Food Chemistry

# Stepwise strategy based on 1H-NMR fingerprinting in combination with chemometrics to determine the content of vegetable oils in olive oil mixtures

--Manuscript Draft--

| | |
|---|---|
| **Manuscript Number:** | FOODCHEM-D-21-01386R1 |
| **Article Type:** | Research Article (max 7,500 words) |
| **Keywords:** | Olive oil; nuclear magnetic resonance; Multivariate data analysis; decision tree; adulteration; Authentication |
| **Corresponding Author:** | Rosa Maria Alonso-Salces, Ph.D.<br>CONICET, Universidad Nacional de Mar del Plata, Facultad de Ciencias Exactas y Naturales, Consejo Nacional de Investigaciones Cientificas y Tecnicas (CONICET)<br>Mar del Plata, ARGENTINA |
| **First Author:** | Rosa Maria Alonso-Salces, Ph.D. |
| **Order of Authors:** | Rosa Maria Alonso-Salces, Ph.D. |
| | Luis Angel Berrueta, Professor |
| | Beatriz Quintanilla-Casas, MSc |
| | Stefania Vichi, PhD |
| | Alba Tres, PhD |
| | María Isabel Collado, PhD |
| | Carlos Asensio-Regalado, MSc |
| | Gabriela Elena Viacava, PhD |
| | Aimará Ayelen Poliero, BSc |
| | Enrico Valli, PhD |
| | Alessandra Bendini, PhD |
| | Tullia Gallina Toschi, Professor |
| | José Manuel Martínez-Rivas, PhD |
| | Wenceslao Moreda, PhD |
| | Blanca Gallo, Professor |
| **Abstract:** | 1H-NMR fingerprinting of edible oils and a set of multivariate classification and regression models organised in a decision tree is proposed as a stepwise strategy to assure the authenticity and traceability of olive oils and their declared blends with other vegetable oils (VOs). Oils of the 'virgin olive oil' and 'olive oil' categories and their mixtures with the most common VOs, i.e. sunflower, high oleic sunflower, hazelnut, avocado, soybean, corn, refined palm olein and desterolized high oleic sunflower oils, were studied. Partial least squares (PLS) discriminant analysis provided stable and robust binary classification models to identify the olive oil type and the VO in the blend. PLS regression afforded models with excellent precisions and acceptable accuracies to determine the percentage of VO in the mixture. The satisfactory performance of this approach, tested with blind samples, confirm its potential to support regulations and control bodies. |

The final published version of this article is available online at: https://doi.org/10.1016/j.foodchem.2021.130588
© 2021 Elsevier. This manuscript version is made available under the Creative Commons Attribution–NonCommercial–NoDerivs (CC BY–NC–ND) 4.0 International License
(https://creativecommons.org/licenses/by–nc–nd/4.0)

## Highlights

- NMR fingerprinting & chemometrics to authenticate pure & legal blends of olive oil
- $^{1}$H-NMR & pattern recognition to detect adulteration olive oil with vegetable oils
- Stepwise strategy based on NMR spectral data and classification & regression models
- Olive oil traceability using decision trees with classification & regression models
- Determination of the botanical nature and the percentage of each oil in a mixture

REVISED Manuscript (text UNmarked)

# Stepwise strategy based on $^{1}$H-NMR fingerprinting in combination with chemometrics to determine the content of vegetable oils in olive oil mixtures

Rosa María Alonso-Salces[1,*], Luis Ángel Berrueta[2], Beatriz Quintanilla-Casas[3], Stefania Vichi[3], Alba Tres[3], María Isabel Collado[4], Carlos Asensio-Regalado[2], Gabriela Elena Viacava[5], Aimará Ayelen Poliero[6], Enrico Valli[7], Alessandra Bendini[7], Tullia Gallina Toschi[7], José Manuel Martínez-Rivas[8], Wenceslao Moreda[9], Blanca Gallo[2]

[1] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Departamento de Biología, Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Mar del Plata (UNMdP), Funes 3350, 7600, Mar del Plata, Argentina.

[2] Departamento de Química Analítica, Facultad de Ciencia y Tecnología, Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU), P.O. Box 644, 48080, Bilbao, Spain.

[3] Food Science and Nutrition Department, XaRTA (Catalonian Reference Network on Food Technology), University of Barcelona (UB), Food and Nutrition Torribera Campus, Av. Prat de la Riba, 171. 08028, Sta. Coloma de Gramanet, Spain.

[4] SGIKER, Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU), Barrio Sarriena s/n, 48940, Leioa, Spain.

[5] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Grupo de Investigación en Ingeniería en Alimentos (GIIA), Departamento de Ingeniería Química y en Alimentos, Facultad de Ingeniería, Universidad Nacional de Mar del Plata (UNMdP), Juan B. Justo 4302, 7600, Mar del Plata, Argentina.

[6] Departamento de Biología, Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Mar del Plata (UNMdP), Funes 3350, 7600, Mar del Plata, Argentina.

[7] Department of Agricultural and Food Sciences, Alma Mater Studiorum University of Bologna (UNIBO), Piazza Goidanich, 60, I-47521, Cesena (FC), Bologna, Italy.

[8] Department of Biochemistry and Molecular Biology of Plant Products, Instituto de la Grasa-CSIC, Ctra. Utrera km 1, Building 46, E-41013 Sevilla, Spain.

[9] Department of Characterization and Quality of Lipids, Instituto de la Grasa-CSIC, Ctra. Utrera km 1, Building 46, E-41013 Sevilla, Spain.

[*] Corresponding author.

E-mail address: rosamaria.alonsosalces@gmail.com (R.M. Alonso-Salces).

**Abstract**

$^{1}$H-NMR fingerprinting of edible oils and a set of multivariate classification and regression models organised in a decision tree is proposed as a stepwise strategy to assure the authenticity and traceability of olive oils and their declared blends with other vegetable oils (VOs). Oils of the 'virgin olive oil' and 'olive oil' categories and their mixtures with the most common VOs, i.e. sunflower, high oleic sunflower, hazelnut, avocado, soybean, corn, refined palm olein and desterolized high oleic sunflower oils, were studied. Partial least squares (PLS) discriminant analysis provided stable and robust binary classification models to identify the olive oil type and the VO in the blend. PLS regression afforded models with excellent precisions and acceptable accuracies to determine the percentage of VO in the mixture. The satisfactory performance of this approach, tested with blind samples, confirm its potential to support regulations and control bodies.

Keywords: olive oil, nuclear magnetic resonance, multivariate data analysis, decision tree, adulteration, authentication

## 1. Introduction

The high price of olive oil, the distinctive sensory profile, and its reputation as a healthy source of dietary fats make olive oil a target for fraud. The most common types of olive oil fraud are illegal blending with other vegetable oils (VOs) or low-quality olive oils, deliberate mislabelling of less expensive classes of olive oils, other vegetable oils or their blends with olive oils, and mislabelling of the geographical origin or Protected Designation of Origin declaration. Indeed, the European Parliament pointed out that olive oil adulteration has become one of the biggest financial fraud in the agricultural sector, and evidenced the need to update and harmonize analytical methods for quality and authenticity control of olive oil (EC, 2020; European Parliament, 2014). In this context, the so-called OLEUM Project was supported by the European Commission with the overall objective of improving existing analytical methods and developing new strategies of analysis for assuring the quality and authenticity of olive oil (OLEUM Project, 2016).

The EU Regulation 29/2012 standardises the labelling of all olive oil categories and their mixtures with other VOs, allowing to highlight the presence of olive oil on the label outside the ingredient list, only if it accounts for at least 50% of the blend (EC, 2012). However, this regulation and its amendments do not refer to any analytical parameter or method to control the percentage of olive oil in the admixture or the botanical origin of oil. The need of analytical methods to confirm the presence of olive oil in the blend, to distinguish pure and adulterated olive oils, to identify the adulterant oils in the mixture, as well as to determine the percentage of olive oil and the adulterants in the blend, is evidenced and is an issue of major concern in order to implement the established regulations (Conte, Bendini, Valli, Lucci, Moret, Maquet, et al., 2020). In literature, few works deal with the verification of the percentage of olive oil in fraudulent blends with VOs with regard to the labelling compliance of Reg. (EU) 29/2012 (De la Mata, Dominguez-Vidal, Bosque-Sendra, Ruiz-Medina, Cuadros-Rodríguez, & Ayora-Cañada, 2012; Gómez-Coca, Pérez-Camino, Martínez-Rivas, Bendini, Gallina Toschi, & Moreda, 2020; Monfreda, Gobbi, & Grippa, 2012; Santos, Kock, Santos, Lobo, Carvalho, & Colnago, 2017).

The chemical methods traditionally used in food analysis are laborious, time-consuming, non-eco-friendly and require sample preparation and skilled operators. In contrast, metabolomic approaches based on advanced instrumental techniques, such as MS and NMR, coupled to chemometrics overcome some of these operational drawbacks and provide useful tools for food quality control and traceability (Lioupi, Nenadis, & Theodoridis, 2020). Most of the NMR approaches developed for olive oil authentication, detection of olive oil adulteration and to determine the composition of olive oil blends with VOs, were based on measuring NMR signals that give quantitative information of certain compounds or are used to calculate some parameters and ratios (i.e. profiling) (Agiomyrgianaki, Petrakis, & Dais, 2010; García-González, Mannina, D'Imperio, Segre, & Aparicio, 2004; Jiang, Li, Chen, & Weng, 2018; Mannina, D'Imperio, Capitani, Rezzi, Guillou, Mavromoustakos, et al., 2009; Popescu, Costinel, Dinca, Marinescu, Stefanescu, & Ionete, 2015; Vigli, Philippidis, Spyros, & Dais, 2003; Zamora, Alba, & Hidalgo, 2001). Instead, NMR fingerprinting was only reported in few studies using low-field NMR spectroscopy (Parker, Limer, Watson, Defernez, Williamson, & Kemsley, 2014; Santos et al., 2017; Wang, Wang, Hou, & Nie, 2020). To the authors' knowledge, high-field NMR fingerprinting has been used to study mixtures of olive oil with other VOs for the first time in the present work. This study aimed to develop an analytical strategy based on $^{1}$H-NMR fingerprinting together with multivariate classification and regression models organised in a decision tree to determine the composition of an oil blend from both points of view, the botanical nature of the oils and the percentage of each oil in the blend. The performance of the complete stepwise analytical strategy is evaluated by the prediction results obtained for an external set of blind oil samples and commercial oils. It is worth noting that this analytical approach addresses some issues not considered in previous studies: (*i*) the discrimination between oil samples containing oil of the 'virgin olive oil' category (VOO) and the 'olive oil' category (OO); (*ii*) the distinction of pure and blended oils; and (*iii*) the study of a large sample set with pure oils and blends of the most common VOs used for olive oil adulteration, and a wide range

of % VO in the blend (including the percentages for the labelling verification in compliance with Reg. (EU) 29/2012).

## 2. Material and methods

### 2.1. Samples

Genuine samples of virgin (VOO) and extra virgin olive (EVOO) oils (n=176), olive oils (OO, n=3), refined conventional sunflower oil (normal type sunflower oil, NTSO, n=17), refined high oleic sunflower oil (HOSO, n=16), desterolized and deodorized high oleic sunflower oil (DOSO, n=1), refined hazelnut oil (HR, n=11), virgin hazelnut oil (HV, n=6), refined soybean oil (S, n=10), virgin avocado oil (EVAO, n=1), refined avocado oil (RAO, n=1), refined palm olein oil (RPOO, n=1) and refined corn oil (CO, n=1) were used to prepare binary mixtures at different percentages (2–90%) of VOs in VOOs or OOs (1007 blends). Samples were obtained in the framework of the AUTENFOOD and OLEUM projects. Oils from the sample banks of both projects were produced during two consecutive harvest years (2016/17 and 2017/18). Besides, eight commercial oil samples collected in the Swedish market were analysed. According to their labels, the commercial oils were described as mixtures of VOO and other VO such as rapeseed oil, sunflower oil, or non-identified vegetable oil.

Blends were prepared and preserved under controlled temperature conditions. All pure and blended oil samples were bottled with nitrogen headspace or minimal air headspace, kept at -20 ºC and protected from light. Before analysis, oil samples were taken from the cold storage, left to equilibrate at room temperature at least for 12 h, and shaken vigorously before sampling the oil aliquot for analysis.

### 2.2. Chemicals

Deuterated chloroform for NMR analysis (99.8 atom % D) was provided by Sigma-Aldrich Chemie (Steinheim, Germany).

## 2.3. NMR analysis

Aliquots of 150 µL of each oil sample were dissolved in 750 µL of deuterated chloroform, shaken in a vortex, and placed in a 5 mm NMR capillary. The $^{1}$H-NMR experiments were performed at 300K on a Bruker (Rheinstetten, Germany) Avance 500 (nominal frequency 500.13 MHz) equipped with a 5 mm broadband inverse probe with Z-gradients. The spectra were recorded using a 6.1 µs pulse (90°), an acquisition time of 3.5 s (50k data points) and a total recycling time of 7.0 s, a spectral width of 7100 Hz (14 ppm), 32 scans (+ 4 dummy scans), with no sample rotation. Prior to Fourier transformation, the free induction decays (FIDs) were zero-filled to 64k and a 0.3 Hz line-broadening factor was applied. The chemical shifts were expressed in δ scale (ppm), referenced to the residual signal of chloroform (7.26 ppm). The spectra were phase- and baseline-corrected manually, binned with 0.02 ppm-wide buckets, and normalized to total intensity over the region 4.10–4.26 ppm (glycerol signal). The region of the NMR spectra studied comprised from 0 ppm to 11 ppm. TopSpin 2.1 (2013) and Amix-Viewer 3.7.7 (2006) from Bruker BioSpin GMBH (Rheinstetten, Germany) were used to perform the processing of the spectra. The data table generated with the spectra of all samples, excluding the eight buckets in the reference region 4.10–4.26 ppm, was then submitted to multivariate data analysis.

## 2.4. Data analysis

Datasets were made up of the 542 buckets of the $^{1}$H-NMR spectra (variables in columns) measured on the oil samples (samples in rows). A total number of 1239 pure and blended oil samples were analysed by $^{1}$H-NMR. Depending on the aim of the multivariate model to be developed, the dataset contained the NMR spectral data of the corresponding studied samples. Datasets were analysed by univariate procedures (ANOVA, Fisher index and Box & Whisker plots); and by multivariate techniques, unsupervised such as principal component analysis (PCA), and supervised as partial least squares discriminant analysis (PLS-DA) and partial least squares regression (PLS-R) (Berrueta, Alonso-Salces, & Héberger, 2007). Data analysis was performed by means of the

150 statistical software package Statistica 7.0 (StatSoft Inc., Tulsa, OK, USA, 1984–2004) and The
151 Unscrambler v9.7 (Camo Software AS, 1986–2007).

