samples belong and it is used to develop the multivariate methods [**43**]. Within them these are two kinds of multivariate methods depending upon the invoked information tier [**44**].

4.1. Multivariate qualitative methods

They are used to differentiate or classify an object set using data which explain any/some particular characteristics of them. These are an analytical strategy for addressing problems related to food quality or food authentication that cannot be solved with just one variable [**45**]. Three classification strategies could be applied: class-modelling, discriminant analysis and decision rules based on voting [**39**].

Particular types of qualitative methods are the screening methods which are employed to provide fast information about the objects and whose most common response is binary in nature, for instance, 'yes' or 'not'. In short, a screening analytical method is aimed at discriminating objects or samples which hold a target property or feature from those that do not. The concept of screening method is not new and in fact it comes from medical field, in which is used commonly to early detection of diseases. Draws attention, however the non-existence of procedures or protocols to apply this kind of methods in food quality and food authentication field when there is the 'gold rush' in developing faster analytical methods in order to solve problems in this field in recent years. The use the 'vanguard' term to refer to them how rapid and low-cost methods which allow to select the objects with target characteristics has been proposed

[**46**], on this way they can be used for immediate decision. Some papers related to the application of them in analytical chemistry can be found in bibliography [**47**,**48**,**49**].

The development and validation of the screening methods when are employed in the food quality and authentication is decisive and it is different to the conventional way of building of classification method in the same field. Prior to the development of the method the crucial question is to decide what use will be made of the results. From this use, the acceptation criteria in the validation step of the method will establish according to the final aim purposed. This question will be addressed in the following sections in more detail.

Data mining techniques commonly applied for the development of these methods are [**39**,**50**,**51**], SIMCA [**52**], PLS-DA [**53**], SVM [**54**,**55**,**56**], CART [**57**] and RF [**58**].

4.2. Multivariate quantitative methods

The goal is to predict a property of interest from multiple instrumental measurements. These are multivariate regression methods focused on determining the functional relationships between the analytical signal acquired from a set of samples and a characteristic feature of such samples such as their composition [**59**]. The advantages regarding univariate regression are: (i) to perform the quantification in the presence of interfering substances; and (ii) to quantify materials (not analytes), for example the composition of olive oil in blends with other vegetable oils [**60**]. Unlike multivariate qualitative methods these are just quantitative and generally applied to data between which a certain relationship is known to exist. Multivariate quantitative methods could be effectively applied for predicting the value of a quantitative property or magnitude linked to both a chemical component (for example, the concentration of an analyte or the value of a related parameter) and material as a whole (for example, the proportion of particular vegetal oil in a blend of vegetal oils).

The most widely used algorithm is partial least squares regression. PLSR calculates a partial least squares regression model to predict a dependent y-variable from a set of independent x-variables [**61**]. In addition, there are other kinds the multivariate methods that are applied when working with two-way fingerprints, that have the possibility of obtaining information for each sample in two dimensions, [**62**,**63**]. In this case, the most common methods are PARAFAC and MCR-ALS.

5. Validation of fingerprinting analytical methods

Validation is fundamental to ensure the reliability, traceability or comparability of results. Although, some recent studies in bibliography related to recommendations to perform the validation process can be found [**42**,**48**,**64**,**65**], however there are scarcely official or widely recognised protocols or standard criteria for validation of these in food authentication field [**66**]. In addition, it is important to highlight the selection and number of samples affect the model performance and is decisive on the model reliability. In this sense, a recent Eurolab Guide [**67**] focused on fingerprints obtained using NMR recommended a minimum of 50 samples (representative of the class which belong to objects) in order to ensure the representativeness of the validation results of an analytical multivariate method (qualitative and quantitative). In the same context, Musio *et al*. stated the need of establishing metrological traceability as overriding requirement to compare the measurements results produced at different laboratories and at different times using NMR when untargeted approach is employed in food science [**68**]. Even, in the pharmacy field there are already official protocols for the application of multivariate methods as well as their proper validation unlike food quality and food authentication fields [**69**,**70**].

5.1. Multivariate qualitative methods: successes/errors contingency

The scope of multivariate classification methods in analytical chemistry is increasingly and the number of published analytical applications is growing in all fields. Conventionally, the validation of these methods has involved the estimating of several quality performance metrics from contingency chart (validation plot). A generic example of validation plot collecting the results of a binary classification method is showed on a 2×2 square chart where the y-axis represents the number of classifier assignations and the x-axis is the actual value of class (see Figure 2).

Using the symbols displays in Figure 2 are defined the main quality performance metrics follows [**71**]:

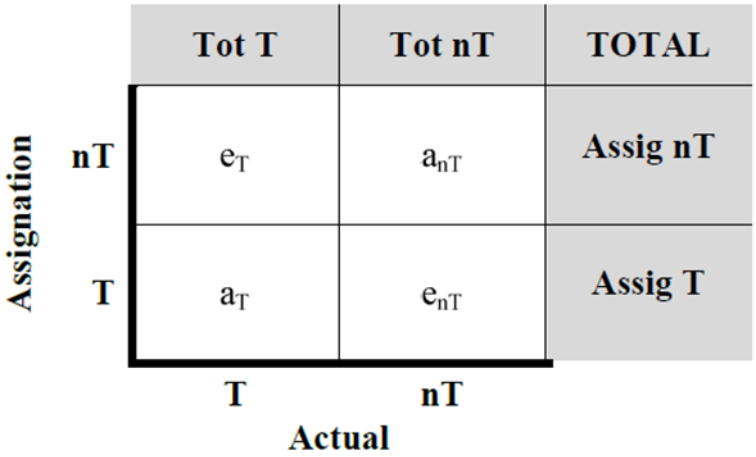


Figure 2. Contingency chart for binary classification.

a_T: number of samples correctly assigned to the target class; a_{nT}: number of samples correctly assigned to the non-target class; e_T: number of samples wrongly assigned to the target class; e_{nT}: number of samples wrongly assigned to the non-target class; T: target class; nT: non-target class.

- *Sensitivity or recall (SENS)*. Proportion of agreements of the target class.

$$\mathrm{SENS} = \frac{a_T}{a_T + e_T} = \frac{a_T}{\mathrm{Tot\ T}}$$

- *Specificity (SPEC)*. Proportion of agreements of the non-target class.

$$\mathrm{SPEC} = \frac{a_{nT}}{a_{nT} + e_{nT}} = \frac{a_{nT}}{\mathrm{Tot\ nT}}$$

- *Positive predicate value or precision (PREC)*. Proportion of agreements in relation to all assignments to the target class.

$$\mathrm{PREC} = \frac{a_T}{a_T + e_{nT}} = \frac{a_T}{\mathrm{Assig\ T}}$$

- *Negative predicate value (NPV)*. Proportion of agreements in relation to all assignments to the non-target class.

$$\mathrm{NPV} = \frac{a_{nT}}{a_{nT} + e_T} = \frac{a_{nT}}{\mathrm{Assig\ nT}}$$

These four metrics signify a probability and their value ranges from 0 to 1, although it only values larger than 0.5 will be suitable since equals or below values to 0.5 is considered that the method classified the samples with a random distribution.

From these main rates, it is possible to deduce supplementary quality metrics to evaluate the goodness of the qualitative multivariate method. Cuadros-Rodríguez *et al*. [**40**] published a tutorial in which the most performance

indicators for classification methods, such as false positive rate, false negative rate, Youden's index, likelihood ratios, classification odds ratio, F-measure, discriminant power, efficiency, misclassification rate, area under the receiver operating curve, Gini coefficient, G-mean, Matthews' correlation coefficient, change agreement rate, chance error rate and Kappa coefficient are collected. Among them, the most common are:

- *False positive rate (FPR) and False negative rate (FNR)*. Supplementary metrics of SPEC and SENS are expressed as the proportion of errors in the assignment of both classes respectively.

$$\mathrm{FPR} = 1 - \mathrm{SPEC}\,; \quad \mathrm{FNR} = 1 - \mathrm{SENS}$$

- *Efficiency or* accuracy *(EFFIC)*. Probability of correct classification in relation to the total number of objects. It represents the skill of the method to correctly classify a sample.

$$\mathrm{EFFIC} = \frac{a_T + a_{nT}}{a_T + e_{nT} + a_{nT} + e_T} = \frac{a_T + a_{nT}}{\mathrm{TOTAL}}$$

- *Area under the receiver operating characteristic curve (AUC)*. This quality metric evaluate the ability of the model to avoid errors during classification using the receiver operating characteristic (ROC) curve.

$$\mathrm{AUC} = \frac{\mathrm{SENS} + \mathrm{SPEC}}{2}$$

4.2. Multivariate quantitative methods

In this kind of methods, the reliability of them is commonly established on the basis of the determination coefficient (R^2) and the root mean square error of validation (RMSEV):

$$\mathrm{RMSEV} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_{ref,i})^2}{n}}$$

Although, the validation errors also can be evaluated estimating mean absolute error of validation (MAEV) and median absolute error of validation (MdAEV):

$$\mathrm{MAEV} = \mathrm{mean}\left(\left|\hat{y}_i - y_{ref,i}\right|\right)\ ; \quad \mathrm{MdAEV} = \mathrm{median}\left(\left|\hat{y}_i - y_{ref,i}\right|\right)$$

Where 'n' is the number of different validation objects and '$\hat{y}$' the predicted values of the property of dependent variable, and 'y_{ref}' the reference values of the same property.

In the case of multivariate quantitative methods, there are some guides published, although the majority of these are based on the validation of multivariate quantitative method using infrared spectrometric techniques [**72**].

The estimating of these metrics is usually carried out when are applied multivariate calibration techniques such as partial least squares regression, principal component regression or support vector machine regression, since the multivariate quantitative method is developed from two-way data. However, when working with analytical techniques which generate data that can be arranged into more complex mathematical such as a matrix or a three-dimensional array, namely multiway-data, the validation indicators usually estimated are the same that in the case of the univariate approach: sensitivity, selectivity, detection limit and quantification limit [**73**,**74**]. As commented in previous section the multivariate techniques employed in this case are PARAFAC and MCR-ALS. Although in food authentication field is fewer common works with this kind of data in quantitative analysis.

4.3. Validation requirements of the classification/screening methods

ISO/IEC 17025 defines the validation step as *provision of objective evidence that a given item fulfils specified requirements, where the specified requirements are adequate for an intended use*" [**75**]. Nowadays laboratories must demonstrate that their analytical methods provide reliable and adequate results for to ensure that it is fit-for-purpose since many decisions that are made are based on the information that this data provides. The validation of methodologies, together with other activities involved in quality assurance control, makes it possible to demonstrate to laboratories that their analytical methods provide reliable results. Validating a method is to verify its suitability to certain requirements, previously established by analysts in order to solve a particular analytical problem.

These requirements are those that define the quality parameters or criteria that a method to use must hold to solve an analytical problem. The general stages of the validation process can be seen in Figure 3.

Usually for validating the multivariate classification/screening methods have mainly determined quality parameters such as sensitivity which indicating the percentage of successes of target class samples classified as such in relation to with the total target samples or the specificity that represents the total percentage of successes but relative to the non-target class, but in the screening methods does not provide information of interest of the method. On the contrary, there is another crucial parameter, the precision, which is not very common and represents the proportion of target samples correctly allocated as such in relation to the total of the samples assigned to that class.

Thus, what are the criteria for accepting validation of the method? Is any precision or sensitivity value valid? Or are they simply taken for good if they approach a value of 1 that would be the ideal case? To answer that question

Cuadros-Rodríguez *et al.* [**76**] have proposed the use of new applicability indicators the method called: error index (I_{ERROR}), saving index (I_{SAVING}), penalty index ($I_{PENALTY}$) and loss index (I_{LOSS}), depending on method applicability scenario: trade/marketing scenario, which presents two possible situations: quality-oriented focus or profit-oriented focus, or conformity scenario.

Figure 3. General process of validation of an analytical method.

- I_{ERROR}: designed for both the conformity scenario and the quality-oriented focus on the trade/marketing scenario and it indicates the risk of erroneously assigning a sample as belonging to the target class. It is the risk of non-compliant products on the market without being detected.

$$\mathrm{I_{ERROR}\ (\%)} = \frac{\mathrm{e_{nT}}}{\mathrm{TOTAL}} \times 100$$

where 'e_{nT}' is the number of samples wrongly assigned to the target class.

- I_{SAVING}: is the percentage of samples that do not require confirmation *i.e*. they are all samples whose assignation is considered reliable, once analysed by the screening method.

$$\mathrm{I_{SAVING}\ (\%)} = \frac{\mathrm{Assig\ T}}{\mathrm{TOTAL}} \times 100$$

where 'Assig T' is the total number of samples assigned as belonging to the target class and 'TOTAL' is the number of samples analysed.

- $I_{PENALTY}$: is similar to the I_{ERROR} but it is only applied in trade/marketing scenario. It is an estimation of the risk of being penalised if the product is subjected to conformity assessment.

$$\mathrm{I_{PENALTY}}\ (\%) = \frac{\mathrm{e_{nT}}}{\mathrm{TOTAL}} \times 100$$

- I_{LOSS}: is an estimation of obtaining false non-compliance and, thus, of the risk of economic losses due to assignment errors.

$$\mathrm{I_{LOSS}}\ (\%) = \frac{\mathrm{e_{T}}}{\mathrm{TOTAL}} \times 100$$

However, for the proper validation of the multivariate screening method, and estimating of the new indicators it is necessary to take in account a new parameter which never considered before in multivariate analysis, named occurrence (OCURR) [**49**]. The knowledge of this parameter is crucial to verify that the screening method works correctly for the intended purpose. The occurrence indicates the proportion of samples that have the characteristic of the target class with respect to the total population, so, it is a population parameter and the most important thing is a parameter imposed on the experimenter, that is, it cannot be modified, and it is also difficult to know. This parameter is defined with following equation:

$$\frac{\mathrm{Tot\ T}}{\mathrm{Tot\ nT}} \approx \frac{\mathrm{OCURR}}{1 - \mathrm{OCURR}}$$

where 'Tot T' and 'Tot nT' symbolise respectively the number of target and non-target samples in the validation set.

Thus, once the validation requirements have been decided and the value of the occurrence is known, the minimum acceptance criteria of the performance metrics of the classification/screening method (precision (PREC), sensitivity (SENS) and specificity (SPEC) are established based on them [**49**].

$$\mathrm{PREC} = 1 - \frac{\mathrm{I_{ERROR}}}{\mathrm{I_{SAVING}}}$$

$$\mathrm{SENS} = (\mathrm{I_{SAVING}} - \mathrm{I_{ERROR}}) \times \frac{1}{\mathrm{OCURR}}$$

$$\mathrm{SPEC} = 1 - \mathrm{I_{ERROR}} \times \frac{\mathrm{PREC}}{1 - \mathrm{OCCUR}}$$

As it has been statement before the SPEC is a validation parameter which is habitually stated, but it is linked to PREC and does not provide significant information for validation purpose of screening methods and could be left out.

Consequently, in order to make the decision on the validity of a classification method, the applicability indicators can be considered as mandatory pre-step to be carried out before carrying out the validation process. Therefore, fit-for-purpose method requirements should be set once the applicability indicators have been fixed which involve a risk. In this sense, the steps to perform the validation stage of a classification method would be those reflected in Figure 4.

Figure 4. More appropiate scenario of the validation process of an analytical method

It is important to highlighted that is necessary to carry out a modification of the step 3 'analytical method application', that is to say in the development of the classification method. As it has been declared in the section 2, the method development involves splitting the data in two set: training set and external validation set. Traditionally the training set has been made up of 70-80% of the total samples and the remaining one has been part of the validation set. Now, the occurrence value must also be taken into account to set the number of samples belonging to the target and non-target class that make up the model validation set, on this way the model will be validated in real conditions.

Concerning this Jimenez-Carvelo and Cuadros-Rodríguez published a recent paper in which they applied the occurrence parameter to validate a multivariate classification-based screening method to authenticate olive oil [**49**]. In the work they set out to have a saving index of 67% *i.e*. the confirmation method goes 1 out of 3 samples and an error index of 5%, that is, it is accepted that 1 out of 20 samples assigned as genuine oil it could be adulterated. Once these model applicability requirements were decided, the occurrence in order to establish the criteria for accepting the method was necessary to take into account. In the case of the authentication of olive oil, the occurrence is 90%, a value of 90% could be considered appropriate in the worst case, *i.e*. of all the oils which will be analysed in the laboratories, at least 90% are actually olive oil while less than 10% is adulterated olive oil. Based on this, the minimum acceptation criteria of precision and sensitivity were set to 0.93 and 0.68 respectively. After that, their challenge was how do to validate the method with this value of

occurrence that is imposed? In this case they analysed 207 samples of oils (87 olive oil and 112 olive oils adulterated) using HPLC-DAD and GC-FID, so for the methods development from data obtained by the two analytical techniques, 147 samples were used for training of which 50 was genuine extra virgin olive oils and 97 adulterated olive oils and for validation 50 samples were used and taking into account the occurrence value 45 were olive oils (*i.e*. 90 %) and 5 were adulterated olive oils (10%). This work displayed a fast and easy methodology to validate a multivariate analytical method of screening, which could be assumed as vanguard method considering the real conditions and the scenario where the method must work.

6. Recent applications of fingerprinting methods

Due to the rising interest in the applications of the fingerprinting methodology, abundant literature has been recently published addressing the application of these analytical techniques for the resolution of problems the authenticity and traceability in different food matrices. In this section a detailed description of the reviews published for the past ten years, from 2010, based on the analytical techniques (vertical approach) is now presented.

6.1. Mass and ionic mobility spectrometries

A review of mass spectrometric fingerprints for food fraud analysis was published in 2018 by Cavanna *et al.* [**77**]. This paper was devoted to review the status of this methodology in the detection of fraud in different food commodities. It includes 48 references of different mass spectrometry modalities between 2011 and 2018. The most widely used were LC-MS (60%), followed by DART-MS (15%). Although, both GC-MS and PTR-MS were also significantly used (11%). Finally MS (6%) and REIMS-MS, MALDI-MS, LESA-MS, SPME-MS were the less employed modalities. Different food commodities were studied: meat (17%), honey (10%), cereals (12%) extra-virgin olive oil (8%), wine (8%), spices and fruit juices (6%) and finally coffee, cheese, fish and vinegar (less than 5%). Besides, the chemometric applied tools were: PCA (77%), PLS-DA (52%), LDA (14%), SIMCA (8%), ANN, kNN and HCA (4%).

The application of AMS in detecting food fraud was reviewed in 2016 [**27**]. DART was the most widely mode used for food authentication, mainly focused on the analysis of meat, fish, milk, dairy products, oils, cereals, fruit, cereals and drink. Notice that an interesting table gathering food authentication issues that have been addressed by AMS and other analytical techniques since 2009 is included.

Ion mobility spectrometry is now being implemented as analytical suitable for application in food authentication studies, especially due to the appearance of news commercial equipment. In 2019, Hernández-Mesa *et al.* [**28**], published a review with 139 references of this technique. Fingerprinting methodology has been applied in nine of them. Olive oil was analysed in four of these applications while green tea, Iberian ham, honey and white wine were the other foods analysed. The instrumental configurations of IMS were HS-GC-IMS (50%), GC-IMS (20%), UV-IMS (20%), PP-IMS (10%). As chemometric tools, PCA (29%), LDA (29%), kNN (24%), PLS-DA (14%) and SVMC (48%) were used.

6.2. Molecular spectrometries

In 2018 Esteki *et al*. [**78**] published a complete reviewing about spectroscopic methods in combination with LDA for authentication of food products. This review collects applications of these techniques: fluorescence (21%), NMR (20%), NIR (20%), Raman (15%), UV (10%), FT-IR (9%) and HSI (5%) and the food commodities analysed were: oils (34%), meat (24%), milk (21%), coffee (9%), wine (8%) and juice (4%).

NMR fingerprints methodologies are widely recognised as important tools for food authentication. In a review published in 2019 by Hatzakis *et al.* [**29**], it is stated how NMR fingerprint methodologies are applied in routine food fraud control. Food commodities matrices that are habitually authenticated by applying NMR, as coffee, honey, oil, wine, vinegar and meat, are discussed. Comparison between high-resolution NMR and benchtop (low-resolution) NMR is also done and it is concluded that high-resolution NMR had been widely used for fingerprint methodology. As chemometric tools, PCA was again the most used (39%), followed by PLS-DA (32%), LDA (21%), SIMCA and kNN (4%).

Two interesting reviews of the applications of vibrational spectrometric fingerprints (FTIR, NIR, Raman) published in 2015 by Cozzolino [**79**] and 2018 by McGrath *et al.* [**80**]. On the one hand, Cozzolino's review is focused on the use of vibrational spectroscopy as a tool to assess economically motivated fraud and counterfeit issues in agricultural products and foods. On the other hand, McGrath's review sum up scientific publications carried out between 2011 and 2017, which are focused on optical spectrometric techniques. They emphasise three approaches when these instrumental techniques are applied: detection of adulteration; differentiation between food commodity types; and proving geographic origin. They concluded that the food commodity with the greatest number of applications was dairy products, followed by oil and meat (32%, 18% and 13%, respectively). Finally honey, drinks, alcohol, spices and nuts also have several applications. In this research 112 publications were

reviewed: 44% were from FTIR, 42% from NIR and 28% from Raman techniques. It indicates that the three optical techniques are widely implanted in the spectroscopic fingerprint methodology. The spectra subjected to fingerprint analyses contain information that needs to be extracted using a wide variety of chemometric methods. Among them, PCA (71%), PLS-DA (40%), SIMCA (24%), LDA (24%) and SVMC (13%) were the most applied. Tools such as kNN, OCPLS and ANN were found to scarcely being applied.

The applications of NIR spectroscopy were again reviewed in 2019 by Mendez *et al.* [**81**] where the applications of an emerging technique such as hyperspectral imaging were also considered. This review was focused on six food groups: vegetable oils, coffee beans, cereals products, meat, fish and honey. 70 NIR applications were reviewed and the most used chemometric tool was PLS-DA (34%) while PCA was surprisingly found to be the least used technique (7%). On the other hand, in HIS applications the most used tool was SVMC (28%). However, it should be noted that the number of HIS fingerprinting applications has increased sharply in recent years. In a 2016 review of this technique by Siche *et al.* [**82**], showed that there was still little use of chemometric tools to extract information from the obtained images. Research was mainly done with RGB image parameters (profiling approach).

A non-relative widespread technique is THz spectroscopy. In 2019, Afsah-Hejri *et al.* [**31**] made a review of its applications in food inspection, food quality control and adulteration detection. Multiple examples of different application of the technique during past two decades were considered and shown in tables. Only a few of these applications are based on fingerprinting in food commodities such as water, tea, milk, milk powder, chocolate, rice, cereals, chicken meat, honey, virgin olive oil and other vegetable oils. Analytical technique had also been used to detect contaminants such as aflatoxin and different families of antibiotics (tetracyclines, quinolones). Likewise, the chemometric tools that were used: PLS (24%), PCA (19%), SVM (17%), ANN (9%), RF, PLS-DA, GA (6%).

Molecular fluorescence spectroscopy, as explained in section 3, could be used as isolated analytical technique suitable for generating fingerprint. In 2011, a review was published by Karoui *et al.* [**83**] showing some applications of fingerprinting for measuring the quality of food of animal origin (milk, meat, fish and eggs) and vegetable origin (oils, cereals, sugar, fruits and vegetables). In this case PCA, PLS, DA and PARAFAC were applied. Both emission and excitation spectra were used to obtain the fluorimetric fingerprints. In 2016 a review was published by Tsuta *et al.* [**84**] defining the fluorescence fingerprint as an excitation-emission matrix acquired at consecutive excitation wavelengths, producing a three-dimensional diagram. They also collect some

applications on fruits, honey, meat and bread where fluorescent images were used as fingerprints. PCA, PLS and SIMCA tools were used.

6.3. Gas and liquid chromatographies

Chromatographic techniques have been the most used techniques to acquire a fingerprint in food field and, therefore, the greatest numbers of publications are found. In 2016, Cuadros *et al.* published a tutorial about chromatographic fingerprints, being its main goal to clarify the difference between the 'profiling' and 'fingerprinting' terms. In addition, a new term, named 'food identitation', is introduced [**7**]. This describes several chromatographic fingerprinting applications related to the authenticity of different foodstuffs such as: edible oils (26% GC and 20% HPLC), tea (15% GC and 43% HPLC), fruits (15% GC and 14% HPLC), alcoholic beverages (15% GC and 20% HPLC), coffee (8% GC) and finally cheese, mushrooms, rice, honey (less than 5%). In GC the main principal detectors employed were MS (72%) and FID (28%), whereas HPLC were UV (33%), DAD (30%), MS (23%) and CAD (10%). Furthermore, the chemometrics applied were PCA (35%), PLS (21%), PLS-DA (16%), SIMCA (8%), HCA (7%), and with less than 5%: LDA, SVM, DA, kNN, and RF.

The use of chromatographic methods coupled to chemometrics for food authentication was also reviewed in 2018 by Esteki et al [**50**]. This paper devotes the applications of the GC and HPLC for authentication of different food commodities: edible oils (19% GC and 26% HPLC), dairy product (10% GC and 6.5% HPLC), honey (12% GC and 15% HPLC), fruit juices (8% GC, and 6.5% HPLC), alcoholic beverages (17% GC and 17% HPLC), coffee (10% GC and 13% HPLC) and finally meat, fruit and vinegar (less than 5%). In GC the main principal detectors used were MS (74%) and FID (22%), whereas HPLC have been DAD (33%), UV (26%), MS (23%) and CAD (13%). In addition, the chemometrics applied were: PCA (41%), LDA (15%), PLS-DA (13%), PLS (11%), SIMCA (8%), HCA (6%), and with less than 5%: SVM, DA, SIMCA, kNN, ANN and RF.

A new review on the application of gas chromatography was reported by the same authors in 2019 [**85**] collecting the main applications of this technique for solving problems on food authentication using fingerprinting. It shows the percentage of papers devoted to each food commodity in recent years. The ranking is headed by edible oils and fats (with 41%), followed by fruits (17%), coffee and tea (11%), and with lower percentages (below 10%) are wines, juices, dairy products, honey, meat and nuts. Authors show graphically the annual increase in the number of applications based on GC fingerprinting merged with chemometrics in food authentications. The study shows that mass spectrometer was the most use coupled measuring device. It has been used

three times more than FID. Among the chemometric tools that used, the PCA was the one with the highest number of applications (with 54%), follow by PLS and LDA (20% and 13%, respectively) and finally in this order SVM, ANN, SIMCA, HCA, MLR and RF (less than 5%).

A review focussed on the detection of food adulteration using HPLC-MS fingerprinting was published in 2016 by Knolhoff *et al*. [**86**]. This paper outlined the current methodologies and challenges for sample preparation, mass spectrometry, chromatographic and chemometric data processing tools. In 2018, newly Esteki *et al.* published a review devoted to liquid chromatography [**87**]. The same bibliometric studies as in the previous article were carried out. The conclusions showed that the most studied food commodities are fruits (23%), followed by coffee and tea (18%), edible oils and fats (16%), dairy products (14%) and finally honey, wine, juices, meat and nuts (less than 10%). Again graphically, they showed the rise of HPLC applications in the food authenticity field in recent years. Greater variety of measuring analytical techniques than GC has been employed. The most used one was DAD (38%), closely followed by MS (25%) and UV-vis absorption at fixed wavelength (19%). It is worth mentioning the increase applications of CAD detector in recent years, becoming the fourth placed (6%). Again PCA and PLS were the most applied chemometric tools for HPLC fingerprinting (59% and 20%, respectively) followed by LDA and SIMCA (6%), while SVM, kNN, HCA, MLR and RF were the least used.

In addition, an overview of the applications of TLC fingerprints was presented in 2013 by Milojkovic-Opsenica *et al.* [**88**]. The most widely used chemometrics tools and their main advantages and disadvantages were considered. Among the applications, only a few ones are devoted to food analysis methodology has been mainly focused on other fields, especially pharmaceutical analysis.

7. Final remarks and future prospects

The complexity of analytical food authentication has resulted in a unique interdisciplinary research field attracting scientists from the areas of food and plant science, molecular biology, analytical chemistry, etc. Consequently, targeted methods based on a few components (chemical markers) still have much to offer in some very particular cases, in which there are very characteristic markers, but it is increasingly acknowledged that food is a complex matrix and should thus be treated and analysed using techniques that can embrace this complexity. Future trends in analytical chemistry on food field, especially on food fraud, will rely on the implementation of non-targeted fingerprinting methods obtained by high informative techniques such as MS and

NMR fingerprints that provide additional information than simply compliance or non-compliance.

Main advantages of the non-targeted fingerprinting approaches are ability to detect multiple small changes in the foodstuff and to extract these changes as valuable information through advanced multivariate statistics and mathematical techniques with the purpose of authentication, differentiation or classification of samples.

Apart from the numerous opportunities and benefits of the food fingerprinting approaches, these applications are suffering a major handicap. For routine analysis, and, in particular, for official control purposes in food surveillance, it is essential that analytical procedures are validated and become legally incontestable. However, the application and validation of non-targeted standard methods, using multivariate techniques, is not yet covered by current legislation, which certainly explains their limited applicability for routine analysis and official control purposes today [**10**].

Besides, there is little information concerning the best sample extraction and analytical methods for food authenticity purpose, which together with suitable multivariate approach will allow us to reach valid conclusions from data that otherwise, would be very hard to understand. This makes clear that joint effort is necessary in official food control to provide solutions for the particular challenges and to furthermore establish these techniques for the consumer protection and/or for quality control purposes.

Indeed, this chapter shows not only the potential of the non-targeted food fingerprinting approach, but also a brief explanation of the basis for developing multivariate qualitative and quantitative methods. Moreover, the conventional validation process and new approach for establishing the validation requirements of the qualitative screening methods are also included and discussed.

8. Abbreviations and acronyms

2D	Two dimensional
AMS	Ambient mass spectrometry
ANN	Artificial neural network
ASAP	Atmospheric pressure solid analysis probe
CAD	Charged aerosol detector
CART	Classification and regression tree
CE	Capillary electrophoresis
DAD	Diode array detector
DART	Direct analysis in real time
DART-MS	Direct analysis in real time - mass spectrometry
DESI	Desorption electrospray ionisation
DIMS	Direct infusion (or injection) mass spectrometry
EEM	Excitation and emission matrix
ELSD	Evaporative light scattering detector
FID	Flame ionisation detector
FTIR	Fourier-transform infrared spectroscopy
GA	Genetic algorithm
GC	Gas chromatography
GC-FID	Gas chromatography-flame ion detector
GC-IMS	Gas chromatography-ion mobility spectrometry
GC-MS	Gas chromatography-mass spectrometry
HCA	Hierarchical cluster analysis
HIS	Hyperspectral imaging spectrometry
HPLC	High performance liquid chromatography
HPLC-DAD	High performance liquid chromatography-diode array detector
HS-GC-IMS	Headspace-gas- chromatography-ion- mobility spectrometry
IMS	Ion mobility spectrometry
IM-MS	Ion mobility-mass spectrometry
IR	Infrared spectrometry
kNN	k-nearest neighbors

LC	Liquid chromatography
LC-DAD	Liquid chromatography-diode array detector
LC-MS	Liquid chromatography-mass spectrometry
LESA-MS	Liquid extraction surface analysis-mass spectrometry
LDA	Lineal discriminant analysis
MALDI	Matrix assisted laser desorption ionisation
MALDI-MS	Matrix assisted laser desorption ionisation-mass spectrometry
MCR-ALS	Multivariate curve resolution-alternating least squares
MIR	Mid infrared spectrometry
MLR	Multiple lineal regression
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
NIR	Near infrared spectrometry
NMR	Nuclear magnetic resonance
OCPLS	One-class partial least squares
PARAFAC	Parallel factor analysis
PCA	Principal component analysis
PDA	Photometric diode array
PDD	Pulsed discharged detector
PDO	Protected designation origin
PLS-DA	Partial least squares-discriminant analysis
PLSR	Partial least squares-regression
PP-IMS	Positive photoionisation-ion mobility spectrometry
PTR-MS	Proton transfer reaction-mass spectrometry
REIMS-MS	Rapid evaporative ionisation mass spectrometry-mass spectrometry
RF	Random forest
RGB	Red/green/blue
SFIT-MS	Selected ion flow tube-mass spectrometry
SIMCA	Soft independent modelling by class analogy
SPME-MS	Solid-phase microextraction-mass spectrometry
SRI-MS	Selective reagent ionisation-mass spectrometry

SVM	Support vector machine
SVMC	Support vector machine classification
THz	Terahertz spectrometry
TLC	*Thin-layer* chromatography
UV	*Ultraviolet*
UV-IMS	Ultraviolet photoionisation-ion mobility spectrometry
UV-Vis	Ultraviolet-visible spectrometry

References

[1] Bádia-Melis R, Mishra P, Ruiz-Gracía L. Food traceability: New trends and recent advances – A review. Food Control 2015; 57:393-401.

[2] Danezis GP, Tsagkaris AS, Camin F, Brusic V, Georgiuo CA. Food authentication: Techniques, trends & emerging approaches. Trends Anal Chem 2016; 57:123-132.

[3] Soon JM, Krzyzaniak SC, Shuttlewood Z, Smith M, Jack L. Food fraud vulnerability assessment tools used in food industry. Food Control 2019; 101:225-232.

[4] Zederkopff Ballin N, Holst Laursen K. To target or not to target? Definitions and nomenclature for targeted versus non-targeted analytical food authentication. Trends Food Sci Tech 2019; 86:537-543.

[5] Booksh KS, Kowalski BR. Theory of analytical chemistry. Anal Chem 1994; 66:782A-791A.

[6] Medina S, Perestrelo R, Silva P, Pereira JAM, Câmara JS. Current trends and recent advances on food authenticity technologies and chemometric approaches. Trends Food Sci Tech 2019; 85:163-176.

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

[8] Bagur-González MG, Pérez-Castaño E, Sánchez-Viñas M, Vázquez-Evangelista D. Using the liquid-chromatographic-fingerprint of sterols fraction to discriminate virgin olive from other edible oils. J Chromatogr A 2015; 1380:64-70.

[9] Medina S, Pereira JAA, Silva P, Perestrelo R, Câmara JS. Food fingerprints – A valuable tool to monitor food authenticity and safety. Food Chem 2019; 278:144-162.

[10] Esslinger S, Riedl J, Fauhl-Hassek C. Potential and limitations of non-targeted fingerprinting for authentication of food in official control. Food Res Int 2014; 60:189-204.

[11] Cubero-Leon E, Peñalver R, Maquet A. Review on metabolomics for food authentication. Food Res Int 2014; 60:95-107.

[12] Ellis DI, Brewster VL, Dunn WB, Allwood JW, Golovanov AP, Goodacre R. Fingerprinting food: current technologies for the detection of food adulteration and contamination. Chem Soc Rev 2012; 41:5706-5727.

[13] Ahad T, Nissar J. Fingerprinting in determining the adulteration of food. J Pharmacogn Phytochem 2017; 6:1543-1553

[14] Gao B, Holroyd SE, Moore JC, Laurvick K, Gendel SM, Xie Z. Opportunities and challenges using non-targeted methods for food fraud detection. J Agric Food Chem 2019; 67:8425-8430.

[15] Zhang J, Zhang X, Dediu L, Victor C. Review of the current application of fingerprinting allowing detection of food adulteration and fraud in China. Food Control 2011; 22: 1126-1135.

[16] Grauwet T, Vervoort L, Colle I, Van Loey A, Hendrickx M. From fingerprinting to kinetics in evaluating food quality changes. Trends Biotechnol 2014; 32:125-131.

[17] Brereton P, editor. New analytical approaches for verifying the origin of food. Cambridge: Academic Press; 2013.

[18] de la Guardia M, Gonzálvez A, editors. Food protected designation of origin – Methodologies and applications. Amsterdam: Elsevier; 2013.

[19] Haynes A, editor. Advances in food analysis research. New York: Nova Science Publishers; 2015.

[20] Downey G, editor. Advances in food authenticity testing. Duxford: Elsevier; 2016.

[21] Georgiou CA, Danezis GP, editors. Food authentication – Management, analysis and regulation. Chichester: John Wiley & Sons; 2017.

[22] Morin JF, Lees M. Food integrity handbook. A guide to food authenticity issues and analytical solutions. 1st. ed. Nantes: Eurofins Analytics France; 2018.

[23] Da-Wen S, editor. Modern techniques for food authentication. 2nd. ed. London: Academic Press / Elsevier; 2018.

[24] Montet D, Ray RC, editors. Food traceability and authenticity – Analytical techniques. Boca Raton: CRC Press; 2018.

[25] Siddiqi K.S, Nollet LM, editors. Fingerprinting techniques in food authentication and traceability. Boca Raton: CRC Press; 2019.

[26] Zhong J, Wang X, editors. Evaluation technologies for food quality. Duxford: Academic Press / Elsevier; 2019.

[27] Black C, Chevallier OP, Elliott CT. The current and potential applications of ambient mass spectrometry in detecting food fraud. Trends Anal Chem 2016; 82:268-278.

[28] Hernández-Mesa M, Ropartz D, García-Campaña AM, Rogniaux H, Dervilly-Pinel G, Le Bizec B. Ion mobility spectrometry in food analysis: principles, current applications and future trends. Molecules 2019; 24:2706; 1-28.

[29] Sobolev AP, Circi S, Capitani D, Ingallina C, Marinna L. Molecular fingerprinting of food authenticity. Curr Opin Food Sci 2017; 16:59-66.

[30] Hatzakis E. Nuclear Magnetic Resonance (NMR) spectroscopy in food science: a comprehensive review. Compr Rev Food Sci Food Saf. 2019; 18:189-220.

[31] Afsah-Hejri L, Hajeb P, Ara P, Ehsani RJ. A comprehensive review on food applications of terahertz spectroscopy and imaging. Compr Rev Food Sci Food Saf 2019; 18:1563-1621.

[32] Airado-Rodríguez D, Durán-Merás I, Galeano-Díaz T, Wold JP. Front-face fluorescence spectroscopy: new tool for control in the wine industry. J. Food Compos Anal 2011; 24:257-264.

[33] Hu L, Ma S, Yin C. Discrimination of geographical origin and detection of adulteration of kudzu root by fluorescence spectroscopy coupled with multi-way pattern recognition. Spectrochim Acta A 2018; 193:87-94.

[34] Amigo JM, Martí I, Gowen A. Hyperspectral imaging and chemometrics: A perfect combination for the analysis of food structure, composition and quality. In Marini F, editor. Chemometrics in food chemistry. Elsevier; 2013, ch. 9, p. 343-370.

[35] Kamruzzaman M, Makino Y, Oshita S. Parsimonious model development for real-time monitoring of moisture in red meat using hyperspectral imaging. Food Chem 2016; 196; 1084-1091.

[36] Oliveira MM, Cruz-Tirado JP, Barbin DF. Nontargeted analytical methods as a powerful tool for the authentication of spices and herbs: a review. Compr Rev Food Sci Food Saf 2019; 18:670-689.

[37] Shmueli G. To explain or to predict? Stat Sci 2010; 25:289-310.

[38] Sriboonchitta S, Longpré L, Kreinovich V, Dumrongpokaphan T. Why the best predictive models are often different from the best explanatory models: a theorical explanation. In: Kreinovich V, Sriboonchitta S, editors. Structural Changes and their Econometric Modeling. TES 2019 Conference Proceedings. Cham: Springer Nature; 2019. p. 163-171.

[39] Jiménez Carvelo AM, González Casado A, Bagur González MG, Cuadros Rodríguez L. Alternative data mining/machine learning methods for the analytical evaluation of food quality and authenticity. Food Res Int 2019; 122:25-39.

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

[41] Westad F, Marini F. Validation of chemometric models – A tutorial. Anal Chim Acta 2015; 893:14-24.

[42] Esbensen KH, Geladi P. Principes of proper validation: use and abuse or re-sampling for validation. J Chemom 2010; 24:168-187.

[43] Brereton RG. Pattern recognition in chemometrics. Chemom Intell Lab Syst 2015; 149:90-96.

[44] Cuadros-Rodríguez L, González-Casado A, Ruiz-Samblás C, Bagur-González MG. Evolution of the quality concept in analytical laboratories. Encyclopedia of Analytical Chemistry. Online: John Wiley & Sons; 2017. p. 1-54.

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

[46] Valcárcel M, Cárdenas S. Vanguard-rearguard analytical strategies. Trends Anal Chem 2005; 24:67-74.

[47] Boggia R, Casolino MC, Hysenaj V, Oliveri P, Zunin P. A screening method based on UV-visible spectroscopy and multivariate analysis to assess addition of filler juices and water to pomegranate juices. Food Chem 2013; 140:735-741.

[48] López MI, Trullols E, Callao MP, Ruisánchez I. Multivariate screening in food adulteration: untargeted versus targeted modelling. Food Chem 2014; 147:177-181.

[49] Jiménez-Carvelo AM, Cuadros-Rodríguez L. The occurrence: a meaningful parameter to be consider in the validation of multivariate classification-based screening methods – Application for authenticating virgin olive oil. Talanta 2020; 208:120467.

[50] Esteki M, Simal-Gandara J, Shahsavari Z, Zandbaaf S, Dashtaki E. A review on the application of chromatographic methods, coupled to chemometrics, for food authentication. Food Control 2018; 93:165-182.

[51] Kumar N, Bansal A, Sarma GS, Rawal RK. Chemometrics tools used in analytical chemistry: an overview. Talanta 2014; 123:186-199.

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

[53] Brereton RG, Lloyd GR. Partial least squares discriminant analysis: taking the magic away. J Chemom 2014; 28:213-225.

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

[55] Brereton RG, Lloyd GR. Support vector machines for classification and regression. Analyst 2010; 135:230-267.

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

[57] Maione C, De Paula SP, Gallimberti M, Lemos Batista B, Dobal Campiglia A, Barbosa Jr F, Baborsa RM. 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.

[58] Bachion de Santana F, Borges Neto W, Poppi RJ. Random Forests as one-class classifier and infrared spectroscopy for food adulteration detection. Food Chem 2019; 293:323-332.

[59] Brereton RG. Introduction to multivariate calibration in analytical chemistry. Analyst 2000; 125:2125-2154.

[60] Jiménez-Carvelo, AM, González-Casado A, Cuadros Rodríguez, L. 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.

[61] Wold S, Sjöström, Eirksson L. PLS-regression: a basic tool for chemometrics. Chemom-Intell Lab Syst 2001; 58:109-130.

[62] Escandar GM, Olivieri AC. Multi-way chromatographic calibration – A review. J Chromatogr A 2019; 1587:2-13.

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

[64] Westad F, Marini F. Validation of chemometric models – A tutorial. Anal Chim Acta 2015; 893:14-24.

[65] Alewijn M, Van der Voet S, Van Ruth S. Validation of multivariate classification methods using analytical fingerprints – concept and case study on organic feed for laying hens. J. Food Compos Anal 2016; 51:15-23.

[66] FCC 11, Appendix XVIII. Guidance on developing and validating non-targeted methods for adulteration detection. Food Chemicals Codex, 12th ed; 2020.

[67] TR No 01/2015, Guide to NMR Method development and validation – Part II: Multivariate data analysis. EUROLAB; 2015.

[68] Gallo V, *and other 54 more*. A Contribution to the harmonization of non-targeted NMR methods for data-driven food authenticity assessment. Food Anal Meth 2020; 13:530-541.

[69] USP 43 – NF 38, <1039> Chemometrics. United States Pharmacopeia (2019).

[70] PhEur 5.21 – Chemometrics methods applied to analytical data. European Pharmacopoeia 9.8 (2018), p. 783-801.

[71] Riedl J, Esslinger S, Fauhl-Hassek C. Review of validation and reporting of non-targeted fingerprinting approaches for food authentication. Anal. Chim. Acta 2015; 885:17-32.

[72] ASTM E1655-05(2012). Standard practices for infrared multivariate quantitative analysis. ASTM International, West Conshohocken (PA), 2012.

[73] Yuan-Na L, Hai-Long W, Xiang-Dong Q, Quan L, Shu-Fand L, Hai-Yan F, Yong-Jie Y, Ru-Qin Y. Quantitative analysis of triazine herbicides in environmental samples by using high performance liquid chromatography and diode array detection combined with second-order calibration based on an alternating penalty trilinear decomposition algorithm. Anal Chim Acta 2010; 678:26-33.

[74] Carabajal MD, Arancibia JA, Escandar GM. Excitacion-emission fluorescence-kinetic thir-order/four-way data: Determination of bisphenol A and nonylphenol in food contact plastics. Talanta 2019; 197:348-355.

[75] ISO/IEC 17025:2017, General requirements for the competence of testing and calibration laboratories. International Organization for Standardization, Geneva, 2017.

[76] Cuadros-Rodríguez L, Valverde-Som, Jiménez-Carvelo AM, Delgado-Aguilar M. Validation requirements of screening analytical methods based on scenario-specified applicability indicators. Trends Anal Chem 2020; 122:115705.

[77] Cavanna D, Righetti L, Elliott C, Suman M. The scientific challenges in moving from targeted to non-targeted mass spectrometric methods for food fraud analysis: A proposed validation workflow to bring about a harmonized approach. Trends Food Sci Technol 2018; 80:223-241.

[78] Esteki M, Shahsavari Z, Simal-Gandara J. Use of spectroscopic methods in combination with linear discriminant analysis for authentication of food products. Food Control 2018; 91:100-112.

[79] Cozzolino D. The role of vibrational spectroscopy as a tool to assess economically motivated fraud and counterfeit issues in agricultural products and foods. Anal Methods 2015; 7:9390-9400.

[80] McGrath TF, Haughey SA, Patterson J, Fauhl-Hassek C, Donarskic J, Alewijnd M, Van Ruthd S, Elliott CT. What are the scientific challenges in moving from targeted to non-targeted methods for food fraud testing and how can they be addressed? - Spectroscopy case study. Trends Food Sci Technol 2018; 76:38-55.

[81] Mendez J, Mendoza L, Cruz-Tirado J.P, Quevedo R, Siche R. Trends in application of NIR and hyperspectral imaging for food authentication. Sci Agropecu 2019; 10:143-161.

[82] Siche R, Vejarano R, Aredo V, Velasquez L, Saldaña E, Quevedo R. Evaluation of food quality and safety with hyperspectral imaging (HSI). Food Eng Rev 2016; 8:306-322.

[83] Karoui R, Blecker C. Fluorescence spectroscopy measurement for quality assessment of food systems – A review. Food Bioprocess Technol 2011; 4:364-386.

[84] Tsuta M. Research on using fluorescence fingerprints for the evaluation of food quality. J Jpn Soc Food Sci 2016; 63:77-381.

[85] Esteki M, Shahsavaria Z, Simal-Gandara J. Gas chromatographic fingerprinting coupled to chemometrics for food authentication. Food Rev Int 2020; 36:384-427.

[86] Knolhoff AN, Croley TR. Non-targeted screening approaches for contaminants and adulterants in food using liquid chromatography hyphenated to high resolution mass spectrometry. J Chromatogr A 2016; 1428:86-96.

[87] Esteki M, Shahsavaria Z, Simal-Gandara J. Food identification by high performance liquid chromatography fingerprinting and mathematical processing. Food Res Int 2019; 122:303-317.

[88] Milojkovic-Opsenica D, Ristivojevic P, Andric F, Trifkovic J. Planar chromatographic system in pattern recognition and fingerprint analysis. Chromatographia 2013; 76:1239-1247.

CAPÍTULO II

APLICACIÓN DE QUIMIOMETRÍA PARA LA EVALUACIÓN ANALÍTICA DE LA AUTENTICIDAD: MATRIZ DE AGUACATE

CAPÍTULO II

Aplicación de la quimiometría para la evaluación analítica de la autenticidad: matriz de aguacate

II.1. Presentación

Este capítulo recoge los resultados obtenidos al aplicar la metodología de huella instrumental cromatográfica de la fracción grasa del aguacate, obtenida por cromatografía líquida y posterior evaluación de ésta huella mediante aplicación de diversas herramientas quimiométricas de discriminación y clasificación de datos.

Los objetivos de este capítulo fueron:

- Obtener una huella cromatográfica representativa de cada variedad botánica de aguacate en estudio.
- Construir modelos multivariables de clasificación que permitan diferenciar las diferentes variedades botánicas.
- Obtener una huella cromatográfica representativa del origen geográfico de los aguacates en estudio.
- Construir modelos multivariables de clasificación que permitan diferenciar los diferentes orígenes geográficos.

Se trabajó en las dos principales modalidades cromatográficas de cromatografía de líquidos, fase normal y fase invertida. Además, para obtener la huella dactilar cromatográfica se aplicaron dos sistemas de detección diferentes, un sistema universal de detección por aerosol cargado (CAD) y un sistema de detección de fila de diodos en la región del UV-VIS (DAD). Se realizaron estudios y clasificaciones multivarianbles de las huellas dactilares cromatográficas de manera dirigida y no dirigida aplicando diversos algoritmos quimiométricos. Todo esto dio lugar a dos publicaciones científicas:

1. PUBLICACIÓN II: Differentiation of avocados according to their botanical variety using liquid chromatographic fingerprinting and multivariate classification tree.
2. PUBLICACIÓN III: Authentication of the geographical origin and the botanical variety of avocados using liquid chromatography fingerprinting and deep learning methods.

II.2. Introducción

La Real Academia Española define el término autenticidad como la cualidad de ser auténtico [**1**]. El concepto de auténtico, sinónimo de original, verdadero, genuino, indudable, etc. aplicado a los alimentos, certifica que éstos son de un origen determinado y que cumplen con las normas vigentes y con todas las declaraciones incluidas en la etiqueta de presentación: propiedades nutricionales, fecha de producción (año de cosecha), fecha de caducidad, certificaciones y/o logotipos de calidad diferenciada, etc. [**2**]. En cuanto al origen de un alimento, éste se refiere al lugar o región en la que se produjo el alimento o sus ingredientes, así como a la especie botánica y/o a la materia prima utilizada.

El aguacate es una fruta tropical de gran importancia económica, rica en vitaminas A y B, minerales y antioxidantes reconocida como un alimento funcional. Aunque quizás lo que más destaca en las características del aguacate es su alto contenido en ácidos grasos insaturados, con un alto porcentaje en poliinsaturados. Numerosos estudios han indicado que el contenido graso, y su composición, varía en función de la localización de los cultivos, la variedad, el número de días entre la floración y la recolección, el contenido en materia seca e incluso de la parte de la fruta sometida a ensayo [**3**]. Se conocen tres variedades botánicas de aguacate: la *Antillana, Mejicana y Guatemalteca*. Estas razas se pueden diferenciar tanto por sus rasgos morfológicos y fisiológicos como por sus rasgos hortícolas [**4**]. De éstas, por hibridación en la mayoría de los casos fortuita, se han obtenido la mayor parte de las variedades comerciales, entre las que destacan las variedades "Hass", "Bacon" y "Fuerte". Respecto al origen geográfico, aunque éste se remonta a

---

1. Real Academia Española. https://dle.rae.es/autenticidad (accedido 05 mayo 2022).
2. Sotirchos, G.G., Danezis, G.P., Georgiou, C.A. Introduction, definitions and legislation (chapter 1). In C.A. Georgiou and P. Georgios (Eds.), Danezis Food Authentication Management, Analysis and Regulation, John Wiley & Sons, Nueva Jersey, 2017.
3. Ferreyra, R., Sellés, G., Saavedra, J., Ortiz, J., Zúñiga, C., Troncosoa, C., Rivera, S.A., González Agüero, M., Defilippi, B.G. Identification of pre-harvest factors that affect fatty acid profiles of avocado fruit (Persea americana Mill) cv. 'Hass' at harvest. S. Afr. J. Bot., 104 (2016), 15-20. https://doi.org/10.1016/j.sajb.2015.10.006
4. Alcaraz, M.L.; Hormaza, J.I. Molecular characterization and genetic diversity in an avocado collection of cultivars and local Spanish genotypes using SSRs. Hereditas, 144 (2007), 244-253. https://doi.org/10.1111/j.2007.0018-0661.02019x

América Central, en la actualidad el aguacate es cultivado ampliamente en todo el mundo [**5**].

El aguacate posee un rendimiento graso, de aceite, similar a la aceituna, y una proporción relativa similar de ácidos grasos, predominando en ambos el ácido oleico. El aceite de aguacate es rico en ácidos grasos omega que son buenos para la salud humana, especialmente en la prevención de enfermedades cardiovasculares [**6**]. Sin embargo, por la heterogeneidad que presenta la matriz lipídica del aguacate pocos estudios se encuentran en la bibliografía referentes al estudio de su perfil de triglicéridos (TG), lo cuales constituyen sus componentes mayoritarios [**7**]. No se ha localizado ninguno enfocado al análisis de las diferencias entre variedades, estudios que sí existen para otros alimentos vegetales de alto valor económico como el aceite de oliva [**8**]. La mayoría de los autores, se han limitado a liberar y estudiar el perfil de ácidos grasos del aguacate, generalmente tras metilación, por cromatografía de gases. Esta manera de actuar, proporciona una información parcial y no permite conocer la variabilidad que se crea en la combinación de tres ácidos grasos en cada triglicérido [**9**].

Hasta el momento varias técnicas cromatográficas se han utilizado con éxito para la determinación cualitativa y cuantitativa de TG, la cromatografía de gases (GC) [**10**], cromatografía líquida de altas prestaciones (HPLC) en

---

5. Statistics Division of the Food and Agriculture Organization for the United Nations, FAOSTAT data, 2022. http://www.fao.org/faostat/es/#data/QC4, (accedido 05 mayo 2022).
6. Krumreich, F.D., Borges, C.D., Mendonça, C.R.B., Jansen Alves, Zambiazi, R.C. Bioactive compounds and quality parameters of avocado oil obtained by different processes. Food Chem., 257 (2018), 376-381. https://doi.org/10.1016/j.foodchem.2018.03.048
7. Hierro, M.T.G., Tomás, M.C., Fernández Martín, F., Santa María, G. Determination of the triglyceride composition of avocado oil by high-performance liquid chromatograpgy using a light-scattering detector. J. Chromatogr., 607, 2 (1992), 329-338. https://doi.org/10.1016/0021-9673(92)87092-m
8. Vera, D.N., Jiménez Carvelo, A.M., Cuadros Rodríguez, L., Ruisánchez, I., Callao, M.P. Authentication of the geographical origin of extra-virgin olive oil of the Arbequina cultivar by chromatographic fingerprinting and chemometrics. Talanta, 203 (2019), 194-202. https://doi.org/10.1016/j.talanta.2019.05.064
9. Bora, P.S., Narain, N., Rocha, R.V.M., Paulo, M.Q. Characterization of the oils from the pulp and seeds of avocado (cultivar: Fuerte) fruits. Grasas y aceites, 52 (2001), 171-174. https://doi.org/10.3989/gya.2001.v52.i3-4.353
10. Andrikopoulos, N.K., Giannakis, I.G., Tzamtzis, V. Analysis of olive oil and seed oil triglycerides by capillary gas chromatography as a tool for the detection of the adulteration of olive oil. J. Chromatograp. Sci., 39 (2001), 137-145. https://doi.org/10.1093/chromsci/39.4.137

modalidad de trabajo fase normal (NP) [**11**] y fase invertida (RP) [**12**, **13**], cromatografía en capa fina (TLC) y cromatografía de fluido supercrítico (SFC) [**14**]. En este capítulo nos centraremos en el uso de HPLC, en sus dos modalidades de trabajo.

HPLC es un proceso de separación basado en la selectividad de la interacción de los solutos entre una fase estacionaria y una fase móvil. La fase estacionaria tiene una adecuada fase "líquida" depositada como una película delgada, o químicamente unida a la superficie de un soporte sólido inerte. En su modalidad de trabajo en fase normal, NP-HPLC, se caracteriza por el reparto de solutos entre grupos polares de la fase estacionaria y un sistema solvente menos polar que esta fase. El orden de elución de TG en NP-HPLC está determinado por el grado de saturación y por el número de total de carbonos suma de las longitudes de la cadenas de alquilo de los tres residuos de ácidos grasos del triglicérido. De este modo, las especies de TG saturadas se eluyen antes que los homólogos más insaturados. Los sistemas de separación en fase normal tienen un poder de resolución inferior al de los de fase invertida (RP) y por este motivo no se usan comúnmente para las separaciones de triglicéridos. Sin embargo, la cromatografía NP, ya sea en capa fina o en formato HPLC, es una herramienta indispensable para la separación de las diferentes clases de lípidos y con ella los TG pueden aislarse fácilmente en presencia de otros grupos de lípidos como diacilgliceroles, monoacilgliceroles, esteroles, ácidos grasos libres y fosfolípidos [**15**].

---

11. 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 chromatography fingerprint of the methyl-transesterified fraction. Food Chem., 221 (2017), 1784-1791
https://doi.org/10.1016/j.foodchem.2016.10.103

12. De la Mata Espinosa, P., Bosque Sendra, J.M., Cuadros Rodríguez, L. Quantification of triacylglycerols in olive oils using HPLC-CAD. Food Anal. Methods, 4 (2011), 574-581.
https://doi.org/10.1007/s12161-011-9207-3

13. De la Mata Espinosa, P., Bosque Sendra, J.M., Brob, R., Cuadros Rodríguez, L. Olive oil quantification of edible vegetable oil blends using triacylglycerols chromatographic fingerprints and chemometric tools. Talanta, 85 (2011), 177-182.
https://doi.org/10.1016/j.talanta.2011.03.049

14. Buchgraber, M., Ulberth, F., Emons, H., Anklam, E. Triacylglycerol profiling by using chromatographic thechniques. Eur. J. Food Sci. Technol., 106 (2004), 621-648.
https://doi.org/10.1002/ejlt.200400986

15. Buchgraber, M., Ulberth, F., Emons, H., Anklam, E. Triacylglycerol profiling by using chromatographic thechniques. Eur. J. Food Sci. Technol., 106 (2004), 621-648.
https://doi.org/10.1002/ejlt.200400986

En el caso de la modalidad en fase invertida, RP-HPLC, la fase móvil presenta una polaridad mayor que la fase estacionaria: típicamente se emplean mezclas de acetonitrilo con acetona o isopropanol como fase móvil (en el caso de análisis de triglicéridos se usan mezclas relativamente apolares debido a la poca polaridad de éstos) y gel de sílice modificado octadecilsilil (C18) como fase estacionaria. RP-HPLC es la técnica cromatográfica más empleada para separar mezclas de TG. El parámetro que se utiliza para explicar la separación de TGs e identificarlos en RP-HPLC es el "número equivalente de carbono" (ECN) [**16**], que equivale al número de carbonos de las cadenas grasas menos el doble del número de insaturaciones (ECN = CN – 2n).

En cuanto a los sistemas de detección empleados, suelen ser detectores universales, cuyas respuestas sean casi independientes de la estructura del analito, permitiendo así cuantificaciones más rápidas y simples. De hecho, el método estándar de la IUPAC para la determinación de TG emplea el sistema de detección de índice de refracción [**17**]. En este caso, teniendo en cuenta la metodología de "huella instrumental", que se aplicará en los estudios de esta tesis, se emplearon dos sistemas de detección diferentes para tratar de conseguir señales inespecíficas de la fracción grasa del aguacate y cumplir con los objetivos de autenticidad específicos: sistema de detección de aerosol cargado (CAD) y sistema de absorción UV-Vis mediante fila de diodos (DAD).

CAD es un sistema de detección universal en el cual se genera un aerosol de partículas cargadas que son detectadas en un electrómetro y cuya carga total depende únicamente de la composición del eluato. Las aplicaciones de este detector se centran principalmente en el ámbito farmacéutico [**18**] aunque también se pueden encontrar trabajos de su empleo para el análisis de lípidos, triglicéridos y ácidos grasos [**19**]. Además, en el ámbito de la autenticidad, el grupo de investigación posee amplia experiencia en la aplicación con éxito de

16. Lange, M., Ni, Z., Criscuolo, A., Fedorova, M. Liquid chromatography techniques in lipidomics research. Chromatographia, 82 (2019), 77-100. https://doi.org/10.1007/s10337-018-3656-4

17. IUPAC 2.324. Standard Methods for the Analysis of Oils, Fats and Derivatives, 7th ed., Oxford Blackwell Scientific Publication, London, 1992.

18. Wahl, O., Holzgrabe, U. Impurity profiling of ibandronate sodium by HPLC-CAD. J. Pharm. Biomed. Anal., 114 (2015), 254-264. https://doi.org/10.1016/j.jpba.2015.06.002

19. Libong, D., Héron, S., Tchapla, A., Chamide, P. Lipid Analysis with the corona CAD. In: Gamache, P.H. (Ed). Charged Aerosol Detection for Liquid Chromarography and Related Separation Techniques, John Wiley & Sons, Inc, Nueva Jersey, 2017.

este sistema de detección con objetivos similares a otras matrices grasas como el aceite de oliva [**20**,**21**].

DAD se utiliza mayoritariamente para medir componentes de una mezcla que muestran un espectro de absorción diferente en la región del ultravioleta y/o el visible, permitiendo obtener información sobre una amplia gama de longitudes de onda a la vez. Esto es posible debido a que simultáneamente puede registrar el espectro UV-Vis a todas las longitudes de onda en el rango especificado, ya que posee una fila de diodos y cada uno es capaz de registrar la señal proveniente de una longitud de onda. Cabe destacar que sólo los ácidos grasos insaturados (más propiamente, los triacilgliceroles que tienen al menos un ácido graso insaturado, principalmente el palmitoleico, C16:1; el oleico, C18:1; y el linoleico, C18:2) producen una señal medible, ya que los ácidos grasos saturados son casi transparentes en la zona del espectro UV-Vis. Por consiguiente, en este caso la huella digital tiene un alto grado de especificidad con respecto a la distribución de estos ácidos grasos insaturados en la cantidad de los distintos triacilgliceroles y proporcionará información diferente a la obtenida con CAD. DAD ha sido ampliamente utilizado con fines de autenticidad alimentaria de distintas matrices vegetales [**22**].

Cualquier señal analítica obtenida de cualquier sistema de medida puede clasificarse como señal específica o inespecífica. Una señal específica es aquella que contiene la información explícita para poder obtener información de componentes individuales o grupos de compuestos de una misma familia. Cuando utilizamos una técnica cromatográfica, se puede definir como una señal específica cuando se obtiene una buena separación de picos cromatográficos y se dispone de un sistema detector capaz de identificar los compuestos presentes en la muestra. Por el contrario, una señal inespecífica se puede definir como aquella señal no necesariamente clara y unívoca que no permite diferenciar cada uno de los componentes de forma individual, aunque generalmente responde a mayor número de especies químicas (universal).

---

20. De la Mata Espinosa, P., Bosque Sendra, J.M., Bro, R., Cuadros Rodriguez, L. Discriminating olive and non-olive oils using HPLC-CAD and chemometrics. Anal. Bioanal. Chem., 399 (2011), 2083-2092. https://doi.org/10.1007/s00216-010-4366-4

21. 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., 100 (2017), 2, 345-350. https://doi.org/10.5740/jaoacint.16-0411

22. Esteki, M., Shahsavari, Z., Simal Gandara, J. Food identification by high performance liquid chromatography fingerprinting and mathematical processing. Food Res. Int., 12 (2019), 303-317. https://doi.org/10.1016/j.foodres.2019.04.025

La metodología 'fingerprinting' o de "huella instrumental" típicamente utiliza señales inespecíficas, donde se utiliza toda la información que se encuentra de forma implícita en las muestras sin necesidad de identificar ni cuantificar los compuestos presentes en ella. Tiene la principal ventaja de necesitar un menor tratamiento de muestra y emplear métodos de análisis sencillos que proporcionan gran cantidad de información implícita en menores tiempos de análisis. Haciendo uso de técnicas cromatográficas se obtienen cromatogramas inespecíficos a los que se les denomina "huella cromatográfica" donde el cromatograma se trata como un vector de datos en su conjunto. Para ello es necesario recurrir a técnicas avanzadas de tratamiento de datos, como es la quimiometría, que permite extraer toda la información útil contenida en cada señal analítica [**23**]. Los diferentes métodos quimiométricos a aplicar, principalmente métodos de reconocimientos de pautas, tienen como objetivo principal la detección de grupos y la clasificación de objetos, así como el modelado de las relaciones entre diferentes datos analíticos. Dependiendo del punto de partida, se pueden establecer dos tipos de métodos de reconocimiento de pautas: no supervisados y supervisados.

Los métodos no supervisados intentan establecer la existencia de grupos de muestras en la matriz de datos teniendo en cuenta las similitudes entre ellos, sin información previa de tal existencia. Se puede trabajar con este tipo de métodos cuando no se dispone de información previa o cuando sea interesante estudiar el agrupamiento para observar las relaciones naturales de los datos, aunque se disponga previamente de la información. Uno de los más utilizados es el análisis de componentes principales (PCA). El análisis PCA es una de las formas más simples de evidenciar pequeñas diferencias en grandes conjuntos de datos refiriéndose a la varianza de los datos [**24**]. Tiene como objetivos extraer la información más importante del conjunto de los datos y simplificar la descripción del set datos quedándose solo con esta información importante. PCA calcula nuevas variables llamadas componentes principales, combinaciones lineales de las variables originales. El primer componente principal (PC1) es el que tiene la mayor varianza posible (inercia) y, por lo tanto, este componente "explicará" la mayor parte de la inercia de la tabla de

23. Hibbert D.B. Vocabulary of concepts and terms in chemometrics (IUPAC Recommendations 2016). Pure Appl. Chem., 88, 4 (2016), 407-443. https://doi.org/10.1016/10.1515/pac-2015-0605

24. Indelicato, S., Bongiorno, D., Pitonzo, R., Di Stefano, V., Calabrese, V., Indelicato, S., Avellone, G. Triacylglycerols in edible oil: determination, characterization, quantitation, chemometric approach and evaluation of adulterations. J. Chromatogr. A, 1515 (2017), 1-16. https://doi.org/ 10.1016/j.chroma.2017.08.002

datos. El segundo componente (PC2) se calcula bajo la restricción de ser ortogonal a PC1 y tener la mayor inercia posible y así sucesivamente [**25**].

Si "X" es una matriz de datos con "m" filas (objetos o muestras) y "n" columnas (variables), PCA descompone "X" como suma de términos en vectores, de menor o igual dimensión que "X":

- Los vectores "scores" (puntuaciones) contienen la información de cómo las muestras se relacionan unas con otras. Son las nuevas coordenadas de las muestras.
- Los vectores "loadings" (pesos o ponderales) contienen información de cómo las variables se relacionan unas con otras.

PCA puede utilizarse, además, junto a intervalos de confianza para identificar datos anómalos ('outliers'), correspondientes a valores atípicos dentro de un conjunto de objetos.

Por el contrario, las técnicas supervisadas hacen uso de una característica conocida a priori para establecer patrones y reglas que permitan predecir nuevos datos. En estos métodos, se conocen muestras pertenecientes a una clase o grupo particular y el objetivo es establecer modelos de clasificación implícitos o explícitos [**26**]. Normalmente, el conjunto de datos es dividido en dos subconjuntos: un subconjunto de entrenamiento y un subconjunto de validación. Con el subconjunto de entrenamiento (o de calibración) se establece el modelo con el objetivo de establecer las reglas o criterios de clasificación que serán aplicados posteriormente en las muestras de validación y en muestras cuya clase se desee conocer. El subconjunto de validación o de evaluación se utiliza para validar el modelo, normalmente utilizando un conjunto de muestras diferentes a las utilizadas en el conjunto de entrenamiento (validación externa). La mejor o la peor clasificación para cada una de las muestras del conjunto de validación miden la calidad del modelo (capacidad de reconocimiento, previsibilidad, robustez, etc. [**27**]).

Entre los métodos supervisados de reconocimiento de pautas o patrones se encuentran los métodos de análisis discriminante como clasificación K-vecino

---

25. Abdi, H., Williams, L.J. Principal component analysis. Wiley Interdisciplinary Reviews, 2 (2010), 433-459. https://doi.org/10.1002/wics.101
26. Smolinska, A., Blanchet, L., Buydens, L.M.C., Wijmenga, S.S. NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review. Anal. Chim. Acta, 750 (2012), 82-97. https://doi.org/10.1016/j.aca.2012.05.049
27. Cuadros Rodríguez, L., Pérez Castaño, E., Ruíz Samblás, C. Quality performance metrics in multivariate classification methods for qualitative analysis, Trends Anal. Chem., 80 (2016), 612-624. https://doi.org/10.1016/j.trac.2016.04.021

más cercano (K-NN), análisis discriminante de mínimos cuadrados parciales (PLS-DA) y máquinas de soporte vectorial (SVM), y los métodos de modelado de clases como el modelado independiente por analogía de clases (SIMCA). En este trabajo se emplearon los métodos PLS-DA, SVM y SIMCA por lo que una información ampliada sobre ambos se proporciona a continuación.

PLS-DA es una técnica quimiométrica de análisis discriminante utilizada para optimizar la separación entre diferentes grupos de muestras, lo que se logra uniendo dos matrices de datos: X (datos brutos) e Y (grupos, clases, etc.) [**28**]. La clasificación se basa en construir un modelo de regresión PLS con diferentes clases a las que se les asocia un valor numérico distinto, por ejemplo 0 y 1 en el caso de una clasificación binaria. La asignación de clase se realiza prediciendo el valor numérico de la muestra por analogía de clases.

Al igual que el PLS-DA, SVM también es aplicable tanto para resolver problemas de clasificación como de regresión. En clasificación, SVM-C tiene como objetivo obtener el límite óptimo de las diferentes clases en un espacio vectorial independientemente de las distribuciones probabilísticas de los vectores de entrenamiento en el conjunto de datos. SVM-C proyecta los datos del conjunto de entrenamiento en un espacio de características que suele tener una dimensión mucho mayor que el espacio de datos original. En el espacio de características se calcula un hiperplano (plano de decisión) que separa los puntos individuales de pertenencia a un grupo conocido. La mejor separación se consigue maximizando el margen entre los grupos. Los puntos en el espacio de características que definen el margen se denominan vectores de soporte. El margen es la distancia mínima desde el hiperplano de separación hasta los puntos de datos más cercanos. Así, SVM busca un hiperplano de separación óptimo en el que el margen sea máximo. Además, el algoritmo SVM utiliza un conjunto de funciones matemáticas denominadas 'kernel' que transforman los datos originales al formato requerido. Cada 'kernel' tiene un conjunto de parámetros que deben ser ajustados. El uso de la función de base radial (RBF) está particularmente extendida porque sólo requiere un parámetro ajustable (la anchura radial σ) [**29**].

SIMCA es una técnica de clasificación supervisada que usa PCA para la clasificación al crear una región de confianza alrededor de cada clase usando los residuos de las muestras pertenecientes al conjunto de entrenamiento.

---

28. Gromski, P.S., Muhamadali, H., Ellis, D.I., Xu, Y., Correa, E., Turner, M.L., Goodacre, R. A tutorial rewiew: metabolomics and partial least squares-discriminant analysis – a marriage of convenience or a shotgun wedding. Anal. Chim. Acta, 879 (2015), 10-23. https://doi.org/10.1016/j.aca.2015.02.012
29. Xu, Y., Zomer, S., Brereton, R.G. Support vector machine: a recent method for classification in chemometrics. Crit. Rev. Anal. Chem. 36 (2006), 177-188. https://doi.org/10.1080/10408340600969486.

Consiste en una técnica de reconocimiento de patrones en la que los objetos nuevos se proyectan como miembros de una clase particular en función de la distancia euclidiana al espacio de su particular componente principal. Por lo tanto, las clases son modeladas una a una por una serie de estructuras lineales (un punto, una línea, un plano, etc.) dependiendo del número de componentes requeridos para reproducir los datos de la clase. Es posible definir superficies límite alrededor de estas estructuras lineales basadas en los residuos de los datos después de ajustar los componentes. SIMCA también se puede tratar como una prueba de valores atípicos multivariable porque verifica valores atípicos en el espacio de las PC seleccionadas [**30**,**31**].

A continuación, se presentan los artículos publicados, donde se describe de forma más detallada el estudio realizado, los resultados obtenidos y las conclusiones a las que se ha llegado.

30. Hopke, P.K. The evolution of chemometrics. Anal. Chim. Acta, 500 (2003), 365-377. https://doi.org/10.1016/S0003-2670(03)00944-9

31. Kumar, N., Bansal, A., Sarma, G.S., Rawal, R.K. Chemometrics tools used in analytical chemistry: an overview. Talanta, 123 (2014), 186-199. https://doi.org/10.1016/j.talanta.2014.02.003

PUBLICACIÓN II.

ARTÍCULO CIENTÍFICO

Differentiation of avocados according to their botanical variety using liquid chromatographic fingerprinting and multivariate classification tree

Journal of the Science of Food and Agriculture, 99, 4932-4941, 2019.
https://doi.org/10.1002/jsfa.9725

Research Article SCI

Received: 1 February 2019 Revised: 31 March 2019 Accepted article published: 5 April 2019 Published online in Wiley Online Library: 13 May 2019

(wileyonlinelibrary.com) DOI 10.1002/jsfa.9725

Differentiation of avocados according to their botanical variety using liquid chromatographic fingerprinting and multivariate classification tree

Sandra Martín-Torres,* Ana M Jiménez-Carvelo, Antonio González-Casado and Luis Cuadros-Rodríguez

Differentiation of avocados according to their botanical variety using liquid chromatographic fingerprinting and multivariate classification tree

Chromatographic fingerprint and multivariate methods to discriminate avocado varieties

Sandra MARTÍN-TORRES✉, 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

BACKGROUND

The oil content, composition and marketing threshold value of an avocado depends on the cultivar hence, identifying the cultivar of the avocado fruit is desirable. However, analytical methods have not been reported with this aim.

RESULTS

A multivariate classification tree method was proposed to discriminate three commercial botanical varieties of avocado: Hass, Fuerte and Bacon, using liquid chromatography coupled to a charged aerosol detector (HPLC-CAD). Prior to the chromatographic analysis the avocados were lyophilised and then the oil fraction was extracted using a pressurised liquid extraction system. Normal and reverse phase liquid chromatography were applied in order to obtain the chromatographic fingerprint for each sample.

Soft independent modelling of class analogies (SIMCA) and partial least-squares discriminant analysis (PLS-DA) were applied. Classification quality metrics were determined to evaluate the performance of the classification. Several strategies to develop the classification models were employed. Finally, the useful application of 'classification trees' methodology, which has been scarcely applied in the field of analytical food control, was evaluated to perform a multiclass classification.

CONCLUSION

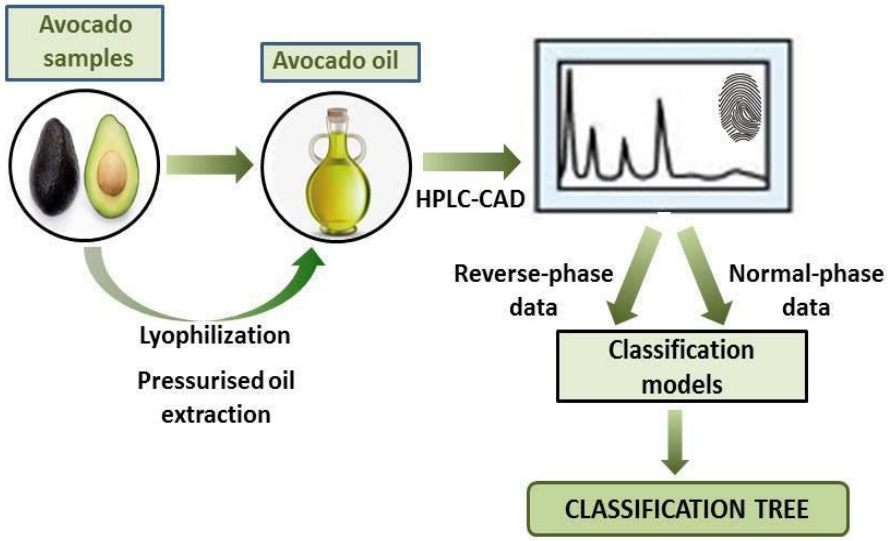
Discrimination of the three botanical varieties was achieved. The best classification was obtained when the PLS-DA is applied on the normal-phase chromatographic fingerprints. Classification trees are showed to be useful tools that provide complementary information to single concatenated models showing different results from the same prediction sample set.

✉ Corresponding author: phone: +34958240797; fax: +34958243328; email: sanmartor@ugr.es

Keywords

Avocado oil; Botanical variety; Chromatographic fingerprinting; Multivariate classification; Classification tree.

Graphical abstract



1. Introduction

The avocado tree, Persea americana Mill., is a member of the Lauraceae, a pantropical family of about 50 genera and 3000 species, and of the genus Persea, which is divided into three different subgenera that contain more than 150 species [**1**]. The avocado fruit, also so-called 'palta' in Quechua langue, is a berry fruit that consists of a large central seed and pericarp, which is the sum of the skin (exocarp), the edible portion (mesocarp) and the inner layer surrounding the seed (endocarp) [**2**].

Although the avocado plant is native from Central America and Mexico, the avocado crop is dispersed worldwide in tropical and subtropical regions and its presence in the world market has been growing steadily in the past two decades (in 2016, the world avocado production was approximately 5.5 million tons, representing an increase of 35% in the last ten years), and it is no longer considered an exotic fruit but part of the everyday diet of many countries [**3**]. This trend has been reinforced by the consumer inclination to look for natural products. The avocado oils are used as cosmetic, soap, and shampoo; as well as processed foods derived from it, such as guacamole, frozen products and avocado paste. Three original botanical races of avocado adapted to different climate conditions have traditionally been recognized: Mexican, Guatemalan and West Indian. The differences between them are basically of morphological, physiological and horticultural traits. Most commercial avocado cultivars are interracial hybrids developed from chance seedlings. Thus, the most important marketed cultivars in subtropical climates are Guatemalan-Mexican hybrids with different degrees of hybridization [**2**,**4**].

The avocado is a climacteric fruit, i.e., its ripening is completely different from that of most other fruits because ripening does not happen in the tree, but only after harvest [**5**]. When an avocado pear reaches the point of physiological ripeness, it contains almost 80% water. From a nutritional point of view, avocado pear is an important and high caloric fruit; moreover, it is rich in vitamin E, ascorbic acid (vitamin C), vitamin B6, β-carotene, and potassium [**6**]. Indeed, avocado fruit present a high nutritional value due to its high content of mostly monounsaturated oils: oleic acid stands out within this group as one of the most characteristic. Other important fatty acids of avocado fruit, although less abundant, are linoleic (polyunsaturated) and palmitic (saturated) acids [2]. According to the American Dietetic Association (ADA), avocado fruit is classified as a functional food owing to its high nutritional value and proven beneficial effects on human health due to the presence of many bioactive lipid-soluble components [**7**].

The oil content and composition of an avocado vary according to multiple factors, such as, the geographical location of the orchard, climate conditions, agricultural practices, growing season, harvesting time, maturity or ripeness or

botanical variety [**8**,**9**,**10**]. Maturity and development of avocados can be assessed by their dry matter contents, but the marketing threshold value depends on the cultivar [**11**], this fact evidences the importance of knowing the cultivar of the avocado fruit.

From the analytical point of view and as any vegetable oil, the avocado oil is composed by two chemical fractions: (i) the saponifiable fraction, constituted by triacylglycerols (TAGs), diacylglycerols, monoacylglycerols, phospholipids, waxes and free fatty acids; and (ii) the unsaponifiable fraction chemically composed by minor components as hydrocarbons, sterols, phenols, tocopherols, terpenes, colorants, volatiles, etc [**12**]. Surprisingly, the scientific literature on the lipid composition of avocado oil is still scarce and insufficient and few papers has been published focused on this matter [**13**,**14**] and almost always they are local studies [**10**,**15**,**16**]. Most of the authors have limited to releasing and studying the fatty acid (FA) profile using gas chromatography after a transmethylation derivatization reaction of the original oil. This approach provides incomplete chemical information and does not allow knowing the variability that is created in the combination of three fatty acids in the TAGs, the mayor components in the avocado oil (95-98% of the whole oil composition).

It is well known that each vegetable fruit, according to its botanical specie, genus or variety, has a characteristic lipid profile [**17**]. Consequently, the TAGs profiling is a tool both useful and reliable in identifying and/or authenticating vegetable oils. However, as it has been above stated, few studies have been found in the literature in which the TAG profile is directly determined in avocado oil and the most current one dates back to 1992 [**18**,**19**]. Moreover, to our knowledge, there is no antecedents describing the comparison and classification of the avocado oils according to its botanical origin in contrast with others edible vegetable oils as olive oil [**20**, **21**].

The analysis of the TAG composition of oil is a very challenging task. TAGs are triesters consisting of a glycerol skeleton linked to three medium or long chain fatty acid residues and, moreover, exits a large variability due to the number of possible FAs combinations on the glycerol skeleton and the different positional isomers [**22**]. For TAG profiling, conventional chromatographic techniques have been used successfully for the qualitative and quantitative determination of TAGs [**23**]: mainly high-temperature gas chromatography ((HT)GC) [**24**], and high-performance liquid chromatography (HPLC) in both normal and reverse phase modes [**25**,**26**,**27**].

Another way to identify each vegetable species is applying the 'fingerprinting' methodology. This uses no-specific signals, where all the implicit, but non-evident, information contained in the analytical signal from the samples is used, and there is no need to identify or quantify each compound present in the test sample. Signal coming directly from the chromatographic instrument is

evaluated as a whole using chemometric tools as exploratory data analysis and pattern recognition analysis [**28**,**29**].

This paper is aimed (i) to obtain liquid chromatographic 'fingerprints' of the oil extracted from the avocado pulp, representative of the major lipid fraction that may be characteristic of each variety of avocado; and (ii) to apply proper chemometric tools to recognize patterns on the data vectors that define each 'chromatographic fingerprints', in order to build multivariate mathematical models capable of differentiating each plant cultivar. Firstly, the reverse phase modality will be used to record the fingerprints, following the tendency found in the literature for TAGs analysis. Then, the normal phase modality will be employed to compare results.

2. Materials and methods

2.1. Chemicals

All solvents used (acetonitrile, isopropanol, n-hexane, methanol) were HPLC grade. Acetonitrile, n-hexane and methanol were supplied by Honeywell (New Jersey, USA) and isopropanol was provided by Panreac Química (Barcelona, Spain). Pelletiser diatomaceous earth, used as inert solid diluent and filter, was provided by Dionex Corporation (Sunnyvale, CA, USA).

TAG standards (trioleyl glycerine or trioleine, tristearyl glycerol or triestearyine and tripalmityl glycerine or tripalmitin) were supplied by Sigma-Aldrich (Steinheim, Germany). In addition, the 5α-colestan-3β-ol reagent, (or simply, cholestanol), used as equipment quality control standard, was supplied by Sigma-Aldrich (Steinheim, Germany).

2.2. Analytical equipment and experimental conditions

A SCANVAC CoolSafe™ freeze dryer was used for sample lyophilisation (48 hours, -106 ºC, reduced pressure). A pressurised liquid extraction system ASE 100 (Dionex Corporation, Sunnyvale, CA, USA), using 34 mL steel extraction cells were used to perform the oil extractions. A Büchi RE 124 rotatory evaporator equipped with a vacuum pump V 700 (Büchi, Flawil, Switzerland) was used to remove the remaining solvents after extraction.

An Agilent 1100 series (Santa Clara, CA, USA), equipped with a CH30 column thermostat (Eppendorf, Hamburg, Germany), a quaternary pump and degasser auto sampler, was used to perform the liquid chromatographic analysis. Detection was performed with a corona charged aerosol detector (CAD) (ESA Bioscenses, Chemlsford, MA, USA); nitrogen (99.9999%, Air Liquid, Madrid, Spain) was used for the aerosol formation. Agilent ChemStation software (rev.

B.02.01-SR1) for LC systems was used. The chromatographic columns used and the chromatographic conditions in both reverse and normal phase modes are shown in table 1.

Table 1. Description of chromatographic columns and chromatographic conditions applied in both reverse and normal phase methods.

	Reverse phase (RP)	Normal phase (NP)
Column	250x4.6 mm i.d., 5 µm Develosil packed column C30-UG-5 (supplier by Nomura Chemica CO, Aichi, Japan)	250×4 mm i.d., 5 µm Lichrospher 100 CN (supplier by Merck, Darmstadt, Germany)
Temperature	45 ºC	30 ºC
Mobile phase	acetonitrile/isopropanol (40:60 v/v)	n-hexane/isopropanol (96:4 v/v)
Injection volume	20 µL	20 µL
Run time	30 min	26 min

2.3. Samples and sample preparation

Thirty-two avocados pear were analysed. They belonged to three different botanical varieties: 16 Hass samples (from Spanish and Mexican origin), 8 Bacon samples and 8 Fuerte samples (purchased in food stores in Granada, Spain)

In addition, four commercial guacamoles (an avocado-based dip or spread) from different marketing brands were acquired in grocery markets in Granada (Spain), always having an avocado content greater than 95%. Moreover, three samples of commercial avocado oil were bought (two from Spanish origin and one from Mexican origin). The avocado varieties were not specified in the labelling and it was not possible to contact the respective producers to have a more detailed knowledge of them. Therefore, these samples were treated as samples of 'unknown class' for the chemometric study.

Previously to the chromatographic analysis, avocados were subjected to a freeze-drying and oil-extraction processes. The detailed procedure was: (i) the avocados were peeled and the kernel was removed to leave only the mesocarp (pulp); (ii) the mesocarp was cut into small cubes; (iii) samples were lyophilized during 48 hours at -106 °C and reduced pressure; (iv) the dried avocado pieces were milled using a household grinder; (v) 2 g of each avocado sample, immediately after being lyophilised, were mixed with 3 g of diatomaceous earth and then extracted using an hexane:isopropanol (3:2, v/v) mixture; a single

extraction cycle at pre-optimized conditions of temperature and static time of 175°C and 5 min respectively [**30**], were applied; (vi) the resulting solution was transferred to a 50 mL round bottom flask and the remaining solvent was evaporated; (vii) the isolated oil, duly weighed, was transferred to a 2 mL screw vial and stored at -20ºC; (viii) from each avocado oil, a test solution was prepared by weighing 0.1 g of oil and 1 g of n-hexane; (ix) the mixture was vortexed and filtered through a 0.22 µm polytetrafluoroethylene (PTFE) filter; and (x) finally, the filtrated solution was cold stored in a 5 mL amber glass vial, in a refrigerator at 4ºC, until its chromatographic analysis.

For guacamole samples the exact process was similar to the process described above (except steps i and ii).

2.4. Chemometrics and software

The raw data files from each chromatogram were obtained in a CSV (comma-separated values) file, and then converted to MATLAB format (R2013b version, The Mathworks Inc. MA, USA) using a home-made MATLAB function named 'csv2workspace' (version 02). The data vector of each sample was described by 2999 variables as a function of time.

Next, the pre-processing of the data was carried out using a home-made MATLAB function, named 'MEDINA' (version 07) [**31**] which employs some algorithms contained in MATLAB Bioinformatics Toolbox™ software to improve the quality of raw chromatographic data, and the 'icoshift' algorithm (version 1.2) to align the chromatographic signals [**32**]. The data pre-processing stages were: (1) raw chromatograms data grouping and overlay; (2) filtered of the raw chromatograms data to eliminate noise of analytical signal; (3) baseline correction. The baseline of the chromatogram corresponds to the detector signal in the absence of eluted compounds. The baseline can vary from one chromatogram to another due to different factors such as temperature, pressure or variations in the detector; (4) alignment of the peaks using the function 'icoshift'; and (5) finally, a mean centring of the data set was applied prior to the statistical analysis.

Once the chromatographic data pre-processing was carried out, it was then possible to use classification and statistical learning tools to build the different multivariate classification models. All chemometric treatments were carried out using the PLS Toolbox (version 8.2) (Eigenvector Research Inc., Wenatchee, WA), for MATLAB software (Mathworks Inc., Natick, MA, USA). Two well-known classification methods were applied: soft independent modelling by class analogy (SIMCA) [**33**] and partial least squares-discriminant analysis (PLS-DA) [**34**]. The original dataset was previously split into a training set and an external validation set applying the Kennard Stone algorithm [**35**] and approximately 30%

of the samples from each class were set to define the external validation set while the remaining samples constitutes the training set.

The label and description corresponding to each classification model appears in the table 2.

Table 2. Labels and description of each classification model mentioned in the text.

Classification	Input class	Designation	Description
SIMCA	Hass	(1i)S-H/nH(i)	Model trained with Hass samples and externally validated with Hass, Fuerte and Bacon samples.
		(1i)S-H/nH(f)	Model trained with all Hass samples, cross-validated and used to predict unknown class samples.
	Fuerte / Bacon	(2i)S-F/B(i)	Model trained with Fuerte and Bacon samples and externally validated with Fuerte and Bacon samples.
		(2i)S-F/B(f)	Model trained with all Fuerte and Bacon samples, cross-validated and used to predict unknown class samples.
PLS-DA	Hass / no Hass	(2i)P-H/nH(i)	Model trained with Hass and noHass (Fuerte and Bacon) samples and externally validated with Hass and noHass samples.
		(2i)P-H/nH(f)	Model trained with all the sample of the study (32 samples), cross-validated and used to predict unknown class samples.
	Fuerte / Bacon	(2i)P-F/B(i)	Model trained with Fuerte and Bacon samples and externally validated with Fuerte and Bacon samples.
		(2i)P-F/B(f)	Model trained with all Fuerte and Bacon samples, cross-validated and used to predict unknown class samples.

The number of samples per class in the training test sets of the models is deeply described in Table 3.

SIMCA

Firstly, the aim was to build a model that differentiates between Hass and the remained varieties. Hass variety is currently considered the standard variety in the avocado marketing, due to the advantages that it presents compared to other varieties. Also, it is widely cultivated in Spain and it is the common variety that fruit stands offer.

SIMCA allows the development of classification models with a one-input class or 'target' class [**36**], which presents a great advantage since it only requires experimental data of a class. Therefore, the initial Hass model ((1i)S-H/nH(i)), was built with a training set made up only with Hass samples and an external validation set composed of all the varieties. The second initial classification model Fuerte/Bacon ((2i)S-F/B(i)) was a conventional two-input (2iC) class model.

Once the different models cited above were validated, they were nurtured with the samples used in the external validation set. Following, these final models ((1i)S-H/nH(f) and (2i)S-F/B(f)) were applied to predict the class belonging of the guacamole and commercial avocado oil samples according to the botanical variety of the avocados used in their production. In this step, venetian blinds cross-validation was applied in order to verify the quality performances features of both final classification models. Venetian blinds select segments of consecutive samples as training and test subsets. It is a very straightforward way of stratifying that is useful when the samples of the training and test sets span the same data space as far as possible [**37**].

Table 3. Description of the samples set used in the development and application of the different classification models.

SIMCA				
	Development of the model		**Application of the model**	
Dataset	(1i)S-H/nH(i)	(2i)S-F/B(i)	(1i)S-H/nH(f)	(2i)S-F/B(f)
Training set	10 Hass	5 Fuerte 5 Bacon	16 Hass	8 Fuerte 8 Bacon
External Validation set	6 Hass 8 Fuerte 8 Bacon	3 Fuerte 3 Bacon	–	–
Prediction set	–	–	4 Guac 3 Avoc.oil	4 Guac 3 Avoc.oil
PLS-DA				
	Development of the model		**Application of the model**	
Dataset	(2i)P-H/nH(i)	(2i)P-F/B(i)	(2i)P-H/nH(f)	(2i)P-F/B(f)
Training set	10 Hass 5 Fuerte 5 Bacon	5 Fuerte 5 Bacon	16 Hass 8 Fuerte 8 Bacon	8 Fuerte 8 Bacon
External validation set	6 Hass 3 Fuerte 3 Bacon	3 Fuerte 5 Bacon	–	–
Prediction set	–	–	4 Guac 3 Avoc.oil	4 Guac 3 Avoc.oil

Guac: commercial guacamole; Avoc. Oil: avocado commercial oil

PLS-DA

Two initial binary classification concatenated models were made Hass/noHass ((2i)P-H/nH(i)) and Fuerte/Bacon ((2i)P-F/B(i)). The second one was made using the samples declared as noHass in the first model.

Continuing with the established pattern, once both models were validated, they were fed with the samples previous used in the external validation set and the two new final models ((2i)P-F/B(f) and (2i)P-F/B(f), respectively) were applied to predict the class belonging of the unknown-variety samples (guacamoles and commercial oil).

Table 3 shows a straightforward description of the sample set used in the development and application of the SIMCA and PLS-DA models.

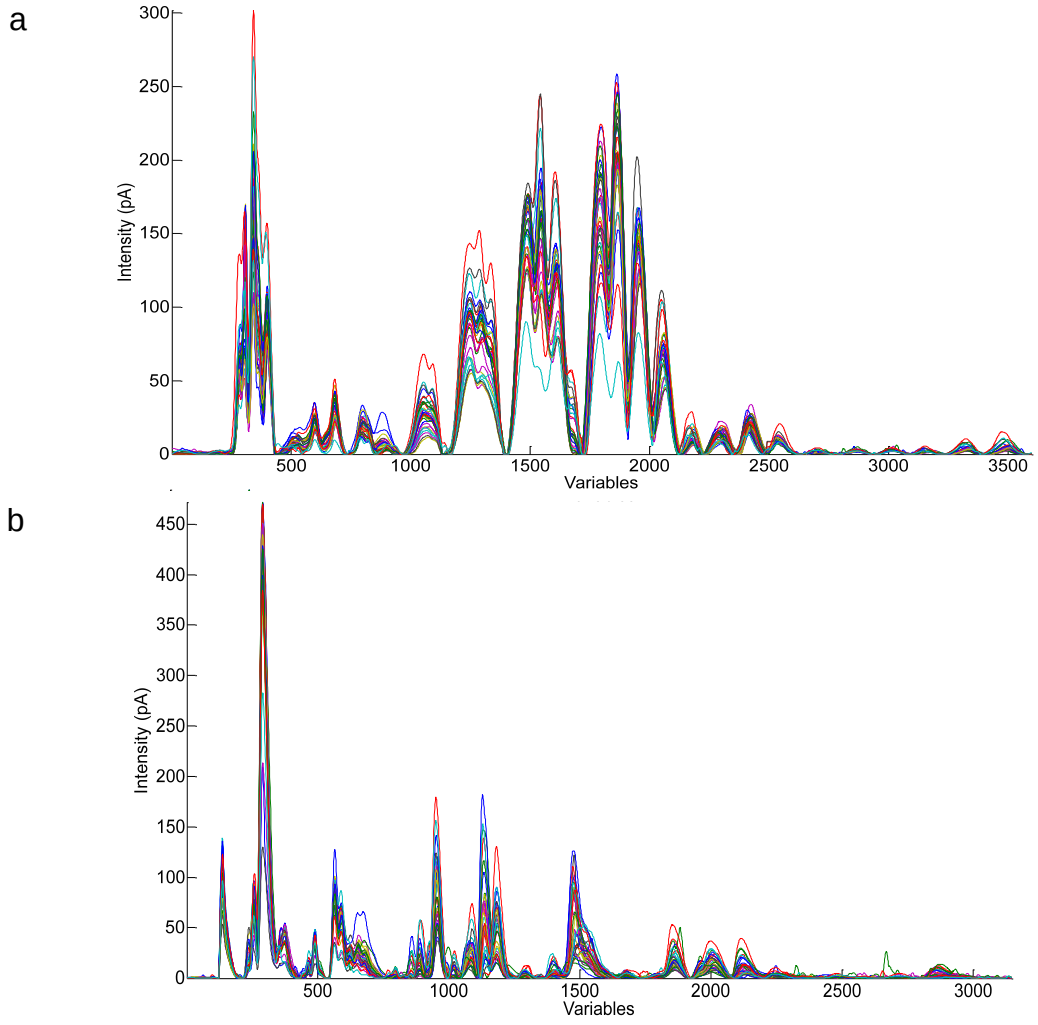


Figure 1. Superposed chromatograms of the 32 avocado samples after pre-processing (except the mean centring step): (a) chromatograms obtained in reverse phase; (b) chromatograms obtain in normal phase.

3. Results and discussion

A chromatogram was recorded for each vegetable oil sample both in reverse and normal phases. As any vegetal oil, the avocado oil is majority composed by TAGs. The behaviour of these compounds is not the same in normal phase (NP) or reverse phase (RP) liquid chromatography (LC). The elution in NPLC is determined by the degree of saturation and by the number of carbons of the fat chains, and saturated species elute before the more unsaturated homologues. On the contrary, in RPLC the components are separated according to the joint effect of the chain-lengths plus their degree of unsaturation, whereby each double bond reduces the retention time by the equivalent of about 2 carbon atoms [**38**]. So, NPLC and RPLC chromatograms are not similar and one of them successfully can be more useful in order to build a classification model.

Figures 1(a) and 1(b) show the superposed chromatograms of all vegetable oil samples in reverse and normal phase modes, respectively.

3.1. Reverse phase

Firstly, a principal component analysis (PCA) was carried out considering the dataset composed of the whole chromatogram from each vegetable oil sample. The aim of this method was to explore if there were natural groupings of different oils. Five principal components (PCs) were enough to explain 94.92% of the total variance (36.97, 31.72, 19.69, 3.87, 2.68% respectively). Figure 2 shows the scores in PC3-PC1 plane for reverse phase fingerprint.

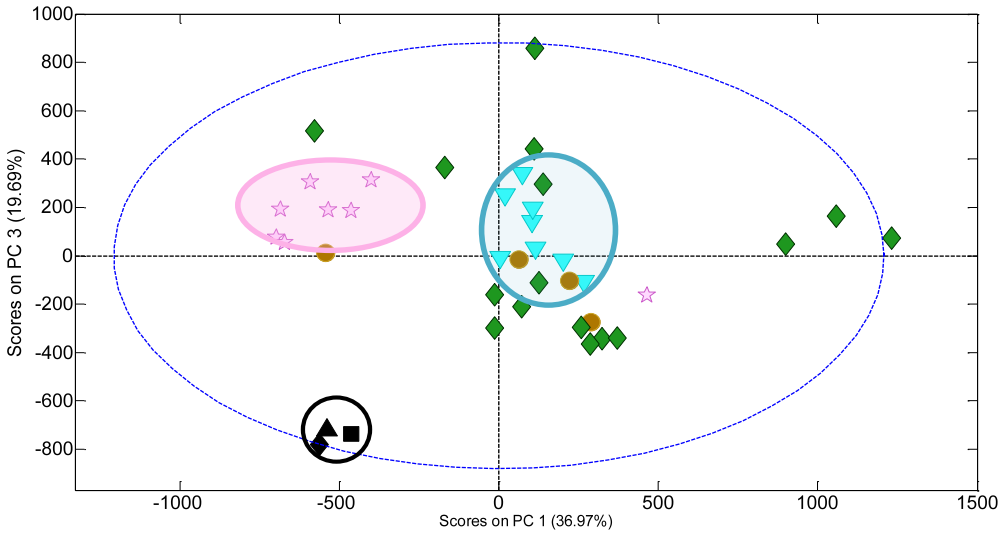


Figure 2. PCA scores biplot in PC3-PC1 plane: Hass samples (green rhombus); Fuerte samples (pink stairs); Bacon samples (blue triangles); guacamole samples (yellow circles); and commercial oil samples (in black).

PCA in reverse phase only shows two natural groups of data. PC1 component grouped the Fuerte variety samples, with negative scores for this component, and Bacon variety with positive scores for it. The plot shows that the samples of Hass variety do not show any tendency, they are scattered throughout the region of confidence of the model.

SIMCA was the first chemometric method used to establish multivariate class modelling models. In this part of the study only one class (1iC) was modelled: 'Hass' and (1i)S-H/nH(i) model was built.

The classification threshold was chosen from the reduced Hotelling T_r^2 and Q_r statistics values. Thus, the oil samples were classified 'within the class model' when their values were lower than 1 for both statistics values. Figure 3a shows the scores of T_r^2 versus Q_r for the initial model (1i)S-H/nH(i). As it is shown, all the Hass variety samples of the training set were correctly classified. On the contrary, any Hass variety sample of the external validation set satisfied the model and they were not classified (sensitivity of the model equal to 0).

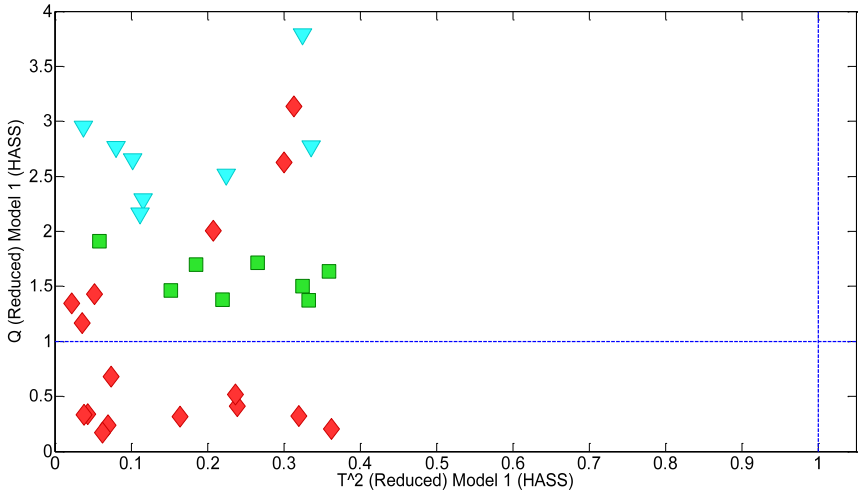
Then, an initial model (2i)S-F/B(i) was built using all the samples declared as 'no Hass', and in this way two classes were modelled: 'Fuerte' and 'Bacon'. The classification of the samples was evaluated looking at Coomans' plot. Coomans' plot is a very useful tool to visualize principal groupings, in which the two axes represent the orthogonal distances of all the samples respect the individual model. In Coomans' plot, PCA model is separately applied to each class and two class models are plotted against each other with the critical levels as straight lines displaying the boundaries allowing visualizing the assignment of the samples to each of the classes simultaneously. In this case, all the training set was properly assigned to each class, but no sample of the external validation set was properly classified and it was not assigned (no class).

PLS-DA was the method then used to test the discrimination process. In order to differentiate Hass variety from other varieties, two input-class classification (2iC) strategy was applied in which the target class was 'Hass' (assigned value iqual to 1) and the alternative class was 'no Hass' (assigned value equal to 0) (initial (2i)P-H/nH(i) model). Then, the samples declared as 'no Hass' were used to build the initial model (2i)P-F/B(i). In this model, 'Bacon' class was assigned a value equal to 1 and 'Fuerte' class with a value equal to 0.

Figure 3b shows the PLS-DA assignation of samples to each of the predefined classes in model (2i)P-H/nH(i). The horizontal red line corresponds to the threshold established by the software. Both initial models, (2i)P-H/nH(i) and (2i)P-F/B(i), allowed the properly classification of 100% of samples: all the Hass samples were predicted to belong to class 'Hass' (figure 3b, prediction values higher than the threshold), all the samples from Bacon variety were predicted to

belong to 'Bacon' class and finally all the Fuerte samples were predicted to belong to 'Fuerte' class.

a



b

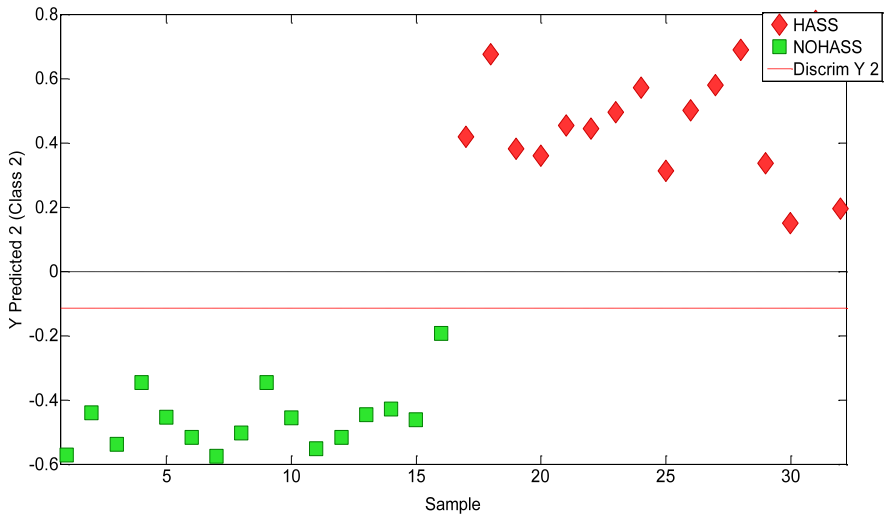


Figure 3. Classification plots from the reverse phase chromatographic fingerprints: (a) (1i)S-H/nH(i) model plot: Hass samples (red rhombus); Fuerte samples (green squares); Bacon samples (blue triangles); (b) (2i)P-H/nH(i) model.

Table 4 shows the specifications and Table 5 shows the different quality performance features of all the different multivariate classification models made up with the dataset obtained from reverse phase HPLC.

Table 4. Characteristics of the PLS-DA and SIMCA models from reverse phase HPLC data.

	SIMCA					PLS-DA	
models	PCs (class value =1)	%var	PCs (class value =0)	% var	**models**	LVs	% var
(1i)S-H/nH(i)	5	94.92	–	–	(2i)P-H/nH(i)	5	96.18
(2i)S-F/B(i)	3	96.25	2	96.78	(2i)P-F/B(i))	4	97.85

The hyphen ‘‘–” is signifying that there is no class.

Table 5. Quality performance features for the different classification models make up with the data obtained by reverse phase.

	SIMCA		PLS-DA	
Performance features	**(1i)S-H/nH(i)**	**(2i)S-F/B(i)**	**(2i)P-H/nH(i)**	**(2i)P-F/B(i)**
Sensitivity(or Recall)	0.00	0.00	1.00	1.00
Specificity	1.00	1.00	1.00	1.00
Positive predictive value (Precision)	–	–	1.00	1.00
Negative predictive value	0.73	0.50	1.00	1.00
Youden index	0.00	0.00	1.00	1.00
Positive likelihood rate	–	–	–	–
Negative likelihood rate	1.00	1.00	0.00	0.00
F-measure	–	–	1.00	1.00
Discriminant power	–	–	–	–
Efficiency (or Accuracy)	0.73	0.50	1.00	1.00
AUC (Correctly classified rate)	0.50	0.50	1.00	1.00
Matthews correlation coefficient	–	–	1.00	1.00
Kappa coefficient	0.00	0.00	1.00	1.00

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

In order to improve these results and with the objective of predict the avocado varieties of different commercial guacamole and oils; it was decided to make up new models using the normal phase chromatographic fingerprints.

3.2. Normal phase

A PCA was carried out considering the dataset composed of the whole chromatograms from each vegetable oil sample in normal phase. Five PCs were enough to explain 90.26% of the variance (37.45, 31.20, 12.91, 5.01 and 3.69% respectively). Figure 4 shows the scores in PC4-PC3 plane for normal phase fingerprint.

In this case, PCA shows four groups of samples which are correlated with Bacon (negative scores for both PC3 and PC4), Fuerte (negative scores for PC3 but positive for PC4), Hass (positive scores for PC3 and negative scores for PC4) varieties and, finally, guacamole samples (positive scores for both PC3 and PC4). On the other hand, the commercial oil samples are not grouped and they are scattered throughout both Fuerte and Hass samples scores.

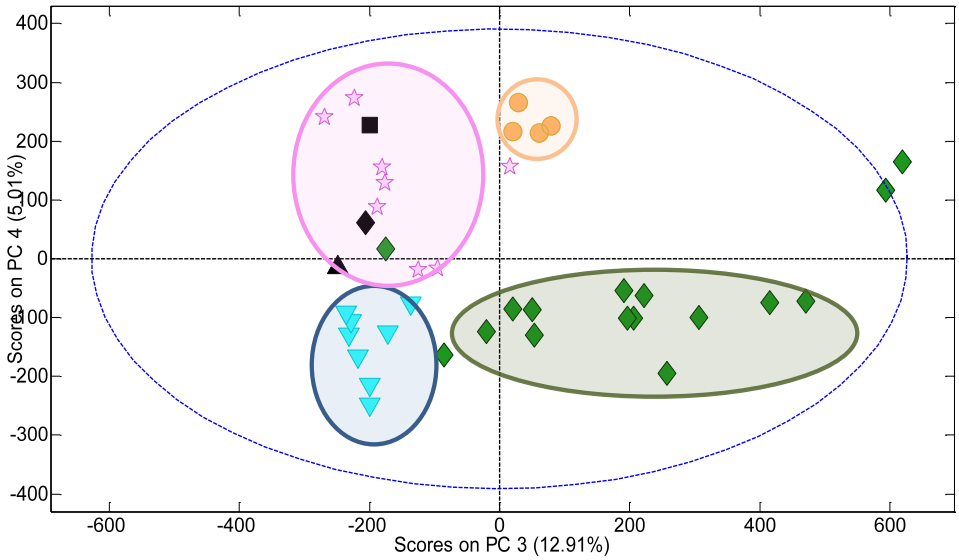


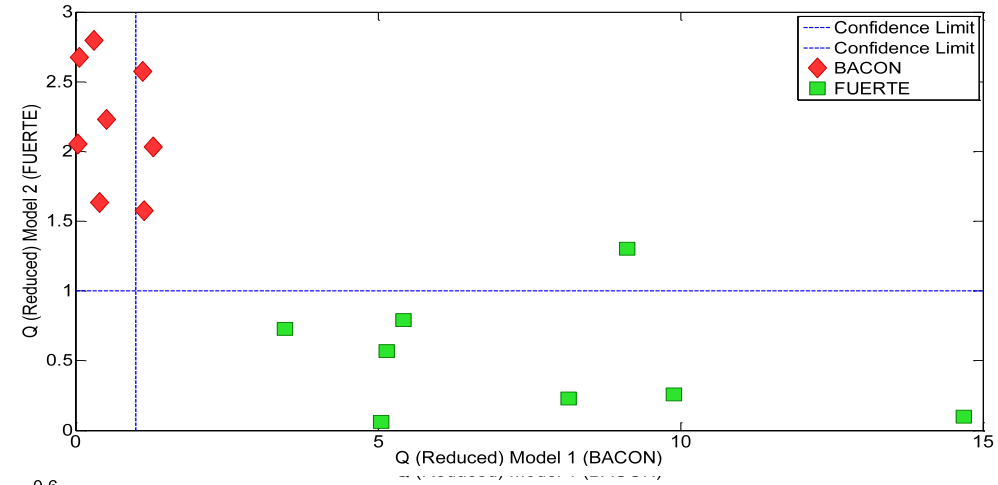
Figure 4. PCA scores biplot in PC4-PC3 plane: Hass samples (green rhombus); Fuerte samples (pink stairs); Bacon samples (blue triangles); guacamole samples (yellow circles); and commercial oil samples (in black).

Using the same strategy previously described, an initial SIMCA model ((1i)S-H/nH(i)) was built. The same criteria previous described was used to establish the classification threshold in this model. The scores plot of T_r^2 versus Q_r showed that all Hass samples, both of training and external validation sets, were classified within the class model whereas none of 'no Hass' samples were (sensitivity and specificity equal to 1).

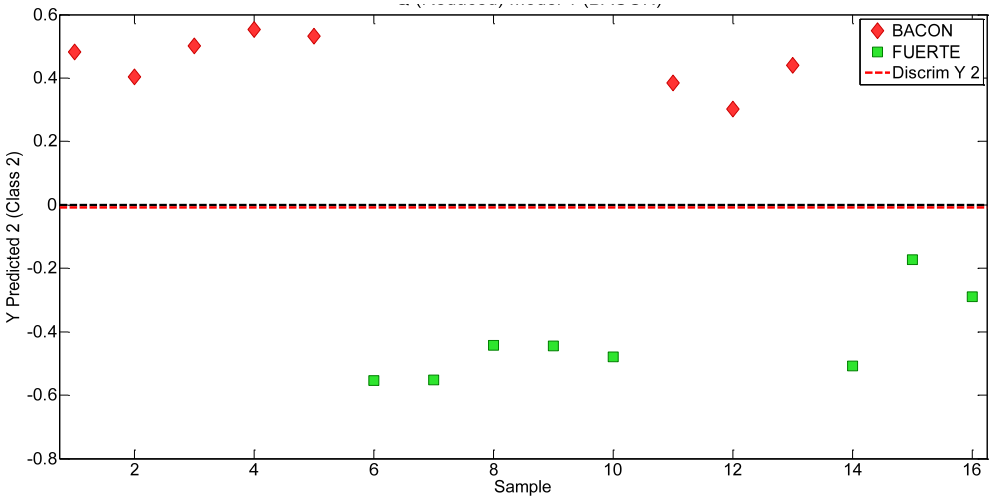
The second initial classification model Fuerte/Bacon ((2i)S-F/B(i)) was built using all the samples declared as 'no Hass'. The classification of the samples was evaluated looking at Coomans' plot showed in figure 5a. It is concluded that no Bacon samples of the validation set were assigned to the model while

most Fuerte samples were properly assigned to its own class.

a



b



c

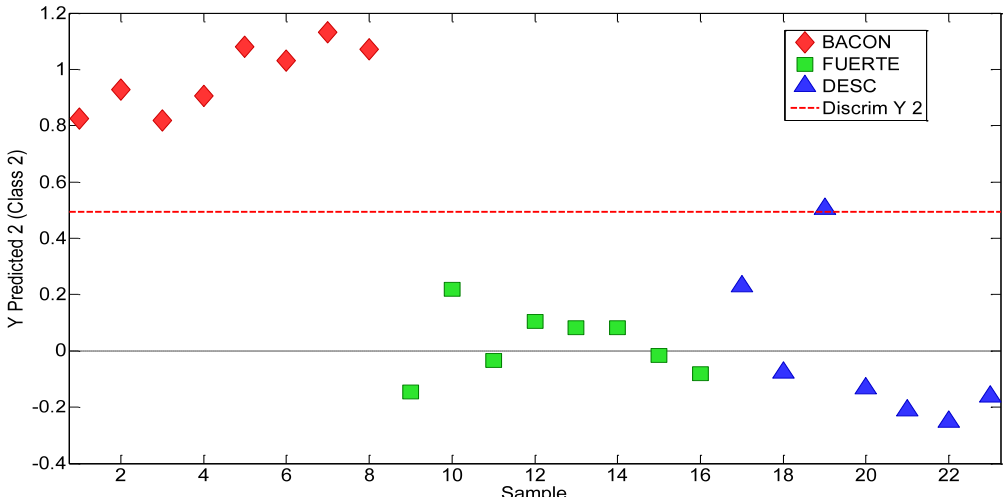


Figure 5. Classification plots from the normal phase chromatographic fingerprints: (a) Coomans' plot of the (2i)S-F/B(i) model. (b) full plot of the (2i)P-F/B(i) model (c) (2i)P-F/B(f) model.

In order to evaluate the prediction capacity of the models, the final models (1i)S-

H/nH(f) and (2i)S-F/B(f) were built. In both final models, the training sets were nurtured with the samples previous used in the respective external validation set and a prediction set was made up with unknown variety commercial guacamoles and avocado oils. Venetian blinds cross-validation was applied in order to verify the quality performances features of both classification models. No model allows the prediction of the unknown samples as they were not included in the confident region of any modelled class.

Similar methodology was used to establish the different PLS-DA multivariate classification models. Firstly, (2i)P-H/nH(i) model was built and all the samples of the external validation set were properly classified to belong to 'Hass' or 'no Hass' predefined class. Secondly, the (2i)P-F/B(i) model was made up with the samples declares as 'no Hass'. Figure 5b show the classification plots of the training and validation samples for the (2i)P-F/B(i) model. All of them were correctly assigned to each class (scores above the threshold line for Bacon samples and below the threshold line for Fuerte samples).

Finally, (2i)P-H/nH(f) and (2i)P-F/B(f) final models were used to predict the class belonging of the guacamole and commercial avocado oil samples based on the botanical variety of the avocados used in their production. Figure 5c shows the scores for the model (2i)P-F/B(f). As it is shown, four 'unknown' samples were classified in 'Fuerte' class while a Mexican oil sample was situated on the discriminant line between the two predefined classes: this is an inconclusive sample [**39**].

Table 6 shows the specifications and Table 7 shows the different quality performance features of all the different multivariate classification models made up with the dataset obtained from normal phase HPLC.

Table 6. Characteristics of the PLS-DA and SIMCA models with normal phase data.

	SIMCA					**PLS-DA**	
models	PCs (class value =1)	%var	PCs (class value =0)	% var	**models**	LVs	% var
(1i)S-H/nH(i)	4	92.78	–	–	(2i)P-H/nH(i)	4	90.06
(2i)S-F/B(i)	2	97.52	2	97.11	(2i)P-F/B(i))	4	97.52
(1i)S-H/nH(f)	4	89.61	–	–	(2i)P-H/nH(f)	5	92.08
(2i)S-F/B(f)	3	96.25	2	96.78	(2i)P-F/B(f)	3	94.19

The hyphen "–" is signifying that there is no class.

Table 7. Quality performance features for the different classification models make up with the data obtained by normal phase.

	SIMCA				PLS-DA			
Performance features	**(1i)S-H/nH(i)**	**(1i)S-H/nH(f)**	**(2i)S-F/B(i)**	**(1i)S-F/B(f)**	**(2i)P-H/nH(i)**	**(2i)P-H/nH(f)**	**(2i)P-F/B(i)**	**(2i)P-F/B(f)**
Sensitivity (or Recall)	1.00	1.00	0,67	1.00	1.00	1.00	1.00	1.00
Specificity	1.00	1.00	1,00	1.00	1.00	1.00	1.00	1.00
Positive predictive value (Precision)	1.00	1.00	1,00	1.00	1.00	1.00	1.00	1.00
Negative predictive value	1.00	1.00	0,75	1.00	1.00	1.00	1.00	1.00
Youden index	1.00	1.00	0,67	1.00	1.00	1.00	1.00	1.00
Positive likelihood rate	–	–	–	–	–	–	–	–
Negative likelihood rate	0.00	0.00	0,33	0.00	0.00	0.00	0.00	0.00
F-measure	1.00	1.00	0,80	1.00	1.00	1.00	1.00	1.00
Discriminant power	–	–	–	–	–	–	–	–
Efficiency (or Accuracy)	1.00	1.00	0,83	1.00	1.00	1.00	1.00	1.00
AUC (Correctly classified rate)	1.00	1.00	0,83	1.00	1.00	1.00	1.00	1.00
Matthews correlation coefficient	1.00	1.00	0,71	1.00	1.00	1.00	1.00	1.00
Kappa coefficient	1.00	1.00	0,83	1.00	1.00	1.00	1.00	1.00

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

Both PLS-DA final models were used to design a classification tree in order to be submitted to a confirmatory analysis to check their results. Arduous classification problems often cannot be solved applying a single classification model. Instead, the problem requires to be divided into sub-groups with individual models handling the increased detail of the problem [**40**]. In this case, (2i)P-H/nH(f) and (2i)P-F/B(f) models were used as rule nodes (also known as decision nodes) to establish the tree and the number of branches were fixed by the number of classes defined in each classification model plus an additional 'otherwise' decision branch which is used when none of the other classes is selected (inconclusive samples).

Figure 6 shows a flow-chart of the designed classification tree. A test sample will not be classified in any model's class if the model does not recognise it as an object from the training classes of the models. This can result in different prediction results when compared to the single PLS-DA model's prediction is based on most-probable class. In this case, all the 'known-variety' avocado samples were properly assigned to the corresponding class but none of the prediction samples set were classified in any of the predefined classes and they were assigned to the 'otherwise' decision branch. Thus, all the 'unknown-samples' are finally declared as inconclusive samples.

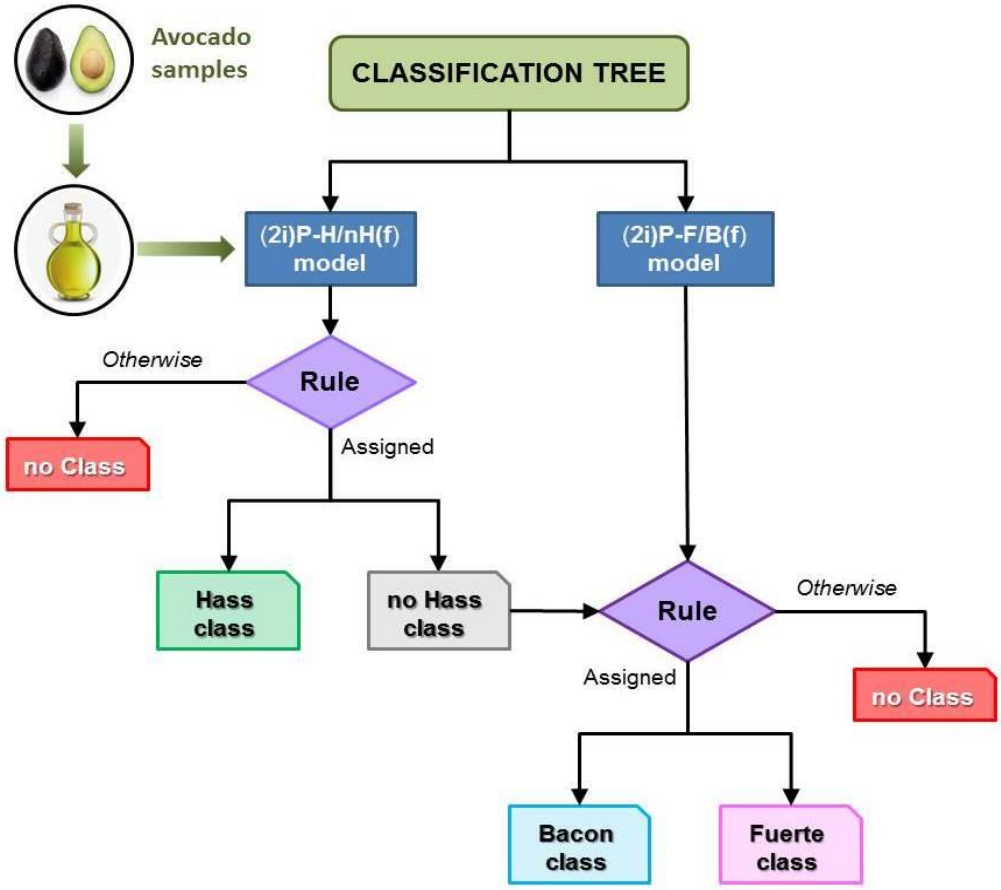


Figure 6. Scheme of a classification tree build using (2i)P-H/nH(f) and (2i)P-F/B(f) models as decision nodes.

4. Conclusion

In this study, a methodology for the discrimination of three commercial varieties of avocado (Hass, Fuerte and Bacon) using the chromatographic fingerprinting methodology is described. Normal and reverse phase liquid chromatography together with different multivariate classification strategies have been tested.

To assess the different classification scenarios, several quality classification metrics were calculated. It has been proved that normal phase conditions are better than reverse phase conditions for this purpose, showing shorter analysis time and better performance of the classification models. The PLS-DA method has achieved to be a better classifier than SIMCA class-modelling method and it should be emphasized that the two-input class strategy yielded better results that one-input class strategy.

Lastly, it has been shown that the 'classification trees', scarcely used in the food control field, are a useful tool to perform a multiclass classification. Classification trees provide complementary information to single concatenated models because they showed different results from the same prediction sample set [**41**].

References

1 Whiley AW, Schaffer B, Wolstenholme BN, The avocado: botany, production and *uses*, 2nd ed. Cab International, Wallingford (2013).

2 Hurtado Fernández E, Fernández Gutiérrez A, Carrasco Pancorbo A, Avocado fruit – Persea americana, in *Exotic fruits – Reference guide*, ed. by Rodrigues S, Silva E, de Brito E. Academic Press, London, pp. 37–48 (2018).

3 FAOSTAT data. *Statistics division of the food and agriculture organization for the United Nations* (2016). Available: http://www.fao.org/faostat/en/#data/QC [31 January 2019]

4 Alcaraz ML and Hormaza JI, Molecular characterization and genetic diversity in an avocado collection of cultivars and local Spanish genotypes using SSRs. *Hereditas* **144**:244–253 (2007).

5 Araújo RG, Rodríguez Jasso RM, Ruiz HA, Pintado ME and Aguilar CN, Avocado by-products: nutritional and functional properties. *Trends Food Sci Technol* **80**:51–60 (2018).

6 Ozdemir F and Topuz A, Changes in dry matter, oil content and fatty acids composition of avocado during harvesting time and post-harvesting ripening period. *Food Chem* **86**:79–83 (2004).

7 Krumreich FD, Borges CD, Mendonça CR, Jansen-Alves C and Zambiazi RC, Bioactive compounds and quality parameters of avocado oil obtained by different processes. *Food Chem* **257**:376–381 (2018).

8 Villa Rodríguez JA, Molina Corral FJ, Ayala Zavala JF, Olivas GI and González Aguilar GA, Effect of maturity stage on the content of fatty acids and antioxidant activity of 'Hass' avocado. *Food Res Int* **44**:1231–1237 (2011).

9 Galvao MS, Narain N and Nigan N, Influence of different cultivars on oil quality and chemical characteristics of avocado fruit. *Food Sci Technol* **34**: 539–546 (2014).

10 Ferreyra R, Sellés G, Saavedra J, Ortiz J, Zúñiga C, Troncoso C, Rivera SA, González-Agüero M and Defilippi BG, Identification of pre-harvest factors that affect fatty acid profiles of avocado fruit (Persea Americana Mill) cv. "Hass" at harvest. *S African J Bot* **104**:15–20 (2016).

11 European Union, Regulation (EC) No 387/2005 amending Regulation (EC) No 831/97 laying down marketing standards applicable to avocados, *Official Journal of the European Union*. L62:5–6 (2005).

12 Ramírez Tortosa MC, Granados S and Quiles JL, Chemical composition, types and characteristics of olive oil, in *Olive oil and health*, ed. by Quiles JL, Ramírez Tortosa MC and Yaqoob P. Cab International, Wallingford, pp. 45–62 (2006).

13 Bora PS, Narain N, Rocha RVM. and Queiroz Paulo M, Characterization of the oils from the pulp and seeds of avocado (cultivar: Fuerte) fruits. *Grasas y Aceites* **52**:171–174 (2001).

14 Dreher ML and Davenport AJ, Hass avocado composition and potential health effects. *Crit Rev Food Sci Nutr* **53**:738–750 (2013).

15 Takenaga F, Matsuyama K, Abe S, Torii Y and Itoh S, Lipid and fatty acid composition of mesocarp and seed of avocado fruits harvested at northern range in Japan. *J Oleo Sci* **57**:591–597 (2008).

16 Ge Y, Ma F, Wu B and Tan L, Morphological and chemical analysis of 16 avocado accessions (Persea americana) from China by principal component analysis and cluster analysis. *J Agric Sci* **10**:80–89 (2018).

17 Lozano YF, Dhuique Mayer C, Bannon C and Gaydou EM, Unsaponifiable matter, total sterol and tocopherol contents of avocado oil varieties. *J Am Oil Chem Soc* **70**:561–565 (1993).

18 Gaydou EM, Lozano Y and Ratovohery J, Triglyceride and fatty-acid compositions in the mesocarp of Persea americana during fruit-development. *Phytochemistry* **26**:1595–1597 (1987).

19 Hierro MTG, Tomás MC, Fernández Martín F and Santa María G, Determination of the triglyceride composition of avocado oil by high-performance liquid chromatography using a light-scattering detector. *J Chromatogr A* **607**:329–338 (1992).

20 Ruiz Samblás C, Cuadros Rodríguez L, González Casado A, Rodríguez García FP, De la Mata Espinosa P and Bosque Sendra JM, Multivariate analysis of HT/GC-(IT)MS chromatographic profiles of triacylglycerol for classification of olive oil varieties. *Anal Bioanal Chem* **399**:2093–2103 (2011).

21 Valdés A, Beltrán A, Mellinas C, Jiménez A, and Garrigós MC, Analytical methods combined with multivariate analysis for authentication of animal and vegetable food products with high fat content. *Trends Food Sci Technol* **77**:120–130 (2018).

22 Marini F, Triacylglycerols: characterization and determination, in *Encyclopedia of food and health, Vol 5*, ed. by Caballero B, Finglas PM, Toldrá F. Academic Press, Oxford, pp. 345–350 (2016).

23 Indelicato S, Bongiornob D, Pitonzo R, Di Stefano V, Calabrese V, Indelicato S and Avellone G, Triacylglycerols in edible oils: determination, characterization, quantitation, chemometric approach and evaluation of adulterations. *J Chromatogr A* **1515**:1–16 (2017).

24 Ruiz Samblás C, González Casado A and Cuadros Rodríguez L, Triacylglycerols determination by high-temperature gas chromatography in the analysis of vegetable oils and foods: a review of the past 10 years. *Crit Rev Food Sci Nutr* **55**:1618–1631 (2015).

25 Lısa M, Lynen F, Holcapek M and Sandra P, Quantitation of triacylglycerols from plant oils using charged aerosol detection with gradient compensation. J Chromatogr A **1176**: 135–142 (2007).

26 Rombaut R, De Clercq N, Foubert I and Dewettinck K, Triacylglycerol analysis of fats and oils by evaporative light scattering detection. *J Am Oil Chem Soc* **86**:19–25.

27 De la Mata Espinosa P, Bosque Sendra JM and Cuadros Rodríguez L, Quantification of triacylglycerols in olive oils using HPLC-CAD. *Food Anal Method* **4**:574–581 (2011).

28 Bosque Sendra, JM, Cuadros Rodríguez L, Ruiz Samblás C and de la Mata AP, Combining chromatography and chemometrics for the characterization and authentication of fats and oils from triacylglycerol compositional data – a review. *Anal Chim Acta* **724**:1–11 (2012).

29 Cuadros Rodríguez L, Ruiz Samblás C, Valverde Som L, Pérez Castaño E and González Casado A, Chromatographic fingerprinting: an innovative approach for food 'identitation' and food authentication – a tutorial. *Anal Chim Acta* **909**:9–23 (2016).

30 Ruiz Samblás C, Cuadros Rodríguez L and González Casado A, Pressurised liquid extraction and quantification of fat-oil in bread and derivatives products. *Talanta* **83**:25–30 (2010).

31 Pérez Castaño E, Ruiz Samblás C, Medina Rodríguez S, Quirós Rodríguez V, Jiménez Carvelo AM, Valverde Som L, González Casado A and Cuadros Rodríguez L, Comparison of different analytical classification scenarios: application for the geographical origin of edible palm oil by sterolic (NP)HPLC fingerprinting. *Anal Methods* **7**:4192–4201 (2015).

32 Tomasi G, Savorani F and Engelsen SB, Icoshift: an effective tool for the alignment of chromatographic data. *J Chromatogr A* **1218**:7832–7840 (2011).

33 Kumar N, Bansal A, Sarma GS and Rawal RK, Chemometrics tools used in analytical chemistry: an overview. *Talanta* **123**:186–199 (2014).

34 Van Ruth SM, Villegas B, Akkermans W, Rozijn M, Van der Kamp H and Koot A, Prediction of the identity of fats and oils by their fatty acid, triacylglycerol and volatile compositions using PLS-DA. *Food Chem* **118**:948–955 (2010).

35 Kennard RW and Stone LA, Computer aided design of experiments. *Technometrics* **11**:137–148 (1969).

36 Cuadros Rodríguez L, Pérez Castaño E and Ruiz Samblás C, Quality performance metrics in multivariate classification methods for qualitative analysis. *Trends Analyt Chem* **80**:612–624 (2016).

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

38 Buchgraber M, Ulberth F, Emons H and Anklam E. Triacylglycerol profiling by using chromatographic techniques. *Eur. J. Lipid Sci. Technol.* **106**:621–648 (2004).

39 López MI, Callao MP and Ruisánchez I, A tutorial on the validation of qualitative methods: From the univariate to the multivariate approach. *Anal Chim Acta* **891**:62–72 (2015).

40 *Hierarchical Model Builder.* Eigenvector Research, Inc. [Online]. Available: http://wiki.eigenvector.com/index.php?title=Hierarchical_Model_Builder#Applying_Hierarchical_Models_to_New_Data [07 Mayo 2019]

PUBLICACIÓN III.

ARTÍCULO CIENTÍFICO

Authentication of the geographical origin and the botanical variety of avocados using liquid chromatography fingerprinting and deep learning methods

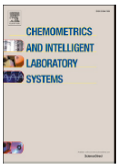
Chemometrics and Intelligent Laboratory Systems, 199, 103960, 1-7, 2020.
https://doi.org/10.1016/j.chemolab.2020.103960

Chemometrics and Intelligent Laboratory Systems 199 (2020) 103960

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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



Authentication of the geographical origin and the botanical variety of avocados using liquid chromatography fingerprinting and deep learning methods

Sandra Martín-Torres *, 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

Authentication of the geographical origin and the botanical variety of avocados using liquid chromatography fingerprinting and deep learning methods

Sandra Martín-Torres✉, Ana M. Jiménez-Carvelo, Antonio González-Casado, Luis Cuadros-Rodríguez

Abstract

The lipid chromatographic fingerprint of different avocado fruits have been acquired and two classification multivariate methods, partial least squares-discriminant analysis (PLS-DA) and support vector machine (SVM), have been successfully tested in order to discriminate and classify a higher variability of avocado samples. Two authentication goals have been achieved attending to: (i) the geographical origin, and (ii) the botanical variety or cultivar. However, to our knowledge, there are no antecedents aimed at comparing and classifying avocado fruits. The pulp oil fraction of the avocado fruit was first extracted using pressurised liquid extraction from the previously lyophilised pulp. Then the 190-400 nm UV-absorption fingerprints were obtained from the avocado oils using normal phase high performance liquid chromatography coupled to an absorption diode-array detector ((NP)HPLC-DAD) and the 220 nm spectra were then selected for classification model building. Several input-class classification strategies were applied and the classification models were externally validated from the specific success/error contingencies. In addition, some quality metrics, *i.e.* sensitivity (or recall), specificity, precision, negative predictive values, efficiency (or accuracy), AUC (area under the receiver operating curve), Mathews correlation coefficient and Kappa coefficient, were determined to evaluate the performance of each classification model (PLS-DA and SVM) and the results clearly show that SVM method is the most proficient.

Keywords

Avocado; Authentication; Chromatographic fingerprinting; Supervised pattern recognition; Deep learning methods.

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

1. Introduction

Avocado (*Persea americana Mill., Lauraceae*) is an ever-green tree originating from Mesoamerica and introduced in southern Spain during the sixteenth century by the Spaniards. It is botanically classified into three groups, which have been termed the Mexican (*Persea americana var. drymifolia*), Guatemalan (*Persea nubigena var. guatemalensis*) and West Indian (*Persea americana var. americana*) races. The differences between them are basically of morphological, physiological and horticultural traits. Each race has unique ecological adaptations and identifiable characteristics. Most commercial avocado cultivars are interracial hybrids, developed from chance seedlings, with different degrees of hybridization. The most well-known and marketed varieties are the Hass and Fuerte varieties [**1**].

The approximate composition of avocado pulp has been reported by the USDA (United States Department of Agriculture) National Nutrient Database for Standard Reference as follows: water, 73.23%; total lipids, 14.66%; proteins, 2%; dietary fibre, 6.70%; sugars, 0.66%; and carbohydrates (by difference) 8.53% [**2**].

Despite the origin of this crop is Central America, avocado has a high nutritional content that has recently aroused increasing global interest and it is now dispersed worldwide in tropical and subtropical regions. In 2017, the world avocado production was almost 6 million of tons. The leading countries in avocado production were Mexico (more than 2 million tons), Dominican Republic, Kenya or Chile among others. American continent conquers the avocado production (74 %) and also have the highest number of harvested hectares (406,464 ha), followed by Africa (11.7 %), Asia (11.2 %) and Europe (1.6 %) [**3**]. Spain is a special case in avocado cultivation since it is the only European country with a significant commercial production (about 90,000 tons in 2017). As more countries produce avocado for export or become familiar with the fruit, the increase in production area, production and consumption is likely to continue. Avocado marketing threshold value depends on the cultivar [**4**]. Moreover, main countries producers of avocado as Mexico or Peru with cheap labour, looser environmental regulations, and large acreage planted in avocados, seemingly are able to undercut the domestic fruit prices, providing consumers with cheaper priced avocados. This fact evidences the importance of knowing the traceability of avocado fruits.

Each vegetable species has a characteristic lipid profile. The fatty acid profile of avocado oil has been relatively consistent among studies, but the relative concentration of each fatty acid component was found to vary considerably. However, the oleic (42-51%) and palmitic (20-25%) fatty acids are always present in a larger proportion in avocado fruits since the predominant triacylglycerols are OOO (21-34%) and OOP (19-24%), where O and P denote oleic and palmitic acids, respectively [**5**]. There are many reports on the lipids contained in avocado fruit [**6**]. Moreover, some of them have proven that the oil content and composition vary

according to the location of the orchard, the variety, the number of days between flowering and harvest, between others [**7**]. Several works have studied and quantified the fatty acid profile of different avocado varieties collected from different geographical regions. Nevertheless, all these studies are based on the quantification of specific compounds, such as oleic acid or tri-unsaturated fatty acids [**8**,**9**].

Fingerprinting is described as a battery of analytical techniques or methods based on treating the entire or a part of the chromatogram as a whole, without identifying or quantifying each compound [**10**]. The fingerprinting methodology uses nonspecific signals, where all the information implicitly found in the samples is used in a non-selective way with the main aim of characterizing or authenticating the food. Fingerprinting is based on obtaining as much useful information of the sample as possible without necessarily identifying or quantifying the compounds present.

The relevance of fingerprinting approach is undeniable too; they offer attractive advantages over other strategies, regarding analysis simplicity, accuracy and rapidity. The most widely used analytical techniques in fingerprinting analysis for the authentication and detection of adulteration of vegetable oils are infrared (IR) [**11**] and nuclear magnetic resonance (NMR) [**12**] spectrometries and both gas and liquid chromatographies coupled to numerous measuring devices [**10**]. Regardless of the approach selected, the obtained data require to be treated using appropriate chemometric tools to extract the unspecific and non-evident information of interest that is implicitly contained on the data. Focusing on chromatographic fingerprint, signal coming directly from the chromatographic instrument is processed and treated as a whole. Some examples of pattern recognition learning algorithms are unsupervised methods, and supervised ones [**13,14**].

In a previous paper it was displayed the discrimination of three avocado varieties using the entire lipid chromatographic profile (mainly related to the triacylglycerol profile) of the fruit applying the fingerprinting methodology [**15**]. Both reverse and normal phase liquid chromatographic coupled to charge aerosol detector fingerprints were recorded and different multivariate classification strategies were tested. Normal phase conditions were proved to be better than reverse phase conditions for this purpose, showing shorter analysis time and better performance of the classification models. The PLS-DA method showed the best classification results.

To our knowledge, there are no other antecedents describing the identification, comparison and/or classification of avocado fruits according to both their botanical and geographical origin using the chromatographic fingerprinting from the fat fraction merging with chemometric tools.

In this paper, experimentally measured multivariate data based on UV-absorption high performance liquid chromatography in normal phase ((NP)HPLC-DAD) fingerprints of the oil fraction of avocado fruits, jointly with deep learning chemometric methods, have been successfully apply with the objective of differentiating and

classifying a higher variability of avocado samples. Two different authentication goals have been pursued attending to: (i) the geographical origin, and (ii) the cultivar. Each classification method has been externally validated by establishing both the success/error contingency and proper quality performance metrics such as sensitivity, specificity, precision, or efficiency among others.

2. Materials and methods

2.1. Chemicals

All solvents used were HPLC grade. N-hexane was supplied by Honeywell (Charlotte, NC) and isopropanol was provided by Panreac Química (Barcelona, Spain). Vitamin D3 99% supplied by Alfa Aesar (Haverhill, MA) was used as control standard.

2.2. Samples and sample preparation

One of the biggest challenges for the development of non-targeted chromatographic methods aimed at food authentication is to become available representative samples of geographic origin, plant cultivar, season, processing conditions and other factors under consideration. A related challenge is the number of samples required for method development: the minimum number of samples needed to develop and validate a robust method depends upon the applicability statement, variation in the food matrix designated, and the characteristics analysed [**16**].

Table 1. Both varietal and geographical description of the collected avocado samples.

Variety	Nº Samples	Country	Origin class
Bacon	14	Spain	Europe
Ettinger	5	Israel	—
Fuerte	14	Spain	Europe
Hass	18	Spain	Europe
	12	Kenya	Africa
	10	South Africa	Africa
	8	Mexico	America
	8	Peru	America
Pinkerton	3	Spain	Europe
Topa-Topa	12	Spain	Europe
Unknown	4	Spain	Europe

Taking that into account, 108 avocados were provided by several orchards or purchased in grocery markets in the domain of Granada (Spain) from different varieties, origin and ripeness characteristics, trying to cover as much variability as possible. A briefly description about the geographical origin and the botanical variety of the samples is shown in Table 1.

Prior to the chromatographic analysis, avocados were lyophilized and the oil fraction was extracted using a n-hexane/isopropanol (3:2, v/v) mixture. The detailed procedure and the equipment used were described in a previous paper [**15**].

2.3. Chromatographic equipment and conditions

(NP)HPLC-DAD analysis was performed using an Agilent 1260 series liquid chromatograph (Santa Clara, CA) equipped with a CH30 column thermostat (Eppendorf, Hamburg, Germany), a quaternary pump and degasser auto sampler. Detection was performed with a diode-array detector (DAD). Agilent ChemStation OpenLab CDS software (rev. C.01.09) for LC systems was used to collect and process data. The UV spectra were recorded from 190 to 400 nm. 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 26 min.

In these circumstances, the obtained fingerprint depends on the composition of the major chemical fraction, *i.e.*, the triglycerides or triacylglycerols. Note that only the unsaturated fatty acids (more properly, the triacylglycerols having at least one unsaturated fatty acid mainly pamitoleic, C16:1; oleic, C18:1; and linoleic, C18:2 [¡Error! Marcador no definido.]) yield a measurable signal as the saturated fatty acids are nearly transparent to the UV-absorption detector. Consequently, in this instance the fingerprint has a high degree of specificity in respect to the distribution of the unsaturated fatty acids amount the various triacylglycerols. This characteristic causes that UV-absorption fingerprints were different from the fingerprints obtained from other universal detectors used in vegetable oil analyses as corona charged aerosol detector (CAD) [**15**].

2.4. Chemometrics

The raw data files from each chromatogram were obtained in a CSV (comma-separated values) file, and then converted to MATLAB format (R2017b version, The Mathworks Inc. Natik, MA). The data matrix of each sample was described by 7501×212 variables. Wavelength and time interval of interest were selected. Previous instrumental studies carried out by authors have shown that 220 nm and 1 to 17 minutes was the region of interest. It was selected for being the interval in which the

samples showed the greater significant absorbance and the greater differences each other, discarding the analysis of variables outside this range.

Then, preprocessing of the data was carried out to eliminate all the excess of information that masks signal of interest. Raw chromatograms need to be submitted to mathematical transformation in order to minimize the no informative spectral variability as well as to maximize the comprehensive interpretation of chemical information extracted from data, to gather deep knowledge about the system. A home-made MATLAB function, named 'MEDINA' (version 07) [**17**] was used with this purpose. The function employs some algorithms contained in MATLAB Bioinformatics Toolbox™ software and the 'icoshift' algorithm (version 1.2) to align the chromatographic signals.

The data pre-processing stages were: (1) Chromatograms data grouping and overlay; (2) De-noising and smoothing the chromatographic signal using a least-squares digital polynomial filter. For it, the MATLAB “mssgolay” function is used; this filter smooths the raw noisy signal data preserving the sharpness (or high-frequency components) of the peaks in the chromatogram; (3) Baseline correction. The baseline of the chromatogram corresponds to the detector signal in the absence of eluted compounds. The baseline can vary from one chromatogram to another due to different factors such as temperature, pressure or variations in the detector. The MATLAB "msbackadj" function was used to estimate a low-frequency baseline, which is hidden among high-frequency noise and signal peaks. It then subtracts this baseline from the all chromatograms; (4) Peaks alignment using 'icoshift' algorithm which employs a fast Fourier transform (FFT) engine that aligns all chromatograms simultaneously; and finally, (5) Mean centring of the data set was applied prior to the statistical analysis.

Once the chromatographic data pre-processing was carried out, then appropriate chemometric tools as pattern recognition analysis including unsupervised methods: PCA, HCA; and supervised ones: PLS-DA, SVM were used. All chemometric treatments were carried out using the PLS_Toolbox (version 8.7, Eigenvector Research Inc., Wenatchee, WA).

2.4.1. Unsupervised pattern recognition techniques

Unsupervised pattern recognition identifies the underlying patterns inherent to a data set. In unsupervised techniques, the grouping is carried out taking into account only similarities and differences between samples, without using any additional prior information about them. In these techniques the rule for grouping samples is not often known, neither the number nor the identity of the classes. These tools are also employed to study the assembling and to observe the natural relationships of the data although the information is not previously available [**18**]. Among them, PCA and

HCA are well-known and commonly used techniques aimed to graphically show the natural grouping of the samples or the similarities between samples in study [**19**,**20**].

2.4.2. Supervised pattern recognition techniques

Supervised pattern recognition techniques use the information about the class membership of the samples to define the criteria that allow, whenever a new unknown sample is analysed, the prediction of their category on the basis of their pattern of measurements. The development of a pattern recognition supervised model involves two stages: (i) in the first stage the model is designed and developed using a set of samples (training set) whose class is known, and (ii) in the second stage the performance of the model is externally validated using additional samples (validation set) which are not part of training set [**21**].

In our case, the original dataset was previously split into a training set and an external validation set applying the CADEX algorithm, develop by Kennard and Stone [**22**], at least 30% of the samples from each class were set to define the external validation set while the remaining samples constitutes the training set.

Traditionally well-known conventional supervised methods such as kNN, SIMCA, PLS-DA, have been used in the area of food analytical chemistry. Nevertheless, in recent years, the applications of new pattern recognition algorithms are growing in the area of food, due to their advantages and potential to solve complex problems related to food authenticity. The most widely used one is support vector machine (SVM). Further explanation of the technique will be given taken into account that it is a method that is not particularly widespread and that it has been the main tool used in this work.

PLS-DA method involves performing a multivariate regression model to establish class limits and placing a numeric value to each object/sample first, and then classifying them into a specific class [**23**].

Support Vector Machines (SVM), is a classifier which aims to find a hyperplane with maximum margin to separate the classes of data. As PLS-DA, SVM is also applicable both to solve classification and regression problems. For classification purposes, SVM aims to obtain the 'optimal' boundary of different classes in a vector space independently on the probabilistic distributions of training vectors in the data set. SVM-C project X-data of the training set into a feature space of usually much higher dimension than the original data space. In the feature space a hyperplane (also called decision plane) is computed that separates individual points of known group membership. The best discriminating separation is achieved by maximizing the margin between groups. Points in the feature space that define the margin are called support vectors.

The margin is the minimal distance from the separating hyperplane to the closest data points. So, SVM learning machine seeks for an optimal separating hyperplane where the margin is maximal. This hyperplane is represented by the following equation, where *w* is a weight vector, *x* is the input data and *b* is a bias:

$$\mathrm{w} \cdot \mathrm{x} + \mathrm{b} = 0$$

SVM algorithm uses a set of mathematical functions named kernel. The kernel functions transform the original data into the required format.

When the classes are separated by a non-linear boundary, the kernel method consists of a transformation of the vector space itself to a higher dimensional space, where the samples are projected by means of a feature function or kernel function. The kernel function Φ(x) performs a non-linear mapping into feature space. Each kernel has a set of parameters that must be tuned. Radial basis function (RBF) is particularly widespread because it requires only one tuneable parameter (the radial width σ). RBF is defined as the following equation where x_i, x_j are two generic training samples.

$$\mathrm{K}(\mathrm{x_i}, \mathrm{x_j}) = \exp\left(-\frac{\left|\mathrm{x_i} - \mathrm{x_j}\right|^2}{2\sigma^2}\right)$$

This kernel function on SVM shows good features and strong learning capability [**21**,**24**].

3. Results and discussion

A two-way data array was acquired and recorded for each avocado sample. As it was stated above, the chromatogram profile at 220 nm for each sample was selected as characteristic fingerprint to build the different multivariate methods. Figure 1a illustrates the heat map of a Spanish avocado sample from Hass cultivar; Figure 1b shows the superposed chromatographic profile at 220 nm of the 108 avocado samples.

Geographical origin differentiation

The first goal was the differentiation of avocados according to geographical origin. Three classes were defined for this purpose: Europe, Africa and America. Ettinger samples, from Israel, were discarded as belonging to Asia class on this goal because Ettinger cultivar is not representative of Asia's avocado production and we were not able to have any other Asiatic cultivar samples.

PCA was carried out for exploring whether there were natural groupings of the different avocado oils. Three principal components (PCs) were enough to explain 90.85% of the total variance (PC1 72.02, PC2 13.42, and PC3 5.41%). The most of

African samples were grouped with negatives values of PC3 while the rest of classes were scattered throughout the region of confidence of the model without showing any natural trend.

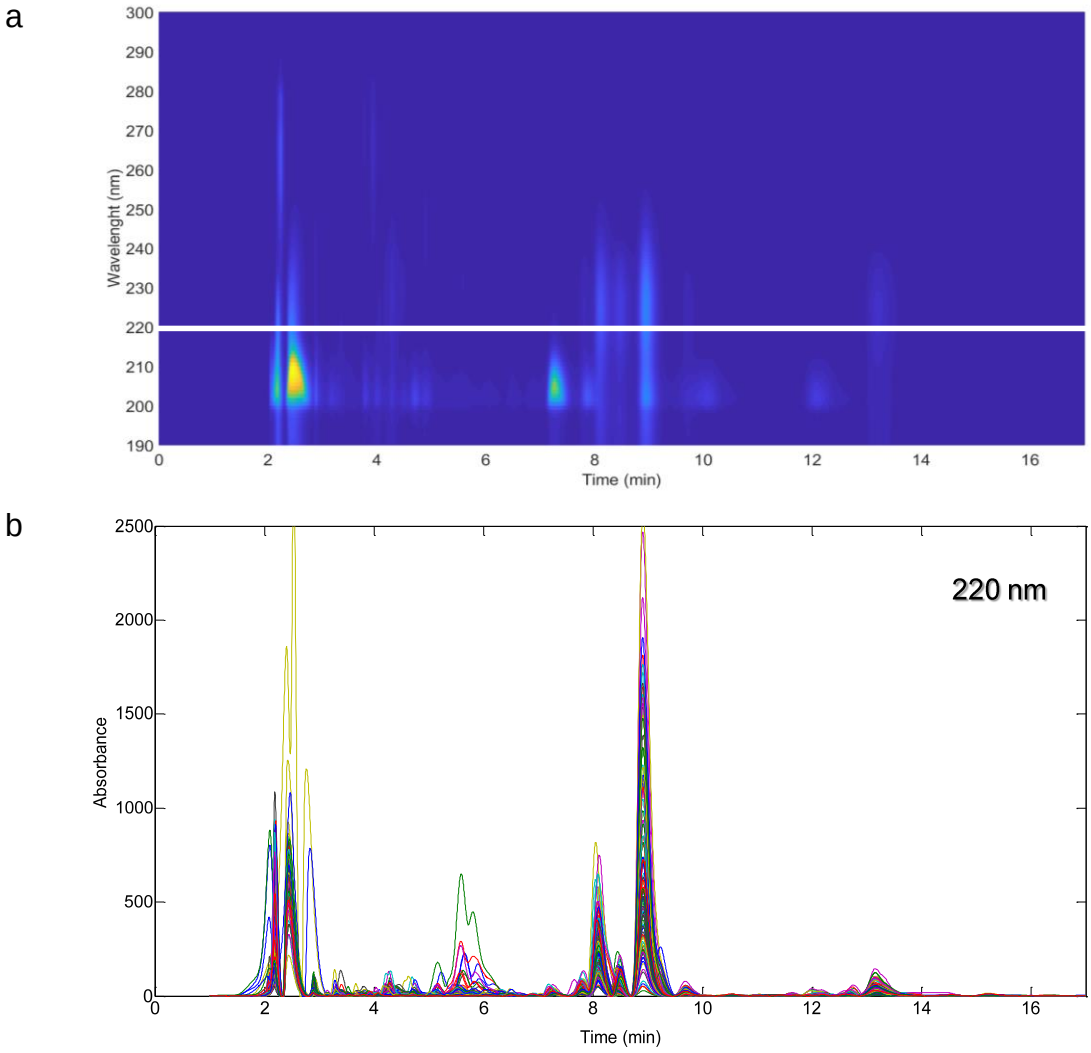


Figure 1. (a) Time-wavelength heat map of the 2D spectro-chromatografic signal of a sample of Hass avocado from Spain. The white line is indicating the wavelength selected; (b) superposed pre-processed (NP)HPLC-DAD chromatograms of the 108 avocado samples at 220 nm.

Then, HCA dendrogram was also perform based on Mahalanobis distance to evaluate the multivariate association between samples but no major groups of samples could be visualized and therefore it is not showed.

PLS-DA was the first technique used to establish suitable classification models for the discrimination between the three classes (three input-class classification, 3iC): Europe (class 1), America (class 2) and Africa (class 3). A training set of 78 samples having the three classes was used and a PLS-DA classification was performed with five latent variables (LVs), which were chosen using venetian blinds cross-validation

to minimize the root mean square error for cross validation (RMSECV) for each class. The remaining samples were used to externally validate the model. The model was able to differentiate the European training samples from the other two classes, however the African and American samples were considered as being the same class, and the 3iC classification model did not found differences between them. Moreover, 50% of the Europe class validation samples were misclassified.

In order to improve the obtained results, we build the 3iC model using SVM algorithm. The main advantage of SVM-C over PLS-DA is that it creates a separation between the regions of the different classes when they are not sufficiently evident. SVM classification was performed using the radial basis function (RBF) as kernel function in order to transform the original data into the required format, without variable reduction. Fig. 2 shows the SVM-C assignation of samples to each of the predefined classes.

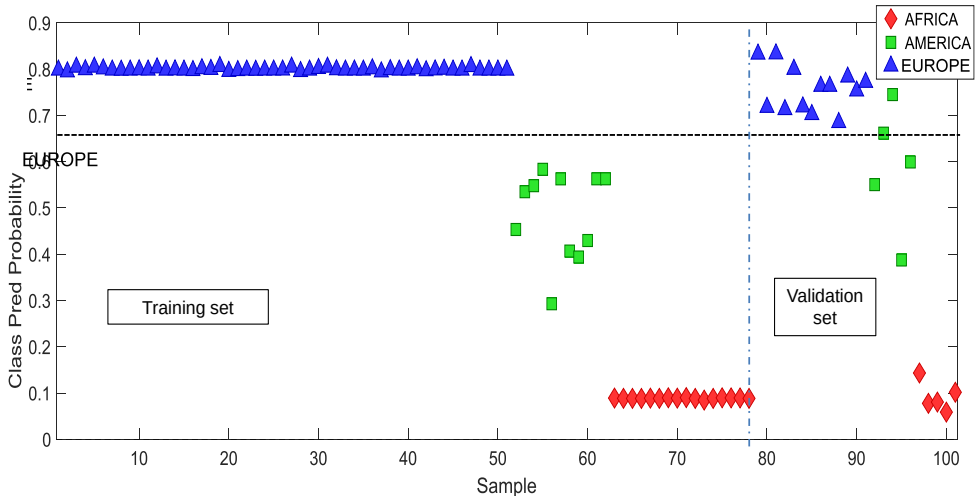


Figure 2. Sample assignment of Europe samples from the SVM-C model developed with three input-classes: Europe, America and Africa. Dashed line marks the classification threshold.

The horizontal line corresponds to the classification threshold. The SVM default classification criterion considers the classification threshold of 0.5 but this value may be *ad-hoc* modified by practitioners. A lower threshold will typically increase the sensitivity and reduce specificity and precision. Depends on the main goal of the classification method, the threshold can be set to balance the two [**16**]. The Europe/America/Africa model was developed searching that any African not American samples were assigned as belonging to the Europe class. Therefore, a high precision in the assignments to the Europe class is wanted in spite of having a lower sensibility. As it is showed in the Figure 2, threshold was established at 0.65; all European samples of the external validation set were predicted as belonging to the class Europa while only two American samples were misclassified. Contingence

table and quality performance metrics of the model are shown in tables 2 and 3 respectively. Each parameter has been estimated for each one of the three classes modelled with regard to the other two.

Table 2. Contingency chart for the three input-class SVM classification model established regarding the geographical origin goal.

		13	**5**	**5**	**23**
Assignment	Africa	0	0	5	**5**
	America	0	3	0	**3**
	Europe	13	2	0	**15**
		Europe	America	Africa	
		Reference			

Table 3. Classification performance features for the three input-class SVM classification model established regarding the geographical origin goal.

Performance features	**Europe**	**America**	**Africa**
Sensitivity (or Recall)	1.00	0.60	1.00
Specificity	0.80	1.00	1.00
Positive predictive value (Precision)	0.87	1.00	1.00
Negative predictive value	1.00	0.90	1.00
Youden index	0.80	0.60	1.00
Positive likelihood rate	5.00	–	–
Negative likelihood rate	0.00	0.40	0.00
F-measure	0.93	0.75	1.00
Discriminant power	–	–	–
Efficiency (or Accuracy)	0.91	0.50	1.00
AUC (Correctly classified rate)	0.90	0.80	1.00
Matthews correlation coefficient	–	–	1.00
Kappa coefficient	0.82	0.70	1.00

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

In order to improve the discrimination, two binaries (or two input-class, 2iC) models were then built: the first one allowed to discriminate between European and the rest of samples (Europe class was assigned with a value of 1 while the noEurope class

was assigned a value of 2) and the second one differentiated between American and African samples (America class was assigned a value of 1 and Africa a value of 2).

In both instances, SVM-C models showed better results than PLS-DA ones, and allowed the right classification of almost all the samples (only one American sample, in the America-Africa binary model, was located over the classification threshold, but having an slightly higher probability of belonging to Africa class, therefore, consider as a misclassified sample). The Europe-noEurope SVM-C model was a non-linear model while America-Africa SVM-C model was set up by selecting a linear function.

Cultivar differentiation

The second goal was aimed at differentiating avocados according to their botanical varieties. For this purpose two concatenated classification models were developed: (i) a binary model to differentiate between Hass/noHass samples, since the Hass cultivar is the most common variety worldwide, available all the year and cultivated around the world, and (ii) a ternary model to discriminate between Bacon, Fuerte and Topa-Topa classes. Before to the development of the classification models, a PCA model and a HCA dendrogram were inspected, but no natural groupings of the samples were intuitively distinguished and accordingly these graphical outlines are not showed.

Firstly, a PLS-DA model was performed. Four LVs were selected to build the model, explaining 91.77% of the variance. The Hass class was assigned with a value of 1, and the noHass class with a value of 0. Model was externally validated and 50% of the samples were wrong classified. All Bacon samples of the validation set were assigned to Hass class. Consequently, the PLS-DA model was no acceptable for the discrimination of Hass samples. Thus, an SVM-C model was carried out using a linear kernel function, without variable reduction. In this case, all the samples were properly classified and precision value of 1.00 regarding Hass class was achieved. The assignment of the samples is shown in Figure 3.

Once all the samples from Hass class were classified, the ternary model was built. As previously, a PCA model was developed but excluding the Hass samples. In this case, PCA allowed to visualize three groups of samples which are correlated with Bacon (negative scores for both PC3 and PC1), Fuerte (negative scores for PC1 and both positive and negatives for PC3) and Topa-Topa (negative scores for PC3 and positive scores for PC1) varieties. PC1-PC3 plot from the PCA model is displayed on Figure 4.

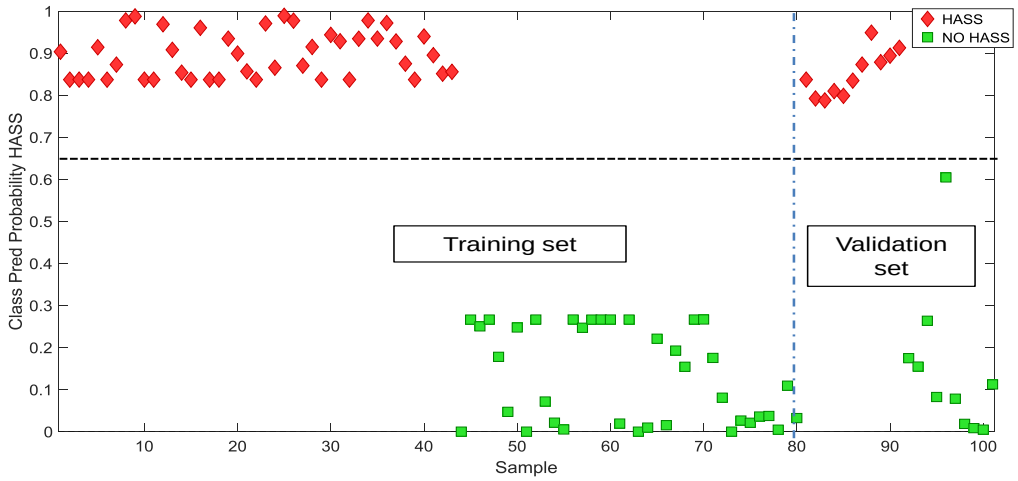


Figure 3. SVM classification plots of samples from Hass-noHass classes. Dashed line marks the classification threshold.