152 PCA, PLS-DA and PLS-R were applied to the autoscaled or centered data matrix of $^{1}$H-NMR
153 spectra of the oil samples. The presence of outliers in the dataset was analysed by PCA. In PLS-DA
154 and PLS-R, the optimal number of PLS-components is estimated by cross-validation by plotting the
155 root mean square error in the prediction (RMSEP) against the number of PLS-components. The
156 model with the smallest number of features should be accepted from among equivalent models on
157 the training set in order to avoid overfitting (according to the principle of parsimony). In PLS-DA,
158 once the number of PLS-components is optimised, the predictions in the training-test set are
159 represented in a box and whisker plot in order to define the half of the distance between the
160 quartiles as the boundary. The regression coefficients (B) of the optimal number of PLS-
161 components denote the importance of the NMR variables on the model: the larger the B-coefficient,
162 the higher the influence of the variable on the PLS-DA or PLS-R model. A large B-coefficient may
163 also indicate a variable with small absolute values but large relative differences (Esbensen, Guyot,
164 Westad, & Houmøller, 2002). PLS-DA and PLS-R models were validated by 3-fold or leave-one
165 out cross-validation for parameter optimization, and by external validation when an external set of
166 samples was available. Binary classification models can lead to artefacts if they are not used and
167 validated properly (Kjeldahl & Bro, 2010). The reliability of the classification models developed
168 was studied in terms of recognition and prediction abilities in the cross-validation, and prediction
169 ability in the external validation (Berrueta et al., 2007). The goodness of the regression model fit
170 was evaluated by means of the prediction error, the correlation coefficient between predicted and
171 measured values in calibration and validation (R-cal, R-val), the determination coefficient in
172 calibration and validation ($R^2$-cal, $R^2$-val), and the evaluation of the residuals. The RMSEP is the
173 practical average prediction error estimated by the validation set (empirical error estimate expressed
174 in the original measurement units). The result is expressed as the predicted Y-value ± 2 RMSEP.

The R-RMSEP is the relative prediction error in % (comparable to the analytical accuracy) (Esbensen et al., 2002).

## 3. Results and discussion

### 3.1. Mixtures of olive oil with vegetable oils

Oils of the VOO and OO categories and their mixtures with the most common VOs used for the adulteration of olive oil or making 'legal' blends, i.e. NTSO, HOSO, DOSO, HR, HV, S, EVAO, RAO, RPOO and CO, were studied. The $^{1}$H-NMR spectra of the oil samples, both pure and blended (binary mixtures of VO with VOO or OO) oils, were recorded. The chemical shifts of the $^{1}$H-signals and their assignments to protons of the different functional groups are shown in Table S1 (supplementary material). The $^{1}$H-NMR profiles of the oil samples presented characteristic patterns of triglycerides, diglycerides and some minor constituents of the unsaponifiable fraction, which are useful for the determination of the botanical origin of oils and the composition of blended oils (Agiomyrgianaki et al., 2010; Alonso-Salces, Segebarth, Garmón-Lobato, Holland, Moreno-Rojas, Fernández-Pierna, et al., 2015; García-González et al., 2004; Guillén & Ruiz, 2003; Mannina et al., 2009; Parker et al., 2014; Popescu et al., 2015; Vigli et al., 2003; Wang et al., 2020).

The proposed approach to detect blends of olive oils (VOOs or OOs) with other VO and quantify the % VO in the blend is based on the use of the $^{1}$H-NMR fingerprint of the oil and a set of multivariate classification and regression models organized in a decision tree (Figures 1 and S1 in supplementary material). The PLS-DA and PLS-R models achieved and their chemical interpretation are described in the next sections. The most influential variables on the models were not completely discriminant unless otherwise specified.

### 3.2. PLS-DA model to confirm the presence of VOO or OO

The first stage of the decision tree (Figure 1) consists in identifying whether the oil sample contains VOO or OO using PLS-DA model-1 with recognition and prediction abilities of 97% and 98% for

the VOO and OO classes respectively (Table 1). The most influential NMR variables on the model were the $^{1}$H-signals of oleic acid (#7b, #9b), linolenic acid (#10c, #13d) and saturated fatty acids (#9a), exhibiting higher intensities in VOO and their blends than in samples containing OO. In contrast, the $^{1}$H-signals of linoleic acid (#12b) and *sn*-1,3-diacylglycerides (#17) presented lower intensities in the VOO class. These observations are consistent with previous studies reporting the differences in the composition of oleic, linolenic and saturated fatty acids and *sn*-1,3-diacylglycerides between VOOs and OOs (Guillén et al., 2003; Jiang et al., 2018).

Once the oil sample is classified as containing VOO or OO, further predictions are made using the binary classification models built separately for each type of olive oil to elucidate whether the olive oil sample is mixed with a VO, in which proportion (low or high) and with which particular VO (Figure 1).

### 3.3. PLS-DA models to discriminate blends of VOO with VO

For blends containing VOO, PLS-DA model-2 classifies the oil sample according to the proportion of VO in the mixture, i.e. low (0–20% VO in VOO) and high (25–90% VO in VOO), with correct prediction abilities of 98% and 97% respectively (Table 1). The most important variables on this model were the $^{1}$H-signals of oleic acid (#9b) and squalene (#11), whose signal intensities were higher in the low class. Indeed, VOO is known to be one of the vegetable oils that presents the highest contents of oleic acid and squalene (Jiang et al., 2018; Popescu et al., 2015; Vigli et al., 2003).

Pure VOOs are distinguished from blends with 2–20% VO in VOO, being identified even 92% of the pure VOOs and 90% of the VO-VOO blends (PLS-DA models 3 and 4 in Table 1). The main $^{1}$H-signals involved in the distinction of both classes were due to saturated fatty acids (#7a, #9a), which exhibited lower intensities in the VO-VOO class. In fact, saturated fatty acids are the second major class of fatty acids in VOO, being present in higher or similar concentrations than in the VOs studied, i.e. NTSO, HOSO, EVAO, HV, HR and S (Contiñas, Martínez, Carballo, & Franco, 2008;

Guillén et al., 2003; Jabeur, Zribi, Makni, Rebai, Abdelhedi, & Bouaziz, 2014; Jiang et al., 2018; Jović, Smolić, Primožič, & Hrenar, 2016; Monfreda et al., 2012; Ranade & Thiagarajan, 2015; Yang, Ferro, Cavaco, & Liang, 2013). Concerning the discrimination of blends of 2% VO in VOO for a certain VO, a satisfactory classification model was only achieved for soybean oil; thus, all blends with 2% S in VOO were detected, and 97% of the blends with 2% of other VO in VOO were correctly predicted (PLS-DA model-5 in Table 1).

The $^{1}$H-NMR fingerprint of an oil sample classified in the low class (0–20% VO in VOO) is then submitted to classification models developed for each VO (PLS-DA models 6–24) to identify which particular VO is contained in the oil sample (Tables 2 and S2–S3 in supplementary material). The classification abilities of the PLS-DA models were better when the dataset contained only the data of blended oils with 5–20% VO in VOO than when data of pure VOO and/or 2% VO in VOO was also included. The prediction abilities ranged between 83% and 98% of hits depending on the VO blended with VOO. Similarly, when an oil sample is classified in the high class (25–90% VO in VOO), its $^{1}$H-NMR fingerprint is submitted to PLS-DA models developed for mixtures of 20–90% VO in VOO (PLS-DA models 25–28 in Table 3) to identify the VO contained in the blend. In the present study, only binary mixtures of NTSO, HOSO, EVAO or HV with VOO were available in the range of 20–90% VO. The recognition and prediction abilities of the classification models built to determine whether the VOO blend contained NTSO, HV or EVAO were 99–100% for both classes, and 100% for the non-HOSO class and 92% for the HOSO class.

Regarding the most influential variables on the models, the $^{1}$H-signal of oleic acid (#9b) was completely discriminant between VOO mixtures with high % NTSO and those with other VOs. The blends of 20–90% NTSO in VOO contained significantly lower amounts of oleic acid than VOO blends with 20–90% HOSO, EVAO or HV. It is well-documented that virgin hazelnut oil, high oleic sunflower oil and virgin avocado oil present significantly higher contents of oleic acid than sunflower oil (Contiñas et al., 2008; Guillén et al., 2003; Jabeur et al., 2014; Jović et al., 2016;

Ranade et al., 2015; Vigli et al., 2003; Yang et al., 2013). Other important variables to discriminate the presence of NTSO in VOO were the $^{1}$H-signals due to linoleic acid (#13c, #12b, #7c) and unsaturated fatty acids (#24), which presented higher intensities in NTSO-VOO mixtures than in most of the other VO-VOO blends (Contiñas et al., 2008; Guillén et al., 2003; Jović et al., 2016; Ranade et al., 2015; Vigli et al., 2003). Concerning the most important $^{1}$H-signals on HOSO models, the signal intensities of linolenic acid (#13d, #12c) and unsaturated fatty acids (#24 at 5.30–5.32 ppm) were lower in the HOSO-VOO mixtures; whereas those of linoleic acid (#13c, #12b, #9c), unsaturated fatty acids (#24 at 5.32–5.34 ppm) and terpenic alcohols or sterols (#2) were higher in HOSO-VOO mixtures. These observations agreed with the fact that HOSO presents higher concentrations of linoleic acid than VOO, HV and EVAO and lower than NTSO; and HOSO contains lower amounts of linolenic acid than NTSO, VOO and EVAO, and similar to HV (Guillén et al., 2003; Jović et al., 2016; Ranade et al., 2015). Moreover, the mixture of HOSO with VOO leads to an increase in the sterol content compared to pure olive oil (Al-Ismail, Alsaed, Ahmad, & Al-Dabbas, 2010). Evaluating the main variables on the EVAO models, it was observed that the $^{1}$H NMR spectra of the mixtures of EVAO in VOO showed higher intensities for the signals of saturated fatty acids (#10a, #7a, #9a), oleic acid (#7b, #12a, #9b), linoleic acid (#12b, #13c, #10c), squalene (#11) and β-sitosterol (#4) than the spectra of the other VO-VOO blends. Meanwhile, the $^{1}$H-signals of unsaturated fatty acids (#24, #9 at 1.32–1.36 ppm) and linolenic acid (#13d, #12c, #9c) presented lower intensities in the EVAO-VOO blends. Indeed, EVAO presents the highest contents of the saturated fatty acids, mainly palmitic acid, of all the VOs blended with VOO in this study; similar intermediate amounts of oleic and linoleic acids as HOSO; and low concentrations of linolenic acid as VOO, HV and HR (Guillén et al., 2003; Jabeur et al., 2014; Jović et al., 2016; Ranade et al., 2015). To distinguish blends with high % HV in VOO, the $^{1}$H-signals of oleic acid (#7b, #9b, #12a), whose intensities were significantly higher in the HV class, were among the most important variables on the HV models. HV presents similar or slightly higher contents of oleic acid than VOO, and considerably higher amounts compared to the other VOs studied (Guillén et al.,

2003). The opposite trend was shown by the $^{1}$H-signals of linoleic (#7c) and linolenic (#12c) acids, which displayed lower intensity values in the HV class than in the non-HV class. Certainly, the concentrations of linoleic acid in HV are lower than in the other VOs and slightly higher than in VOO; and linolenic acid is present in similar amounts in HV and HOSO but lower amounts in HV than in NTSO, VOO and EVAO (Christopoulou, Lazaraki, Komaitis, & Kaselimis, 2004; Jović et al., 2016; Vigli et al., 2003). For the distinction of mixtures of low % HR in VOO from other VO-VOO mixtures, the $^{1}$H-signals of oleic (#12a) and linolenic (#12c, #7d) acids, saturated fatty acids (#7a) and terpenic alcohols or sterols (#2) exhibited lower intensities in the HR class (Guillén et al., 2003; Vigli et al., 2003). The most discriminant variables in the models to detect low % S in VOO were the $^{1}$H-signals of linolenic acid (#15b, #7d, #12c) and unsaturated fatty acids (#24), which presented significantly higher intensities in S-VOO blends than in the other VO-VOO blends. Soybean oil is the oil with the highest contents of linolenic acid among the studied VOs (Contiñas et al., 2008; Christopoulou et al., 2004; Guillén et al., 2003; Jabeur et al., 2014; Vigli et al., 2003). Furthermore, the lower signal intensities of oleic (#7b) and linoleic (#13c) acids in the S class also contributed to the discrimination of both classes, being consistent with the literature reporting that soybean oil presents significantly lower contents of oleic acid than VOO, and similar contents of linoleic acid as other VOs, such as sunflower oil (Guillén et al., 2003; Jović et al., 2016; Vigli et al., 2003).

### 3.4. PLS-DA models to discriminate blends of OO with VO

Satisfactory binary classification models for all the studied VOs (RPOO, CO, HOSO, NTSO, DOSO, RAO and HR) were obtained using the data of the full % range of VO in the OO mixture, i.e. 0–80% VO in OO (PLS-DA models 30–36 in Table S4 (supplementary material). Prediction abilities were 95–100% for both classes in the models developed to discriminate between OO blends with and without RPOO, CO or HOSO; 84–87% for the OO mixtures with NTSO, DOSO or RAO, and 91–97% for the OO blends that did not contain the corresponding specific VO; and 97%

for the HR class and 89% for the non-HR class. These classification results were improved for each VO by further PLS-DA models developed separately for blends with low or high % VO in OO. Hence, the oil sample containing OO is first classified according to its level of VO, i.e. low (0–20% VO in OO) or high (30–80% VO in OO), by PLS-DA model-29 with prediction abilities of 96% and 94% respectively (Table 1). The most influential variables on this model were the $^{1}$H-signals of saturated fatty acids (#7a), β-sitosterol (#4), linoleic acid (#12b, #15a, #13c) and unsaturated fatty acids (#24, #7 at 1.00–1.02 ppm, #9 at 1.32–1.33 ppm), which exhibited lower intensities in the low class; and those of linolenic (#7d, #15b) and oleic (#12a) acids, which displayed higher intensities in the low class. The chemical composition of the blends that constituted each class justified these observations; thus, the low class contained the samples with the highest % of OO, which is the oil that contains the highest concentrations of oleic acid, together with HR; whereas the high class included the samples with high % of VO characterised by high linoleic and β-sitosterol contents (Al-Ismail et al., 2010; Aparicio & Harwood, 2013; Green & Wang, 2020; Guillén et al., 2003; Jović et al., 2016; Parcerisa, Casals, Boatella, Codony, & Rafecas, 2000; Vigli et al., 2003).

An oil sample containing low % VO in OO is then subjected to various classification models (PLS-DA models 37–50) to identify the specific VO contained in the OO blend (Tables 2 and S5 in supplementary material). The recognition and prediction abilities of these models were higher than 95% of hits for detecting RPOO, CO and HOSO in OO; c.a. 90% for NTSO, DOSO and HR in OO; and c.a. 80–85% for RAO in OO. Taking into account that all CO-OO blends, 95% of the RPOO-OO blends, and at least 95% of the OO blends not containing CO or RPOO were identified with the corresponding models for low % VO in OO, further classification models were developed using datasets without the $^{1}$H-NMR spectral data of RPOO-OO and CO-OO mixtures. The PLS-DA models achieved (PLS-DA models 51–55) afforded better classification abilities to detect NTSO and RAO in OO, and similar results to resolve the presence of HOSO, DOSO or HR in OO (Table S6 in supplementary material).