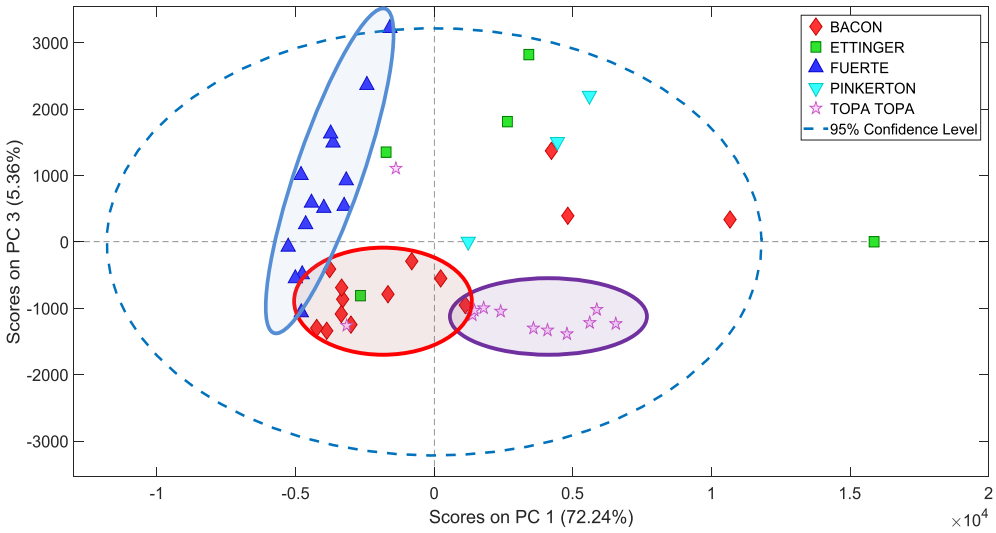


Figure 4. PCA scores plot in PC1-PC3 plane for samples of Fuerte, Bacon, Ettinger, Pinkerton and Topa-Topa classes.

Using the same strategy previously describe, PLS-DA was the first classification algorithm used to establish the model but almost all samples of the validation set were misclassified again. Therefore, a SVM-C model was carried out using a linear kernel function. Three input-classes (3iC): 'Bacon', 'Fuerte' and 'Topa-Topa' were considered to build the model and finally the 100% of the validation set samples were correctly assigned to their related class. Precision and sensitivity values of the three classes were equal to 1.00. As example, the SVM-C assignation to class Bacon is shown in Figure 5.

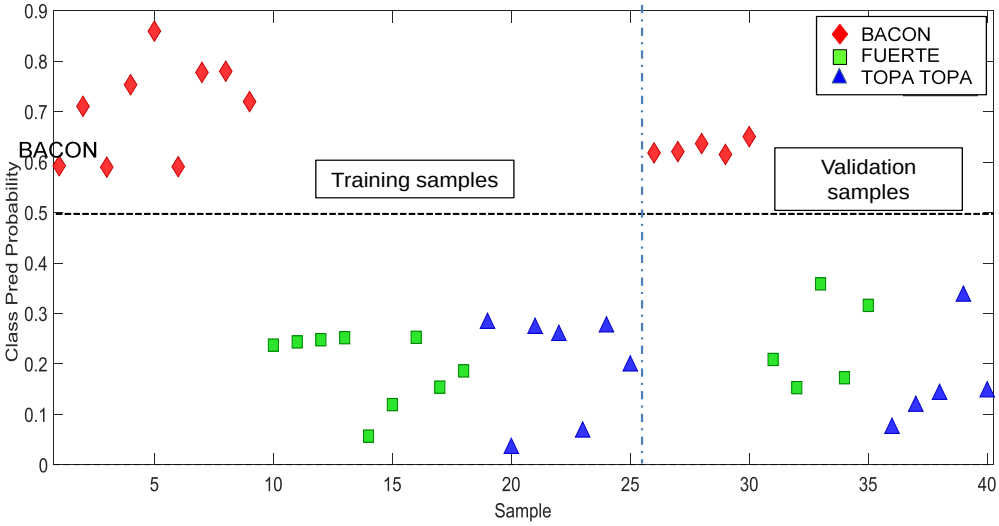


Figure 5. Sample assignment of Bacon samples from the SVM-C model developed with three input-classes: Fuerte, Bacon and Topa-Topa. Dashed line marks the classification threshold.

Two varieties, Ettinger and Pinkerton, had not been considered in the previous 3iC model. The low number of available samples from these varieties made no possible developing and properly validating a robust classification method. However, a testing set was set up with 5 Ettinger and 3 Pinkerton samples to check whether these samples were similar to any of the previous modelling classes (Fuerte, Bacon and Topa-Topa).

A new 3iC SVM-C lineal model was established with a training set feedback with all the samples previously used in respective training and validation sets. All Pinkerton samples were assigned with a higher probability as belonging to Topa-Topa class while Ettinger samples were equally distributed between the three classes. The results showed that Ettinger samples were clearly different to any of the previous classes but Pinkerton samples seem to be quite similar to Topa-Topa ones. Further studies will be needed, with a high number of samples, to verify whether if it is possible to distinguish between them.

4. Conclusion

In this study, a methodology for distinguishing avocado fruits using their entire lipid chromatographic fingerprint jointly with deep learning chemometric methods is described. Different multivariate classification strategies have been tested and the differentiation of avocados according to their geographical origin and their cultivar has been achieved. To assess the different classification scenarios and describe the performance of the applied classification methods, several quality classification metrics have been calculated. The SVM classification methods have been compared to PLS-DA ones. On our previous paper, PLS-DA was highly successful in the

classification of avocado samples. In this study the performance of PLS-DA appears to be any poor. A higher number of avocado samples was used (108 fruits), more varieties had been considered (6 different cultivars) and a different not universal detector (DAD) had been employed to record the fingerprints. Moreover, fingerprints of two avocados of the same cultivar but harvested in different continents show higher similarity than the fingerprints of two avocados from different varieties. All these facts make necessary the use of appropriate classification and prediction methods capable of finding significant differences among very similar fingerprints, furthermore when both different origin and cultivar are merged in the same model. According to the obtained results, SVM-C has been proved to be a better classifier than PLS-DA and it should be emphasized that only one avocado of the entire external validation data set has been misclassified.

Also, it should be noticed that the established models allow predicting whether two different cultivars, not previously trained, show similarities with any of the modelling classes.

Acknowledgements

We are grateful to Salvador Salazar for freely providing a significant number of samples from his avocado plantation located at Salobreña (Granada, Spain) to be included in this study. In addition, one of the authors, S.M.T. wants to express their sincere gratitude to the Spanish Ministry of Sciences. Innovation and Universities for a pre-doctoral fellowship (project RTC-2017-6170-2).

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] R.G. Araújo, R.M. Rodriguez-Jasso, H.A. Ruiz, M.M.E. Pintado, Avocado by-products: Nutritional and functional properties, Trends Food Sci Tech. 80 (2018) 51–60. Doi: 10.1016/j.tifs.2018.07.027.

[2] U.S.D.A. Research Service, Food Data Central Search Results, https://fdc.nal.usda.gov/fdc-app.html#/?query=avocado, 2019 (accessed 13 September 2019).

[3] Statistics Division of the Food and Agriculture Organization for the United Nations. FAOSTAT data. http://www.fao.org/faostat/en/#data/QC, 2017 (accessed 12 September 2019).

[4] European Union, Regulation (EC) No 387/2005 amending Regulation (EC) No 831/97 laying down marketing standards applicable to avocados, Official Journal of the European Union. L62 5–6, 2005.

[5] M. Flores, C. Saravia, C. E. Vergara, F. Avila, H. Valdés, J. Ortiz Viedma, Avocado Oil: Characteristics, Properties, and Applications, Molecules. 24 (2019) 1–21. Doi: 10.3390/molecules24112172.

[6] F. Takenaga, K. Matsuyama, S. Abe, Y. Torii, S. Itoh, Lipid and fatty acid composition of mesocarp and seed of avocado fruits harvested at northern range in japan, J. Oleo. Sci. 57 (2008) 591–597. Doi:10.5650/jos.57.591.

[7] R. Ferreyra, G. Sellés, J. Saavedra, J. Ortiz, C. Zúñiga, C. Troncoso, S.A. Rivera, M. González-Agüero, B.G. Defilippi, Identification of pre-harvest factors that affect fatty acid profiles of avocado fruit (Persea americana Mill) cv. 'Hass' at harvest, S. African. J. Bot. 104 (2016) 15–20. Doi: 10.1016/j.sajb.2015.10.006.

[8] M. Donetti, L.A. Terry, Biochemical markers defining growing area and ripening stage of imported avocado fruit cv. Hass, J. Food Compos. Anal. 34 (2014) 90–98. Doi: 10.1016/j.jfca.2013.11.011.

[9] N. Yanty, J. Marikkar, K. Long, Effect of varietal differences on composition and thermal characteristics of avocado oil, J. Am. Oil Chem. Soc. 88 (2011) 1997–2003. Doi: 10.1007/s11746-011-1877-x.

[10] L. Cuadros Rodríguez, C. Ruiz Samblás, L. Valverde Som, E. Pérez Castaño, A. González Casado, Chromatographic fingerprinting: An innovative approach for food 'identitation' and food authentication – A tutorial, Anal. Chim. Acta 909 (2016) 9–23. Doi: 10.1016/j.aca.2015.12.042.

[11] A. Malechaux, N. Dupuy, J. Artaud, Applications of vibrational spectroscopy techniques, in: M. G. Kontominas (Ed.), Authentication and Detection of Adulteration of Olive Oil, Nova Science Publishers, 2018, pp. 7-28. ISBN: 978-1-53614-596-0.

[12] S. del Caño Ochoa, A. Ruiz Aracama, M. D. Guillén Lorén, Potential of nuclear magnetic resonance for a discriminant characterization of PDO VOOs, Eur. J. Lipid Sci. Technol. 121 (2019) 1800137. Doi: 10.1002/ejlt.201800137.

[13] J.M. Bosque Sendra, L. Cuadros Rodríguez, C. Ruiz Samblás, A.P. de la Mata, Combining chromatography and chemometrics for the characterization and authentication of fats and oils from triacylglycerol compositional data - A review, Anal. Chim. Acta 724 (2012) 1–11. Doi: 10.1016/j.aca.2012.02.041.

[14] F. Marini, R. Bucci, A.L. Magrì, A.D. Magrì, An overview of the chemometric methods for the authentication of the geographical and varietal origin of olive oils, in: V. R.

Preedy, R. R. Watson (Eds.) Olives and Olive Oil in Health and Disease Prediction, Academic Press, 2010, pp. 569–579. Doi: 10.1016/B978-0-12-374420-3.00062-0.

[15] S. Martín Torres, A. M. Jiménez Carvelo, A. González Casado, L. Cuadros Rodríguez, Differentiation of avocados according to their botanical variety using liquid chromatographic fingerprinting and multivariate classification tree, J. Sci. Food Agric. 99 (2019) 4932–4941. Doi: 10.1002/jsfa.9725.

[16] B. Gao, S.E. Holroyd, J.C. Moore, K. Laurvick, S.M. Gendel, Z. Xie, Opportunities and challenges using non-targeted methods for food fraud detection, J. Agric. Food Chem. 67 (2019) 8425–8430. Doi: 10.1021/acs.jafc.9b03085.

[17] 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. Anal Methods 7 (2015) 4192–4201. Doi: 10.1039/C5AY00168D.

[18] A. Gredilla, S. Fdez Ortiz de Vallejuelo, A. de Diego, J.M. Madariaga, J.M. Amigo, Unsupervised pattern-recognition techniques to investigate metal pollution in estuaries, Trends Anal. Chem. 46 (2013) 59–69. Doi: 10.1016/j.trac.2013.01.014.

[19] S. Mishra, U. Sarkar, S. Taraphder, S. Datta, D. Swain, R. Saikhom, S. Panda, M. Laishram, Principal component analysis, Int. J. Livest. Res. 7 (2017) 60–78. Doi: 10.5455/ijlr.20170415115235.

[20] M. Anandarajan, C. Hill, T. Nolan, Cluster analysis: modeling groups in text, in: M. Anandarajan, C. Hill, T. Nolan (Eds.), Practical Text Analytics: Maximizing the Value of Text Data, Springer International Publishing, Cham, 2019, pp. 93–115. Doi: 10.1007/978-3-319-95663-3_7.

[21] L. A. Berrueta, R.M. Alonso-Salces, K. Héberger, Supervised pattern recognition in food analysis, J. Chromatogr. A 1158 (2007) 196–214. Doi: 10.1016/j.chroma.2007.05.024.

[22] R. W. Kennard, L. A. Stone, Computer aided design of experiments. Technometrics 11 (1969) 137–148. Doi: 10.1080/00401706.1969.10490666.

[23] A. M. Jiménez Carvelo, A. González Casado, E. Pérez Casraño, L. Cuadros rodríguez, LC fingerprinting approach for authenticating olive oil, in: M. G. Kontominas (Ed), Authentication and Detection of Adulteration of Olive Oil, Nova Science Publishers, 2018, pp. 7-28. ISBN: 978-1-53614-596-0.

[24] Y. Xu, S. Zomer, R. G. Brereton, Support vector machine: a recent method for classification in chemometrics, Crit. Rev. Anal. Chem. 36:3-4 (2006) 177-188. Doi: 10.1080/10408340600969486.

CAPÍTULO III

AGNOSTIZACIÓN DE SEÑALES CROMATOGRÁFICAS

CAPÍTULO III

Agnostización de señales en cromatografía líquida

III.1. Presentación

El grupo de investigación "Análisis en Alimentación y Medio Ambiente" (Cod. PAIDI: FQM-232), ha propuesto, desarrollado y aplicado una metodología para conseguir señales cromatográficas independientes del instrumento, y del estado del propio instrumento, que han sido denominadas como "señales agnostizadas", tanto para cromatografía de gases, como para cromatografía de líquidos. Este capítulo muestra la propuesta aplicada a la modalidad de cromatografía líquida acoplada a diversos sistemas de detección.

Los objetivos de este capítulo fueron:

- Elegir una serie de compuestos patrón que cubran un alto rango de polaridad y cubran los tiempos de retención más habituales (considerando un método referencia que presente un gradiente lineal de fases móviles).
- Desarrollar una metodología para obtener índices relacionados con el tiempo de retención-elución e independientes del sistema cromatográfico.
- Agnostizar huellas cromatográficas. Normalizar los valores de intensidad y los tiempos de retención de forma que para una muestra en particular y siempre que se aplique el mismo método cromatográfico, cada singularidad de su cromatograma tenga una intensidad similar y le corresponda una posición similar en el eje que define la retención, independientemente del estado cromatográfico del instrumento, o del propio instrumento utilizado.

Para la consecución de estos objetivos se trabajó con cromatografía líquida en modalidad fase invertida acoplado a un detector de absorción UV-Vis de fila de diodos (DAD) y a un espectrómetro de masas (MS). Las publicaciones obtenidas fueron:

1. PUBLICACIÓN IV: Standardization of chromatographic signals – Part II: Expanding instrument-agnostic fingerprints to reverse phase liquid chromatography.
2. PUBLICACIÓN V: Instrumental-agnostizing methodology for liquid chromatography-mass spectrometry systems.

III.2. Introducción

Desde su invención, la cromatografía ha evolucionado hasta convertirse en una técnica analítica flexible con múltiples aplicaciones tanto en el ámbito académico como en el industrial, considerándose el método más versátil de todos los métodos de análisis químico [**1**]. Considerando la cromatografía de líquidos (LC), un analito se distribuye entre una fase estacionaria y una fase móvil. El tiempo que tarda el analito en migrar a través de la columna se denomina tiempo de retención, el cual depende del tiempo muerto de la columna y el factor de retención del analito en las dos fases que varía en función del pH, la temperatura, la composición de la fase móvil, etc. [**2**].

El avance de la cromatografía especialmente en las últimas décadas, con instrumentos cada vez más sofisticados que permiten trabajar en condiciones de ultra-alta eficiencia (UHPLC), permitiendo separaciones más rápidas y eficientes, y el desarrollo de sistemas de medida y 'software' que permiten manejar gran cantidad de datos, así como la adopción de nuevos enfoques de trabajo como la modalidad de huella instrumental o "fingerprinting" [**3**] y sumando a todo esto el empleo de herramientas quimiométricas, han solucionado muchos de los problemas analíticos existentes, sobre todo en las áreas de las ciencias de la vida y de los alimentos. Sin embargo, cualquier conjunto de condiciones cromatográficas encontradas tras aplicar un proceso de optimización será inútil si son tan críticas que el cromatograma obtenido no puede volver a reproducirse experimentalmente.

Cuando en LC se trabaja en la modalidad de gradiente, por ejemplo, es necesario que el equipo cromatográfico tenga un dispositivo capaz de realizar mezclas de la fase móvil con un control preciso y reproducible. En cualquier condición, la falta de robustez del método desarrollado puede deberse a fuentes de error relacionadas con el equipo, la columna, las variables ambientales no controladas, fallos en la mezcla de disolventes y de los

1. Azevedo Calderón, L. Chromatography. The Most Versatile Method of Chemical Analysis, Intech, Rijeka, 2012.
2. Den Uijil, M.J., Schoenmakers, P.J., Pirok, B.W.J., Van Bommel, M.R. Recent applications of retention modelling in liquid chromatography. Journal Sep. Sci., 44 (2021), 88-114. https://doi.org/10.1002/jssc.202000905
3. Cuadros Rodríguez, L., Ruiz Samblás, C., Valverde Som, L., Pérez Castaño, E., González Casado, A. Chromatographic fingerprinting: an innovative approach for food 'identitation' and food authentication – a tutorial. Anal. Chim. Acta, 909 (2016), 9-23. https://doi.org/10.1016/j.aca.2015.12.042

modelos de retención [**4**]. Por ello, la misma muestra analizada repetidamente en el mismo sistema (equipo) y con el mismo método puede mostrar una dispersión aleatoria en los valores de los tiempos de retención y en la intensidad de la señal registrada en un plazo corto de tiempo, y una evidente deriva en sus volares a largo plazo.

De igual forma, si repetimos un análisis cromatográfico en otro equipo o en el mismo equipo en el transcurso del tiempo, es habitual encontrar desplazamientos en los tiempos de retención de los analitos y variaciones en la intensidad de la señal cromatográfica. Esto puede dar lugar a inconvenientes cuando la cromatografía se acopla a un espectrómetro de masas (MS) para identificar compuestos desconocidos, cuantificar compuestos conocidos y/o dilucidar la estructura y las propiedades químicas de las moléculas. Pero además, pequeñas desviaciones muestran aún más importancia cuando todo el cromatograma se utiliza como dato tensorial de entrada para crear un modelo multivariable, ya que estos desplazamientos en los ejes de tiempos e intensidades, dificultan el desarrollo y validación de modelos creando clasificaciones erróneas y pudiendo llegar a predicciones no confiables.

En bibliografía existen diferentes antecedentes para intentar solventar el problema del eje de tiempos (t_R) en LC. La mayoría se basan en el uso de índices de retención ampliamente descritos para su uso en cromatografía de gases (GC) y conocidos como índices de Kovats [**5**]. Kovats propuso el uso de una serie homóloga de n-alcanos como patrones externos para fijar el comportamiento de retención en GC de cualquier analito que se analice bajo las mismas condiciones experimentales. A cada compuesto estándar, le atribuyó arbitrariamente un índice de retención en función del número de carbonos de su cadena alquílica. Por analogía con la GC, al principio, cuando los detectores espectrométricos (absorción UV) eran los más utilizados en LC, algunos estudios examinaron el uso de n-alcanos como estándares para aplicaciones de LC, pero su difícil detección con este tipo de detectores, así como su baja polaridad, llevaron a los autores a probar la utilización de otras series homólogas como estándares alternativos.

---

4. García Álvarez Coque, M.C., Torres Lapasió, J.R., Baeza Baeza, J.J. Models and objective functions for the optimisation of selectivity in reversed-phase liquid chromatography. Anal. Chim. Acta, 579 (2006), 125-145. https://doi.org/10.1016/j.aca.2006.07.028
5. Kováts, V.E. Gas-chromatographische charakterisierung organischer verbindung. teil 1: retentionsindices aliphatischer halogenide, alkohole, aldehyde und ketone. Helv. Chim. Acta, 206 (1958), 1915-1932. https://doi.org/10.1002/hlca.19580410703

Basándose en este primer antecedente, Baker and Ma [**6**] propusieron el uso de una serie de cetonas como serie estándar en condiciones isocráticas; Smith *et al.* [**7**] apostaron por el uso de alquil aril cetonas; Bogust *et al.* [**8**] describieron el uso de 1-nitroalcanos para establecer nuevos índices de retención; mientras que otros autores propusieron el uso de una escala de índice de retención basada en una serie de hidrocarburos aromáticos policíclicos (PAH) o incluso de sus derivados nitrados (Nitro-PAH) que contienen un número creciente de compuestos aromáticos [**9**]; entre otros.

A pesar de todos estos esfuerzos para aplicar el sistema de identificación basado en los índices de retención en LC durante las décadas de los 80-90, la menor repetibilidad de los tiempos de retención con respecto a la GC, la mayor gama de compuestos susceptibles de LC, el mayor número de opciones de elución (diferente composición de la fase móvil) y las diferencias en tipo de equipos y columnas han limitado el número de aplicaciones satisfactorias de estos índices en LC. La elección de los compuestos de referencia juega un papel clave, ya que el índice de retención calculado para el analito se refiere a los de dos compuestos estándar eluídos inmediatamente antes y después de él, garantizando así una compensación más precisa de las variaciones instrumentales y/o analíticas. En GC la búsqueda de una serie homóloga de referencia adecuada no es una tarea compleja, ya que la elución se produce básicamente en función del aumento del punto de ebullición. Sin embardo en LC entran en juego más interacciones del analito con ambas fases, lo cual juega un papel fundamental en el proceso de elución, y cada aplicación podría necesitar una serie de estándares específica, que eluya satisfactoriamente en las condiciones empleadas, más aún con las dos modalidades de trabajo posibles, cromatografía en fase normal y en fase invertida [**10**].

---

6. Baker, J.K., Ma, C.Y. Retention index scale for liquid-liquid chromatography. J. Chromatogr. A, 169 (1979), 107-115. https://doi.org/10.1016/0021- 9673(75)85036- 9
7. Smith, R.M. Alkylarylketones as a retention index scale in liquid chromatography. J. Chromatogr. A, 236 (1982), 313-320. https://doi.org/10.1016/S0 021-9673(0 0) 84880-3
8. Bogusz, M., Aderjan, R. Improved standardization in reversed-phase high-performance liquid chromatography using 1-nitroalkanes as a retention index scale. J. Chromatogr. A, 435 (1988), 43-53. https://doi.org/10.1016/S0021-9673(01)82161-0
9. Smith, R.M. Retention index scales used in high-performance liquid chromatography, in: R.M. Smith (Ed.), Retention and Selectivity in Liquid Chromatography, Elsevier Science B.V., Amsterdam, (107-122), 1995.
10. Rigano, F., Arigò, A., Oteri, M., La Tella, R., Dugo, P., Mondello, L. The retention index approach in liquid chromatography: an historical review and recent advantages. J. Chromatogr. A, 1640 (2021), 461963, 1-13. https://doi.org/10.1016/j.chroma.2021.461963

En la última década, tras los notables avances en el campo de la MS y en el empleo de detectores de MS en tándem, se han publicado trabajos con objetivo de estandarizar el tiempo de retención de LC-MS y proporcionar un parámetro de identificación adicional para la detección [**11**, **12**, **13**] en el área de la alimentómica (o 'foodomica'). Sin embargo, tanto los antecedentes más antiguos como los recientes intentos señalaron la imposibilidad de establecer en LC una escala de índice de retención universal. Además, no se han encontrado publicaciones que usen los compuestos propuestos para ser usados como dichas series homólogas para el cálculo de los correspondientes índices de retención, probablemente por su falta de disponibilidad, alto coste, además de, en algunos casos, presentar una toxicidad elevada. En este capítulo se presentan un capítulo de libro y un artículo publicados, donde se propone por primera vez la metodología de agnostización para la obtención de señales cromatográficas invariantes e independientes del sistema cromatográfico.

En el artículo, se propone por primera vez, el uso de esta metodología, no con fines de identificación, sino para el alineamiento de huellas cromatográficas obtenidas por LC en fase invertida, haciendo que la señal o huella instrumental obtenida, sea independiente del estado cromatográfico o de la fecha de análisis y que las huellas cromatográficas de dos estados instrumentales diferentes tengan un alto grado de similitud. Esto permitiría establecer bases de datos cromatográficos característicos para armonizar, recoger, analizar y compartir información de análisis comparable para cada matriz alimentaria, que permita combinar los resultados obtenidos por diferentes estados cromatográficos en un escenario interlaboratorio.

En el capítulo de libro, se presenta la aplicación de esta metodología, basada en la retención de una nueva serie de patrones, que presentan unos requisitos específicos, y que han resultado útiles para la detección y cuantificación de pesticidas cuando se emplean diferentes plataformas de espectrometría de masas (en modos de ionización positiva y negativa) acopladas a equipos UHPLC.

---

11. Quilliam, M.A. Retention index standards for liquid chromatography, WO2013134862A1, 2013.

12. Rigano, F., Oteri, M., Russo, M., Dugo, P., Mondello, L. Proposal of a linear retention index system for improving identification reliability of triacylglycerol profiles in lipid samples by liquid chromatography methods. Anal. Chem., 90 (2018), 3313-3320. https://doi.org/10.1021/acs.analchem.7b04837

13. Celma, A., Bijlsma, L., López, F.J., Sancho, J.V. Development of s retention time interpolation scale (RTi) for liquid chromatography coupled to mass spectrometry in both positive and negative ionization modes. J. Chromatogr. A, 1565 (2018), 101-107. https://doi.org/10.1016/j.chroma.2018.07.030

PUBLICACIÓN IV.

ARTÍCULO CIENTÍFICO

Standardization of chromatographic signals – Part II: Expanding instrument-agnostic fingerprints to reverse phase liquid chromatography

Journal of Chromatography A, 1641, 461973, 1-14, 2021.
https://doi.org/10.1016/j.chroma.2021.461973

Journal of Chromatography A 1641 (2021) 461973

Contents lists available at ScienceDirect

Journal of Chromatography A

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

Standardization of chromatographic signals – Part II: Expanding instrument-agnostic fingerprints to reverse phase liquid chromatography

Luis Cuadros-Rodríguez [a], Sandra Martín-Torres [a,*], Fidel Ortega-Gavilán [a], Ana M. Jiménez-Carvelo [a], Rosalía López-Ruiz [b], Antonia Garrido-Frenich [b], M. Gracia Bagur-González [a], Antonio González-Casado [a]

[a] Department of Analytical Chemistry, Faculty of Science, University of Granada, C/ Fuentenueva s/n, E-18071, Granada, Spain
[b] Department of Chemistry and Physics, University of Almería, Agri-food International Campus of Excellence, CeiA3, E-04120, Almería, Spain

Standardization of chromatographic signals – Part II: Expanding instrument-agnostic fingerprints to reverse phase liquid chromatography

Luis CUADROS-RODRÍGUEZ[(1)], Sandra MARTÍN-TORRES[(1,✉)], Fidel ORTEGA-GAVILÁN[(1)], Ana M. JIMÉNEZ-CARVELO[(1)], Rosalía LÓPEZ-RUIZ[(2)], Antonia GARRIDO-FRENICH[(2)], M. Gracia BAGUR-GONZÁLEZ[(1)], Antonio GONZÁLEZ-CASADO[(1)]

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

[(2)] *Department of Chemistry and Physics, University of Almeria, Agri-food International Campus of Excellence, CeiA3, E-04120, Almeria, Spain.*

Highlights

- Instrumental-agnostizing methodology proposal for RPLC-DAD fingerprints
- Internal standard intensity is used as reference for signal intensities normalization
- Standard retention scores are established as reference for signal alignment
- Different chromatographic states are tested
- Good results are obtained on similarity indices calculated after agnostizing

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

ABSTRACT

There is a large amount of literature relating to multivariate analytical methods using liquid chromatography together with multivariate chemometric/data mining methods in the food science field. Nevertheless, dating the obtained results cannot be compared as they are based on data acquired by a particular analytical instrument, thus they are instrument-dependent. Therefore, this creates difficulties in generating a database large enough to gather together all the variability of the samples. The solution to this problem is to obtain an instrument-agnostic chromatographic signal that is independent of the chromatographic state, i.e., measuring instrument or particular condition of the same instrument from which it was acquired.

This paper describes the methodology to be followed to obtain standardized instrumental fingerprints when liquid chromatography is used for prior separation. For this purpose both internal and external chemical standards series are used as references. As an application example, we have applied this methodology for the determination of biophenols in olive oil by liquid chromatography coupled to ultraviolet-visible detector (LC-UV), using three different LC-UV instruments. The instrument-agnostic fingerprints obtained show a high grade of similarity, regardless of the state of the chromatographic system or the time of acquisition.

Key words

Liquid chromatography; fingerprint; instrument-agnostizing; standard retention scores; signal alignment.

1. INTRODUCTION

This paper is the second part of a series of articles devoted to the *standardization of instrumental fingerprints* [**1**]. In this part, we have expanded the proposed methodology to reverse phase liquid chromatography (RPLC). As already indicated in the first part, instrumental fingerprints recall a comprehensive description of a test material that is carried out in a non-selective (untargeted) way. The untargeted approach for chemical analysis and instrumental fingerprinting are closely related, but they do not express exactly the same thing. Only when the complete raw signal is used, and no compound is identified/quantified, should the untargeted approach be considered as an extension of the instrumental fingerprinting methodology. It should be noted that the instrumental fingerprint is the raw signal after having been subjected to all signal treatments [**2**].

The general comment indicated above is obviously extendable to liquid chromatography in any usual mode (RPLC, normal phase NPLC, hydrophilic interaction HILIC, size exclusion SEC and cationic or anionic ion exchange IEC). Esteki *et al.* [**3**,**4**] have reviewed applications for food authentication as well as the different chemometric tools used for this purpose. Most of these applications (85%) are focused on RPLC, whilst only 15% corresponds to NPLC. In addition, HILIC fingerprinting has been described to determine the harvest year and varieties of white wines [**5**] and several applications devoted to the analysis of traditional Chinese medicinal herbs [**6**,**7**] have been recently published. There is a limited history of SEC fingerprinting. In 2009 two papers reported the characterization of oligomers [**8**] and polymers [**9**] but the authors only made a visual comparison of the SEC fingerprints. Only one fingerprinting application using SEC for the evaluation of polysaccharides [**10**] has been published. IEC is mainly used to quantify anions and cations by applying a targeted approach.

Liquid chromatographic fingerprinting depends on both separation and measuring performances. The detection/measuring device for acquiring chromatographic fingerprints could be classified according to their degree of selectivity and the grade of information they provide. Thus, low-selectivity universal detectors, such as the refractive index detector (RID), evaporative light scattering detector (ELSD), charged aerosol detector (CAD), UV-visible absorption detector at a specific wavelength (UV-Vis) and diode array detector (DAD), the so-called photodiode array detector (PDA), have been used in approximately 6%, 6%, 8%, 22% and 48% respectively, of the applications. Note that the DAD is the most widely used and has been used in almost half of the reported papers describing fingerprinting analytical applications. Electrochemical detectors (ECDs), based on both amperometric and coulometric measurements, are the second category since only

electrochemically active substances able to be oxidized or reduced could be detected (1% use). Finally, the fluorescence detector (FLD), which shows a high selectivity (*i.e.*, third category), is only used in 6% of the applications. Mass spectrometry (MS) measuring system already constitute the more promising and powerful analytical techniques. LC-MS is used for many applications, especially in the authenticity field, which has very high sensitivity and selectivity. MS operates by the formation of positively or negatively charged ions and subsequent detection and measurement of the mass to charge (m/z) ratio. It produces a mass spectrum representing the fingerprint of the sample components. A number of different MS fingerprinted set-ups based on the ionization technique and the mass analysers used are possible.

Untargeted studies may include hundreds or thousands of data. Their acquisition can often take a long time and both systematic and random errors may occur during the measurement stage. The information contained in the analytical signal (*i.e.*, chromatogram) depends on the chromatographic state. Note that chromatographic state refers to both the instrument and the particular conditions of the chromatographic equipment itself (*e.g.,* type and age of chromatographic column, status of coupled detector, time the analysis is performed, or software used for instrument command and data acquisition). In the case of liquid chromatography, undesirable information can be due to: temporary drift in the sensitivity of the instrument, changes in the flow or composition of the mobile phase, or fluctuations in column temperature. Each of these facts contaminates the analytical signal and reduces the quality of the information obtained.

As mentioned in part 1 of this series [**1**], data extracted from instrumental fingerprints can be arranged into a mathematical structure defined by an intensity data tensor INT where each i-th element of the tensor may be expressed as $INT_i = f(x_{1i}, x_{2i}, x_{3i}, \ldots)$ in which INT_i is the signal intensity of the i-th element from the tensor and $x_{1i}, x_{2i}, x_{3i}, \ldots$ are the scores which describe the position of the same element within the tensor. The number of scores defines the 'order' of the tensor.

In the case of liquid chromatography, it is possible to obtain data tensors of zeroth, first, second and third orders, depending on the detector/measuring device coupled and the data extracted from the acquired signal. The simplest universal detectors (RID, ELSD, CAD, UV-Vis) provide zero-order tensor (scalar) corresponding to a single peak area or height. They can also produce first order tensors such as $INT_i = f(t_{Ri})$ defining a continuous signal (1D chromatogram) or a sequence of scalars (peak areas or heights, after an integration process). The versatility of DAD can provide 3D signals (absorption spectrum- chromatograms) from which zero, first and second order data tensors could be extracted. Note that an absorption spectrum-chromatogram is

identified with a second order data tensor defined by two scores, x_1 retention time (t_R) and x_2 absorption wavelength (λ_{abs}).

FLD would bring an additional spectral dimension to the raw signal and if a total fluorescence spectrum-chromatogram is acquired, it consists of a third order tensor, which is characterised by three scores, x_1 retention time (t_R), x_2 excitation wavelength (λ_{exc}) and x_3 emission wavelength (λ_{em}). This 4D signal cannot be visualized, however it could be mathematically treated in order to obtain analytical outcomes [**11**]. Table 1 shows the types of data which could be extracted from each type of chromatographic signal described above.

The wide operational variability inherent to the control of liquid chromatograph results in the shifts/warping of retention times (t_R) (x_1, time dependent score) or drifts/fluctuations of the intensity values (I, detector dependent measurements). These variations do not affect the spectral scores (x_2, x_3) of the tensors from DAD and FLD signals as they depend exclusively on the optical system which is not subject to run-to-run variations. However, if these chromatographic dissimilarities are ignored, chemometric models can confuse the variability linked to the instrumental effects with the variability related to the sample specificity. The same peaks in different chromatograms, as well as spectral information related to the same molecular structures but in different samples, should be aligned at the same acquisition point or represented in the same column of a matrix. It has been demonstrated, for example, that imperfections in this requirement influence the number of principal components (PCs) when a PCA is carried out.

To minimize the influence of these errors to be able to extract reliable information (hidden behind the raw chromatographic fingerprint tensor), it is necessary to apply a set of pre-processing operations on the acquired raw chromatographic signal. Two groups of pre-processing could be considered: (i) those that are applied systematically, regardless of the type of signal (*e.g.*, de-noising, smoothing, baseline correction or re-sampling amongst others; and (ii) those involving the use of a reference (*e.g.*, scaling, normalization or alignment). Alignment, the most popular approach, requires finding a suitable transformation of the time axis (x-score in tensor notation) using the so-called alignment or warping function developed through an alignment algorithm. It consists of constructing a function that warps the content of the runs to most closely match a given reference signal. As a consequence, the selection of this reference is a critical task, since not all the signals will have a similar shape or the same complexity. Additional drawbacks can be considered if the liquid chromatographic separation process itself is considered, where not only the stationary phase plays an important role in retention times values but also the mobile phase (i.e., chemical composition particularly when gradient elution is used to solve the separation of complex mixtures).

Table 1. Types of data to be extracted from liquid chromatographic signals.

Tensor order	Elements of the intensity tensor	Instrumental output
	Simplest universal detectors (RID, ELSD, CAD, UV-visible)	
Zeroth	**INT** = A(set $\Delta t_{R\text{-}peak}$)	Single peak area
	INT = h(set $t_{R\text{-}peak}$)	Single peak height
First	**INT** = A_1(set $\Delta t_{R\text{-}peak1}$), A_2(set $\Delta t_{R\text{-}peak2}$), …	Sequence of peak areas
	INT = h_1(set $t_{R1\text{-}peak1}$), h_2(set $t_{R\text{-}peak2}$), …	Sequence of peak heights
First	**INT** = $I_1(t_{R1})$, $I_2(t_{R2})$, …	1D chromatogram
	Diode array detector (DAD)	
Zeroth	**INT** = A(set $\Delta t_{R\text{-}peak}$, set λ_{abs})	Single peak area
	INT = h(set $t_{R\text{-}peak}$, set λ_{abs})	Single peak height
First	**INT** = A_1(set $\Delta t_{R\text{-}peak1}$, set λ_{abs1}), A_2(set $\Delta t_{R\text{-}peak2}$, set λ_{abs2}),	Sequence of peak areas Sequence of peak heights
	INT = h_1(set $t_{R\text{-}peak1}$, set λ_{abs1}), h_2(set $t_{R\text{-}peak2}$, set λ_{abs2}), …	
First	**INT** = $I_1(t_{R1}$, set $\lambda_{abs})$, $I_2(t_{R2}$, set $\lambda_{abs})$, …	1D chromatogram
First	**INT** = I_1(set t_R, λ_{abs1}), I_2(set t_R, λ_{abs2}), …	Absorbance spectrum
Second	**INT** = $I_{1,1}(t_{R1}, \lambda_{abs1})$, $I_{1,2}(t_{R1}, \lambda_{abs2})$, …	Absorption spectrum-chromatogram
	Fluorescence detector (FLD)*	
Zeroth	**INT** = A(set $\Delta t_{R\text{-}peak}$, set λ_{exc}, set λ_{em})	Single peak area
	INT = h(set $t_{R\text{-}peak}$, set λ_{exc}, set λ_{em})	Single peak height
First	**INT** = A_1(set $\Delta t_{R\text{-}peak1}$, set λ_{exc}, set λ_{em}), A_2(set $\Delta t_{R\text{-}peak2}$, set λ_{exc}, set λ_{em}), …	Sequence of peak areas Sequence of peak heights
	INT = h_1(set $t_{R\text{-}peak1}$, set λ_{exc}, set λ_{em}), h_2(set $t_{R\text{-}peak2}$, set λ_{exc}, set λ_{em}), …	
First	**INT** = $I_1(t_{R1}$, set λ_{em}, set $\lambda_{exc})$, $I_2(t_{R2}$, set λ_{em}, set $\lambda_{exc})$, …	1D chromatogram
First	**INT** = I_1(set t_R, λ_{exc1}, set λ_{em}), I_2(set t_R, λ_{exc2}, set λ_{em}), …	Excitation spectrum Emission spectrum
	INT = I_1(set t_R, set λ_{exc}, λ_{em1}), I_2(set t_R, set λ_{exc}, λ_{em2}), …	
Second	**INT** = $I_{1,1}(t_{R1}, \lambda_{exc1}$, set $\lambda_{em})$, $I_{1,2}(t_{R1}, \lambda_{exc2}$, set $\lambda_{em})$, …	Excitation spectrum-chromatogram
	INT = $I_{1,1}(t_{R1}$, set $\lambda_{exc}, \lambda_{em1})$, $I_{1,2}(t_{R1}$, set $\lambda_{exc}, \lambda_{em2})$, …	Emission spectrum-chromatogram
Third	**INT** = $I_{1,1,1}(t_{R1}, \lambda_{exc1}, \lambda_{em1})$, $I_{1,1,2}(t_{R1}, \lambda_{exc1}, \lambda_{em2})$, …	Total fluorescence spectrum-chromatogram*

INT: intensity tensor (scalar, vector, matrix, cube), t_R: retention time, λ: wavelength, A: peak area, h: peak height, I: signal intensity, abs: absorption, exc: excitation, em: emission.

***** *Note that conventional FLD are not capable of on-line recording total fluorescence spectra and a fast spectrofluorometer must be coupled. Ref **11** states that 10 s is required to acquire an excitation-emission total fluorescence spectrum having a resolution of 4.5 nm (emission) and 5 nm (excitation).*

Further interactions among analytes and both mobile and stationary phases play an important role in this process, which vary depending on the work-mode of the LC considered.

Taking into account these considerations, considerable efforts have been made to develop more reproducible chromatographic systems, mainly devoted to minimizing the lack of reproducibility related to the type and stability of columns, reliability of pumps and other mechanical devices (*e.g.,* valves) amongst others, in order to improve the consistency of retention times, which have not always succeeded.

Focusing on the efforts made to control the inherent retention time variability, two principal approaches emerged relatively early: (i) the application of relative retention times which are calculated in respect of particular internal standards or set of internal standards [**12**], and (ii) the correction or adjustment of the differences through a set of test compounds, such as a homologous retention index scale, in a similar way to that proposed by Kovats for GC [**13**].

When chemical subrogates are used as internal standards, they are added to the sample prior to the analysis as a way to standardize the chromatographic signal. Although this has been the most fruitful approach and has been the most studied, some authors have proposed using subrogates as external standards (external standardization) as an alternative, taking advantage of removing one step from the sample preparation sequence or avoiding the possibility that a component in the samples could mask the peak for the standards [**14**]. Whatever the type of standardization used, subrogates should be compounds as closely related to the analytes of interest as possible and show a parallel behaviour at each stage of the analytical procedure *e.g.*, sample preparation (separation, derivatization, etc.) and measuring (detector sensitivity). Moreover, they are unlikely to be present as a contaminant of the sample.

By analogy with GC, at the beginning when spectroscopic detectors were the most commonly used, some studies examined the use of n-alkanes as standards for RPLC applications but their difficult detection using conventional detector as well as their low polarity led authors to try alternative homologous series as standards. In the early stages of these studies, a number of homologous series of compounds were studied as retention index standards specifically for HPLC. Amongst others, n-alkylbenzenes which have the advantage of been detected spectroscopically; alkan-2-ones (or 2-keto alkanes which were used under isocratic condition allowing linear interpolation between the retention factors of the standards; and alky aryl ketones or nitroalkanes which have stronger chromophores so detection at short wavelengths or a diode array detector is recommended. [**14**]. Other authors proposed the use of a retention index scale based on a series of polycyclic aromatic hydrocarbons (PAHs) or nitrated polynuclear aromatic hydrocarbons (Nitro-PAHs) containing

increasing numbers of aromatic compounds. Alternative RI scales have been examined but most of them have scarcely been used by the analytical community, as is the case of aliphatic esters which have been used for the characterization of lipids and related samples of similar polarity (*e.g.*, saturated and unsaturated triglycerides, lecithins or phospholipids) when non-spectroscopic detectors are used [**15**].

In spite of all the efforts made to obtain a standardized scale of retention index, the main observed drawbacks are related to the strong dependency of RI with both eluent composition and different stationary phases. Thus, in order to avoid the influence of different stationary phases on the RI, Bogusz *et al.* [**16**] proposed to scale the results using standard reference compounds. In spite of the proposal producing suitable results in many examples it cannot be effective for compensating the differences between columns which results in changes in the order of elution.

The application of RI scales has been developed in RPLC only to characterize the separation system (stationary and mobile phases) or to identify analytes [**15**]. The principal difference from GC is that the LC separation system is strongly influenced by both the mobile and stationary phases which provoke a wide elution range resulting in a wide range in RI scales with little direct correlation. Under these bases and bearing in mind the crucial fact that the LC conventional retention values are sensitive to small changes in separation conditions of temperature, mobile phase composition and flow rates, differences in the stationary phase with time (ageing) and between columns from the same manufacturers, a harmonization of chromatographic outcomes is desirable.

To date, dozens of correspondence algorithms have been proposed to deal with retention time alignment and signal normalization of LC fingerprints, especially when coupled to MS systems. Amongst others, the COW algorithm matches all sample signals towards a common reference signal [**17**] and the fast Fourier transform cross-correlation algorithm is used to compute the cross correlations between the segments of the spectra and a reference to find out how many points must be shifted to maximize these correlations [**18**]. Several comprehensive reviews on the main methods of chromatogram alignment have recently been published [**19**,**20**]. Nevertheless, most of them are analyte dependent, based on inner reference variables and none of these approaches can be considered as suitable procedures to obtain reproducible absolute responses to solve the universal standardization of data.

Thus, the advent of an independent standardization method such as agnosticism methodology, referred to an external reference, could be crucial bearing in mind that the combination of external and internal standard references is a useful methodology to obtain comparable fingerprints, regardless of the chromatographic state and the sample analyte type. In this

paper we propose the use of this new methodology for harmonising chromatographic signals and thus solve the problems associated with non-targeted fingerprinting methodology when reverse phase liquid chromatography is used.

In a similar way to the study we previously published about signals obtained by GC, a new methodology for LC fingerprints agnostizing is featured in part 2. The paper is structured into three general sections, first of all a discussion about how LC-agnostizing methodology can be performed using Excel. Then, the application of this methodology for the determination of biophenols in olive oil is described and finally some conclusions and final remarks are presented.

2. METHODOLOGY: LIQUID CHROMATOGRAPHY AGNOSTIZING

The hybridization of separation techniques such as liquid chromatography together with suitable measuring techniques is one of the most widespread, versatile and powerful analytical tools to achieve specific signals from a sample. Most analytical methods, in the most favourable cases, show a high degree of selectivity but are not specific. Therefore, especially in the case of complex samples, the step of separating the analyte/s from matrix components in order to avoid possible interferences is essential. However, the advantages from the use of this separative stage can lead to negative effect on the harmonization of results if the dissimilarities in both x-scores and intensities of the signals obtained from different instruments or from the same instrument in different time periods are ignored.

For this purpose, a new methodology is described based on two main stages: a single stage for setting up standard retention scores, which is only applied once; and the stage of chromatogram agnostizing in which specific normalizations, of both intensities and retention scores, are carried out using the previously established scores. A general description of the whole procedure is in a flowchart in Figure 1 and detailed in the following sub-sections.

2.1. Setting up standard retention scores (SRS)

The first step is to establish an invariant reference chemical system for normalizing the retention values. To make it possible, a proper standard mixture of suitable chemical compounds must be designed. An empirical value, known as standard retention scores (SRSs), is assigned to each chemical compound related to their retention time and elution order and included in the standard reference mix. The main difference from the gas chromatography agnostizing methodology, previously reported [**1**], is in this stage. For GC retention indices related to the number of carbon atoms in the molecule of each chemical compound (*e.g.*, Kovats indices) are used. In LC, there are no recent

antecedents of the applicability of these indices aimed at aligning chromatograms. The definition and use of the SRS are proposed for the first time in this paper.

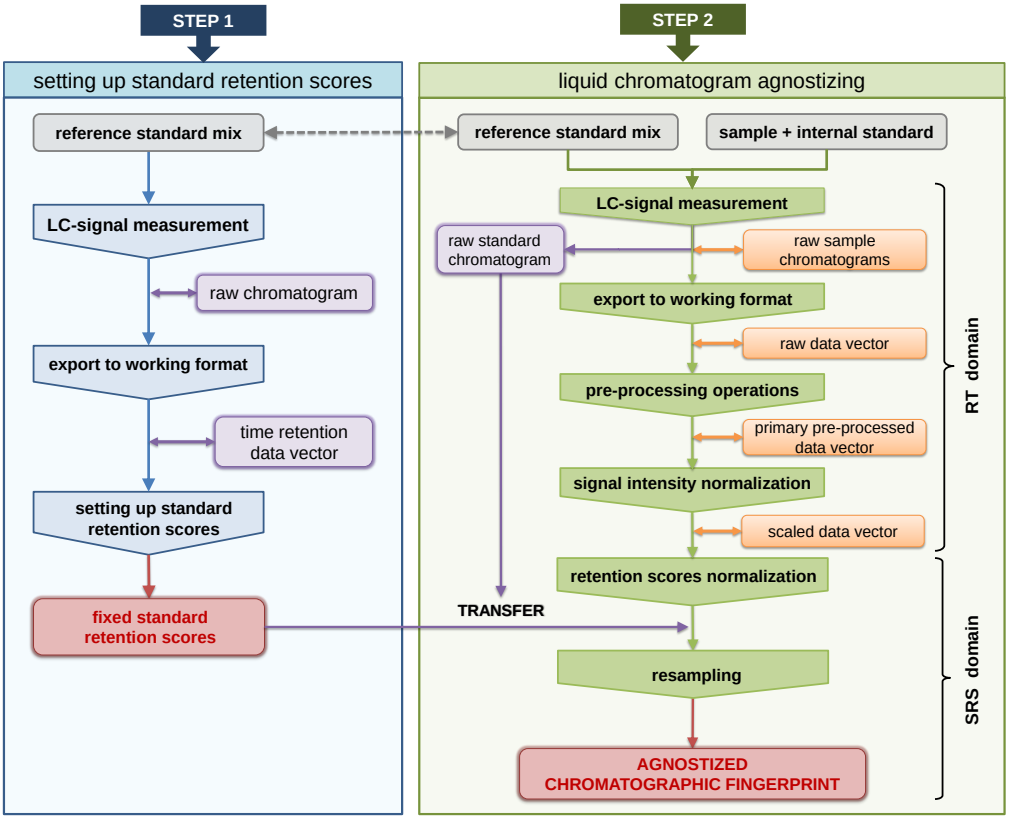


Figure 1. Flowchart showing a general description of steps in liquid chromatography agnostizing. Step 1: Setting up standard retention scores (SRSs); Step 2: Agnostizing liquid chromatogram samples.

Note that this step is only carried out at the beginning of the study and, once the SRSs have been established for each constituent of the standard mix, they are used continuously for retention score normalization of the chromatograms obtained for the test samples. The process should be carried out, in principle, for each particular method. However, the originality of the proposal is supported by the use of a standard reference mix that may be common to several methods in reverse phase conditions.

Two main aspects should be considered when setting up the empirical SRSs:

(i) The reference chromatographic method. In the first instance, the standard retention reference system would have to be established for each particular method, analysing the standard mix with the selected working method. However, it is possible to select fixed reference analysis conditions based on a simple linear gradient (or isocratic) mode and representing all those methods with the same mobile phases, the methods could differ in their

profile of applied gradient as long as the mobile phases and the composition of the solvent mixture at the beginning and at the end of the chromatographic run are the same.

(ii) An adequate reference standard mix. Ideally it should meet these requirements: (i) cover all the chromatographic run time (enough standards are available to cover a wide range of retention scores); (ii) show a regular elution profile under the selected experiment conditions; (ii) each chemical constituent should be well characterized and be pure enough; (iv) commercially available and (if possible) inexpensive; (v) have a chemical behaviour similar to endogenous components of interest in the sample. Likewise, the elution order of standard relative peak-spacing should not change from one stationary phase to another similar. This last requirement could be difficult to meet due to the large number of interactions that occur in LC.

Once the analysis conditions and the standards have been selected, the standard mix is analysed at least 10 times under nearly reproducible instrumental conditions, i.e., analysing different aliquots of the standard mix using the same method and same analysis conditions (same chromatographic state), in order to obtain consistent and representative retention time values. Even so, a robust statistic based on the median is applied to the retention times of each i-th chemical compound along analyses (median RT_i) to eliminate the influence of possible outliers. The calculated medians are representative values of the retention time (RT) of each compound depending on the chromatographic method applied.

The chemical constituents are then ordered according to their order of elution (represented by integers from 1 to n). To explore the behaviour of the standard mixture, the RT values are plotted *vs* the order of elution. As can be seen in Figure 2, a linear trend is already found. Based on this successful outcome, the SRS values are assigned as follows:

- The absolute differences on retention times between adjacent chemical constituents are calculated ($\Delta(\text{medianRT})_{i,i-1}$) and from them, the overall median value-(median($\Delta(\text{medianRT})_{i,i-1}$).
- An SRS of 1 is always assigned to the first eluted compound of the reference standard mix.
- For the rest of the chemical compounds the SRS is defined as:

$$\text{SRS}_i = \frac{\Delta(\text{medianRT})_{i,i-1}}{\text{median}\left(\Delta(\text{medianRT})_{i,i-1}\right)} + \text{SRS}_{i-1}$$

Finally, by plotting the experimental RT values, on the Y-axis versus established SRS values on the X-axis, a perfect linear fitting is found (the same cannot be

said when the elution order is considered). Figure 2 shows, in a schematic way, the operations that have to be carried out to set up the SRS values.

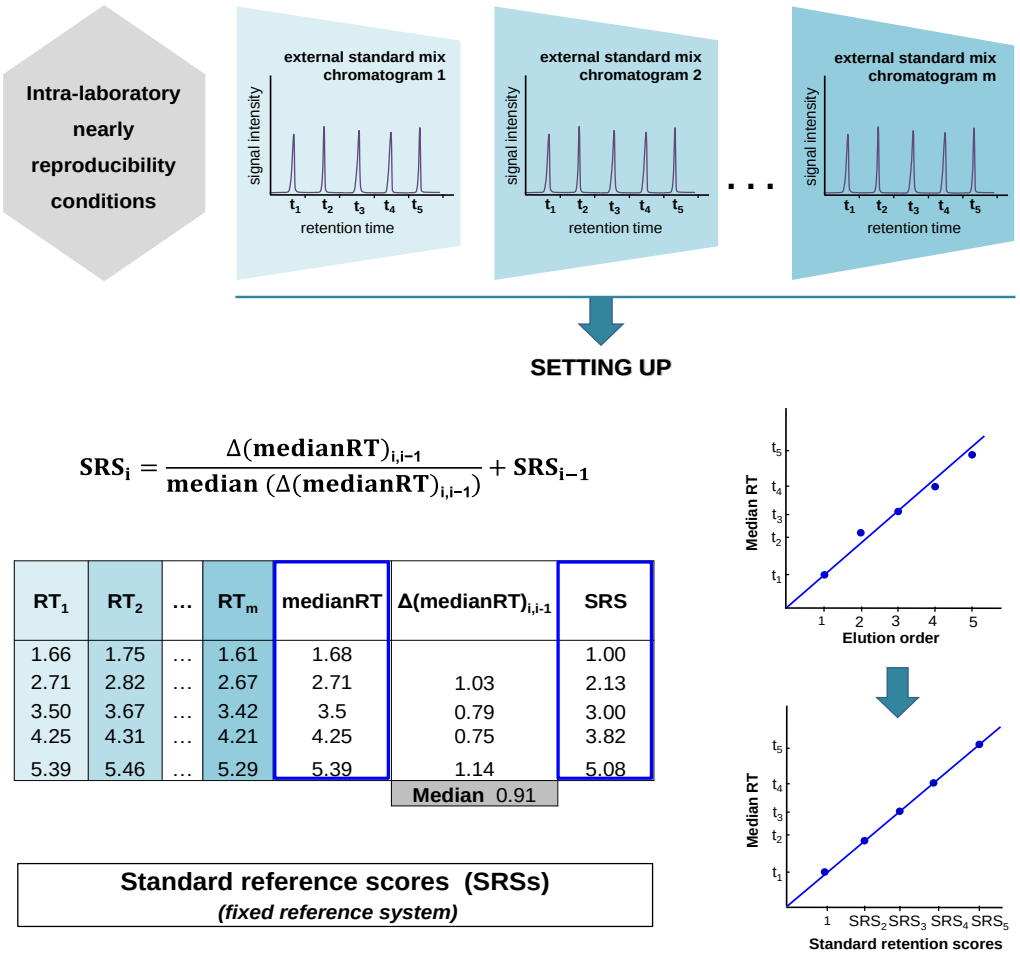


Figure 2. Graphic overview of operations applied for setting up SRS values.

Once the SRSs have been established its values are invariant as long as the mobile phase does not change. Thus, they remain as a fixed reference system for all analytical methods that are the same or similar to the one used as the reference chromatographic method.

2.2. Agnostizing of sample liquid chromatograms

Harmonised data generation, collection and recording are crucial for the optimisation of the sharing and universal use of dataset. Liquid chromatography agnostizing involves the minimization of the influence of the type of column, stationary phase, detector, mobile phase, etc. or even the instrument from which the chromatographic signal is acquired. Finally, for a particular sample and provided that the same chromatographic method is applied, each singularity of the chromatogram (*e.g.*, a peak) has a similar intensity and

corresponds to a similar position in the axis that defines the retention feature. In tensor notation, it is expected that the corresponding values of INT and x-scores agree within a limited tolerance given by the errors associated with the measurements and the data processing process itself. There is, therefore, a double objective that is developed in two stages: normalization of signal intensities and normalization of the retention scores.

It should be noted that, before using the agnostizing methodology, raw chromatographic signals must be exported from the instrument software and arranged in two proper data vectors (intensity vector and retention vector) in a file format compatible with the data treatment software to be used (*e.g.*, CSV or CDF). Then, the data vector should be cleaned-up by applying a suitable pre-processing step: *truncation and cutting data*, *de-noising and smoothing* for removing/minimising noise; *baseline correction* for drifting or rungs along the chromatogram*;* or *re-sampling* in order to fix the number of tensor elements to a given number.

2.2.1. Normalization of signal intensities: use of ratio intensities

Even if previous analytical procedures (*e.g.*, sample collection, storage and handling, sample preparation) have been standardized and properly conducted, instrumental drifts inevitably occur due to changes in the chromatographic state, including, amongst others, condition of the columns and ageing, decay in the detector response, or progressive contamination of the system, and they are the main causes of the unwanted variation of signal intensities. These variations in the raw data are minimised through a normalization step [**21**]. The intensity normalization involves correcting the lack of reproducibility in the intensity value and, optionally, scaling to provide values within the same intensity range for all the chromatographic fingerprints. Internal standard-based normalization has to be applied to obtain intensity-normalized chromatograms.

The internal standard (IS) is added at a well-defined and constant amount to each test solution from samples. Ideally, the concentration of the IS in the test solution may provide the highest signal intensity (peak height) in the sample chromatogram, so that once the intensity-normalized chromatogram is obtained, all the intensity values of the corresponding data vector are within the range 0–1. Each selected IS should fulfil the following features: (i) be chemically similar to the endogenous compounds of the target fraction and it should not be present in the sample(s), and (ii) have the same or similar functional groups, polarity, and chromatographic behaviour as the endogenous compounds.

Whenever possible, the use of at least two internal standards is strongly recommended, selected so that they elute into each of the two halves of the sample chromatogram. Hence, the ratio-intensity from the intensities of the two

IS peaks could be useful as a potential quality control metric for each sample chromatogram run (QC metrics = INT_{IS1} / $INT_{IS2;}$ where INT_{IS1} and INT_{IS2} are the corresponding intensity values or peak heights). For this purpose, this ratio-intensity must remain constant in successive analyses within a pre-determined tolerance range; a significant deviation from the established value suggests a problem throughout the chromatographic run. In this case, the result should be discarded and the analysis repeated.

Once the IS(s) is(are) selected, the intensity normalization process of the sample is conducted calculating a relative intensity for each i-th element of the intensity data vector from the samples ($\mathrm{INT_i^{rel}}$) dividing the original intensity value ($\mathrm{INT_i}$) by the corresponding reference intensity value of the internal standard, denoted by $\mathrm{INT_i^{ref}}$, using the following expression:

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

If only a single IS is used the $\mathrm{INT_i^{rel}}$ is always the same ($\mathrm{INT_i^{rel}} = \mathrm{INT_{IS}}$), but if two ISs are available, it is better to take the pooled intensity from both the internal standards ($\mathrm{INT_i^{pooled}}$) as follows:

$$\mathrm{INT_i^{ref}} = \mathrm{INT_i^{pooled}}(IS1, IS2) = \frac{(\mathrm{INT_{IS_1}} \times |\Delta t|_{IS_{2,i}}) + (\mathrm{INT_{IS2}} \times |\Delta t|_{IS_{1,i}})}{\left|\Delta t_{IS_2,IS_1}\right|}$$

where $\mathrm{INT_{IS_1}}$ and $\mathrm{INT_{IS_2}}$ are the intensity values (peak heights) of both internal standards IS_1 and IS_2, and $|\Delta t|_{IS_{2,i}}$, $|\Delta t|_{IS_{1,i}}$ and $\left|\Delta t_{IS_2,IS_1}\right|$ are the absolute values of the differences between the retention times for IS_1 and the i-th element, IS_2 and the i-th element, and IS_1 and IS_2 respectively.

2.2.2. Normalization of retention scores: transfer of standard retention scores

Normalization of retention scores is the key to building multivariate analytical models from fingerprints that will be more robust against outliers. The final objective is to make retention data which are only dependent on the chromatographic phenomenon, i.e., on the three-term interaction: analyte-stationary phase-mobile phase and therefore independent of the instrument.

Once the SRSs are established (see section 2.1) the second step in retention time normalization is the transference of retention standard scores from the external standard mix to any sample chromatogram and involves the transformation of the chromatographic intensity vectors from the instrumental-dependent RT domain to an instrumental-agnostic SRS domain.

In order to apply this stage, the standard mix should be analysed at least at the beginning and at the end of the chromatographic batch and the experimental retention time for each chemical compound recorded. The standard mix is used as an external standard since it is analysed in an independent test solution. Firstly, the average between the RTs obtained in the initial and final analyses is calculated (provided that their difference does not exceed a tolerance value previously established as a control parameter). Secondly, the average RTs of all the chemical compounds making up the standard mix are plotted against predetermined invariant SRSs. A nearly linear behaviour is usually found, although in some cases a curved pattern may appear.

The transference to each retention time of the recorded sample data vector can be carried out by applying different mathematical strategies: (i) linear spline interpolation between the two values of external standard data vector that cover the range of times which are going to be normalized, (ii) curve interpolation (quadratic or cubic) using at least three or four elements of the external standard data vector, (iii) interpolation from a previously fitted (regressed) polynomial function representative of the whole interval defined by the selected constituents of the external standard mix. Although all of these interpolation procedures allow the assignation of the SRS vector to the samples, using a piecewise function that is transferred by linear spline interpolation between the pairs to each section of the function is the easiest way and provides the best results. The spacing between the consecutive retention times of the standard mix constituents is small and the piecewise linear model is a good approximation of the real chromatographic retention performance. Finally, the resampling algorithm is again applied in order to create a single SRS vector with a fixed number of variables, common to all the samples.

To sum up, we start from a signal on RT domain that is transformed into a signal for the SRS domain and that can be treated by a single multivariate model so that the slight differences that remain in the chromatographic fingerprints could be considered as irrelevant information (or noise) by the model.

This step needs to be carried out each time a new chromatographic batch is carried out.

2.2.3. Use of the Excel spreadsheet

The proposed normalization methodology can be applied using an Excel spreadsheet. Once the raw chromatograms have been exported from the chromatographic software and arranged in data vectors they can be primarily pre-processed (see Figure 1) and directly transferred to the Excel. The intensity normalization is carried out by the direct programme of the expression

described in section 2.1 for each element of the sample data vectors, where the reference intensity value for each of these vectors depends on the internal standard(s) added to the sample.

The spline interpolation/transferring for retention time normalization is carried out using specific Excel functions. FORECAST predicts a value within a data set previously available using the whole data set for mathematical computing. Although this function performs linear interpolations, it is also applicable when the data sets show non-linear trends if a linear behaviour is assumed in each interpolation range. For this purpose, FORECAST is combined with two other Excel functions: MACHT and INDEX. MATCH searches for a given element within a range of cells and returns the relative position of that element. In addition, INDEX returns the value of a specific position in the cell range.

According to the FORECAST programming formula (=FORECAST(t_i,x_{know},y_{know})), three input arguments are required to obtain the interpolation values: (i) the cell containing the t_i element whose value is predicted, (ii) the cell range which contains the known data of the X-axis, and (iii) the cell range which contains the known data of the Y-axis. t_i takes the retention time values of the i-th element one by one from the sample data vector, whilst x_{know} and y_{know} are made up of the range of the cells which contain the fixed standard retention scores (SRSs) and the experimental retention time of each compound present in the external standard mix respectively. When FORECAST is combined with the MATCH and INDEX the expression programmed in Excel should be:

=FORECAST(t_i;INDEX(x_{know};MATCH(t_i;y_{know};1)):INDEX(x_{know};MATCH(t_i; y_{know};1)+1); INDEX(y_{know};MACTH(t_i;y_{know};1)):INDEX(y_{know};MATCH(t_i;y_{know};1)+1))

This expression is used independently of the trend described by the data set. As an example, Figure 3 shows the retention time-retention index plot obtained after applying the FORECAST, MATCH and INDEX functions on two simulated data sets based on hypothetical experimental results obtained specifically for this purpose.

3. APPLICATION EXAMPLES

We have tested the applicability of the proposed agnostizing methodology. In this section examples of reverse phase chromatography are described. Note that these examples have been performed by applying the general methodology in an intra-laboratory scenario so that the results obtained by a particular LC-UV system which has been set as the reference one and obtained results are compared with ones found using others LC equipment. This study lays the groundwork for a future inter-laboratory study.

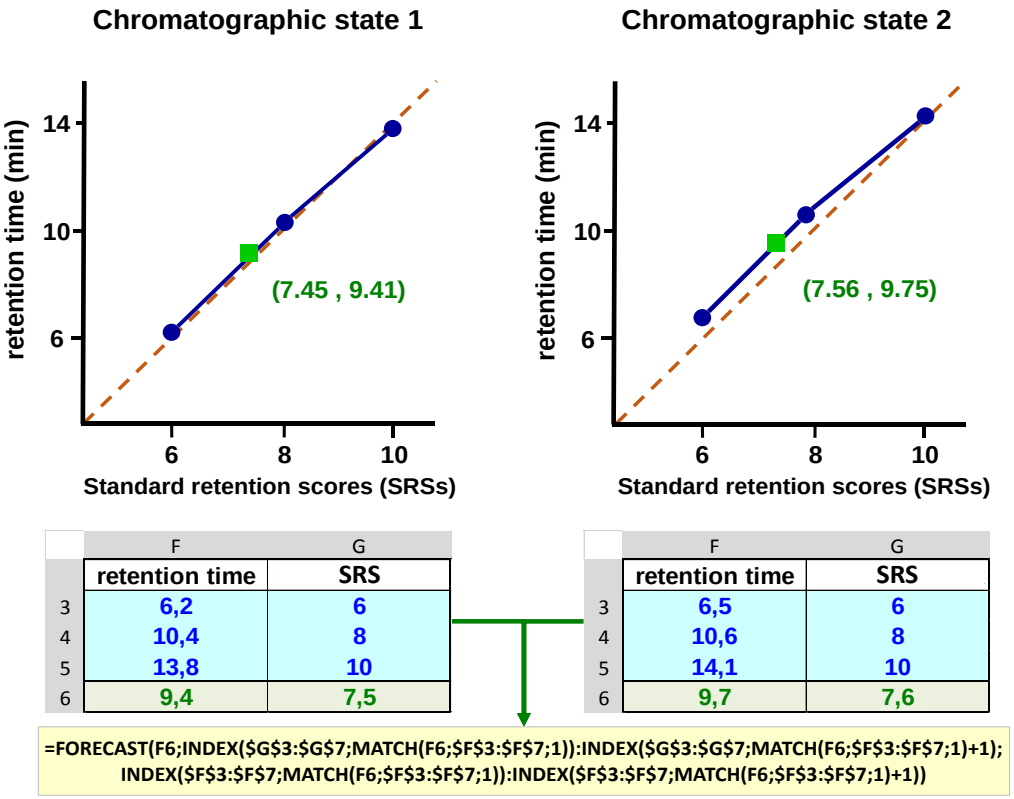


Figure 3. Descriptive example of the applicability of FORECAST, MATCH and INDEX Excel functions to carry out the agnostizing methodology.

3.1. RPLC method for characterizing biophenols in virgin olive oil

Determination of biophenols in olive oils by HPLC following the International Olive Council (IOC) method [**22**], with slight modifications, was chosen as the first case study. Biophenol determination is a common practice in the olive oil industry, the author's own research field, and it was chosen due to its widely used conditions of both mobile and stationary phases and simple gradient condition.

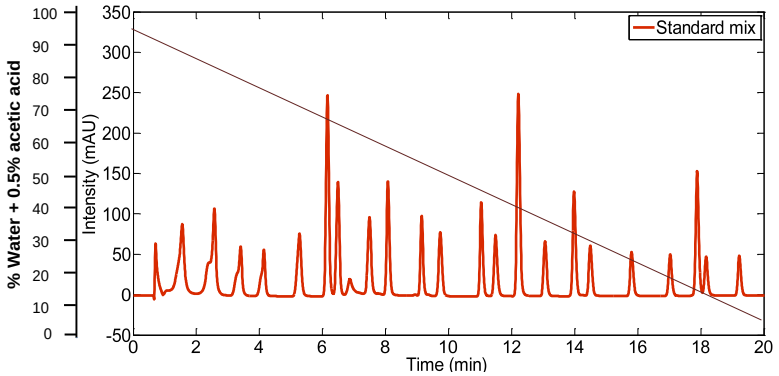
All the solvents used for extraction and analyses were HPLC grade and supplied from Panreac (Barcelona, Spain). First, the phenolic compounds were extracted from the olive oil samples using a liquid-liquid extraction procedure briefly described as follows: 2.0 g of oil was weighed into a centrifuge tube, and 1 mL of n-hexane and 2.0 mL of methanol/water (60/40, v/v) were added. Syringic acid solution (20 µL; 0.75 mg mL^{-1}), prepared from the pure solid supplied by Sigma-Aldrich (purity ≥95%), was added to the oil as the internal standard. The mixture was stirred for 2 min in a vortex apparatus and then the tube was centrifuged at 604 g for 6 min. The methanol/water layer was then separated and filtered through 0.22 µm nylon membrane filters. After that, 20 µL

of the test solution was taken and the chromatographic analysis was carried out.

3.1.1. Materials and methods

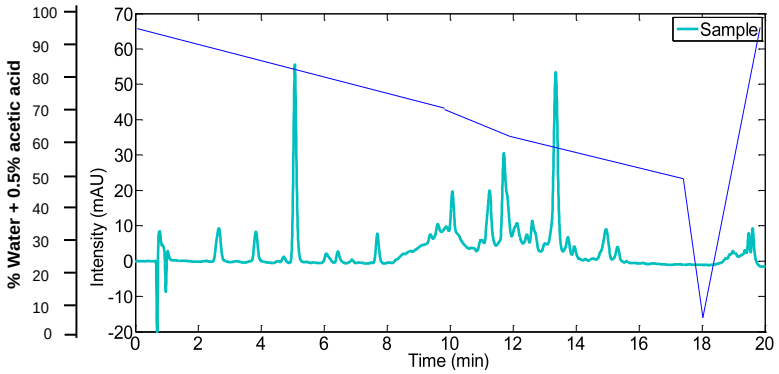
RPLC conditions were re-optimized from the one stated in the IOC method in order to reduce time analysis. Thus, the original chromatographic gradient was modified and the total time analysis was set as 20 minutes plus 3 min to recover initial equilibrium conditions.

However, the gradient of the original method is complex and a simpler gradient is selected as standard reference conditions for setting up SRS values. Figure 4 shows the gradient profiles of both the reference and testing methods superimposed on two chromatograms of the standard mix and an olive oil sample, respectively. In both cases, the flow rate was 1 mL min^{-1} and the column temperature maintained at 25 ºC. Data acquisition was set at 280 nm.



Reference gradient

Time (min)	Water + 0.5 % acetic acid (%)	ACN (%)
0	95	5
20	5	95



Test gradient

Time (min)	Water + 0.5 % acetic acid (%)	ACN (%)
0	95	5
10	70	30
12	62	38
17	50	50
18	5	95
20	95	5

Figure 4. Chromatographic gradient profiles of both reference and testing methods graphically superimposed on the corresponding chromatograms.

In order to test the agnostizing of the instrumental signals obtained by RPLC-DAD, chromatographic analyses were carried out in three different LC systems, using two C18 different columns and pre-columns (the same specifications and

supplier but different ageing and batch number). In addition, comparison of signals obtained from different but similar stationary phases, C8 and C18, were done.

Table 2. LC equipment to determine the different chromatographic states applied in example 1.

Equipment	Specifications
HPLC chromatograph 1 *(reference system)*	Agilent 1260 series liquid chromatograph (Agilent Technologies, Santa Clara, CA, USA) equipped with a CH30 column thermostat (Eppendorf, Hamburg, Germany), a G1311A quaternary pump, a G1322A degasser and G1329A autosampler. Detection was performed with a G7115A Infinity II diode-array detector (DAD). Agilent ChemStation OpenLab CDS software (rev. C.01.09) for LC systems was used to export data to CSV format.
HPLC chromatograph 2	Agilent 1100 series liquid chromatography (Agilent Technologies, Santa Clara, CA, USA) 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) for HPLC systems was used to export data to CVS format.
UPLC chromatograph	Acquity UPLC™ I-Class chromatography system (Waters, Manchester, UK), equipped with an Acquity™ sample manager and an Acquity UPLC™ binary solvent manager was used. Detection was performed with an Acquity™ UPLC PDA detector. OpenChrom Open Source software was used to export data to CSV format.
UHPLC chromatograph	Dionex Ultimate 3000 UHPLC^{+} Focused chromatography system (Thermo Scientific, Waltham, MA, USA) equipped with a RS autosampler and column compartment. Detection was performed with a RS variable wavelength detector. Chromeleon^{TM} version 7.0 software was used to visualize and export data.
Column 1	Zorbax Eclipse Plus C18; 4.6 × 50 mm, 5 µm (Agilent Technologies, Santa Clara, CA, USA).
Column 2	Column 2 is a different column having the same specifications as column 1 but it differs on the lot number, the ageing and the nature of the methods it has been previously applied to.
Column 3	LiChrospher RP8-5; 4.6 mm × 12.5 cm, 5 µm, (Supelco, Darmstadt, Germany).

Four different chromatographic states were defined depending on the chromatographic equipment (instruments and columns) used the features of which are listed in Table 2:

- Chromatographic state #1: HPLC chromatograph1 and column1
- Chromatographic state #2: HPLC chromatograph2 and column2
- Chromatographic state #3: UPLC chromatograph and column2
- Chromatographic state #4: UHPLC chromatograph and column2
- Chromatographic state #5: HPLC chromatograph1 and column3

3.1.2. Establishment of the particular standard retention scores

Chromatographic state #1 was set as the reference system and the fixed standard retention scores (SRSs) were established from it. The choice of the chemicals which would be part of the reference standard mix was an essential step. In an attempt to find a number of chemical compounds which met the characteristics previously described, the homologous series of 1-benzoyl alkanes (the so-called alkyl aryl ketones, from acetophenone to decanophenone) were tested. In order to cover all the chromatographic run time, functional group contributions to the retention of analytes in RPLC proposed by Smith were examined [**23**,**24**]. Taking into account the mobile and stationary phase selected, the empirical elution order of several 1-benzoyl alkanes and related organic compounds were valued and compounds of interest were chromatographed to verify the elution order empirically decided. Finally, to minimize the retention spaces between two contiguous compounds and to achieve elution at a desired position (compounds that elute at earlier or later times under the selected reference analysis method conditions), simple organic compounds, with a similar number of carbons but different functional groups which were not considered in the Smith's paper (for example carboxyl group) or benzenes di- o tri-substituted to modify the polarity, were tested. The list of chemical compounds which were finally selected to constitute the standard mix is set out in Table 3.

Once the compounds had been selected, the standard mix was prepared by dissolving it in 10 mL of methanol. All the test solutions were prepared from the corresponding solid at ≥ 97 % purity. Concentrations varied from 12 ng mL^{-1} (syringic acid) to 320 ng mL^{-1} (benzyl benzoate), attempting to obtain similar values of peak intensities. The solution was stored at –20ºC as a stock standard mix. On each analysis day, an aliquot of 100 µL of the stock standard mix was re-dissolved in 10 mL of methanol, filtered through 0.22 µm nylon membrane filter and 20 µL were injected into the chromatographic equipment. The analyses were repeated 27 times over four selected months. The LC system

was switched on/off several times and the mobile phases were prepared several times during that period of time.

The SRSs were calculated on Excel following the methodology described in section 2.1. Experimental data and calculations are shown in Table 4. Estimated SRSs were fixed and they were used on chromatographic states #1, #2, #3 and #4 to achieve chromatography-agnostic sample fingerprints.

Table 3. List of chemical compounds selected to constitute the standard mix, ordered by their experimental retention time.

Chemical compound	Molar mass (Da)	Molecular formula	Empirical* elution order
3,5-Dihydroxybenzyl alcohol	140.14	$C_7H_8O_3$	(1)
3,4-Dihydroxybenzoic acid	154.12	$C_7H_6O_4$	(2)
Tyrosol	138.16	$C_8H_{10}O_2$	(3)
Syringic acid	198.17	$C_9H_{10}O_5$	–
Phenol	94.11	C_6H_6O	(5)
Benzoic acid	122.12	$C_7H_6O_2$	–
Methyl 4-hidroxybenzoate	152.15	$C_8H_8O_3$	(4)
Benzaldehyde	106.13	C_7H_6O	(6)
Acetophenone	120.15	C_8H_8O	(7)
Ethyl 4-hydroxybenzoate	166.17	$C_9H_{10}O_3$	–
Naphthaleneacetic acid	186.21	$C_{12}H_{12}O_2$	–
Propiophenone	134.18	$C_9H_{10}O$	(8)
Butyl 4-hydroxybenzoate	194.23	$C_{11}H_{14}O_3$	–
Butyrophenone	148.20	$C_{10}H_{12}O$	(9)
Benzophenone	182.22	$C_{13}H_{10}O$	(10)
Valerophenone	162.23	$C_{11}H_{14}O$	(11)
Benzyl benzoate	212.24	$C_{14}H_{12}O_2$	(12)
Hexanophenone	176.25	$C_{12}H_{16}O$	(13)
Heptanophenone	190.28	$C_{13}H_{18}O$	(14)
Octanophenone	204.31	$C_{14}H_{20}O$	(15)
Butylhydroxytoluene	220.35	$C_{15}H_{24}O$	–
Nonanophenone	218.34	$C_{15}H_{22}O$	(16)
Decanophenone	232.36	$C_{16}H_{24}O$	(17)

* Elution order estimated by considering capacity factors and functional group contributions for aromatic substituents measured in acetonitrile-water proposed by Smith [**23**]. The hyphen "–" signifies that it could not be estimated with available published data.

3.1.3. Agnostizing RPLC chromatograms

In order to reach chromatography-agnostic fingerprints for any sample, intensity and time normalizations were carried out in each chromatographic state. The same chromatographic conditions were used on each of the states for analysing the standard mix and samples respectively. Firstly, the standard mix was externally analysed before and after the samples in the same chromatographic batch. Figure 5 shows both the external standard mix and the biophenol olive oil raw chromatograms recorded from each chromatographic state.

The sample raw chromatogram data files were exported in a CSV file and converted to MATLAB format (R2013b version, The Mathworks Inc. MA, USA). The data were pre-processed in the Matlab environment with three main objectives: i) to generate a single time vector common to all chromatograms: different chromatographic equipment may provide different sampling intervals, therefore, an equivalent time vector was generated and signals were interpolated using the *spline* function; ii) to select and truncate the region of interest using the *resample* function, interval [1, 20] min was selected; and iii) to correct potential defects in the data vectors using the *mssgolay* and *msbackadj* algorithms: *mssgolay* function smoothes raw noisy signal data, intensities, using a least-squares digital polynomial filter (Savitzky-Golay filter) whilst *msbackadj* corrects the baseline of the signal with peaks.

The sample intensity vectors were then normalized by dividing the entire sample intensity vectors by the intensities of the internal standard, i.e., syringic acid, recorded on each chromatographic state. Figure 6 shows overlapped scaled sample signals obtained from each chromatographic state. Finally, the averaged retention time dataset, experimentally obtained from the external standard mix on the same chromatographic batch as samples, was plotted against previously established corresponding SRS values. As different retention times were obtained in different chromatographic states, different piecewise linear models were fitted in order to interpolate the sample time vectors.

As a result, the sample chromatograms were transformed from the instrumental-dependent RT domain to an instrumental-agnostic SRS domain following the methodology described in sections 2.2.2 and 2.2.3 and the chromatographic peaks were aligned. Finally, the chromatograms were re-sampled to generate a single common SRS vector to obtain a harmonised database of instrument-agnostic chromatographic fingerprints. The fingerprints recorded from C8 column (chromatographic state #5) clearly differ from those obtained from C18 column. Figure 7 shows overlapped agnostic sample fingerprints obtained in each chromatographic state excluding state #5 fingerprint (for better visualizing).

Table 4. Overview of data recorded on standard mix analyses in order to set up fixed standard retention scores (SRSs).

Date	12 Feb	13 Feb	14 Feb	19 Feb	19 Feb	20 Feb	20 Feb	21 Feb	21 Feb	28 May	28 May	02 Jun	03 Jun	03 Jun	03 Jun	04 Jun	04 Jun	04 Jun	05 Jun	05 Jun	05 Jun	08 Jun	08 Jun	08 Jun	09 Jun	09 Jun	10 Jun			
Peak	**Retention times (RT)**																										**Median RT**	**ΔMedian (RT)$_{i,i-1}$**	**SRS**
Peak01	1.66	1.65	1.64	1.64	1.67	1.63	1.65	1.63	1.65	1.52	1.55	1.57	1.61	1.62	1.71	1.66	1.66	1.66	1.66	1.66	1.66	1.68	1.67	1.67	1.66	1.64	1.66		**1.00**
Peak02	2.70	2.69	2.68	2.70	2.73	2.69	2.71	2.69	2.68	2.53	2.57	2.60	2.65	2.66	2.74	2.70	2.70	2.70	2.70	2.71	2.71	2.71	2.71	2.71	2.71	2.68	2.71	2.70	1.04	**2.27**
Peak03	3.49	3.48	3.47	3.49	3.51	3.48	3.50	3.48	3.48	3.38	3.41	3.43	3.47	3.47	3.53	3.49	3.49	3.50	3.50	3.50	3.50	3.51	3.51	3.50	3.50	3.49	3.50	3.49	0.79	**3.25**
Peak04	4.23	4.22	4.22	4.24	4.25	4.23	4.24	4.22	4.22	4.13	4.17	4.19	4.22	4.22	4.27	4.24	4.24	4.25	4.25	4.25	4.30	4.25	4.26	4.26	4.25	4.24	4.25	4.24	0.75	**4.16**
Peak05	5.44	5.42	5.41	5.44	5.46	5.41	5.43	5.41	5.39	5.26	5.32	5.35	5.40	5.41	5.48	5.44	5.44	5.46	5.45	5.45	5.47	5.46	5.46	5.46	5.45	5.43	5.45	5.44	1.20	**5.63**
Peak06	6.30	6.28	6.27	6.30	6.31	6.27	6.29	6.27	6.25	6.16	6.21	6.24	6.27	6.28	6.34	6.31	6.31	6.33	6.32	6.31	6.33	6.32	6.32	6.32	6.31	6.29	6.31	6.30	0.86	**6.69**
Peak07	6.64	6.62	6.61	6.65	6.66	6.62	6.64	6.61	6.60	6.49	6.55	6.59	6.61	6.62	6.68	6.65	6.65	6.68	6.66	6.66	6.67	6.67	6.66	6.66	6.66	6.64	6.66	6.65	0.34	**7.11**
Peak08	7.02	7.00	6.98	7.01	7.03	6.99	7.01	6.98	6.97	6.85	6.91	6.95	6.99	7.00	7.06	7.03	7.03	7.05	7.04	7.04	7.06	7.05	7.05	7.05	7.07	7.02	7.04	7.02	0.37	**7.57**
Peak09	7.64	7.62	7.61	7.64	7.65	7.62	7.63	7.61	7.60	7.51	7.57	7.60	7.62	7.63	7.68	7.65	7.65	7.68	7.66	7.66	7.67	7.66	7.66	7.66	7.65	7.64	7.65	7.64	0.62	**8.33**
Peak10	8.23	8.21	8.20	8.23	8.24	8.20	8.22	8.20	8.19	8.10	8.16	8.18	8.21	8.21	8.27	8.24	8.24	8.26	8.25	8.25	8.26	8.25	8.25	8.25	8.24	8.22	8.24	8.23	0.59	**9.05**
Peak11	9.31	9.29	9.27	9.31	9.31	9.28	9.30	9.27	9.26	9.18	9.24	9.26	9.28	9.29	9.34	9.32	9.31	9.33	9.33	9.32	9.34	9.33	9.33	9.32	9.32	9.30	9.31	9.31	1.08	**10.37**
Peak12	9.91	9.89	9.88	9.91	9.92	9.88	9.90	9.87	9.87	9.79	9.85	9.87	9.89	9.90	9.95	9.93	9.92	9.94	9.93	9.93	9.94	9.93	9.93	9.93	9.92	9.90	9.92	9.91	0.61	**11.11**
Peak13	11.20	11.17	11.16	11.20	11.20	11.17	11.20	11.17	11.17	11.10	11.15	11.16	11.18	11.18	11.22	11.21	11.21	11.23	11.22	11.21	11.22	11.22	11.22	11.21	11.21	11.19	11.20	11.20	1.29	**12.69**
Peak14	11.67	11.64	11.63	11.67	11.67	11.64	11.66	11.63	11.63	11.56	11.61	11.62	11.65	11.65	11.70	11.68	11.67	11.70	11.69	11.69	11.69	11.69	11.69	11.69	11.68	11.65	11.67	11.67	0.47	**13.26**
Peak15	12.39	12.37	12.36	12.39	12.39	12.37	12.38	12.36	12.36	12.28	12.34	12.34	12.37	12.37	12.42	12.40	12.40	12.42	12.41	12.41	12.41	12.41	12.41	12.40	12.40	12.38	12.39	12.39	0.72	**14.15**
Peak16	13.25	13.22	13.21	13.25	13.24	13.22	13.24	13.21	13.22	13.14	13.19	13.20	13.22	13.22	13.27	13.25	13.25	13.27	13.26	13.26	13.26	13.26	13.26	13.25	13.25	13.23	13.24	13.24	0.85	**15.19**
Peak17	14.17	14.14	14.13	14.17	14.17	14.15	14.16	14.14	14.14	14.07	14.12	14.12	14.13	14.14	14.18	14.17	14.16	14.18	14.18	14.18	14.17	14.17	14.17	14.16	14.16	14.14	14.15	14.16	0.91	**16.31**
Peak18	14.70	14.67	14.66	14.70	14.70	14.68	14.69	14.67	14.67	14.60	14.65	14.65	14.67	14.67	14.72	14.70	14.70	14.71	14.71	14.71	14.71	14.66	14.70	14.69	14.69	14.67	14.68	14.69	0.53	**16.96**
Peak19	16.05	16.03	16.02	16.06	16.06	16.03	16.04	16.02	16.01	15.94	15.99	15.99	16.02	16.02	16.08	16.05	16.05	16.07	16.06	16.06	16.06	16.05	16.05	16.05	16.04	16.02	16.03	16.04	1.35	**18.62**
Peak20	17.33	17.30	17.29	17.33	17.33	17.29	17.31	17.29	17.26	17.19	17.25	17.24	17.29	17.29	17.37	17.32	17.32	17.34	17.33	17.33	17.33	17.32	17.32	17.32	17.32	17.30	17.31	17.32	1.28	**20.18**
Peak21	18.17	18.15	18.14	18.18	18.18	18.15	18.16	18.14	18.12	18.08	18.13	18.11	18.15	18.15	18.21	18.16	18.16	18.18	18.17	18.17	18.17	18.16	18.16	18.16	18.16	18.14	18.15	18.16	0.84	**21.21**
Peak22	18.52	18.50	18.48	18.53	18.54	18.48	18.51	18.48	18.45	18.36	18.43	18.42	18.48	18.48	18.57	18.51	18.50	18.53	18.52	18.52	18.53	18.52	18.52	18.51	18.52	18.49	18.51	18.51	0.35	**21.64**
Peak23	19.63	19.60	19.58	19.63	19.64	19.58	19.61	19.58	19.55	19.45	19.52	19.51	19.58	19.57	19.67	19.60	19.61	19.63	19.61	19.61	19.62	19.61	19.61	19.61	19.62	19.59	19.61	19.61	1.10	**22.98**
																												MEDIAN	0.82	

Peak01: 3,5-Dihydroxybenzyl alcohol; **peak02:** 3,4-Dihidroxibenzoic acid; **peak03:** Tyrosol; **peak04:** Syringic acid; **peak05:** Phenol; **peak06:** Benzoic acid; **peak07:** Methyl 4-hidroxybenzoate; **peak08:** Benzaldehyde; **peak09:** Acetophenone; **peak10:** Ethyl 4-hydroxybenzoate; **peak11:** Naphthaleneacetic acid; **peak12:** Propiophenone; **peak13:** Butyl 4-hydroxybenzoate; **peak14:** Butyrophenone; **peak15:** Benzophenone; **peak16:** Valerophenone; **peak17:** Benzyl benzoate; **peak18:** Hexanophenone; **peak19:** Heptanophenone; **peak20:** Octanophenone; **peak21:** Butylhydroxytoluene; **peak22:** Nonanophenone; **peak23:** Decanophenone.

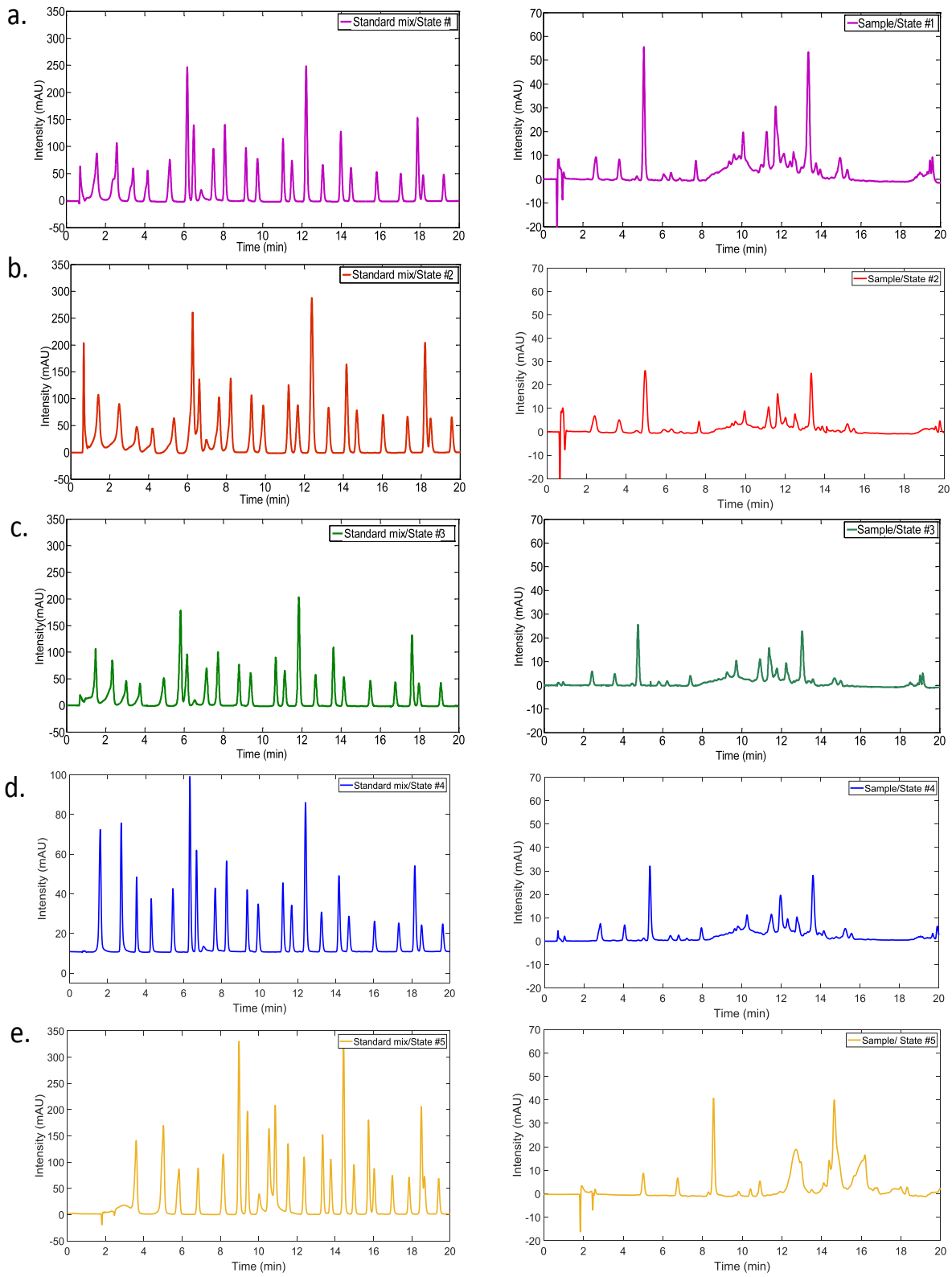


Figure 5. External standard mix and (same) sample raw chromatograms recorded in each chromatographic state: a) chromatographic state #1; b) chromatographic state #2; c) chromatographic state #3; d) chromatographic state #4; e) chromatographic state #5.

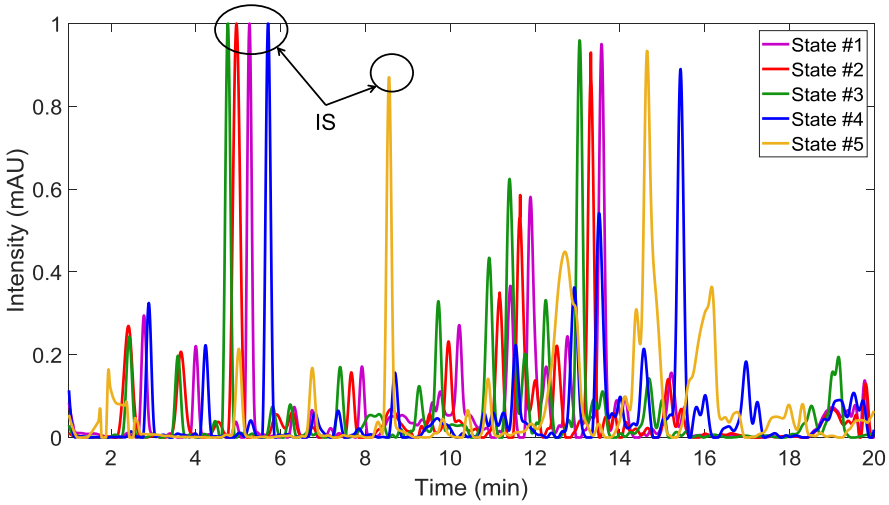


Figure 6. Overlapped chromatograms of the same sample recorded in the five chromatographic states after intensity normalization (time domain).

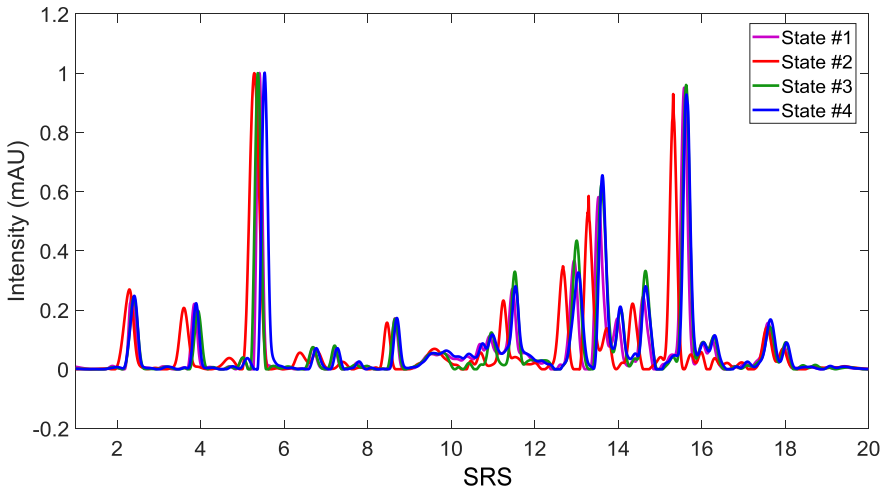


Figure 7. Overlapped chromatograms of the same sample recorded in the four chromatographic states after agnostizing methodology on the SRS domain (excluding state #5 fingerprint for better visualizing).

3.1.4. Results and discussion

In order to validate the applicability of the proposed agnostizing methodology, similarity indices were then calculated from the biophenolic liquid chromatographic fingerprints from each chromatographic state before and after agnostizing. The chromatograms transformed into data vectors were firstly pre-processed and compared by pair with chromatographic the state #1 reference fingerprint for the time domain. Then, the fingerprints were instrument-agnostic and again they need to be re-sampled to obtain the same SRS vector. After that, they were again compared by pair with reference for the SRS domain.

As a representative similarity index, the cosine index (cosθ), a measurement of the cosine of the angle between the considered and reference fingerprint vectors was used [**25**]. The cosθ will be 1 when the angle is 0°, i.e., the two fingerprints have the same orientation:

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

where, x_{ci} and x_{ri} symbolise each element of the considered and reference fingerprints, respectively. A high level of similarity can be assumed when no differences between the fingerprints were found. In addition, the closer to 1 the values of the index are, the greater the similarity between the considered and the reference fingerprints is. Furthermore, reference and each considered fingerprint vector were subtracted, element by element (x_{ri} - x_{ci}), and the mean values of the differences were calculated. The results are shown in Table 5.

As Figure 4 and the calculations in Table 5 it is show, the fingerprints prior to agnostizing clearly differ. Considering chromatographic state #1 as reference, the mean difference between non-instrument-agnostic fingerprints are in the order of 100 times greater than after agnostizing. Calculated cosθ indices also improve after agnostizing in all cases. As shown in Figure 7 almost no differences were found in fingerprints (SRS domain).

Cosθ index from state #2 indicates a worse agnostization. As can be seen in Figure 7, the first and last peaks are well aligned, whilst in the intermediate zone of the chromatographic fingerprint, the peaks appear slightly earlier and therefore the calculated similarity index has a lower value. Agnostizing may correct early or delayed elution behaviour that occurs constantly or following a trend, but not random variations due to non-constant instrumental defects throughout the analysis. The state #2 chromatograph is widely used equipment that does not pass most of the common qualification tests. It is used extensively for teaching practice with students. It has been found to have problems in performing the gradient.

The fingerprint recorded in state #5, after liquid chromatographic separation in the C8 column, clearly differs from the other ones (separation using C18 columns) so that it makes no sense to compare the instrument-agnostic chromatograms. However, if we only take the IS peak into account, its retention time differs more than 3 minutes from C8 to C18 raw fingerprints whilst after agnostization, the SRS of the IS peak in chromatography state #5 only differs 0.7 units from the SRS of the IS peak in the reference chromatography state. This fact verifies the power of this methodology for the alignment of similar fingerprints recorded under widely different states.

Table 5. Similarity indices calculated to validate agnostizing methodology applicability: state#1 is considered as reference.

	Index	Pre-processed chromatograms	Instrument-agnostic chromatograms
State#2	Cosθ	0.15	0.36
	Mean difference ($\times 10^2$)	-33	0.25
State#3	Cosθ	0.24	0.90
	Mean difference ($\times 10^2$)	-52	0.42
State#4	Cosθ	0.13	0.70
	Mean difference ($\times 10^2$)	81	0.56

4. CONCLUSIONS AND FUTURE TRENDS

In view of the results attained within this study, we can conclude that a novel methodology has been established for standardizing 1D chromatograms, making the new instrument-agnostic signal or fingerprint independent from the chromatographic state or date of analysis and that the fingerprints from two different instruments states have a high degree of similarity. This would allow the establishment of characteristic chromatographic databases to harmonize, collect, analyse and share comparable analysis information for each food matrix. Instrument agnostizing provides a response to Horizon 2020 EFSA's Priority Research Topics related to food interests which aims to design, develop and test an integrated system for harmonising data generation [**26**].

Furthermore, working with instrument-agnostic chromatographic fingerprints provides the possibility of creating a single multivariate model, either for use in qualitative (classification) or quantitative (quantification) applications, which would allow the results obtained by different chromatographic states in an inter-laboratory scenario to be combined. In this context, agnostizing could be considered as a further step in the pre-processing operations of data tensors for the establishment of multivariate models. The methodology is simple, does not require great efforts in sample preparation, chromatographic optimization or the use of specific platforms for data processing and offers really effective harmonised results. Our future research is to conduct a comprehensive inter-laboratory study to establish representative (by agreement) standard retention scores. The authors are currently working on other LC modes not covered in this article such as NPLC, HILIC, IEC or SEC.

Acknowledgments

This work was financially supported by the European Regional Development Fund (ERDF) and Ministry of Science and Technology within the framework of the funding corresponding to 'Retos-Colaboracion' (Challenges-Collaboration Program) call of the State Program for Research, Development and Innovation Oriented to Challenges of Society (RTC-2017-6170-2). In addition, SMR author acknowledges the PhD grant supported by the Ministry of Universities for the Training of University Teachers (FPU19/02078).

REFERENCES

[1] L. Cuadros Rodríguez, F. Ortega Gavilán, S. Martin Torres, S. Medina Rodríguez, A.M. Jiménez Carvelo, A. Gozález Casado, M.G. Bagur González, Standardization of instrumental fingerprints – Part I: Towards an instrumental agnosticism in gas chromatography, J. Chromatogr. A1641 (2021) 461983. https://doi.org/10.1016/j.chroma.2021.461983

[2] M.G. Bagur González, A.M. Jiménez Carvelo, F. Ortega Gavilán, A. González Casado, Chromatographic methods, in: C.M. Galanakis (Ed), Food Traceability and Authentication, Academic Press / Elsevier, Oxford, 2020, ch. 5, *(in press)*. https://doi.org/10.1016/B978-0-12-821104-5.00010-6

[3] M. Esteki, Z. Shahsavari, J. Simal Gandara, Food identification by high performance liquid chromatography fingerprinting and mathematical processing, Food Res. Int. 122 (2019) 303–317. https://doi.org/10.1016/j.foodres.2019.04.025

[4] M. Esteki, J. Simal Gandara, Z. Shahsavari, S. Zandbaaf, E. Dashtaki, Y. Vander Heyden, A review on the application of chromatographic methods, coupled to chemometrics, for food authentication, Food Control 93 (2018) 165–182. https://doi.org/10.1016/j.foodcont.2018.06.015

[5] C. Pérez Ràfols, X. Subirats, N. Serrano, J. M. Díaz Cruz, New discrimination tools for harvest year and varieties of white wines based on hydrophilic interaction liquid chromatography with amperometric detection, Talanta 201 (2019) 104–110. https://doi.org/10.1016/j.talanta.2019.03.099

[6] Q. Wang, L. Tong, L. Yao, P. Zhang, L. Xu, Fingerprinting of traditional Chinese medicines on the C18-Diolmixed-mode column in online or offline two-dimensional liquid chromatography on the single column modes, J. Pharm. Biomed. Anal. 125 (2016) 205–211. http://dx.doi.org/10.1016/j.jpba.2016.03.033

[7] Y. Chen, W. Bicker, J. Wu, M. Yong Xie, W. Lindner, Ganoderma species discrimination by dual-mode chromatographic fingerprinting: A study on stationary phase effects in hydrophilic interaction chromatography and reduction of sample misclassification rate by additional use of reversed-phase chromatography, J. Chromatogr. A, 1217 (2010) 1255–1265. http://dx.doi.org/10.1016/j.chroma.2009.12.024

[8] A.M. Striegel, W.W. Yau, J.J. Kirkland, D.D. Bly, Modern Size-exclusion Liquid Chromatography, second ed., John Wiley & Sons, Inc., Hoboken, 2009, ch. 13. ISBN 978-0-471-20172-4.

[9] S. Simon, B. Païro, M. Villain, P. D'Abzac, E. Van Hullebusch, P. Lens, G. Guibaud, Evaluation of size exclusion chromatography (SEC) for the characterization of extracellular polymeric substances (EPS) in anaerobic granular sludges, Bioresource Technol. 100 (2009) 6258–6268.

https://doi:10.1016/j.biortech.2009.07.013
[10] W. Liu, J. Xu, R. Zhu, Y. Zhu, Y. Zhao, P. Chen, C. Pan, W. Yao, X. Gao, Fingerprinting profile of polysaccharides from *Lycium barbarum* using multiplex approaches and chemometrics, Int. J. Biol. Macromol. 78 (2015) 230–23. https://dx.doi.org/10.1016/j.ijbiomac.2015.03.062
[11] M.D. Carabajal, J.A. Arancibia, G.M. Escandar, Multivariate curve resolution strategy for non-quadrilinear type 4 third-order/four-way liquid chromatography-excitation-emission fluorescence matrix data, Talanta 189 (2018) 509–516. https://doi.org/10.1016/j.talanta.2018.07.017
[12] D.W. Hill, A.J. Kind, Reversed-phase solvent-gradient HPLC retention indexes of drugs, J. Anal. Toxicol. 18 (1994) 233–242. https://doi.org/10.1093/jat/18.5.233.
[13] E. Kovats, Gas-chromatographische charakterisierung organischer verbindungen. Teil 1: retentions indices aliphatischer halogenide, alkohole, aldehyde und ketone, Helv. Chim. Acta 41 (1958) 1915–1932. https://doi.org/10.1002/hlca.19580410703-2.
[14] R.M. Smith, Retention index scales used in high-performance liquid chromatography, in: R.M. Smith (Ed.), Retention and Selectivity in Liquid Chromatography, Elsevier Science B.V., Amsterdam, 1995, pp. 107–122. ISBN: 0-444-81539-2
[15] F. Rigano, M. Oteri, M. Russo, P. Dugo, L. Mondello, Proposal of a linear retention index system for improving identification reliability of triacylglycerol profiles in lipid samples by liquid chromatography methods, Anal. Chem. 90 (2018), 3313–3320. https://doi.org/10.1021/acs.analchem.7b04837.
[16] M. Bogusz, M. Erkens, Reversed-phase high-performance liquid chromatographic database of retention indices and UV spectra of toxicologically relevant substances and its interlaboratory use, J. Chromatogr. A, 674 (1994) 97–126. https://doi.org/10.1016/0021-9673(94)85220-0.
[17] N.V. Nielsen, J.M. Carstensen, J. Smedsgaard, Aligning of simple and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimized warping, J. Chromatogr. A, 805 (1998), 17–35. https://doi.org/10.1016/S0021-9673(98)00021-1
[18] J.W.H. Wong, C. Durante, H.M. Cartwright, Application of fast Fourier transform cross-correlation for the alignment of large chromatographic and spectral datasets, Anal. Chem., 77 (2005), 5655–5661. https://doi.org/10.1021/ac050619p
[19] R.H Jellema, A. Folch-Fortuny, M. MWB Hendriks, Variable shift and alignment, in: S.D. Brown, R. Tauler, B. Walczak (Eds), Comprehensive Chemometrics: Chemical and Biochemical Data Analysis, second edition, Elsevier B.V., Amsterdam, 2020, vol. 3, ch. 3.05, pp. 115–136. https://doi.org/10.1016/B978-0-12-409547-2.14886-3
[20] R. Smith, D. Ventura, J.T. Prince, LC-MS alignment in theory and practice: a comprehensive algorithmic rewiew, Brief. Bioinform., 16 (2013) 104-107. https://doi.org/10.1093/bib/bbt080
[21] P. Filzmoser, B. Walczak, What can go wrong at the data normalization step for identification of biomarkers? J. Chromatogr. A 1362 (2014) 194–205. http://dx.doi.org/10.1016/j.chroma.2014.08.050
[22] COI/T.20/Doc No. 29/Rev.1 2017, Determination of biophenols in olive oil by HPLC, International Olive Council, Madrid, 2017.
[23] R.M. Smith, Functional group contributions to the retention of analytes in reverse-phase high-performance liquid chromatography, J. Chromatogr. A 656 (1993) 381–415. https://doi.org/10.1016/0021-9673(93)80811-L

[24] L.R. Snyder, J.J. Kirkland, J.W. Dolan, Introduction to Modern Liquid Chromatography, third edition, Wiley, New Jersey, 2010. pp. 80–85. ISBN: 978-0-470-16754-0

[25] F. Ortega Gavilán, L. Valverde Som, F.P. Rodríguez García, L. Cuadros Rodríguez, M.G. Bagur González, Homogeneity assessment of reference materials for sensory analysis of liquid foodstuffs. The virgin olive oil as case study, Food Chem. 322 (2020) 126743. https://doi.org/10.1016/j.foodchem.2020.126743

[26] G. Kass, J. Moon, T. Robinson, Horizon 2020: EFSA's Priority Research Topics, Technical Report EFSA, European Food Safety Authority, 2017. http://dx.doi.org/10.2903/sp.efsa.2017.EN-1166

PUBLICACIÓN V.

CAPÍTULO DE LIBRO

Instrument-agnostizing methodology for liquid chromatography-mass spectrometry systems

(Chapter 22). Aceptado para publicación en el libro: *Mass Spectrometry for Metabolomics-Methods in Molecular Biology, González Domínguez, R. (ed.), Springer US, 1st Edition 2022. ISBN: 9781071626986*

Instrument-agnostizing methodology for liquid chromatography-mass spectrometry systems

Rosalía López-Ruíz [(1,✉)], Sandra Martín-Torres [(2,✉)], Ana M. Jiménez-Carvelo [(2)], Roberto Romero-González [(1)], Luis Cuadros-Rodríguez [(2)]

[(1)] *Department of Chemistry and Physics, Research Centre for Mediterranean Intensive Agrosystems and Agri-Food Biotechnology (CIAIMBITAL), University of Almeria, Agri-food International Campus of Excellence, CeiA3, E-04120, Almeria, Spain.*

[(2)] *Department of Analytical Chemistry, University of Granada, c/ Fuentenueva, s.n. E-18071 Granada, Spain.*

Running head: Instrument-agnostizing methodology

Abstract

Mass spectrometry is a powerful analytical technique used to identify unknown compounds, to quantify known compounds, and to elucidate the structure and chemical properties of molecules.

Nevertheless, the transfer of data from one instrument to another is one of the main problems, and obtaining the same or similar information from an analogous instrument but from a different manufacturer or even with the same instrument after carrying out the analyses in different time spacing is not possible. Hence a general methodology to provide a chromatographic signal (or chromatogram) independent of the instrument is needed. In this sense, this book chapter describes the standardisation procedure of chromatographic signals obtained from mass spectrometry platforms to obtain instrument-agnostic chromatographic signals for the determination of standard retention scores. This parameter may be used for the quantification of compounds when different mass spectrometry platforms coupled to ultra-hight performance liquid chromatography are employed.

Keywords

Instrument-agnostizing methodology; standardised chromatogram; instrument-independent analytical signals; mass spectrometry.

✉ Corresponding author, e-mail: rlr468@ual.es; phone: +34 950015168.

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

1. INTRODUCTION

Nowadays, the use of pesticides to reduce crop losses due to pests and keep food quality is still increasing. Their use has increased worldwide because of their rapid and efficient action despite the fact that organic farming is also expanding. Nevertheless, the presence of pesticide residues in fruits and vegetables, among other foods, has become a major concern for consumer's health, especially in developing countries [**1**].

To solve this problem, liquid chromatography coupled to mass spectrometry is the technique widely used for the determination of pesticide residues in food [**2**]. In this field, laboratory routine analysis must have the capacity to analyse a wide range of pesticides (more than 500) to ensure food safety and quality. This becomes a great concern due to the necessity to obtain analytical standards of each pesticide to detect them by chromatographic techniques coupled to mass spectrometry. In addition, parameters as retention time or peak area can be difficult to reproduce in different equipment or laboratories [**3**], and therefore, it is difficult to compare these relevant parameters among different analytical platforms used by the same or different laboratories. Considerable efforts have been made to develop more reproducible chromatographic systems, which have not always succeeded.

To date, dozens of correspondence algorithms have been proposed to deal with retention time alignment and signal normalization of LC fingerprints, especially when LC is coupled to MS systems [**4**]. Nevertheless, most of them are analyte dependant, based on inner reference variables and none of these approaches can be considered as suitable procedures to obtain reproducible absolute responses to solve the universal standardization of data.

In this field, instrument-agnostizing methodology takes part as the ability to provide standardized and comparable chromatographic fingerprints regardless of the instrument, the particular instrumental conditions or the time of analysis, in order to ensure better reproducibility. This new methodology is described based on two main stages: a single stage for setting up standard retention scores, which is only applied once and referred to an external reference; and the stage of chromatogram agnostizing in which specific normalizations, of both intensities and retention scores, are carried out using the previously established scores. A general description of the whole procedure is given in recent published articles [**5**,**6**].

2. MATERIALS

2.1. External standard solutions (ES)

1. Purchase analytical standards of solid contaminants, parabens and drugs. Prepared several stock solutions:
   1.1. Standard solutions of each analyte (Table 1) (10 mL) at 1000 mg/L in MeOH.
   1.2. Standard mix solution (Table 1) (10 mL) at 10 mg/L in MeOH.
   1.3. Standard mix solution (Table 1) (10 mL) at 1 mg/L in MeOH (working solution).
2. Store solutions 1, 2 and 3 in the freezer at -21ºC

2.2. Internal standard solutions (IS)

1. Purchase solid analytical standards and prepared several stock solutions:
   1.1. Standard of each caffeine and triphenylphosphate (TPP) (10 mL) at 1000 mg/L in MeOH.
   1.2. Standard mix solution of caffeine and TPP (10 mL) at 10 mg/L in MeOH.
2. Store solutions 1 and 2 in the freezer at -21ºC.

2.3. Mass calibration

Accurate the mass calibration of the LC-Q-Orbitrap and LC-Orbitrap mass analysers using a mixture of acetic acid, caffeine, Met-Arg-Phe-Ala-acetate salt and Ultramark 1621, and a mixture of acetic acid, sodium dodecyl sulfate, taurocholic acid sodium salt hydrate and Ultramark 1621 (fluorinated phosphazines). Calibration mix solutions were provided by the commercial brand.

2.4. Separation systems

1. Vanquish flex Quaternary LC pump (Thermo Scientific Transcend™, Thermo Fisher Scientific, San Jose, CA, USA).
2. Thermo Fisher Scientific Transcend 600 LC (Thermo Scientific Transcend™, Thermo Fisher Scientific, San Jose, CA, USA).
3. Column: Zorbax Eclipse Plus C_{18} column (100 mm × 2.1 mm, 1.8 µm particle size).

2.5. *Mass spectrometry systems*

2.5.1. MS system #1

1. Q-Exactive Orbitrap analyser (Thermo Fisher Scientific, Bremen, Germany) using an electrospray interface (ESI) (HESI-II, Thermo Fisher Scientific, San Jose, CA, USA) in positive and negative mode.

2.5.2. MS system #2

1. Exactive Orbitrap analyser (Thermo Fisher Scientific, Bremen, Germany) using an electrospray interface (ESI) (HESI-II, Thermo Fisher Scientific, San Jose, CA, USA) in positive and negative mode.

3. METHODS

3.1. Sample treatment

1. Select cucumber as a representative matrix of vegetables with high water content.
2. Use extraction method well-known QuEChERS procedure [**7**], commonly used for pesticide extraction.
3. Crush and homogeny 1 kg of cucumber.
4. Place 10 g of sample in a 50 mL Falcon® tube.
5. Add 10 mL of ACN. Shake for 1 min in vortex.
6. Add 1 g of NaCl and 4 g of $MgSO_4$. Shake 1 min in a vortex.
7. Centrifuge the sample for 10 min at 5000 rpm (RCF = 4480 *g*).
8. Collect the supernatant to prepare the samples for the instrument-agnostizing procedure.
9. Select two internal standards (IS) and add them at a well-defined and constant amount to each test solution from samples. Ideally the concentration of the IS in the test solution may provide the highest signal intensity (peak height) in the sample chromatogram and they should elute into each of the two halves of the sample chromatogram.

3.2. LC conditions

1. Mobile phase A: Water solution of 0.1% formic acid. Mobile phase B: MeOH.
2. Elution mode gradient: 0-1 min, 50% of A, decreased to 0% of A in 10 min, kept 2 min and returned to 50% of A in 0.5 min, finally from 13.5 to 20 min kept constant to equilibrate the analytical column.
3. Column temperature set on 30 °C.
4. Injection volume 10 µL.
5. Flow rate of 0.2 mL/min.
6. The total running time 20 min.

3.3. MS parameters

3.3.1. MS system #1

1. ESI parameters: spray voltage, 4 kV; sheath gas (N2, >95%), 35 (arbitrary units); auxiliary gas (N2, >95%), 10 (arbitrary units); heater temperature, 300 °C; capillary temperature, 300 °C and S-lens RF level, 50 (arbitrary units).

3.3.2. MS system #2

1. ESI parameters: spray voltage, 4 kV; sheath gas (N2, >95%), 35 (arbitrary units); auxiliary gas (N2, >95%), 10 (arbitrary units); skimmer voltage, 18 V; capillary voltage, 35 V; tube lens voltage, 95 V; heater temperature, 305 °C; capillary temperature, 300 °C.

3.4. Step 1: Setting up standard retention scores (SRS)

1. Select a reference chromatographic method based on simple linear gradient (or isocratic) mode and represent all those methods with the same mobile phases and the same solvent composition mixture at the beginning and at the end of the chromatographic run (see **Note 1**).
2. Establish an invariant reference chemical system: select a proper standard mixture of chemical compounds. Ideally, the compounds should cover all the chromatographic run time; show a regular elution profile under the selected experiment conditions; be well characterized and be pure enough; commercially available and (if possible) inexpensive; have a chemical behaviour similar to endogenous components of interest in the sample.
3. Analyse the standard mix at least 10 times under nearly reproducible instrumental conditions to obtain representative retention time values.
4. Apply a robust statistic based on the median to the retention times of each i-th chemical compound along analyses (median RT_i) to remove possible outliers (see **Note 2**).
5. Sort out chemical constituents according to their elution order, represented by integers from 1 to n. A nearly linear trend must be found between RT values and elution order.
6. Calculate standard retention scores (SRSs), understood as empirical values related to retention time and elution order:

$$\mathrm{SRS_i} = \frac{\Delta(\mathrm{medianRT})_{\mathrm{i,i-1}}}{\mathrm{median}\left(\Delta(\mathrm{medianRT})_{\mathrm{i,i-1}}\right)} + \mathrm{SRS_{i-1}} \ ;$$

where $\Delta(\mathrm{medianRT})_{i,i-1}$ denotes absolute differences on retention times between adjacent chemical constituents and $(\mathrm{median}(\Delta(\mathrm{medianRT})_{i,i-1})$ overall median value from them.

7. Assign a SRS to each standard: a SRS of 1 is always assigned to the first eluted compound of the standard mix (see **Note 3**).

   NOTE. SRSs are invariant as long as the mobile phase does not change. Thus, they remain as a fixed reference system for all analytical methods that are the same or similar to the one used as the reference chromatographic method. That is, step 1 is carried out only once, before implementation of the analytical method.

3.5. Step 2: liquid chromatography-mass spectrometry agnostizing

3.5.1. Chromatographic analyse of samples

1. Design the chromatographic batch: each day it must include the analysis of the standard mix at the beginning and at the end of the batch.
2. Analyse the external standard mix solution applying the reference method.
3. Analyse the sample solutions, which include the internal standards, applying the specific method (see **Note 4-5**).
4. Export chromatographic signals to working format and obtain an intensity data vector for each chromatogram.

3.5.2. Normalization of signal intensities

1. Clean-up data vector by applying a suitable pre-processing step: truncation and cutting data, de-noising and smoothing for removing/minimising noise; baseline correction for drifting or rungs along the chromatogram; or re-sampling to fix the number of elements in the data vector to a given number.
2. Calculate a relative intensity for each i-th element of the intensity data vector from the samples (INT_i^{rel})

$$\mathrm{INT_i^{rel}} = \frac{\mathrm{INT_i}}{\mathrm{INT_i^{ref}}} \ ;$$

   where INT_i denotes original intensity value; $INT_{i\ ref}$ is reference intensity value of the internal standard.

3. Pooled intensity from both internal standards is calculated as:

$$\mathrm{INT_i^{ref}} = \mathrm{INT_i^{pooled}}(IS1, IS2) = \frac{(\mathrm{INT_{IS_1}} \times |\Delta t|_{\mathrm{IS_{2,i}}}) + (\mathrm{INT_{IS2}} \times |\Delta t|_{\mathrm{IS_{1,i}}})}{|\Delta t_{\mathrm{IS_2,IS_1}}|} \ ;$$

   where INT_{IS_1} and INT_{IS_2} are the intensity values (peak heights) of both internal standards IS_1 and IS_2, and $|\Delta t|_{IS_{2,i}}$, $|\Delta t|_{IS_{1,i}}$ and $|\Delta t_{IS_2,IS_1}|$ are the absolute values of the differences between the retention

times for IS_1 and the i-th element, IS_2 and the i-th element, and IS_1 and IS_2 respectively.

3.5.3. Normalization of retention scores: transfer to SRS domain

1. Record and list the experimental retention times for each chemical standard of the two analyses (beginning and ending times) each day.
2. Calculate the average between the RTs obtained in the initial and final analyses a day.
3. Plot the average RTs of all the chemical compounds making up the standard mix against predetermined invariant SRSs.
4. Establish a piecewise function that is transferred by linear spline interpolation between the pairs to each section of the function.

   $$=FORECAST(t_i;INDEX(x_{know};MATCH(t_i;y_{know};1)):INDEX(x_{know};MATCH(t_{ii};y_{know};1)+1);INDEX(y_{know};MACTH(t_i;y_{know};1));INDEX(y_{know};MATCH(t_{ii};y_{know};1)+1))$$

   Indeed, the transferring for retention time normalization is carried out using three specific Excel functions:

   - *FORECAST:* it predicts a value within a data set previously available using the whole data set for mathematical computing. Although this function performs linear interpolations, it is also applicable when the data sets show non-linear trends if a linear behaviour is assumed in each interpolation range.
   - *MACHT:* it searches for a given element within a range of cells and returns the relative position of that element.
   - *INDEX:* it returns the value of a specific position in the cell range.

   Further information can be found on [**6**].
5. Assign and record a SRS vector to each sample.
6. Apply a resampling algorithm to fix the number of elements which constitutes each instrument-agnostic data vector (see Notes 6-7).

3.6. Instrument-agnostizing of samples chromatograms: general overview

1. Set up standard retention scores (SRS).
2. LC-MS signal acquisition of samples.
3. Export signal to working format.
4. Carry out pre-processing operations.
6. Normalize signal intensities by scaling the chromatographic data vector using internal standard intensities.

7. Normalize retention scores by transferring the retention time scale to SRS domain.
8. Resampling.
9. Verify results (see **Notes 8-19**)

Table 1. Chemical compounds chosen to constitute the standard mixture.

Compound name	CAS	Molecular formula	Monitored ion (*m/z*)	SRS
Trimethoprim	738-70-5	$C_{14}H_{18}N_4O_3$	291.14505	1.000
3,5-Dihydroxybenzoic	99-10-5	$C_7H_6O_4$	153.01933	1.114
Syringic acid	530-57-4	$C_9H_{10}O_5$	197.04555	1.805
Sulfadimethoxine	122-11-2	$C_{12}H_{14}N_4O_4S$	311.08081	2.750
Methylparaben	99-76-3	C_7H_6O	151.04007	3.750
Benzoic acid	65-85-0	$C_7H_6O_2$	121.02950	4.140
Ethylparaben	120-47-8	$C_9H_{10}O_3$	165.05572	6.417
Propiophenone	202-257-6	$C_9H_{10}O$	135.08044	8.034
Benzaldehyde	100-52-7	C_7H_6O	107.04914	8.725
Butanophenone (*Butyrophenone*)	207-799-7	$C_{10}H_{12}O$	149.09609	11.113
Benzophenone	119-61-9	$C_{13}H_{10}O$	183.08044	12.234
Pentanophenone (*Valerophenone*)	213-767-3	$C_{11}H_{14}O$	163.11174	13.415
Benzyl benzoate	204-402-9	$C_{14}H_{12}O_2$	213.09101	13.668
Hexanophenone	213-394-6	$C_{12}H_{16}O$	177.12739	15.101
Heptanophenone	216-802-0	$C_{13}H_{18}O$	191.14113	16.391
Octanophenone (*Caprylophenone*)	216-817-2	$C_{14}H_{20}O$	205.15869	17.429
Nonanophenone	227-861-7	$C_{15}H_{22}O$	219.17434	18.294
Decanophenone	227-946-9	$C_{16}H_{24}O$	233.18999	19.018

4. NOTES

1. Select analytical standard mixture (Table 1) according to the criterion explained in section 3.4.
2. Obtain representative retention times, after the analysis of the standard mix 15 times in the Q-Exactive mass spectrometer (MS system #1).
3. Calculate SRS according to the procedure described in steps 5-7 (Section 3.4) to obtain a linear equation (Table 1 and Figure 1).

4. Pesticides were chosen (see Table 2) as example of target analytes in order to verify the advantages of instrument-agnostizing methodology.

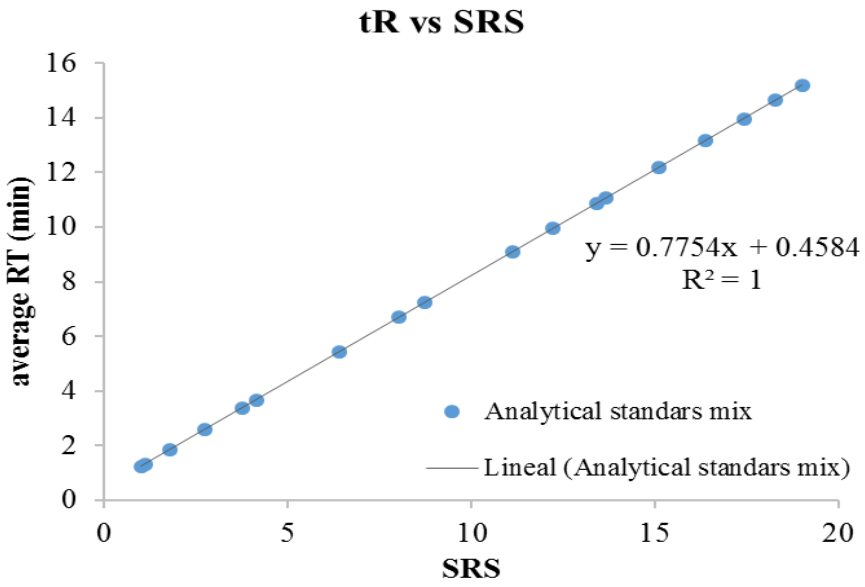


Figure 1. Average RT (min) vs SRS values for analytical standard mixture.

Table 2. Compounds selected to test instrument-agnostizing methodology.

Compound name	Molecular formula	Monitored ion (*m/z*)
Benalaxil	$C_{20}H_{23}NO_3$	326.17507
Desmedipham	$C_{16}H_{16}N_2O_4$	301.11828
Epoxiconazole	$C_{17}H_{13}ClFN_3O$	330.08039
Fenbuconazole	$C_{19}H_{17}ClN_4$	337.12145
Fluquinconazole	$C_{16}H_8Cl_2FN_5O$	376.01627
Imazapyr	$C_{13}H_{15}N_3O_3$	262.11862
Metconazole	$C_{17}H_{22}ClN_3O$	320.15241
Penconazole	$C_{13}H_{15}Cl_2N_3$	284.07157
Pyriproxyfen	$C_{20}H_{19}NO_3$	322.14377
Triticonazole	$C_{17}H_{20}ClN_3O$	318.13676

5. Analyse the mix of pesticides and IS in Q-Exactive mass spectrometer (MS system #1), selecting retention time and peak area at 0.06 and 0.15 mg/L (n=20) (Table 3).

6. Once the RT vector of each sample is collected, an SRS vector should be calculated for each pesticide by applying the methodology described in Section 3.6 (Table 4).

Table 3. Mean of all measurements for RT and peak area in Q-Exactive (MS system *#1)*.

Compound name	RT	Peak area	
		0.06 mg/L	0.15 mg/L
Benalaxil	13.31	599101141	1433235309
Desmedipham	8.94	165665040	690995903
Epoxiconazole	11.29	519462959	1154567356
Fenbuconazole	11.44	388352804	870524681
Fluquinconazole	10.96	111765021	214692592
Imazapyr	2.19	577931822	218545375
Metconazole	12.44	773035186	177327019
Penconazole	11.96	662118942	669196703
Pyriproxyfen	13.83	1465751430	396817976
Triticonazole	11.13	722781016	404038253

7. Follow the same procedure previously described in Exactive mass spectrometer (MS system #2), establishing SRS (Table 4).

Table 4. SRS used for instrument-agnostizing methodology (both MS systems).

Compound name	SRS from MS system #1	SRS from MS system #2
Benalaxil	16.57	16.31
Desmedipham	10.93	10.76
Epoxiconazole	13.97	13.46
Fenbuconazole	14.16	13.49
Fluquinconazole	13.55	13.20
Imazapyr	2.24	2.29
Metconazole	15.46	15.19
Penconazole	14.83	13.60
Pyriproxyfen	17.25	16.99
Triticonazole	13.76	13.40

8. Once pesticides are modelled according to instrument-agnostizing methodology, information of both mass spectrometers is compared to determine the SRS deviation and the possibilities of the instrument-agnostizing (Table 5).

Table 5. SRS deviation between MS system #1 and MS system #2 (Q-Exactive and Exactive mass spectrometers).

Compound name	SRS deviation
Benalaxil	-1.56
Desmedipham	-1.61
Epoxiconazole	-3.70
Fenbuconazole	-4.70
Fluquinconazole	-2.57
Imazapyr	2.48
Metconazole	-1.79
Penconazole	-8.33
Pyriproxyfen	-1.45
Triticonazole	-2.67

9. Criteria used to classify pesticides are:

   $|SRS| > 5\% \Rightarrow$ not valid for instrument-agnostizing methodology

   $|SRS| < 5\% \Rightarrow$ valid for instrument-agnostizing methodology

10. For pesticide studies and according to the criteria exposed, it is possible to detect and determine five out six between the two mass spectrometers (compounds in red, Table 5).
11. Quantification is also studied, using the caffeine and TPP internal standards to normalize the peak area in both mass spectrometers using the procedure described in section 3.7.
12. Evaluate two concentrations to determine the possibilities of quantification (0.06 mg/L and 0.15 mg/L) (Table 6).
13. To test the normalized values, the quantification errors are calculate using the Q-Exactive as reference (MS system #1), and testing concentrations in Exactive (MS system #2) (Table 7).

Table 6. Peak area normalized.

Compound name	Peak area normalized (PAN) in MS system #1		Peak area normalized (PAN) in MS system #2	
	0.06 mg/L	0.15 mg/L	0.06 mg/L	0.15 mg/L
Benalaxil	0.125	0.316	0.109	0.255
Desmedipham	0.048	0.112	0.040	0.117
Epoxiconazole	0.134	0.314	0.117	0.281
Fenbuconazole	0.100	0.235	0.095	0.222
Fluquinconazole	0.029	0.059	0.026	0.060
Imazapyr	0.28	0.577	0.210	0.480
Metconazole	0.044	0.415	0.152	0.365
Penconazole	0.165	0.374	0.137	0.315
Pyriproxyfen	0.309	0.640	0.248	0.599
Triticonazole	0.176	0.110	0.133	0.309

Table 7. Concentration estimated and errors calculated.

Compound name	Ce		%Error	
	0.06 mg/L	0.15 mg/L	0.06 mg/L	0.15 mg/L
Benalaxil	0.053	0.121	-12.58	-19.48
Desmedipham	0.049	0.157	-17.45	4.83
Epoxiconazole	0.052	0.134	-12.71	-10.40
Fenbuconazole	0.052	0.142	-4.53	-5.53
Fluquinconazole	0.054	0.153	-9.22	2.15
Imazapyr	0.049	0.157	-17.75	4.83
Metconazole	0.206	0.132	243.57	-11.97
Penconazole	0.050	0.127	-16.71	-15.50
Pyriproxyfen	0.048	0.140	-19.81	-6.68
Triticonazole	0.046	0.418	-24.23	178.37

14. Calculate the concentrations estimated (C_e) from Exactive using the following formula:

$$C_e = \frac{\text{PAN Exactive} \times C_T}{\text{PAN QExactive}} ;$$

where: PAN is peak area normalized and C_T is theoretical concentration.

15. Estimate errors, using:

$$\%\ \text{Error} = \frac{(C_e - C_T) \times 100}{C_T} ;$$

16. Criteria used to quantify pesticides are:

$|\text{Error}| > 40\% \Rightarrow$ no quantifiable

$|\text{Error}| < 40\% \Rightarrow$ quantifiable

17. According to the criteria, only two pesticides are not quantifiable at lower concentrations (metconazole) and at higher concentrations (triticonazole) (compounds in red, Table 7).
18. Finally, mixing both concepts SRS and quantification, seven compounds of ten are valid for the instrumental-agnostizing methodology.
19. With that seven, it is possible the detection and quantification only using the instrument-agnostizing analytical standards (Table 1) and the IS. Analytical standards of the pesticides are not required.

Acknowledgements

RLR acknowledges "Plan Propio of Investigation" of the University of Almería, co-funded by CAJAMAR and the Operational Program Funds Europeans of Regional Development of Andalusia (2014-2020) (FEDER), for financial support.

SMR author acknowledges the PhD grant supported by the Ministry of Universities for the Training of University Teachers (FPU19/02078).

Reference

1. López-Ruiz R, Romero-González R, Serra B, Garrido-Frenich A (2019) Dissipation kinetic studies of fenamidone and propamocarb in vegetables under greenhouse conditions using liquid and gas chromatography coupled to high-resolution mass spectrometry. Chemosphere 226:36-46. https://doi.org/10.1016/j.chemosphere.2019.03.118
2. López-Ruiz R, Romero-González R, Garrido-Frenich A (2019) Ultrahigh-pressure liquid chromatography-mass spectrometry: an overview of the last decade. Trends Anal Chem 118:170-181. https://doi.org/10.1016/j.trac.2019.05.044
3. Brunius C, Shi L, Landberg R (2016) Large-scale untargeted LC-MS metabolomics data correction using between-batch feature alignment and cluster-based within-batch signal intensity drift correction. Metabolomics 12(173):1-13. https://doi.org/10.1007/s11306-016-1124-4
4. Smith R, Ventura D, Prince JT (2013) LC-MS alignment in theory and practice: a comprehensive algorithmic rewiew. Brief Bioinform 16:104-107. https://doi.org/10.1093/bib/bbt080
5. Cuadros-Rodríguez L, Ortega-Gavilán F, Martín-Torres S, Medina-rodríguez S, Jiménez-Carvelo A, González-Casado A, Bagur-González MG (2021) Standardization of instrumental fingerprints 1 – Part I: Towards an instrumental agnosticism in gas chromatography. J Chromatogr A 1641(461983):1-11. https://doi.org/10.1016/j.chroma.2021.461983
6. Cuadros-Rodríguez L, Martín-Torres S, Ortega-Gavilán F, Jiménez-Carvelo A, López-Ruiz R, Garrido-Frenich A, Bagur-González MG, González-Casado A (2021) Standardization of chromatographic signals –Part II: Expanding instrument-agnostic fingerprints to reverse phase liquid chromatography. J Chromatogr A 1641(461973):1-14. https://doi.org/10.1016/j.chroma.2021.461973
7. Lehotay SJ, Son KA, Kwon H, Koesukwiwat U, Fu W, Mastovska K, Hoh E, Leepipatpiboon N (2010) Comparison of QuEChERS sample preparation methods for the analysis of pesticide residues in fruits and vegetables. J Chromatogr A 1217:2548-2560. https://doi.org/10.1016/j.chroma.2010.01.044

CAPÍTULO IV

ESTUDIO DE ESTABILIDAD Y VIDA MEDIA DE ACEITES VEGETALES COMESTIBLES

CAPÍTULO IV

Estudio de estabilidad y vida media de aceites vegetales comestibles

IV.1. Presentación

Este capítulo recoge los resultados obtenidos al llevar a cabo diferentes estudios multivariables de estabilidad y predicción de la vida útil representativa de diferentes aceites vegetales vírgenes y refinados, destacando el aceite de oliva virgen extra (AOVE).

Los estudios descritos en este capítulo fueron desarrollados en colaboración con la empresa "Laboratorio Juan Antonio Tello", con sede en Jaén, en el marco de un proyecto RETOS-COLABORACIÓN, titulado AVANCES ANALÍTICOS PARA LA MEJORA DE LA INFORMACIÓN SOBRE CALIDAD Y SEGURIDAD DE ACEITES VEGETALES COMESTIBLES Y OTROS PRODUCTOS ALIMENTICIOS VEGETALES DE ALTO CONTENIDO GRASO (CASEITE) (Ref. RTC-2017-6170-2), del Programa Estatal de Investigación, Desarrollo e Innovación orientada a los Retos de la Sociedad, Plan Estatal de Investigación Científica y Técnica y de Innovación 2013-2016.

Los objetivos de este capítulo fueron:

- Obtener un banco de muestras de aceites vegetales representativo de la variabilidad de aceites que se encuentran en el mercado y representativos del tiempo cero de vida del aceite, es decir, aceites recién producidos y filtrados (y, en su caso, justo tras el proceso de refino). Las muestras de aceite serán almacenadas reproduciendo las condiciones en las que se encuentran en los supermercados y se dejarán envejecer durante un período de 12-24 meses.
- Estudiar la estabilidad de diferentes aceites vegetales (destacando el AOVE) a lo largo de su periodo de consumo realizando medidas analíticas aisladas (por ejemplo, determinación de índice de refracción, anisidina, tocoferoles, etc.) y medidas de huella instrumental (por ejemplo, huella obtenida por cromatografía de gases de compuestos volátiles).
- Desarrollar modelos cinéticos multivariables para la determinación del tiempo de vida útil de aceites vegetales comestibles. Emplear herramientas quimiométricas de tratamiento de datos.
- Establecer una serie de índices de calidad/estabilidad frente a la rancidez que permita predecir su fecha de caducidad en cada momento, es decir, número de meses que le quedan a un aceite para dejar de ser considerado como apto según su categoría comercial.

- Comparar aceites, en términos de estabilidad, empleando métodos de clasificación multivariable: discriminación de aceites en función de su vida útil.

Para la consecución de estos objetivos, se realizaron ensayos de envejecimientos tanto en condiciones ambientales normales como en condiciones forzadas que permitían acelerar el proceso. Se emplearon numerosas técnicas analíticas como la espectroscopía, cromatografía de líquidos, cromatografía de gases, entre otras. Todo esto dio lugar a cuatro artículos científicos:

1. PUBLICACIÓN VI: Applications of multivariate data analysis in shelf life studies of edible vegetable oils – A review of the past years.
2. PUBLICACIÓN VII: Multivariate stability monitoring and shelf life models of deterioration of vegetable oils under real time ageing conditions – Part 1: Extra virgin olive oil as a main case of study.
3. PUBLICACIÓN VIII: Multivariate stability monitoring and shelf life models – Part 2: The matter of further edible oils.
4. PUBLICACIÓN IX: A comparison of the stability of refined edible vegetable oils under frying conditions: multivariate fingerprinting approach.

IV.2. Introducción

Los consumidores de productos alimentarios valoran cada día más la calidad alimentaria y, en consecuencia, es de interés avanzar en la información analítica suministrada a éstos, con un buen soporte científico, para la caracterización, en término de su calidad, de aceites y otros productos alimentarios con alto contenido graso.

El concepto de estabilidad de un aceite generalmente se acepta como la vida útil de almacenamiento del producto hasta que la rancidez se hace evidente [**1**]. La oxidación de los lípidos es la causa mayor del deterioro de los productos alimentarios, limitando su vida útil, y una de las principales preocupaciones de la industria alimentaria. Es por ello que el estudio de la estabilidad oxidativa de lípidos resulta una tarea fundamental para garantizar que los productos alimentarios ofrezcan la calidad esperada por el consumidor final en términos de nutrición, aspecto, textura, sabor y aroma.

La degradación oxidativa se refiere a los olores y sabores indeseables que se desarrollan cuando el aceite comestible se expone al oxígeno del aire, donde fundamentalmente los ácidos grasos insaturados de los aceites comestibles reaccionan con el oxígeno para producir peróxidos o hidroperóxidos. Éstos se descomponen para dar lugar a una compleja mezcla de compuestos, entre los que se encuentran aldehídos, cetonas, productos volátiles, entre otros, que son responsable de mal sabor, de los olores y sabores "rancios" en el aceite [**2**]. Además de la alteración de sus propiedades sensoriales, el proceso de oxidación provoca la pérdida de sus compuestos antioxidantes y la reducción de sus propiedades saludables. Estos cambios pueden llevar incluso a una pérdida de la categoría (por ejemplo, de "aceite de oliva virgen extra" a "aceite de oliva virgen" o incluso "aceite de oliva lampante"), con la consiguiente reducción del valor del producto, de la aceptación por parte de los consumidores y la posible inclusión en fraude alimentario.

La mayoría de los ensayos desarrollados hasta el momento para evaluar la oxidación de los lípidos se basan en la detección y cuantificación de productos de oxidación primarios o secundarios: índice de peróxidos, determinación de pirofeofitinas, contenido en 1.2-diglicéridos, absortividad en el ultravioleta, estudio de compuesto volátiles, o la evolución de compuestos fenólicos, entre

---

1. T. Bosso, Olive oil Handbook volumen I, Academic Pages, New York, 2017.
2. Sharma, S., Cheng, S.F., Bhattacharya, B., Chakkaravarthi, S. Efficacy of free and encapsulated natural antioxidants in oxidative stability of edible oil: special emphasis on nanoemulsion-based encapsulation. Trends Food Sci. Technol., 91 (2019), 305-318. https://doi.org/10.1016/j.tifs.2019.07.030

otros [**3**,**4**,**5**]. Sin embargo, el número de parámetros que pueden ser considerados para evaluar la estabilidad o grado de deterioro de aceites es excesivamente amplio para que examinando la evolución de cada uno de forma aislada sea posible en la práctica, adquirir una idea clara de su comportamiento, lo que limita las conclusiones sobre el progreso de la oxidación de los lípidos.

La oxidación de los lípidos es un proceso sumamente complejo que implica a numerosas reacciones que se producen de modo simultáneo y de forma competitiva y que dan lugar a una gran variedad de cambios físicos y químicos. Además, los métodos que combinan la detección simultánea de productos de oxidación primarios y secundarios son escasos. Una revisión reciente sobre los métodos de predicción de la vida útil de aceites propuestos en los últimos años por diversos autores [**6**] pone de manifiesto las debilidades que éstos presentan, fundamentalmente la falta de reproducibilidad por insuficientes muestras en condiciones reales de oxidación, y la necesidad para la industria del aceite de oliva de modelos prácticos y efectivos que puedan ser usados o adoptados fácilmente para predecir razonablemente la vida útil de los aceites, especialmente AOVE, y asegurar que cumplen con los estándares vigentes para su categoría.

A pesar de esto, la normativa actual obliga a los envasadores a señalar una fecha de consumo preferente en la etiqueta de los aceites vegetales comestibles que normalmente, por convenio de la mayoría de grandes envasadoras, se fija entre 12 y 24 meses. Después de ese periodo un aceite vegetal bien conservado, por ejemplo de oliva virgen o virgen extra, se puede consumir con plena garantía sobre su seguridad alimentaria, ya que no es un producto perecedero, lo que sí ocurre es que sus atributos sensoriales empeoran con el tiempo y se merman sus propiedades saludables. Por ello, es común encontrar en el mercado aceites de oliva etiquetados bajo una categoría comercial superior a la que realmente pertenecen y se pueden encontrar

3. Guillaume, C., Ravetti, L. Shelf-Life prediction of extra virgin olive oils using an empirical model based on standard quality tests. J. Chem., 6393962 (2016), 1-7. http://dx.doi.org/10.1155/2016/6393962
4. Aparicio Ruiz, R., Roca, M., Gandul Rojas, B. Mathematical model to predict the formation of pyropheophytin in a virgin olive oil during storage. J. Agr. Food Chem., 60 (2012), 7040−7049. http://dx.doi.org/10.1021/jf3010965
5. Psomiadou, E., Konstantinos, X., Blekas, K.G., Tsimidou, M.Z., Boskou, D. Proposed parameters for monitoring quality of virgin olive oil (Koroneiki cv). Eur. J. Lipid Sci. Technol., 105 (2003), 403-409. http://dx.doi.org/10.1002/ejlt.200300780
6. Xueqi, L., Wang, C.S.C. Shelf life of extra virgin olive oil and its prediction models. J. Food Qual., 1639260 (2018), 1-15. https://doi.org/10.1155/2018/1639260

alertas lanzadas por la OCU (Organización de Consumidores y Usuarios) de aceites de mercado que no cumplen los requisitos de la reglamentación europea [**7**] que regula las características de los aceites para ser considerados como tales, sino que pertenecían a una categoría inferior, más barata.

En los últimos años, se ha avanzado en la comprensión de reacciones complejas relacionadas con la calidad de los alimentos basándose en estudios cinéticos multiparamétricos, en la aplicación de nuevos enfoques de análisis no dirigidos como la metodología de "huella instrumental" [**8**,**9**] que proporcionan información menos sesgada aunque no puede reconocerse de forma evidente o explícita, por lo que es necesario conjuntarla con métodos de modelización mediante la aplicación de algoritmos específicos de extracción de datos (actualmente denominados de minería de datos) basados en el uso de la "quimiometría" [**10**]. Los modelos cinéticos y de vida útil multiparamétricos suelen ser más aplicables genéricamente que los modelos de un solo parámetro y podrían extrapolarse más fácilmente a otros productos o procesos.

Es por ello que en este capítulo se incluye una revisión bibliográfica detallada de los estudios de estabilidad de aceites vegetales en los que se aplica un enfoque multivariable. En ella se describen los parámetros analíticos estudiados, las conclusiones obtenidas, las ventajas y debilidades que presentan y se proponen futuras direcciones para el desarrollo de modelos adecuados de vida útil. Además, en este capítulo se recogen nuevos resultados obtenidos en estudios multivariables de estabilidad y predicción de vida útil de los aceites vegetales comestibles más comunes realizados bajo dos condiciones de envejecimiento diferentes: condiciones de oxidación normales, equivalentes a las condiciones en las que pueden encontrarse los aceites una vez envasados y conservados en las estanterías de un lineal de supermercado; y condiciones de oxidación aceleradas, simulando el proceso de fritura doméstica.

---

7. Reglamento (CEE) Nº 2568/91 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. Última versión consolidada 2016.
8. Alonso Salces, R.M., Holland, M.V., Guillou, C. 1H-NMR fingerprinting to evaluate the stability of olive oil. Food Control, 22 (2011), 2041-2046. https://doi.org/10.1016/j.foodcont.2011.05.026
9. Grauwet, T., Vervoort, L., Colle, I., Van Loey, A., Hendrickx, M. From fingerprinting to kinetics in evaluating food quality changes. Trends Biotechnol., 32 (2014), 3, 125-131. https://doi.org/10.1016/j.tibtech.2014.01.002
10. Derossi, A., Mastrandrea, L., Amodio, M.L., de Chiara, M.L.V., Colelli, G. application of multivariate accelerated test for the shelf life estimation of fresh-cut lettuce. J. Food Eng., 169 (2016), 122-130. http://dx.doi.org/10.1016/j.jfoodeng.2015.08.010

PUBLICACIÓN VI.

ARTÍCULO CIENTÍFICO (REVIEW)

Applications of multivariate data analysis in shelf life studies of edible vegetal oils – A review of the few past years

Food Packaging and Shelf Life, 31, 100790, 1-13, 2022.
https://doi.org/10.1016/j.fpsl.2021.100790

Food Packaging and Shelf Life 31 (2022) 100790

Contents lists available at ScienceDirect

Food Packaging and Shelf Life

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

Applications of multivariate data analysis in shelf life studies of edible vegetal oils – A review of the few past years

Sandra Martín-Torres *, Laura Ruiz-Castro, Ana M. Jiménez-Carvelo, Luis Cuadros-Rodríguez

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

ARTICLE INFO

Keywords:
Food shelf life
Oxidative stability
Multivariate approach
Vegetable oils
Accelerated testing
Real-time testing

ABSTRACT

The physical-chemical and organoleptic properties of vegetal oils are subject to frequent change caused by different degradation process. Hence, producers must provide accurate information on shelf-life prediction. Due to the complexity of chemical interactions between the oil phase and environment, multiple studies under real-time storage and forced conditions have been conducted. This paper reviews the application of methods based on multivariate data analysis, and its advantages, in shelf life and stability studies of edible both virgin and refined vegetable oils. Three multivariate approaches could mainly be differentiated regarding previous studies on foodstuff stability data: unsupervised methods; multivariate pattern recognition methods; and other chemometric resolution methods. Note that, multivariate approaches bring numerous opportunities, these applications currently show also limitations, especially for official control purposes in food surveillance. Future trends will rely on the practical implementation of such promising chemometric approaches combining high informative advanced techniques and comprehensive multivariate data analysis.

Applications of multivariate data analysis in shelf life studies of edible vegetal oils – A review of the few past years

Multivariate data analysis in shelf life studies – Review

Sandra MARTÍN-TORRES✉, Laura RUIZ-CASTRO, Ana M. JIMÉNEZ-CARVELO, Luis CUADROS-RODRÍGUEZ

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

Highlights

- Dietary lipid oxidation is a major cause of food quality deterioration
- Wide number of parameters can be monitored to assess the stability of oils
- Univariate stability models are extremely condition dependent
- Detailed review of multivariate approaches on oil stability and shelf life studies

ABSTRACT

The physical-chemical and organoleptic properties of vegetal oils are subject to frequent change caused by different degradation process. Hence, producers must provide accurate information on shelf-life prediction. Due to the complexity of chemical interactions between the oil phase and environment, multiple studies under real-time storage and forced conditions have been conducted. This paper reviews the application of methods based on multivariate data analysis, and its advantages, in shelf life and stability studies of edible both virgin and refined vegetable oils. Three multivariate approaches could mainly be differentiated regarding previous studies on foodstuff stability data: unsupervised methods; multivariate pattern recognition methods; and other chemometric resolution methods. Note that, multivariate approaches bring numerous opportunities, these applications currently show also limitations, especially for official control purposes in food surveillance. Future trends will rely on the practical implementation of such promising chemometric approaches combining high informative advanced techniques and comprehensive multivariate data analysis.

Key words

Food shelf life; Oxidative stability; Multivariate approach; Vegetable oils; Accelerated testing; Real-time testing.

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

1. INTRODUCTION

Vegetable oils are a key and valuable foodstuff, which is commonly used for frying, salad dressing, preservative, and other applications. A unique case is extra virgin olive oil due to its chemical composition and healthy properties. The European Food Safety Authority (EFSA) claimed that olive oil polyphenols contribute to the protection of blood lipids from oxidative stress (Regulation (EC) No 1924/2006) and recommended a 20 g daily intake of olive oil only if it contains at least 5 mg of hydroxytyrosol and its derivatives (e.g., oleuropein complex and tyrosol) per 20 mg of olive oil. However, the physico-chemical and organoleptic characteristics of edible vegetable oils are not stable but are subject to frequent change caused by different degradation processes. Lipids, consisting mainly of triacylglycerols (>98%) and other minor constituents, are susceptible to oxidative processes such as autoxidation, photo-oxidation, thermal oxidation, and enzymatic oxidation under different conditions, most of which involve some type of free radicals or oxygen species. The rate of oxidation depends mainly on chemical composition (fatty acid profile, natural antioxidants such as phenols or tocopherols, moisture and other minor impurities) and quality of processing, as well as external storage factors such as temperature, air (oxygen), and light.

Among processes, autoxidation is the most common one leading to oxidative deterioration and is defined as the spontaneous reaction of atmospheric oxygen with lipids (**Shahidi, and Zhong, 2005**). Autoxidation is carried out by a free radical chain reaction, involving chain initiation, propagation, and termination, and is a primary pathway for the formation of lipid hydroperoxides, a tasteless and odourless primary oxidation product. Hydroperoxides are, however, generally unstable and lead to the formation of secondary lipid oxidation products such as aldehydes, ketones, alcohols, epoxides, and hydrocarbons, that significantly impact the sensory quality of oils/fats and oils/fats-containing foods such as the well-known rancid perception (**Syed, 2016**). Lipid oxidation not only can produce rancid odours, unpleasant flavours, and discoloration but also can decrease the nutritional quality and safety due to degradation products, resulting in harmful effects on human health. Excellent reviews and books have presented detailed studies about oxidation pathways (**Gunstone, Harwood, and Dijkstra, 2007; Logan, Nienaber, & Pan, 2013; Li & Wang, 2018**) and modelling chemical reaction kinetic (**Marangoni, 2017; Aktar & Adal, 2019**) of lipid and high-fat food system degradation process in an attempt to understand degradation reactions of oils during storage and to establish their shelf life.

Shelf life of foodstuffs is an engaging concept that results from merging scientifically based issues with economic, regulatory, and consumer-related concerns. Institute of Food Science and Technology (IFST) defined shelf life as

the time during which a food product remains safe, comply with label declaration of nutritional data and retain desired sensory, chemical, physical and microbiological characteristics when stored under the recommended conditions (**IFST, 1993**). It involves that a quality threshold should be defined in order to discriminate food products that are still tolerable for consumption from those no longer suitable. This quality threshold is generally defined as the acceptability limit. The time needed to reach the quality threshold corresponding to the acceptability limit is the primary shelf life. This is different from the secondary shelf life which is the time after package the opening that a food product maintains an acceptable quality level and is typically shorter than the primary one (**Piergiovanni & Limbo, 2019**). Figure 1 shows a graphical outline of primary and secondary shelf lives. In this regard, EU Regulation 1169/2011 (**European Commission, 2011**) mandates a statement of how it should be properly stored or used after opening of the package as well as the time limit for product consumption, where appropriate. Information on shelf life after opening should be communicated, such as "best within XX days of opening".

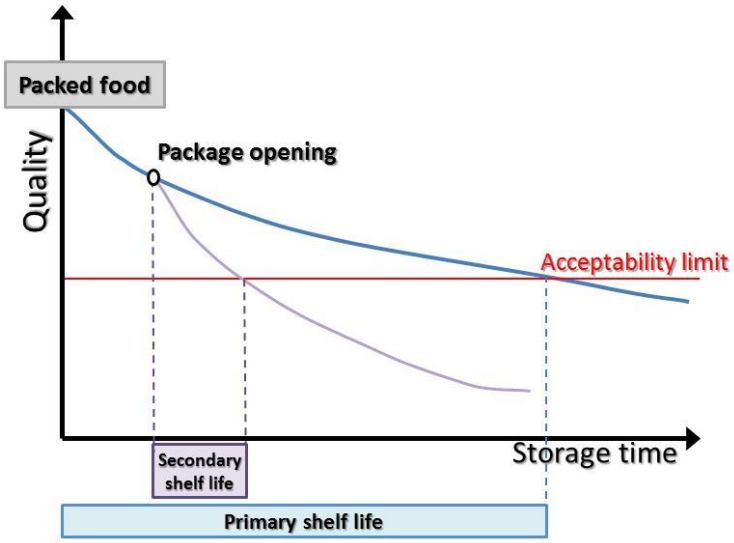


Figure 1. Graphical outline of primary and secondary shelf life extensions.

In recent decades, shelf life meaning has varied reflecting social and economic evolution. Based on this definition, attention should be paid to the difference existing between a shelf life study, aiming to correctly estimate the product shelf life and a stability study, addressed to establish quality depletion kinetics. In other words, stability studies focus on degradation rates whereas shelf life studies address decay times. The difficulty in clearly define the boundaries between the two goals has had repercussions on the methodologies developed to carry out such studies (**Nicoli, 2012**).

Dietary lipid oxidation is a major cause of food quality deterioration and has been a challenge for manufacturers and food scientists alike. As previously noted, the acceptability limit is related to quality and never to safety issues, and, in most cases, it may be freely chosen according to the quality policy of the producer. Several methodologies for evaluating the shelf life of edible vegetal oils have been developed by monitoring lipid oxidation during food storage. Usually, physical and chemical analytical methods and sensory tests are used to assess the oxidative stability and shelf life of oils and fats, and foodstuffs containing oils or fats. There are several analytical methods but only a limited number are routinely used due to the convenience and effectiveness they provide.

One question that persists in lipid oxidation analysis is which oxidation product is the best to be monitored. As a consequence, available methods to monitor lipid oxidation in foods can be classified into three main groups based on what they measure: oxygen absorption, formation of free radicals, and formation of primary and secondary oxidation products. Peroxide value as well as conjugated dienes and trienes measurement (K_{232} and K_{270} specific absorptivity, respectively), tocopherols and pigments profile characterization, determination of total phenol content (**Psomiadou & Tsimidou, 2002**), pyropheophytin evolution (**Aparicio-Ruiz, Roca, & Gandul-Rojas, 2012**), fatty acid composition, diacylglicerols content (**Guillaume & Ravetti, 2016**), gas chromatography (GC) analysis of rancidity marker volatile compounds (**Oeslati *et al.*, 2018**) or empirical based on oxidative stability instruments (e.g., Rancimat) (**Aparicio, Roda, Albi, & Gutiérrez, 1999**) have been the most common analytical approaches used by several authors for measuring edible vegetal oil oxidation. However, there is no uniform and standard method for detecting all oxidative changes in lipidic food systems. In general, these early studies focused on a particular (set of) quality parameter(s), following a 'targeted approach'. The number of experimental setting that can be considered to assess the stability or degree of deterioration of oils is extremely wide. By examining the evolution of each one separately, it is difficult to acquire a clear understanding of their evolution in order to develop valid shelf life models. As an example, kinetic parameters obtained through a single experimental setting are empirical ones linked to the inherent deterioration process for which was established and the related kinetic models are extremely test condition dependent (**Grauwet, Vervoort, Colle, Loey, & Hendrickx, 2014**).

In recent years, relying on multiparametric kinetic studies and modelling approaches, progress has been made in the understanding of complex food quality-related reactions. In this targeted analytical approach, multiple related analytical parameters, rather than a single parameter or measurement, of a particular food quality-related reaction are evaluated as a function of the

extrinsic process variables. Multiparametric kinetic and stability models are often more generically applicable than single-parameter models and could be more easily applied to other products or processes, although parameter re-estimation and even model reformulation might be necessary to take into account the complex reaction environment the food system offers (**Grauwet, Vervoort, Colle, Loey, & Hendrickx, 2014**).

In order to have more accurate and convenient understanding of vegetable oil stability, which could be applied to properly predict reliable values of edible oil shelf life, this paper focuses on reviewing the background on the application of multivariate-based chemometric methods on vegetable oil stability studies which allows simultaneous monitoring of many parameters. In addition, some future directions of shelf life prediction models are proposed.

2. CHEMOMETRIC-SUPPORTED SHELF LIFE ASSESSMENT

A shelf life study of food includes three fundamental steps. The first one involves the identification of the most critical events leading to product quality depletion and the choice of a relevant acceptability limit. The second step is to evaluate the changes in the selected quality experimental indicators as a function of time under storage conditions. The evaluation can be conducted emulating the expected storage conditions (real-time shelf life testing) or under environmentally-forced conditions able to accelerate quality depletion, so-called accelerated shelf life testing (ASLT). Finally, data are modelled to obtain a shelf life estimation or prediction (**Calligarid, Manzocco, Anese, & Nicoli, 2015**). When conducting an oil-shelf life study, selecting appropriate analytical methods and oxidative indicators (markers) is of critical importance. As mentioned previously, although numerous analytical methods exist, a limited array of methods is used routinely due to the convenience and effectiveness they provide.

The edible oil industry has a great need to obtain, in a relatively short time, the necessary information for determining the shelf life of oil products. Univariate methods are still the most used for shelf life assessment, although they generally only offer a very limited overview of the real problem. For decades, the evolution of change of a specific quality characteristic has been studied, typically performing a kinetic study in which the processing variables were completely controlled and monitored. Advances in analytical technology and the increasing availability of powerful instrumentation now offer food analytical chemists the potential of obtaining high amounts of data in an acceptable time frame (**Pedro & Ferreira, 2006**).

In the recent decade, the fingerprinting strategies are gaining more and more popularity. Fingerprinting approach has been a common practice on food quality

and authentication issues. Instrumental fingerprint refers to the characteristic unspecific instrumental signal from the analysed sample which can be related to its properties, complex chemical composition and authenticity. The term thus recalls a comprehensive description of a test material that is carried out in a non-selective (untargeted) way. Instrumental fingerprints are acquired and recorded by an analytical instrument and they require further mathematical data treatment (**Jiménez Carvelo, Martín Torres, Cuadros Rodríguez, & González Casado, 2021**). However, fewer examples of shelf life studies have been reported integrating fingerprinting in kinetic modelling (**Grauwet, Vervoort, Colle, Loey, & Hendrickx, 2014; Kedebe *et al.*, 2015; Chaudry *et al.*, 2018**).

Multivariate strategies within the framework of chemometrics are suitable for carrying out such operations, allowing a more complete interpretation of data structures. Chemometrics has been defined as the science of relating measurements made on a chemical system or process to the state of the system via the application of mathematical or statistical methods. It should not involve theoretical calculations but deal primarily with the extraction of useful chemical information from measured data (**Hibbert, 2016**). However, the availability of large data set does not mean the availability of accessible information of the analysed sample: properly interpretation of the hidden and non-evident information embedded within the data is necessary. Chemometric tools try to find the relationships between sample features and variables in a given data set and converting to new latent variables. Multivariate data analysis (MVDA) involves the analysis of data consisting of a large number of variables measured from a particular sample bank. Three approaches could mainly be differentiated regarding multivariate data analysis of edible oil stability data: unsupervised methods; multivariate pattern recognition methods (PRMs); and chemometric resolution methods (CRMs).