For oil samples with high % VO in OO, the classification models developed for blends with 20–80% VO in OO (PLS-DA models 56–62) presented recognition and prediction abilities of 98–100% for both classes in RPOO, CO, DOSO and HR models; ≥91% for both classes in NTSO and RAO models; and 86% for the HOSO class and 99% for the non-HOSO class (Table 3). Since all blends were correctly classified by the RPOO and CO models, further PLS-DA models to detect 20–80% VO in OO were built using a dataset without the $^{1}$H-NMR spectral data of RPOO-OO and CO-OO blends (PLS-DA models 63–67 in Table S7 in supplementary material). These models provided the same or better classification abilities than the previous ones, except for HR-OO blends. Indeed, the NTSO and HOSO models allowed the correct classification of all samples of both classes; and the RAO model identified all samples containing RAO and 92% of the samples in the non-RAO class. The main $^{1}$H-signals responsible for the identification of OO blends containing RPOO were those of saturated fatty acids (#9a), which presented significantly higher intensities in the RPOO-OO blends; and those of linoleic acid (#9c, #12b), which showed lower intensities in the RPOO class. The $^{1}$H-signals #9a and #9c were completely discriminants between OO blends containing ≥20% RPOO and the other VO-OO blends with high % VO. As a result, the measurement of just one of these two variables would be enough to confirm whether an OO is mixed with RPOO in percentages ≥20%. Palm oil is the oil that contains the highest amounts of saturated fatty acids among the VOs studied (Vigli et al., 2003). Palmitic acid is the major saturated fatty acid in palm oil and is contained in similar amounts as oleic acid. Meanwhile, linoleic acid is a minor compound in palm oil, present in similar concentrations as in OO, and in lower amounts than in the rest of VOs (Montoya, Cochard, Flori, Cros, Lopes, Cuellar, et al., 2014). The CO-OO blends were distinguished from the other VO-OO mixtures due to the $^{1}$H-signals of linoleic (#7c) and linolenic (#15b, #7d) acids, saturated fatty acids (#7a) and β-sitosterol (#4), which presented higher intensities in the blends containing CO; and to the signal of oleic acid (#9b) with lower intensities in the CO class. Actually, corn oil presents linoleic acid in amounts similar to sunflower oil and significantly higher than refined avocado, refined hazelnut, palm and olive oils; linolenic acid and

β-sitosterol in slightly higher concentrations than the other oils studied; saturated fatty acids in lower contents than palm oil but similar or slightly higher than the rest of the oils considered in the model; and the lowest content of oleic acid, together with sunflower oil. (Aparicio et al., 2013; Guillén et al., 2003; Monfreda et al., 2012; Vigli et al., 2003). The major contributors to the discrimination of HOSO from other VOs in OO were the $^{1}$H-signals of oleic (#9b, #12a) and linoleic (#12b, #9c) acids and saturated (#9a) and unsaturated (#24, #9 at 1.30–1.34 ppm) fatty acids, which exhibited higher intensities in the OO blends with HOSO. Indeed, HOSO contains higher amounts of oleic acid than sunflower, corn and palm oils; similar to avocado oil; and lower than hazelnut and olive oils. Linoleic acid is present in larger concentrations in HOSO than in palm, olive, hazelnut and avocado oils, and smaller than in sunflower and corn oils. The content of saturated fatty acids (#9a) in HOSO is intermediate-high with respect to other VOs but far from those of RPOO, which exhibit the largest contents (Green et al., 2020; Guillén et al., 2003; Jović et al., 2016; Vigli et al., 2003). As in NTSO-VOO models, the most influential variables on the classification models achieved for the detection of NTSO in OO were the $^{1}$H-signals of linoleic acid (#7c, #15a, #12b) and unsaturated fatty acids (#24, #7 at 1.00–1.02 ppm, #9 at 1.32–1.36 ppm), displaying higher intensities in the OO blends with NTSO; and oleic acid (#12a, #7b, #9b), showing the opposite trend. For OO blends with 20–80% NTSO, once the presence of RPOO and CO in the OO blend was discarded by the PLS-DA models 56 and 57 respectively (Table 3), not only the signal of oleic acid (#9b) but also several other signals (#15a, #12b, #9 at 1.34–1.36 ppm, #24) were completely discriminant between both classes; therefore any of them can be used as markers to determine whether an OO blend contains NTSO at concentrations ≥20%. Sunflower oil is characterised by the largest contents of linoleic and unsaturated fatty acids, and the lowest contents of oleic acid with regard to the other VOs studied (Guillén et al., 2003; Jabeur et al., 2014; Jović et al., 2016; Monfreda et al., 2012; Yang et al., 2013). The DOSO models disclosed that the intensities of the $^{1}$H-signals due to oleic acid (#12a, #9b) were significantly higher in DOSO-OO blends, in contrast with linoleic acid (#12b, #7c, #24) signals exhibiting higher intensities in the non-DOSO

class. During the desterolization process, it takes place the dehydration of sterols and the elimination of the acid group of sterol esters by bleaching, producing olefinic degradation products and di-steryl ethers; meanwhile the profiles of triacylglycerides and fatty acids are practically unaltered (Grob, Biedermann, Bronz, & Giuffré, 1994). Therefore, it would be expected that DOSO presents relatively high contents of oleic and linoleic acids as HOSO. However, the deodorization process may affect the composition of triglycerides, diglycerides, fatty acids and minor components of the unsaponifiable fraction, depending mainly on the temperature and time of the process (Aparicio et al., 2013), which could be responsible for the lower content of linoleic acid observed in DOSO blends in relation to the other VOs, including HOSO. The main $^{1}$H-signals on the RAO models were linoleic (#7c, #12b, #13c, #10c) and oleic (#9b) acids and β-sitosterol (#4), exhibiting similar or higher intensities in RAO-OO blends; linolenic acid (#13d, #9c) and unsaturated fatty acids (#9 at 1.32−1.34 ppm, #24), displaying similar or lower intensities in the RAO class; and saturated fatty acids (#9 at 1.20−1.22 ppm) with intermediate intensities. In fact, refined avocado oil, compared to the other VOs studied, presents intermediate compositions of fatty acids (Guillén et al., 2003; Jabeur et al., 2014; Jović et al., 2016; Vigli et al., 2003; Yang et al., 2013) and sterol contents, in particular, β-sitosterol (Al-Ismail et al., 2010; Green et al., 2020; Parcerisa et al., 2000). The most contributing variables to the identification of HR in OO were the $^{1}$H-signals of oleic (#7b, #12a, #9b) and linoleic (#12b) acids, presenting higher intensities in the HR class; and the signals of linolenic acid (#7d, #15b, #12c, #13d), unsaturated (#24) and saturated (#10a, #7a) fatty acids and terpenic alcohols or sterols (#2), showing lower intensities in the HR-OO mixtures. The trend of oleic and linoleic signals observed in HR-OO is opposite to that in HR-VOO. Refined hazelnut oil contains the highest amounts of oleic acid among the VOs studied, comparable to those in OO but lower than VOO; the lowest linolenic contents, similar to those found in HOSO (Green et al., 2020; Guillén et al., 2003; Jović et al., 2016; Parcerisa et al., 2000; Vigli et al., 2003); and characteristic profiles of sterols and terpenic alcohols (Al-Ismail et al., 2010; Aparicio et al., 2013; Parcerisa et al., 2000).

### 3.5. PLS-R models to determine the percentage of VO in a blend with VOO or OO

PLS regression models to determine the % VO contained in a binary mixture with VOO or OO (PLS-R models 1–27) were successfully built for all VOs studied (Table 4). The PLS-R models developed for different sub-ranges of % VO in VOO or OO provided more accurate predictions than those constructed for the full % VO range. The most influential variables on the regression models coincided with those on the classification ones. Therefore, the regression results were explained by the characteristic composition in fatty acids, triacylglycerides and squalene of the oils present in the blend. In VO-VOO models, diacylglycerides, terpenic alcohols and sterols were also decisive.

All regression models presented excellent precisions; yielding $R^2$ values 0.93–0.990, except for the low % range models of VOO mixtures with NTSO, HOSO, HR and S. The PLS-R models for low % NTSO, HOSO and S in VOO presented $R^2$ values <0.70, indicating that the equation can only be used for screening purposes, which enables to distinguish between low, medium and high values of % VO. The PLS-R model for low % HR in VOO showed $R^2$ values <0.50, so the equation only discriminates between high and low values (Priego Capote, Ruiz Jiménez, & Luque De Castro, 2007), in the same way as PLS-DA model-73 distinguishes 2–5% HR and 10% HR in VOO (Table 5).

The regression models achieved allow to determine the % VO in a VOO blend with uncertainties under 5% R-RMSEP for contents of ≥10% NTSO, ≥34% EVAO, ≥39% HOSO and ≥45% HV; 5–10% R-RMSEP for contents of 13–45% HV; 5–15% R-RMSEP for contents of 8–10% NTSO, 7–34% EVAO, 20–39% HOSO and 10–26% HV; 15–20% R-RMSEP for contents of 6–8% NTSO, 5–7% EVAO, 17–20% HOSO and 5% S; and with uncertainty of 28% R-RMSEP for contents of 10% HR. Considering VO-OO blends, the % VO in OO was quantified with uncertainties under 5% R-RMSEP for contents of ≥5% RPOO, ≥6% CO, ≥10% HR, ≥16% DOSO, ≥16% HOSO, ≥9% NTSO and ≥31% RAO; 5–15% R-RMSEP for contents of 2–5% RPOO, 2–6% CO, 3–10% HR,

5–16% DOSO, 7–16% HOSO, 3–9% NTSO and 5–31% RAO; and 15–20% R-RMSEP for contents of 2–3% HR, 4–5% DOSO, 5–7% HOSO, 2–3% NTSO and 4–5% RAO.

The classification abilities of the PLS-DA models to identify blends with low % HV, HR, HOSO and NTSO in VOO and low % RAO in OO were considerably improved when the samples of 2% VO in VOO and/or pure olive oil (VOO or OO) were removed from the dataset used to develop the models (Table 2), indicating that these samples were close to the boundary and therefore could be misclassified. Regarding this fact and the precisions and accuracies of the regression models built, the experimental detection limits were established in the ranges between 2–5% VO for blends of HV, HR, HOSO or NTSO in VOO; between 2–4% VO for blends of RAO in OO; and under 2% VO for blends of EVAO or S in VOO and RPOO, CO, HOSO, NTSO, DOSO or HR in OO. The present results are similar or outperform those reported in the literature using NMR (Parker et al., 2014; Wang et al., 2020) or other analytical techniques (De La Mata-Espinosa et al., 2011; Grob et al., 1994; Jabeur et al., 2014; Jović et al., 2016; Monfreda et al., 2012). In previous high-field NMR studies, the adulteration of refined hazelnut oil in olive oil was detected at a proportion of 10% using $^{1}$H-NMR and linear discriminant analysis (Mannina et al., 2009), 8% using $^{1}$H and $^{13}$C-NMR and artificial neural networks (García-González et al., 2004), 1% using $^{1}$H and $^{31}$P-NMR and canonical discriminant analysis or classification trees (Agiomyrgianaki et al., 2010), and 5% of hazelnut oil in VOO using $^{13}$C-NMR and discriminant data analysis (Zamora et al., 2001). $^{1}$H and $^{31}$P-NMR together with discriminant analysis allowed the detection of adulterations as low as 5% of hazelnut, corn, sunflower and soybean oils in VOO (Vigli et al., 2003). $^{13}$C-NMR and discriminant data analysis distinguished palm oil at 5% in OO (Guyader, Thomas, Portaluri, Jamin, Akoka, Silvestre, et al., 2018). The determination of the contents of oleic, linoleic, linolenic and saturated fatty acids and squalene by $^{1}$H-NMR enabled the detection of 4.5% soybean oil in VOO (Jiang et al., 2018). Nevertheless, chromatographic techniques afforded the lowest limits of detection for sunflower, soybean, corn and palm oils in VOO, detecting even 0.1% adulteration (Jabeur, Zribi, & Bouaziz, 2016).

### 3.6. PLS-DA models to discriminate between ‘legal’ and ‘illegal’ blends of VOO or OO with VO

The potential of the present multivariate approach to implement Reg. (EU) 29/2012 and its amendments is demonstrated with a case study. The most common vegetable oil used to be blended with olive oil is sunflower oil. Therefore NTSO and HOSO were considered as model VOs in ‘legal’ blends with VOO or OO, as done in previous studies (Gómez-Coca et al., 2020; Monfreda et al., 2012). The olive oil blends with the other VOs studied were regarded as ‘illegal’ blends. Binary classification models were developed to first distinguish between ‘legal’ and ‘illegal’ blends, and then differentiate which of the two types of sunflower oils, i.e. NTSO or HOSO, is in the ‘legal’ blend with VOO or OO (Figure S1 in supplementary material). The percentage of NTSO or HOSO in the mixture is determined by the regression models that are reported in the previous section (Table 4).

The PLS-DA model discriminating between ‘legal’ and ‘illegal’ blends provided prediction abilities of 77% for both classes concerning blends with VOO (PLS-DA model-68), and 86% and 98% respectively for blends with OO (PLS-DA model-70 in Table 5). The most discriminant variables on these models are shown in Table S8 (supplementary material). The trends observed for the $^{1}$H-signals involved were consistent with the known differences in the chemical composition of NTSO and HOSO with respect to the VOs in the ‘illegal’ class and both categories of olive oils, already mentioned above.

In addition, classification models were constructed to distinguish ‘legal’ blends containing NTSO from those with HOSO, affording prediction abilities of 83–85% for blends with VOO (PLS-DA model-69), and 97% for blends with OO (PLS-DA model-71 in Table 5). HOSO contains higher amounts of oleic acid and lower concentrations of linoleic and linolenic acids (polyunsaturated fatty acids) than NTSO (Jović et al., 2016), which is reflected on the most influential $^{1}$H-signals on these models (Table S8 in supplementary material).

### 3.7. PLS-DA models to discriminate between blends of VOO or OO with different compositions

Further binary classification models can be built using datasets containing only the information related to specific VOs or % VO in the blends. These complementary models are useful whenever an oil sample is predicted to contain a certain VO by more than one of the classification models described above. Likewise, in the case that the determination of the % VO is not enough accurate by the corresponding regression model for low percentages, it is interesting to be able to discriminate between mixtures with different % VO. As a proof of concept, binary classification models were developed to distinguish blends of different % S or HR in VOO (PLS-DA models 72 and 73); and OO mixtures containing DOSO or HR (PLS-DA model-74), RAO or HR (PLS-DA model-75), RAO or DOSO (PLS-DA model-76) and DOSO or HOSO (PLS-DA model-77), with satisfactory classification abilities (Table 5). The most influential $^{1}$H-signals on these models are gathered in Table S8 (supplementary material). Depending on the class and model considered, different trends were observed in the signal intensities, which are in accordance with the relative chemical composition of each kind of oil in the blend previously reported. The major fatty acids in S and VOO are linoleic acid and oleic acid respectively (Vigli et al., 2003). VOO contains higher amounts of squalene and linolenic acid than HR, and the opposite occurs for linoleic acid (Guillén et al., 2003; Vigli et al., 2003). HR presents higher contents of oleic acid, similar concentrations of linoleic acid and lower amounts of saturated fatty acids than RAO (Green et al., 2020; Parcerisa et al., 2000). In respect of the main variables on the models obtained for the discrimination of DOSO-OO blends from other VO-OO mixtures, DOSO-OO blends contained higher concentrations of oleic acid than OO blends of HR, RAO and HOSO, which are the VOs that present the highest contents of oleic acid (Green et al., 2020; Guillén et al., 2003; Jović et al., 2016; Parcerisa et al., 2000); and lower amounts of linoleic acid than OO blends of HR, RAO and HOSO. Taking into account that DOSO is obtained from the desterolization and deodorization of HOSO, these results evidenced that during the deodorization and/or desterolization processes the fatty acid profile of the

oil was altered, resulting in lower linoleic and higher oleic contents. In this sense, it has been already reported that the drastic conditions used during raffination processes lead to olefinic degradation of sterols, the isomerization of squalene and linoleic and linolenic acids, among other changes in the chemical composition of the oil (Aparicio et al., 2013; Grob et al., 1994).