Unsupervised methods do not require training input to find the output: no additional knowledge (e.g., Y-variable) besides raw data (X-variable) is required to describe the data set. Models work on their own to improve understanding and accessibility of the intrinsic features of the data. Among them, multivariate exploratory data analysis (MEDA) and comparative methods can be differentiated. MEDA refers to the process of revealing hidden and unknown structures from data in such a form that the analyst obtains an immediate, direct and easy-to-understand graphical overview (**Li Vigni, Durante, & Cocchi, 2013**). Of all such methods, principal component analysis (PCA) (**Bro & Smilde, 2014**) and cluster analysis (CA) (**Drab & Daszykowski, 2014**) are the most prevalent in the literature probably because of their ability to reduce large and complex data matrices to a more readable outcome. PCA looks for a smaller number of underlying factors that explain most of the variability exhibited by the larger number of measurements made on the objects/samples.

This is based on the degree of correlation between the variables: inter-correlated variables can 'represent' one another and constitute the first principal component (PC1). A second PC2 (uncorrelated with PC1) can then be derived to examine more variability and successively. CA groups similar objects based on an assessment of the distance between measured variables, it is also a widely used exploratory method that searches with no prior definition (**Bower, 2013**).

Multivariate analysis of variance (MANOVA) is a very useful comparative unsupervised tool for the analysis of data (both one-way and two-way MANOVA is used). It is a statistical method to compare the population means of two or more groups by analysing variance. The variance would differ only when the means are significantly different (**Gower & Krzanowski, 1999**). However, MANOVA has been scarcely used in food stability and shelf life studies and only a history of univariate version of the analysis of variance (ANOVA) is available.

PRMs are supervised learning/prediction methods based on multivariate regression analysis. Regression fits mathematical relationships between variables or groups of variables and provides functional models for qualitative or quantitative predictions. Regression techniques can be linear or non-linear, depending on the type of relationship they are able to model. Both linear and non-linear approaches have been successfully applied on different edible oil studies for the estimation of days of storage and the prediction of shelf life values. Although the classical multivariate least-squares regression (MLR) method can be used, the most widely used PRMs algorithm for quantitative predictions is partial least-squares (PLS) regression. PLS is a supervised multivariate method because apart from the information on the X-variables measured, the available knowledge on a dependent response Y-variable (usually, storage time) is applied to obtain a latent variable model that optimally describes the response variable. PLS computes a partial least squares regression model to predict a dependent y-variable from a set of independent x-variables (**Jiménez Carvelo, Martín Torres, Cuadros Rodríguez, & González Casado, 2021**). Linear discriminant analysis (LDA), closely related to MANOVA and regression analysis, which also attempt to express one dependent variable as a linear combination of other features or measurements had also been commonly used. Similarly, it is also feasible to perform non-linear modelling by applying quadratic discriminant analysis (QDA) (**Berrueta, Alonso-Salces, & Héberger, 2007**).

Qualitative prediction methods are frequently referred to as classification methods. k-Nearest neighbour (k-NN) is one of the simplest and widely used methods for qualitative prediction. Nearest neighbour methods are based on the determination of the distances between an unknown object and each of the objects of a training set. Then, the lowest distance is selected for the

assignment of the class membership. Its mathematical simplicity, which does not prevent it from achieving classification results as good as other more complex pattern recognition techniques. k-NN can be also used for quantification (**Song, Liang, Lu, & Zhao, 2017**) but is by far more popularly used for classification problems. In addition, PLS discriminant analysis (PLS-DA) is a widespread classification method based on PLS linear regression (**Lee, Liong, & Jemain, 2018**).

The common purpose of all CRMs is to provide a linear model of individual component contributions from unspecific instrumental fingerprints using solely the raw experimental measurements provided that the signal profile for each concerned component is known in advance. The most well-known CRM method is multivariate curve resolution (MCR) (**de Juan, Jaumot, & Tauler, 2014**). When 2D analytical signals are obtained, e.g., spectrum-chromatograms, specific methods are available that decompose mathematically a mixed signal into the pure contributions due to each of the components in the system when no prior information about the nature and composition (**Jiménez Carvelo & Cuadros Rodríguez, 2021**). Two commonly examples, which can combine the three approaches of multivariate data analysis, are multivariate curve resolution-alternating least squares (MCR-ALS) and parallel factor analysis (PARAFAC) (**Olivieri, 2012; Anzardi, Arancibia, & Olivieri, 2021**).

Special cases are deep learning methods of which artificial neural networks (ANNs) are the finest example. ANNs are a family of nonparametric versatile tools that can be used both for qualitative and quantitative predictive modelling. ANN are defined as structures comprised of densely interconnected adaptive simple processing elements, called artificial neurons (or nodes) that are capable of performing massively parallel computations for data processing and knowledge representation. They are generally well suited for nonlinear problems, and the related software is easily available (**Berrueta, Alonso-Salces, & Héberger, 2007**). In addition, support vector machines (SVM) and random forest (RF) are two supervised data mining methods that are increasingly being used in food science to build both qualitative and quantitative multivariate models (**Jiménez Carvelo, González Casado, Bagur González, & Cuadros Rodríguez, 2019**).

2.1. Multivariate approach on the determination of shelf life and stability studies: Forced conditions

ASLT applies to any deterioration process that has a valid kinetic model. ASLT allows estimation of the shelf life data at storage conditions usually experienced by the product on the market by using experimental data from the accelerated depletion process acquired at forced storage conditions. When performing

ASLT, a proper accelerating factor to be used in the test should be chosen. The absolute necessary requirement to apply ASLT is the knowledge of the relationship between the accelerating factor and oxidation/degradation rate. Only if this requisite is complied, the extrapolation to actual conditions could be used to predict shelf life. The main environmental factors exploited to perform ASLT of oil and oil-based products are temperature and light irradiation (**Syed, 2016**).

Several authors have applied multivariate tools in order to reach shelf life or to carry out comparison stabilities studies of edible vegetable oils. The first attempt to assess virgin olive oil (VOO) stability under forced conditions using chemometric tools was carried out by **Aparicio, Roda, Albi, & Gutiérrez (1999)**. They measured oxidative stability by the Rancimat method at 100 ºC and evaluated the contribution of different parameters, such as tocopherols and phenolic compounds, with the stability measured using PCA and linear regression methods. **Maggio *et al.* (2011)** also studied oxidative stability of extra-virgin olive oil (EVOO) but using oil stability index measurements (110 ºC). Their main contribution was the use of spectroscopic profiles such as near-infrared (NIR) or Fourier-transform mid-infrared (FT-MIR) together with PLS to distinguish among the different oils and the different heating times. Similar thermal-oxidative stability of EVOO study was carried out by **Selaimia, Oumeddour, & Nigri (2017)**. However, neither of them comes to valid conclusions on comparison of relative stability and shelf life time. **Le Dréau, Dupuy, Artaud, Ollivier, & Kister (2009)** used FT-MIR spectra to compare different vegetable refined oils (such as rapeseed, sunflower, etc.) according to their stability. They introduce the use of MCR-ALS to determine the contribution of a chemical compound to the oxidation process and **Wójcicki, Khmelinskii, Sikorski, & Sikorska (2015)** combine MCR-ALS with PLS to correlate quantitatively FT-MIR attenuated total reflection and FT-NIR spectral changes with peroxide values.

The applicability of electronic sensing (e-sensing) devices to monitor lipid oxidation under forced conditions has been also described. **Mildner-Szkudlarz, Jelen, & Zawirska-Wojtasiak (2008)** reported the use of electronic and human nose to monitor rapeseed oil oxidation; **Xu, Yu, Liu, & Zhang (2016)** concluded that e-nose combined with CA, PCA and LDA models could be used for differentiation of non-oxidized and oxidized oils. **Buratti, Malegori, Benedetti, Oliveri, & Giovanelli (2018)** proposed a multivariate strategy that enabled to characterize olive oils (OO) of different categories on the basis of their sensorial properties, and also to follow the evolution of OO samples during storage at different temperatures from e-nose, e-tongue and e-eye fussed data and classified samples into oxidized and fresh using the k-NN algorithm. All of them conclude on the opportunity to use e-sensing methods as a promising tool in

quality control of oils to verify sensory and rancidity changes but reliable conclusions on stability time are lacking. Volatile compound changes have been monitored by **Poyato, Ansorena, Navarro-Blasco, & Astiasarán (2014)** to estimate virgin and refined vegetable oil intensity of oxidation. They used MLR methods to predict the formation and semi-quantification of volatile aldehydes. On the contrary, **Ennouri *et al.* (2019)** have recently reported that models based on MLR cannot predict oxidative stability of studied olive oils with similar accuracy as the obtained using ANNs. **Upadhyay & Mishra (2015)** described a PLS model to predict, with low error, sunflower oil shelf life at 60 ºC only monitoring the formation of conjugated dienes (60 ºC) and measuring the Rancimat induction time (100-130 ºC). They extrapolated the results and estimated that shelf life of oils (blended with antioxidants at different concentration levels) at 25 ranges from 145 to 337 days. That is the only ASLT precedent in which the shell life of edible oil other than virgin olive oils (VOOs) is estimated using multivariate tools.

Table 1 provides a comprehensive summary of oil type, forced storage conditions, chemical-physical and sensory features, chemometric tool, and shelf life/stability conclusions reported during the last years.

Table 1. Summary of shelf life and stability models of vegetable oils developed by several authors under forced conditions of oxidation.

Oil	Storage conditions	Chemical/physical and sensory features	Chemometric tool	Shelf life/Stability conclusions	Ref.
Virgin olive	Not apply	Chlorophylls Carotenoids Tocopherols Phenolic compounds Fatty acids Oxidative stability: Rancimat (100 ºC)	PCA Stepwise and piecewise MLR	Linear regression based on the oleic/linoleic ratio and the contents of phenols and tocopherols showed a good correlation with the Rancimat stability outcomes. The contribution of phenolic compounds to stability was around 51%. No effect, or very little, was shown by tocopherols.	**Aparicio, Roda, Albi, and Gutiérrez (1999)**
Refined rapeseed	- 12 days at 60 ºC darkness - 6 month room Tª; sampling every month	HS-e-nose Sensory analysis Volatile compounds Peroxide value Acid value	PCA PLS	The applicability of e-nose to monitor lipid oxidation.	**Mildner-Szkudlarz, Jelen, and Zawirska-Wojtasiak (2008)**
Peanut Rapeseed (virgin and refined) Walnut Grapeseed Soybean Sunflower	Heating from 25ºC up to 130ºC under argon at 11ºC min^{-1} on an ageing cell developed by authors	Infrared spectra	PCA MRC-ALS	Induction time, half time and final time determined by oxidative spectroscopy index to compare oils according to their stability. MCR-ALS to determine the contribution of chemical compound to oxidation.	**Le Dréau, Dupuy, Artaud, Ollivier, and Kister (2009)**
Extra virgin olive	Conventional heating at 180 ºC in an oven during 180 min; sampling every 30 min	Fatty acid composition UV spectrometric indices (K_{232} and K_{270}) Oil stability index (110 ºC) FT-MIR spectroscopy Vis-NIR spectroscopy	PLS PCA	Significant increase of K_{270} with heating, while a common trend for K_{232} among samples was not evidenced. PLS from spectroscopic data was able to distinguish among different oils	**Maggio *et al*. (2011)**

				and different heating times: useful for monitoring the oxidative status of cooked oils.	
Rapeseed Sunflower Virgin olive	170ºC, 190ºC and 210 ªC during 180 min; Sampling every 30 min	^{1}H NMR	PCA ANOVA	A model to monitor the appearance of aldehydes. Constant of thermal stability: olive oil is more stable than sunflower, and that, in turn, more than rapeseed.	**Cordella, Tekye, Rutledge, and Leardi (2012)**
Virgin linseed Algae Sunflower High-oleic sunflower Extra virgin olive Refined olive	Firstly 0, 2 and 4 h at 180 ºC (air exchange during heating) Secondly, cooling for 15 min in and ice bath.	Fatty acid profile Volatile aldehydes Thiobarbituric acid reactive substances (TBARS) value	ANOVA PCA MLR	A regression model to predict the formation and semi-quantification of volatile aldehydes, so that, estimate intensity of oxidation. Using TBARS for assessing oxidation in heated oils.	**Poyato, Ansorena, Navarro-Blasco, and Astiasarán (2014)**
Soya Corn Sunflower Canola Olive	Heated from 30 until 170 ºC by steps of 10 ºC	UV-VIS spectroscopy	MCR-ALS	MCR-ALS method provided information over tocopherol degradation and oxidation products appearance due to the heating even when evaluated by UV-Vis spectroscopy. Sunflower, colza and olive oils offered more stability to increasing temperatures, while soybean and corn oils were less resistant	**Gonçalves, Março, and Valderrama, (2014)**
Extra virgin olive Pomace Refined olive Rapeseed	Heating into a convection oven, 60 ºC during 15 days. Sampling	Peroxide value MIR and NIR spectra	PCA MCR-ALS PLS	Oxidative stability decreases in the order: olive to rapeseed to sunflower oil. Spectral changes evaluated were quantitatively correlated with peroxide value.	**Wójcicki, Khmelinskii, Sikorski, and Sikorska (2015)**

(cold-pressed and refined)	every third day			Calibration models underline the correlations between the entire spectra and peroxide value.	
Sunflower (cold-pressed and refined)					
Sunflower oil-sage extracts blends	Not apply	Photochemiluminescence assay Rancimat test (100, 110, 120 and 130 ºC)	ANOVA PCA HCA	PCL assay and Rancimat test were correlated and could be potential instrumental methods of antioxidant activity measurement. Shelf life model generated using the induction period as a function of temperature for estimating the shelf life of sunflower oil blended with sage extracts under accelerated conditions	**Upadhyay, & Mishra (2015)**
Sunflower blended with oleoresin rosemary (ROS) and ascorbyl palmitate (AP)	Incubation maintained at 60 ± 1 ºC under dark and with headspace-controlled oxygen content; sampling every 5 days	Peroxide value Acid value Total polar matter Antioxidant capacity Conjugated diene value Rancimat stability (100, 110, 120 and 130 ºC)	ANOVA PLS	ROS and AP were found to be directly associated with the oxidative stability of oil. Significant correlation between chemical composition parameters and oxidative stability. PLS model to predict shelf life at 60 ºC with low error: unified model which required only the measuring of formation of conjugated dienes (60ºC) and induction period (100-130 ºC). Shelf lives extrapolated at 25 ºC were estimated to be from 145 to 337 days.	**Upadhyay, & Mishra (2015)**

Rapeseed	110 ºC in an oven test (analysed immediately after opening and at the end of induction time previously evaluated at 110 ºC)	Moisture content Peroxide value Total carotenoids content Total chlorophylls Total phenolic compounds Tocopherols Phytosterols Fatty acid composition Rancimat stability (110 ºC)	PCA	Induction time correlated mostly with phenolics content and chlorophylls while carotenoids acted as pro-oxidants. The degradation degree of tested compounds during oxidation was: phytosterols > tocopherols > carotenoids > phenolic compounds > chlorophylls.	**Roszkowska, Tanska, Czaplicki, and Konopka (2015)**
Olive Peanut Soybean Rapeseed Camelia Corn Sunflower Linseed Walnut	12-36 h at 105 ºC	Peroxide value Acid value E-nose	CA PCA LDA	E-nose has adequate selectivity and sensitivity to discriminate the degree of oxidation of edible oils. Results from CA, PCA and LDA models indicated that the e-nose technique could be used for differentiation of non-oxidized and oxidized oils. LDA produced slightly better results than CA and PCA.	**Xu, Yu, Liu, and Zhang (2016)**
Palm Camellia Sunflower Perilla	35 days at 62±1 ºC randomly inside an oven; sampling every 7 days	Flourescence spectroscopy analysis Tocopherol Acid value Peroxide value Anisidine value Nonpolar carbonyl compounds	ANOVA PCA Pearson correlation analysis	Fluorescence spectroscopy combined with PCA could be used differentiate different oils and to monitor their oxidation evolution. Fluorescence combined with regression could further determine the relationships between fluorescence information of oils and primary or secondary oxidation products.	**Cao, Li, Liu, Fan, and Deng (2017)**
Extra virgin olive	90 days in stove set at 60 ºC; sampling every 10 days	Free fatty acid Peroxide value Saponification number Iodine value FTIR spectra	PCA PLS	The IR spectra show that the oxidation process starts with the formation of hydroperoxides. FTIR-PLS models were able to distinguish different oils' regions	**Selaimia, Oumeddour, and Nigri (2017)**

				quality and the effect of heating times: useful for monitoring the oxidative status of heating oils.	
Chia	48 h under Schaal oven test conditions (65 ºC)	Acidity value Peroxide value Humidity Fatty acid profile FTIR-ATR spectra TGA DSC UV-VIS spectroscopy	MCR-ALS	Comparison of stability of chia oil and oil-loaded microparticles. Degradation was evaluated by the UV-Vis spectroscopy coupled to chemometrics and also by the extinction coefficients, corroborating the increase in the oxidative stability.	**Guimaräes-Inácio *et al.* (2018)**
Corn Olive (virgin and extra virgin) Sunflower High oleic sunflower Rapeseed (refined and virgin) Sesame Toasted sesame Coconut (refined and virgin) Red pal (virgin)	Dark cold place before analysis	Fatty acid composition: saturated (SFA), unsaturated (UFA) and polyunsaturated (PUFA) Tocopherols β-carotene Chlorophylls Total phenolic content (TPC) Rancimat induction period (IP) (120 ºC)	ANOVA PCA MLR	The oils showed a huge diversity in initial composition. Samples are grouped into three different clusters according to their compositional data and oxidative stability. Type of fatty acid is critically important for the oxidative stability, but to a lesser extent than previously described for refined oils. TPC and SFA were the most important individual factors, which correlated positively with IP. UFA, PUFA and total tocopherols showed a negative correlation with oxidative stability.	**Redondo-Cuevas, Castellano, Torrens, and Raikos (2018)**
Extra virgin olive (EVOO) Olive (OO) Pomace olive	EVOO: open in dark at: - 60 ºC, 37 days - 40 ºC, 67 days	Acidity Peroxide value UV absorption indices (K_{232}, K_{270} and ΔK) Total phenols Antioxidant activity	ANOVA PCA k-NN	Multivariate strategy enabled to characterize olive oils of different categories on the basis of their sensorial properties, and also to follow the evolution of EVOO and OO samples during storage at	**Buratti, Malegori, Benedetti, Oliveri, and Giovanelli (2018)**

		Tocopherols		different temperatures.	
	OO: open in dark at 40 ºC, 74 days	Tyrosol, hydroxytyrosol and oleuropein content Chlorophylls and carotenoids E-senses analyses: e-nose, e-tongue and e-eye		Samples are grouped on the basis of their freshness and two classes were identified: fresh and oxidized using physic-chemical data. The k-NN classification model has evidenced the e-senses ability to classify samples in the two classes.	
Olive (in-house extracted)	Not apply	Rancimat stability (50-220 ºC) Fatty acid profile Phenol profile Sterols	ANOVA MLR ANN	Models based on MLR cannot predict oxidative stability of studied olive oils with similar accuracy as the obtained for the selected neural networks. Modelling based on artificial neural networks brings the possibility of the use of networks for prediction of more complex parameters.	**Ennouri *et al.* (2019)**

ANN: artificial neural network
ANOVA: analysis of variance
DSC: differential scanning calorimetry
FT-MIR: Fourier-transform mid infrared
FTIR-ATR: Fourier-transform mid infrared-attenuated total reflection
HCA: hierarchical cluster analysis
^{1}H MNR: nuclear magnetic resonance of ^{1}H
k-NN: k-nearest neighbour
LDA: linear discriminant analysis
MLR: multivariate least-squares regression
MCR-ALS: multivariate curve resolution-alternating least squares
PCA: principal component analysis
PLS: partial least-squares regression
TGA: thermogravimetric analysis
Vis-NIR: visible-near infrared

2.2. Multivariate approach on the determination of shelf life and stability studies: ambient conditions

Real-time shelf life testing might be ideally exploited for any food category, but it is actually advantageous only for perishable foods, which undergo quick changes. For practical reasons, especially when the actual storage time is long, the industry resorts to accelerated test techniques that considerably shorten the process of obtaining the necessary experimental data. That is the main reason why the majority of stability studies of refined edible oil are accelerated studies. On the contrary, the majority of real time conditions stability studies are about commercially valuable virgin oils, in particular extra virgin olive oil (EVOO).

EVOO stability was studied by **Zanoni, Bertuccioli, Rovellini, Marotta, & Mattei (2005).** They proposed a phenomenological model of EVOO stability based on acidity value (indirectly related to stability), oleic acid content and bitter taste (directly related to stability) and stated the possibility of setting up predictive models of oil degradation extent monitoring only peroxide value, K_{232} and lipid oxidation status (dienoic and trienoic conjugated isomer and oxidised fatty acid contents) using liquid chromatography coupled to an ultraviolet detector. They checked their results only on 11 oil samples and the model was able to predict the rate of oil degradation, though in an indirect, semi-quantitative way. More experimental experience is missing**. Aparicio-Ruiz, Roca, & Gandul-Rojas (2016)** devised a suitable mathematical multivariate model for predicting the stability and loss of freshness of VOO at room temperature only by analyzing chlorophyll pigments using a liquid chromatography coupled to a corona aerosol detector. They described and validated an Arrhenius model for mathematically expressing the percentage of pyropheophytin a, according to the time and temperature of storage. They stated that the kinetic prediction model is useful for the VOO producers and wholesalers so they can have, a priori, an estimation of the maximum storage time for VOO under controlled temperature. Nevertheless, examining the evolution of pigments separately, a non-regulatory parameter, it is not possible in practice to acquire a clear understanding of oil evolution and to develop a real shelf life model. Visible-ultraviolet (UV-Vis) absorption spectroscopy was also used by **Gonçalves *et al.* (2018)** who stated that UV-Vis spectra in conjunction with MCR-ALS are a feasible tool to monitor autoxidation processes in edible oils through storage time and in the monitoring of the EVOO quality in different packaging systems. They proved that the glass bottle system provides more protection for the autoxidation processes with the time for the EVOO.

Sensory analysis has been a key factor for real conditions oil degradation studies. Several authors have reported the use of e-senses and compared them with trained panel member estimations. **Savarese *et al*. (2013)** described the advantages of e-nose concerning organoleptic characteristics and instrumental

traditional techniques and discriminated between oils of different quality (EVOO and lampante) and different intensity for a fruity or rancid attribute. **Rodrigues, Dias, Veloso, Pereira, & Peres (2016)** used e-tongue sensors and proposed that a single electrochemical assay enable the simultaneous quantification of physicochemical olive oil parameters. Their results showed that EVOOs stored at ambient temperature and mimic supermarket conditions suffered degradation in such a level that no longer could be classified as EVOOs after 9-12 months of storage. However, they used 36 dark amber glass bottles samples to draw their conclusions while most of the edible oil on the market is stored in transparent PET so that biased outcomes are supposed. **Karami, Rasekh, & Mirzaee-Ghaleh (2020)** employed e-nose on the stability study of refined vegetable oils (sunflower, canola and soy) stored at 24°C for 5 months and classified samples into oxidized and non-oxidized oils comparing their results with the ones obtained by AOCS methods. They also concluded that the SVM method possessed higher accuracy in the determination of the shelf life of the edible oils as compared with LDA and QDA methods.

Ün & OK (2018) studied specific changes in the chemical composition of VOOs during their shelf life time by ^{1}H NMR. **Alonso-Salces *et al*. (2021)** applied ^{1}H NMR analysis and obtained a metabolic fingerprint of VOO that provides valuable approaches for the quality control of VOOs, as well as to assess their stability. They also set up predictive models that disclosed the chemical compounds responsible for the compositional in VOO, which helps to understand the quality changes that experience during its shelf life. Both studies agree on the chemical compounds causing VOO oxidation but no prediction of stability time is given.

Near-infrared spectrum analysis merged with appropriate chemometric methods was tested to be an efficient tool to realize the rapid and high-precision detection of refined edible oils in different storage periods by **He, Jiang, & Chen (2020).** Their conclusions agreed that the SVM model displayed the best stability and generalization performance in the qualitative identification of the actual storage periods of edible oil. **Kharbach *et al*. (2021)** studied virgin argan oils and proposed FTIR spectra combined with chemometrics to form an interesting approach to survey the shelf life of this vegetal oil as well as the good predictive ability for physicochemical parameters. Finally, fluorescence spectra of monovarietal VOOs were obtained by **Lobo-Prieto, Tena, Aparicio-Ruiz, García-González, & Sikorska (2020)** who demonstrated that excitation-emission fluorescence spectroscopy combined with the PARAFAC model provides a degradation map of the oil sample and assessed the relationship between fluorescence spectra and different quality parameters, including organoleptic attributes.

Table 2. Summary of shelf life and stability models of vegetable oils developed by several authors under ambient conditions of oxidation.

Oil	Storage conditions	Chemical/physical and sensory features	Chemometric tool	Shelf life/Stability conclusions	Ref.
Virgin olive oil (VOO) (in-house extracted of olives affected by *Dacus oleae* infestation differently)	Not apply	Polar compounds	PCA PLS	PLS model shows a significant correlation between the composition of polar compounds and the state of health of the olives. Orthodiphenolic compounds are useful tool to differentiate oils according to the state of health of the olives.	**Evangelisti *et al.* (1997)**
Extra virgin olive oil (EVOO)	Different commercial activities simulations: - dark glass bottles at 20 ºC for 21 months - stored in Australian and Italian supermarkets under uncontrolled ambient conditions for 14-16 months Sampling every 2 months	Acidity Peroxide value Phenol content Tyrosol and hydroxytyrosol contents α-Tocopherol Rancimat stability (120 ºC) K_{232} and K_{270} Spectroscopic indices related to chlorophylls and carotenoids Organoleptic characteristics	PCA PLS	A few parameters were found to be significant: hydroxytyrosol and tyrosol contents, carotenoid absorbance at 475 and 448 nm, α-tocopherol content, Rancimat induction time, and $K_{232.}$ Stability of EVOO was not significantly influenced by different uncontrolled bottling, transport, and storage conditions. Empirical models were set up to predict the time to reach a reference value for K_{232}.	**Pagliatini, Zanoni, and Giovanelli (2000)**

Extra virgin olive oil (EVOO)	Not apply	Acidity Fatty acid content Bitter taste intensity Peroxide value K_{232} and K_{270} Total minor polar compounds Lipid oxidation status: dienoic and trienoic conjugated isomer contents were measured by HPLC-UV at 230 nm and oxidised fatty acid contents Antioxidant activity Organoleptic characteristics	PCA PLS	EVOO stability model: stability/instability indices are acidity value (indirectly related to stability), oleic acid content and bitter taste (directly related to stability). They may cause sensitivity to degradation during oil shelf-life. Degradation parameters to be monitored: peroxide value, K_{232} and lipid oxidation. The more acidity the more degradation, the more oleic acid content the less degradation, and the more bitter the taste the less degradation was observed. Phenomenological model was applied to set up some predictive models of oil degradation extent.	**Zanoni, Bertuccioli, Rovellini, Marotta, and Mattei (2005)**
Monovarietal extra virgin olive oil (EVOO) Rancid olive oil	18 months in optimal storage conditions	Volatile compounds Organoleptic characteristics E-nose	PCA	Advantages of e-nose respect to sensory analysis and instrumental traditional techniques: short time; no pre-treatments; simple response; synthetic easy interpretable way; possibility to monitoring processes on line. Discrimination between oils of different quality (EVOO and lampante) and between different intensity for fruity or rancid attribute. The e-nose proved to be a useful tool for the monitoring of oil flavour evolution during storage.	**Battimo *et al.* (2007)**

Monovarietal extra virgin olive oil (EVOO) Rancid olive oil	18 month at room temperature in the dark	Organoleptic characteristics E-nose Volatile compounds	PCA	Discrimination among olive oils of different quality (EVOO and lampante oil) and among different intensities of the fruity and rancid attributes. E-nose was proved to be a useful tool for the monitoring of aroma evolution during storage of bottled EVOO.	**Savarese *et al.* (2013)**
Extra virgin olive oil (EVOO) blends	Ambient temperature (17-25 ºC) mimic supermarket conditions	Free acidity Peroxide value K_{232}, K_{270}, and ΔK Rancimat stability (120 ºC) E-tongue	ANOVA MLR PCR PLS	Physicochemical parameters varied with the storage lighting conditions and more significantly with time. Some of EVOOs studied suffered degradation in such a level that no longer could be classified as EVOOs after 9-12 months of storage. Single electrochemical assay enables the simultaneously quantification of physicochemical olive oil parameters, reducing time and cost; prior to its implementation a significant larger number of samples must be evaluated.	**Rodrigues, Dias, Veloso, Pereira, and Peres (2016)**
Monovarietal virgin olive oil (VOO)	Amber glass jars closed to airtight at rom temperatures (10.4-28.6 ºC) in darkness during one year; Sampling every month	Chlorophyll pigments	ANOVA PCA MLR	Mathematical model created using multivariate statistical procedures that describes the changes of pyropheophytin a and is used in the prediction of the stability and loss of freshness of VOO. Arrhenius model for mathematically expressing the percentage of pyropheophytin a, according to the time and temperature of storage (validated).	**Aparicio-Ruiz, Roca, and Gandul-Rojas (2016)**

Extra virgin olive oil (EVOO) blends	Laboratory at 17 to 25 °C during 1 year, under two storage lighting conditions that tried to mimic real storage conditions of supermarkets: - dark - open shelves exposed to natural daylight and artificial light (14 h/day); Sampling every 3 months	Free acidity Peroxide value K_{232}, K_{270} and ΔK Rancimat stability Organoleptic characteristics E-tongue analysis	ANOVA LDA-SA	E-tongue coupled to LDA-SA approach showed satisfactory predictive potential to classify EVOO stored under different storage conditions: light/dark exposition and time. E-tongue device could monitor the storage time of olive oils in glass bottles and could be used to assess their freshness under the usual commercial light exposition conditions during the first year of storage.	**Rodrigues, Dias, Veloso, Pereira, and Peres (2017)**
Virgin olive oil (VOO)	Dark at 18 ºC	^{1}H NMR	ANOVA PCA	The sterol and alcohol derivatives are not stable during shelf life time. Cyclic compounds such as cyclopropene and cycloartenol do not evaporate during the period of shelf life time. The stability of aldehyde compounds is independent of shelf life but related to production of olive oil itself. Fatty acyls are stable during the shelf life time.	**Ün, and OK (2018)**
Monovarietal extra virgin olive oil (EVOO)	- 18 samples stayed in dark green glass bottles	UV-Vis spectra	MCR-ALS PCA	UV-Vis spectroscopy coupled with MCR-ALS is a feasible tool to monitor autoxidation processes in edible oils through storage time and	**Gonçalves *et al.* (2018)**

	- 6 samples stayed in tinplate cans 12 months at room temperature, darkness. Sampling at 30, 90, 180 and 365 days			in the monitoring of the EVOO quality in different package systems. Glass bottle system provides more protection for the autoxidation processes with the time for the EVOO.	
Olive oil	125 ml dark bottles in the dark at room temperature within two months	Peroxide value Rancimat stability Total phenols Total tocopherols CIELAB colour scale parameters Organoleptic characteristics E-tongue	MLR	Feasibility of using a potentiometric e-tongue, comprising non-specific lipid polymeric and cross-sensitive sensor membranes, coupled with chemometric tools, to predict key quality parameters of olive oils based on single-run assays. Multivariate linear models established allowed predicting peroxide value, oxidative stability, total phenols and tocopherols contents, CIELAB scale parameters, as well as 11 organoleptic positive attributes with satisfactory accuracy.	**Rodrigues *et al.* (2019)**
Virgin olive oil (VOO) varying widely in quality	Glass bottles closed in the presence of air, and stored at different temperatures (25 and 37 ºC), in darkness, using thermostatic incubators for	Fatty acid profile Diacylglycerols Tocopherols Total phenols UV absorption	ANN MLR PCA	All of the physical-chemical parameters changed during long-term storage resulting in VOO quality deterioration. Oxidative stability of samples was strongly correlated with the initial composition of the VOO samples. ANFIS was employed to predict the oxidative stability of VOO during storage as a function of time, storage temperature, and the rest	**Arabameri *et al.* (2019)**

	36 months (humidity 60%)			of chemical/physical parameters. The minor constituents of VOO strongly influenced the oxidative stability of the samples.	
Sunflower Canola Soy (mixed of new and 6-month production dates)	Dry and dark conditions at 24°C and relative humidity of 34-85% during 5 months Sampling every month.	Acidity value Peroxide value Anisidine value E-nose	CA LDA QDA SVM	Based on the e-nose and CA results, all the samples were classified into oxidized and non-oxidized oils which are in line with the AOCS method. SVM method possessed higher accuracy in the determination of the shelf life of the edible oils as compared with QDA and LDA methods.	**Karami, Rasekh, and Mirzaee-Ghaleh (2020)**
Soybean Rapeseed Corn Sunflower	Dark room at 20 ± 1°C during 8 months Sampling every month.	NIR spectrum	PCA k-NN RF SVM	NIR spectral analysis combined with appropriate chemometric methods is an efficient tool to realize the rapid and high-precision detection of edible oil in different storage periods. SVM model displayed excellent stability and generalization performance in the qualitative identification of the actual storage periods of edible oil.	**He, Jiang, and Chen (2020)**
Monovarietal virgin olive oil (VOO)	Transparent PET bottles. Simulated conditions of supermarket: - light intensity of 1000 lx in 12 h light/dark cycles, - temperature	Peroxide value Free fatty content K_{232} and K_{270} Organoleptic characteristics Phenol content α-Tocopherol content Pigments Fluorescence	PARAFAC PCA SLDA	Ability of fluorescence spectroscopy to monitor the chemical changes of VOO during storage and assessed the relationship with the different quality parameters. Excitation-emission fluorescence spectroscopy combined with PARAFAC model provides a degradation map of the oil. The sensory analysis of the samples	**Lobo-Prieto, Tena, Aparicio-Ruiz, García-González, and Sikorska (2020)**

	(16.3 - 29.7 ºC) - humidity (21 - 70%) for 21 months Sampling every month.	spectra		revealed that the changes of the PARAFAC components occurred at the same time as, or even before, the changes of the sensory characteristics.	
Virgin argan oil (VAO) from roasted and unroasted kernels	Laboratory at ambient conditions: - dark glass bottles - clear glass bottles Temperature varied between 12 ± 4 °C and 25 ± 5 °C, exposed (not directly) 12 h to natural daylight during 2 years Sampling 6-monthly.	Free acidity Peroxide value K_{270} and K_{232} UV absorption Fatty acid profile Tocopherols Sterol analysis Rancimat stability: (110 ºC) FTIR measurements	PCA PLS-DA PLS ANOVA	Shelf-life of VAO was assessed by both chemical properties and FTIR spectra. FTIR spectra and chemometrics form an interesting approach to survey the shelf life of VAO. FTIR spectra also showed good predictive ability for four physicochemical parameters. The roasting process seems to have a conservation effect, and the shelf stability of 'roasted' VAO is better than of 'unroasted'. The dark-glass bottles exhibited a good ability to conserve the oil quality.	**Kharbach *et al*. (2021)**
Virgin olive oil (VOO)	Glass UVA grade bottles on shelves. Light exposure was 500 lux for 12 h/day: - 20 EVOOs at 25 °C both in light and dark conditions during 12	Free acidity Peroxide value K_{232}, K_{270} and ΔK. Fatty acid profile α-Tocopherol Phenols Volatile profile ^{1}H NMR analysis	PCA PLS-DA PLS	Metabolic fingerprinting of VOO by ^{1}H NMR spectroscopy combined with chemometrics provides valuable approaches for the quality control of VOOs, as well as to evaluate their stability. Predictive models that disclosed the chemical compounds responsible for the VOO composition, which helps to understand the quality changes that experience during its	**Alonso-Salces *et al*. (2021)**

	months. Sampling monthly - 5 EVOOs at 25 °C, 30 °C and 35 °C during 24 months in the dark. Sampling every two months.	shelf life. Relatively small changes in the original ^{1}H signal intensities and the emergence of new low intensity signals indicated that some oxidative and hydrolytic degradation of VOO was occurring under ambient market conditions. Light and increasing temperature accelerate VOO degradation during shelf life period.

ANN: artificial neural network
ANOVA: analysis of variance
CA: cluster analysis
FT-MIR: Fourier-transform mid-infrared
^{1}H MNR: nuclear magnetic resonance of ^{1}H
k-NN: k-nearest neighbour
LDA: linear discriminant analysis
LDA-SA: linear discriminant analysis-simulated annealing
MCR-ALS: multivariate curve resolution-alternating least squares
MLR: multivariate least-squares regression
NIR: near-infrared
PCA: principal component analysis
PCR: principal component regression
PLS: partial least-squares regression
PLS-DA: partial least square-discriminant analysis
QDA: quadratic discriminant analysis
RF: random forest
SVM: support vector machine
SLDA: stepwise linear discriminant analysis
Vis-UV: visible-ultraviolet

Table 2 provides a comprehensive summary of oil type, ambient storage conditions, chemical-physical and sensory features, chemometric tool, and shelf life/stability conclusions reported during the last years.

2.3. Multivariate approach on the determination of stability of thermal oxidized oils: frying process

Frying is one of the most popular food processing methods to produce palatable and desirable foods with unique characteristics of flavour, odour and colour. Frying oil acts as a homogeneous heat transfer liquid medium and contributes to the development of texture and flavour of fried foods. During the frying process, due to the presence of oxygen, the moisture content of the food, the high oil temperature, and the leaching of components from the food, a variety of reactions occur. Hydrolysis of oil, oxidation and polymerisation are the main chemical reactions that usually take place in oils during this process (**Choe & Min, 2007**).

As the quality of fried foods is affected by that of the frying oil, regulations or guidelines have been established in many countries to guarantee high quality fried foods. The most widely-used limitation establishes that the used frying oil must be replaced when the total content of polar compounds, representative of the new compounds formed during frying, is around 25% expressed by weight of oil. Much analytical work has been done on the identification and quantification of the specific compounds formed to approach the key pending issues of intake estimations and safety assessments. Currently, analysis of used frying oils and fats is successfully achieved by chromatographic techniques that allow quantitative determination of the main groups of compounds formed or by rapid test correlated with standard methods (**Dobarganes & Márquez-Ruiz, 2013**).

Despite that, advanced techniques combined with chemometric tools have been applied to the analysis of stability and usage life of used frying oil. **Zhang *et al.* (2012)** simulated the frying process of corn oil, peanut oil, rapeseed oil and soybean oil. They acquired an FTIR spectrum of oils with different times of frying and divided it into 22 regions according to the different absorption peaks. Cluster analysis for calibration set and discriminant analysis for validation set was accomplished based on areas of these regions and got a reasonable vegetable oils authentication result.

Zribi *et al.* (2013) evaluated physicochemical behaviour of refined olive (ROO) and refined soybean oils during potatoes pan-frying simulating the household coking process. They determined standard quality and composition parameters, phenolic compounds, fatty acid content and induction time, and carried out

liquid chromatography analysis coupled to both UV spectroscopy and mass spectrometry. PCA was applied to characterize the effect of the enrichment of ROO and soybean oil with olive leaf hydroxytyrosol-rich hydrolysate extract after five successive pan-frying sessions according to all the parameters investigated. They concluded that such enrichment had shown remarkable resistance to oxidative deterioration and the formation of *trans*-fatty acids was found to be lower in the enriched refined oil samples.

Zribi *et al.* (2014) used ROO, corn, soybean, and sunflower oils as cooking oils for deep-frying at two different temperatures, 160 and 190 ºC, and for pan-frying of potatoes at 180 ºC for 10 successive sessions under the usual domestic practice. They determined standard quality and composition parameters, total polar compounds, polymeric triacylglycerols, anisidine value, fatty acid composition and Rancimat oxidative stability. PCA was applied to characterize the effect of the frying method and temperatures on the quality and stability characteristics and compositions of ROO, corn, soybean and sunflower oils before and after 10 successive frying sessions according to all of the parameters investigated without any rotation type (ANOVA was applied to confirm PCA results). Their results showed that ROO is the most stable oil during deep-frying at 160 ºC, whereas sunflower oils are the least stable during pan-frying at 180 ºC. Higher number of variability of samples it would be needed to support its conclusions.

Upadhyay, Sehwag, & Mishra (2017) aimed to develop frying stable sunflower oil blends stabilized with oleoresin rosemary and ascorbyl palmitate. They prepared different combinations of sunflower oil with both stabilizers at varied concentrations and PCA and hierarchical cluster analysis (HCA) were used to successfully classify sunflower oil blends into different groups, based on the alteration of compositional parameters and formation of polar compounds. An attempt was also made to explore the applicability of combined PCA and response surface methodology approaches for the optimization of the synergistic blend composition. Even though, the approach presented in their study was related to sunflower oil, methodology may be useful to develop newer blends of cooking oils with higher stability.

Dordevic *et al*. (2020) modelled the effect of oven heat treatment cycles on fatty acid composition and antioxidant profile of olive oil. Correlations between heating cycles and fatty acid profile changes were measured by Pearson's correlation analysis. The overall differences among independent variances were checked by PCA. Their results unambiguously emphasized how relatively small differences in heating temperatures result in different fatty acid profiles and, in consequence, different stability times.

Lastly, **Amit *et al.* (2020)** used Fourier-transform mid-infrared-attenuated total

reflection spectroscopy coupled to multivariate regression modelling to detect pure coconut oil adulteration with fried coconut oil. They reproduced the frying of potatoes in pure coconut oil at 180 ºC and oil was analysed within two weeks of frying. PCA was applied to analyse the differences among the samples. For qualitative analysis, the most informative variables obtained from the PCA loading spectra were then subjected to discriminant analysis for determining the possible outcome of a sample to pertain to a formerly determined cluster. Further, principal component regression (PCR) and PLS regression models were used to satisfactory develop calibration models for the quantification of adulterants.

3. CONCLUSION AND FUTURE TREND

Multivariate qualitative and quantitative methods are increasingly used in many fields (chemistry, process monitoring, food science, etc.). An important trend in shelf life evaluation is to develop rapid, simple, low-cost analytical tools that are suitable for online monitoring or quality control of the oil oxidation process. Although a wide number of studies dealing with food stability are available in the literature, they often do not allow shelf life data to be obtained due to a lack of a clearly defined acceptability limit.

Chemometric techniques and multivariate approaches used in shelf life and stability studies of edible both virgin and refined vegetable oils under forced and ambient conditions of storage have been detailed reviewed. Apart from the numerous opportunities and benefits of the multivariate approaches, these applications currently show also limitations. In particular, for official control purposes in food surveillance, analytical procedures must be validated and become legally incontestable. However, widely recognized guidelines for method validation concerning non-targeted analysis are not available yet, which certainly explains the currently still limited applicability for routine analysis and shelf life estimation.

Future trends will rely on the implementation of fingerprinting methods obtained by high informative techniques such as mass or nuclear magnetic resonance spectrometric fingerprints that provide additional information on real time oxidative evolution.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgments

This work was financially supported by the European Regional Development Fund (ERDF) and Ministry of Science and Technology within the framework of the funding corresponding to 'Retos-Colaboración' (Challenges-Collaboration Program) call of the State Program for Research, Development and Innovation Oriented to Challenges of the Society (RTC-2017-6170-2). In addition, SMR author acknowledges the PhD grant supported by the Ministry of Universities for the Training of University Teachers (FPU19/02078).

References

Aktar, T., & Adal, E. (2019). Determining the Arrhenius kinetics of avocado oil: oxidative stability under Rancimat test conditions. *Foods, 8,* 236; 1-13. https://doi.org/10.3390/foods8070236

Alonso Salces, R. M., Gallo, B., Collado, M. I., Sasía Arriba, A., Viacava, G. E., García González, D. L., Toschi, T., Servili, M., & Berrueta, L.Á. (2021). ^{1}H-NMR fingerprinting and supervised pattern recognition to evaluate the stability of virgin olive oil during storage. *Food Control, 123*, article 107831. https://doi.org/10.1016/j.foodcont.2020.107831

Amit, Jamwal, R., Kumari, S., Kelly, S., Cannavan, A., & Singh, D.K. (2020). Rapid detection of pure coconut oil adulteration with fried coconut oil using ATR-FTIR spectroscopy coupled with multivariate regression modeling. *LWT-Food Science and Technology, 125*, article 109250. https://doi.org/10.1016/j.lwt.2020.109250

Anzardi, M.B., Arancibia, J.A., & Olivieri A.C. (2021). Processing multi-way chromatographic data for analytical calibration, classification and discrimination: A successful marriage between separation science and chemometrics. *Trends in Analytical Chemistry*, *134*, article 116128. https://doi.org/10.1016/j.trac.2020.116128

Aparicio Ruiz, R., Roca, M., & Gandul Rojas, B. (2012). Mathematical model to predict the formation of pyropheophytin in a virgin olive oil during storage. *Journal of Agricultural and Food Chemistry, 60*, 7040−7049. https://doi.org/10.1021/jf3010965.

Aparicio, R., Roda, L., Albi, M.A., & Gutiérrez, F. (1999). Effect of various compounds on virgin olive oil stability measured by Rancimat. *Journal of Agricultural and Food Chemistry, 47*, 4150-4155. https://doi.org/10.1021/jf9812230

Arabameri, M., Nazari, R.R., Abdolshahi, A., Abdollahzadeh, M., Mirzamohammadi, S., Shariatifar, N., Barba, F.J., & Mousavi Khaneghah, A. (2019). Oxidative stability of virgin olive oil: evaluation and prediction with an adaptive neuro-fuzzy inference system (ANFIS). *Journal of the Science of Food and Agriculture, 99*, 5358-5367. https://doi.org/10.1002/jsfa.9777

Battimo, I., Savarese, M., Parisini, C., Malagrinò, G., Paduano, A., de Marco, E., Della Medaglia, D.A., Terminiello, R., Ambrosino, M.L., & Sacchi, R. (2007). Rancidity evaluation and shelf-life monitoring of virgin olive oil by electronic-nose. *Italian Journal of Food Science, 19,* 278-287.

Berrueta, L.A., Alonso-Salces, R.M., & Héberger, K. (2007). Supervised pattern recognition in food analysis. *Journal of Chromatography A, 1158*, 196-214. https://doi.org/10.1016/j.chroma.2007.05.024

Bower, J.A. (Ed.). (2013). *Statistical methods for food science* (2nd ed.). Oxford: Wiley Blackwell.

Bro, R., & Smilde, A.K. (2014). Principal component analysis. *Analytical Methods*, *6*, 2812-2831. https://doi.org/10.1039/c3ay41907j

Buratti, S., Malegori, C., Benedetti, S., Oliveri, P., & Giovanelli, G. (2018). E-nose, e-tongue and e-eye for edible olive oil characterization and shelf life assessment: a powerful data fusion approach. *Talanta, 182*, 131-141. https://doi.org/10.1016/j.talanta.2018.01.096

Calligaris, S., Manzocco, L., Anese, M., & Nicoli, M.C. (2015). Shelf life assessment of food undergoing oxidation - A review. *Critical Reviews in Food Science and Nutrition,* 1903-1912.
http://dx.doi.org/10.1080/10408398.2013.807222

Cao, J., Li, C., Liu, R., Liu, X.R., Fan, Y., & Deng, Z.Y. (2017). Combined application of fluorescence spectroscopy and chemometrics analysis in oxidative deterioration of edible oils. *Food Analytical Methods, 10,* 649-658.
https://doi.org/10.1007/s12161-016-0587-2

Chaudry, M.M.A., Amodio, M.L., Babellahi, F., De Chiara, M.L.V., Amigo Rubio, J.M., & Colelli, G. (2018). Hyperspectral imaging and multivariate accelerated shelf life testing (MASLT) approach for determining shelf life of rocket leaves. *Journal of Food Engineering, 238*, 122-133.
https://doi.org/10.1016/j.jfoodeng.2018.06.017

Choe, E., & Min, D.B. (2007). Chemistry of deep-fat frying oils. *Journal of Food Science*, 72, 5, 77-86.
https://doi.org/10.1111/j.1750-3841.2007.00352.x

Cordella, C.B.Y., Tekye, T., Rutledge D.N., & Leardi, R. (2012). A multiway chemometric and kinetic study for evaluating the thermal stability of edible oils by ^{1}H NMR analysis: comparison of methods. *Talanta, 88,* 358-368.
https://doi.org/10.1016/j.talanta.2011.11.001

De Juan, A., Jaumot, J., & Tauler, R. (2014). Multivariate curve resolution (MCR) – Solving the mixture analysis problem. *Analytical Methods*, *6*, 4964-4974.
https://doi.org/10.1039/c4ay00571f

Dobarganes, C., & Márquez-Ruiz, G. (2013). Analysis of used frying oils. Lipid *Technology, 25*, 159-162.
https://doi.org/10.1002/lite.201300284

Dordevic, D., Kushkevych, I., Jancikova, S., Zeljkovic, S.C., Zdarsky, M., & Hodulova, L. (2020). Modeling the effect of heat treatment on fatty acid composition in home-made olive oil preparations. *Open Life Sciences, 15*, 606-618.
https://doi.org/10.1515/biol-2020-0064

Drab, K., & Daszykowski, M. (2014). Clustering in analytical chemistry. *Journal of AOAC International*, *97*, 29-38.
https://doi.org/10.5740/jaoacint.SGEDrab

Ennouri, K., Hlima, H.B., Ayed, R.B., Braïek, O.B., Mazzarello, M., Ottaviani, E., Mallouli, L., & Smaoui, S. (2019). Assessment of Tunisian virgin olive oils via synchronized analysis of sterols, phenolic acids, and fatty acids in combination with multivariate chemometrics. *European Food Research and Technology, 245*, 1811-1824.
https://doi.org/10.1007/s00217-019-03303-2

European Commission (2011, consolidated version 2019). Regulation No. 1169/2011 of the European Parliament and of the Council on the provision of food information to consumers. Official Journal of European Communities, 02011R1169, 1-60.

Evangelisti, F., Zunin, P., Tiscornia, E., Petacchi, R., Drava, G., & Lanteri, S. (1997). Stability to oxidation of virgin olive oils as related to olive conditions: Study of polar compounds by chemometric methods. *Journal of the American Oil Chemists' Society, 74*, 1017-1023.
https://doi.org/10.1007/s11746-997-0019-y

Gonçalves, R.P., Março, P.H., & Valderrama, P. (2014). Thermal edible oil evaluation by UV–Vis spectroscopy and chemometrics. *Food Chemistry, 163,* 83-86.
http://dx.doi.org/10.1016/j.foodchem.2014.04.109

Gonçalves, T., Rosa, L., Gonçalves, R., Torquato, A., Março, P.H., Gomes, S.T., Matsushita, M., & Valderrama, P. (2018). Monitoring the oxidative stability of monovarietal extra virgin olive oils by UV–Vis spectroscopy and MCR-ALS. *Food Analytical Methods, 11*, 1936-1943. https://doi.org/10.1007/s12161-018-1149-6

Gower J.C., & Krzanowski W.J. (1999). Analysis of distance for structured multivariate data and extensions to multivariate analysis of variance. *Journal of the Royal Statistical Society: Series C (Applied Statistics),* 48, 505–519. https://doi.org/10.1111/1467-9876.00168

Grauwet, T., Vervoort, L., Colle, I., Van Loey, A., & Hendrickx, M. (2014). From fingerprinting to kinetics in evaluating food quality changes cell press. *Trends in Biotechnology, 32*, 125-131. http://dx.doi.org/10.1016/j.tibtech.2014.01.002

Guillaume, C., & Ravetti, L. (2016). Shelf-life prediction of extra virgin olive oils using an empirical model based on standard quality tests. *Journal of Chemistry,* article 6393962, (1-7). https://doi.org/10.1155/2016/6393962.

Guimaräes-Inácio, A., Lopes Francisco, C.R., Maeda Rojas, V., de Souza Leone, R., Valderrama, P., Bona, E., Leimann, F.V., Coelho Tanamati, A.A., & Gonçalves, O.H. (2018). Evaluation of the oxidative stability of chia oil-loaded microparticles by thermal, spectroscopic and chemometric methods. *LWT-Food Science and Technology, 87*, 498-506. https://doi.org/10.1016/j.lwt.2017.09.031

Gunstone, F.D., Harwood, J.L., & Dijkstra, A.J. (Eds.). (2007). *The lipid handbook* (3rd ed.). New York: CRC Press.

He, Y., Jiang, H., & Chen, Q. (2020). High-precision identification of the actual storage periods of edible oil by FT-NIR spectroscopy combined with chemometric methods. *Analytical Methods, 12*, 3722-3728. https://doi.org/10.1039/d0ay00779j

Hibbert, D.B. (2016). Vocabulary of concepts and terms in chemometrics (IUPAC Recommendations). *Pure and Applied Chemistry, 88*, 407-443. https://doi.org/10.1515/pac-2015-0605

IFST (1993). Shelf life of foods: guidelines for its determination and prediction. London: Institute of Food Science and Technology.

Jiménez Carvelo, A.M., & Cuadros Rodríguez, L. (2021). Data mining/machine learning methods in foodomics. *Current Opinion in Food Science, 37*, 76-82. https://doi.org/10.1016/j.cofs.2020.09.008

Jiménez Carvelo, A.M., González Casado, A., Bagur González, M.G. & Cuadros Rodríguez, L. (2019). Alternative data mining/machine learning methods for the analytical evaluation of food quality and authenticity – A review. *Food Research International*, *122,* 25-39. https://doi.org/10.1016/j.foodres.2019.03.063

Jiménez Carvelo, A.M., Martin Torres, S., Cuadros Rodríguez, L., & González Casado, A. (2021). Nontargeted fingerprinting approaches. In C.M. Galanakis (Ed.), *Food authentication and traceability* (pp. 163-193). London: Academic Press / Elsevier.

Karami, H., Rasekh, M., & Mirzaee–Ghaleh, E. (2020). Comparison of chemometrics and AOCS official methods for predicting the shelf life of edible oil. *Chemometrics and Intelligent Laboratory Systems, 206*, 104165. https://doi.org/10.1016/j.chemolab.2020.104165

Kebede, B.T., Grauwt, T., Magpusao, J., Palmers, S., Michiels, C., Hendrickc, M., & Van Loey, A. (2015). Chemical changes of thermally sterilized broccoli puree during shelf-life: investigation of the volatile fraction by fingerprinting-kinetics. *Food Research International, 67*, 264-271. http://dx.doi.org/10.1016/j.foodres.2014.10.017

Kharbach, M., Marmouzi, I., Kamal, R., Yu, H., Barra, I., Cherrah, Y., Alaoui, K., Heyden, Y.V., & Bouklouze, A. (2021). Extra virgin argan oils' shelf-life monitoring and prediction based on chemical properties or FTIR fingerprints and chemometrics. *Food Control, 121*, article 107607. https://doi.org/10.1016/j.foodcont.2020.107607

Le Dréau, Y., Dupuy, N., Artaud, J., Ollivier, D., & Kister, J. (2009). Infrared study of aging of edible oils by oxidative spectroscopic index and MCR-ALS chemometric method. *Talanta, 77*, 1748-1756. https://doi.org/10.1016/j.talanta.2008.10.012

Lee, L.C., Liong, C.Y., & Jemain, A.A. (2018). Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps. *Analyst*, *143*, 3526-3539. https://doi.org/10.1039/c8an00599k

Li Vigni, M., Durante, C., & Cocchi, M. (2013). Exploratory data analysis. In F, Marini (Ed.), *Chemometrics in food chemistry* (pp. 55-126). Amsterdam: Elsevier.

Li, X., & Wang, S.C. (2018). Shelf life of extra virgin olive oil and its prediction models. *Journal of Food Quality*, article 1936260 (1-15). https://doi.org/10.1155/2018/1639260

Lobo Prieto, A., Tena, N., Aparicio Ruiz, R., García González, D. L., & Sikorska, E. (2020). Monitoring virgin olive oil shelf-life by fluorescence spectroscopy and sensory characteristics: a multidimensional study carried out under simulated market conditions. *Foods, 9*, article 1846. https://doi.org/10.3390/foods9121846

Logan, A., Nienaber, U., & Pan, X.S. (2013). *Lipid Oxidation Challenges in Food Systems*. Urbana, Illinois: AOCS Press.

Maggio, R.M, Valli, E., Bendini, A., Gómez Caravaca, A.M., Toschi, T.G., & Cerretani, L. (2011). A spectroscopic and chemometric study of virgin olive oils subjected to thermal stress. *Food Chemistry, 127*, 216-221. https://doi.org/10.1016/j.foodchem.2010.12.01

Marangoni, A.G. (2017). *Kinetic analysis food systems*. Cham: Springer.

Mildner Szkudlarz, S., Jelen, H.H., & Zawirska Wojtasiak R. (2008). The use of electronic and human nose for monitoring rapeseed oil autoxidation. *European Journal of Lipid Science and Technology, 110*, 61-72. http://dx.doi.org/10.1002/ejlt.200700009

Nicoli, M.C. (Ed.). (2012). *Shelf life assessment of food*. Boca Raton: CRC Press.

Olivieri, A.C. (2012). Recent advances in analytical calibration with multi-way data. *Analytical Methods*, *4*, 1876-1886. https://doi.org/10.1039/c2ay25064k

Oueslati, I., Manaï, H., Madrigal Martínez, M., Martínez Cañas, M.A., Sánchez Casas, J., Zarrouk, M., & Flamini G. (2018). Volatile molecular markers of VOO thermo-oxidation: effect of heating processes, macronutrients composition, and olive ripeness on the new emitted aldehydic compounds. *Food Research International, 106*, 654–665. https://doi.org/10.1016/j.foodres.2018.01.029.

Pagliarini, E., Zanoni, B., & Giovanelli, G. (2000). Predictive study on Tuscan extra virgin olive oil stability under several commercial conditions. *Journal of Agricultural and Food Chemistry, 48,* 1345-1351. https://doi.org/10.1021/jf990372v

Pedro, A.M.K., & Ferreira, M.M.C. (2006). Multivariate accelerated shelf-life testing: a novel approach for determining the shelf-life of foods. *Journal of Chemometrics*, 20, 76-83. https://doi.org/10.1002/cem.995

Piergiovanni, L., & Limbo, S. (2019). Food shelf-life models. In R. Accorsi, R. Manzini (Eds.), *Sustainable Food Supply Chains* (pp. 49-60). London: Academic Press. https://doi.org/10.1016/C2016-0-05053-2

Poyato, C., Ansorena, D., Navarro-Blasco, I., & Astiasarán, I. (2014). A novel approach to monitor the oxidation process of different types of heated oils by using chemometric tools. *Food Research International, 57*, 152-161. http://dx.doi.org/10.1016/j.foodres.2014.01.033

Psomiadou, E., & Tsimidou, M. (2002). Stability of virgin olive oil. 1. Autoxidation studies. *Journal of Agricultural and Food Chemistry, 50,* 716-721. https://doi.org/10.1021/jf0108462

Redondo Cuevas, L., Castellano, G., Torrens, F., & Raikos, V. (2018). Revealing the relationship between vegetable oil composition and oxidative stability: a multifactorial approach. *Journal of Food Composition and Analysis, 66,* 221-229. https://doi.org/10.1016/j.jfca.2017.12.027

Rodrigues, N., Dias, L.G., Veloso, A.C., Pereira, J.A., & Peres, A.M. (2016). Monitoring olive oils quality and oxidative resistance during storage using an electronic tongue. *LWT – Food Science and Technology, 73*, 683-692. https://doi.org/10.1016/j.lwt.2016.07.002

Rodrigues, N., Dias, L.G., Veloso, A.C., Pereira, J.A., & Peres, A.M. (2017). Evaluation of extra-virgin olive oils shelf life using an electronic tongue-chemometric approach. *European Food Research and Technology, 243*, 597-607. https://doi.org/10.1007/s00217-016-2773-2

Rodrigues, N., Marx, Í.M., Casal, S., Dias, L.G., Veloso, A.C., Pereira, J.A., & Peres, A.M. (2019). Application of an electronic tongue as a single-run tool for olive oils' physicochemical and sensory simultaneous assessment. *Talanta, 197*, 363-373. https://doi.org/10.1016/j.talanta.2019.01.055

Roszkowska, B., Tariska, M., Czaplicki, S., & Konopka, I. (2015). Variation in the composition and oxidative stability of commercial rapeseed oils during their shelf life. *European Journal of Lipid Science and Technology, 117*, 673-683. https://doi.org/10.1002/ejlt.201400271

Savarese, M., Caporaso, N., Parisini, C., Paduano, A., De Marco, E., & Sacchi, R. (2013). *Application of an electronic nose for the evaluation of rancidity and shelf life in virgin olive oil*. In Electronic International Interdisciplinary Conference, 12, 361-366.

Selaimia, R., Oumeddour, R., & Nigri, S. (2017). The chemometrics approach applied to FTIR spectral data for the oxidation study of Algerian extra virgin olive oil. *International Food Research Journal 24*, 1301-1307.

Shahidi, F., & Zhong, Y. (2005). Lipid oxidation: measurement methods. In F. Shahidi (Ed.), *Bailey's Industrial Oil and Fat Products* (6th ed., pp. 357-385). Canada: John Wiley & Sons, Inc.

Song, Y., Liang, J., Lu, J., & Zhao, X. (2017). An efficient instance selection algorithm for k nearest neighbor regression. *Nuerocomputing, 251*, 26-34. https://doi.org/10.1016/j.neucom.2017.04.018

Syed, A. (2016). Oxidative stability and shelf life of vegetable oils. In M. Hu, C. Jacobsen (Eds.), *Oxidative Stability and Shelf Life of Foods Containing Oils and Fats* (pp. 187-207). London: Academic Press.

Ün, İ., & Salim, O.K. (2018). Analysis of olive oil for authentication and shelf life determination. *Journal of food science and technology, 55*, 2476-2487. https://doi.org/10.1007/s13197-018-3165-3

Upadhyay, R., & Mishra, H.N. (2015a). Multivariate analysis for kinetic modeling of oxidative stability and shelf life estimation of sunflower oil blended with sage (Salvia officinalis) extract under rancimat conditions. *Food Bioprocess Technology, 8,* 801-810. http://dx.doi.org/ 10.1007/s11947-014-1446-z

Upadhyay, R., & Mishra, H.N. (2015b). Predictive modeling for shelf life estimation of sunflower oil blended with oleoresin rosemary (Rosmarinus officinalis L.) and ascorbyl palmitate at low and high temperatures*. LWT-Food Science and Technology, 60*, 42-49. http://dx.doi.org/10.1016/j.lwt.2014.09.029

Upadhyay, R., Sehwag, S., & Mishra, H.N. (2017). Chemometric approach to develop frying stable sunflower oil blends stabilized with oleoresin rosemary and ascorbyl palmitate. *Food Chemistry, 218*, 496-504. http://dx.doi.org/10.1016/j.foodchem.2016.09.105

Wójcicki, K., Khmelinskii, I., Sikorski, M., & Sikorska, E. (2015). Near and mid infrared spectroscopy and multivariate data analysis in studies of oxidation of edible oils. *Food Chemistry, 187*, 416-423. http://dx.doi.org/10.1016/j.foodchem.2015.04.046

Xu, L., Yu, X., Liu, L., & Zhang R. (2016). A novel method for qualitative analysis of edible oil oxidation using an electronic nose. *Food Chemistry, 202*, 229-235. http://dx.doi.org/10.1016/j.foodchem.2016.01.144

Zanoni, B., Bertuccioli, M., Rovellini, P., Marotta, F., & Mattei, A. (2005). A preliminary approach to predictive modelling of extra virgin olive oil stability. *Journal of the Science of Food and Agriculture, 85*, 1492-1498. https://doi.org/10.1002/jsfa.2135

Zhang, Q., Liu, C., Sun, Z., Hu, X., Shen, Q., & Wu, J. (2012). Authentication of edible vegetable oils adulterated with used frying oil by Fourier Transform Infrared Spectroscopy. *Food Chemistry, 132*, 1607-1613. https://doi.org/10.1016/j.foodchem.2011.11.129

Zribi, A., Gargouri, B., Jabeur, H., Rebaï, A., & Bouaziz M. (2013). Enrichment of pan-frying refined oils with olive leaf phenolic-rich extract to extend the usage life. *European Journal of Lipid Science and Technology, 115*, 1443-1453. https://doi.org/10.1002/ejlt.201300037

Zribi, A., Jabeur, H., Aladedunye, F., Rebaï, A., Matthäus, B., & Bouaziz M. (2014). Monitoring of quality and stability characteristics and fatty acid compositions of refined olive and seed oils during repeated pan- and deep- frying using GC, FT-NIRS, and chemometrics. *Journal of Agricultural and Food Chemistry, 62*, 10357-10367. https://doi.org/10.1021/jf503146f

PUBLICACIÓN VII.

ARTÍCULO CIENTÍFICO

Multivariate stability monitoring and shelf life models of deterioration of vegetable oils under real time ageing conditions – Part 1: Extra virgin olive oil as a main case of study

EVOO multivariate stability monitoring and shelf life modelling

Food Packaging and Shelf Life, (under review)

Multivariate stability monitoring and shelf life models of deterioration of vegetable oils under real time ageing conditions – Part 1: Extra virgin olive oil as a main case of study

EVOO multivariate stability monitoring and shelf life modelling

Sandra MARTÍN-TORRES[1,✉], Juan Antonio TELLO-JIMÉNEZ[2], Rafael LÓPEZ-BLANCO[2], Antonio GONZÁLEZ-CASADO[1], Luis CUADROS-RODRÍGUEZ[1]

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

[2] *Laboratory 'J.A. Tello', S.L., Industrial Park "Los Olivares", C/ La Iruela, No 8, E-23009 Jaén, Spain.*

Highlights

- Monitoring of EVOO quality depletion under standard trade storage conditions
- PCA scores are used for modelling the EVOO oxidation kinetic
- PLS scores and loading vector are used to predict current shelf life of EVOO
- Multivariate acceptability limit includes regulatory and non-regulatory parameters

ABSTRACT

Modelling quality depletion of oils and establishing their shelf life (or ageing rate) remains a challenge for food industry. A persistent issue in the control of lipid oxidation is deciding which oxidation products are the most suitable to be monitored. Several attempts have been done in this regard, however, the number of parameters that can be considered to assess oil shelf life is excessively wide and the proposed univariate models have proven to be extremely dependent on experimental conditions. For this reason, the methodology for carrying out a multivariate kinetic model is presented, combining physico-chemical and sensory parameters from experimental analytical determinations, and chemometric data processing tools. The main objective is to develop a multivariate shelf life model that allow predicting the number of months in which an extra virgin olive oil complies with the requirements of its commercial category when ageing takes place under standard trading conditions.

Key words

Multivariate deterioration monitoring, multiparametric-kinetic studies, chemometrics, shelf life, quality and stability indices, extra virgin olive oil.

1. INTRODUCTION

Edible olive oil is a food commodity marketed under different designations referring to different categories, being the extra virgin olive oil (EVOO) the highest quality which is determined by threshold values of physico-chemical and sensory characteristics which in the latter case requires that must be free of defective organoleptic attributes (**European Commission, 2019**). In addition, EVOO, defined as a juice obtained from the fruit of the olive tree (*Olea europea L.*) which has not undergone any treatment other than milling, washing, decanting, centrifuging and filtering, is found at the base of the 'Mediterranean Diet' pyramid and has proven considerable beneficial effects on human health (**Sánchez Quesada et al., 2020**).

However, sensory characteristics and the most of physico-chemical properties of EVOO are not stable through the time. Among the deterioration processes that EVOO could undergo from the moment it is produced, they can be grouped into three general mechanisms: (i) enzymatic oxidation, (ii) photo-oxidation, (iii) autoxidation; favoured by the incidence of various factors, all of which oxygen, temperature and light are the most important (**Talbot, 2011; Morales, & Przybylski, 2013**). Once EVOO is extracted, filtered and packaged, or stored in suitable containers, and provided that it is kept away from direct sunlight and heat, the main alteration is due to autoxidation, also known as rancidity. During the autoxidation reaction, a number of new compounds are formed, while minor components are degraded, causing off-flavours, loss of nutritional value and finally consumer rejection.

The concept of stability of vegetable edible oils is generally accepted as the shelf life of the product until rancidity becomes evident (**Moschopoulou, Moatsou, & Drosinos, 2019**). Current legislation demands that producers must indicate a best-before date on the label of EVOO, which is usually set at around 18 to 24 months by agreement (**Government of Spain, 2021**). After this period, an EVOO could be consumed with full health guarantee, as it is not a perishable product, but its sensory attributes and healthy attributes are diluted over time. Therefore, it is easy to find olive oils in the market labelled as a higher-quality product than it really is. These edible oils would be engaging in non-conformity food fraud. Consumers increasingly appreciate food quality and, consequently, it is necessary to provide advance analytical information, with scientific support, for the quality characterization of oils and other high fat food products.

One question that persists in lipid oxidation control is deciding which oxidation products are the most suitable to be monitored. There are a lot of advances and research papers in this topic (**Pignitter, & Somoza, 2012**), however, the number of parameters that could be considered for assessing the shelf life of EVOO is excessively wide: total phenol content (**Psomiadou, & Tsimidou, 2002**), pyropheophytins trend (**Aparicio, Roca, & Gandul Rojas, 2012**),

changes and evolving of the fatty acid composition, volatile compounds, ultraviolet absorbance or peroxide value (**Guillaume, & Ravetti, 2016**), among others; so that, in practice, it is not feasible acquire a clear idea of their behaviour by examining separately the evolution of each one.

In recent years, relying on multiparametric kinetic studies and modelling approaches by applying specific data mining methods, which formally referred to as 'chemofoodmetrics', progress has been made in the understanding of complex food quality-related reactions such as EVOO. Formally, 'chemofoodmetrics' could be defined as the science that relates measurements of chemical or physico-chemical parameters to the quality features of a food by applying mathematical or statistical methods. Multiparametric kinetic and shelf life models are often more generically applicable than single-parameter models and could be more easily applied to other products or processes (**Grauwet, Vervoort, Colle, Van Loey, & Hendrickx, 2014**). A detailed review of multivariate approaches applied on vegetable oil stability studies monitoring wide selection of parameters have been recently reported (**Martín Torres, Ruiz Castro, Jiménez Carvelo, & Cuadros Rodríguez, 2022**). Although a wide number of studies dealing with EVOO stability are available in the literature, both single-parameter monitoring models and multiparametric models, they often do not allow shelf life data to be obtained due to lack of a clearly defined acceptability limit.

For this reason, this paper deals with explanation and discussion of fundamental and framework for carrying out a multivariate shelf life model that allows an estimation of the shelf life of the vegetable oil under consideration. EVOO oxidation kinetics and shelf life results, when oxidised under standard storage conditions simulating supermarket shelves, are reported.

2. METHODOLOGY

2.1. EVOO samples

83 freshly produced and filtered EVOO samples were supplied by different producers during 2019/2020 harvest season just after milling, so that its ageing time could be initially considered as zero. Several olive varieties, degree of ripeness and geographical regions (mainly from Spain, but also Morocco and Portugal) were considered in the aim of being representative of the product variability found on the market.

2.2. Storage

As mentioned above, oxygen, temperature and exposure to light are the main causes of the oxidation of vegetable oils. So that, the control of storage conditions for shelf life studies is a critical step. Two different strategies can be pursued: (i) actual shelf life testing, that mimics the expected real storage conditions carefully chosen to realistically simulate storage and which must be kept constant until the end of the study; (ii) accelerated shelf life testing, under forced light and temperature conditions capable of increasing the rate at which food quality is lost.

In this study, seven batches were considered, consisting of the 83 different EVOOs, which were packaged in transparent food-grade PET 60 mL bottles (labelled from A to G) and were stored in a temperature-controlled room (20 ± 5 ºC) exposed to 12 hours of LED light, simulating supermarket store conditions during 14 months. The first lot (A), it was analysed at the beginning of the study, consider as zero month of ageing. After 2 months of storage, lot (B) was analysed. Consecutively, every two months, an aliquot of each oil sample was analysed (up to lot (G) after 14 months of ageing) for physico-chemical and sensory parameters.

2.3. Analytical equipment

Refractive indexes were obtained using an Abbe refractometer ORT1RS. Spectroscopic UV absorption values were measured on Genesys 10SUV-Vis while anisidine value was determined using Agilent 8453 spectrophotometers. Oxidative stability was performed on Methrom 892 Professional Rancimat. Varian 3800 gas chromatographer equipped with a flame ionisation detector (Varian 450 GC) and an RXI-5HT capillary column was used to the quantification of the relative amounts of 1,2- and 1,3-diglycerides. Same equipment with Agilent HP-INNOWAX column was used to acquire the volatile compound fingerprints. The analytical determination of tocopherols was performed on Agilent 1100 Series liquid chromatograph equipped with a G1321A fluorescence detector using Ultrabase Sil column. Zorbax eclipse plus C18 and Agilent 1260 multiple wavelength detector were used for pyropheophytin a content measurement. Polar (phenolic) compounds quantification was performed on Agilent 1260 equipped with G7111B diode array detector.

2.4. Testing

Once samples and storage conditions are selected, is then necessary to carefully chosen quality indicators to monitor the changes as a function of time. It is advisable to considerer primary and secondary oxidation analytical

parameters to have a clearer picture of the on-going oxidation situation together with sensory evaluation in order to identify off-flavours that are easily appreciable by the consumer's senses.

In order to carry out a multivariate-based estimation of the EVOO shelf life, the following parameters were monitored.

2.4.1. Physico-chemical analyses

Refractive index was determined according to ISO 280 standard (**ISO 280, 1998**). Spectroscopic UV absorptivity values, K_{232}, K_{270} and ΔK were determined in compliance with COI/T.20/Doc. No 19 standard (**COI/T.20/Doc. No 19/ Rev.5, 2019**). The oxidative stability was measured by using the Rancimat method at 120 ºC as described in ISO ^{6886} standard (**ISO 6886, 2016**). Both peroxide and anisidine values were determined according to the recognised methods described in COI/T.20/Doc. No 35 (**COI/T.20/Doc. No 35/ Rev.1, 2017**) and ISO 6885 standards (**ISO 6885, 2016**) respectively.

Chromatographic profiling. The quantification of the relative amounts of 1,2- and 1,3-diglycerides was performed by gas chromatography as described in the ISO 29822 standard (**ISO 29822, 2009**). Total, α-, β-, γ- y δ-tocopherols were quantified after high performance liquid chromatography (HPLC) and subsequent fluorescence detection according to ISO 9936 standard (**ISO 9936, 2016**). 'Pyropheophytin a' content was determined in accordance with ISO 29841 standard (**ISO 29841, 2009**).

The extraction, separation and subsequent detection of polar (phenolic) compounds by HPLC-DAD were carried out following the principles outlined in COI/T.20/Doc No29 method (**COI/T.20/Doc. No 29, 2009**) with slight modifications (previously published) (**Cuadros Rodríguez et al., 2021**). Ratio of areas of phenolic alcohols, phenolic acids, secoiridoids, lignans and flavonoids, referred to the total biophenols area, was expressed as a percentage.

Chromatographic fingerprinting. Volatile compounds were characterized by headspace solid-phase microextraction followed by GC-FID analysis. A more detailed explanation of both sample preparation and chromatographic conditions are described in a previous paper (**Ortega Gavilán, Valverde Son, Rodríguez García, Cuadros Rodríguez, & Bagur González, 2020**). After volatile fingerprints acquisition, chromatograms were exported from the instrument software and embedded in a vector of data, and further pre-processed: (i) selection of region of interest; (ii) alignment using an average chromatogram as reference; and (iii) normalization of intensities with respect to the internal standard. Each data vector containing the related chromatographic fingerprint of an EVOO sample consisted of 8,391 elements, each of which is a

single variable indicating a normalised signal intensity. When fingerprinting methodology is applied, the compounds referred to are neither identified nor quantified in the conventional way. It is therefore necessary to apply a process of selection of the variables that best define the evolution of the oil. This issue will be suitably explained later.

2.4.2 Sensory Analyses

EVOO samples were sensory evaluated of rancidity. Rancidity value was agreed by a panel of 2-3 trained and experienced tasters in 'open-tasting' sessions, following the procedure described in COI/T.20/Doc No15 (**COI/T.20/Doc. No 15/Rev. 10, 2018**). It should be noted that the tasters are part of a sensory panel authorised by the IOC and accredited by the ISO 17025 standard.

2.5. Kinetic modelling and shelf life estimation

The availability of large data set does not mean at the immediate time availability of information promptly accessible. A properly use of chemometric tools that allow the interpretation of the hidden and non-evident information embodied within the data is necessary. Two approaches could mainly be differentiated regarding multivariate data analysis of foodstuff stability data: unsupervised and supervised multivariate pattern recognition methods.

Principal component analysis (PCA) is aimed at finding the simplest mathematical model able to describe the data set satisfactorily. PCA looks for a smaller number of underlying factors, named principal components (PCs), that explain most of the variability exhibited by the larger number of measurements made on the objects/samples. It is an unsupervised method because it does not require training input to find the output: no additional knowledge (e.g., y-variable) besides raw data (x-variable) is needed to describe the data set. Models work on its own to improve understanding and accessibility of the intrinsic features of the data. PCA model was built in order to detect the presence or absence of outlier samples and to evaluate relationships between samples and variables, and between variable themselves. After that, PCs scores were used to model kinetic of EVOO oxidation.

Partial least-squares regression analysis (PLS) is aimed at detecting cause-effect relationships. PLS is a supervised multivariate method because, apart from the information on the x-variables measured, the available knowledge on a dependent response y-variable (usually, storage time) is applied to obtain a latent variable (LV) model that optimally describes the response variable. PLS computes a partial least-squares regression model to predict a dependent y-

variable from a set of independent x-variables. PLS regression was used to evaluate importance of individual variables in the quality depletion of EVOO using loading plot of the variable's importance in projection (VIP) scores in the PLS model. The VIP formulation as originally proposed is a parameter varying in a fixed range since the sum of squared VIP for all variables sum to the number of variables. A variable with a VIP value close to or greater than one can be considered important in a given model. Variables with VIP values significantly less than one are less important and might be good candidates for exclusion from the model. It works for x-variables with n-dimensions and y-variables up to two-dimensions and assumes that the samples are in the first model (**Cocchi, Biancolillo, & Marini, 2018**). After that, LV scores were used to obtain a multivariate equation of EVOO quality depletion and, therefore, to evaluate shelf life time.

PCA may be more responsive to other sources of variation whereas a PLS prediction model based on ageing months will in general behave more robustly as PLS only takes into account the covariance between spectral profiles and predicted values and which may provide the advantage of being more flexible for future validations.

Acceptability limit. Shelf-life studies require the identification of an acceptability limit; this is what makes it different from a stability study. This is a very hard decision to make when selecting the cut-off value of a critical attribute, *i.e.*, the attribute that has the highest impact on the quality depletion of the oil, or shows the most change over the shortest time period. Normally it is referred to legal or regulatory requirements. When multivariate shelf life model is conducted, acceptability limit vector includes all the attributes that show change over time and gives rise to a single scalar coefficient, Qc, which is traduced to shelf life time.

Once the PLS regression model of shelf life is built, loading matrix is necessary to calculate Q_C, defined as follows:

$$Q_C = \max(\mathbf{Q_A} \cdot \mathbf{L})$$

where $\mathbf{Q_A}$ is the autoscaled acceptability limit vector and **L** the loading matrix of the time related latent variable of the model. Q_C is then interpolated into shelf life model equation to obtain cut off criteria, t_C, or shelf life time.

3. RESULT AND DISCUSSION

3.1. Experimental data matrix of ageing

Table 1. EVOO physico-chemical and sensory parameters being monitored in this study.

No	Parameter	Brief description
1.	Peroxide value	Peroxide content expressed in terms of milliequivalents of active oxygen per kilogram of EVOO oil
2.	K_{270}	The specific extinctions are calculated for a concentration of 1% (m/V) in a 10 mm cell (absorbance·(g/100mL)$^{-1}$·cm$^{-1)}$
3.	K_{232}	
4.	ΔK	
5.	Refractive index	Expressed at 20 °C (as reference)
6.	Rancidity (sensory)	Median of the negative attribute rancidity given by a panel of expert tasters
7.	Oxidative stability (Rancimat)	Induction period (in hours) at 120ºC
8.	Anisidine value	Rate of increase of absorbance, at 350 nm in a 10 mm cell, when reacted with p-anisidine under specific conditions (no units)
9.	1,2-Diglycerides	Percent mass fraction of the peak areas of all 1,2-diacylglycerols relative to the sum of the peaks of all diacylglycerols, determined by gas chromatography
10.	1,3-Diglycerides	
11.	Total tocopherols	Absolute content of each type of tocopherol expressed in milligrams per kilogram of oil mg/kg oil, determined by liquid chromatography
12.	α-Tocopherol	
13.	β-Tocopherol	
14.	γ-Tocopherol	
15.	δ-Tocopherol	
16.	Total biophenols	Milligrams of total biophenol content per kilogram of oil, determined by liquid chromatography
17.	Phenolic alcohols	Ratio of areas of the different biophenol families to the total biophenol area, expressed in percentages
18.	Phenolic acids	
19.	Secoiridoids	
20.	Lignans + flavonoids	
21.	Pyropheophytin a	Relative proportion of pyropheophytin a, expressed as percentage of the mass fraction, based on the ratios of the peak areas of pheophytin a and pheophytin a', determined by liquid chromatography
22.	Volat1	Normalized chromatographic intensities (heights) at specific retention time values previously selected, extracted from the corresponding volatile chromatographic fingerprint (no units).
23.	Volat2	
24.	Volat3	
25.	Volat4	
26.	Volat5	
27.	Volat6	
28.	Volat7	
29.	Volat8	
30.	Volat9	

Firstly, data were arranged in a 664×30 matrix: each row corresponds to each sample at specific storage time batch (0, 2, 4, …, 14 months). Each column corresponds to the value of a single determined parameter. Table 1 shows all parameters or variables, and a brief description of them, which have been monitored over 14 months.

In order to carry out the previous selection of the volatile variables, all volatile fingerprints were integrated into a data matrix. Then, a preliminary PLS regression model against storage time was built and the VIP coefficients were calculated. Variables (linked to retention times) showing an absolute VIP coefficient exceeding the threshold value of 1.00 were selected as relevant influential ones and the corresponding chromatographic intensity (height) was considered. Table 2 shows the nine selected volatile variables and the corresponding retention times for each one.

Table 2. Influential volatile variables, related to the volatile chromatographic fingerprint, on the storage time.

Name	Variable number	VIP scores for Y1	Retention time interval (min)
Volat1	19	1.34	1.89 – 2.36
Volat2	36	1.04	2.39 – 2.42
Volat3	55	1.01	2.50 – 2.51
Volat4	83	1.04	2.71 – 2.75
Volat5	142	1.22	2.95 – 3.06
Volat6	178	1.04	4.86 – 5.11
Volat7	203	1.02	5.33 – 5.34
Volat8	254	1.06	7.27 – 7.30
Volat9	280	1.09	8.25 – 8.37

3.2. Modelling of EVOO ageing

Experimental data were processed by multivariate analysis procedures to select significant parameters on stability of our EVOO lots. All data were mean-centred and the variables were weighed by their standard deviation to give them equal variance (autoscaled).

A multidimensional map of the 83 EVOO samples in relation to the 30 physicochemical parameters and sensory characteristics was obtained by PCA. 6 PCs were selected that explain 75% of the total data variability. The PC1 (40.6 % of variance) seemed to grouped and order the samples according to the ageing of storing (Figure 1) showing the higher scores a greater number of

months stored oils. In order to inspect for 'outlier' samples by checking the residual values and the leverage effect they generate in the model, a Q-T^2 screening plot is shown on figure 2. The particular data for three possible outliers are reviewed. No reason for outliers was detected. It was concluded that these must be measurement experimental errors and it was decided to continue without excluding any sample.

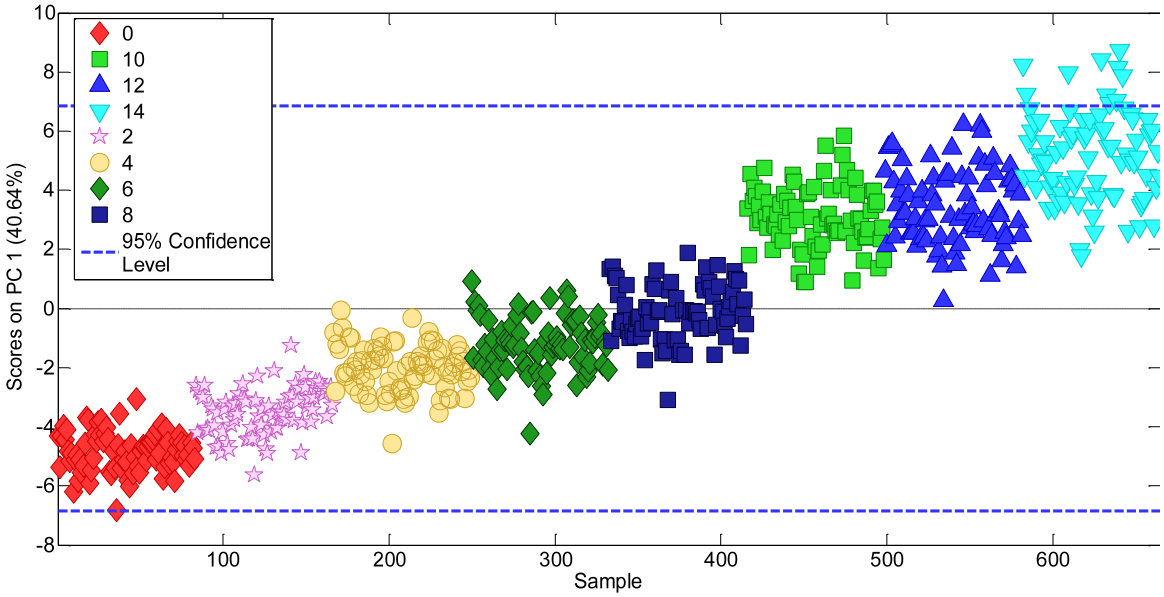


Figure 1. PC1 scores for the 83 EVOO samples: storage time is used only as class marker.

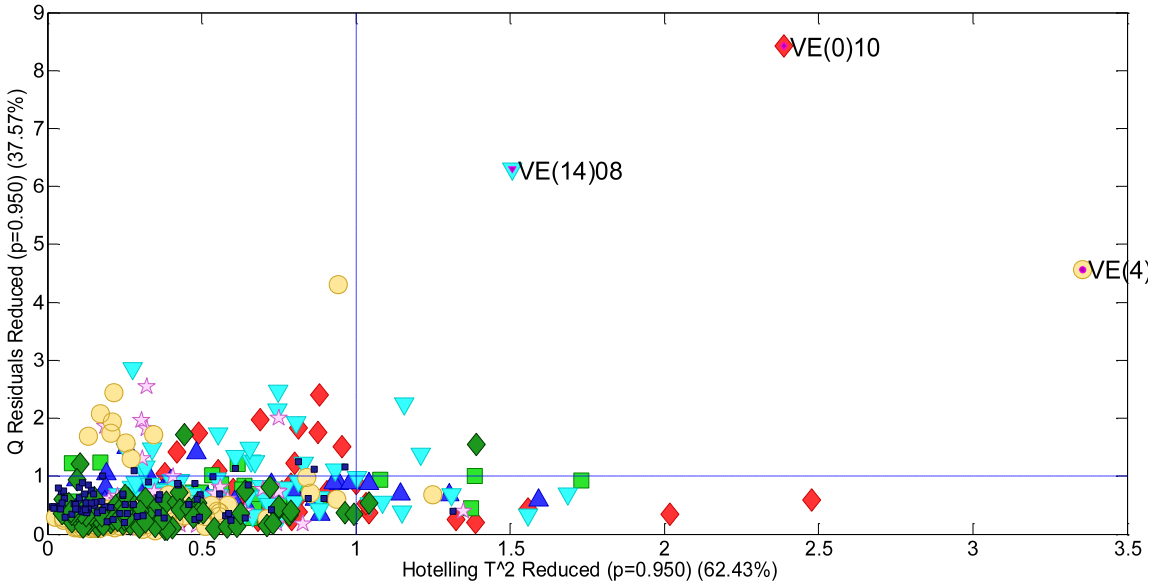


Figure 2. Q-T^2 screening plot obtained from the PCA model considering all variables.

Selection of significant variables. A preliminary PLS model, considering all the variables, was then applied to evidence possible relationships between storage time (y-variable) and the potentially significant dependent variables (x-variables). 4 LV were selected; LV1 explained 91% of the cumulative variance of the y-variable which stated a high correlation between among influential analytical parameters and storage time. Figure 3 shows that samples were linearly distributed with a correlation coefficient of 0.95. This leads to the option of using PLS (instead of PCA) to estimate the life time. In addition, figure 4 shows VIP coefficients for the PLS model.

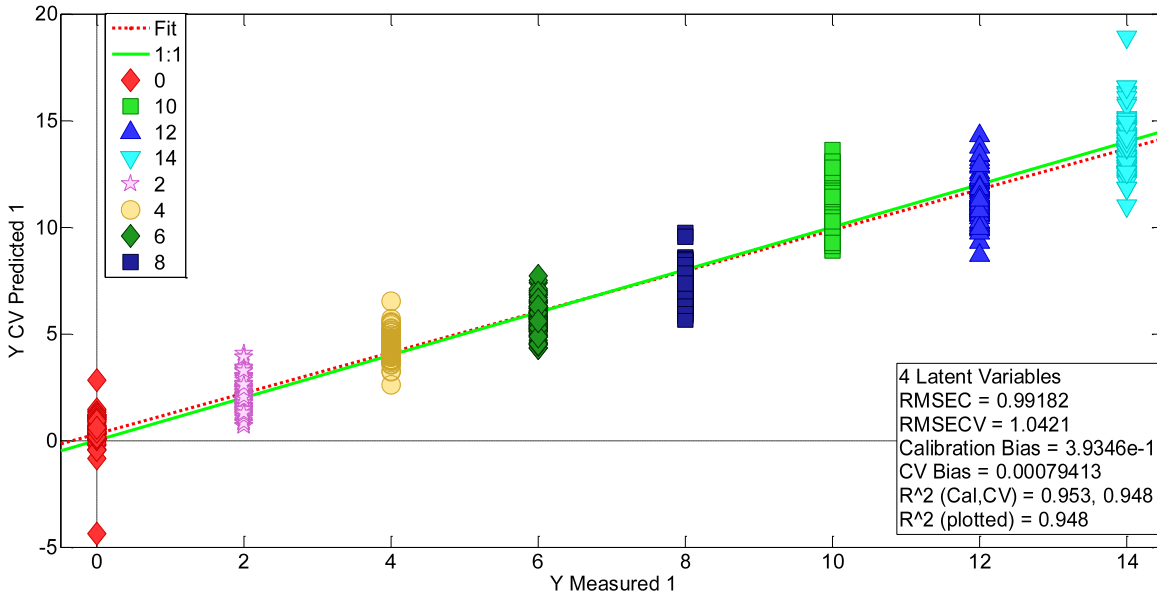


Figure 3. Relationship between predicted and known values of the independent y-variable (ageing time) obtained from the PLS model considering all variables.

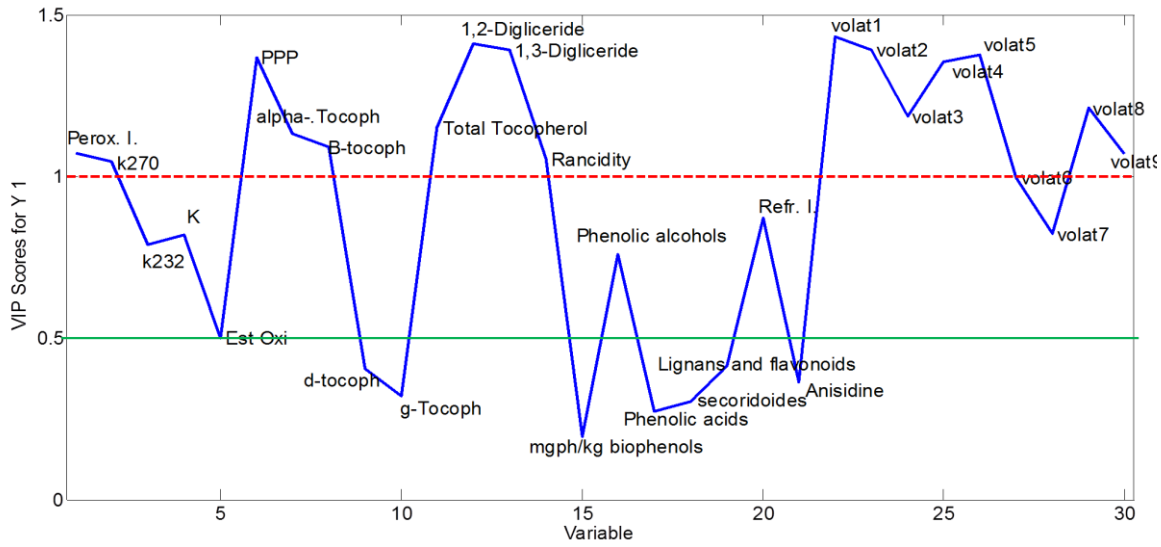


Figure 4. VIP scores profile obtained from the PLS model considering all variables.

Clearly γ-tocopherol and δ-tocopherol, total biophenols, phenolic acids, secoridoids, lignans+flavonoids, and anisidine value variable do not contribute with relevant information to the model and were not further considered (VIP coefficient <0.5). K_{232}, ΔK, oxidative stability, phenolic alcohols and refractive index are questionable. On the contrary, pyropheophytin a, 1,2-diglycerides, 1,3-diglycerides, volat1, volat2, volat3, volat4, volat5 and volat8 are the variables with the highest correlation with ageing.

Pearson's correlations between pair of variables included in the model were evaluated. Highly correlated variables (Pearson correlation coefficient >0.70) provide similar information and there is a risk of overfitting the model. Table 3 shows a summary of excluded and selected variable on the multivariate models for the rest of the study.

Table 3. Summary of both selected and excluded variables on the EVOO multivariate shelf life model.

Selected variables		Excluded variables	
Peroxide value	1,2-diglicerides	Refractive index*	1,3-diglicerides*
K_{270}	Volat1	Anisidine value	Volat3*
K_{232}	Volat2	Total biophenols	Volat4*
ΔK	Volat5	Secoridoides	Volat6*
Oxidative stability	Volat7	Phenolic acids	
Rancidity	Volat8	Lignans + flavonoids*	
Phenolic alcohols	Volat9	Total tocopherols*	
α-Tocopherol	Pyropheophytin a	γ-Tocopherol	
β-Tocopherol		δ-Tocopherol	

** Excluded variables for having correlation coefficients greater than 0.7 with any of the selected variables.*

Kinetic parameters. A new PCA model was built but considering only selected relevant variables (664×17 data matrix). On this occasion, 6 PCs were selected increasing both the total cumulative variance explained (>80%) and the variance explained by PC1 (>50%) with respect to the former model. It clearly shows that PC1 is time-structured, making it suitable for estimating the kinetic parameters.

Typically, degradation reactions could follow a zeroth-order, first-order or second-order kinetics, as well as the cumulative form of the Weibull model

which has proven useful in numerous studies that improved the fit of experimental observations to various quality indices as a function of time due to its extreme flexibility (**Amodio, Derossi, Mastrandrea, & Colelli, 2015**).

Zeroth order kinetics

$$\frac{-dQ}{dt} = k; \qquad Q = Q_0 - k \cdot t$$

First order kinetics

$$\frac{-dQ}{dt} = k \cdot Q; \qquad \ln Q = \ln Q_0 - k \cdot t$$

Second order kinetics

$$\frac{-dQ}{dt} = k \cdot [Q]^2; \qquad \frac{1}{Q} = \frac{1}{Q_0} + 2k \cdot t$$

Weibull kinetics

$$Q(t) = Q_0 \cdot e^{[-b_m(T)n_m(T)]}$$

where Q (or Q(t)) is a variable which collects the PC1 scores of the samples as a function of time, and Q_0 the score of the fresh sample, *i.e.*, a parameter representative of initial conditions at time zero. In all cases, k is the oxidation kinetic constant.

The best model fitting was obtained when the logarithm of PC1 scores is plotted *vs.* the storage time (coefficient of determination, R^2 = 0.894; root mean square error, RMSE = 0.0489) so we can stay that the data follow pseudo-first order degradation kinetics with Q_0 = 1.165 and k = 0.031.

Shelf life modelling. A subsequent PLS model, considering just the selected variables, was set up with 3 LVs (the first variable explains more than 90% of the cumulative variance of the y-variable, *i.e.*, storage time) considering only relevant x-variables. Figure 5a shows the LV1-LV2 scores plot. It can be perceived that the samples were ordered from negative scores for LV1 in the youngest oils, to positive values when storage time increases. It can be also seen how the biggest differences were found after month 10 of ageing, showing a wider dispersion of the samples, with most of the samples being outside the confidence region (95%) of the LV1-LV2 plot. Moreover, within the samples of the same class, a greater data scatter is found as ageing time progresses, *i.e.*, the samples degraded during 10 months show a greater variability between them than the samples aged 2 months; this suggests unequal ageing rates of the different samples under study. The greatest differences in the score values

appear between months 8 and 10; where the slope of growth of LV1 scores changes (see figure 5b).

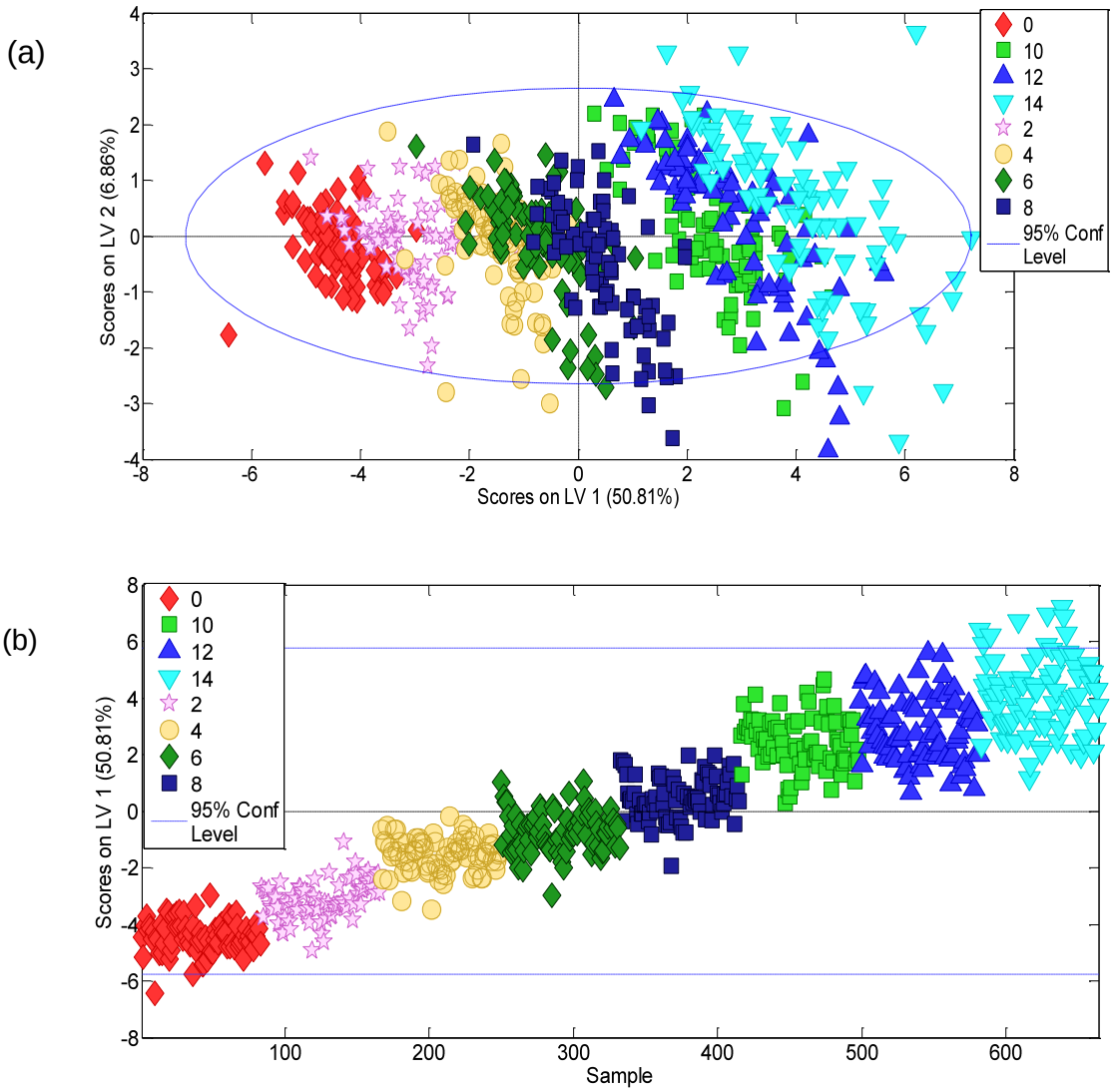


Figure 5. (a) LV1-LV2 score plot (PLS model considering all variables); (b) Evolving of LV1 scores.

Finally, a linear regression model is established between predicted and known values of the storage time, obtaining a good fit (R^2 = 0.94, RMSE = 0.004) as it is shown on figure 6.

LV1 is time-structured (as it can be seen on figure 5b) and accordingly, LV1 scores were used to establish a linear model ($Q(t) = a + b \cdot t$) of the autoscaled values of the experimental data *vs.* the storage time (R^2 = 0.91; RMSE = 0.89). The estimated coefficients (95% confidence interval) were: a = -4.267 *(-4.393, -4.141)* and b = 0.6096 *(0.5945, 0.6247)*.

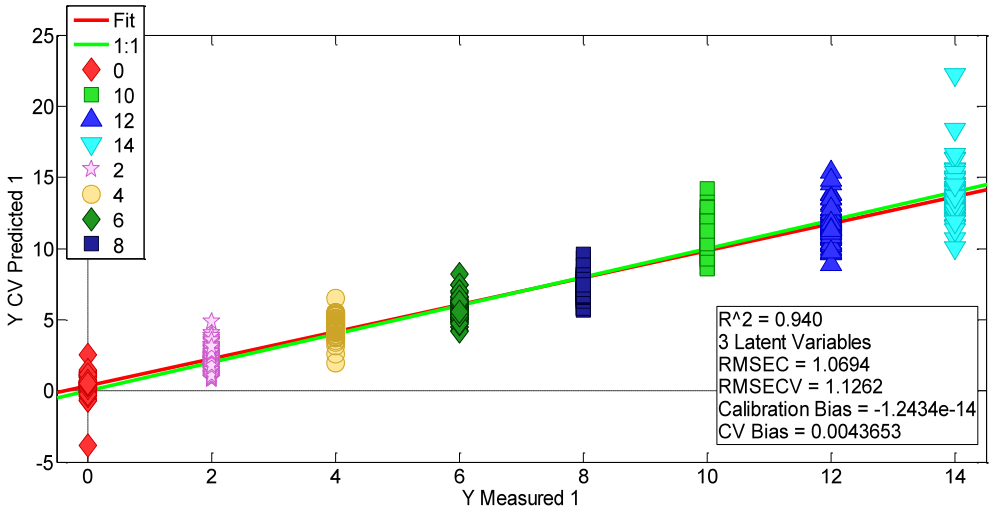


Figure 6. Relationship between predicted and known values of the independent y-variable (ageing time) by obtained from the PLS model considering the selected significant variables.

In order to estimate the shelf-life or 'critical time', t_C, of EVOO, is necessary to define the acceptability limit vector collecting the regulatory threshold values of the monitored physico-chemical or sensory characteristics under study (**European Commission, 2019; AS 5264, 2011**) beyond an olive oil can no longer be considered as EVOO: peroxide value ≤20; K_{232} ≤2.5; K_{270} ≤0.22; ΔK ≤0.01; median of rancid defect =0; pyropheophytin a ≤17%; 1,2-diacylglycerols ≥35%.

At this point the question that emerges was: *what limit values should be set for those variables that have significance for the shelf life model but they are not covered by regulations?* To solve that question, we decided to inspect for those scalars in 'unacceptable oils', *i.e.*, olive oils that has exceeded the compulsory limits in one or more physico-chemical or sensory analyses. Due to the unequal ageing (as expected) of the 83 EVOO samples in this study, it was decided to average the data for each ageing time. The robust median statistics was applied to find a representative value for all samples for each storage time. As a result, 14 vectors were obtained providing the median of the 17 selected analytical parameters of the 83 samples in months 0, 2, 4, ... up to 14 months. When those vectors were compared with the regulatory limits (where applicable), it should note that the median intensity of the rancid attribute exceeds the regulatory parameter from month 10 onwards and consequently may be considered that the half of the olive oils have lost EVOO category.

The acceptability limit vector, $\mathbf{Q_A}$, is then defined from the limit values for all the regulated parameters and the values obtained in the median vector of 10-month aged samples for the non-regulatory parameters. To this regard, the criteria for the definition of the shelf life acceptability limits for EVOO undergoing oxidation

are based on (1) compliance to legal requirements, and (2) robust statistics mean values. Table 4 shows acceptability limit vector values.

Table 4. Acceptability limit values used for building the scores vector to be applied to the EVOO shelf life model.

Variable	Acceptability limit	Variable	Acceptability limit
Peroxide value	20.0	1,2-diglicerides	35.00
K_{270}	0.22	Volat1	0.870
K_{232}	2.50	Volat2	0.382
ΔK	0.01	Volat7	0.045
Oxidative stability	11.10	Volat5	0.002
Rancidity	1.0*	Volat8	0.025
Phenolic alcohols	4.90	Volat9	0.003
α-Tocopherol	165.06	Pyropheophytin a	17.0
β-Tocopherol	2.47		

** lowest rancid organoleptic intensity being considered perceptible and non-zero*

For estimating a descriptive value of the Qc scalar, $\mathbf{Q_A}$ vector was autoscaled and multiplied by **L**, the loading vector of LV1, which explains the largest variance of the y-variable, *i.e.*, ageing time, regarding the PLS shelf life model. A value of 3.92 was obtained for Q_C.

Finally, Q_C was interpolated in the regression equation of multivariate shelf life model to obtain a t_C value, *i.e.*, the maximum ageing time at which an olive oil, provided it has been stored and analysed under the same experimental conditions, cannot be classified as EVOO anymore. 13 months were finally estimated to be a representative shelf life time for EVOO stored under the experimental conditions described in this study. Notice that a two-month uncertainty in the predictions is taken into account; this corresponds to the time that elapses from two consecutive analytical controls (experimental work frequency). Therefore, the shelf life time, given in months, should be rounded down to the nearest whole number to give each EVOO the benefit of the doubt and to ensure it complies with requirements.

3.3. Shelf-Life index and ageing rate

Once the model is established, and t_C is estimated, the equivalent ageing time, t_i, of every single sample could be calculated as:

$$Q_i = \mathbf{X_i} \cdot \mathbf{L} \quad \Rightarrow \quad Q_i = -4.267 + 0.6096 \cdot t_i$$

where $\mathbf{X_i}$ symbolizes the auto-scaled vector of the experimental data of each sample at a specific time, and **L** is the vector of loadings of the first latent variable, LV1, from the regression PLS model. Q_i is interpolated in the shelf life model to obtain t_i.

From this values, a 'Shelf Life Index', I_{SL}, is then defined as a cardinal reflecting the number of months in which an EVOO continues in compliance with the requirements of its category. It is calculated from:

$$I_{SL} = t_C - t_i$$

where the value of t_C is set to 13 months.

Likewise, an 'Ageing Rate', %Age, of the EVOO is calculated as follows:

$$\%\text{Age} = \frac{t_i}{t_C} \cdot 100$$

3.4. Verifying the shelf life prediction ability

In order test the reliability of the shelf life model and verify the reliability of the predicted critical times, 5 EVOO samples were randomly selected from among those included in the study:

- **VE003**. 'Arbequina' monovarietal EVOO from Granada (Andalusia, South of Spain). It lost the EVOO category after 8 months of storing, exceeding the regulatory value of K272 and rancid defect.
- **VE038.** 'Hojiblanca' monovarietal EVOO from Granada (Andalusia, South of Spain). It remained as EVOO throughout the study (14 months), *i.e.*, it did not exceed any of the regulated limit values.
- **VE040.** 'Picual' monovarietal EVOO from Jaen (Andalusia, South of Spain). It remained as EVOO until the 14th month, where it presence a rancid defect of 1.
- **VE053. '**Coupage' blend EVOO from Seville and Jaen (Andalusia, South of Spain). It lost the EVOO category in 12 months, as from which the value of rancid organoleptic intensity is higher than threshold value.
- **VE075**. 'Hojiblanca' monovarietal EVOO from Seville (Andalusia, South of Spain). It reached a rancid intensity value higher than zero from the month 8 of ageing.

For these samples, Q_i, t_i, I_{SL} and Age(%) were calculated at each ageing time (see table 5).

Table 5. Estimated shelf life index, ISL, and ageing rate, %Age, values for 5 randomly selected EVOO to be used as examples on the reliability of the predictions.

Sample	Ageing time	Q_i	t_i	I_{SL}	%Age
VE003	0	-3.98	0.47	12.5	4
***** *(Month 8)*	2	-2.61	2.72	10.3	21
	4	-0.89	5.54	7.5	43
	6	0.69	8.13	4.9	63
	8	1.97	10.24	2.8	79
	10	3.97	13.51	**-0.5**	**104**
	12	4.83	14.93	**-1.9**	**115**
	14	6.13	17.05	**-4.1**	**131**
VE038	0	-4.33	-0.10	13.1	-1
***** *(Stable)*	2	-3.45	1.33	11.7	10
	4	-1.94	3.82	9.2	29
	6	-1.42	4.68	8.3	36
	8	-0.42	6.32	6.7	49
	10	0.44	7.72	5.3	59
	12	1.08	8.77	4.2	67
	14	1.03	8.70	4.3	67
VE040	0	-5.84	-2.59	15.6	-20
***** *(Month 14)*	2	5.00	-1.20	14.2	-9
	4	-3.61	1.07	11.9	8
	6	-3.10	1.92	11.1	15
	8	-2.04	3.66	9.3	28
	10	0.74	8.22	4.8	63
	12	1.03	8.70	4.3	67
	14	2.41	10.95	2.0	84
VE053	0	-4.99	-1.18	14.2	-9
***** *(Month 12)*	2	-3.72	0.90	12.1	7
	4	-1.82	4.01	9.0	31
	6	-0.83	5.64	7.4	43
	8	0.30	7.49	5.5	58
	10	2.37	10.89	2.1	84
	12	2.58	11.24	1.8	86
	14	4.78	14.84	**-1.8**	**114**
VE075	0	-4.58	-0.51	13.5	-4
***** *(Month 8)*	2	-3.35	1.51	11.5	12
	4	-2.42	3.04	10.0	23
	6	-1.70	4.21	8.8	32
	8	0.28	7.46	5.5	57
	10	2.16	10.55	2.5	81
	12	2.76	11.52	1.5	89
	14	4.06	13.66	**-0.7**	**105**

** The month in which the olive oil loses EVOO category when the regulatory parameters are monitored independently (univariate approach), is indicated in parentheses; "stable" specifies that the olive oil still retains the EVOO category at the end of 14 months.*

The predicted lifetime from the model, taking as a reference the state of each of the olive oils at months 2 and 6, is then compared with the actual date on which the olive oil loses EVOO category. The objective was to check whether the prediction ability about the predicted time of each EVOOs to meet the requirements when no oxidation (or minimal, *i.e.*, 2 months of storage), and for the same EVOO when ageing is more evident (6 months of storage), was fulfilled.

At month 2 of storage, VE003 is predicted to keep its quality parameters within accepted limits for 10 months. Provided that oil is kept under described ambient condition and analyses are conducted after 10 months of storage, VE003 have just exceeded the acceptability limit and sample is considered non-conforming to the category.

In a similar way, VE038, VE053 and VE075 are predicted to have almost 12 to be in compliance with requirements. From month 14 of ageing onwards, they could either remain acceptable for a few months or have just failed to meet the standard. VE040 is predicted to be suitable up to month 16 and at month 14 of storing, it is predicted to be within two months until reaching the end of its shelf life.

Considering predictions at month 6 of storage, VE003 showed a shelf life index of 5 while it is non-compliant with the category in month 12 of storage. VE053 and VE075 are predicted to have shelf life indexes of 7 and 8 months respectively. Measurements carried out in month 14 predicted that both of them are already oxidised and have just exceeded the limit of acceptability fixed for its commercial category (note that these are true predictions taking into account the uncertainty). Finally, VE038 and VE040 showed indexes of 8 and 11 months to be in compliance with the legal requirements, and, at month 14 of storage (end of the study), they continue in accordance with the EVOO category.

4. CONCLUSIONS

The multivariate shelf life test was successfully applied to EVOOs. The results of the present study indicate that the actual shelf life of EVOO could be accurate predicted by measuring some key quality parameters, evaluating their significance and using multivariate data analysis tools to model real time EVOO degradation. Multivariate analysis is based in observation and analysis of more than one experimental variable at a time and taking into account the effects of all variables on the modelling of the feature of interest. The nature of multivariate approach is to reveal the inherent structure and meaning embedded within original set of variables through the application and interpretation of a variety of statistical methods. Conducting a multivariate study is not synonymous to applying multiple univariate approaches simultaneously,

as people might think. Therefore, the shelf life of EVOO is predicted in 13 months by considering the changes it undergoes as a whole, just as the value of a work of art (a painting) is measured not only by looking at each parameter (composition, colours, perspective, ...) but also by looking at the sense of harmony of the whole artwork.

Research is currently being carried out to verify the application of our multivariate modelling of kinetic and shelf life to directly predict the rate of oil degradation and remaining shelf life of any EVOO when stored under the experimental conditions described in this study. In our opinion, under the same storage conditions the same combination of stability/instability indices may result in an equal constant rate of kinetic degradation parameters.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgments

This work was financially supported by the European Regional Development Fund (ERDF) and Ministry of Science and Technology within the framework of the funding corresponding to 'Retos-Colaboración' (Challenges-Collaboration Program) call of the State Program for Research, Development and Innovation Focussed to Challenges of Society [RTC-2017-6170-2]. In addition, SMR author acknowledges the PhD grant supported by the Ministry of Universities for the Training of University Teachers (FPU 19/02078).

References

Amodio, M.L., Derossi, A., Mastrandrea, L. & Colelli, G. (2015). A study of the estimated shelf life of fresh rocket using a non-linear model. *Journal of Food Engineering, 150,* 19-28. https://doi.org/10.1016/j.jfoodeng.2014.10.030

Aparicio Ruiz, R., Roca, M., & Gandul Rojas, B. (2012). Mathematical model to predict the formation of pyropheophytin in a virgin olive oil during storage. *Journal of Agricultural and Food Chemistry, 60,* 7040−7049. https://doi.org/10.1021/jf3010965

AS 5264 (2011). Olive oils and olive-pomace oils. Standards Australia, Sidney.

Cocchi, M., Biancolillo, A. & Marini, F. (2018). Chemometric methods for classification and features selection. In J. Jaumot, C. Bedia, & R. Tauler (Eds), *Data analysis for omic sciences – Methods and applications* (pp. 265-299). Elsevier. https://doi.org/10.1016/bs.coac.2018.08.006

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

COI/T.20/Doc. No 19/ Rev.5 (2019). Spectrophotometric investigation in the ultraviolet. International Olive Council (IOC), Madrid.

COI/T.20/Doc. No 29 (2009). Determination of biophenols in olive oils by HPLC. International Olive Council (IOC), Madrid.

COI/T.20/Doc. No 35/ Rev.1 (2017). Determination of peroxide value. International Olive Council (IOC), Madrid.

Cuadros Rodríguez, L., Martín Torres, S., Ortega Gavilán, F., Jiménez Carvelo, A.M., López Ruiz, R., Garrido Frenich, A., Bagur González, M.G. & González Casado, A. (2021). Standardization of chromatographic signals – Part II: Expanding instrument-agnostic fingerprints to reverse phase liquid chromatography. *Journal of Chromatography A, 1641,* Article 461973, 1-14. https://doi.org/10.1016/j.chroma.2021.461973

European Commission (2019). Regulation (EEC) No 2568/91 on the characteristics of olive oil and olive-pomace oil and on the relevant methods of analysis (consolidated version 01991R2568-20191020). *Official Journal of the European Union, 032.001*, 1-128.

Government of Spain (2021). Royal Decree 760/2021, approving the quality standard for olive and olive-pomace oils *(in Spanish). Official Spanish Gazette, 209,* 106518-106532. https://www.boe.es/eli/es/rd/2021/08/31/760

Grauwet, T., Vervoort, L., Colle, I., Van Loey, A., & Hendrickx, M. (2014). From fingerprinting to kinetics in evaluating food quality changes. *Trends in Biotechnology, 32*, 125-131. http://dx.doi.org/10.1016/j.tibtech.2014.01.002

Guillaume, C., & Ravetti, L. (2016). Shelf-life prediction of extra virgin olive oils using an empirical model based on standard quality tests. *Journal of Chemistry, 2016*, Article 6393962, 1-7. http://dx.doi.org/10.1155/2016/6393962

ISO 280 (1998). Essentials oils – Determination of refractive index. International Organization for Standardization (ISO), Geneva.

ISO 29822 (2009). Vegetable fats and oils. Isomeric diacylglycerols. Determination of relative amount of 1,2- and 1,3-diacylglycerols. International Organization for

Standardization (ISO), Geneva.

ISO 29841 (2009). Vegetable fats and oils. Determination of the degradation products of chlorophylls a and a' (pheophytins a, a' and pyropheophytins). International Organization for Standardization (ISO), Geneva.

ISO 6885 (2016). Animal and vegetable fats and oils - Determination of anisidine value. International Organization for Standardization (ISO), Geneva.

ISO 6886 (2016). Animal and vegetable fats and oils – Determination of oxidative stability (accelerated oxidation test). International Organization for Standardization (ISO), Geneva.

ISO 9936 (2016). Animal and vegetable fats and oils. Determination of tocopherol and tocotrienol contents by high-performance liquid chromatography. International Organization for Standardization (ISO), Geneva.

Martin Torres, S, Ruiz Castro, L., Jiménez Carvelo, A.M., & Cuadros Rodríguez, L. (2022). Applications of multivariate data analysis in shelf life studies of edible vegetal oils – A review of the few past years. *Food Packaging and Shelf Life, 31*, Article 100790, 1-13. https://doi.org/10.1016/j.fpsl.2021.100790

Morales, M.T. & and Przybylski, R. (2013). Olive Oil Oxidation. In R. Aparicio & J. Harwood (Eds), *Handbook of olive oil – Analysis and properties* (4th ed.) (pp. 480-522), Springer Science.

Moschopoulou E., Moatsou G. & Drosinos E.H. (2019). Food quality changes during shelf life. In C.M. Galanakis (Ed.), *Food quality and shelf life* (pp. 1-31), Academic Press.

Ortega Gavilan, F., Valverde Son, L., Rodríguez García, F.P., Cuadros Rodríguez, C. & Bagur González, M.G. (2020). Homogeneity assessment of reference materials for sensory analysis of liquid foodstuffs. The virgin olive oil as case study. *Food Chemistry, 322*, Article 126743, 1-12. https://doi.org/10.1016/j.foodchem.2020.126743

Pignitter, M. & Somoza, V. (2012). Critical evaluation of methods for the measurement of oxidative rancidity in vegetable oils. *Journal of Food and Drug Analysis, 20*, 727-777. https://doi.org/10.10.6227/jfda.2012200305

Psomiadou, E. & Tsimidou, M. (2002). Stability of virgin olive oil. 1. Autoxidation studies. *Journal of Agricultural and Food Chemistry, 50*, 716-721. https://doi.org/10.1021/jf0108462

Sánchez Quesada, C. & 41 more. (2020). Relationship between olive oil consumption and ankle-brachial pressure index in a population at high cardiovascular risk. *Artherosclerosis, 314*, 48-57. https://doi.org/10.1016/j.atherosclerosis.2020.10.008

Talbot, G. (2011). The stability and shelf life of fats and oils. In D. Kilcast & P. Subramaniam (Eds), *Food and beverage stability and shelf life* (pp. 683-715), Woodhead Publishing.

PUBLICACIÓN VIII.

ARTÍCULO CIENTÍFICO

Multivariate stability monitoring and shelf life models – Part 2: The matter of further edible vegetable oils

Refined oils multivariate stability and shelf life modelling

Food Packaging and Shelf Life, (under review)

Multivariate stability monitoring and shelf life models – Part 2: The matter of further edible vegetable oils

Refined oils multivariate stability and shelf life modelling

Sandra MARTÍN-TORRES[1,✉], Juan Antonio TELLO-JIMÉNEZ[2], Rafael LÓPEZ-BLANCO[2], Antonio GONZÁLEZ-CASADO[1], Luis CUADROS-RODRÍGUEZ[1]

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

[2] *Laboratory 'J.A. Tello', S.L., Industrial Park "Los Olivares", C/ La Iruela, No 8, E-23009 Jaén, Spain.*

Highlights

- Monitoring of refined oil quality depletion under actual shelf life testing.
- Multivariate oxidation kinetic modelling using PCA scores.
- Multivariate acceptability limit includes regulatory and non-regulatory parameters.
- Comparison of shelf life prediction from multivariate and univariate approaches.

ABSTRACT

The oxidative stability of most commonly consumed refined edible vegetable oils was assessed. Oxidation simulating supermarket store conditions during 24 months ('actual' shelf life testing) was monitored by measuring physico-chemical and sensory parameters, combining instruments that generate a single datum per sample and instruments capable of generating multiple measurements in a single test (data vector). Methodology for carrying out a multivariate kinetic model is described, taking into account all analytical time-related variables and chemometric data processing tools. Multiparametric equations allowed predicting time (number of months) for each vegetable oil to be consistent with requirements. Results pointed out the accuracy of multivariate shelf life modelling with regard to univariate peroxide value-based model. Discrepancies were found in the oxidation rates of oils extracted from different plant seeds.

Key words

Multivariate deterioration monitoring, multiparametric-kinetic studies, chemometrics, shelf life, quality and stability indices, refined vegetable oil.

1. INTRODUCTION

Vegetable oils are important constituents of food and are essential components of human diet. Vegetable oils are obtained by mechanical expelling or solvent extraction of oleaginous seeds (soybeans, rapeseed, sunflower, etc.) or oleaginous fruit like palm and olive (**McGuire, 2022**). Among them, consumption of refined vegetable oils has been growing strongly in recent years. Attractive international edible vegetable oil prices (cheaper than virgin oils such as olive oil), together with improved gross producer profit margins, increase in cultivation area, as well as multiple industrial and culinary applications make palm, sunflower, soybean, rapeseed, canola, corn and grapeseed oils the most consumed edible vegetable oils worldwide (**Statista, 2022**).

Knowledge of the information about the potential or expected shelf life (oxidative stability) of culinary refined vegetable oils is therefore of great importance in terms of palatability, nutritional quality and, in some instances, toxicity before they are marketed, incorporated into food products, and/or used by the consumers as salad oils/dressings. Oxidative stability is defined as the resistance to oxidation during processing and storage and can be expressed as the period of time necessary for the critical point of oxidation, whether it is a sensorial change or a sudden acceleration of the oxidative process (**Roszkowska, Tanska, Czaplicki, & Konopka, 2015**). Because a large portion of phospholipids is removed by degumming in the refining of vegetable oils, oxidation of refined edible vegetable oils primarily involves the oxidation of triacylglycerols. During oxidation, in the presence of initiators such as light, unsaturated fatty chains yield to form hydroperoxides, primary oxidation products. Hydroperoxides, tasteless and odourless, are generally unstable and react to form secondary lipid oxidation products such as aldehydes, ketones, alcohols, organic acids, epoxides, and hydrocarbons. These secondary oxidation products would contribute off-flavour notes that significantly impact the sensory quality of edible vegetable oils/fats and foods containing oils/fats. In general, this negative characteristic of the product is known as rancidity and it is main cause of consumer's rejection. Additionally, oxidation products, together with free radicals, may have an adverse effect on the human body, or even be toxic after intake (**Hu, & Jacobsen, 2016; Tatum, & Chow, 2008**). Moreover, it should be noticed that the temperature enhances the kinetics of these processes.

One question that persists in lipid oxidation analytical control is which oxidation products are the most effective to monitor. Most food companies define shelf life in their own ways and follow their own criteria of defining a quality informative index for considering a product acceptable or unacceptable. That fact implies same vegetable oil bottled by two different companies could have different

expiration dates. One criterion for bottled vegetable oil is that the consumer should not be able to smell any oxidized or rancid aroma during handling and cooking so that only sensory perception is monitored. But, do all consumers perceive the same threshold? Other common approach is that vegetable oils cannot exceed a peroxide value (PV) regulated in definitive guidelines (**CXS 19-1981, 2019**). However, PV only provides information about primary oxidation compound. It does not take into account the secondary oxidation compounds, which are actually responsible for the off-odours.

On a previous work (**Part 1 of series**), authors proposed chemofoodmetrics as the best way to carry out a shelf life study and discussed framework and fundamentals for carrying out a multivariate shelf life model that allows an estimation of the shelf life of virgin olive oil. Multiparametric monitoring kinetic and shelf life models have been proved to be more generically applicable than single-parameter models and could be more easily applied to other products. This paper, which constitutes Part 2 of the series, is aimed at estimating shelf life of most commonly consumed refined edible vegetable oils in Spain when oxidised under standard storage conditions relying on multiparametric modelling. The basics of applied methodology have been adequately described and explained in Part 1. However, the most critical aspects are also presented in this article so that they can be read and understood independently of the former.

2. METHODOLOGY

2.1. Refined vegetable oil samples

60 refined vegetable oil samples were used in this study, distributed as follow: 29 samples of sunflower oil; 1 sample of grapeseed oil; 1 sample of rapeseed oil; 4 samples of corn oil; 4 samples of vegetable oil blend constituted by different (not stated) seeds; 17 samples of olive oil (marketed-blend of virgin and refined olive oil); and 4 samples of pomace oil; were bought on grocery supermarkets. The requirement was established that the vegetable oils should have been bottled for the shortest possible time (maximum 2 months).

2.2. Storage

Twelve batches were considered, consisting of the 60 different refined vegetable oils, which were packaged in transparent food-grade PET 60 mL bottles (labelled by A to L letters) and were stored in a temperature-controlled room (20 ± 5 ºC) exposed to 12 hours of LED light, simulating supermarket store conditions during 24 months. The first lot (A), it was analysed at the

beginning of the study, consider as two month of ageing. After 2 months of storage, lot (B) was analysed. Consecutively, every two months, an aliquot of each vegetable oil sample was analysed (up lot (C) to lot (L) after 24 months of ageing) for determining physico-chemical and sensory (volatile-based) parameters.

2.3. Analytical equipment

Analytical equipment (refractometer, molecular absorption spectrophotometers, Rancimat instrument and chromatographic systems) used are properly described in Part 1 of this series (**Part 1 of series**)

2.4. Physico-chemical testing

Characteristic primary and secondary oxidation analytical parameters were carefully chosen to be monitored as feasible quality indicators to monitor oxidative changes as a function of time.

Refractive index was determined according to ISO 280 standard (**ISO 280, 1998**). Spectroscopic UV absorptivity values, K_{232}, K_{270} and ΔK were determined in compliance with COI/T.20/Doc. No 19 standard (**COI/T.20/Doc. No 19/ Rev.5, 2019**). The oxidative stability was measured by using the Rancimat method at 120 ºC, described in ISO 6886 standard (**ISO 6886, 2016**). In addition, both peroxide and anisidine values were determined according to the recognised methods described in COI/T.20/Doc. No 35 (**COI/T.20/Doc. No 35/ Rev.1, 2017**) and ISO 6885 standards (**ISO 6885, 2016**), respectively. Total α-, β-, γ- y δ-tocopherols were quantified according to ISO 9936 standard (**ISO 9936, 2016**).

Authorised antioxidant additives such as butylated hydroxytoluene (BHT), butylated hidroxyanisole (BHA), tert-butyl-hydroquinone (TBHQ) and propylgallate (PG) were quantified by high-performance liquid chromatography (Agilent 1260 series) and successive UV molecular absorption measurements using a diode-array detection system following the principles outlined in the IUPAC method (**Dieffenbacher, & Poclington, 1992**) with slight modifications on the chromatography conditions. A Zorbax Eclipse Plus C18 4.6 × 50 mm, 5 µm column was used. A flow rate of 1.2 mL/min and a column temperature of 30 ºC were set. 20 µL of sample was injected, 5% (v/v) solution of acetic acid in water (A) and 5% (v/v) solution of acetic acid in acetonitrile (B) were used as mobile phases, and the following solvent program was applied: from 70% of A to 100% of B over 7 min, with a 10-min hold at 100% of B. Return to initial conditions in 1 min and hold until 22 min.

For each vegetable oil, volatile compounds were firstly fractioned by headspace solid-phase microextraction (HS-SPME) and further a representative

chromatographic fingerprint were acquired using a gas chromatographer equipped with a flame ionization detector (GC-FID). A more detailed explanation of both sample preparation and chromatographic conditions are described in a previous paper (**Ortega Gavilán, Valverde Son, Rodríguez García, Cuadros Rodríguez, & Bagur González, 2020**). After acquisition, volatile fingerprints were exported from the instrument software and embedded in a data vector, and further pre-processed: (i) selection of region of interest; (ii) alignment using an average chromatogram; and (iii) normalization of intensities with respect to the internal standard. Each data vector containing the related chromatographic fingerprint of a refined vegetable oil sample consisted of 8391 elements, each of which is single variable indicating a normalised signal intensity. When fingerprinting methodology is applied, the compounds referred to are neither identified nor quantified in the conventional way. It is therefore necessary to apply a selection of the variables that best define the evolving of vegetable oil ageing.

2.5. Kinetic modelling and shelf life estimation

As argued previously, modelling food quality features means modelling changes, the quality of a food nearly always changes over time. Food quality modelling is therefore almost synonymous with kinetic modelling. The consequence is that certain differential equations frequently form the basis for mathematical models; these can sometimes be solved analytically but if not, it is relatively easy nowadays to solve them numerically with the available software, or even using conventional spreadsheets (**Van Boekel, 2008**).

As explained in Part 1 of this series (**Part 1 of the series**) two chemometric multivariate tools are used for kinetic modelling and shelf life prediction respectively: principal component analysis (PCA) and partial least-squares linear regression (PLS). The two chemometric tools also allow detecting outliers, to assess the relationships between variables and to evaluate the importance of individual variables in the overall quality depletion model.

One of the critical steps in assessing shelf life and distinguishing from a stability study is to properly establish an acceptability limit from which shelf life values could be calculated. When multivariate shelf life model is conducted, acceptability limit consist of a vector that includes the limit values, normally referred to legal or regulatory requirements, of all the concerned physico-chemical features that show change over time and gives rise to a single scalar coefficient, Qc, which is traduced to shelf life time:

$$Q_C = \max(\mathbf{Q_A} \cdot \mathbf{L})$$

where $\mathbf{Q_A}$ is the autoscaled acceptability limit vector and **L** the loading matrix of

the time related latent variable of the model. Q_C is then interpolated into shelf life model equation to obtain cut off criteria, t_C, or shelf life time.

Attending to refined vegetable oils, different regulatory limits are found whether refined vegetable oils from olives, *i.e.*, olive and pomace oils (**European Commission, 2019**), or other seed oils (**CXS 210-1999, 2019**) are concerned. Because of that, samples were modelled separately divided in two groups: (i) olive-pomace refined oils, and (ii) refined seed (sunflower, grapeseed, rapeseed, corn and blend) vegetable oils. Results of both groups of samples are discussed below.

3. RESULT AND DISCUSSION

3.1. Experimental data matrix of ageing

As stated before, data were arranged in two different data matrices: (i) 252×22 olive-pomace oil data matrix; and (ii) 468×20 sunflower-grapeseed-rapeseed-corn-blended oil data matrix.

Each row corresponds to each sample at specific storage time and each column corresponds to the value of a single determined parameter. Table 1 shows all parameters or variables monitored for olive-pomace oil, and a brief description of them. Regarding refined seed oils, peroxide value, K_{270}, K_{232}, ΔK, refractive index, oxidative stability, anisidine value, total tocopherols, α-tocopherol, β-tocopherol, γ-tocopherol, δ-tocopherol, volat1, volat2, volat3, volat4 and volat5 variables were considered. In addition BHA, BHT and PG variables, which refer to absolute content of each type of antioxidant, are given in mg/kg of vegetable oil. As the addition of preservatives is not allowed in olive-pomace oils, they were not tested for them.

In order to carry out a previous selection of the volatile variables, all volatile fingerprints were integrated into a data matrix. Then, a preliminary PLS regression model against storage time was built and the 'variable importance in projection' (VIP) coefficients were calculated. Variables (linked to retention times) showing an absolute VIP coefficient exceeding the threshold value of 1.00 were selected as influential ones and the corresponding chromatographic intensity (height) was considered. Table 2 shows the selected volatile variables and the corresponding retention times for each matrix.

Table 1. Physico-chemical and sensory parameters being monitored in olive-pomace study.

No	Parameter	Brief description
1.	Peroxide value	Peroxide content expressed in terms of milliequivalents of active oxygen per kilogram of vegetable oil.
2.	K_{270}	The specific absorbances are calculated for a concentration of 1% (m/V) in a 10 mm cell (absorbance units× $(g/100mL)^{-1}\times cm^{-1}$).
3.	K_{232}	
4.	ΔK	
5.	Refractive index	At 20 °C as reference temperature (no units).
6.	Oxidative stability	Rancimat induction period (in hours) at 120ºC.
7.	Anisidine value	Rate of increase of absorbance, at 350 nm in a 10 mm cell, when reacted with p-anisidine under specific conditions (no units).
8.	Total tocopherols	Absolute content of each type of tocopherol expressed in milligrams per kilogram of vegetable oil, determined by liquid chromatography.
9.	α- tocopherol	
10.	β- tocopherol	
11.	γ- tocopherol	
12.	δ- tocopherol	
13.	Volat1	Normalized chromatographic intensities (heights) at specific retention time values previously selected, extracted from the corresponding volatile chromatographic fingerprint (no units).
14.	Volat2	
15.	Volat3	
16.	Volat4	
17.	Volat5	
18.	Volat6	
19.	Volat7	
20.	Volat8	
21.	Volat9	
22.	Volat10	

3.2. Modelling of ageing of refined olive-pomace oils

Experimental data matrix was autoscaled: variables were mean-centred and weighed by their standard deviation to give them equal variance. A multidimensional map of the 21 olive-pomace samples in relation to the 22 physico-chemical parameters was obtained by PCA. 6 principal components (PCs) were selected that explain 73% of the total data variance. The PC1-PC2 scores plot, which explains the greatest variability in the data, shows a grouping

of the samples according to the type of vegetable oil involved, discriminating between olive and pomace oil (figure 1). However, based on the PC2 scores, a natural grouping of samples according to storage (ageing) time is evident (figure 2).

Table 2. Influential volatile variables, related to the volatile chromatographic fingerprint, on the storage time.

(i) Olive-pomace oil matrix			(ii) Seed oil matrix		
Name	**Variable number**	**Retention time interval (min)**	**Name**	**Variable number**	**Retention time interval (min)**
Volat1	77	1.60-1.68	Volat1	76	1.58-1.69
Volat2	167	1.77-1.79	Volat2	196	1.82-1.84
Volat3	195	1.81-1.84	Volat3	521	2.34-2.48
Volat4	337	2.04-2.12	Volat4	846	2.89-2.93
Volat5	491	2.31-2.32	Volat5	871	2.93-3.01
Volat6	518	2.33-2.47			
Volat7	845	2.90-2.92			
Volat8	869	2.93-2.97			
Volat9	2092	4.96-5.01			
Volat10	2124	5.03-5.06			

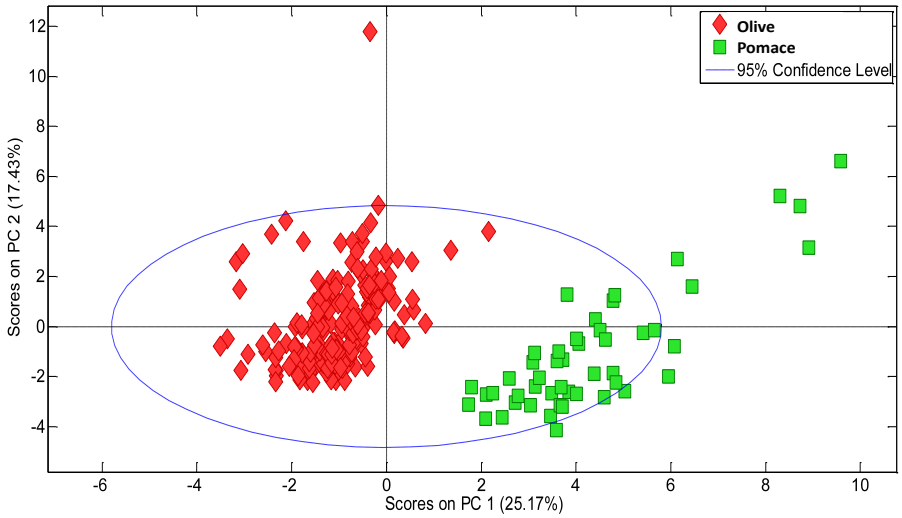


Figure 1. PC1-PC2 scores plot obtained from the PCA model considering all variables for the olive-pomace oil.

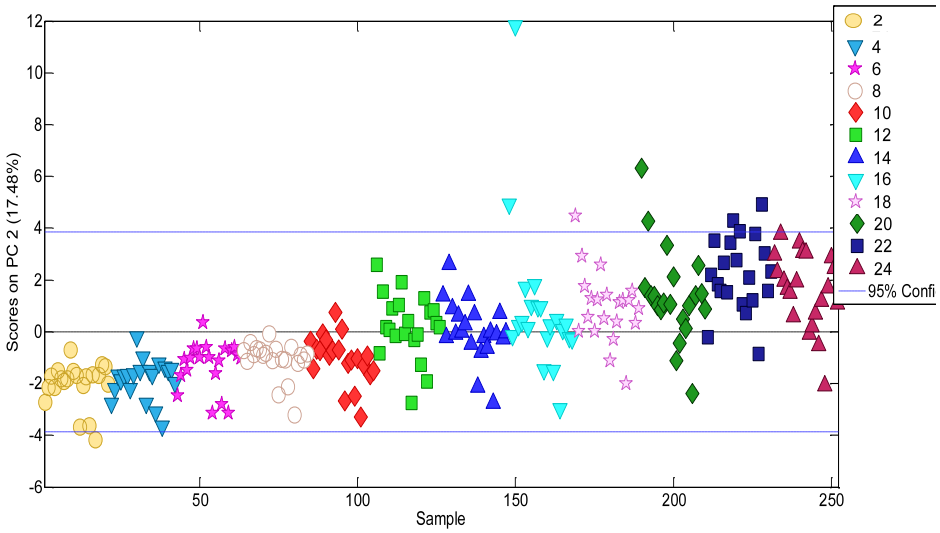


Figure 2. PC2 scores obtained from the PCA model considering all variables for the olive-pomace oil: storage time is used only as class marker.

When inspecting the residual-leverage (Q-T^2) screening plot, two different samples (from 16 and 20 month of storing) were shown as possible outliers. No specific cause was found and it was concluded that they must have been random experimental errors in one or several of the performed analyses. Consequently, it was decided not to consider these samples as outliers and to keep them in the model.

Selection of significant variables. A second PLS modelling was then applied to evidence possible relationships between storage time (y-variable) and the potentially significant dependent variables (x-variables). 4 latent variables (LVs) were then selected; LV1 explained 64% of the variance of the y-variable which stated a high correlation between analytical parameters and storage time.

The importance of the variables for the model is estimated by evaluating the loadings profile of the PLS model. For this purpose, the VIP coefficients are plotted against the variables (figure 3). VIP formulation is a combined measure of how much a variable contributes to describe the two data sets; the dependent variable (y-variable) and the independent variable (x-variable). A variable with a VIP coefficient close to, or greater than, one can be considered relevant in a given model. Variables with VIP coefficients lower than one are less important and might be good candidates for exclusion from the model. For previous experience, 0.5 was set as a threshold value: any variable with VIP below 0.5 was considered to be unrelated over time. Variables with VIP between 0.5 and 1 were considered 'borderline' and a Pearson correlation analysis with reference to storage time was then applied in order to verify the potential significance on the model. Pearson correlations between pair of variables were also evaluated to avoid highly correlated variables which could cause overfitting of the model. Table 3 shows a summary of excluded and included variable on the multivariate models for the rest of the study.

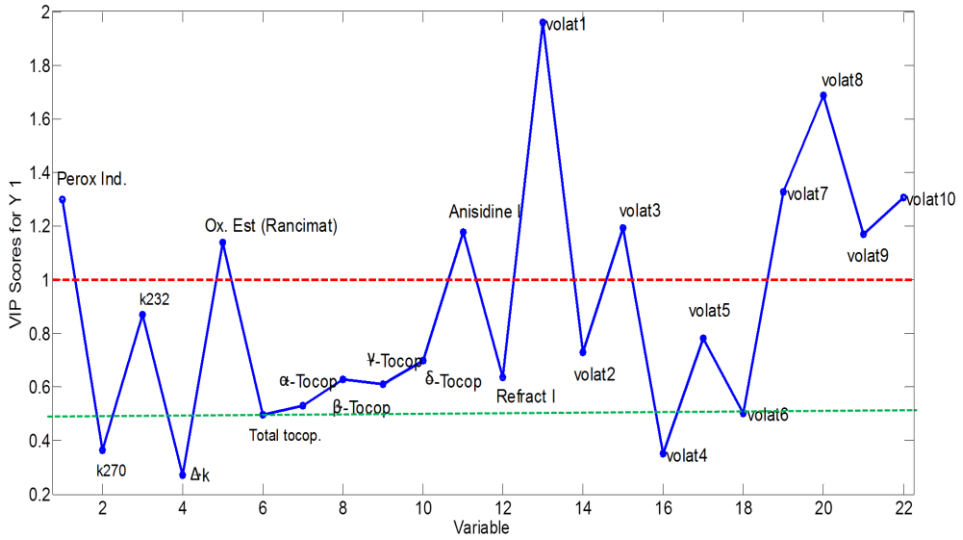


Figure 3. VIP scores profile obtained from the PLS model considering all variables.

Table 3. Summary of both selected and excluded variables on the olive-pomace multivariate shelf life model.

Selected variables		Excluded variables	
Peroxide value	Volat1	K_{270}	Volat4
Oxidative stability	Volat2	K_{232}*	Volat6
Anisidine value	Volat3	ΔK	Volat10*
α-Tocopherol	Volat5	Total tocopherols	
γ-Tocopherol	Volat7	β-Tocopherol*	
	Volat8	δ-Tocopherol	
	Volat9	Refractive index*	

** Excluded variables for having correlation coefficients greater than 0.7 with any of the selected variables.*

Kinetic parameters. A new PCA model was built but considering only selected relevant variables (252×12 data matrix). This time, 5 PCs were selected increasing both the total cumulative variance explained by the model (>72%) and the variance explained by PC1 (28%) with respect to the former model. Figure 4 clearly show that PC1 is now time-structured, making it suitable for estimating the kinetic parameters: samples were ordered from negative scores for PC1 in the youngest olive-pomace oils, to positive greater values when storage time increases. PC1-PC2 scores plot no longer shows discrimination between olive and pomace oils.

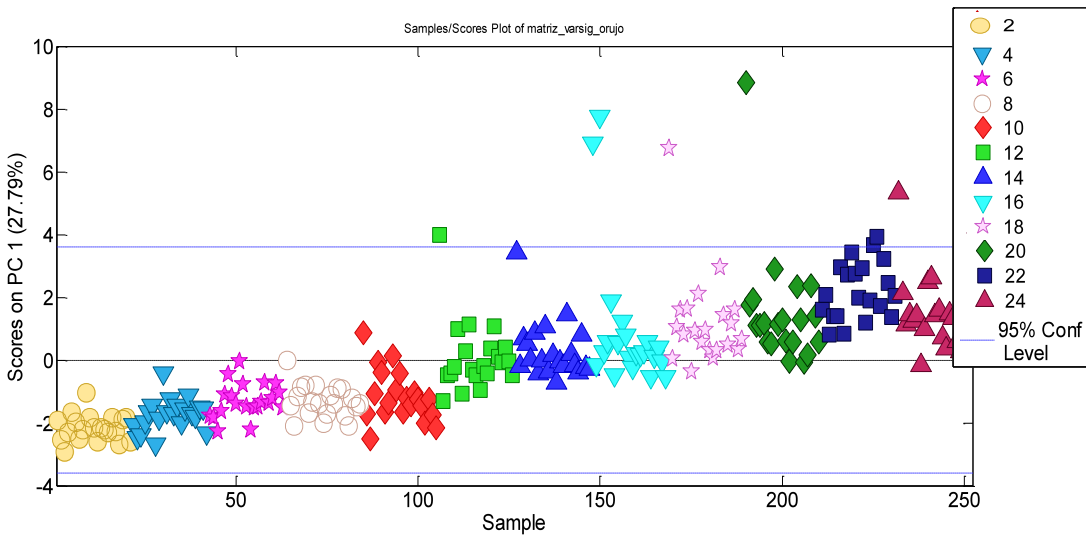


Figure 4. PC1 scores obtained from the PCA model considering only ageing related variables for the olive-pomace oil: storage time is used only as class marker.

Fitting a multivariate kinetic model involves the kinetic description of the important degradation reactions based on the PC1 scores, assuming that the degradation reactions are the main sources of variability in the data set. Robust bisquare fitting was applied. The best model fitting was obtained when the logarithm of PC1 scores are plotted *vs.* the storage time (coefficient of determination, R^2 = 0.82, root mean square error, RMSE = 0.0506) so that a pseudo-first order degradation kinetics could be regarded:

$$-\frac{dQ}{dt} = k \cdot Q \; ; \qquad \ln Q = \ln Q_0 - k \cdot t$$

where Q is a multivariate parameter that gathers the PC1 scores of the samples as a function of time, Q_0 is the score of the fresh sample and k is the oxidation kinetic constant. Values of Q_0 = 1.165 and k = 0.031 were calculated.

Shelf life modelling. A last PLS regression model was set up considering only relevant x-variables. 4 LVs were selected, taking into account that LV1 explained more than 65% of the variance of y-variable, *i.e.*, storage time. Finally, the linear model between the y-variable predicted by the model and the y-variable (ageing time) proved to be suitable, showing a goodness of fit of 0.73 with an RMSE of 0.08 (figure 5a). However, two trends in the data could be distinguished; samples from 2 to 10 months of ageing have similar scores. Thus, it seems that in these first months the vegetable oils did not almost evolve and it is after 10 months of storage that the slope of the LV1 scores increases (see figure 5b).

(a)

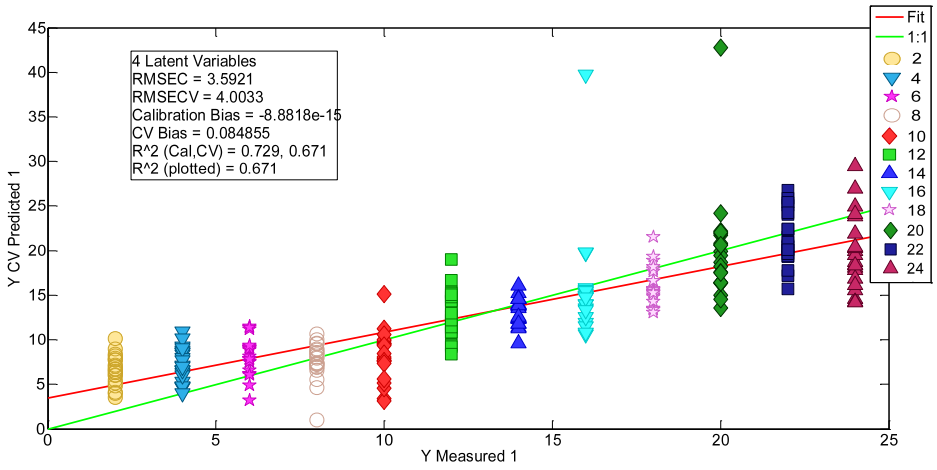


Figure 5. (a) Relationship between predicted and known values of the independent y-variable (ageing time) obtained from the PLS model considering the selected significant variables.

(b)

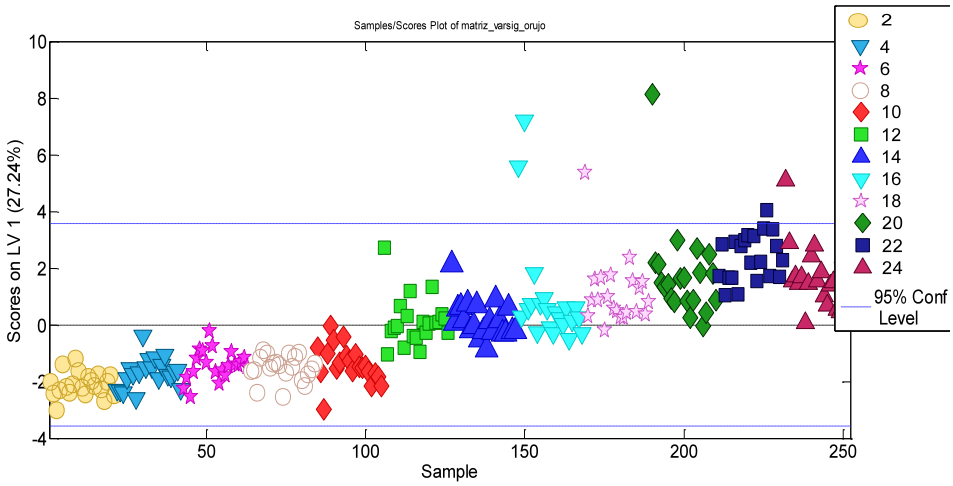


Figure 5. (b) LV1 scores used to establish shelf life model equation.

LV1 scores were used to establish a linear model (Q(t) = a + b·t) of autoscaled experimental data *vs.* storage time (R^2 = 0.84; RMSE = 0.73). The estimated coefficients *(95% confidence interval)* were: a = -2.702 *(-2.896, -2.509)* and b = 0.1977 *(0.1845, 0.2108)*.

As argued in the first part of this series (**Part 1 of this serial**), the acceptability limit vector collects the regulatory threshold values of the monitored physico-chemical characteristics under study as well as robust statistics mean values of non-regulatory parameters at storage time in which the vegetable oils can no longer be considered as complying with the requirements. In this case, only PV is regulated. An olive-pomace oil sample should be considered unfit for consumption if the PV is 15 mEq/kg or more [¡Error! Marcador no definido.].

When comparing the evolving of the median calculated from the peroxide values of all the samples with the regulatory value, we could see that it

increases from 12.2 at 14 months of storage to 19 when samples had been stored for 16 months. Furthermore, observing the evolution of the oxidative stability, it decreased from 18.8 to 13.5 h in month 14 of storage and then remain stable (taking into account the error of the method) until the end of the study. Therefore, the acceptability limit vector, $\mathbf{Q_A}$, is defined from the median values of each variable included in the fitted model considering all the samples at a storage time of 14 months, with the only exception of the peroxide value, which is set at 15 according to the current legislation. Table 4 shows acceptability limit vector values.

$\mathbf{Q_A}$ vector was autoscaled and multiplied by **L**, the loading vector of LV1, which explains the largest variance of the y-variable, *i.e.*, storage time, regarding the PLS shelf life model. A value of 0.155 was obtained for Q_C which was interpolated in the regression equation of multivariate shelf life model to obtain a t_C value, representative of the vegetable oil shelf life. 14 months were finally estimated to be a representative shelf life time for olive-pomace oils stored under the experimental conditions described in this study.

Table 4. Acceptability limit values used for building the scores vector to be applied to the olive-pomace oil shelf life model.

Variable	Acceptability limit	Variable	Acceptability limit
Peroxide value	15.00	Volat1	0.196
Oxidative stability	13.37	Volat2	0.143
Anisidine value	5.39	Volat3	0.130
α-Tocopherol	174.70	Volat5	0.086
γ-Tocopherol	10.30	Volat7	0.108
		Volat8	0.084
		Volat9	0.073

Peroxide value-based modelling – univariate approach. At this point the question arises as to what results we would have obtained based only on the empirical modelling of the PV data, in a univariate approach, taking into account that it is the only parameter whose value is regulated and has been taken into account to establish the limit vector. A conventional least-squares regression of the peroxide data with respect to ageing time was performed in order to obtain the (purely empirical) model that best fits and explains the variation of the PV over ageing time. The best fit of the data is achieved with the multiplicative form of the model (R^2 = 0.67; mean absolute error = 0.23):

$$PV = a \cdot t^{b}$$

Observations with residuals greater than 3 were considered outliers and were removed from the model (5 outliers). The equation of the fitted (linearised) model was:

$$\ln PV = 1.021 + 0.605 \cdot \ln t$$

From this model, the lifetime was calculated to be 16 months when the PV is interpolated to 15 mEq/kg. This one result an overestimation compared to the value estimated from the multivariate approach.

3.3. Modelling of ageing of refined seed oils

As already explained, a PCA model is now built considering all the variables under study. 6 PCs, explaining the 65% of the total data variance, were selected. PC1-PC2 scores plot shows a sample grouping according to the type of vegetable oil involved (notice that blend seed oils appear with the same scores as sunflower oils, which leads to the conclusion that sunflower could be the predominant type of vegetable oil into the blend). PC3 shows the highest scores with regard to increase on storage time (see figure 6). No outliers were considered.

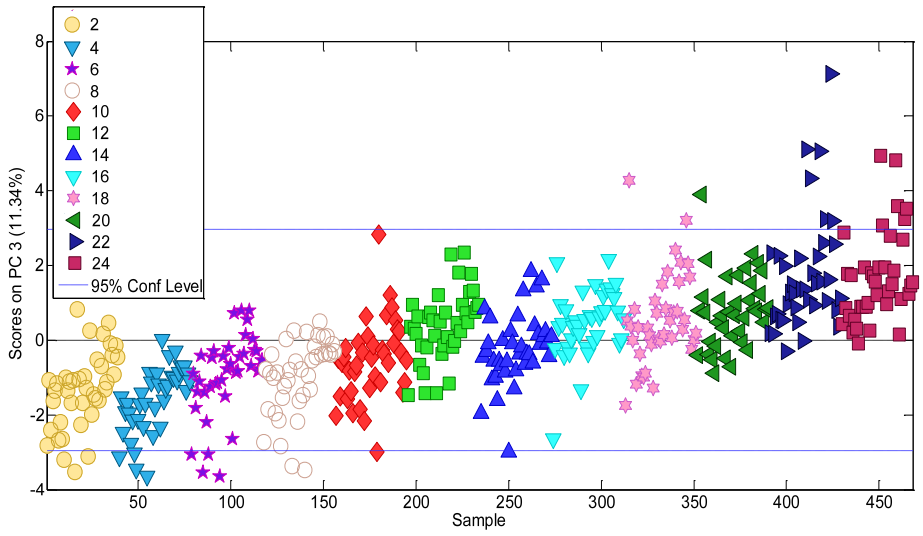


Figure 6. PC3 scores for the refined seed oils PCA modelling: storage time is used only as class marker.

Selection of significant variables. PLS modelling VIPs and between-variables Pearson's correlations were evaluated. Significant selected variables related to storage time were reduced to: PV; K_{270}; oxidative stability; total tocopherols; β-tocopherol; anisidine value; volat1; volat2; volat3; and volat5.

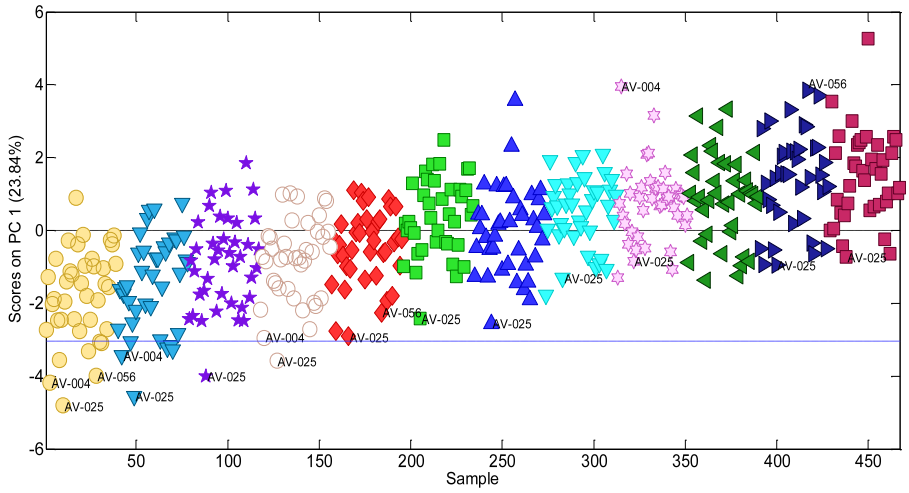


Figure 7. PC1 scores obtained from the PCA model considering only ageing related variables for the refined seed oil.

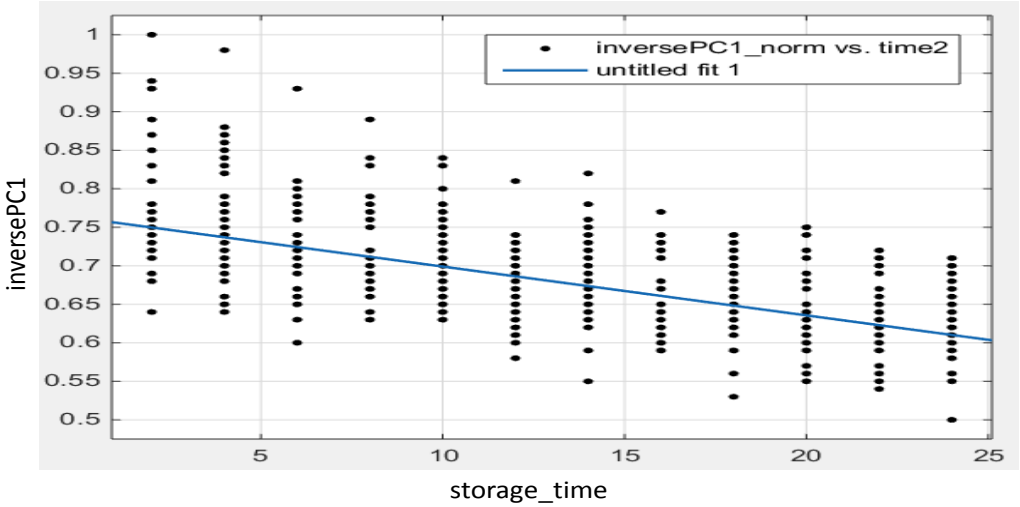
Kinetic parameters. A PCA model considering only relevant x-variables (468×10 data matrix) was build. 6 PCs were selected which showed an increase of both the total model-explained cumulative variance (>80%) and the PC1-expained variance (24%) with respect to the former model. This time, PC1-PC2 scores plot no longer shows grouping by vegetable oil type as before. PC1 is time-structured and therefore suitable for estimating the kinetic parameters. It was also observed that three samples have lower PC1 scores than expected, particularly within the first months (see figure 7). These samples agree in being high oleic sunflower oils, so it could be thought that their degradation kinetics are lower than the remaining samples at the beginning of the storage time, while as the oxidation progresses these differences are not so clear.

A multivariate kinetic model with PC1 scores was fitted. The best results were obtained when fitting a pseudo-second order degradation kinetics equation (R^2 = 0.4437, RMSE = 0.0559):

$$-\frac{dQ}{dt} = k \cdot [Q]^2 \; ; \qquad \frac{1}{Q} = \frac{1}{Q_0} + 2k \cdot t$$

where Q collects the PC1 scores of the samples as a function of the storage time, Q_0 the score of the fresh sample and k is the oxidation kinetic constant. Values of Q_0 = 1.312 and k = -0.0032 were calculated. In this case, the fit to the model was not good. During the first months of ageing a large data scattering can be observed, and thus high residual values, although these differences were not as pronounced at the end of the study (see Figure 8). This could be a consequence of the great variability of the samples and, therefore, of discrepancies in the oxidation rates.

(a)



(b)

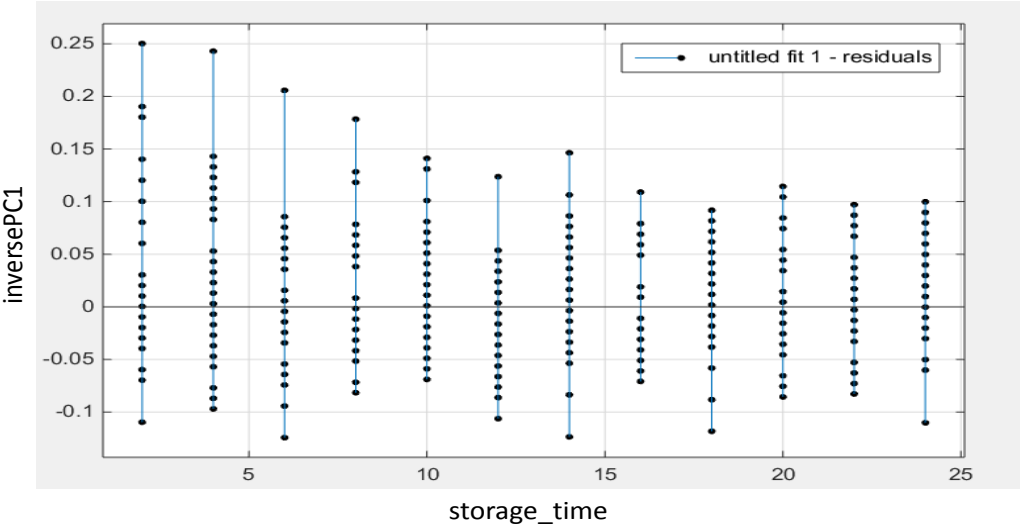


Figure 8. (a) Relationship between the inverse of the PC1 scores and storage time (multivariate kinetic modelling). (b) Residual plot for the kinetic model.