### 3.8. Prediction of blends of olive oil with other vegetable oils

The composition of thirty-six blind oil samples provided within the OLEUM Project and eight commercial oils was predicted by the classification and regression models developed for blends of olive oil with other vegetable oils following the decision trees shown in Figures 1 and S1 (supplementary material). For each blind sample, Table S9 (supplementary material) gathers *i*) the PLS-DA and PLS-R models applied; *ii*) the PLS-DA predictions related to the category of the olive oil (VOO or OO), the VO contained, and the low/high level of VO in the blend (Tables 1–3, S2–S7 in supplementary material); *iii*) the % VO in the blend determined by the corresponding PLS-R model (Table 4); and *iv*) the predictions of the complementary PLS-DA models (Table 5). Most of the blind samples were predicted satisfactorily according to the description provided (Table S9 in supplementary material); thus, the category of olive oil, i.e. VOO or OO, the particular VO and the % VO in the oil sample were accurately determined. All mixtures of VOO or OO with 40–60% NTSO or HOSO (**1**–**12**), all the blends (containing 5–30% VO) of RPOO-OO (**29**–**32**) and HV-VOO (**17**–**20**), and the blends of EVAO-VOO (**14**–**16**) and HR-OO (**26**–**28**) with ≥10% VO were correctly identified and the % VO properly figured out. Only blind samples **16**, **17** and **19** were predicted to present slightly higher % VO in VOO, and sample **26** scarcely lower % HR in OO, than those percentages given in the description. The DOSO-OO blends (**33**–**36**) were satisfactorily determined by the corresponding classification and regression models; the % DOSO in OO in sample **36** was barely lower than predicted. The blend of 10% DOSO in OO (**34**) was confused with mixtures of 2–11% of HOSO in OO. For the blend of 5% EVAO in VOO (**13**), the contained VO was not recognised by any of the classification models, but the calculated % VO was within the

calibration range of the regression model developed for EVAO-VOO blends; and this model predicted correctly the % EVAO in the mixture, even with better precisions than the other models built for HOSO-VOO and HR-VOO blends. The VO in the blend of 5% HR in OO (**25**) was not identified by any of the HR-OO classification models. Indeed, the detection of the adulteration of OO with HR is still one of the main challenges in fraud detection due to the close composition of both refined oils (Agiomyrgianaki et al., 2010; García-González et al., 2004; Mannina et al., 2009). Even blends with ≤10% HR in OO can be confused with RAO-OO blends. The composition of blind samples **21**–**24** were determined by the classification and regression models built for both RAO-OO and DOSO-OO blends; however, the PLS-DA model-76 (Table 5), which distinguishes these two OO mixtures, predicted satisfactorily that these blind samples contained RAO, except for the mixture of 10% RAO in OO (**22**).

Regarding the commercial oils analysed, samples **37**, **38** and **44** were declared to be mixtures of vegetable oils or NTSO with EVOO or VOO. Samples **37** and **38** were confirmed to contain VOO, whereas sample **44** was classified as an OO blend. Furthermore, the three samples were predicted to contain NTSO, in accordance with their label specifications. All the other commercial oil samples (**39**–**43**) were labelled as mixtures of VOO or EVOO with rapeseed oil; however, all of them were classified as blends of OO. These results are not conclusive since no blends of rapeseed oil with VOO or OO were available to be included in the modelling step of the present study.

## 4. Conclusion

A stepwise strategy based on $^{1}$H-NMR fingerprinting of an oil sample in combination with chemometrics is proposed to determine the content of mixtures of oils of the 'virgin olive oil' or 'olive oil' categories and vegetable oils, providing a chemical tool to (*i*) confirm the presence of VOO or OO in an oil sample; (*ii*) discriminate between pure olive oils and their blends with VOs to a certain extent, given by the detection limit disclosed for each VO; (*iii*) identify the VO in the blend with VOO or OO; (*iv*) differentiate between blends made with different VOs in VOO or OO;

555 (*v*) distinguish blends made with the same VO in different proportions; and (*vi*) determine the %
556 VO blended with VOO or OO.

557 $^{1}$H-NMR spectral data of olive oils and their mixtures with the VOs most commonly used to make
558 blends, i.e. sunflower oil, high oleic sunflower oil, desterolized high oleic sunflower oil, virgin and
559 refined avocado oil, virgin and refined hazelnut oil, refined palm olein oil, corn oil and soybean oil,
560 was used to optimize and validate classification and regression models built by PLS-DA and PLS-R
561 respectively. The classification models achieved were satisfactory, robust and stable. Excellent
562 precisions and acceptable accuracies were afforded by the regression models developed for the
563 determination of the % VO in VOO or OO. The reliability of the classification and regression
564 models was supported by the chemical interpretation of the most influential variables on the
565 validated models. The % VO in the blend is determined with uncertainties under the 20% of R-
566 RMSEP for contents as low as 5% EVAO or S, 6% NTSO, 10% HV and 17% HOSO in VOO; and
567 2% RPOO, CO, NTSO or HR, 4% DOSO or RAO and 5% HOSO in OO. The detection limits are
568 under 2% EVAO or S and between 2–5% NTSO, HOSO, HV or HR in VOO; and under 2% RPOO,
569 CO, HOSO, NTSO, DOSO or HR and 2–4% RAO in OO. The performance and effectiveness of the
570 proposed strategy were validated by a set of blind samples, which confirmed its feasibility to
571 support Reg. (EU) 29/2012. Further studies should be carried out with larger balanced sample sets
572 covering the variability of olive oils of both categories (VOO and OO) and the vegetable oils of
573 interest. The different possible sources of variability, such as the varieties of each botanical oil
574 species, the agronomical and climatic conditions, the geographical origins and harvests, should be
575 considered. The implementation of this approach requires a databank of $^{1}$H-NMR fingerprints of
576 oils. The databank has to include pure oils comprising olive oils of the different categories,
577 vegetable oils used to make legal blends and adulterant oils, and their mixtures; because it has to be
578 representative of oil variability in order to guarantee robust models for both authentication and
579 fraud detection. It is worth noting that this requirement is feasible in practice since the creation of
580 the OLEUM Databank and the OLEUM Network are among the objectives of the OLEUM Project

that are being accomplished. The OLEUM Databank is an online integrated quality assurance database of olive oil analytical methods and chemical data, which is currently being developed. The OLEUM Network is a worldwide community of proficient analytical laboratories involved in olive oil analysis, and it is expected to expand and may also contribute to the feeding and updating of the databank over time.

## Acknowledgments

This work was developed in the framework of the project OLEUM "Advanced solutions for assuring authenticity and quality of olive oil at global scale" funded by the European Commission within the Horizon 2020 Programme (2014–2020), grant agreement No. 635690; and the project AUTENFOOD funded by ACCIÓ-Generalitat de Catalunya and the European Union through the Programa Operatiu FEDER Catalunya 2014-2020 (Ref COMRDI-15-1-0035). The information contained in this article reflects the authors' views; the European Commission is not liable for any use of the information contained herein. The authors would like to thank all producers that supplied the olive oils, virgin olive oils and vegetable oils for this study, and the technical and staff support provided by SGIker (UPV/EHU, MICINN, GV/EJ, ESF).

## References

Agiomyrgianaki, A., Petrakis, P. V., & Dais, P. (2010). Detection of refined olive oil adulteration with refined hazelnut oil by employing NMR spectroscopy and multivariate statistical analysis. *Talanta, 80*(5), 2165-2171. https://doi.org/10.1016/j.talanta.2009.11.024.

Al-Ismail, K. M., Alsaed, A. K., Ahmad, R., & Al-Dabbas, M. (2010). Detection of olive oil adulteration with some plant oils by GLC analysis of sterols using polar column. *Food Chemistry, 121*(4), 1255-1259. https://doi.org/10.1016/j.foodchem.2010.01.016.

Alonso-Salces, R. M., Segebarth, N., Garmón-Lobato, S., Holland, M. V., Moreno-Rojas, J. M., Fernández-Pierna, J. A., et al. (2015). $^{1}$H-NMR and isotopic fingerprinting of olive oil and its unsaponifiable fraction: Geographical origin of virgin olive oils by pattern recognition.

606 *European Journal of Lipid Science and Technology, 117*(12), 1991-2006.
607 https://doi.org/10.1002/ejlt.201400243.
608 Aparicio, R., & Harwood, J. (2013). *Handbook of olive oil: Analysis and properties* (2nd ed.). USA:
609 Springer US.
610 Berrueta, L. A., Alonso-Salces, R. M., & Héberger, K. (2007). Supervised pattern recognition in
611 food analysis. *Journal of Chromatography A, 1158*(1-2), 196-214.
612 https://doi.org/10.1016/j.chroma.2007.05.024.
613 Conte, L., Bendini, A., Valli, E., Lucci, P., Moret, S., Maquet, A., et al. (2020). Olive oil quality
614 and authenticity: A review of current EU legislation, standards, relevant methods of
615 analyses, their drawbacks and recommendations for the future. *Trends in Food Science &*
616 *Technology, 105*, 483-493. https://doi.org/10.1016/j.tifs.2019.02.025.
617 Contiñas, A., Martínez, S., Carballo, J., & Franco, I. (2008). Detection of contaminations and/or
618 adulterations of the extra virgin olive oil with seeds oils (sunflower and soybean) and olive
619 pomace oil. *Grasas y Aceites, 59*(2), 97-103. https://doi.org/10.3989/gya.2008.v59.i2.496.
620 Christopoulou, E., Lazaraki, M., Komaitis, M., & Kaselimis, K. (2004). Effectiveness of
621 determinations of fatty acids and triglycerides for the detection of adulteration of olive oils
622 with vegetable oils. *Food Chemistry, 84*(3), 463-474. https://doi.org/10.1016/s0308-
623 8146(03)00273-5.
624 De la Mata, P., Dominguez-Vidal, A., Bosque-Sendra, J. M., Ruiz-Medina, A., Cuadros-Rodríguez,
625 L., & Ayora-Cañada, M. J. (2012). Olive oil assessment in edible oil blends by means of
626 ATR-FTIR and chemometrics. *Food Control, 23*(2), 449-455.
627 EC. (2012). European Commission Regulation 29/2012 on marketing standards for olive oil.
628 *Official Journal of the European Union, January 13, 2012*, 12-21.
629 EC. (2020). European Commission website. Food Fraud.
630 https://ec.europa.eu/knowledge4policy/food-fraud-quality/food-fraud-data-bases_en,
631 Accessed date: December 1, 2020.

Esbensen, K. H., Guyot, D., Westad, F., & Houmøller, L. P. (2002). *Multivariate data analysis - in practice: An introduction to multivariate data analysis and experimental design* (5th ed.). Oslo: Camo Process AS.

European Parliament. (2014). Committee on the Environment, Public Health and Food Safety. Resolution of 14 January 2014 on the food crisis, fraud in the food chain and the control thereof (2013/2091 (INI)).

Fregapane, G., Gómez-Rico, A., Inarejos, A. M., & Salvador, M. D. (2013). Relevance of minor components stability in commercial olive oil quality during the market period. *European Journal of Lipid Science and Technology, 115*(5), 541-548. https://doi.org/10.1002/ejlt.201200209.

García-González, D. L., Mannina, L., D'Imperio, M., Segre, A. L., & Aparicio, R. (2004). Using $^{1}H$ and $^{13}C$ NMR techniques and artificial neural networks to detect the adulteration of olive oil with hazelnut oil. *European Food Research and Technology, 219*(5), 545-548. https://doi.org/10.1007/s00217-004-0996-0.

Gómez-Coca, R. B., Pérez-Camino, M. d. C., Martínez-Rivas, J. M., Bendini, A., Gallina Toschi, T., & Moreda, W. (2020). Olive oil mixtures. Part one: Decisional trees or how to verify the olive oil percentage in declared blends. *Food Chemistry, 315*, Article 126235. https://doi.org/https://doi.org/10.1016/j.foodchem.2020.126235.

Green, H. S., & Wang, S. C. (2020). First report on quality and purity evaluations of avocado oil sold in the US. *Food Control, 116*, Article 107328. https://doi.org/https://doi.org/10.1016/j.foodcont.2020.107328.

Grob, K., Biedermann, M., Bronz, M., & Giuffré, A. M. (1994). The Detection of Adulteration with Desterolized Oils. *Lipid / Fett, 96*(9), 341-345. https://doi.org/10.1002/lipi.19940960905.

Guillén, M. D., & Ruiz, A. (2003). Rapid simultaneous determination by proton NMR of unsaturation and composition of acyl groups in vegetable oils. *European Journal of Lipid Science and Technology, 105*(11), 688-696. https://doi.org/10.1002/ejlt.200300866.

Guyader, S., Thomas, F., Portaluri, V., Jamin, E., Akoka, S., Silvestre, V., et al. (2018). Authentication of edible fats and oils by non-targeted $^{13}C$ INEPT NMR spectroscopy. *Food Control, 91*, 216-224. https://doi.org/10.1016/j.foodcont.2018.03.046.

Jabeur, H., Zribi, A., & Bouaziz, M. (2016). Extra-virgin olive oil and cheap vegetable ols: Distinction and dtection of aulteration as dtermined by GC and chemometrics. *Food Analytical Methods, 9*(3), 712-723. https://doi.org/10.1007/s12161-015-0249-9.

Jabeur, H., Zribi, A., Makni, J., Rebai, A., Abdelhedi, R., & Bouaziz, M. (2014). Detection of chemlali extra-virgin olive oil adulteration mixed with soybean oil, corn oil, and sunflower oil by using GC and HPLC. *Journal of Agricultural and Food Chemistry, 62*(21), 4893-4904. https://doi.org/10.1021/jf500571n.

Jiang, X. Y., Li, C., Chen, Q. Q., & Weng, X. C. (2018). Comparison of $^{19}F$ and $^{1}H$ NMR spectroscopy with conventional methods for the detection of extra virgin olive oil adulteration. *Grasas y Aceites, 69*(2), Article e249. https://doi.org/10.3989/gya.1221172.

Jović, O., Smolić, T., Primožič, I., & Hrenar, T. (2016). Spectroscopic and chemometric analysis of binary and ternary edible oil mixtures: Qualitative and quantitative study. *Analytical chemistry, 88*(8), 4516-4524. https://doi.org/10.1021/acs.analchem.6b00505.

Kjeldahl, K., & Bro, R. (2010). Some common misunderstandings in chemometrics. *Journal of Chemometrics, 24*(7-8), 558-564. https://doi.org/10.1002/cem.1346.

Lioupi, A., Nenadis, N., & Theodoridis, G. (2020). Virgin olive oil metabolomics: A review. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences, 1150*, Article 122161. https://doi.org/10.1016/j.jchromb.2020.122161.

Mannina, L., D'Imperio, M., Capitani, D., Rezzi, S., Guillou, C., Mavromoustakos, T., et al. (2009). $^{1}H$ NMR-based protocol for the detection of adulterations of refined olive oil with refined hazelnut oil. *Journal of Agricultural and Food Chemistry, 57*(24), 11550-11556. https://doi.org/10.1021/jf902426b.

683 Monfreda, M., Gobbi, L., & Grippa, A. (2012). Blends of olive oil and sunflower oil:
684 Characterisation and olive oil quantification using fatty acid composition and chemometric
685 tools. *Food Chemistry, 134*(4), 2283-2290. https://doi.org/10.1016/j.foodchem.2012.03.122.
686 Montoya, C., Cochard, B., Flori, A., Cros, D., Lopes, R., Cuellar, T., et al. (2014). Genetic
687 architecture of palm oil fatty acid composition in cultivated oil palm (*Elaeis guineensis*
688 Jacq.) compared to its wild relative *E. oleifera* (H.B.K) Cortés. *PLoS ONE, 9*(5), Article
689 e95412. https://doi.org/10.1371/journal.pone.0095412.
690 OLEUM Project. (2016). About OLEUM: Aims and Objectives. http://www.oleumproject.eu/about-
691 oleum/aims-and-objectives, Accessed date: December 1, 2020.
692 Parcerisa, J., Casals, I., Boatella, J., Codony, R., & Rafecas, M. (2000). Analysis of olive and
693 hazelnut oil mixtures by high-performance liquid chromatography-atmospheric pressure
694 chemical ionisation mass spectrometry of triacylglycerols and gas-liquid chromatography of
695 non-saponifiable compounds (tocopherols and sterols). *Journal of Chromatography A,*
696 *881*(1-2), 149-158. https://doi.org/10.1016/S0021-9673(00)00352-6.
697 Parker, T., Limer, E., Watson, A. D., Defernez, M., Williamson, D., & Kemsley, E. K. (2014).
698 60MHz $^{1}$H NMR spectroscopy for the analysis of edible oils. *TrAC - Trends in Analytical*
699 *Chemistry, 57*, 147-158. https://doi.org/10.1016/j.trac.2014.02.006.
700 Popescu, R., Costinel, D., Dinca, O. R., Marinescu, A., Stefanescu, I., & Ionete, R. E. (2015).
701 Discrimination of vegetable oils using NMR spectroscopy and chemometrics. *Food Control,*
702 *48*, 84-90. https://doi.org/10.1016/j.foodcont.2014.04.046.
703 Priego Capote, F., Ruiz Jiménez, J., & Luque De Castro, M. D. (2007). Sequential (step-by-step)
704 detection, identification and quantitation of extra virgin olive oil adulteration by
705 chemometric treatment of chromatographic profiles. *Analytical and Bioanalytical*
706 *Chemistry, 388*(8), 1859-1865. https://doi.org/10.1007/s00216-007-1422-9.
707 Ranade, S., & Thiagarajan, P. (2015). A review on *Persea Americana* Mill. (Avocado) - Its fruit
708 and oil. *International Journal of PharmTech Research, 8*(6), 72-77.

Santos, P. M., Kock, F. V. C., Santos, M. S., Lobo, C. M. S., Carvalho, A. S., & Colnago, L. A. (2017). Non-invasive detection of adulterated olive oil in full bottles using time-domain NMR relaxometry. *Journal of the Brazilian Chemical Society, 28*(2), 385-390. https://doi.org/10.5935/0103-5053.20160188.

Vigli, G., Philippidis, A., Spyros, A., & Dais, P. (2003). Classification of edible oils by employing $^{31}P$ and $^{1}H$ NMR spectroscopy in combination with multivariate statistical analysis. A proposal for the detection of seed oil adulteration in virgin olive oils. *Journal of Agricultural and Food Chemistry, 51*(19), 5715-5722. https://doi.org/10.1021/jf030100z.

Wang, X., Wang, G., Hou, X., & Nie, S. (2020). A rapid screening approach for authentication of olive oil and classification of binary blends of olive oils using low-field nuclear magnetic resonance spectra and support vector machine. *Food Analytical Methods, 13*, 1894–1905. https://doi.org/10.1007/s12161-020-01799-z.

Yang, Y., Ferro, M. D., Cavaco, I., & Liang, Y. (2013). Detection and identification of extra virgin olive oil adulteration by GC-MS combined with chemometrics. *Journal of Agricultural and Food Chemistry, 61*(15), 3693-3702. https://doi.org/10.1021/jf4000538.

Zamora, R., Alba, V., & Hidalgo, F. J. (2001). Use of high-resolution $^{13}C$ nuclear magnetic resonance spectroscopy for the screening of virgin olive oils. *Journal of the American Oil Chemists' Society, 78*(1), 89-94. https://doi.org/10.1007/s11746-001-0225-z.

## Figure captions

**Figure 1.** Decision tree constituted of PLS-DA classification and PLS-R regression models to determine the composition of binary mixtures of oils of the 'virgin olive oil' or 'olive oil' categories and other vegetable oils. Abbreviations: VOO, virgin olive oil; OO, olive oil; VO, vegetable oil; NTSO, refined conventional sunflower oil (normal type sunflower oil); HOSO, refined high oleic sunflower oil; DOSO, desterolized and deodorized high oleic sunflower oil; HR, refined hazelnut oil; HV, virgin hazelnut oil; S, refined soybean oil; EVAO, virgin avocado oil; RAO, refined avocado oil; RPOO, refined palm olein oil; CO, refined corn oil.

## Supplementary material

**Figure S1.** Decision tree constituted of PLS-DA classification and PLS-R regression models for a case-study: Discrimination between 'legal' (containing NTSO or HOSO) and 'illegal' (not containing NTSO or HOSO) blends, and determination of % NTSO or HOSO in binary mixtures with oils of the 'virgin olive oil' or 'olive oil' categories. Abbreviations: VOO, virgin olive oil; OO, olive oil; VO, vegetable oil; NTSO, refined conventional sunflower oil (normal type sunflower oil); HOSO, refined high oleic sunflower oil.

REVISED Manuscript (text with changes Marked) Click here to view linked References

# Stepwise strategy based on $^{1}$H-NMR fingerprinting in combination with chemometrics to determine the content of vegetable oils in olive oil mixtures

Rosa María Alonso-Salces[1,*], Luis Ángel Berrueta[2], Beatriz Quintanilla-Casas[3], Stefania Vichi[3], Alba Tres[3], María Isabel Collado[4], Carlos Asensio-Regalado[2], Gabriela Elena Viacava[5], Aimará Ayelen Poliero[6], Enrico Valli[7], Alessandra Bendini[7], Tullia Gallina Toschi[7], José Manuel Martínez-Rivas[8], Wenceslao Moreda[9], Blanca Gallo[2]

[1] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Departamento de Biología, Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Mar del Plata (UNMdP), Funes 3350, 7600, Mar del Plata, Argentina.

[2] Departamento de Química Analítica, Facultad de Ciencia y Tecnología, Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU), P.O. Box 644, 48080, Bilbao, Spain.

[3] Food Science and Nutrition Department, XaRTA (Catalonian Reference Network on Food Technology), University of Barcelona (UB), Food and Nutrition Torribera Campus, Av. Prat de la Riba, 171. 08028, Sta. Coloma de Gramanet, Spain.

[4] SGIKER, Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU), Barrio Sarriena s/n, 48940, Leioa, Spain.

[5] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Grupo de Investigación en Ingeniería en Alimentos (GIIA), Departamento de Ingeniería Química y en Alimentos, Facultad de Ingeniería, Universidad Nacional de Mar del Plata (UNMdP), Juan B. Justo 4302, 7600, Mar del Plata, Argentina.

[6] Departamento de Biología, Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Mar del Plata (UNMdP), Funes 3350, 7600, Mar del Plata, Argentina.

[7] Department of Agricultural and Food Sciences, Alma Mater Studiorum University of Bologna (UNIBO), Piazza Goidanich, 60, I-47521, Cesena (FC), Bologna, Italy.

[8] Department of Biochemistry and Molecular Biology of Plant Products, Instituto de la Grasa-CSIC, Ctra. Utrera km 1, Building 46, E-41013 Sevilla, Spain.

[9] Department of Characterization and Quality of Lipids, Instituto de la Grasa-CSIC, Ctra. Utrera km 1, Building 46, E-41013 Sevilla, Spain.

[*] Corresponding author.

E-mail address: rosamaria.alonsosalces@gmail.com (R.M. Alonso-Salces).

**Abstract**

$^{1}$H-NMR fingerprinting of edible oils and a set of multivariate classification and regression models organised in a decision tree is proposed as a stepwise strategy to assure the authenticity and traceability of olive oils and their declared blends with other vegetable oils (VOs). Oils of the 'virgin olive oil' and 'olive oil' categories and their mixtures with the most common VOs, i.e. sunflower, high oleic sunflower, hazelnut, avocado, soybean, corn, refined palm olein and desterolized high oleic sunflower oils, were studied. Partial least squares (PLS) discriminant analysis provided stable and robust binary classification models to identify the olive oil type and the VO in the blend. PLS regression afforded models with excellent precisions and acceptable accuracies to determine the percentage of VO in the mixture. The satisfactory performance of this approach, tested with blind samples, confirm its potential to support regulations and control bodies.

Keywords: olive oil, nuclear magnetic resonance, multivariate data analysis, decision tree, adulteration, authentication

## 1. Introduction

The high price of olive oil, the distinctive sensory profile, and its reputation as a healthy source of dietary fats make olive oil a target for fraud. The most common types of olive oil fraud are illegal blending with other vegetable oils (VOs) or low-quality olive oils, deliberate mislabelling of less expensive classes of olive oils, other vegetable oils or their blends with olive oils, and mislabelling of the geographical origin or Protected Designation of Origin declaration. Indeed, the European Parliament pointed out that olive oil adulteration has become one of the biggest financial fraud in the agricultural sector, and evidenced the need to update and harmonize analytical methods for quality and authenticity control of olive oil (EC, 2020; European Parliament, 2014). In this context, the so-called OLEUM Project was supported by the European Commission with the overall objective of improving existing analytical methods and developing new strategies of analysis for assuring the quality and authenticity of olive oil (OLEUM Project, 2016).

The EU Regulation 29/2012 standardises the labelling of all olive oil categories and their mixtures with other VOs, allowing to highlight the presence of olive oil on the label outside the ingredient list, only if it accounts for at least 50% of the blend (EC, 2012). However, this regulation and its amendments do not refer to any analytical parameter or method to control the percentage of olive oil in the admixture or the botanical origin of oil. The need of analytical methods to confirm the presence of olive oil in the blend, to distinguish pure and adulterated olive oils, to identify the adulterant oils in the mixture, as well as to determine the percentage of olive oil and the adulterants in the blend, is evidenced and is an issue of major concern in order to implement the established regulations (Conte, Bendini, Valli, Lucci, Moret, Maquet, et al., 2020). In literature, few works deal with the verification of the percentage of olive oil in fraudulent blends with VOs with regard to the labelling compliance of Reg. (EU) 29/2012 (De la Mata, Dominguez-Vidal, Bosque-Sendra, Ruiz-Medina, Cuadros-Rodríguez, & Ayora-Cañada, 2012; Gómez-Coca, Pérez-Camino, Martínez-Rivas, Bendini, Gallina Toschi, & Moreda, 2020; Monfreda, Gobbi, & Grippa, 2012; Santos, Kock, Santos, Lobo, Carvalho, & Colnago, 2017).

The chemical methods traditionally used in food analysis are laborious, time-consuming, non-eco-friendly and require sample preparation and skilled operators. In contrast, metabolomic approaches based on advanced instrumental techniques, such as MS and NMR, coupled to chemometrics overcome some of these operational drawbacks and provide useful tools for food quality control and traceability (Lioupi, Nenadis, & Theodoridis, 2020). Most of the NMR approaches developed for olive oil authentication, detection of olive oil adulteration and to determine the composition of olive oil blends with VOs, were based on measuring NMR signals that give quantitative information of certain compounds or are used to calculate some parameters and ratios (i.e. profiling) (Agiomyrgianaki, Petrakis, & Dais, 2010; García-González, Mannina, D'Imperio, Segre, & Aparicio, 2004; Jiang, Li, Chen, & Weng, 2018; Mannina, D'Imperio, Capitani, Rezzi, Guillou, Mavromoustakos, et al., 2009; Popescu, Costinel, Dinca, Marinescu, Stefanescu, & Ionete, 2015; Vigli, Philippidis, Spyros, & Dais, 2003; Zamora, Alba, & Hidalgo, 2001). Instead, NMR fingerprinting was only reported in few studies using low-field NMR spectroscopy (Parker, Limer, Watson, Defernez, Williamson, & Kemsley, 2014; Santos et al., 2017; Wang, Wang, Hou, & Nie, 2020). To the authors' knowledge, high-field NMR fingerprinting has been used to study mixtures of olive oil with other VOs for the first time in the present work. This study aimed to develop an analytical strategy based on $^{1}$H-NMR fingerprinting together with multivariate classification and regression models organised in a decision tree to determine the composition of an oil blend from both points of view, the botanical nature of the oils and the percentage of each oil in the blend. The performance of the complete stepwise analytical strategy is evaluated by the prediction results obtained for an external set of blind oil samples and commercial oils. It is worth noting that this analytical approach addresses some issues not considered in previous studies: (*i*) the discrimination between oil samples containing oil of the 'virgin olive oil' category (VOO) and the 'olive oil' category (OO); (*ii*) the distinction of pure and blended oils; and (*iii*) the study of a large sample set with pure oils and blends of the most common VOs used for olive oil adulteration, and a wide range

of % VO in the blend (including the percentages for the labelling verification in compliance with Reg. (EU) 29/2012).

## 2. Material and methods

### 2.1. Samples

Genuine samples of virgin (VOO) and extra virgin olive (EVOO) oils (n=176), olive oils (OO, n=3), refined conventional sunflower oil (normal type sunflower oil, NTSO, n=17), refined high oleic sunflower oil (HOSO, n=16), desterolized and deodorized high oleic sunflower oil (DOSO, n=1), refined hazelnut oil (HR, n=11), virgin hazelnut oil (HV, n=6), refined soybean oil (S, n=10), virgin avocado oil (EVAO, n=1), refined avocado oil (RAO, n=1), refined palm olein oil (RPOO, n=1) and refined corn oil (CO, n=1) were used to prepare binary mixtures at different percentages (2–90%) of VOs in VOOs or OOs (1007 blends). Samples were obtained in the framework of the AUTENFOOD and OLEUM projects. Oils from the sample banks of both projects were produced during two consecutive harvest years (2016/17 and 2017/18). Besides, eight commercial oil samples collected in the Swedish market were analysed. According to their labels, the commercial oils were described as mixtures of VOO and other VO such as rapeseed oil, sunflower oil, or non-identified vegetable oil.

Blends were prepared and preserved under controlled temperature conditions. All pure and blended oil samples were bottled with nitrogen headspace or minimal air headspace, kept at -20 ºC and protected from light. Before analysis, oil samples were taken from the cold storage, left to equilibrate at room temperature at least for 12 h, and shaken vigorously before sampling the oil aliquot for analysis.

### 2.2. Chemicals

Deuterated chloroform for NMR analysis (99.8 atom % D) was provided by Sigma-Aldrich Chemie (Steinheim, Germany).