Shelf life modelling. Supervised PLS regression model was set up considering only relevant x-variables. 4 LVs were selected, taking into account that LV1 explained almost 60% of the variance of y-variable, *i.e.*, storage time. A goodness of fit of 0.63 with an RMSE of 0.05 was obtained when y-predicted *vs.* y-measured is plotted. LV1 scores were used to establish a linear model ($Q(t) = a + b \cdot t$) between autoscaled experimental data and storage time (R^2 = 0.67; RMSE = 0.87). The estimated coefficients (*95% confidence interval*) were: a = -2.135 (*-2.304, -1.965*) and b = 0.1602 (*0.1487, 0.1717*).

Similarly, only a PV less than 10 mEq/kg is required for a refined seed oil to be considered as meeting the mandatory requirements of its category [¡Error! Marcador no definido.]. The median of all the peroxide values exceeds the regulatory value when samples had been ageing for 10 months. Therefore, the acceptability limit vector, $\mathbf{Q_A}$, is defined considering the median values of each variable included in the model at month 10 except for the PV which is set at 10

mEq/kg to be compliant with the current regulation. Table 5 shows acceptability limit vector values.

Table 5. Acceptability limit values used for building the scores vector to be applied to the refined seed oil shelf life model.

Variable	Acceptability limit	Variable	Acceptability limit
Peroxide value	10.00	Anisidine value	4.65
K_{270}	4.24	Volat1	0.125
Oxidative stability	4.50	Volat2	0.023
Total tocopherols	665.11	Volat3	0.789
β-Tocopherol	28.66	Volat5	0.044

$\mathbf{Q_A}$ vector was autoscaled and multiplied by **L**, the loading vector of LV1 corresponding to the PLS shelf life model. A value of -0.625 was obtained for the Q_C scalar. This estimated value was interpolated in the regression equation of the multivariate PLS shelf life model to obtain a representative t_C value. 9 months were finally estimated to be a suitable shelf life time for refined seed oils stored under the experimental conditions described in this study.

3.4. Shelf-Life index and ageing rate

Once the shelf life model is established, and t_C is estimated, indexes proposed in Part 1 of this series (**Part 1 of serial**) are calculated. Equivalent ageing time, t_i; is extrapolated from the shelf life model equation by calculating Q_i:

$$Q_i = \mathbf{X_i} \cdot \mathbf{L}$$

where $\mathbf{X_i}$ symbolizes the autoscaled vector of the experimental data of each sample at a specific time, and **L** is the vector of loadings of the first latent variable, LV1, from the PLS model.

Shelf life Index, I_{SL}, reflecting the number of months an oil continues in compliance with the requirements of its category, defined as:

$$I_{SL} = t_C - t_i$$

And finally ageing rate, %Age, calculated as follows:

$$\%\text{Age} = \frac{t_i}{t_C} \cdot 100$$

Similarly, these indices are calculated for the univariate PV-based approach in order to compare the results obtained by both approaches.

Comparison of results (multivariate and univariate approaches). Table 6 shown the results obtained for 5 olive-pomace samples randomly selected from among those included in the study.

Table 6. Estimated shelf life index, I_{SL}, and ageing rate, %Age, values for 5 randomly selected refined olive or pomace oils to be used as examples on the reliability of the predictions.

		Multivariate approach PLS (LV1 scores) model			*Univariate approach Empirical (peroxide value) model*			
Sample	**Ageing time**	**t_i**	**I_{SL}**	**%Age**	**PV**	**t_i**	**I_{SL}**	**%Age**
AV002	2	3.7	10.3	26	2.5	0.8	15.2	5
	4	2.5	11.5	18	3.8	1.7	14.3	11
	6	3.5	10.5	25	6.2	3.8	12.2	24
	8	10.0	4.0	72	15.8	17.7	*-1.7*	*111*
	10	14.3	*-0.3*	*102*	26.7	42.1	*-26.1*	*263*
AV019	2	2.0	12.0	15	5.1	2.7	13.3	17
	4	4.0	10.0	28	7.0	4.6	11.4	29
	6	5.5	8.5	39	8	5.7	10.3	36
	8	7.7	6.3	55	8.8	6.7	9.3	42
	10	8.3	5.7	59	9.1	7.1	8.9	44
	12	12.2	1.8	87	9.2	7.2	8.8	45
	14	16.6	*-2.6*	*118*	12.8	12.5	3.5	78
AV026	2	3.3	10.7	24	2	0.6	15.4	4
	4	5.8	8.2	41	12.0	11.2	4.8	70
	6	10.6	4.4	75	8	5.7	10.3	36
	8	9.2	4.8	66	10	8.3	7.7	52
	10	11.4	2.6	81	10.5	9	7.0	56
	12	18.2	*-4.3*	*130*	10.5	9	7.0	56
	14	18.0	*-4.0*	*129*	9.4	7.5	8.5	47
	16	23.0	*-9.0*	*164*	12	11.2	4.8	70
AV030	14	13.7	0.3	98	10.5	9	7.0	56
	16	16.9	*-2.9*	*121*	20	26.1	*-10.1*	*163*
	24	19.6	*-5.6*	*140*	12.9	12.7	3.3	79
AV031	2	8.1	5.9	58	10	8.3	7.7	52
	4	11.5	2.5	82	14	14.5	1.5	91
	6	13.1	0.9	94	15	16.2	*-0.2*	*101*
	8	11.4	2.6	81	16.3	18.6	*-2.6*	*116*
	10	13.0	1.0	93	19.4	24.8	*-8.8*	*155*
	12	19.6	*-5.6*	*140*	15	16.2	*-0.2*	*101*
	14	19.6	*-5.6*	*140*	15	16.2	*-0.2*	*101*
	16	19.3	*-5.3*	*138*	22.5	31.7	*-15.7*	*198*
	18	23.7	*-9.7*	*169*	21.4	29.2	*-13.2*	*183*
	20	28.6	*-14.6*	*204*	27.5	44.2	*-28.2*	*276*

Values in red italics are indicating that the vegetable oil under consideration is not suitable for consumption.

Considering only the PV, vegetable oil AV002 would no longer comply with the

standard as of month 8 of ageing. However, the multivariate model predicts that this vegetable oil is compliant until month 10. The opposite applies to AV019. The multivariate model detects that vegetable oil is suitable only up to month 14. Considering the PV-based empirical model, the vegetable oil could remain acceptable for 3.5 months beyond month 14. A similar behaviour to the latter is predicted for AV026 vegetable oil.

AV030 is a vegetable oil that has aged for 14 months before being subjected to analytical controls. The multivariate model detects that the oil is already oxidised and close to the limit of acceptability. However the PV-based univariate model predicts that it would remain compliant for another seven months.

Finally, AV031, which would no longer be compliant from month 6 onwards considering only the PV. On the contrary, multivariate model showed that sample is suitable until month 12 of storage, when the analytical values exceed the acceptability limits.

Verifying the shelf life prediction capability (multivariate and univariate approaches). The predicted lifetime from the multivariate shelf life model, taking as a reference the state of each of the oils at months 2, is then compared with the actual date on which the olive-pomace oil do not meet the required quality parameters. Notice that a two-month uncertainty in the predictions is taken into account; this corresponds to the time that elapses from two consecutive analytical controls (experimental work frequency).

In all cases of study, the shelf life indexes at month 2 predict the number of months for each vegetable oil to be consistent with requirements. As example, AV002 is predicted be suitable for an additional 10 months and just after 10 months of ageing it has just exceeded acceptability limit of oxidative-related quality parameters. AV031 is predicted to keep 8 months of shelf life while in month 10 of ageing it presents a shelf life index of 1, about to be considered unsuitable. On the contrary, empirical PV-based model does not perform successfully predictions about suitability of vegetable oils. AV002 is predicted to conserve 15 months of to reach the end of its shelf life when predicted after two months of storing. However, already in month 8 of ageing it presents an ageing rate above 100%, which is not in compliance with the quality parameters corresponding to its category (peroxide value 15.8 mEq/kg).

A similar prediction result comparison between univariate and multivariate seed oil models is missing. As state before, degradation model fitting is not as good, showing a large dispersion of data supposing different rate of degradation of considered samples. Any prediction of oxidation times from a model that does not fit the reality of the data will be biased.

4. CONCLUSIONS

Shelf life of refined edible vegetable oils has been evaluated by considering the changes they undergo as a whole, applying a suitable multivariate approach. Refined vegetable oils obtained from the olive fruit such as pomace and olive (marketed-blend of virgin and refined olive oil) oils showed longer shelf life time that seed obtained oils. The established shelf life regression models are an easy, simple and good tool to assess the actual oxidation status and time remaining in compliance with the accepted requirements for each refined vegetable oil sample. The predictability of the multivariate approach has been proved satisfactory as compared to and empirical model related only to PV. Discrepancies were found in the oxidation rates of oils extracted from different plant seeds particularly during the first months of ageing. Future research will be carried out increasing number of samples of each seed type and modelling analytical data separately, to obtain a better model fitting for purpose.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgments

This work was financially supported by the European Regional Development Fund (ERDF) and Ministry of Science and Technology within the framework of the funding corresponding to 'Retos-Colaboración' (Challenges-Collaboration Program) call of the State Program for Research, Development and Innovation Focussed to Challenges of Society [RTC-2017-6170-2]. In addition, SMR author acknowledges the PhD grant supported by the Ministry of Universities for the Training of University Teachers [FPU 19/02078].

References

COI/T.20/Doc. No 19/ Rev.5 (2019). Spectrophotometric investigation in the ultraviolet. International Olive Council (IOC), Madrid.

COI/T.20/Doc. No 35/ Rev.1 (2017). Determination of peroxide value. International Olive Council (IOC), Madrid.

CXS 19-1981 (2019). Standard for edible fats and oils not covered by individual standards. Codex Alimentarius, Food and Agriculture Organization of the United Nations (FAO), Rome.

CXS 210-1999 (2019). Standard for named vegetable oils CXS 210-1999. Codex Alimentarius, Food and Agriculture Organization of the United Nations (FAO), Rome.

Dieffenbacher A.M., & Poclington W.D. (1992). Determination of antioxidants by high performance liquid chromatography, in: *Standard Methods for the Analysis of Oils, Fats and Derivatives*. 1st supplement to the 7th edition.

European Commission (2019). Regulation (EU) 2019/1604 of 27 September 2019 amending Regulation (EEC) No 2568/91 on the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis. Official Journal of the European Union, L250/14-48.

Hu, M., & Jacobsen, C. (2016). *Oxidative stability and shelf life of foods containing oils and fats*. Academic Press, London.

ISO 280 (1998). Essentials oils – Determination of refractive index. International Organization for Standardization (ISO), Geneva.

ISO 6885 (2016). Animal and vegetable fats and oils - Determination of anisidine value. International Organization for Standardization (ISO), Geneva.

ISO 6886 (2016). Animal and vegetable fats and oils – Determination of oxidative stability (accelerated oxidation test). International Organization for Standardization (ISO), Geneva.

ISO 9936 (2016). Animal and vegetable fats and oils. Determination of tocopherol and tocotrienol contents by high-performance liquid chromatography. International Organization for Standardization (ISO), Geneva.

McGuire, P.C. (2022). Refining vegetable oils: chemical and physical refining. *The Scientific World Journal, 6627013*, 1-10. https://doi.org/10.1155/2022/6627013

Ortega Gavilán, F., Valverde Son, L., Rodríguez García, F.P., Cuadros Rodríguez, C., & Bagur González, M.G. (2020). Homogeneity assessment of reference materials for sensory analysis of liquid foodstuffs. The virgin olive oil as case study. *Food Chemistry, 322*, 126743, 1-12. https://doi.org/10.1016/j.foodchem.2020.126743

Roszkowska, B., Tanska, M., Czaplicki, S., & Konopka, I. (2015). Variation in the composition and oxidative stability of comercial rapeseed oils during their shelf life. *European Journal of Lipid Science and Tecnology*, *117,* 673-683. https://doi.org/10.1002/ejlt.201400271

Statista (2022). *Consumo doméstico de los principales aceites vegetales en el mundo en la campaña de 2021/2022 según tipo*. Retrieved from: https://es.statista.com/estadisticas/564768/consumo-domestico-de-los-principales aceites-vegetales-segun-tipo/

Tatum, V., & Chow, C.K. (2008). Effects of processing and storage on fatty acids in edible oils. In C.K. Chow (Ed.), *Fatty acids in foods and their health implication,* 3rd edition, pp. 493-510, CRC Press.

Van Boekel, M.A.J.S. (2008). Kinetic modelling of food quality: a critical review. *Comprehensive Reviews in Food Science and Food Safety, 7*, 144-158. https://doi.org/10.1111/j.1541-4337.2007.00036.x

PUBLICACIÓN IX.

ARTÍCULO CIENTÍFICO

A comparison of the stability of refined edible vegetable oils under frying conditions: multivariate fingerprinting approach

Food Control, (under review)

A comparison of the stability of refined edible vegetable oils under frying conditions: multivariate fingerprinting approach

Sandra Martín-Torres✉, Antonio González-Casado, Miriam Medina-García, M. Soledad Medina-Vázquez, Luis Cuadros-Rodríguez

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

Highlights

- Refined vegetable oils were compared in term of oxidative stability after frying
- Oils which contain additives and oil which do not were considered
- Total polar compounds were monitored using a dielectric constant sensor
- GC-MS fingerprinting approach was used to compared stability for each vegetable oil

ABSTRACT

The stability of highly consumed vegetable refined oils after discontinuous frying of potatoes was compared. Vegetable oil samples were evaluated for refractive index, anisidine value, peroxide value, UV absorbance and dielectric constant-based determination of the content of total polar compounds. Changes along frying time were recorded so multivariate modelling of data was carried out. A new gas chromatographic-mass spectroscopy method was intended to record a fingerprint of both polar and non-polar (after silica cartridge separation) compound fractions. Modelling of fingerprints was also developed and similar results were statistically verified. Models allowed differentiating among the kind of vegetable oils and the stability against oxidation; (also between vegetable oils containing synthetic antioxidants and/or anti-foaming agents and vegetable oils which do not. The oxidative stabilities of the edible vegetable oils were in order olive oils > marketed-blends of seed oils > sunflower oils. Olive oils were found to have the highest natural thermo-oxidative stability compared with other seed oils but not significantly different to pomace oils and seed oils containing synthetic additives.

Key words

Multivariate deterioration monitoring, multiparametric-kinetic studies, chemometrics, shelf life, quality and stability indices, extra virgin olive oil.

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

1. Introduction

Edible vegetable oil constitutes a complex multi-component food matrix, the analysis of which is not an easy task. It is difficult to determine with any degree of certainty all the minority constituents in a vegetable oil matrix, due to their chemical complexity and low concentration. In addition, some of these minority constituents are only present in the crude oil, and technological processing such as refining remove them (**Bosso, 2017**). A reliable and accurate determination of vegetable oil composition is essential to assess its quality and authenticity. Overall, the quality of an edible oil is related to its oxidative stability, storage history, sensory characteristics, nutritional properties and culinary use (**Dais & Hatzakis, 2013**).

Among methods of cooking, deep-frying is highly popular food processing method to produce palatable and desirable foods with unique characteristics of flavour, odour and colour. The frying oil acts as a homogeneous heat transfer liquid medium and contributes to the development of specific texture and flavour of fried foods. Oxidative and hydrolytic deterioration of the vegetable oil occurs during frying: hydrolysis of lipids, breakdown of double bond, and oxidation of the unsaturated bonds of fatty chains due to the presence of oxygen, the moisture, the high oil temperature, and the leaching of components from the food (**Upadhyay, Sehwag & Mishra, 2017**). This results in the generation and accumulation of several degradation products, volatile or non-volatile compounds. Most of the volatile compounds are produced during the frying process and lost in the atmosphere with the steam. However, remaining non-volatile decomposition products, collectively known as total polar compounds (TPC), produced primarily by thermal oxidation and polymerization reactions of unsaturated fatty acids, they are of special concern due to accumulation in the frying oil, which promotes further degradation, and the fact that they are absorbed by the fried food (**Arslan et al., 2013**).

Standards or guidelines have been established in many countries to ensure the high quality of fried foods. The most widespread limitation states that used frying oil should be replaced when TPC, representative of the new compounds formed during frying, is approximately 25% (w/w) (limited value in Spain (**Government of Spain, 1989**), Belgium, France, Portugal and Italy legislation, while 24% is set in Germany and 27% in Australia, China and Switzerland (**Chen et al., 2021**). Determination of TPC is usually performed by a previous silica gel column chromatographic fractionation followed by a further gravimetric analysis, according to the standard methods of IUPAC (**IUPAC, 1992**) and AOAC (**Flores, Avendaño, Bravo, Valdés & Firestone, 2013**). However, the reference methods are chemical reagents and time-consuming and need reasonable experience. Because of that, a series of attempts have been done to find simple and rapid substitute methods. Alternatively, there are

commercialized tests to determine the TPC of frying oils is usually determined by monitoring their dielectric constant since an increase in the concentration of polar molecules instantly causes an increment in the dielectric constant of the vegetable oil matrix. This electrochemical method requires simple and inexpensive equipment, and allows to quickly addressing issues such as healthy intake estimations and safety assessments (**Flores et al., 2021**). However, these devices, which need specific and careful calibration, are not suitable to be employed for all type of vegetable oils. In addition, temperature could influence the measurements and interferences such as water, salt, and minerals could affect the polarity of samples and could give false information about the TPC content (**Cascant, Garrigues & de la Guardia, 2017**). Other alternative methods have been developed to determine the TPC content of edible oils; in this respect, FTIR is an attractive and growing analytical technique, although there are still challenges to overcome before it can be used with sufficient confidence (**Chen et al., 2021**).

Application of chemometric approaches in testing the quality status of frying edible oils has not been attempted widely (**Martín Torres, Ruiz Castro, Jiménez Carvelo & Cuadros Rodríguez, 2022**). Thorough this study the time-dependent monitoring of the formation of polar and polymeric compounds and changes in physicochemical parameters over time was performed. Discontinuous chips-batch deep-frying process (over several heating and cooling cycles) was simulated. The most commonly used edible vegetable oils, both containing synthetic antioxidants and anti-foaming agents (since they have been historically added to frying oil to extend its shelf life) and not, were used in the simulated frying processes. A characteristic instrumental 'fingerprint' combining the polar and non-polar fractions of the oils was obtained using gas chromatography-mass spectrometry (GC-MS). A multivariate study of the stability and shelf life of these vegetable oils has been carried out by comparing chromatographic fingerprints and the results corresponding to classical determinations such as the peroxide value and the anisidine index, among others, together with the use of proper chemometric tools. As a result of this study, conclusions will be drawn about the suitability of using one type of oil or another for the usual domestic deep-frying process.

2. Materials and methods

2.1. Samples

The test materials used during the study were 4 sunflower oils (SO), 2 olive-pomace oils (PO), 2 olive oils (marketed-blends of virgin and refined olive oil) (OO) and 2 seed oils blends (constituted by high oleic sunflower, corn and

soybean) (BO). Among them, 3 sunflower oils and 1 blend oil contain synthetic antioxidants and/or anti-foaming agents (vitamin C, vitamin E, ascorbyl palmitate, propyl gallate, dimetilsiloxano). All of them were purchased from local supermarket and grocery shops; they were all qualified products.

2.2. Experimental procedure

During the frying experiments, 1.5 L of edible vegetable oil was placed in an electric domestic stainless clean fryer (1000W, Cecotec, Spain) and heated up to 180 ± 10 ºC (manufacturer's recommended temperature for chips frying). When the temperature reached the stated value, 200 g of strip-cut fresh and raw potatoes (approximately 10 mm × 10 mm × 5 mm thick, previously paper-dried) were placed in the basket fryer and immersed in the hot vegetable oil and fried for 5 min. During the frying process, the lid of the fryer was open (as advised in the fryer's instructions). After 5 min, the chips were removed from the fryer and the operation was continued for a new potato batch during 2 hours (200 g of potatoes each 30 minutes). After 2 hours of frying the fryer is turned off and the oil is allowed to cool below 100 ºC. A 10.00 g sample of vegetable oil, exactly weighted, is previously taken to a new heating cycle.

The heated vegetable oil sample was taken after each two hours, cooled at room temperature and the whole process is repeated. The potato batches were fried successively in the same vegetable oil for consecutive frying treatments until stopped when the regulatory limit of 25% polar content of the oil is exceeded or the oil content into the fryer is below the lower limit necessary to ensure that the potato strips are immersed for at least 2 cm. As summary, the frying experiment involved intermittent heating of the vegetable oil for two hours, at 180 ± 10 °C and exposed to air, without renewing the oil throughout the frying cycles.

Immediately after cooling, each vegetable oil aliquot was packed in an amber glass-vial, stored in a refrigerator and kept at 4°C until analysis in order to avoid further chemical changes.

2.3. Analytical equipment

Abbe refractometer ORT1RS and Mettler Toledo G20 Compact Titrator were used to measure refractive index and peroxide value, respectively. Molecular UV absorption values were measured using an Agilent 8453 spectrophotometer. Agilent Technologies 7820A gas chromatographer equipped with an autosampler (7693) and a mass spectrometer (5977B MSD), using Rtx-65TG (30 m × 0.25 mm i.d., 0.1 μm) capillary column, which was applied for chromatographic analyses. An oil monitor meter FOM320 Ebro

device was employed to quickly measure total polar compound (%TPC). Technical data provides by the device supplier ensures good correlation with the recognised chromatographic method for the most common oils running in the 'semiliquid' equipment mode (**Flores, Avendaño, Bravo, Valdés & Firestone, 2013**), and its performance was periodically checked with a reference oil (Testo 05542650).

2.4. Physical-chemical analyses

Different methods have been developed to evaluate the quality of frying oil. The assessment methods rely on the measurement of physical changes and particular chemical composition, which are arising as a consequence of the deteriorative reactions (**Al-Khusaibi & Rahman, 2021**; **Mudawi, Elhassan & Sulieman, 2014**).

Refractive index (RI)

Refractive Index (RI) refers to the ratio of light in a vacuum to its velocity in a specified medium. The RI was determined as described by the ISO 280 standard (**ISO, 1998**). A double-prism was opened and few drops of oil were placed on the prism and the determination of RI was done using a reference temperature of 20 ± 2 ºC. The refractometer was cleaned between readings by wiping off the oil with a smooth tissue paper; the prism was regularly cleaned with petroleum ether and then allowed to dry by use of a clean tissue. The process was repeated twice.

Peroxide value (PV)

Peroxide value (PV) is the most popular index concerning the oxidation of oils. It is a direct measure of the primary oxidation taking place in oils. Oil peroxide values were determined according to the recognised methods described in COI/T.20/Doc. No 35 (**COI, 2017**) by potentiometric titration.

Anisidine value (AV)

Different types of oxidized triacylglycerols can be produced from lipid hydroperoxides during secondary oxidation such as aldehydes, epoxides and ketones. As a measure of secondary oxidation products, anisidine value (AV) is used to assess thermally stressed oils, which indicates the concentration of aldehydes present in the oil. AVs were determined according to the recognised method described in ISO 6885 standard (**ISO, 2016**). A solution of the vegetable oil in hexane is reacted with *p*-anisidine in glacial acetic acid and the absorbance is measured at 350 nm both before and after reaction.

UV absorption

Specific absorptivities K_{232}, K_{268} and $\Delta K_{268\pm4}$ were determined in compliance with COI/T.20/Doc. No 19 (**COI, 2016**). Absorptivities at 232 nm are related to the formation of hydroperoxides (primary stage of oxidation) and conjugated dienes (intermediate stage of oxidation), while absorptivities at 268 nm are associated with the formation of carbonyl compounds (secondary stage of oxidation) and conjugated trienes (technological treatments), respectively. Both metrics have been proved to increase during frying process (**Abenoza, De las Heras, Benito, Oria & Sánchez Gimeno, 2016**).

Total polar compounds (TPC)

Monitoring vegetable frying oil for total polar compounds (TPC) will ensure that the product is of consistent quality, and will not have a negative impact on consumer health. Extended use of rapid tests correlated with standard methods is crucial in the fast-food segment, characterized by the practice of discontinuous frying. Because of that, the %TPC was directly measured in hot vegetable oil by using the TPC meter FOM 320. It is a direct measuring device that allows measuring the quality status of the vegetable oil through the actual dielectric constant. It incorporates a sensor that is submerged in the vegetable oil subjected to the frying process for one minute, yielding a value of %TPC (**Ebro, 2021**).

2.5. Chromatographic analysis

Chromatographic techniques have been successfully applied to characterize the products formed during food-frying or during a simulated frying process (**Zhang, Qin, Li, Shen & Saleh, 2015**). Polar compounds produced during the frying process include, among others, hydrolysis products, oxidized decomposition compounds, oxidized triacylglycerol monomers, oxygenated triacylglycerol polymers, and oxidized sterols. The measurement of TPC by chromatographic techniques has been extensively reported. More sophisticated methods include complex sample pre-treatments and are only used to analyse the known products. A number of earlier trace amounts of undetected products need to be investigated using more efficient detection techniques. Moreover, some nonpolar compounds such as unchanged triacylglycerols, cyclic products, and trans-isomers, also present in the frying system, are also closely related to the quality of frying vegetable oils or fried foodstuffs in respect of taste and safety (**Brühl, 2014**).

Therefore, a simple method based on GC-MS non-targeted analysis is proposed to monitor the vegetable oil fraction of polar and non-polar compounds produced in the frying process. As result, two particular chromatographic fingerprints are acquired for each frying vegetable oils which will subsequently be considered in order to reach conclusions on their shelf life.

Previously to chromatographic analyses, the non-polar and polar fractions were separated by silica-solid phase extraction cartridges (SPHE-S61-030). 250 mg of oil were diluted in 1 mL of n-hexane (methyl oleate 8 mg/mL was added as internal standard). The non-polar fraction, containing the non-polar triacylglycerols and the internal standard (IS), was eluted with 10 ml of n-hexane/diethyl ether (90:10, v/v). A second fraction, which comprises the total polar compounds, was eluted with 10 ml of ethyl ether. After solvent evaporation, the contents (% w/w) of the nonpolar fraction are determined gravimetrically. Subsequently, polar fraction is estimated by differential weighing.

After that, polar fraction was diluted in 1 mL of chloroform while 20 mL of chloroform were needed to dilute non-polar fraction, filtered (PTFE 0.22 µm) and both fractions were chromatographically analysed individually, using the same chromatographic method. Separation was performed on a Rtx-65TG capillary column (65% diphenyl – 35% dimethylpolysiloxane; 30 m × 0.32 mm i.d. × 0.1 µm, Restek). 2 µL of diluted polar and non-polar fraction (for each oil aliquot), split ratio 10:1, were injected at 320 ºC. Oven temperature was increased from 200 ºC to 290 ºC in 3 min and then raised to 370 at 10 ºC/min and hold during 6 min to a total run time of 17 min. MSD transfer line was set at 350 ºC. Solvent delay of 1.20 minutes, starting m/q of 50 and ending m/q of 1000 were fixed in the detection step. Efficiency of the separation and analysis process was checked out by relative area of internal standard. Run speed and low solvent consumption are the major advantages of this method considering separation and chromatographic stages.

2.6. Chemometrics

Multivariate data evaluation was carried out using MATLAB (R2017b version, The Mathworks Inc. Natik, MA) and PLS_Toolbox (Eigenvector Research Inc., Wenatchee, 193 WA).

Principal component analysis (PCA) is a well-known technique to visualize data and to reduce the real dimension of a data set. PCA is aimed at finding the simplest mathematical model able to describe the data set satisfactorily. It looks for a smaller number of underlying new variables, named principal components (PCs), which explain most of the variability exhibited by the larger number of measurements made on the objects/samples. It is an unsupervised method because it does not require training input to find the output: no additional knowledge (*e.g.*, y-variable) besides raw data (x-variable) is needed to describe the data set. The significant PCs can be used in place of the original variables for successive treatment or to visualize the information contained in the data set. Loadings are the estimated coefficients that define the linear combination of

the original variables originating each principal component. Scores are the projections of the objects on the new PC axes (**Abdi & Williams, 2020**).

Partial least-squares (PLS) regression is a multivariate linear regression technique aimed at fitting cause-effect relationships. PLS computes latent variables (LVs), linear combinations of predictors that are similar to PCs and finds the maximum correlation direction between response and LVs. PLS is a supervised multivariate method because, apart from the information on the x-variables measured, the available knowledge on a dependent y-variable (in this study, storage time) is applied. PLS simultaneously maximizes both the LVs variance and the correlation with respect to the y-variables. Complexity of the model depends on the number significant LVs, according to the maximum percentage of cross-validated explained variance (**Cocchi, Biancolillo & Marini, 2018**).

3. Results and discussion

As mentioned above, each oil sample was exposed to a different number of frying hours until exceeded the regulatory limit of 25% TPC (measured with FOM 320) or up to a maximum of 32 hours when the remaining vegetable oil result insufficient to carry out a new frying cycle in a suitable way. Figure 1 bar graph shows comparison among discrete type of vegetable oil %TPC content concerning frying hours. Different performance can be observed with regard to the type of vegetable oil and whether they contain synthetic additives or not.

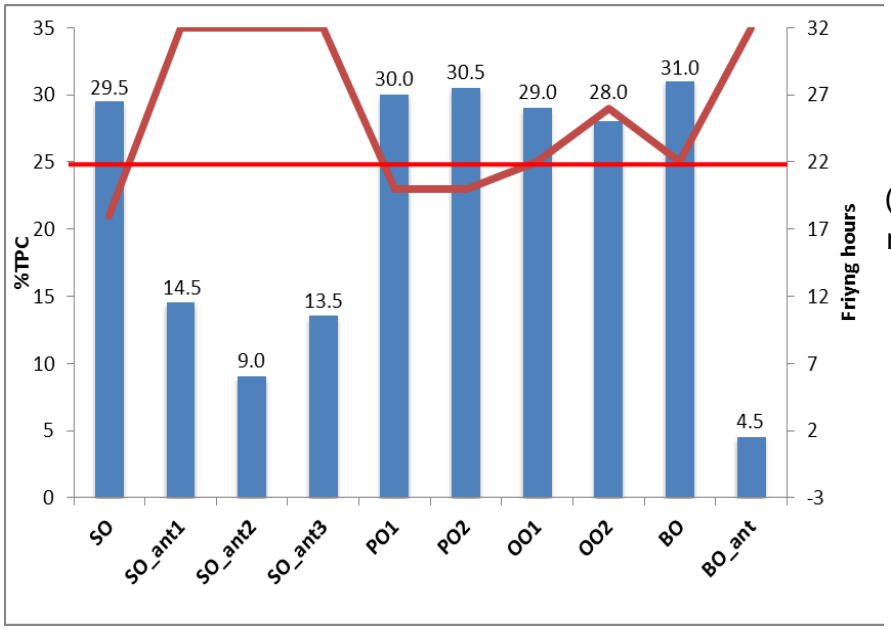


Figure 1. Bar graph of %TPC with respect to frying hour (represented by the red line) of different oil samples: sunflower oil (SO), pomace oil (PO), olive oil (OO) and blended oil (BO) (ant is referring to synthetic antioxidants and anti-foaming's).

SO reach the higher %TPC value in the shortest time, similar performance is shown by PO and BO oils, which allowed one additional frying cycle (compared with SO) without failing to meet the requirement. On the contrary, OO showed the higher natural resistance to oxidation needing the higher time to exceed the regulatory limit. PO containing synthetic additives do not reach 25% TPC even after 32 hours of frying.

Multivariate modelling. As introduced, temperature and interference substances could influence the polarity of samples and can give false information about %TPC content in the dielectric constant-based measurement of heated vegetable oil. Multiparametric monitoring of the kinetic of oxidation and multivariate stability models have been proved to be more generically applicable than single-parameter models (**Martín Torres, Ruiz Castro, Jiménez Carvelo & Cuadros Rodríguez, 2022)**.

Slight increase of RI of frying vegetable oils (as also reported for soybean oil (**Tyagi & Vasishtha, 1996**), as well as PV, which changes from 6 to 11 mEq/Kg (formely, milliequivalents of active oxygen per kilogram of vegetable oil) for OO, were recorded. AVs highly differ between heating oils and fresh oils. UV absorbance measurements notably increased during frying finding the lowest value for olive oils and the highest one for sunflower oil, in agreement with other results (**Abenoza, De las Heras, Benito, Oria & Sánchez Gimeno, 2016**).

Multivariate analysis of discrete experimental data was carried out. Data were arranged in a 106×7 matrix: each row corresponds to each oil aliquot taken (from 0 to total frying hours, respectively). Each column corresponds to the value of a single determined parameter: %TPC, RI, PV, AV, K_{232}, K_{270}, and ΔK.

Data were autoscaled (subtraction of the mean and division by standard deviation for each variable) before PCA computation to assign the same numerical weight to each variable. 4 PCs were selected which explained 90% of cumulative variance. Figure 2a shows PC1-PC2 scores plot. PC2 scores allow grouping for vegetable oil types. The PC1 (41% of variance) seemed to grouped and order the samples according to frying time from negative scores for unheated oils to positive as heating hours increase. If only the behaviour of PC1 scores against the sample number is considered (see figure 2b), as frying time increases, different slopes of scores can be observed depending on the vegetable oil type and the presence/absence of synthetic additive.

Similar results were obtained by PLS modelling when the numbers of hours of frying are used as dependent y-variable. LV1 scores (49% of variance for y-variable) increase with frying time but this rise is not uniform and varies depending on the vegetable oil. Differences on data fitting could be also observed when a linear regression model (see figure 3) is fitted between predicted variables and frying time (R^2_{cal} 0.59).

a.

b.

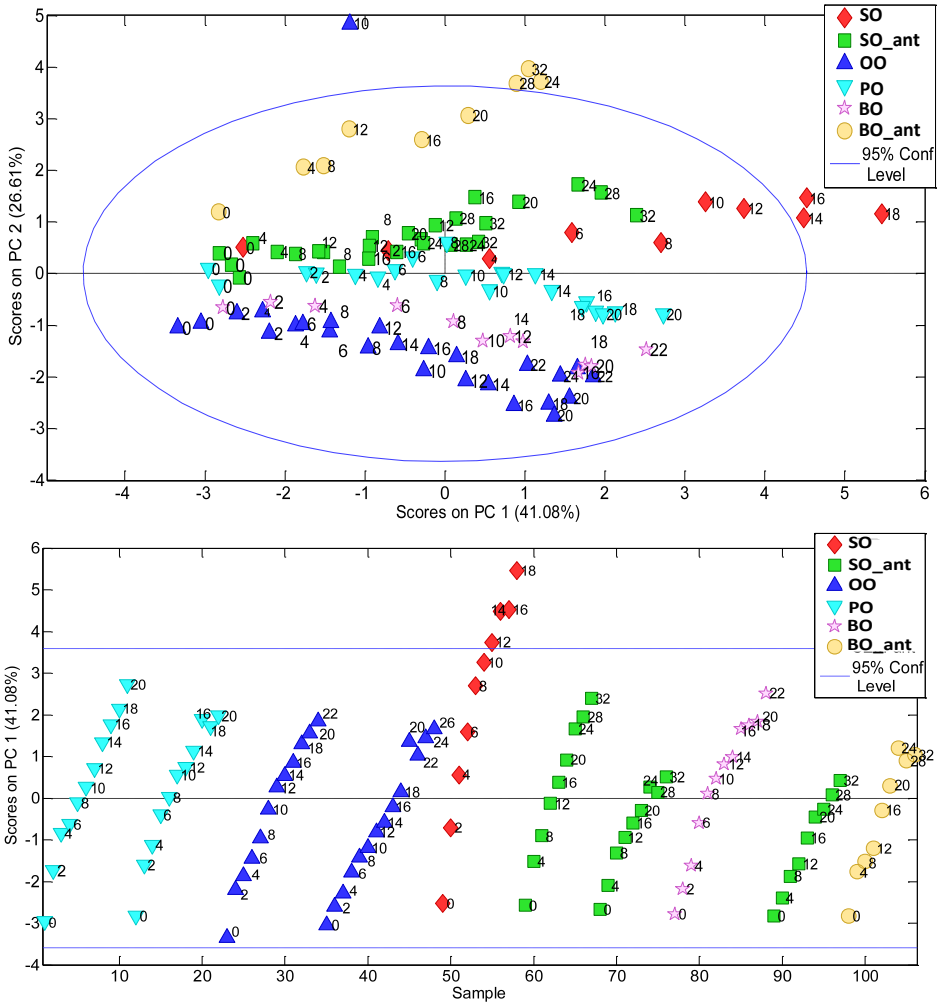


Figure 2. a. PC2-PC1 scores plot from the physico-chemical analytical data values modelling. **b.** PC1 scores against the sample number plot clearly showing the frying time dependence of each type of vegetable oil.

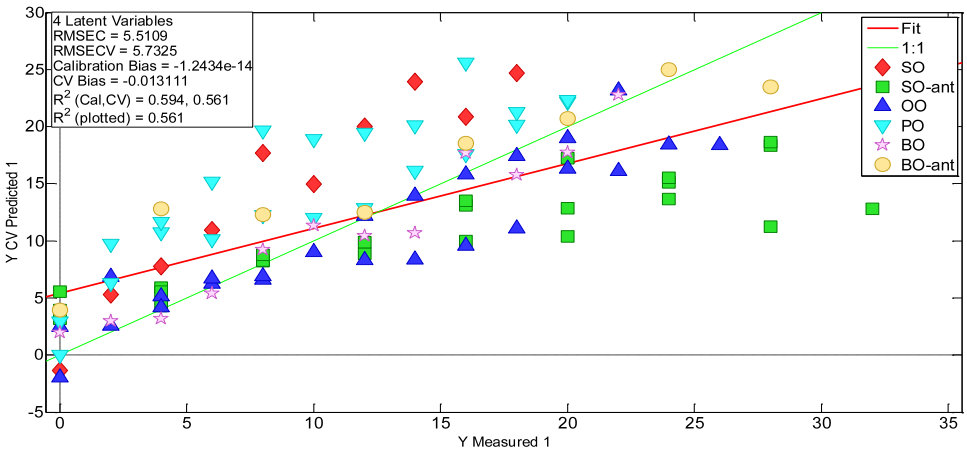


Figure 3. Relationship between dependent predicted discrete variables and independent variable Y (frying hours) by PLS method.

Fingerprinting. Pre-processing of chromatographic fingerprints was carried out. A low-level data fusion strategy was developed: total ion current (TIC) chromatograms from polar and non-polar fractions of same oil aliquot were concatenated on a single chromatographic fingerprint (1466×2 variables, total fictitious retention time of 32 min). Figure 4 shows a comparison between a polar and non-polar fussed fingerprint obtained from a olive oils after 2 and 22 hours of frying, respectively.

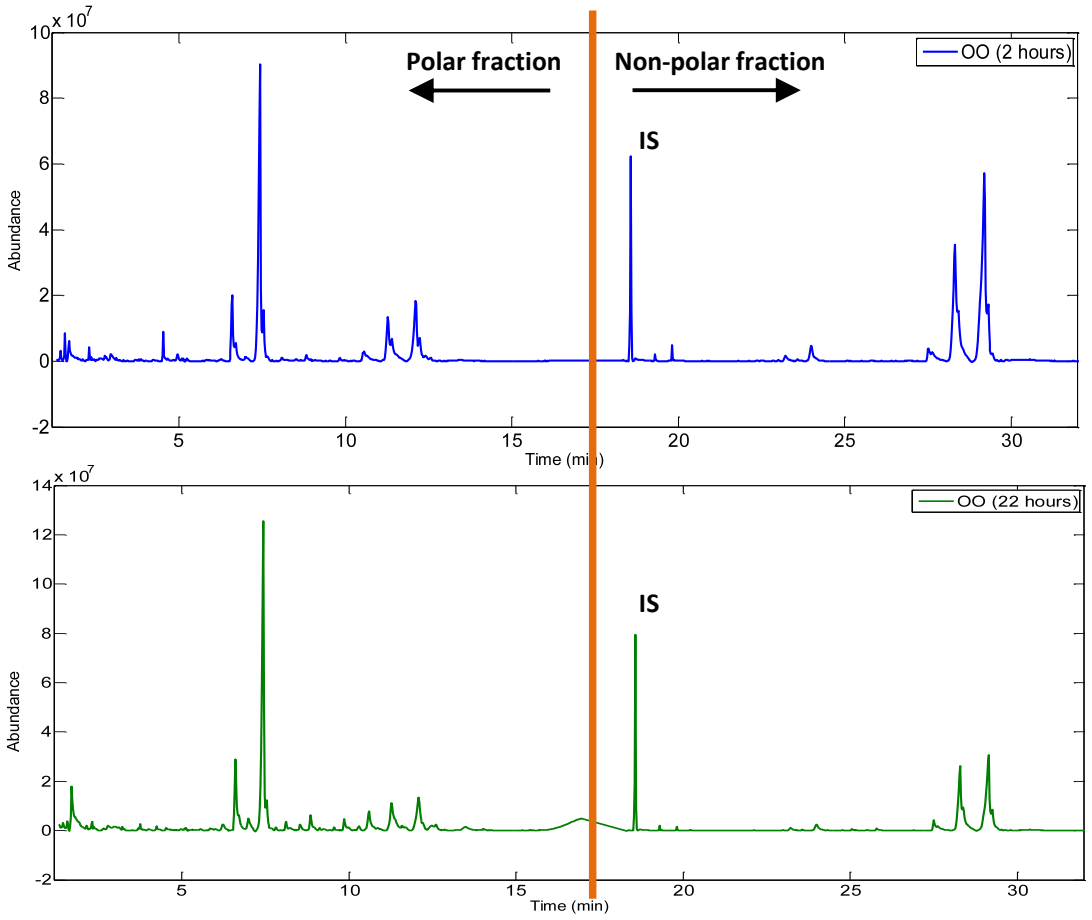


Figure 4. Fussed fingerprints of an olive oil (OO) after two and twenty-two hours of frying respectively (IS symbolizes the internal standard).

Sample fingerprints were embedded in a data matrix and subsequently pre-processed using a home-made MATLAB function, named 'MEDINA' (version 07) (**Pérez Castaño et al., 2015**) and PLS toolbox preprocessing available methods. Indeed, each fingerprint were baseline-corrected using 'Whittaker filter' (Λ=100, p=0.001), overlaid and filtered, normalised of intensities with respect to the internal standard, aligned using 'icoshift' algorithm, and finally mean centred. A new PCA unsupervised model was set up using the entire pre-processed fussed chromatographic fingerprints as input data set. 7 Pcs were

selected explaining 73.48% of the cumulative variance. Differences were found regarding both to frying time and type of vegetable oil. As example, figure 5 shows the PC2-PC5 scores plot in which differences are clearly displayed. Similar results were found by PLS modelling with regard to frying time (R^2_{cal} 0.78).

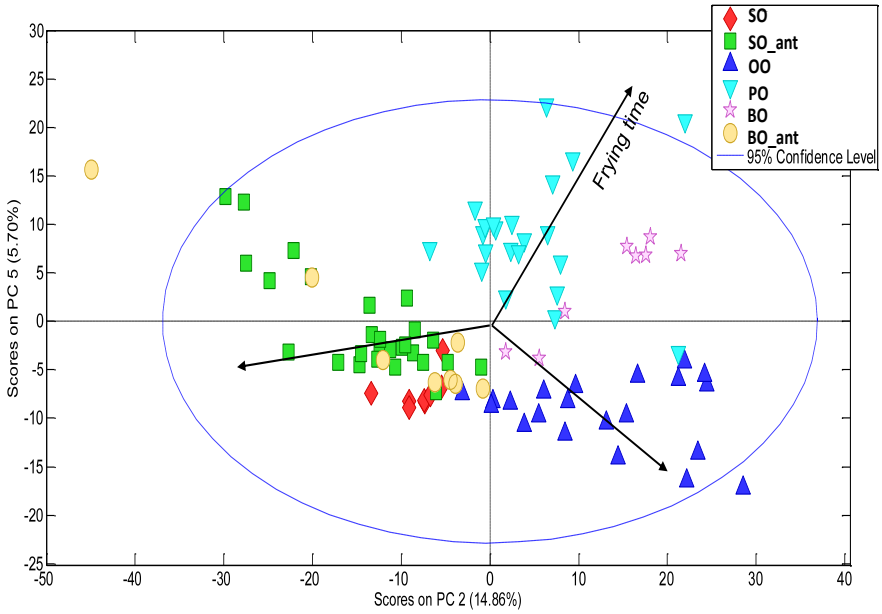


Figure 5. PC2-PC5 scores plot from fussed chromatographic fingerprinting data modelling.

Comparison between physico-chemical variables-based and chromatographic fingerprinting-based multivariate models. In consideration with these findings, two different PLS models were set up for each vegetable oil: the recorded physico-chemical variables (discrete variables, including %TPC measured by the rapid tester) and chromatographic fingerprints were modelled separately, using frying time as input y-variable for each type of oil. Some performance metrics of both models was evaluated and main results are shown of table 1.

The slopes of the regression lines obtained from both methods must be compared to ensure equivalence and, therefore, that the developed GC fingerprinting method provides equal information as all other experimental measurements. Assuming that analytical signals are normally distributed, the comparison of the slopes of two regression lines could be performed by means of a Student t-test statistic whose null hypothesis is defined by $H_0 \equiv \beta_1 = \beta_2$, in which β_1 and β_2 symbolizes the slopes. More comprehensive details and full mathematical treatments, which are out of scope of this work, can be found elsewhere (**Andrade & Estévez Pérez, 2014**). Table 2 shows the statistics on the application of t-test formulation to the specific models. Firstly, F-test was applied to compare the squared standard errors (or residual variance, $s_{y/x}^2$) of

the two regression lines of each example. In all cases, p-values were higher than 0.05 or 0.1 so the null hypothesis could not be rejected (the slopes are nor significantly different) regardless of the significance level being considered.

Table 1. Metrics of PLS methods carry out for each vegetable oil type.

	Fingerprinting modelling			Discrete variables modelling		
Vegetable oil	PLS model	RMSE CV	Y CV predicted *vs* frying time	PLS model	RMSE CV	Y CV predicted *vs* frying time
Sunflower oil (SO)	3 LV 73.2% X-var 99.4% Y-var	3.2145	y = 1.0553 x + 0.5364 R^2 = 0.817	1 LV 80.1% X-var 95.8% Y-var	1.3383	y = 0.9583 x + 0.3689 R^2 = 0.951
Pomace oil (PO)	3 LV, 65.3% X-var 99.5% Y-var	4.2901	y = 0.9621 x + 0.8200 R^2 = 0.681	2 LV 73.6% X-var 98.2% Y-var	1.3256	y = 0.9891 x - 0.0552 R^2 = 0.960
Blended oil (BO)	2 LV 62.2% X-var 97.2% Y-var	2.9666	y = 0.8845 x + 1.337 R^2 = 0.945	2 LV 78.9% X-var 98.9% Y-var	2.7933	y = 0.8518 x + 0.7544 R^2 = 0.872
Olive oil (OO)	2LV 48.2% X-var 88.1% Y-var	3.0873	y = 0.7670 x + 3.1144 R^2 = 0.833	4 LV 98.6% X-var 68.1% Y-var	6.8781	y = 0.6653 x + 2.5972 R^2 = 0.378
Sunflower oil containing additives (SO_ant)	5 LV 73.8% X-var 98.9% Y-var	4.5728	y = 0.8245 x + 3.1474 R^2 = 0.775	2 LV 81.9% X-var 89.6% Y-var	4.2074	y = 0.8256 x - 2.9714 R^2 = 0.741
Blended oil containing additives (BO_ant)	3 LV 8.8% X-var 97.3% Y-var	3.8987	y = 0.9804 x - 0.0605 R^2 = 0.814	2 LV 90.1% X-var 97.2% Y-var	3.7846	y = 1.0238 x - 1.2166 R^2 = 0.892

Table 2. Statistical comparison of results obtained by fingerprinting modelling and discrete variable modelling.

Vegetable oil	Model	Slope ± s_b	$S_{y/x}$	p-value
Sunflower oil (SO)	Fingerprinting	1.055 ± 0.189	3.42	0.645
	Discrete variables	0.985 ± 0.082	1.49	
Pomace oil (PO)	Fingerprinting	0.962 ± 0.150	3.84	0.865
	Discrete variables	0.989 ± 0.046	1.37	
Blended oil (BO)	Fingerprinting	0.884 ± 0.081	1.65	0.827
	Discrete variables	0.851 ± 0.115	2.66	
Olive oil (OO)	Fingerprinting	0.767 ± 0.076	3.16	0.619
	Discrete variables	0.665 ± 0.177	6.55	
Sunflower oil containing additives (SO_ant)	Fingerprinting	0.824 ± 0.088	4.77	0.993
	Discrete variables	0.825 ± 0.073	3.94	
Sunflower oil containing additives (SO_ant)	Fingerprinting	0.980 ± 0.177	5.49	0.848
	Discrete variables	1.023 ± 0.134	4.17	

s_b = standard deviation of slope; $s_{y/x}$ = residual standard deviation; p-value: calculated following the t-statistic Welch's test (**Andrade & Estévez Pérez, 2014**) when the null hypothesis of the F-test can be rejected.

Stability modelling. Scores of LV1 from fingerprinting modelling, which clearly were time-structured, were used to establish a stability model for each vegetable oil. Linear fitting was proved satisfactory when plotting LV1 *vs.* frying time. Regression goodness of fit summary is shown on table 3. Slopes of regression models were compared in order to draw conclusions about the relative stability of each type of oil and thus reach a decision on the suitability of each particular vegetable oil for discontinuous (domestic) deep-frying process. Clearly, seed oils (SO and BO) degradation occurs the fastest. Notice that BO containing additives (anti-foaming agents) show higher resistance to degradation, similar to that shown by the PO and OO. The best performance against oxidative reactions is revealed by SO containing antioxidants. These results differ from the initial univariate conclusions by %TPC measurements from the sensor tester. %TPC content was much lower for vegetable oils containing additives at higher number of frying hours. On the contrary, by modelling the stability from the chromatographic fingerprints, appears that the oxidation behaviour is similar to that of olive oils (and so similar to pomace oils). This confirms that model developed from the fussed polar and non-polar chromatographic fingerprints is equivalent to the multivariate modelling of all

physico-chemical parameters described (discrete variables, including %TPC by the rapid tester). Figure 6 is showing LV1 scores plotting *vs.* frying time.

Table 3. Goodness of fitting statistics for stability models of edible vegetable oil.

Vegetable oil	Regression model	R^2	RMSE
Sunflower oil (SO)	t = 3.665 · LV1 - 35.89	0.9082	7.989
Pomace oil (PO)	t = 3.600 · LV1 - 38.40	0.9443	6.744
Blended oil (BO)	t = 2.053 · LV1 - 23.48	0.8835	7.818
Olive oil (OO)	t = 1.996 · LV1 - 27.17	0.8089	7.526
Sunflower oil containing additives (SO_ant)	t = 0.815 · LV1 - 19.97	0.6297	16.11
Blended oil containing additives (BO_ant)	t = 2.171 · LV1 - 33.91	0.8419	11.02

t = frying hours; LV1 = scores of the latent variable 1 from PLS model; R^2 = coefficient of determination; RMSE = root mean squared error.

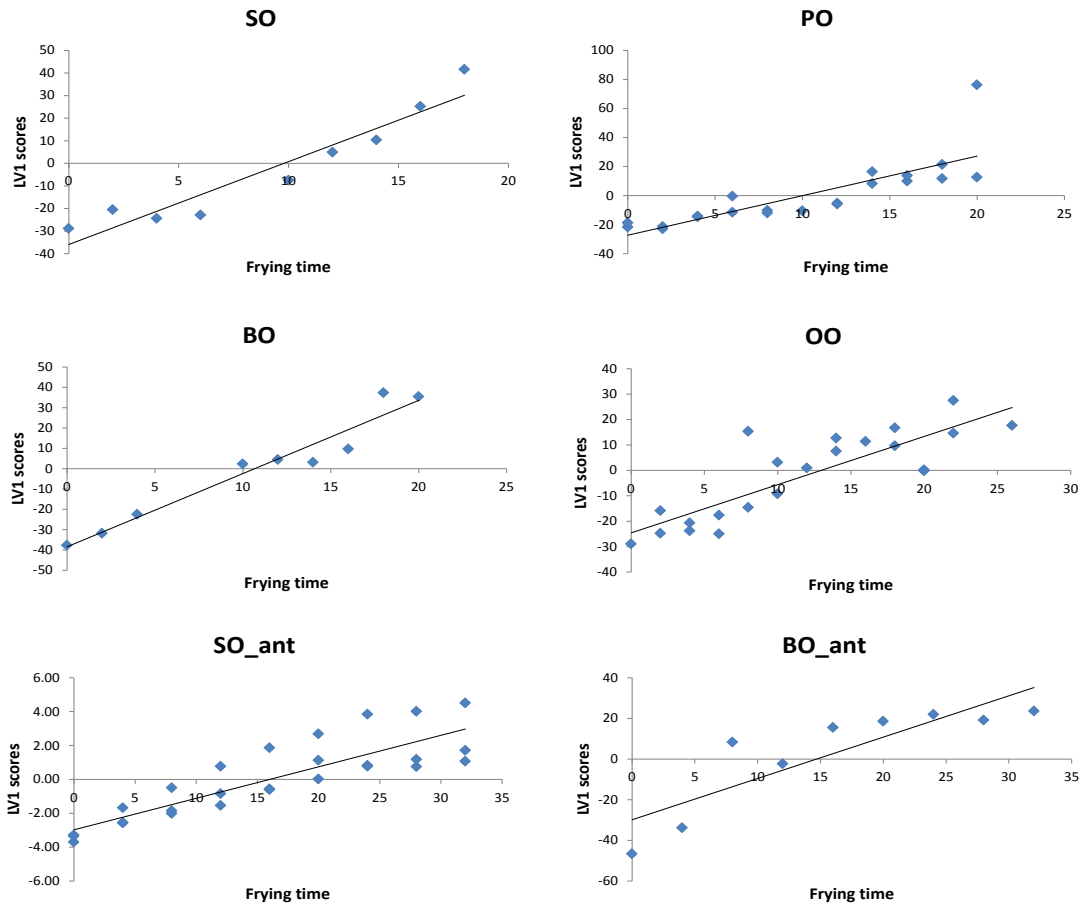


Figure 6. LV1 scores plotting vs. time of frying. SO is sunflower oil; PO is pomace oil; BO is blend of different seed oil; OO is olive oil; and _ant symbolises an oil containing additives.

4. Conclusion

This study described the frying stability of most commonly consumed edible vegetable oils. A chromatographic non-targeted fingerprinting methodology has been proposed for monitoring of polar and non-polar fractions produced during frying process. The increase in the contents of total polar compounds, refractive index, peroxide value, anisidine value and UV absorptivities were found to be dependent to the frying time. Multivariate modelling of discrete measurement leads to the same conclusions as modelling of fingerprinting data. The chemometric PCA and PLS models allow to differentiate among types of vegetable oils and stability against oxidation. In addition, vegetable oils containing synthetic antioxidants and/or anti-foaming agents may be distinguished from the ones do not.

The results obtained in this study were found to be in good agreement with the literature findings. From the chromatographic fingerprint data, olive oils were found to have the highest thermo-oxidative stability compared with seed oils, such as sunflower oils or blended oils. However, this difference is not significant with regard to pomace oils and those seed oils containing synthetic additives. Nevertheless, data analysis revealed that univariate approach models, *i.e.*, rapid measurement of %TPC using a sensor tester, may lead to wrong conclusion about stability and nutritional shelf-life of vegetable oils.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgments

This work was financially supported by the European Regional Development Fund (ERDF) and Ministry of Science and Technology within the framework of the funding corresponding to 'Retos-Colaboración' (Challenges-Collaboration Program) call of the Spanish State Program for Research, Development and Innovation Focussed on the Societal Challenges (RTC-2017-6170-2). In addition, SMR author acknowledges the PhD grant supported by the Ministry of Universities for the Training of University Teachers (FPU 19/02078).

References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *WIREs Computational Statistics, 2*(4), 433-459.
https://doi.org/10.1002/wics.101

Abenoza, M., De Las Heras, P., Benito, M., Oria, R., & Sánchez Gimeno, A. C. (2016). Changes in the physicochemical and nutritional parameters of Picual and Arbequina olive oils during frying. *Journal of Food Processing and Preservation, 40*(3), 353-361.
https://doi.org/10.1111/jfpp.12612

Al-Khusaibi, M., & Shafiur Rahman, M. (2021). Quality assessment of frying oil degradation. In M. Samad Khan, & M. Safiur Rahman (Eds.) *Techniques to measure food safety and quality* (pp. 329-345). Springer.
https://doi.org/10.1007/978-3-030-68636-9

Andrade, J. M., & Estévez Pérez, M. G. (2014). Statistical comparison of the slopes of two regression lines: a tutorial. *Analytica Chimica Acta, 838*, 1-12.
http://dx.doi.org/10.1016/j.aca.2014.04.057

Arslan, F. N., Kara, H., Ayildiz, H. F., Topkafa, M., Tarhan, I., & Kenar, A. (2013). A chemometric approach to assess the frying stability of cottonseed oil blends during deep-frying process: I. Polar and polymeric compound analyses. *Journal of the American Oil Chemists' Society, 90* (8), 1179-1193.
http://dx.doi.org/10.1007/s11746-013-2266-4

Bosso, T. (2017). *Olive oil Handbook* (volumen I). Academic Pages, New York.

Brühl L. (2014). Fatty acid alterations in oils and fats during heating and frying. *European Journal of Lipid Science and Technology, 116*(6), 707-15.
https://doi.org/10.1002/ejlt.201300273

Cascant, M. M., Garrigues, S., & de la Guardia, M. (2017). Comparison of near and mid infrared spectroscopy as green analytical tools for the determination of total polar materials in fried oils. *Microchemical Journal, 135*, 55-59.
https://doi.org/10.1016/j.microc.2017.07.012

Chen J., Zhang, L., Li, Y., Zhang, N., Gao, Y., & Yu, X. (2021). The formation, determination and health implications of polar compounds in edible oils: current status, challenges and perspectives. *Food Chemistry, 364,* 130451.
https://doi.org/10.1016/j.foodchem.2021.130451

Cocchi, M., Biancolillo, A. & Marini, F. (2018). Chemometric methods for classification and features selection. In J. Jaumot, C. Bedia, & R. Tauler (Eds.), *Data analysis for omic sciences – Methods and applications* (pp. 265-299). Elsevier.
https://doi.org/10.1016/bs.coac.2018.08.006.

COI/T.20/Doc. No 19/ Rev.5 (2019). *Spectrophotometric investigation in the ultraviolet.* International Olive Council (IOC), Madrid.

COI/T.20/Doc. No 35/ Rev.1. (2017). *Determination of peroxide value.* International Olive Council (IOC), Madrid.

Dais, P., & Hatzakis, E. (2013). Quality assessment and authentication of virgin olive oil by NMR spectroscopy: a critical rewiew. *Analytica Chmica Acta, 765*, 1-27. http://dx.doi.org/10.1016/j.aca.2012.12.003

Ebro. (2021). *FOM 320 Manual*. Ebro Electronic GmbH Co, KG.

Flores, M., Avendaño, V., Bravo, J., Valdés, C. Firestone. (2013). *Determination of polar compounds in frying fats* (6th ed.). AOCS Official Method Cd 3rd printing (pp. 20-91).

Flores, M., Avendaño, V., Bravo, J., Valdés, C., Forero-Doria, O., Quitral. V., Vilcanqui, Y., & Ortiz Viedma, J. (2021). Edible oil parameters during deterioration processes. *International Journal of Food Science, 7105170*, 1-16.
https://doi.org/10.1155/2021/7105170

Government of Spain (1989). Quality standard for oils and fats heated. *Boletín Oficial del Estado (BOE)*, No 26 (pp. 2665-2667) *(original in Spanish)*.

ISO 280. (1998). *Essentials oils – Determination of refractive index*. International Organization for Standardization (ISO), Geneva.

ISO 6885. (2016). *Animal and vegetable fats and oils - Determination of anisidine value*. International Organization for Standardization (ISO), Geneva.

IUPAC 1992 a. *Standard method 2.507: determination of polar compounds in frying fats, in standard methods for the analysis of oils, fats and derivatives* (7th ed.). International Union of Pure and Applied Chemistry, Blackwell, Oxford, England.

Martín Torres, S., Ruiz Castro, L., Jiménez Carvelo, A.M., & Cuadros Rodríguez, L. (2022). Applications of multivariate data analysis in shelf life studies of edible vegetal oils – a review of the few past years. *Food Packaging and Shelf Life, 3,* 100790.
https://doi.org/10.1016/j.fpsl.2021.100790

Mudawi, H. A., Elhassan, M. S. M., & Sulieman A. M. E. (2014). Effect of frying process on physicochemical characteristics of corn and sunflower oils. *Food and Public Health, 4*(4), 181-184.
https://doi.org/10.5923/j.fph.20140404.01

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. *Analitycal Methods, 7*, 4192-4201.
https://doi.org/10.1039/C5AY00168D.

Tyagi, V. K., & Vasishtha, A. K. (1996). Changes in the characteristics and composition of oils during deep-fatfrying. *Journal of the American Oil Chemists' Society, 73*(4), 499-506.
https://doi.org/10.1007/BF02523926

Upadhyay, R., Sehwag, S., & Mishra, H. N. (2017). Chemometric approach to develop frying stable sunflower oil blends stabilized with oleoresin rosemary and ascorbyl palmitate. *Food Chemistry, 2018*, 496-504.
http://dx.doi.org/10.1016/j.foodchem.2016.09.105

Zhang, Q., Qin, W., Li, M., Shen, Q., & Saleh, A. S. M. (2015). Application of chromatographic techniques in the detection and identification of constituents formed during food frying: a review. *Comprehensive Reviews in Food Science and Food Safety, 14*(5), 601-633.
https://doi.org/10.1111/1541-4337.12147

CAPÍTULO V

DISCUSIÓN INTEGRADA

CAPÍTULO V

Discusión integrada y conclusiones

V.1. Aplicación de la metodología de huellas instrumentales en problemas de calidad de aceites.

El término calidad alimentaria engloba la suma de todas las propiedades y atributos que hacen que un alimento sea aceptable para el cliente. Históricamente, la calidad se ha entendido como la ausencia de defecto, fraude o adulteración. Más recientemente, se ha considerado atributos que incluyen la apariencia, el tamaño, forma, brillo, el color y la consistencia, la textura, el sabor y olor y el contenido nutricional y las propiedades saludables esperadas por el consumidor. Finalmente, se ha considerado también como calidad alimentaria las características deseables que pueden justificar un valor añadido; por ejemplo, las formas de producción (agricultura ecológica, consideración medioambiental y bienestar animal), las zonas de producción (denominación de origen) y sus tradiciones asociadas. La calidad es una de las características más valoradas a la hora de elegir los alimentos, sin embargo por todos estos factores, su descripción objetiva es una tarea muy difícil.

El desarrollo de metodologías cada vez más eficaces para la detección de la adulteración de los alimentos, la confirmación tanto de la trazabilidad como de la autenticidad de los mismos y la verificación de su conformidad con el etiquetado, es un reto para los químicos analíticos de todo el mundo, especialmente cuando se abordan matrices analíticas complejas como los aceites y alimentos con alto contenido graso. Para abordar este problema se han llevado a cabo diferentes enfoques químicos: tradicionalmente se ha hecho uso de metodologías dirigidas (detección de marcadores, elaboración de perfiles de compuestos) y más recientemente el auge en la descripción y empleo de metodologías no dirigidas (huellas instrumentales) que están ganando cada vez más popularidad gracias a los avances en los instrumentos que son capaces de generar una enorme cantidad de datos a la vez y a la aplicación de técnicas quimiométricas. Es por eso que, en la ***Publicación I****,* se describe el potencial de los análisis no dirigidos basados en huellas instrumentales, las técnicas analíticas rápidas e inespecíficas que pueden usarse para obtenerlas, así como las correspondientes herramientas quimiométricas para poder aislar y utilizar la información que en ellas se recoge de forma implícita.

El término huella instrumental se refiere a la señal característica e inespecífica (espectro, voltamperograma, termograma, cromatograma, electroferograma o

imagen) de la muestra analizada, obtenida de forma no selectiva (no dirigida) que puede relacionarse con su composición química compleja, sus propiedades, y, por tanto, con su autenticidad, del mismo modo que una huella dactilar humana es específica de una determinada persona y la identifica inequívocamente. La información que una huella instrumental recoge de forma implícita y no evidente, normalmente requiere un tratamiento matemático basado en la aplicación de herramientas quimiométricas, en el que se incluyen técnicas de proyección, agrupamiento, modelización, etc.; con el objetivo de aislar la información y caracterizar el alimento. Estas huellas instrumentales pueden obtenerse mediante diversas técnicas analíticas (detectores): algunas como la espectrometría de masas (MS) o la resonancia magnética nuclear (NMR) proporcionan información discriminatoria útil y fácilmente disponible de toda la muestra mientras que otras como el detector de aerosol de partículas cargadas (CAD) o medidas de absorbancia UV-VIS proporcionan huellas poco informativas ya que responden de forma similar a la mayoría de compuestos o familias de compuestos en la muestra. Es por eso que suelen emplearse acopladas a una técnica cromatográfica para así obtener la señal de compuestos que eluyen consecutivamente y no simultáneamente.

Los resultados de numerosas investigaciones han demostrado que las huellas cromatográficas proporcionan la información necesaria para ser consideradas una herramienta poderosa y eficaz para el control y verificación de la calidad de los alimentos con alto contenido graso [**1**]. En la huella cromatográfica no es necesaria la identificación de cada uno de los picos si no que tras una separación total, o parcial, de los componentes se obtiene una medición relacionada con cada analito individual o coeluido, dependiendo por tanto de las condiciones en la separación y de las prestaciones del sistema de medición. Es por ello que en las publicaciones ***Publicación II*** y ***Publicación III*** se muestra la aplicación de distintas modalidades cromatográficas acopladas a dos detectores diferentes para la obtención de una huella instrumental representativa de la fracción grasa del aguacate con un doble objetivo de autenticidad referida a la variedad botánica y al origen geográfico.

El contenido graso y la composición de un aguacate varían en función de múltiples factores, como la ubicación geográfica de la planta, las condiciones climáticas, las prácticas agrícolas, el momento de la cosecha, el grado de madurez o la variedad botánica. Además, uno de los principales factores que influyen en la calidad de los aguacates es el estado de maduración al cosechar, determinado como materia seca, cuyo contenido está estrechamente vinculado

1. Esteki, M., Simal Gandara, J., Shahsavari, Z., Zandbaaf, S., Dashtaki, E., Heyden, Y.V. A review on the application of chromatographic methods, coupled to chemometrics, for food authentication. Food Control, 93 (2018), 165-182. https://doi.org/10.1016/j.foodcont.2018.06.015

con el de aceite. El valor umbral de comercialización para considerar que un aguacate posee una determinada calidad, y por tanto un valor económico distinto, depende de la variedad y de este contenido en aceite, lo que evidencia la importancia de conocer el cultivar del fruto del aguacate.

Se conocen tres variedades botánicas de aguacate: Antillana, Mejicana y Guatemalteca. Estas variedades se pueden diferenciar tanto por sus rasgos morfológicos y fisiológicos como por sus rasgos hortícolas. De estas, por hibridación en la mayoría de los casos fortuita, se han obtenido la mayor parte de las variedades comerciales existentes en la actualidad como "Hass", "Bacon" y "Fuerte". La variedad Hass es con diferencia la más extendida en nuestro país, la que presenta un mayor área de cosecha, un mayor consumo y además un precio de mercado más elevado que las otras dos [**2**]. Sin embargo, es habitual encontrar la venta indiscriminada de estas variedades, sin garantizar su trazabilidad y habitualmente con un precio común. Por ello, el estudio se centró en la discriminación de las variedades citadas priorizando la discriminación de la clase "Hass" frente al resto mediante modelos de clasificación multivariable. Se adquirieron muestras de Hass, Bacon y Fuerte en tiendas de alimentación de Granada. Además, se emplearon guacamoles (con un contenido de aguacate superior al 95%) y aceites comerciales de aguacates de variedad no especificada para verificar el poder de predicción de los modelos quimiométricos establecidos.

En primer lugar los aguacates fueron sometidos a un proceso de liofilización y extracción del aceite de la pulpa. Tras ello, se obtuvieron las huellas cromatográficas de los aceites extraídos mediante las dos modalidades de trabajo de la cromatografía de líquidos: en primer lugar se empleó la modalidad de fase invertida ((RP)HPLC), siguiendo la tendencia encontrada en literatura para el análisis de triglicéridos (TAG) [**3**] y a continuación se empleó la cromatografía en fase normal ((NP)HPLC) para comparar los resultados. Los sistemas de separación en NP no tienen el poder de resolución de los de RP y no se usan comúnmente para las separaciones de triglicéridos. Sin embargo, la cromatografía NP es una herramienta indispensable para la separación de las diferentes clases de lípidos y con ella los TAG pueden caracterizarse fácilmente en presencia de otros grupos de lípidos como diacilgliceroles,

2. Agbar. Evolución del precio del aguacate en 2022. https://agbaragriculture.com/evolucion-precio-aguacate/. (accedido 06 junio 2022).
3. Wang, M., Yu, P., Chittiboyina, A.G., Chen, D., Zhao, J., Avula, B., Wang, Y.H., Khan, I.A. Characterization, quantification and quality assessment of avocado (Persea Americana Mill.) oils. Molecules, 25, 6 (2020), 1453. https://doi.org/10.3390/molecules25061453

monoacilgliceroles, esteroles, ácidos grasos libres y fosfolípidos [**4**]. La descripción de los métodos y las condiciones aplicadas en ambas modalidades cromatográficas se muestran en la Tabla 1. En ambos casos, la detección se llevó a cabo empleando un sistema CAD.

Tabla 1. Descripción de las columnas cromatográficas y de las condiciones aplicadas en los métodos cromatográficos fase invertida (RP-) y fase normal (NP-).

	(RP)HPLC	**(NP)HPLC**
Columna	250mm × 4.6 mm i.d., 5 µm. Columna empacada Develosil C30-UG-5 (Nomura Chemica CO, Aichi, Japan)	250 mm × 4 mm i.d., 5 µm Lichrospher 100 CN (Merck, Darmstadt, Germany)
Temperatura	45 ºC	30 ºC
Fase móvil	Acetonitrilo/Isopropanol (40:60 v/v)	n-Hexano/Isopropanol (96:4 v/v)
Modalidad	Isocrático	Isocrático
Volumen de inyección	20 µL	20 µL
Tiempo de análisis	30 min	26 min

CAD se ha aplicado para el análisis de compuestos estructuralmente diversos (neutros, ácidos, básicos y 'zwitteriónicos', no volátiles y semivolátiles, tanto polares como no polares) utilizados en las industrias farmacéutica, química, alimentaria y de productos de consumo, así como en la investigación de las ciencias de la vida (por ejemplo, lípidos, proteínas, esteroides, polímeros, carbohidratos y péptidos). Las principales ventajas del sistema de detección CAD son la detección universal de analitos, una respuesta independiente de las propiedades químicas, un amplio rango de respuesta dinámica con alta sensibilidad desde bajas a altas concentraciones de analitos, buena precisión para una diversa gama de analitos y un funcionamiento sencillo y fiable [**5**]. Además, el empleo de CAD acoplado a HPLC empleando la metodología

4. Buchgraber, M., Ulberth, F., Emons, H., Anklam, E. Triacylglycerol profiling by using chromatographic techniques. Eur. J. Lipid Sci. Technol., 106, 9 (2004), 621-648. https://doi.org/10.1002/ejlt.200400986

5. Vehovec, T., Obreza, A. Review of operating principle and applications of the charged aerosol detector. J. Chrom. A., 1217, 10 (2010), 1549-1556. https://doi.org/10.1016/j.chroma.2010.01.007

'fingerprinting' con fines de autentificación de otros alimentos de alto contenido graso ha sido ampliamente descrito [**6**].

En la Figura 1 se muestran las huellas cromatográficas obtenidas para una misma fracción grasa (aceite) de aguacate aplicando HPLC-DAD en fases normal e invertida, respectivamente. Como cualquier aceite vegetal, el aceite de aguacate está compuesto principalmente por TAG. La elución de TAG no será la misma bajo las condiciones de NP y RP. En (NP)HPLC las especies saturadas eluyen antes que los homólogos más insaturados. Por el contrario, en la (RP)HPLC los componentes se separan según el efecto conjunto de las longitudes de cadena más su grado de insaturación. Por ello se obtienen huellas diferentes que aportan de manera implícita información distinta.

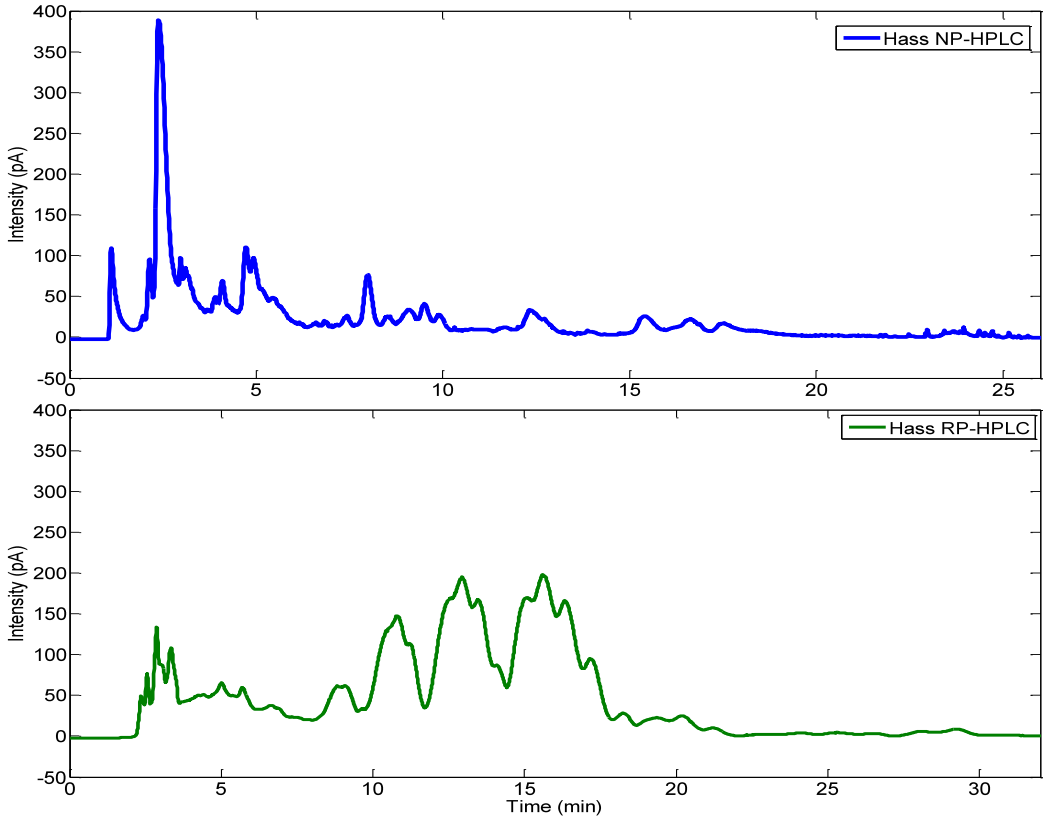


Figura 1. Huella cromatográfica de la fracción grasa de un aguacate de la variedad Hass obtenida por cromatografía de líquidos en modalidades fase normal (NP) y fase invertida (RP) respectivamente.

6. Bosque Sendra, J.M., Cuadros Rodríguez, L., Ruiz Samblás, C., De la Mata, A.P. Combining chromatography and chemometrics for the characterization and authentication of fats and oils from triacylglycerol compositional data – a review. Anal. Chim. Acta, 724 (2012), 1-11. https://doi.org/10.1016/j.aca.2012.02.041

La señal procedente directamente del instrumento cromatográfico se preprocesó de manera adecuada (agrupación y superposición de cromatogramas, filtrado, corrección de línea base, alineamiento con función 'icoshift' y centrado en la media) y su aplicación se evalúo en su conjunto utilizando herramientas quimiométricas de diagnóstico y clasificación. En este caso se aplicaron dos enfoques de modelado, un enfoque no supervisado, para verificar la existencia o no de subgrupos de muestras en la matriz de datos teniendo en cuenta las similitudes entre ellos, sin información previa de tal existencia, y un enfoque supervisado con el fin establecer modelos de clasificación y reglas para asignar una clase a nuevas muestras. En primer lugar se llevó a cabo un modelado PCA del conjunto de los datos para explorar si existían agrupaciones naturales de los distintos aceites en función de su variedad botánica. En segundo lugar, SIMCA y PLS-DA fueron usados para establecer modelos de clasificación de los datos obtenidos en ambas modalidades cromatográficas, validados externamente con el 30% de las muestras.

Para evaluar los diferentes escenarios de clasificación, se calcularon varias métricas de calidad: sensibilidad, especificidad, precisión, valor predictivo negativo, índice de Youden, eficiencia, correlación de Matthews, entre otros, en función de los resultados de validación externa [**7**]. Además, se llevó a cabo un árbol de decisión basado en los modelos PLS-DA con el fin de comprobar los resultados predictivos de los mismos para las muestras de variedad desconocida: en un árbol de decisión una muestra no se clasificará en ninguna clase si no es reconocida como un objeto de las clases de entrenamiento de los modelos. Esto puede dar lugar a resultados de predicción diferentes a los obtenidos mediante un modelo PLS-DA único, que siempre asigna las muestras a la clase más probable.

De todo este estudio se obtuvieron diversas conclusiones. Por un lado se demostró que la metodología de huella cromatográfica obtenida en condiciones NP y RP con un sistema de detección CAD, conjuntamente con diferentes estrategias de clasificación multivariable, permiten la discriminación de variedades botánicas de aguacate y por tanto su autenticación. Hasta donde sabemos, no existían antecedentes destinados a comparar y clasificar los frutos del aguacate, menos aún mediante métodos cromatográficos sencillos y rápidos en modalidad de trabajo 'fingerprinting'. Además, se demostró que las huellas obtenidas bajo condiciones NP incluyen un mayor poder de discriminación implícito que las obtenidas por RP, mostrando un menor tiempo

7. Cuadros Rodríguez, L., Pérez Castaño, E., Ruiz Samblás, C. Quality performance metrics in multivariate classification methods for quality analysis. Trends Anal. Chem., 80 (2016), 612-621. http://dx.doi.org/10.1016/j.trac.2016.04.021

de análisis y un mejor rendimiento de los modelos de clasificación. El método PLS-DA ha mostrado mejores métricas de calidad que el método de modelización de clases SIMCA y cabe destacar que la estrategia de dos clases de entrada dio mejores resultados que la estrategia de clase de una entrada. Por último, se demostró que los árboles de decisión son una herramienta útil para realizar una clasificación multiclase y pueden utilizarse como información complementaria a los modelos simples concatenados porque muestran resultados diferentes a partir del mismo conjunto de muestras de predicción.

La Tabla 2 recoge de manera resumida una comparación de los hitos/resultados obtenidos mediante los diferentes métodos en función de la modalidad cromatográfica empleada para el análisis de las muestras.

Tabla 2. Comparación de los resultados obtenidos por los diferentes modelos para las huellas obtenidas en distintas condiciones cromatográficas.