## 2.3. NMR analysis

Aliquots of 150 µL of each oil sample were dissolved in 750 µL of deuterated chloroform, shaken in a vortex, and placed in a 5 mm NMR capillary. The $^{1}$H-NMR experiments were performed at 300K on a Bruker (Rheinstetten, Germany) Avance 500 (nominal frequency 500.13 MHz) equipped with a 5 mm broadband inverse probe with Z-gradients. The spectra were recorded using a 6.1 µs pulse (90°), an acquisition time of 3.5 s (50k data points) and a total recycling time of 7.0 s, a spectral width of 7100 Hz (14 ppm), 32 scans (+ 4 dummy scans), with no sample rotation. Prior to Fourier transformation, the free induction decays (FIDs) were zero-filled to 64k and a 0.3 Hz line-broadening factor was applied. The chemical shifts were expressed in δ scale (ppm), referenced to the residual signal of chloroform (7.26 ppm). The spectra were phase- and baseline-corrected manually, binned with 0.02 ppm-wide buckets, and normalized to total intensity over the region 4.10–4.26 ppm (glycerol signal). The region of the NMR spectra studied comprised from 0 ppm to 11 ppm. TopSpin 2.1 (2013) and Amix-Viewer 3.7.7 (2006) from Bruker BioSpin GMBH (Rheinstetten, Germany) were used to perform the processing of the spectra. The data table generated with the spectra of all samples, excluding the eight buckets in the reference region 4.10–4.26 ppm, was then submitted to multivariate data analysis.

## 2.4. Data analysis

Datasets were made up of the 542 buckets of the $^{1}$H-NMR spectra (variables in columns) measured on the oil samples (samples in rows). A total number of 1239 pure and blended oil samples were analysed by $^{1}$H-NMR. Depending on the aim of the multivariate model to be developed, the dataset contained the NMR spectral data of the corresponding studied samples. Datasets were analysed by univariate procedures (ANOVA, Fisher index and Box & Whisker plots); and by multivariate techniques, unsupervised such as principal component analysis (PCA), and supervised as partial least squares discriminant analysis (PLS-DA) and partial least squares regression (PLS-R) (Berrueta, Alonso-Salces, & Héberger, 2007). Data analysis was performed by means of the

150 statistical software package Statistica 7.0 (StatSoft Inc., Tulsa, OK, USA, 1984–2004) and The
151 Unscrambler v9.7 (Camo Software AS, 1986–2007).

152 PCA, PLS-DA and PLS-R were applied to the autoscaled or centered data matrix of $^{1}$H-NMR
153 spectra of the oil samples. The presence of outliers in the dataset was analysed by PCA. In PLS-DA
154 and PLS-R, the optimal number of PLS-components is estimated by cross-validation by plotting the
155 root mean square error in the prediction (RMSEP) against the number of PLS-components. The
156 model with the smallest number of features should be accepted from among equivalent models on
157 the training set in order to avoid overfitting (according to the principle of parsimony). In PLS-DA,
158 once the number of PLS-components is optimised, the predictions in the training-test set are
159 represented in a box and whisker plot in order to define the half of the distance between the
160 quartiles as the boundary. The regression coefficients (B) of the optimal number of PLS-
161 components denote the importance of the NMR variables on the model: the larger the B-coefficient,
162 the higher the influence of the variable on the PLS-DA or PLS-R model. A large B-coefficient may
163 also indicate a variable with small absolute values but large relative differences (Esbensen, Guyot,
164 Westad, & Houmøller, 2002). PLS-DA and PLS-R models were validated by 3-fold or leave-one
165 out cross-validation for parameter optimization, and by external validation when an external set of
166 samples was available. Binary classification models can lead to artefacts if they are not used and
167 validated properly (Kjeldahl & Bro, 2010). The reliability of the classification models developed
168 was studied in terms of recognition and prediction abilities in the cross-validation, and prediction
169 ability in the external validation (Berrueta et al., 2007). The goodness of the regression model fit
170 was evaluated by means of the prediction error, the correlation coefficient between predicted and
171 measured values in calibration and validation (R-cal, R-val), the determination coefficient in
172 calibration and validation ($R^2$-cal, $R^2$-val), and the evaluation of the residuals. The RMSEP is the
173 practical average prediction error estimated by the validation set (empirical error estimate expressed
174 in the original measurement units). The result is expressed as the predicted Y-value ± 2 RMSEP.

The R-RMSEP is the relative prediction error in % (comparable to the analytical accuracy) (Esbensen et al., 2002).

## 3. Results and discussion

### 3.1. Mixtures of olive oil with vegetable oils

Oils of the VOO and OO categories and their mixtures with the most common VOs used for the adulteration of olive oil or making 'legal' blends, i.e. NTSO, HOSO, DOSO, HR, HV, S, EVAO, RAO, RPOO and CO, were studied. The $^{1}$H-NMR spectra of the oil samples, both pure and blended (binary mixtures of VO with VOO or OO) oils, were recorded. The chemical shifts of the $^{1}$H-signals and their assignments to protons of the different functional groups are shown in Table S1 (supplementary material). The $^{1}$H-NMR profiles of the oil samples presented characteristic patterns of triglycerides, diglycerides and some minor constituents of the unsaponifiable fraction, which are useful for the determination of the botanical origin of oils and the composition of blended oils (Agiomyrgianaki et al., 2010; Alonso-Salces, Segebarth, Garmón-Lobato, Holland, Moreno-Rojas, Fernández-Pierna, et al., 2015; García-González et al., 2004; Guillén & Ruiz, 2003; Mannina et al., 2009; Parker et al., 2014; Popescu et al., 2015; Vigli et al., 2003; Wang et al., 2020).

The proposed approach to detect blends of olive oils (VOOs or OOs) with other VO and quantify the % VO in the blend is based on the use of the $^{1}$H-NMR fingerprint of the oil and a set of multivariate classification and regression models organized in a decision tree (Figures 1 and S1 in supplementary material). The PLS-DA and PLS-R models achieved and their chemical interpretation are described in the next sections. The most influential variables on the models were not completely discriminant unless otherwise specified.

### 3.2. PLS-DA model to confirm the presence of VOO or OO

The first stage of the decision tree (Figure 1) consists in identifying whether the oil sample contains VOO or OO using PLS-DA model-1 with recognition and prediction abilities of 97% and 98% for

199 the VOO and OO classes respectively (Table 1). The most influential NMR variables on the model
200 were the $^{1}$H-signals of oleic acid (#7b, #9b), linolenic acid (#10c, #13d) and saturated fatty acids
201 (#9a), exhibiting higher intensities in VOO and their blends than in samples containing OO. In
202 contrast, the $^{1}$H-signals of linoleic acid (#12b) and *sn*-1,3-diacylglycerides (#17) presented lower
203 intensities in the VOO class. These observations are consistent with previous studies reporting the
204 differences in the composition of oleic, linolenic and saturated fatty acids and *sn*-1,3-
205 diacylglycerides between VOOs and OOs (Guillén et al., 2003; Jiang et al., 2018).

206 Once the oil sample is classified as containing VOO or OO, further predictions are made using the
207 binary classification models built separately for each type of olive oil to elucidate whether the olive
208 oil sample is mixed with a VO, in which proportion (low or high) and with which particular VO
209 (Figure 1).

210 **3.3. PLS-DA models to discriminate blends of VOO with VO**

211 For blends containing VOO, PLS-DA model-2 classifies the oil sample according to the proportion
212 of VO in the mixture, i.e. low (0–20% VO in VOO) and high (25–90% VO in VOO), with correct
213 prediction abilities of 98% and 97% respectively (Table 1). The most important variables on this
214 model were the $^{1}$H-signals of oleic acid (#9b) and squalene (#11), whose signal intensities were
215 higher in the low class. Indeed, VOO is known to be one of the vegetable oils that presents the
216 highest contents of oleic acid and squalene (Jiang et al., 2018; Popescu et al., 2015; Vigli et al.,
217 2003).

218 Pure VOOs are distinguished from blends with 2–20% VO in VOO, being identified even 92% of
219 the pure VOOs and 90% of the VO-VOO blends (PLS-DA models 3 and 4 in Table 1). The main
220 $^{1}$H-signals involved in the distinction of both classes were due to saturated fatty acids (#7a, #9a),
221 which exhibited lower intensities in the VO-VOO class. In fact, saturated fatty acids are the second
222 major class of fatty acids in VOO, being present in higher or similar concentrations than in the VOs
223 studied, i.e. NTSO, HOSO, EVAO, HV, HR and S (Contiñas, Martínez, Carballo, & Franco, 2008;

Guillén et al., 2003; Jabeur, Zribi, Makni, Rebai, Abdelhedi, & Bouaziz, 2014; Jiang et al., 2018; Jović, Smolić, Primožič, & Hrenar, 2016; Monfreda et al., 2012; Ranade & Thiagarajan, 2015; Yang, Ferro, Cavaco, & Liang, 2013). Concerning the discrimination of blends of 2% VO in VOO for a certain VO, a satisfactory classification model was only achieved for soybean oil; thus, all blends with 2% S in VOO were detected, and 97% of the blends with 2% of other VO in VOO were correctly predicted (PLS-DA model-5 in Table 1).

The $^{1}$H-NMR fingerprint of an oil sample classified in the low class (0–20% VO in VOO) is then submitted to classification models developed for each VO (PLS-DA models 6–24) to identify which particular VO is contained in the oil sample (Tables 2 and S2–S3 in supplementary material). The classification abilities of the PLS-DA models were better when the dataset contained only the data of blended oils with 5–20% VO in VOO than when data of pure VOO and/or 2% VO in VOO was also included. The prediction abilities ranged between 83% and 98% of hits depending on the VO blended with VOO. Similarly, when an oil sample is classified in the high class (25–90% VO in VOO), its $^{1}$H-NMR fingerprint is submitted to PLS-DA models developed for mixtures of 20–90% VO in VOO (PLS-DA models 25–28 in Table 3) to identify the VO contained in the blend. In the present study, only binary mixtures of NTSO, HOSO, EVAO or HV with VOO were available in the range of 20–90% VO. The recognition and prediction abilities of the classification models built to determine whether the VOO blend contained NTSO, HV or EVAO were 99–100% for both classes, and 100% for the non-HOSO class and 92% for the HOSO class.

Regarding the most influential variables on the models, the $^{1}$H-signal of oleic acid (#9b) was completely discriminant between VOO mixtures with high % NTSO and those with other VOs. The blends of 20–90% NTSO in VOO contained significantly lower amounts of oleic acid than VOO blends with 20–90% HOSO, EVAO or HV. It is well-documented that virgin hazelnut oil, high oleic sunflower oil and virgin avocado oil present significantly higher contents of oleic acid than sunflower oil (Contiñas et al., 2008; Guillén et al., 2003; Jabeur et al., 2014; Jović et al., 2016;

Ranade et al., 2015; Vigli et al., 2003; Yang et al., 2013). Other important variables to discriminate the presence of NTSO in VOO were the $^{1}$H-signals due to linoleic acid (#13c, #12b, #7c) and unsaturated fatty acids (#24), which presented higher intensities in NTSO-VOO mixtures than in most of the other VO-VOO blends (Contiñas et al., 2008; Guillén et al., 2003; Jović et al., 2016; Ranade et al., 2015; Vigli et al., 2003). Concerning the most important $^{1}$H-signals on HOSO models, the signal intensities of linolenic acid (#13d, #12c) and unsaturated fatty acids (#24 at 5.30–5.32 ppm) were lower in the HOSO-VOO mixtures; whereas those of linoleic acid (#13c, #12b, #9c), unsaturated fatty acids (#24 at 5.32–5.34 ppm) and terpenic alcohols or sterols (#2) were higher in HOSO-VOO mixtures. These observations agreed with the fact that HOSO presents higher concentrations of linoleic acid than VOO, HV and EVAO and lower than NTSO; and HOSO contains lower amounts of linolenic acid than NTSO, VOO and EVAO, and similar to HV (Guillén et al., 2003; Jović et al., 2016; Ranade et al., 2015). Moreover, the mixture of HOSO with VOO leads to an increase in the sterol content compared to pure olive oil (Al-Ismail, Alsaed, Ahmad, & Al-Dabbas, 2010). Evaluating the main variables on the EVAO models, it was observed that the $^{1}$H NMR spectra of the mixtures of EVAO in VOO showed higher intensities for the signals of saturated fatty acids (#10a, #7a, #9a), oleic acid (#7b, #12a, #9b), linoleic acid (#12b, #13c, #10c), squalene (#11) and β-sitosterol (#4) than the spectra of the other VO-VOO blends. Meanwhile, the $^{1}$H-signals of unsaturated fatty acids (#24, #9 at 1.32–1.36 ppm) and linolenic acid (#13d, #12c, #9c) presented lower intensities in the EVAO-VOO blends. Indeed, EVAO presents the highest contents of the saturated fatty acids, mainly palmitic acid, of all the VOs blended with VOO in this study; similar intermediate amounts of oleic and linoleic acids as HOSO; and low concentrations of linolenic acid as VOO, HV and HR (Guillén et al., 2003; Jabeur et al., 2014; Jović et al., 2016; Ranade et al., 2015). To distinguish blends with high % HV in VOO, the $^{1}$H-signals of oleic acid (#7b, #9b, #12a), whose intensities were significantly higher in the HV class, were among the most important variables on the HV models. HV presents similar or slightly higher contents of oleic acid than VOO, and considerably higher amounts compared to the other VOs studied (Guillén et al.,

2003). The opposite trend was shown by the $^{1}$H-signals of linoleic (#7c) and linolenic (#12c) acids, which displayed lower intensity values in the HV class than in the non-HV class. Certainly, the concentrations of linoleic acid in HV are lower than in the other VOs and slightly higher than in VOO; and linolenic acid is present in similar amounts in HV and HOSO but lower amounts in HV than in NTSO, VOO and EVAO (Christopoulou, Lazaraki, Komaitis, & Kaselimis, 2004; Jović et al., 2016; Vigli et al., 2003). For the distinction of mixtures of low % HR in VOO from other VO-VOO mixtures, the $^{1}$H-signals of oleic (#12a) and linolenic (#12c, #7d) acids, saturated fatty acids (#7a) and terpenic alcohols or sterols (#2) exhibited lower intensities in the HR class (Guillén et al., 2003; Vigli et al., 2003). The most discriminant variables in the models to detect low % S in VOO were the $^{1}$H-signals of linolenic acid (#15b, #7d, #12c) and unsaturated fatty acids (#24), which presented significantly higher intensities in S-VOO blends than in the other VO-VOO blends. Soybean oil is the oil with the highest contents of linolenic acid among the studied VOs (Contiñas et al., 2008; Christopoulou et al., 2004; Guillén et al., 2003; Jabeur et al., 2014; Vigli et al., 2003). Furthermore, the lower signal intensities of oleic (#7b) and linoleic (#13c) acids in the S class also contributed to the discrimination of both classes, being consistent with the literature reporting that soybean oil presents significantly lower contents of oleic acid than VOO, and similar contents of linoleic acid as other VOs, such as sunflower oil (Guillén et al., 2003; Jović et al., 2016; Vigli et al., 2003).

### 3.4. PLS-DA models to discriminate blends of OO with VO

Satisfactory binary classification models for all the studied VOs (RPOO, CO, HOSO, NTSO, DOSO, RAO and HR) were obtained using the data of the full % range of VO in the OO mixture, i.e. 0–80% VO in OO (PLS-DA models 30–36 in Table S4 (supplementary material). Prediction abilities were 95–100% for both classes in the models developed to discriminate between OO blends with and without RPOO, CO or HOSO; 84–87% for the OO mixtures with NTSO, DOSO or RAO, and 91–97% for the OO blends that did not contain the corresponding specific VO; and 97%

for the HR class and 89% for the non-HR class. These classification results were improved for each VO by further PLS-DA models developed separately for blends with low or high % VO in OO. Hence, the oil sample containing OO is first classified according to its level of VO, i.e. low (0–20% VO in OO) or high (30–80% VO in OO), by PLS-DA model-29 with prediction abilities of 96% and 94% respectively (Table 1). The most influential variables on this model were the $^{1}$H-signals of saturated fatty acids (#7a), β-sitosterol (#4), linoleic acid (#12b, #15a, #13c) and unsaturated fatty acids (#24, #7 at 1.00–1.02 ppm, #9 at 1.32–1.33 ppm), which exhibited lower intensities in the low class; and those of linolenic (#7d, #15b) and oleic (#12a) acids, which displayed higher intensities in the low class. The chemical composition of the blends that constituted each class justified these observations; thus, the low class contained the samples with the highest % of OO, which is the oil that contains the highest concentrations of oleic acid, together with HR; whereas the high class included the samples with high % of VO characterised by high linoleic and β-sitosterol contents (Al-Ismail et al., 2010; Aparicio & Harwood, 2013; Green & Wang, 2020; Guillén et al., 2003; Jović et al., 2016; Parcerisa, Casals, Boatella, Codony, & Rafecas, 2000; Vigli et al., 2003).

An oil sample containing low % VO in OO is then subjected to various classification models (PLS-DA models 37–50) to identify the specific VO contained in the OO blend (Tables 2 and S5 in supplementary material). The recognition and prediction abilities of these models were higher than 95% of hits for detecting RPOO, CO and HOSO in OO; c.a. 90% for NTSO, DOSO and HR in OO; and c.a. 80–85% for RAO in OO. Taking into account that all CO-OO blends, 95% of the RPOO-OO blends, and at least 95% of the OO blends not containing CO or RPOO were identified with the corresponding models for low % VO in OO, further classification models were developed using datasets without the $^{1}$H-NMR spectral data of RPOO-OO and CO-OO mixtures. The PLS-DA models achieved (PLS-DA models 51–55) afforded better classification abilities to detect NTSO and RAO in OO, and similar results to resolve the presence of HOSO, DOSO or HR in OO (Table S6 in supplementary material).

For oil samples with high % VO in OO, the classification models developed for blends with 20–80% VO in OO (PLS-DA models 56–62) presented recognition and prediction abilities of 98–100% for both classes in RPOO, CO, DOSO and HR models; ≥91% for both classes in NTSO and RAO models; and 86% for the HOSO class and 99% for the non-HOSO class (Table 3). Since all blends were correctly classified by the RPOO and CO models, further PLS-DA models to detect 20–80% VO in OO were built using a dataset without the $^{1}$H-NMR spectral data of RPOO-OO and CO-OO blends (PLS-DA models 63–67 in Table S7 in supplementary material). These models provided the same or better classification abilities than the previous ones, except for HR-OO blends. Indeed, the NTSO and HOSO models allowed the correct classification of all samples of both classes; and the RAO model identified all samples containing RAO and 92% of the samples in the non-RAO class. The main $^{1}$H-signals responsible for the identification of OO blends containing RPOO were those of saturated fatty acids (#9a), which presented significantly higher intensities in the RPOO-OO blends; and those of linoleic acid (#9c, #12b), which showed lower intensities in the RPOO class. The $^{1}$H-signals #9a and #9c were completely discriminants between OO blends containing ≥20% RPOO and the other VO-OO blends with high % VO. As a result, the measurement of just one of these two variables would be enough to confirm whether an OO is mixed with RPOO in percentages ≥20%. Palm oil is the oil that contains the highest amounts of saturated fatty acids among the VOs studied (Vigli et al., 2003). Palmitic acid is the major saturated fatty acid in palm oil and is contained in similar amounts as oleic acid. Meanwhile, linoleic acid is a minor compound in palm oil, present in similar concentrations as in OO, and in lower amounts than in the rest of VOs (Montoya, Cochard, Flori, Cros, Lopes, Cuellar, et al., 2014). The CO-OO blends were distinguished from the other VO-OO mixtures due to the $^{1}$H-signals of linoleic (#7c) and linolenic (#15b, #7d) acids, saturated fatty acids (#7a) and β-sitosterol (#4), which presented higher intensities in the blends containing CO; and to the signal of oleic acid (#9b) with lower intensities in the CO class. Actually, corn oil presents linoleic acid in amounts similar to sunflower oil and significantly higher than refined avocado, refined hazelnut, palm and olive oils; linolenic acid and

β-sitosterol in slightly higher concentrations than the other oils studied; saturated fatty acids in lower contents than palm oil but similar or slightly higher than the rest of the oils considered in the model; and the lowest content of oleic acid, together with sunflower oil. (Aparicio et al., 2013; Guillén et al., 2003; Monfreda et al., 2012; Vigli et al., 2003). The major contributors to the discrimination of HOSO from other VOs in OO were the $^{1}$H-signals of oleic (#9b, #12a) and linoleic (#12b, #9c) acids and saturated (#9a) and unsaturated (#24, #9 at 1.30–1.34 ppm) fatty acids, which exhibited higher intensities in the OO blends with HOSO. Indeed, HOSO contains higher amounts of oleic acid than sunflower, corn and palm oils; similar to avocado oil; and lower than hazelnut and olive oils. Linoleic acid is present in larger concentrations in HOSO than in palm, olive, hazelnut and avocado oils, and smaller than in sunflower and corn oils. The content of saturated fatty acids (#9a) in HOSO is intermediate-high with respect to other VOs but far from those of RPOO, which exhibit the largest contents (Green et al., 2020; Guillén et al., 2003; Jović et al., 2016; Vigli et al., 2003). As in NTSO-VOO models, the most influential variables on the classification models achieved for the detection of NTSO in OO were the $^{1}$H-signals of linoleic acid (#7c, #15a, #12b) and unsaturated fatty acids (#24, #7 at 1.00–1.02 ppm, #9 at 1.32–1.36 ppm), displaying higher intensities in the OO blends with NTSO; and oleic acid (#12a, #7b, #9b), showing the opposite trend. For OO blends with 20–80% NTSO, once the presence of RPOO and CO in the OO blend was discarded by the PLS-DA models 56 and 57 respectively (Table 3), not only the signal of oleic acid (#9b) but also several other signals (#15a, #12b, #9 at 1.34–1.36 ppm, #24) were completely discriminant between both classes; therefore any of them can be used as markers to determine whether an OO blend contains NTSO at concentrations ≥20%. Sunflower oil is characterised by the largest contents of linoleic and unsaturated fatty acids, and the lowest contents of oleic acid with regard to the other VOs studied (Guillén et al., 2003; Jabeur et al., 2014; Jović et al., 2016; Monfreda et al., 2012; Yang et al., 2013). The DOSO models disclosed that the intensities of the $^{1}$H-signals due to oleic acid (#12a, #9b) were significantly higher in DOSO-OO blends, in contrast with linoleic acid (#12b, #7c, #24) signals exhibiting higher intensities in the non-DOSO

class. During the desterolization process, it takes place the dehydration of sterols and the elimination of the acid group of sterol esters by bleaching, producing olefinic degradation products and di-steryl ethers; meanwhile the profiles of triacylglycerides and fatty acids are practically unaltered (Grob, Biedermann, Bronz, & Giuffré, 1994). Therefore, it would be expected that DOSO presents relatively high contents of oleic and linoleic acids as HOSO. However, the deodorization process may affect the composition of triglycerides, diglycerides, fatty acids and minor components of the unsaponifiable fraction, depending mainly on the temperature and time of the process (Aparicio et al., 2013), which could be responsible for the lower content of linoleic acid observed in DOSO blends in relation to the other VOs, including HOSO. The main $^{1}$H-signals on the RAO models were linoleic (#7c, #12b, #13c, #10c) and oleic (#9b) acids and β-sitosterol (#4), exhibiting similar or higher intensities in RAO-OO blends; linolenic acid (#13d, #9c) and unsaturated fatty acids (#9 at 1.32−1.34 ppm, #24), displaying similar or lower intensities in the RAO class; and saturated fatty acids (#9 at 1.20−1.22 ppm) with intermediate intensities. In fact, refined avocado oil, compared to the other VOs studied, presents intermediate compositions of fatty acids (Guillén et al., 2003; Jabeur et al., 2014; Jović et al., 2016; Vigli et al., 2003; Yang et al., 2013) and sterol contents, in particular, β-sitosterol (Al-Ismail et al., 2010; Green et al., 2020; Parcerisa et al., 2000). The most contributing variables to the identification of HR in OO were the $^{1}$H-signals of oleic (#7b, #12a, #9b) and linoleic (#12b) acids, presenting higher intensities in the HR class; and the signals of linolenic acid (#7d, #15b, #12c, #13d), unsaturated (#24) and saturated (#10a, #7a) fatty acids and terpenic alcohols or sterols (#2), showing lower intensities in the HR-OO mixtures. The trend of oleic and linoleic signals observed in HR-OO is opposite to that in HR-VOO. Refined hazelnut oil contains the highest amounts of oleic acid among the VOs studied, comparable to those in OO but lower than VOO; the lowest linolenic contents, similar to those found in HOSO (Green et al., 2020; Guillén et al., 2003; Jović et al., 2016; Parcerisa et al., 2000; Vigli et al., 2003); and characteristic profiles of sterols and terpenic alcohols (Al-Ismail et al., 2010; Aparicio et al., 2013; Parcerisa et al., 2000).

### 3.5. PLS-R models to determine the percentage of VO in a blend with VOO or OO

PLS regression models to determine the % VO contained in a binary mixture with VOO or OO (PLS-R models 1–27) were successfully built for all VOs studied (Table 4). The PLS-R models developed for different sub-ranges of % VO in VOO or OO provided more accurate predictions than those constructed for the full % VO range. The most influential variables on the regression models coincided with those on the classification ones. Therefore, the regression results were explained by the characteristic composition in fatty acids, triacylglycerides and squalene of the oils present in the blend. In VO-VOO models, diacylglycerides, terpenic alcohols and sterols were also decisive.

All regression models presented excellent precisions; yielding $R^2$ values 0.93–0.990, except for the low % range models of VOO mixtures with NTSO, HOSO, HR and S. The PLS-R models for low % NTSO, HOSO and S in VOO presented $R^2$ values <0.70, indicating that the equation can only be used for screening purposes, which enables to distinguish between low, medium and high values of % VO. The PLS-R model for low % HR in VOO showed $R^2$ values <0.50, so the equation only discriminates between high and low values (Priego Capote, Ruiz Jiménez, & Luque De Castro, 2007), in the same way as PLS-DA model-73 distinguishes 2–5% HR and 10% HR in VOO (Table 5).

The regression models achieved allow to determine the % VO in a VOO blend with uncertainties under 5% R-RMSEP for contents of ≥10% NTSO, ≥34% EVAO, ≥39% HOSO and ≥45% HV; 5–10% R-RMSEP for contents of 13–45% HV; 5–15% R-RMSEP for contents of 8–10% NTSO, 7–34% EVAO, 20–39% HOSO and 10–26% HV; 15–20% R-RMSEP for contents of 6–8% NTSO, 5–7% EVAO, 17–20% HOSO and 5% S; and with uncertainty of 28% R-RMSEP for contents of 10% HR. Considering VO-OO blends, the % VO in OO was quantified with uncertainties under 5% R-RMSEP for contents of ≥5% RPOO, ≥6% CO, ≥10% HR, ≥16% DOSO, ≥16% HOSO, ≥9% NTSO and ≥31% RAO; 5–15% R-RMSEP for contents of 2–5% RPOO, 2–6% CO, 3–10% HR,

5–16% DOSO, 7–16% HOSO, 3–9% NTSO and 5–31% RAO; and 15–20% R-RMSEP for contents of 2–3% HR, 4–5% DOSO, 5–7% HOSO, 2–3% NTSO and 4–5% RAO.

The classification abilities of the PLS-DA models to identify blends with low % HV, HR, HOSO and NTSO in VOO and low % RAO in OO were considerably improved when the samples of 2% VO in VOO and/or pure olive oil (VOO or OO) were removed from the dataset used to develop the models (Table 2), indicating that these samples were close to the boundary and therefore could be misclassified. Regarding this fact and the precisions and accuracies of the regression models built, the experimental detection limits were established in the ranges between 2–5% VO for blends of HV, HR, HOSO or NTSO in VOO; between 2–4% VO for blends of RAO in OO; and under 2% VO for blends of EVAO or S in VOO and RPOO, CO, HOSO, NTSO, DOSO or HR in OO. The present results are similar or outperform those reported in the literature using NMR (Parker et al., 2014; Wang et al., 2020) or other analytical techniques (De La Mata-Espinosa et al., 2011; Grob et al., 1994; Jabeur et al., 2014; Jović et al., 2016; Monfreda et al., 2012). In previous high-field NMR studies, the adulteration of refined hazelnut oil in olive oil was detected at a proportion of 10% using $^{1}$H-NMR and linear discriminant analysis (Mannina et al., 2009), 8% using $^{1}$H and $^{13}$C-NMR and artificial neural networks (García-González et al., 2004), 1% using $^{1}$H and $^{31}$P-NMR and canonical discriminant analysis or classification trees (Agiomyrgianaki et al., 2010), and 5% of hazelnut oil in VOO using $^{13}$C-NMR and discriminant data analysis (Zamora et al., 2001). $^{1}$H and $^{31}$P-NMR together with discriminant analysis allowed the detection of adulterations as low as 5% of hazelnut, corn, sunflower and soybean oils in VOO (Vigli et al., 2003). $^{13}$C-NMR and discriminant data analysis distinguished palm oil at 5% in OO (Guyader, Thomas, Portaluri, Jamin, Akoka, Silvestre, et al., 2018). The determination of the contents of oleic, linoleic, linolenic and saturated fatty acids and squalene by $^{1}$H-NMR enabled the detection of 4.5% soybean oil in VOO (Jiang et al., 2018). Nevertheless, chromatographic techniques afforded the lowest limits of detection for sunflower, soybean, corn and palm oils in VOO, detecting even 0.1% adulteration (Jabeur, Zribi, & Bouaziz, 2016).

### 3.6. PLS-DA models to discriminate between ‘legal’ and ‘illegal’ blends of VOO or OO with VO

The potential of the present multivariate approach to implement Reg. (EU) 29/2012 and its amendments is demonstrated with a case study. The most common vegetable oil used to be blended with olive oil is sunflower oil. Therefore NTSO and HOSO were considered as model VOs in ‘legal’ blends with VOO or OO, as done in previous studies (Gómez-Coca et al., 2020; Monfreda et al., 2012). The olive oil blends with the other VOs studied were regarded as ‘illegal’ blends. Binary classification models were developed to first distinguish between ‘legal’ and ‘illegal’ blends, and then differentiate which of the two types of sunflower oils, i.e. NTSO or HOSO, is in the ‘legal’ blend with VOO or OO (Figure S1 in supplementary material). The percentage of NTSO or HOSO in the mixture is determined by the regression models that are reported in the previous section (Table 4).