Modelo	**RP-HPLC-CAD**	**NP-HPLC-CAD**
PCA	Mostró la discriminación entre las variedades Fuerte (scores negativos de PC1) y Bacon (escores positivos de PC1) mientras que los scores de la variedad Hass aparecían dispersos por todo el modelo.	Mostró el agrupamiento natural de las 3 variedades en estudio: Bacon (scores negativos para PC3 y PC4), Fuerte (scores negativos para PC3 pero positivos para PC4) y Hass (scores positivos para PC3 y negativos para PC4).
SIMCA	*Modelo Hass-no Hass – 1 clase de entrada:* ninguna muestra Hass del conjunto de validación se clasifico de forma correcta. *Modelo Fuerte-Bacon – 2 clases de entrada:* ninguna muestra del conjunto de validación se asignó de forma correcta a ninguna clase modelada.	*Modelo Hass-no Hass – 1 clase de entrada:* todas las muestras de Hass se clasificaron en el modelo de clase, mientras que ninguna de las muestras "no Hass" fue mal clasificada. *Modelo Fuerte-Bacon – 2 clases de entrada:* ninguna muestra Bacon se asignó correctamente, mientras que la mayoría de las muestras Fuerte fueron correctamente asignadas. Ningún modelo permite predecir las muestras de variedad desconocida.
PLS-DA	*Modelo Hass no Hass:* todas las muestras Hass se clasificaron en la clase Hass mientras que todas las muestras Bacon y Fuerte fueron clasificadas como "no Hass". *Modelo Fuerte-Bacon:* todas las muestras fueron bien clasificadas en la clase Fuerte o Bacon correspondiente.	*Modelo Hass no Hass:* todas las muestras Hass se clasificaron en la clase Hass mientras que todas las muestras Bacon y Fuerte fueron clasificadas como "no Hass". *Modelo Fuerte-Bacon:* todas las muestras fueron bien clasificadas en la clase Fuerte o Bacon correspondiente. Cuatro muestras "desconocidas" fueron predichas de la variedad Fuerte y una se consideró no concluyente, ya que se localizaba en la zona de incertidumbre.

En base a estos resultados, una segunda investigación fue llevada a cabo con el objetivo doble de discriminar entre un mayor número de variedades botánicas de aguacate y diferentes orígenes geográficos del cultivar aplicando de nuevo la cromatografía de líquidos en modalidad 'fingerprinting' para llevar a cabo el análisis de las muestras.

A pesar de que el origen de este cultivo es Centroamérica, el aguacate ha despertado un creciente interés mundial y su producción está dispersa por todo el mundo en regiones tropicales y subtropicales. Es muy importante garantizar la trazabilidad de los aguacates ya que su origen afecta de gran manera al precio de mercado. Los principales países productores como México o Perú, con mano de obra barata, regulaciones medioambientales más laxas y una gran superficie de cultivo de aguacates, pueden incurrir en competencia desleal en el mercado de precios internacional. Es por ello que se adquirieron un mayor número de muestras de aguacate de 6 variedades botánicas diferentes y pertenecientes a diversos orígenes geográficos. La descripción de las muestras incluidas en este estudio se muestra en la Tabla 3.

Tabla 3. Descripción del origen varietal y geográfico de las muestras de aguacate incluidas en este estudio.

Clase referida a la variedad botánica	Nº de muestras	País	Clase referida al origen
Bacon	14	España	Europa
Ettinger	5	Israel	—
Fuerte	14	España	Europa
Hass	18	España	Europa
	12	Kenia	África
	10	Sudáfrica	África
	8	México	América
	8	Perú	América
Pinkerton	3	España	Europa
Topa-Topa	12	España	Europa
Desconocida	4	España	Europa

Se aplicó el mismo tratamiento de muestra descrito anteriormente: liofilización de la pulpa del aguacate y extracción grasa. En este nuevo estudio, se aplicó el método de análisis cromatográfico y condiciones cromatográficas descritas

anteriormente solamente en la modalidad (NP)HPLC, puesto que los resultados obtenidos mostraron la idoneidad de estas condiciones cromatográficas frente a las obtenidas en (RP)HPLC y se empleó el sistema de detección de fila de diodos (DAD) mediante absorción molecular en la región del UV. En estas circunstancias, la huella obtenida depende de la composición de la fracción química principal, es decir, TAG. Pero además, sólo los TAG que tienen al menos un ácido graso insaturado, principalmente palmitoleico, C16:1; oleico, C18:1; y linoleico, C18:2 producen una señal medible, ya que los ácidos grasos saturados son casi transparentes para el detector de absorción UV en las longitudes de onda de trabajo. Por lo tanto, en este caso la huella registrada tiene un alto grado de especificidad con respecto a la distribución de los ácidos grasos insaturados y la información que incluye implícita es diferente a las huellas obtenidas con otros detectores universales como el CAD.

A continuación, se realizó el preprocesamiento de las huellas para eliminar todo el exceso de información que pueda enmascarar la señal de interés. Se seleccionó la longitud de onda de 220 nm y 1 a 17 minutos como intervalo de tiempo de análisis por ser las condiciones en las que las muestras mostraban la mayor absorbancia significativa y las mayores diferencias entre ellas, descartando el análisis de las variables fuera de este rango. Además, se realizaron las etapas de suavizado de la señal, corrección de la línea base, alineamiento de picos y centrado en la media de los conjunto de datos (etapas de preprocesado típicas para un cromatograma), antes del análisis quimiométrico. La Figura 2 muestra la señal registrada para un aguacate de variedad Hass y origen España (cromatograma 2D desde 190 a 490 nm) y la misma tras la selección de la región de interés en tiempo (1-17min), longitud de onda (220 nm) y el preprocesado de la señal. Las huellas cromatográficas preprocesadas fueron utilizadas para construir un modelo de clasificación que permitiera la mejor discriminación entre clases referidas al origen geográfico y a las variedades botánicas.

PLS-DA, fue la primera técnica utilizada para establecer modelos de clasificación adecuados para la discriminación entre las tres clases que describen el origen geográfico: Europa, América y África. Se realizó un modelado con tres clases de entrada (3iC) validado externamente con el 30% de las muestras. El modelo fue capaz de diferenciar las muestras de entrenamiento europeas de las de las otras dos clases. Sin embargo, las muestras africanas y americanas se consideraron de la misma clase y el modelo de clasificación 3iC no encontró diferencias entre ellas. Además, la mitad de las muestras de validación de la clase Europa fueron clasificadas erróneamente.

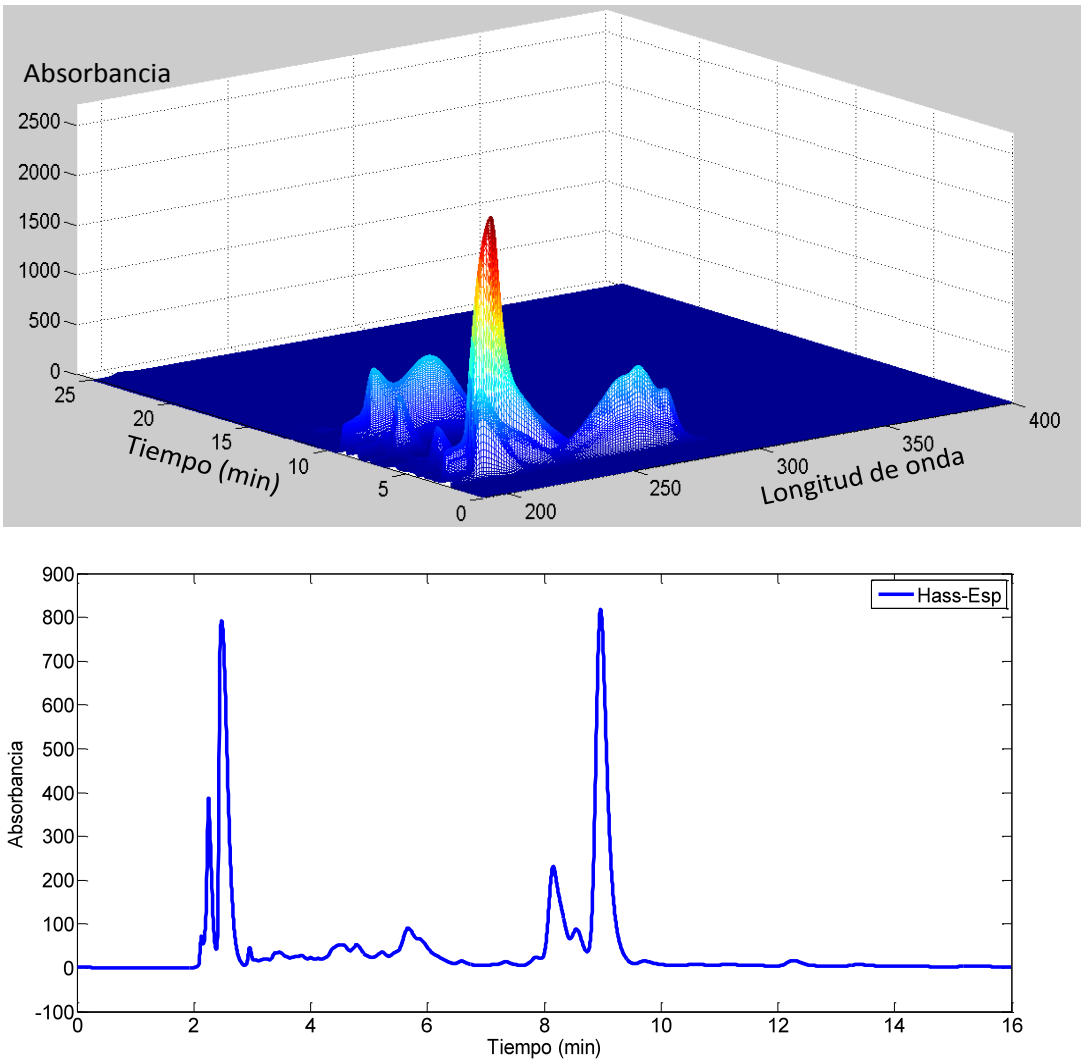


Figura 2. Cromatograma 2D registrado desde 190 a 490 nm para un aguacate Hass de origen español y el cromatograma (huella instrumental 1D) tras la selección de la región de interés: 220 nm e intervalo de tiempo 1 a 17min.

Con el fin de mejorar los resultados de validación obtenidos, se empleó el algoritmo SVM. La principal ventaja de SVM-C sobre PLS-DA es que crea una separación entre las regiones de las diferentes clases cuando no son lo suficientemente evidentes. Los modelos SVM-C mostraron mejores resultados que los PLS-DA y permitieron la clasificación correcta de casi todas las muestras. De entre todas, sólo una muestra de origen americano fue mal clasificada en la clase África.

Resultados similares se obtuvieron para la discriminación de las diferentes variedades botánicas. Se realizó una primera diferenciación entre la variedad Hass y el resto de variedades, seguida de un modelado de tres clases de entrada, Fuerte, Bacon y Topa-topa, variedades de las que se disponía de un suficiente número de muestras. En ambos casos, los modelos SVM mostraron mejores resultados que los de PLS-DA, consiguiéndose la clasificación correcta

de todas las muestras y permitiendo además la comparación de las clases modeladas con otras como Ettinger o Pinkerton, no consideradas como clase de entrada para ningún modelo ya que el bajo número de muestras disponibles de estas variedades no hizo posible desarrollar y validar de forma adecuada un método de clasificación y discriminación de estas variedades.

En nuestro trabajo anterior, PLS-DA mostró las mejores métricas de calidad de clasificación. En este estudio el rendimiento del modelado PLS-DA resultó ser deficiente. Hay que tener en cuenta que se utilizó un mayor número de muestras de aguacate, se incluyeron más variedades y se empleó un detector no universal diferente (DAD) para registrar las huellas. Además, las huellas cromatográficas de dos aguacates de la misma variedad pero cosechados en diferentes continentes muestran mayor similitud que las huellas de dos aguacates de diferentes variedades. Todos estos hechos hacen necesario el uso de métodos de clasificación y predicción capaces de encontrar diferencias significativas entre huellas instrumentales muy similares, más aun cuando se aumenta la variabilidad de los datos y diferentes orígenes y variedades se fusionan en un mismo modelo. De acuerdo con los resultados, SVM demostró ser un mejor clasificador que PLS-DA y se consiguió la autenticación de aguacates de acuerdo a su variedad botánica y origen geográfico.

Por último se llevó a cabo un nuevo estudio en el que se adquirieron las huellas cromatográficas de la fracción grasa de los diferentes frutos de aguacate utilizando la cromatografía gaseosa a altas temperaturas con un detector de ionización de llama (GC-FID) y éstas se emplearon para construir modelos de clasificación. De nuevo, la metodología de la huella cromatográfica demostró ser un enfoque útil para autentificar muestras de aguacate en términos de origen geográfico y el tipo de cultivar. Sin embargo, sólo algunas de las variables o región específica de la huella digital (región de interés) resultan realmente informativas para definir la característica particular de una clase cuando se considera un gran conjunto de datos. Se compararon los resultados obtenidos mediante un análisis discriminante de mínimos cuadrados parciales convencional y disperso (PLS-DA y sPLS-DA). PLS-DA no selecciona automáticamente las variables predictoras significativas y sus resultados dependen de diferentes estadísticos de selección de variables ('variable influence on projection' (VIP), 'genetic algorithm' (GA), etc.). En este sentido sPLS-DA resulta una herramienta útil para examinar datos de alta dimensión y para seleccionar la región de interés de la huella digital, que es característica de una muestra de aguacate. Los resultados de este estudio dieron lugar a una publicación científica no incluida en los contenidos de esta tesis doctoral y que se muestra en el Anexo 1.

Los enfoques 'fingerprinting' se han convertido en una herramienta muy potente para obtener huellas alimentarias en los procesos de autentificación que tienen

como objetivo una caracterización exhaustiva de matrices alimentarias complejas en el ámbito académico. Mediante el análisis químico no dirigido ('untargeted'), empleando posteriormente las herramientas quimiométricas de tratamiento de datos multivariable para extraer la información de interés, se pueden clasificar las matrices alimentarias en cuanto a su origen geográfico, variedad de especies o posibles adulteraciones [**8**]. Su aplicación con éxito en investigación es innegable y ha demostrado la viabilidad de los enfoques de huellas dactilares no dirigidas [**9**]. Sin embargo, la transferencia del mundo académico al mundo real, su adopción e implementación en los controles rutinarios de calidad, autenticidad y seguridad de los alimentos es todavía limitada. La razón principal puede ser la falta de confianza que sigue existiendo en la fiabilidad de aquellos resultados científico-técnicos que no se basan en información concreta y explícitamente percibida. Otro inconveniente adicional es la dificultad de integrar los métodos 'fingerprinting' en los sistemas de Buenas Prácticas de Laboratorio (BPL) o en la norma ISO 17025, lo que impide su aplicación efectiva.

Una consecuencia directa de esta situación es la ausencia casi generalizada de métodos no dirigidos en el catálogo de métodos analíticos oficiales para el control de los alimentos. Desde un punto de vista práctico, la transferibilidad y aplicación de los métodos analíticos basados en la huella instrumental adolecen aún de una correcta armonización (incluida la validación del método) lo que implica establecer normas y requisitos para garantizar que los métodos analíticos sean iguales o eventualmente coherentes cuando se aplican en diferentes laboratorios de modo que los resultados sean comparables. En este sentido, el Codex de Productos Químicos para la Alimentación (FCC) junto con la Farmacopea de los Estados Unidos (USP), mantiene desde 2017 lo que se puede considerar la primera directriz para el desarrollo y la validación de métodos no dirigidos orientados a detectar la adulteración de alimentos [**10**].

Además, la fiabilidad de los resultados obtenidos a partir de la correcta interpretación de las huellas instrumentales depende de la disponibilidad de bases de datos universales, invariantes, reconocidas y accesibles que recojan

---

8. Riedl, J., Esslinger, S., Fauhl-Hassek, C. Review of validation and reporting of non-targeted fingerprinting approaches for food authentication. Anal. Chim. Acta, 885 (2015), 17-32. https://doi.org/10.1016/j.aca.2015.06.003
9. Murti, A.A., Hastuti, B., Rohman, A. Application of elemental fingerprinting for the authentication of tea: a review. J. App. Pharm. Sci., 12, 03 (2022), 45-54. https://doi.org/10.7324/JAPS.2022.120305
10. United States Pharmacopeia (USP). Food Chemicals Codex (FCC) 12, Appendix XVIII. Guidance on developing and validating non-targeted methods for adulteration detection. Food Chemicals Codex, 12th ed., USP:Rockville, MD, (pp 1540), 2020.

un número suficiente de señales representativas del alimento o del conjunto de alimentos en cuestión. Sin embargo, hasta la fecha no existen bases de datos similares de huellas instrumentales de alimentos, y cada laboratorio o entidad genera y utiliza su propia base de datos para alimentar sus propios modelos de autenticación. En principio, estas bases de datos serían más fáciles de crear a partir de señales espectroscópicas o espectrométricas que son estables y altamente reproducibles si se mantienen las condiciones instrumentales.

Sin embargo, en lo que respecta a huellas cromatográficas, la tarea presenta numerosos problemas adicionales ya que los cromatogramas dependen del estado y tipo de instrumento (columna) y el detector empleado. De hecho, los cromatogramas de la misma muestra alimentaria obtenidos en diferentes laboratorios o instrumentos o en el mismo instrumento pero en un gran intervalo de tiempo raramente son iguales y muestran desviaciones tanto en los tiempos de retención como en las intensidades de las señales. Esto puede generar pequeños inconvenientes cuando la cromatografía se acopla a un espectrómetro de masas (MS) para identificar compuestos desconocidos, cuantificar compuestos conocidos y/o dilucidar la estructura y las propiedades químicas de las moléculas. Pero además, pequeñas desviaciones muestran mucha más importancia cuando todo el cromatograma (huella) se utiliza como matriz tensorial de entrada para crear un modelo multivariable, ya que estos desplazamientos en los ejes de tiempos e intensidades, dificultan el desarrollo y validación de modelos creando clasificaciones erróneas y pudiendo llegar a predicciones no confiables. Por ello, las señales cromatográficas deben ser previamente estandarizadas, de modo que puedan incluirse en una base de datos de fácil acceso.

Todos estos hechos se recogieron en una publicación científica de perspectiva sobre las posibilidades y desafíos que actualmente presentan el empleo de las huellas cromatográficas en el campo de la autenticidad y calidad alimentaria (incluida en el Anexo 1 de esta tesis doctoral). Hasta la fecha, no existen muchos enfoques que describan cómo debe realizarse dicha normalización, especialmente para la cromatografía de líquidos. Es por eso que en las ***Publicaciones IV*** *y* ***V*** se presentan el desarrollo y aplicación de una metodología para conseguir señales cromatográficas independientes del instrumento, y del estado del propio instrumento, que han sido denominadas como "señales agnostizadas".

Los cromatogramas en bruto necesitan ser sometidos a una transformación matemática para minimizar la variabilidad informativa no debida a características de la muestra, minimizar la influencia de errores y poder extraer información fiable. Es necesario aplicar un conjunto de operaciones de preprocesamiento, las cuales se pueden dividir en: (i) las que se aplican sistemáticamente, independientemente del tipo de señal (por ejemplo,

eliminación de ruido, suavizado, corrección de la línea base, entre otros) y (ii) las que implican el uso de una referencia (por ejemplo, escalado, normalización o alineación). La alineación, el enfoque más popular, requiere encontrar una transformación adecuada del eje temporal mediante un algoritmo de alineación para que se ajuste lo más posible a una señal de referencia determinada. En consecuencia, la selección de esta referencia es una tarea crítica, ya que no todas las señales tendrán una forma similar o la misma complejidad. Este enfoque fue el utilizado en los trabajos anteriores para la autenticación de aguacates, donde algoritmos como 'icoshift' y 'msalign' se utilizaron para alinear los cromatogramas en función de uno o varios picos de referencia del conjunto total de las huellas. Este alineamiento está limitado a la obtención de señales similares en un mismo sistema cromatográfico y requiere su repetición cada vez que una muestra nueva es incorporada al modelo (entrenamiento o validación) pues todos los cromatogramas son alineados de manera conjunta. Esto tiene el inconveniente de que podría dar lugar a resultados de clasificación diferentes cada ver que una muestra nueva es incorporada.

Hasta la fecha, se han propuesto decenas de algoritmos para tratar la alineación y la normalización de huellas cromatográficas, especialmente cuando están acopladas a sistemas de MS [**11**]. Sin embargo, la mayoría de ellos dependen del analito en estudio, se basan en variables de referencia internas y ninguno de estos enfoques puede considerarse como procedimientos adecuados para obtener respuestas absolutas reproducibles para resolver la armonización universal de los datos. Por ello, la aparición de un método de estandarización independiente, referido a una referencia externa, podría ser crucial teniendo en cuenta que la combinación de referencias estándar externas e internas es una metodología útil para obtener huellas cromatográficas comparables, con independencia del estado cromatográfico y del tipo de analito. En este trabajo se propuso el uso de esta nueva "metodología de agnostización" para armonizar las señales cromatográficas y resolver así los problemas asociados a la metodología 'fingerprinting' cuando se utiliza (RP)HPLC (una revisión reciente sobre la aplicación de las huellas obtenidas por cromatografía de líquidos para la autentificación de alimentos señaló que en un 85% estas eran obtenidas bajo condiciones de fase invertida por lo que se decidió trabajar bajo estas condiciones).

La Figura 3 muestra el diagrama incluido en la ***Publicación IV*** donde se refleja de forma esquemática una descripción general de las etapas necesarias para

11. Jellema, R.H., Folch Fortuny, A., Hendriks M.M.W.B. Variable shift and alignment. In: S.D. Brown, R. Tauler, B. Walczak (Eds.), Comprehensive Chemometrics: Chemical and Biochemical Data Analysis, 3, 2nd ed., Elsevier B.V., Amsterdam, (115-136), 2020. http:/doi.org/10.1016/B978-0-12-409547-2.14886-3

llevar a cabo la metodología propuesta de agnostización de huellas instrumentales obtenidas por HPLC.

El primer paso es establecer una mezcla estándar de compuestos químicos apropiados que sirva como referencia invariable para normalizar los valores de retención. A cada compuesto químico se le asigna un valor empírico, conocido como puntuación de retención estándar (SRS) (Ecuación 5.1), relacionado con su tiempo de retención y su orden de elución:

$$\mathrm{SRS_i} = \frac{\Delta(\mathrm{medianRT})_{\mathrm{i,i-1}}}{\mathrm{median}\left(\Delta(\mathrm{medianRT})_{\mathrm{i,i-1}}\right)} + \mathrm{SRS_{i-1}} \qquad (5.1)$$

donde $(\Delta(\mathrm{medianRT})_{i,i-1})$ se refiere a las diferencias absolutas en los tiempos de retención de constituyentes químicos adyacentes y $(\mathrm{median}(\Delta(\mathrm{medianRT})_{i,i-1})$ a el valor mediano global de estas diferencias. Un valor SRS igual a 1 se asigna siempre al primer compuesto eluído de la mezcla estándar de referencia.

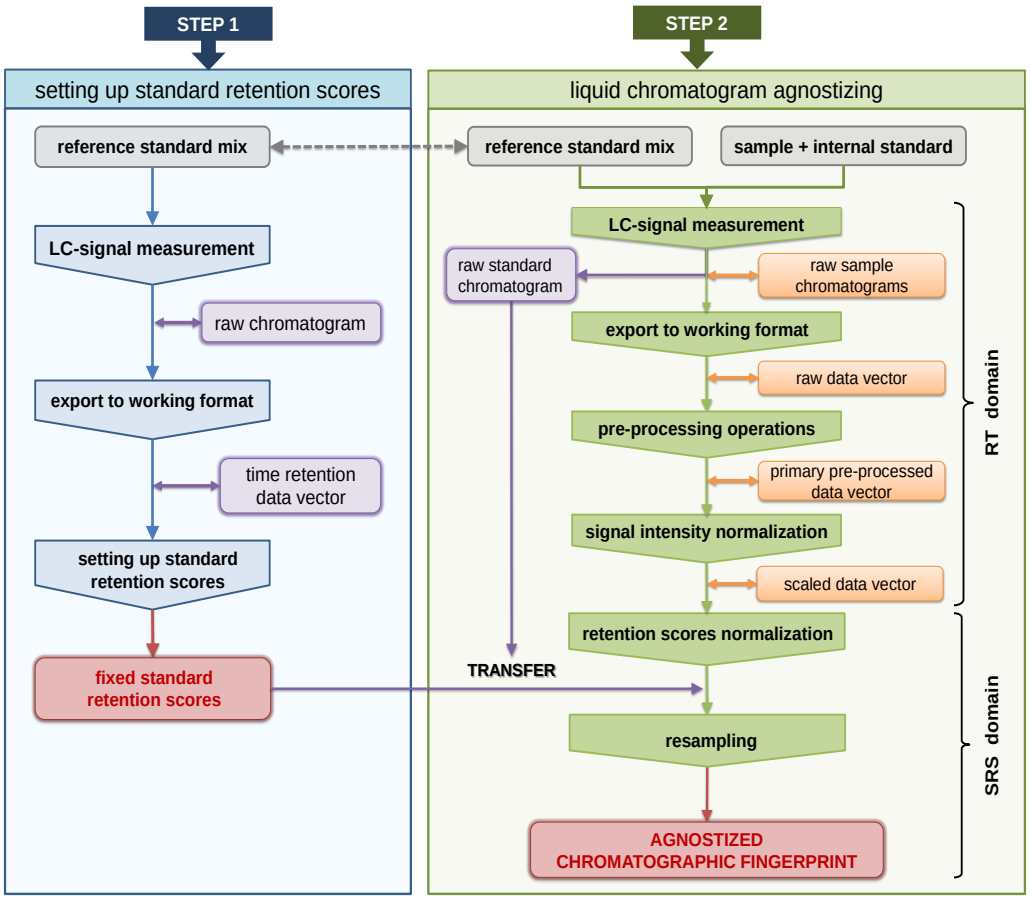


Figura 3. Diagrama de flujo que muestra una descripción general de los pasos de la agnostización para cromatografía de líquidos. Figura incluida en la ***Publicación IV***.

Una vez asignados los SRS invariantes, la mezcla estándar se analiza al menos al principio y al final de cada tanda cromatográfica de análisis y se

registra el tiempo de retención (RT) experimental de cada compuesto químico (la mezcla estándar se utiliza como estándar externo, ya que se analiza en una solución de ensayo independiente). En segundo lugar, se representa el promedio de los RT de todos los compuestos químicos que componen la mezcla estándar frente a los SRS invariantes predeterminados. Normalmente se encuentra un comportamiento casi lineal, aunque en algunos casos puede aparecer un patrón curvo. La huella cromatográfica es preprocesada (operaciones básicas de filtrado y corrección de línea base) y normalizada según la metodología de patrón interno. Finalmente, el vector de tiempos de la huella es reemplazado por un vector de SRS, independiente del estado cromatográfico, mediante interpolación lineal por segmentos ('spline'), es decir, mediante el uso de una función que se transfieren por interpolación entre los dos valores del vector de datos estándar externo de SRS invariantes que cubren el rango de tiempos que van a ser normalizados.

La aplicabilidad de la metodología se comprobó en el análisis del contenido en compuestos biofenólicos presentes en los aceites de oliva virgen extra (AOVE). Uno de los principales desafíos a los que se enfrenta la industria en el análisis de AOVE es el desarrollo de un procedimiento confiable que permita obtener el consenso de todas las partes interesadas con respecto a la cuantificación de las concentraciones de polifenoles presentes en aceites de oliva vírgenes para otorgar a aquellos aceites que cumplan el valor mínimo requerido (de 5 mg de hidroxitirosol y sus derivados por 20 g de aceite de oliva) la declaración de propiedades saludables introducida por el Reglamento (UE) 432/2012 [**12**].

La metodología adoptada por el Consejo Oleícola Internacional (COI) para la determinación de los biofenoles de AOVE consiste en una separación cromatográfica de un extracto hidroalcohólico, seguida de una detección UV a 280 nm y da un resultado global (que abarca los fenoles simples, los secoiridoides, los lignanos flavonoides y los ácidos fenólicos) en función de un patrón interno (ácido siríngico) que se refiere al tirosol. Este método muestra largos tiempos de análisis y un enfoque de cuantificación bastante ambiguo. En esta tesis se describe un nuevo método para la determinación de compuestos biofenólicos en AOVE con tiempos reducidos tanto de preparación de muestra como de análisis cromatográfico. Este método emplea fases móviles comunes y un gradiente relativamente sencillo, que no obstante, se simplificó aún más como método de referencia para establecer los valores de SRS para los compuestos químicos estándar seleccionados para la agnostización en (RP)HPLC. La Tabla 4 muestra las especificaciones para ambos métodos.

12. Reglamento (UE) No 432/2012 por el que se establece una lista de declaraciones autorizadas de propiedades saludables de los alimentos distintas de las relativas a la reducción del riesgo de enfermedad y al desarrollo y la salud de los niños. DOUE L136/1-39, 2012.

Tabla 4. Condiciones cromatográficas de los métodos empleados.

	Método de referencia para establecer SRS			**Método de determinación de biofenoles en AOVE**		
	Tiempo (min)	**% A**	**% B**	**Tiempo (min)**	**% A**	**% B**
Gradiente	0	95	5	0	95	5
	20	5	95	10	70	30
				12	62	38
				17	50	50
				18	5	95
				20	95	5
Fases móviles	(A) Agua + 0.5% ácido acético; (B) Acetonitrilo					
Fase estacionaria de referencia	Zorbax Eclipse Plus C18; 4.6 × 50 mm, 5 µm					
Flujo	1 mL/min					
Temperatura	25 ºC					
Detección	DAD, 280 nm					
Tiempo de análisis	20 min					
Volumen de inyección	20 µL					

La elección de los compuestos químicos que formarían parte de la mezcla estándar de referencia fue un paso esencial. Idealmente, se persiguió que cubrieran todo el tiempo de ejecución, mostraran un perfil de elución regular, fueran puros, disponibles y baratos y tuvieran un comportamiento químico similar a los componentes endógenos de interés en la muestra. En base a esto, y a los antecedentes encontrados se registraron los tiempos de retención, en las condiciones especificadas de una serie de alquil-aril cetonas (también denominados 1-benzoil alcanos) y otros compuestos orgánicos similares para crear la mezcla estándar definitiva y establecer sus SRS invariantes para (RP)HPLC siempre que las condiciones cromatográficas se mantengan.

Finalmente, se llevó a cabo la determinación de biofenoles para una misma muestra en diferentes sistemas cromatográficos: distintos modelos de HPLC, UPLC y UHPLC, equipados con columnas en distinto grado de uso, fase estacionaria diferente (C8), y diferentes modelos de detector. Las huellas obtenidas fueron normalizadas en base a un patrón interno y alineadas en términos de SRS, completando la agnostización de las mismas. Para validar la bondad de los resultados y por tanto la aplicabilidad de la metodología propuesta, se calcularon los índices de similitud de las huellas cromatográficas biofenólicas de cada estado cromatográfico antes y después de la

agnostización en el dominio de tiempo y en el dominio de SRS. Como índice de similitud representativo, se calculó el índice de coseno (cosθ) [**13**], una medida del coseno del ángulo entre los vectores de huellas cromatográficas considerados. Cuanto más cercanos a 1 sean los valores del índice, mayor será la similitud entre las huellas consideradas. Además, se restaron los vectores de huellas cromatográficas elemento por elemento considerando una de ellas como la de referencia y cada una de las huellas obtenidas en los diferentes sistemas cromatográficos, y se calcularon los valores medios de las diferencias como un segundo índice de similitud entre huellas.

En todos los casos, la agnostización mejoró la similitud entre huellas, obteniéndose en la mayoría de los casos huellas casi idénticas (variaciones de cosθ de 0.24 antes de la agnostización a 0.90 cuando ésta se ha llevado a cabo). En base de los resultados que se obtuvieron en este estudio, se pudo demostrar que la metodología es válida para armonizar cromatogramas 1D, haciendo que la nueva señal o 'fingerprint' sea independiente del estado cromatográfico. Esto permitiría establecer bases de datos cromatográficos característicos para armonizar, recopilar, analizar y compartir información de análisis comparables para cada matriz alimentaria. Además, trabajar con huellas cromatográficas agnostizadas proporciona la posibilidad de crear un único modelo multivariable, ya sea para su uso en aplicaciones cualitativas (clasificación) o cuantitativas (cuantificación). Nuestra investigación futura consistirá en elevar la aplicabilidad de la metodología a un estudio interlaboratorio. La metodología es sencilla, no requiere grandes esfuerzos en la preparación de las muestras, la optimización cromatográfica o el uso de plataformas específicas para el procesamiento de datos y ofrece resultados realmente eficaces.

Además, el trabajo conjunto con otros investigadores de nuestro grupo ha permitido el desarrollo y publicación de la metodología de agnostización de huellas obtenidas por cromatografía de gases y por cromatografía de líquidos en condiciones NP (los artículos publicados correspondientes no forman parte de esta tesis, pero se encuentran recogidos en el Anexo 1 de esta Memoria).

Por último, la aplicabilidad de la metodología con fines identificativos ha sido también evaluada. La espectrometría de masas (MS) es una potente técnica analítica utilizada para identificar compuestos desconocidos, cuantificar compuestos conocidos y dilucidar la estructura y las propiedades químicas de las moléculas. Sus aplicaciones son incontables como sistema de detección

13. Ortega Gavilán, F., Valverde Som, L., Rodríguez García, F.P., Cuadros Rodríguez, L., Bagur González, M.G. Homogeneity assessment of reference materials for sensory analysis of liquid foodstuffs. The virgin olive oil as case study. Food Chem., 322 (2020), 126743. https://doi.org/10.1016/j.foodchem.2020.126743

tras una separación cromatográfica [**14**,**15**]. Sin embargo, la propia complejidad de la técnica hace que la transferencia de datos de un instrumento a otro sea uno de los principales problemas, no siendo posible, en la mayoría de los casos, obtener la misma información o una información similar con un instrumento análogo pero de un fabricante diferente o incluso con el mismo instrumento después de realizar los análisis en diferentes espacios de tiempo. Esto complica cualquier análisis que sea llevado a cabo de una manera no dirigida e incluso la identificación de nuevos compuestos. De ahí que sea necesaria una metodología general que proporcione una señal cromatográfica (o cromatograma) independiente del instrumento.

En este sentido, la ***Publicación V*** describe el mismo procedimiento de agnostización de señales cromatográficas adaptado a su uso en la armonización de señales obtenidas a partir de diferentes plataformas de espectrometría de masas acopladas a sistemas LC con fines identificación y cuantificación. Un método para el análisis de plaguicidas fue elegido como ejemplo de analitos objetivo para verificar las ventajas de la metodología de agnostización (pepino como matriz representativa de las hortalizas con alto contenido de agua). La misma serie de compuestos químicos fueron utilizadas como mezcla estándar de referencia con ligeras modificaciones (cambios por la disponibilidad de patrones lo más parecidos posibles a los analitos objetivo). La Tabla 5 muestra la composición de la mezcla estándar de referencia utilizada satisfactoriamente en RP-LC, tanto para el método de biofenoles descrito anteriormente como para el método de determinación de plaguicidas por MS así como sus correspondientes SRS calculados.

La mezcla estándar y las muestras se analizaron en dos sistemas de MS diferentes y las señales fueron comparadas para su identificación en términos de SRS y cuantificación a alta y baja concentración después de la agnostización.

14. Brigante, F.I., Podio, N.S., Wunderlin, D.A., Baroni, M.V. Comparative metabolite fingerprinting of chia, flax and sesame seeds using LC-MS untargeted metabolomics. Food Chem., 371, 11355 (2022), 1-11. https://doi.org/10.1016/j.foodchem.2021.131355
15. Hergueta Castillo, M.E, López Rodríguez, E., López Ruiz, R., Romero González R., Garrido Frenich, A. Targeted and untargeted analysis of triazole fungicides and their metabolites in fruits and vegetables by UHPLC-orbitrap-MS^2. Food Chem., 368, 130860 (2022), 1-9. https://doi.org/10.1016/j.foodchem.2021.130860

Tabla 5. Compuestos químicos elegidos para constituir la mezcla patrón y puntuaciones de retención estándar (SRS) calculados.

Mezcla estándar método biofenoles (RP)HPLC-DAD		Mezcla estándar método plaguicidas (RP)UHPLC-MS	
Nombre	**SRS**	**Nombre**	**SRS**
Ác. 3,5-dihidroxibenzoico	1.000	Trimetoprim	1.000
Ác. 3,4-dihidroxibenzoico	2.275	Ác. 3,5-dihidroxibenzoico	1.114
Tirosol	3.246	Ác. Siríngico	1.805
Ác. Siríngico	4.162	Sulfadimetoxina	2.750
Fenol	5.633	Metillparabeno	3.750
Ác. Benzoico	6.691	Ác. Benzoico	4.140
4-hidroxibenzoato de metilo	7.113	Etillparabeno	6.417
Benzaldehido	7.570	Propiofenona	8.034
Acetofenona	8.330	Benzaldehido	8.725
4-hidroxibenzoato de etilo	9.051	Butirofenona	11.113
Ac. Naftalenoacético	10.370	Benzofenona	12.234
Propiofenona	11.113	Valerofenona	13.415
4-hidroxibenzoato de butilo	12.688	Benzoato de bencilo	13.668
Butirofenona	13.263	Hexanofenona	15.101
Benzofenona	14.149	Heptanofenona	16.391
Valerofenona	15.192	Octanofenona	17.429
Benzoato de bencilo	16.312	Nonanofenona	18.294
Hexanofenona	16.963	Decanofenona	19.018
Heptanofenona	18.618		
Octanofenona	20.182		
Butilhidroxitolueno	21.211		
Nonanofenona	21.638		
Decanofenona	22.984		

En base a los resultados se concluyó que los valores de SRS proporcionaron una mejor reproducibilidad y transferibilidad entre diferentes instrumentos de LC que el parámetro de tiempo de retención convencional, y pueden utilizarse como información de retención fiable. Por lo tanto, podría añadirse como un parámetro adicional en las bases de datos caseras para ser utilizado con fines de identificación cuando se compara la información de diferentes instrumentos de LC. Además, el uso de respuestas normalizadas, utilizando al menos dos estándares internos, proporcionó resultados adecuados para la cuantificación de los residuos de plaguicidas, minimizando el esfuerzo dedicado a este paso,

teniendo en cuenta que una curva de calibración, inyectada en un equipo, podría utilizarse para cuantificar muestras inyectadas en otro. Sin embargo, debido al limitado número de compuestos y matrices evaluados, estos resultados pueden considerarse preliminares y aún es necesario realizar trabajos futuros, en los que participen diferentes laboratorios para comparar los resultados en una situación real, aumentando el número de plaguicidas controlados, así como el control de otros tipos de compuestos en diferentes plataformas. Además, la colaboración conjunta con el grupo de investigación "Química Analítica de Contaminantes" de la Universidad de Almería hizo posible la aplicación de esta metodología para la identificación y cuantificación fiable de los residuos de un mayor número de plaguicidas en diferentes matrices utilizando diferentes plataformas de LC-MS, lo que dio resultado a un artículo publicado en la revista *J. Chromatogr. A* que se encuentra recogido en el Anexo 1 de esta Tesis.

V.2. Estudios multivariable de estabilidad y vida útil.

En la línea de garantizar a los consumidores un producto de máxima calidad, la autentificación y la determinación de su estabilidad o vida útil desempeñan un papel muy importante en el control de calidad de los alimentos por un lado para verificar que un producto se ajusta a las descripciones de la etiqueta para detectar adulteraciones y fraudes y, por otro, hacer un seguimiento del producto a través de cada una de las fases de fabricación y periodo de utilización o consumo. Es en este punto donde la Química Analítica juega un papel fundamental, ya que los analistas deben encargarse de desarrollar metodologías que satisfagan todas estas demandas.

La oxidación de los aceites y grasas es una de las principales causas del deterioro de la calidad de los alimentos, y su monitorización supone un reto para fabricantes y responsables del control de los alimentos por igual. El Instituto de Ciencia y Tecnología de los Alimentos (IFST) define el concepto de vida útil como el tiempo durante el cual un producto alimentario permanece seguro, cumple con la declaración de datos nutricionales del etiquetado y conserva las características sensoriales, químicas, físicas y microbiológicas deseadas cuando se almacena en las condiciones recomendadas [**16**]. Un estudio de vida útil consiste en definir un umbral de calidad para discriminar los productos alimentarios que todavía son tolerables para el consumo de los que ya no son adecuados (siempre en términos de calidad, no de seguridad alimentaria). Este umbral de calidad se define generalmente como el límite de

16. IFST. Shelf life of foods: guidelines for its determination and prediction. Institute of Food Science and Technology, London, 1993.

aceptabilidad y en aceites suele elegirse libremente según la política de calidad del productor.

Existen varias metodologías para evaluar la vida útil de los aceites vegetales comestibles mediante el seguimiento de la oxidación de los lípidos durante el su almacenamiento. Normalmente, se utilizan métodos analíticos físicos y químicos y pruebas sensoriales que proporcionan datos aislados para evaluar la estabilidad oxidativa y establecer su vida útil en base al cumplimiento de las normas relativo a sus características de calidad correspondiente [**17**, **18**].

En los últimos años, se ha avanzado en la comprensión de reacciones complejas relacionadas con la vida útil de los alimentos basándose en estudios cinéticos multiparamétricos [**19**, **20**] y en la aplicación de nuevos enfoques de análisis no dirigidos como la metodología de huella instrumental o 'fingerprinting' [**21**] que proporcionan información menos sesgada, así como en enfoques de modelización mediante la aplicación de métodos específicos de extracción de datos o quimiometría. Una tendencia importante en la evaluación de la vida útil de alimentos es el desarrollo de herramientas analíticas rápidas, sencillas y de bajo coste, adecuadas para la supervisión en línea o el control de calidad del proceso de oxidación del aceite.

Es por ello que en la ***Publicación VI*** se realizó una revisión detallada sobre los antecedentes en la aplicación de métodos de análisis y de tratamiento multivariable de datos, y las ventajas e inconvenientes que esto conlleva en los estudios de estabilidad de los aceites vegetales comestibles, tanto aceites vírgenes como refinados, en condiciones normales de envejecimiento y condiciones que aceleran este proceso. Aunque se encontraron en literatura un

---

17. Reglamento (UE) 2019/1604 de la comisión de 27 de septiembre de 2019 por el que se modifica 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, 2019.
18. Codex Alimentarius. Norma para los aceites vegetales con denominación CXS 210-1999. Normas alimentarias internacionales, 2019.
19. Chaudhry, M.M.A., Amodio, M.L., Babellahi, F., de Chiara, M.L.V., Amigo Rubio, J.M., Colelli, G. Hyperspectral imaging and multivariate accelerated shelf life testing (MASLT) approach for determining shelf life of rocket leaves. J. Food Eng., 238 (2018), 122-133. https://doi.org/10.1016/j.jfoodeng.2018.06.017
20. Kebede, B.T., Grauwet, T., Magpusao, J., Palmers, S., Michiels, C., Hendrickx, M., Van Loey, A. Chemical changes of thermally sterilized broccoli puree during shelf-life: investigation of the volatile fraction by fingerprinting-kinetics. Food Res. Int., 67 (2015), 264-271. http://dx.doi.org/10.1016/j.foodres.2014.10.017
21. Grauwet, T., Vervoort, L., Colle, I., Van Loey, A., Hendrickx, M. From fingerprinting to kinetics in evaluating food quality changes. Trends Biotechnol., 32, 3 (2014), 125-131. https://doi.org/10.1016/j.tibtech.2014.01.002

gran número de estudios que han aplicado herramientas multivariables para realizar estudios de comparación de estabilidad de aceites, en la mayoría de los casos no permitieron obtener datos sobre la vida útil de una forma práctica, debido a la falta de un límite de aceptabilidad claramente definido, o realizan predicciones altamente empíricas y fuertemente dependientes de las condiciones experimentales.

En base a esto en las ***Publicaciones VII*** y ***VIII*** se llevó a cabo un estudio de estimación de la vida útil para aceites de oliva virgen (AOV), y otros aceites refinados como oliva, orujo de oliva y aceites de semillas como girasol y maíz mediante la combinación de medidas físico-químicas y sensoriales aisladas y huellas instrumentales, empleando para ello métodos no supervisados y métodos de reconocimiento de patrones para el tratamiento de datos multivariables. El estudio se llevó a cabo bajo condiciones normales de oxidación, intentando reproducir las condiciones de luz y temperatura de los lineales de supermercados y grandes superficies de venta en las que se encuentran habitualmente expuestos.

Para ello, el primer objetivo fue conseguir un banco de muestras de aceites recién producidos, representativos de los aceites que habitualmente se encuentran a disposición del consumidor y con la información de trazabilidad necesaria para garantizar su fecha de producción, origen, variedad de fruto o tipo de semilla del que proceden, etc. Se consiguieron 95 muestras de AOV recien producidos, es decir, tiempo de envejecimiento 0 meses y 79 muestras de orujo, oliva (refinados) y semillas adquiridos en su mayoría de supermercados, por lo que se verificó que su fecha de producción y/o refinado fuese inferior a 2 meses.

El siguiente paso consistió en el almacenamiento de las muestras bajo condiciones controladas: 20 ± 5 ºC y expuestos a 12 horas de luz LED. Las muestras se dividieron en alícuotas de 60 mL en botes PET transparentes con un volumen de espacio de cabeza constante. Esto se hizo con el objetivo de que cada vez que se fuese a realizar la caracterización analítica en un mes determinado se tomara de un bote cerrado con volumen constante, además, de para que todas las muestras estuvieran en el mismo tipo de envase. Si se hubiesen dejado los aceites en su envase original, cada mes que se realizaran controles analíticos, el volumen de aceite es menor y por tanto el espacio de cabeza relleno de gases oxidantes favorecedores del enranciamiento sería mayor y no se estaría controlando de forma correcta la vida útil de los aceites. Los AOV se mantuvieron almacenados en estas condiciones durante 14 meses mientras que los demás aceites refinados y aceites de semillas se mantuvieron durante dos años. Los aceites fueron caracterizados mediante las siguientes determinaciones:

- **Índice de refracción (RI).** RI se refiere a la relación entre la luz en el vacío y su velocidad en un medio determinado. Se utiliza como índice para informar del grado de hidrogenación del aceite. Se ha reportado que RI de grasas y aceites aumentan durante el proceso de autoxidación [**22**].
- **Índice de peróxidos (PV).** Es una medida directa de la oxidación primaria que tiene lugar en los aceites y determina el contenido en hidroperóxidos. Es un método muy empírico y cualquier variación en el procedimiento puede afectar a los resultados.
- **Índice de anisidina (AV).** Es una medida de la oxidación secundaria de los aceites. Durante la oxidación secundaria se pueden producir diferentes tipos de triacilgliceroles oxidados a partir de hidroperóxidos lipídicos, como aldehídos, epóxidos y cetonas. AV es una medida espectrofotométrica que indica la concentración de aldehídos presentes en el aceite.
- **Absorbancia UV.** La oxidación de los ácidos grasos poliinsaturados va acompañada de un aumento de la absorción ultravioleta de los aceites. Los lípidos insaturados muestran un cambio en la posición del doble enlace durante la oxidación debido a la isomerización y la formación de conjugados. Los dienos conjugados resultantes (oxidación primaria) muestran un aumento en la absorción a 232 nm; de forma similar, los trienos conjugados producidos (oxidación secundaria) absorben a 270 nm.
- **Estabilidad oxidativa o tiempo de inducción Rancimat.** Consiste en una medida de la oxidación en condiciones aceleradas. Se llevó a cabo en el equipo Methrom 892 Professional Rancimat que registra los cambios de conductividad causada por los ácidos orgánicos volátiles iónicos, principalmente el ácido fórmico, de forma automática y continua. El punto final se selecciona donde la conductancia comienza un rápido aumento.
- **Análisis sensorial.** Los aceites vírgenes fueron analizados sensorialmente por catadores expertos en la búsqueda de defectos (especialmente atributo negativo rancio, principal defecto desarrollado en las condiciones de este estudio).
- **Cuantificación de las cantidades relativas de 1,2- y 1,3-diglicéridos por GC-FID.** Los diacilgliceroles o diglicéridos, son componentes menores de los AOV. En los aceites recién extraídos de la pulpa de la aceituna sana, casi la totalidad se encuentra en la disposición isomérica 1,2. Una medida

22. Shahidi, F., Wanasundara, U. Methods for Measuring Oxidative Rancidity in Fats and Oils. In: C. Akoh, D. Min, (Eds.). Food Lipids: Chemistry, Nutrition and Technology, 3rd ed., CRC Press, New York, 2008.

del cambio en el ratio 1,2- y 1,3-diglicéridos es un indicativo de la frescura, edad y calidad del aceite [**23**].

- **Cuantificación de tocoferoles: totales, α, β, γ y δ por HPLC-UV.** Son antioxidantes naturales presentes en el AOV. Se ha observado una disminución del contenido en tocoferoles en aceites expuestos a luz, como medida de la estabilidad que estos presentan [**24**].
- **Cuantificación de la proporción de pirofeofitina a por HPLC-DAD.** Las pirofeofitinas en el aceite de oliva se forman debido a la degradación de los pigmentos de clorofila (feofitinas). Los pigmentos se descomponen debido a un proceso que implica la descarbometoxilación de la clorofila y las feofitinas para formar pirofeofitinas.
- **Cuantificación de compuestos fenólicos por HPLC-DAD.** Los fenoles hidrofílicos influyen en gran medida no sólo en la actividad biológica y la estabilidad oxidativa del AOV, sino también en la calidad del sabor (atributos amargos y picantes). Incluso en pequeñas cantidades, los compuestos fenólicos del AOV son fundamentales para proteger de la oxidación, inhibiéndola mediante diversos mecanismos basados en la eliminación de radicales, la transferencia de átomos de hidrógeno y la quelación de metales [**25**]. Se han monitorizado las variaciones cuantitativas de los compuestos fenólicos y sus productos de degradación a lo largo del tiempo. Varios estudios han demostrado los efectos de las condiciones de almacenamiento sobre los fenoles del aceite, informando de una disminución del contenido fenólico total, de la oxidación de los derivados secoiridoides y del aumento del hidroxitirosol y del tirosol, entre otros [**26**].
- **Determinación de antioxidantes añadidos por HPLC-DAD:** Los antioxidantes fenólicos sintéticos, como el galato de propilo, 2,4,5-

---

23. Salvo, A., Rotondo, A., La Torre, G.L., Cicero, N., Dugo, G. Determination of 1,2/1,3-diglycerides in Sicilian extra-virgin olive oils by ^{1}H-NMR over a one-year storage period. Nat. Prod. Res., 31, 7 (2016), 822-828. https://doi.org/10.1080/14786419.2016.1247084
24. Caponio, F., Bilancia, M.T., Pasqualone, A., Sikorska, E., Gomes, T. Influence of the exposure to light on extra virgin olive oil quality during storage. Eur. Food Res. Technol., 221 (2005), 92-98. https://doi.org/10.1007/s00217-004-1126-8
25. Kotsiou, K., Tasioula Margari, M. Monitoring the phenolic compounds of Greek extra-virgin olive oils during storage. Food Chem., 200 (2016), 255-262. https://doi.org/10.1016/j.foodchem.2015.12.090
26. Castillo Luna, A., Criado Navarro, I. Ledesma Escobar, C., López Bascón, M., Priego Capote, F. The decrease in the health benefits of extra virgin olive oil during storage is conditioned by the initial phenolic profile. Food Chem., 336 (2020), 127730. https://doi.org/10.1016/j.foodchem.2020.127730

trihidroxibutirofenona, tert-butilhidroquinona, butilhidroxianisol, e hidroxitolueno butilado se han utilizado como aditivos alimentarios para retrasar o prevenir la oxidación de los lípidos en los aceites refinados de semillas, ya que su adición está prohibida en aceites vírgenes de oliva. Su presencia o no en el aceite puede estar directamente relacionado con la resistencia frente a la oxidación.

- ❒ **Caracterización de la fracción de compuestos volátiles por GC-FID en modalidad 'fingerprinting'.** En el proceso de oxidación secundaria de los aceites se producen numerosos compuestos volátiles causantes de los malos flavores (sensación conjunta de olor+sabor) que se desarrollan en el aceite, habitualmente relacionados con la percepción del atributo rancio. La interpretación de los cambios sensoriales durante el almacenamiento por medio de los compuestos volátiles llevada a cabo por diversos autores [**27**] requiere el estudio de cada uno de los compuestos, en particular de los que tienen propiedades odorantes mediante el empleo de patrones (no siempre disponibles) o técnicas complejas de análisis. En este caso, se obtuvieron las huellas cromatográficas completas de la fracción volátil tanto de los compuestos volátiles iniciales (relacionados con flavores positivos) y los compuestos volátiles producidos durante la oxidación y se evaluó sus cambios a lo largo del tiempo.

Con periodicidad bimestral, una alícuota de cada aceite era retirada de las condiciones de envejecimiento y analizada con cada uno de los métodos correspondientes. De esta forma, se obtuvo una gran cantidad de datos que se dispuso en forma de matriz para los distintos tipos de aceites en los diferentes tiempos de envejecimiento: matriz de datos de aceites de refinados de semillas (468 vectores de muestras $\times$ 20 variables); matriz de datos de aceites de orujo y oliva (252 vectores de muestras $\times$ 22 variables); y matriz de datos de aceites de oliva virgen (760 vectores de muestras $\times$ 30 variables). Estas matrices de datos fueron utilizadas para el establecimiento de las cinéticas de degradación y el modelo global de estabilidad desde un enfoque multivariable para cada tipo de aceite. Para que un análisis se considere multivariable todas las variables deben ser tenidas en cuenta y relacionadas de tal manera que el efecto que producen en conjunto no pueda ser interpretado de manera individual. Un modelo multivariable nos permitirá representar los datos de forma global y sencilla, establecer la distribución real e importancia de varias variables simultaneamente, desarrollar un modelo de predicción y toma de decisiones basado en todas las variables en su conjunto así como hallar las relaciones

27. Lobo Prieto, A., Tena, N., Aparicio Ruiz, R., Morales, M.T., García González, D.L. Tracking sensory characteristics of virgin olive oils during storage: interpretation of their changes from a multiparametric perspective. Molecules, 25 (2020), 1686. https://doi.org/10.3390/molecules25071686

causa-efecto entre las variables y, en el caso que nos ocupa, el tiempo de envejecimiento.

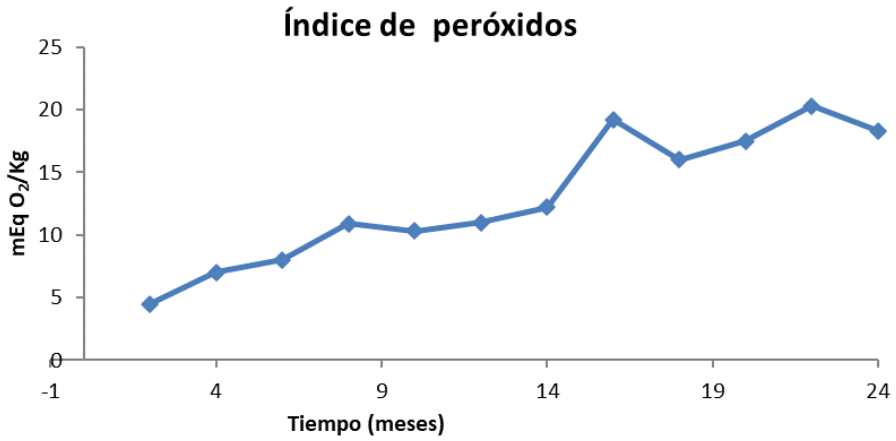
Dos herramientas quimiométricas fueron empleadas con este fin, análisis de componentes principales (PCA) y regresión de mínimos cuadrados parciales (PLSR). PCA como técnica de análisis no supervidado no tiene en cuenta la variable respuesta, Y, ya que el objetivo no es predecir sino extraer información empleando los predictores, por ejemplo, para identificar subgrupos. El principal problema al que se enfrentan los métodos no superviados es la dificultad para validar los resultados dado que no se dispone de una variable respuesta que permita contrastarlos. Por el contrario en PLSR se recoge la mayor variación entre las variables predictoras y respuestas en unas nuevas variables latentes (LV), por lo que consiste en una reducción de la dimensión de los datos supervisada por el resultado. Estos componentes se utilizan luego para ajustar el modelo de regresión. Se puede suponer que el modelo PLSR consiste en una relación externa y una relación interna donde la relación externa describe la variables almacenadas en las matrices de datos X e Y individualmente mientras que la relación interna une los dos matrices.

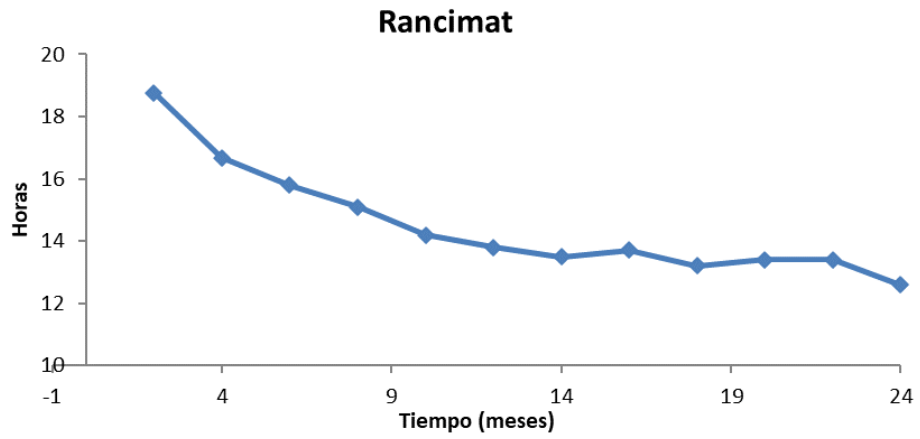
Por ello, para cada una de las matrices de datos descritas se realizó un PCA así como un PLSR considerando todas las variables con un doble objetivo: PCA para comprobar los posibles outliers y ver el agrupamiento natural de los datos y PLSR para obtener la importancia de las variables mediante los valores de "importancia de la variable en la proyección" (VIP) que se le otorgan a cada una de ellas en el modelo de regresión de los datos frente al tiempo de envejecimiento. Una variable con una puntuación VIP cercana o superior a 1 (uno) puede considerarse importante en un modelo determinado. Las variables con puntuaciones VIP significativamente inferiores a 1 son menos influyentes y pueden ser buenas candidatas para ser excluidas del modelo. Tal y como era de esperar, no todas las variables aportaban información relevante a los modelos establecidos, mostrando VIP cercanas a cero, mientras que otros pares de variables aportaban información similiar (medida en base al *coeficiente de correlación de Pearson* entre pares) y cabía el riesgo de estar sobreajustando el modelo. Por ello, se decidió no tener en cuenta aquellas variables que no presentaban una correlación significativa con el tiempo de envejecimiento y volver a repetir los modelos PCA y PLSR solo con estas variables para cada una de las matrices.

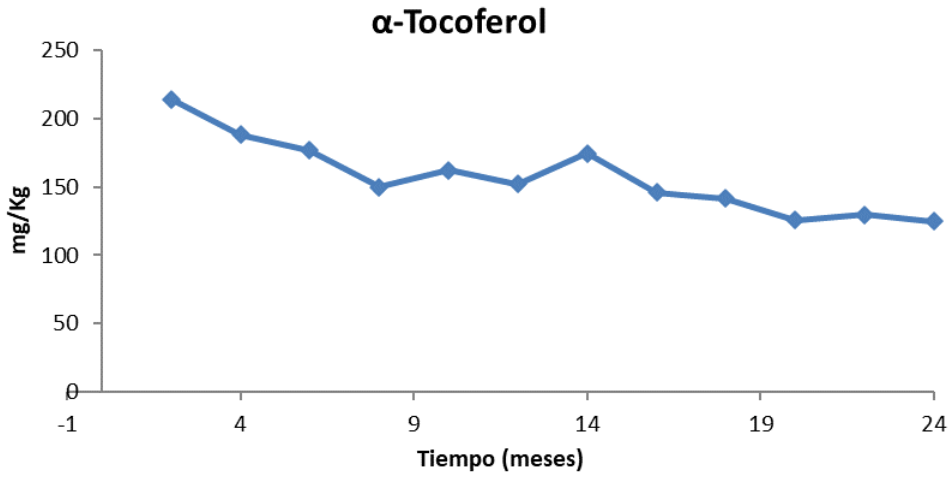
La Tabla 6 recoge las variables (o determinaciones analíticas) con mayor importancia y consideradas para definir los modelos de estabilidad para cada matriz de datos. En la Figura 4 se muestra la evolución en el tiempo de cada uno de los parámetros analíticos (o variables) seleccionados para la matriz de orujo-oliva: cada punto del gráfico representa el valor de la mediana de todos los aceites.

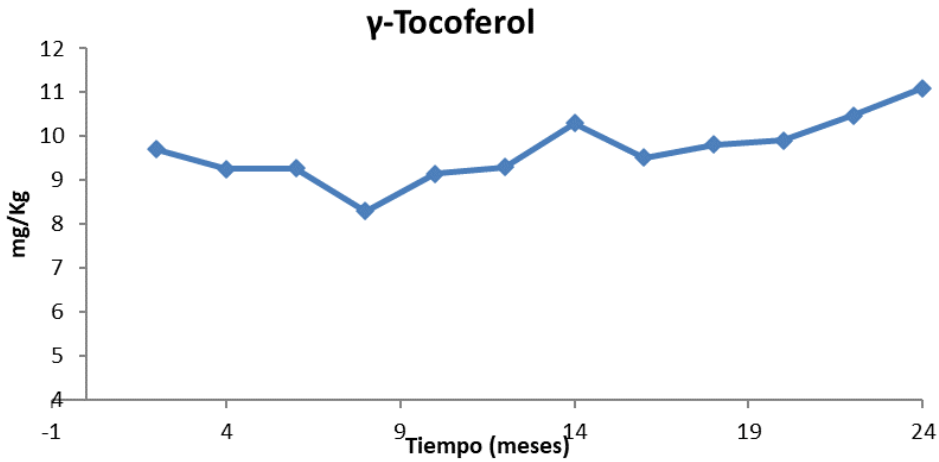
Tabla 6. Variables seleccionadas para establecer los modelos multivariables PCA y PLSR para cada matriz de datos.

Aceites de oliva vírgenes	Aceites refinados de orujo y oliva	Aceites refinados de semillas
Índice de peróxidos	Índice de peróxidos	Índice de peróxidos
K_{270}		K_{270}
K_{232}		
ΔK		
Estabilidad oxidativa	Estabilidad oxidativa	Estabilidad oxidativa
Pirofitina a	.	
α-Tocoferol	α-Tocoferol	Tocoferoles totals
β-Tocoferol	γ-Tocoferol	β-Tocoferol
	Índice de anisidina	Índice anisidina
1,2-Digliceridos		
Alcoholes fenólicos		
Rancidez (sensorial)		
Variables de la huella de volátiles referidas a los tiempos de retención (min): 2.12; 2.41; 3.01; 5.33; 7.28 y 8.31.	Variables de la huella de volátiles referidas a los tiempos de retención (min): 1.63; 1.78; 1.82; 2.32; 2.91; 2.95 y 4.99	Variables de la huella de volátiles referidas a los tiempos de retención (min): 1.63; 1.83; 2.37 y 2.95.




Rancimat
Horas
20
18
16
14
12
10
-1
4
9
14
19
24
Tiempo (meses)


α-Tocoferol
mg/Kg
250
200
150
100
50
0
-1
4
9
14
19
24
Tiempo (meses)


γ-Tocoferol
mg/Kg
12
11
10
9
8
7
6
5
4
-1
4
9
14
19
24
Tiempo (meses)

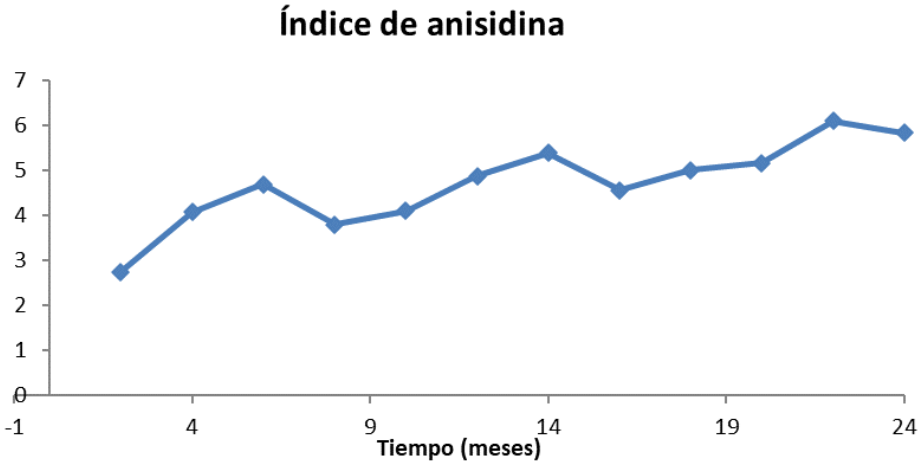


Figura 4. Evolución durante el almacenamietno de cada uno de los parámetros analíticos seleccionados para la matriz de orujo-oliva. Cada punto del gráfico representa el valor mediana de todos los aceites.

Los scores de PLSR se utilizaron para establecer un modelo multivariable de estabilidad o vida útil representativo para cada tipo de aceite siempre que estos se encuentran en condiciones de oxidación a temperatura ambiente y expuestos a luz. Este modelo nos permitirá calcular el índice de vida útil restante o tiempo en el que un aceite sigue manteniendo sus características mínimas de pureza y calidad según su denominación legal.

Para ello primero es necesario establecer el límite de aceptabilidad. Este límite se calculó multiplicando el vector de scores de una muestra considerada "no aceptable", es decir, que ha sobrepasado los límites establecidos en normativa, por cada uno de los vectores de 'loadings' o vector de ponderales de las variables para el modelo de estabilidad correspondiente, considerando todas las LVs del modelo. Para ello, se realizó un estudio particular de cada uno de los tipos de aceite a cada uno de los diferentes tiempos de envejecimiento y de los parámetros reglamentarios que las definen. Como ejemplo, las características para un AOVE están reguladas por el Reglamento 2568/91 así como por la norma australiana AS 5264 que establecen los siguientes valores críticos de calidad de estos aceites: valor de peróxidos < 20; K_{232} < 2.5; K_{270} < 0.22; ΔK < 0.01; mediana del defecto = 0; pirofeofitina a < 17% y 1,2-diacilglicéridos < 35%.

El vector de scores de las muestras "no conformes" para cada tipo de aceite se definió considerando los valores reglamentarios para aquellas variables que sí están reguladas y el valor de la mediana (considerando todas las muestras) de los no reglamentarios calculado para el mes que más se ajusten a los valores reglamentarios. Esta fue la propuesta principal de la investigación, tratándose

de un nuevo enfoque ya que no se encontraron antecedentes en bibliografía que mostraran propuestas similares para establecer un límite de aceptabilidad multivariable para la determinación del tiempo de vida útil. Se realizarón todas las estimaciones y cálculos necesarios y se concluyó que los aceites de orujo y oliva refinado eran los que presentaban mayor tiempo de vida útil, 14 meses, seguidos por AOV con 13 meses y los aceites de semillas con un tiempo de 9 meses (de acuerdo a las disposiciones que regulan sus características de denominación).

Una vez establecidos los modelos, y estimados el tiempo de vida útil de cada tipo de aceite se definieron una serie de índices que nos permite conocer el estado de oxidación (t_i) de cualquier aceite, independientemente de la fecha de producción y las condiciones en las que haya estado almacenado; así como el número de meses de vida útil restantes, siempre que, una vez calculado su t_i, dicho aceite se mantenga almacenado en las condiciones que se detallan en este estudio.

Actualmente se han realizado estimaciones para la validación interna de los resutados obtenidos, comprobando si efectivamente los modelos son capaces de predecir cuantos meses puede mantenerse un aceite almacenado en las condiciones descritas sin perdida de su categoría comercial y por tanto incurrir en fraude. Los resultados obtenidos están siendo satisfactorios, sobre todo para los AOV y orujo-oliva. Resultados algo menos precisos se está obteniendo para los aceites refinados de semillas, probablemente por la gran variabilidad tenida en cuenta en las muestras y por que las métricas de calidad de ajuste de los modelos establecidos eran peores, con mayores residuales y por tanto comportamientos dispares entre las muestras consideradas para entrenar los modelos. Recientemente se ha comenzado ha realizar una validación externa, con un mayor número de muestras, que garantice la fiabilidad de los resultados obtenidos y proporcione una seguridad tanto a productores como a consumidores sobre el índice que en este proyecto se ha propuesto. Esta fase se encuentra aún en desarrollo.

El último estudio incluído en esta tesis doctoral, cuyos resultados fueron plasmados en la ***Publicación IX***, consistió en una comparación de la estabilidad que presentan algunos de los aceites de mayor consumo cuando estos son utilizados en condiciones que aceleran el proceso de oxidación, concretamente, usados en un proceso de fritura. Para llevar a cabo el estudio se adquirieron directamente desde el supermercado aceites que contenian aditivos antioxidantes y antiespumantes y otros que no.

Entre los métodos de cocinado de alimentos, la fritura es un método de procesamiento de alimentos muy popular para producir alimentos sabrosos y con características únicas de sabor, olor y color. En este proceso el aceite

actúa como un medio líquido de transferencia homegénea de calor a una temperatura constante y contribuye al desarrollo de la textura y el sabor específicos de los alimentos fritos. Sin embargo, debido a las altas temperaturas a la que se lleba a cabo, a la presencia de oxígeno, al contenido de humedad del alimento y a la lixiviación de algunos de sus componentes se produce un deterioro oxidativo e hidrolítico del aceite: hidrólisis de los lípidos, ruptura del doble enlace y oxidación de los enlaces insaturados de las cadenas carbonadas [**28**]. Esto da lugar a la generación y acumulación de varios productos de degradación, favorecidndo la formación de compuestos polares totales (TPC) que tienen un mayor peso molecular y una mayor polaridad que los TAG normales inalterados y con efectos perjudiciales sobre la salud si son consumidos en altas cantidades.

Por ello, la determinación de los niveles de TPC puede proporcionar una primera indicación sobre si los aceites de fritura usados son todavía aptos para el consumo [**29**]. Los compuestos oxidados pueden estar presentes decenas de combinaciones por lo que la determinación cuantitativa de cada compuesto específico es relativamente difícil. Como alternativa, existen pruebas comerciales rápidas para determinar el TPC de los aceites de fritura, mediante medidas electroquímicas que requieren un equipo sencillo y barato, y permite abordar rápidamente cuestiones como la estimación de la ingesta saludable y la evaluación de la seguridad. Uno de los equipos más comunmente utilizados se basa en medidas de la constante dieléctrica utilizando un sensor electroquímico. Sin embargo, estos dispositivos necesitan una calibración específica, y no son adecuadas para ser empleadas en todo tipo de aceites. Además, la temperatura puede influir en las mediciones y las interferencias, como el agua, la sal y los minerales, pueden afectar a la polaridad de las muestras y ofrecer información falsa sobre el contenido de TPC de los aceites.

En este estudio, se simuló el proceso de fritura discontinuo de patatas por inmersión en el aceite, utilizando una freidora doméstica. Cada dos horas, se tomó una alícuota de cada aceite y se midió %TPC empleando un medidor de la constante dielétrica (FOM 320 [**30**]). Cada muestra de aceite fue expuesta a un número diferente de horas de fritura hasta superar el límite reglamentario del 25 %TPC o hasta un máximo de 32 horas cuando el contenido del aceite vegetal restante en la freidora fue insuficiente para realizar un nuevo ciclo de

---

28. Khor, Y.P., Hew, K.S., Abas, F., Lai, O.M., Cheong, L.Z., Nehdi, I.A., Sbihi, H.M., Gewik, M.M., Tan, C.P. Oxidation and polymerization of triacylglycerols: in-depth investigations towards the impact of heating profiles. Foods, 8, 475 (2019), 1-15. https://doi.org/10.3390/foods8100475
29. Boletín Oficial del Estado. Quality standard for oils and fats heated, dated 26th of January of 1989. BOE-A-1989-2265, (pp. 1-6), 1989.
30. Ebro. (2021). FOM 320 Manual. Ebro Electronic GmbH Co, KG.

fritura de forma adecuada (debido a la perdida de aceite natural por absorción en el alimento y a las multiples alícuotas ya extraídas). Según los datos obtenidos a partir de estas medidas un aceite de girasol alcanza el mayor %TPC en el menor tiempo mientras que el aceite de oliva mostró la mayor resistencia natural a la oxidación necesitando el mayor tiempo para superar el límite legal. Los aceites vegetales que contienen aditivos sintéticos no alcanzan el 25 %TPC ni siquiera después de 32 horas de fritura.

Para evitar elaborar conclusiones erroneas a partir de un modelado univariante, se realizaron determinaciones del índice de refracción, valor de peróxidos, índice de anisidina y absorbancia al UV de los aceites, factores cuya evolución en función del tiempo de fritura había sido reportada por otros autores [**31**]. Con todas estas variables, incluída los valores %TPC, se realizó un modelo multivariable PLSR para cada tipo de aceite frente al tiempo de fritura, que mostró buenos resultados en la predicción de los datos.

Además, como alternativa se describió un método de cromatografía de gases, en modalidad no dirigida, para monitorizar los cambios en las huellas cromatográficas correspondientes a las fracciones de los compuestos polares y no polares de los aceites, a medida que aumentaba el tiempo de fritura a los que éstos eran sometidos. Los cromatogramas registrados para cada fracción polar y no polar (17 minutos de análisis) se fusionaron en una única huella instrumental que sirvió como dato da entrada para elaborar un segundo modelo multivariable PLSR para cada tipo de aceite.

Ambos modelos fueron estadísticamente comparados danto lugar a resultados no diferentes. Los modelos quimiométricos realizados permitieron diferenciar entre los diferentes tipos de aceites vegetales y su estabilidad frente a la oxidación. Además, tambien fue posible distinguir entre los que contienen antioxidantes sintéticos y antiespumantes y los que no. Las conclusiones obtenidas en este estudio coincidieron con las de los encontrados en la literatura científica [**32**]. A partir de los datos cromatográficos, se comprobó que el aceite de oliva tenía la mayor estabilidad termo-oxidativa en comparación con los aceites de semillas como el de girasol o el de mezcla, pero no se diferenciaba significativamente del aceite de orujo y de los aceites de semillas que contenían aditivos sintéticos. No obstante, el análisis de los datos reveló que los modelos basados en la aplicación de un enfoque univariante, es decir,

---

31. Sahidi, F., Wanasundara, U.N. Methods for Measuring Oxidative Rancidity in Fats and Oils. In: C.C. Akoh & D.B. Min (Eds.). Food Lipids: Chemistry, Nutrition and Biotechnology, 3rd ed., CRC Press, Boca Raton, 2008.
32. Holgado, F., Ruiz Méndez, M.V., Velasco, J., Marquéz Ruiz, G. Performance of olive-pomace oils in discontinuous and continuous frying. Comparative behavior with sunflower oils and high-oleic sunflower oils. Foods, 10 (2021), 3081. https://doi.org/10.3390/foods10123081

la medición rápida del %TPC, puede llevar a conclusiones erróneas sobre la estabilidad y la vida útil nutricional de los aceites vegetales.

Por tanto, nuevamente quedaron demostradas las ventajas que proporciona el modelado multivariante desarrollado a partir de datos obtenidos de huellas instrumentales frente al enfoque tradicional basado en datos experimentales discretos.

CONCLUSIONES FINALES

CONCLUSIONES FINALES

En esta sección se recogen las conclusiones derivadas de los estudios presentados en esta Tesis Doctoral.

I. Se ha revisado el potencial del enfoque de la huella instrumental no dirigida en el ámbito alimentario. Se ha recogido el fundamento de las técnicas más adecuadas para la obtención de huellas instrumentales en función de la información que estas proporcionan. Se ha incluido una breve explicación de las bases para el desarrollo de métodos multivariables cualitativos y cuantitativos. Además, se ha discutido el proceso de validación convencional y un nuevo enfoque para establecer criterios de validación de los métodos multivariables de clasificación/discriminación. Todo esto dio lugar a un capítulo de libro publicado.

II. Se describió una metodología adecuada para la discriminación de diferentes variedades y distintos orígenes geográficos de aguacates en base a la huella cromatográfica lipídica obtenida por HPLC con dos sistemas de detección distintos: CAD y DAD. Se demostró que las condiciones de (NP)HPLC son mejores que las de (RP)HPLC para este propósito, mostrando un menor tiempo de análisis y un mejor rendimiento de los modelos de clasificación. SVM mostró ser un mejor clasificador que PLS-DA y SIMCA con mejores resultados en lo que respecta a las métricas de validación generalmente aplicadas. Todo este trabajo resultó en la publicación de dos artículos científicos.

III. Se ha establecido una metodología novedosa para "agnostizar" cromatogramas 1D, haciendo que la nueva señal o huella cromatográfica sea independiente del estado cromatográfico. De este modo, las huellas de dos instrumentos diferentes estarían estandarizadas frente a una referencia común y tendrían un alto grado de similitud. Esto permitirá establecer bases de datos cromatográficos característicos para armonizar y compartir información de análisis comparable para cada matriz alimentaria. Además, se describió la aplicación del mismo procedimiento de agnostización para la armonización de señales obtenidas a partir de diferentes plataformas de LC-MS con fines identificación y cuantificación. Se publicaron un artículo y un capítulo de libro que recogían estas aplicaciones.

IV. Se revisó detalladamente las técnicas quimiométricas y los enfoques multivariables utilizados en estudios de estabilidad y vida útil de aceites vegetales comestibles, vírgenes y refinados, bajo almacenamiento en condiciones ambientales y condiciones forzadas de oxidación lo que resultó en una publicación.

IV.1. Se han llevado cabo estudios multivariable que han permitido predecir la vida útil de aceites de oliva virgen, oliva refinado, orujo y aceites de semillas almacenados bajo condiciones ambientales. Los resultados de esta investigación se recogieron en dos artículos científicos.

IV.2. Se ha propuesto un nuevo método de "huella instrumental" GC ('fingerprinting') que permitió la monitorización de la oxidación de aceites refinados tras ser sometidos a un proceso domestico de fritura discontinuo. Con los datos obtenidos, se realizaron modelos quimiométricos que permitieron diferenciar entre los diferentes tipos de aceites vegetales y su estabilidad frente a la oxidación tal y como se ha mostrado en el correspondiente artículo científico.

V. Se ha realizado una discusión integrada de toda la investigación realizada en la etapa predoctoral en base a los objetivos generales con los que se planteó esta Tesis Doctoral centrada en la autentificación de alimentos vegetales de alto contenido graso, en términos de calidad y estabilidad. Se concluye que el trabajo presentado cumple con los objetivos establecidos inicialmente y abre camino hacia investigaciones futuras.

ANEXO 1

ANEXO 1

Artículos no incluidos en esta tesis doctoral

Aparte de los artículos recogidos en la tesis Doctoral y ampliamente comentados en esta memoria, se ha llevado a cabo un trabajo de investigación complementario, en colaboración con otros investigadores del propio grupo de investigación, y que fue realizado durante mi etapa predoctoral. Esto ha dado lugar a publicaciones científicas no recogidas en los capítulos anteriores de esta Tesis, y cuya referencia se detalla a continuación:

- Ana M. Jiménez-Carvelo, Sandra Martín-Torres, Fidel Ortega-Gavilán, J. Camacho. PLS-DA vs sparse PLS-DA in food traceability. A case study: authentication of avocado samples. *Talanta, 224,* 2021, 121904. https://doi.org/10.1016/j.talanta.2020.121904
- Luis Cuadros-Rodríguez, Fidel Ortega-Gavilán, Sandra Martín-Torres, Alejandra Arroyo-Cerezo, Ana M. Jiménez-Carvelo. Chromatographic fingerprinting and food identity/quality: potentials and challenges. *Journal of Agricultural and Food Chemistry, 61*, 2021, 14428-14434. https://doi.org/10.1021/acs.jafc.1c05584
- Luis Cuadros-Rodríguez, Fidel Ortega-Gavilán, Sandra Martín-Torres, Santiago Medina-Rodríguez, Ana M. Jiménez-Carvelo, Antonio González-Casado, M. Gracia Bagur-González. Standardization of chromatographic signals – Part I: towards obtaining instrument-agnostic fingerprints in gas chromatography. *Journal of Chromatography A*, 1641, 2021, 461983. https://doi.org/10.1016/j.chroma.2021.461983
- Christian H. Pérez-Beltrán, Ana M. Jiménez-Carvelo, Sandra Martín-Torres, Fidel Ortega-Gavilán, Luis Cuadros-Rodríguez. Instrument-agnostic multivariate models from normal phase liquid chromatographic fingerprinting. A case study: authentication of olive oil. *Food Control, 137*, 2022, 108957. https://doi.org/10.1016/j.foodcont.2022.108957
- Rosalía López-Ruiz, Roberto Romero-González, Sandra Martín-Torres, Ana M. Jiménez-Carvelo, Luis Cuadros-Rodríguez, Antonia Garrido Frenich. Applying an instrument-agnostizing methodology for the standardization of pesticide quantitation using different liquid chromatography-mass spectrometry platforms: a case study. *Journal of Chromatography A, 1664*, 2022, 462791. https://doi.org/10.1016/j.chroma.2021.462791

UNIVERSIDAD
DE GRANADA

Sandra Martín Torres