The PLS-DA model discriminating between ‘legal’ and ‘illegal’ blends provided prediction abilities of 77% for both classes concerning blends with VOO (PLS-DA model-68), and 86% and 98% respectively for blends with OO (PLS-DA model-70 in Table 5). The most discriminant variables on these models are shown in Table S8 (supplementary material). The trends observed for the $^{1}$H-signals involved were consistent with the known differences in the chemical composition of NTSO and HOSO with respect to the VOs in the ‘illegal’ class and both categories of olive oils, already mentioned above.

In addition, classification models were constructed to distinguish ‘legal’ blends containing NTSO from those with HOSO, affording prediction abilities of 83–85% for blends with VOO (PLS-DA model-69), and 97% for blends with OO (PLS-DA model-71 in Table 5). HOSO contains higher amounts of oleic acid and lower concentrations of linoleic and linolenic acids (polyunsaturated fatty acids) than NTSO (Jović et al., 2016), which is reflected on the most influential $^{1}$H-signals on these models (Table S8 in supplementary material).

### 3.7. PLS-DA models to discriminate between blends of VOO or OO with different compositions

Further binary classification models can be built using datasets containing only the information related to specific VOs or % VO in the blends. These complementary models are useful whenever an oil sample is predicted to contain a certain VO by more than one of the classification models described above. Likewise, in the case that the determination of the % VO is not enough accurate by the corresponding regression model for low percentages, it is interesting to be able to discriminate between mixtures with different % VO. As a proof of concept, binary classification models were developed to distinguish blends of different % S or HR in VOO (PLS-DA models 72 and 73); and OO mixtures containing DOSO or HR (PLS-DA model-74), RAO or HR (PLS-DA model-75), RAO or DOSO (PLS-DA model-76) and DOSO or HOSO (PLS-DA model-77), with satisfactory classification abilities (Table 5). The most influential $^{1}$H-signals on these models are gathered in Table S8 (supplementary material). Depending on the class and model considered, different trends were observed in the signal intensities, which are in accordance with the relative chemical composition of each kind of oil in the blend previously reported. The major fatty acids in S and VOO are linoleic acid and oleic acid respectively (Vigli et al., 2003). VOO contains higher amounts of squalene and linolenic acid than HR, and the opposite occurs for linoleic acid (Guillén et al., 2003; Vigli et al., 2003). HR presents higher contents of oleic acid, similar concentrations of linoleic acid and lower amounts of saturated fatty acids than RAO (Green et al., 2020; Parcerisa et al., 2000). In respect of the main variables on the models obtained for the discrimination of DOSO-OO blends from other VO-OO mixtures, DOSO-OO blends contained higher concentrations of oleic acid than OO blends of HR, RAO and HOSO, which are the VOs that present the highest contents of oleic acid (Green et al., 2020; Guillén et al., 2003; Jović et al., 2016; Parcerisa et al., 2000); and lower amounts of linoleic acid than OO blends of HR, RAO and HOSO. Taking into account that DOSO is obtained from the desterolization and deodorization of HOSO, these results evidenced that during the deodorization and/or desterolization processes the fatty acid profile of the

oil was altered, resulting in lower linoleic and higher oleic contents. In this sense, it has been already reported that the drastic conditions used during raffination processes lead to olefinic degradation of sterols, the isomerization of squalene and linoleic and linolenic acids, among other changes in the chemical composition of the oil (Aparicio et al., 2013; Grob et al., 1994).

### 3.8. Prediction of blends of olive oil with other vegetable oils

The composition of thirty-six blind oil samples provided within the OLEUM Project and eight commercial oils was predicted by the classification and regression models developed for blends of olive oil with other vegetable oils following the decision trees shown in Figures 1 and S1 (supplementary material). For each blind sample, Table S9 (supplementary material) gathers *i*) the PLS-DA and PLS-R models applied; *ii*) the PLS-DA predictions related to the category of the olive oil (VOO or OO), the VO contained, and the low/high level of VO in the blend (Tables 1–3, S2–S7 in supplementary material); *iii*) the % VO in the blend determined by the corresponding PLS-R model (Table 4); and *iv*) the predictions of the complementary PLS-DA models (Table 5). Most of the blind samples were predicted satisfactorily according to the description provided (Table S9 in supplementary material); thus, the category of olive oil, i.e. VOO or OO, the particular VO and the % VO in the oil sample were accurately determined. All mixtures of VOO or OO with 40–60% NTSO or HOSO (**1**–**12**), all the blends (containing 5–30% VO) of RPOO-OO (**29**–**32**) and HV-VOO (**17**–**20**), and the blends of EVAO-VOO (**14**–**16**) and HR-OO (**26**–**28**) with ≥10% VO were correctly identified and the % VO properly figured out. Only blind samples **16**, **17** and **19** were predicted to present slightly higher % VO in VOO, and sample **26** scarcely lower % HR in OO, than those percentages given in the description. The DOSO-OO blends (**33**–**36**) were satisfactorily determined by the corresponding classification and regression models; the % DOSO in OO in sample **36** was barely lower than predicted. The blend of 10% DOSO in OO (**34**) was confused with mixtures of 2–11% of HOSO in OO. For the blend of 5% EVAO in VOO (**13**), the contained VO was not recognised by any of the classification models, but the calculated % VO was within the

calibration range of the regression model developed for EVAO-VOO blends; and this model predicted correctly the % EVAO in the mixture, even with better precisions than the other models built for HOSO-VOO and HR-VOO blends. The VO in the blend of 5% HR in OO (**25**) was not identified by any of the HR-OO classification models. Indeed, the detection of the adulteration of OO with HR is still one of the main challenges in fraud detection due to the close composition of both refined oils (Agiomyrgianaki et al., 2010; García-González et al., 2004; Mannina et al., 2009). Even blends with ≤10% HR in OO can be confused with RAO-OO blends. The composition of blind samples **21**–**24** were determined by the classification and regression models built for both RAO-OO and DOSO-OO blends; however, the PLS-DA model-76 (Table 5), which distinguishes these two OO mixtures, predicted satisfactorily that these blind samples contained RAO, except for the mixture of 10% RAO in OO (**22**).

Regarding the commercial oils analysed, samples **37**, **38** and **44** were declared to be mixtures of vegetable oils or NTSO with EVOO or VOO. Samples **37** and **38** were confirmed to contain VOO, whereas sample **44** was classified as an OO blend. Furthermore, the three samples were predicted to contain NTSO, in accordance with their label specifications. All the other commercial oil samples (**39**–**43**) were labelled as mixtures of VOO or EVOO with rapeseed oil; however, all of them were classified as blends of OO. These results are not conclusive since no blends of rapeseed oil with VOO or OO were available to be included in the modelling step of the present study.

## 4. Conclusion

A stepwise strategy based on $^{1}$H-NMR fingerprinting of an oil sample in combination with chemometrics is proposed to determine the content of mixtures of oils of the 'virgin olive oil' or 'olive oil' categories and vegetable oils, providing a chemical tool to (*i*) confirm the presence of VOO or OO in an oil sample; (*ii*) discriminate between pure olive oils and their blends with VOs to a certain extent, given by the detection limit disclosed for each VO; (*iii*) identify the VO in the blend with VOO or OO; (*iv*) differentiate between blends made with different VOs in VOO or OO;

555 (*v*) distinguish blends made with the same VO in different proportions; and (*vi*) determine the %
556 VO blended with VOO or OO.

557 $^{1}$H-NMR spectral data of olive oils and their mixtures with the VOs most commonly used to make
558 blends, i.e. sunflower oil, high oleic sunflower oil, desterolized high oleic sunflower oil, virgin and
559 refined avocado oil, virgin and refined hazelnut oil, refined palm olein oil, corn oil and soybean oil,
560 was used to optimize and validate classification and regression models built by PLS-DA and PLS-R
561 respectively. The classification models achieved were satisfactory, robust and stable. Excellent
562 precisions and acceptable accuracies were afforded by the regression models developed for the
563 determination of the % VO in VOO or OO. The reliability of the classification and regression
564 models was supported by the chemical interpretation of the most influential variables on the
565 validated models. The % VO in the blend is determined with uncertainties under the 20% of R-
566 RMSEP for contents as low as 5% EVAO or S, 6% NTSO, 10% HV and 17% HOSO in VOO; and
567 2% RPOO, CO, NTSO or HR, 4% DOSO or RAO and 5% HOSO in OO. The detection limits are
568 under 2% EVAO or S and between 2–5% NTSO, HOSO, HV or HR in VOO; and under 2% RPOO,
569 CO, HOSO, NTSO, DOSO or HR and 2–4% RAO in OO. The performance and effectiveness of the
570 proposed strategy were validated by a set of blind samples, which confirmed its feasibility to
571 support Reg. (EU) 29/2012. Further studies should be carried out with larger balanced sample sets
572 covering the variability of olive oils of both categories (VOO and OO) and the vegetable oils of
573 interest. The different possible sources of variability, such as the varieties of each botanical oil
574 species, the agronomical and climatic conditions, the geographical origins and harvests, should be
575 considered. The implementation of this approach requires a databank of $^{1}$H-NMR fingerprints of
576 oils. The databank has to include pure oils comprising olive oils of the different categories,
577 vegetable oils used to make legal blends and adulterant oils, and their mixtures; because it has to be
578 representative of oil variability in order to guarantee robust models for both authentication and
579 fraud detection. It is worth noting that this requirement is feasible in practice since the creation of
580 the OLEUM Databank and the OLEUM Network are among the objectives of the OLEUM Project

581 that are being accomplished. The OLEUM Databank is an online integrated quality assurance
582 database of olive oil analytical methods and chemical data, which is currently being developed. The
583 OLEUM Network is a worldwide community of proficient analytical laboratories involved in olive
584 oil analysis, and it is expected to expand and may also contribute to the feeding and updating of the
585 databank over time.

586 **Acknowledgments**

587 This work was developed in the framework of the project OLEUM "Advanced solutions for
588 assuring authenticity and quality of olive oil at global scale" funded by the European Commission
589 within the Horizon 2020 Programme (2014–2020), grant agreement No. 635690; and the project
590 AUTENFOOD funded by ACCIÓ-Generalitat de Catalunya and the European Union through the
591 Programa Operatiu FEDER Catalunya 2014-2020 (Ref COMRDI-15-1-0035). The information
592 contained in this article reflects the authors' views; the European Commission is not liable for any
593 use of the information contained herein. The authors would like to thank all producers that supplied
594 the olive oils, virgin olive oils and vegetable oils for this study, and the technical and staff support
595 provided by SGIker (UPV/EHU, MICINN, GV/EJ, ESF).

596 **References**

597 Agiomyrgianaki, A., Petrakis, P. V., & Dais, P. (2010). Detection of refined olive oil adulteration
598 with refined hazelnut oil by employing NMR spectroscopy and multivariate statistical
599 analysis. *Talanta, 80*(5), 2165-2171. https://doi.org/10.1016/j.talanta.2009.11.024.
600 Al-Ismail, K. M., Alsaed, A. K., Ahmad, R., & Al-Dabbas, M. (2010). Detection of olive oil
601 adulteration with some plant oils by GLC analysis of sterols using polar column. *Food*
602 *Chemistry, 121*(4), 1255-1259. https://doi.org/10.1016/j.foodchem.2010.01.016.
603 Alonso-Salces, R. M., Segebarth, N., Garmón-Lobato, S., Holland, M. V., Moreno-Rojas, J. M.,
604 Fernández-Pierna, J. A., et al. (2015). $^{1}$H-NMR and isotopic fingerprinting of olive oil and
605 its unsaponifiable fraction: Geographical origin of virgin olive oils by pattern recognition.

606 *European Journal of Lipid Science and Technology, 117*(12), 1991-2006.
607 https://doi.org/10.1002/ejlt.201400243.
608 Aparicio, R., & Harwood, J. (2013). *Handbook of olive oil: Analysis and properties* (2nd ed.). USA:
609 Springer US.
610 Berrueta, L. A., Alonso-Salces, R. M., & Héberger, K. (2007). Supervised pattern recognition in
611 food analysis. *Journal of Chromatography A, 1158*(1-2), 196-214.
612 https://doi.org/10.1016/j.chroma.2007.05.024.
613 Conte, L., Bendini, A., Valli, E., Lucci, P., Moret, S., Maquet, A., et al. (2020). Olive oil quality
614 and authenticity: A review of current EU legislation, standards, relevant methods of
615 analyses, their drawbacks and recommendations for the future. *Trends in Food Science &*
616 *Technology, 105*, 483-493. https://doi.org/10.1016/j.tifs.2019.02.025.
617 Contiñas, A., Martínez, S., Carballo, J., & Franco, I. (2008). Detection of contaminations and/or
618 adulterations of the extra virgin olive oil with seeds oils (sunflower and soybean) and olive
619 pomace oil. *Grasas y Aceites, 59*(2), 97-103. https://doi.org/10.3989/gya.2008.v59.i2.496.
620 Christopoulou, E., Lazaraki, M., Komaitis, M., & Kaselimis, K. (2004). Effectiveness of
621 determinations of fatty acids and triglycerides for the detection of adulteration of olive oils
622 with vegetable oils. *Food Chemistry, 84*(3), 463-474. https://doi.org/10.1016/s0308-
623 8146(03)00273-5.
624 De la Mata, P., Dominguez-Vidal, A., Bosque-Sendra, J. M., Ruiz-Medina, A., Cuadros-Rodríguez,
625 L., & Ayora-Cañada, M. J. (2012). Olive oil assessment in edible oil blends by means of
626 ATR-FTIR and chemometrics. *Food Control, 23*(2), 449-455.
627 EC. (2012). European Commission Regulation 29/2012 on marketing standards for olive oil.
628 *Official Journal of the European Union, January 13, 2012*, 12-21.
629 EC. (2020). European Commission website. Food Fraud.
630 https://ec.europa.eu/knowledge4policy/food-fraud-quality/food-fraud-data-bases_en,
631 Accessed date: December 1, 2020.

Esbensen, K. H., Guyot, D., Westad, F., & Houmøller, L. P. (2002). *Multivariate data analysis - in practice: An introduction to multivariate data analysis and experimental design* (5th ed.). Oslo: Camo Process AS.

European Parliament. (2014). Committee on the Environment, Public Health and Food Safety. Resolution of 14 January 2014 on the food crisis, fraud in the food chain and the control thereof (2013/2091 (INI)).

Fregapane, G., Gómez-Rico, A., Inarejos, A. M., & Salvador, M. D. (2013). Relevance of minor components stability in commercial olive oil quality during the market period. *European Journal of Lipid Science and Technology, 115*(5), 541-548. https://doi.org/10.1002/ejlt.201200209.

García-González, D. L., Mannina, L., D'Imperio, M., Segre, A. L., & Aparicio, R. (2004). Using $^{1}H$ and $^{13}C$ NMR techniques and artificial neural networks to detect the adulteration of olive oil with hazelnut oil. *European Food Research and Technology, 219*(5), 545-548. https://doi.org/10.1007/s00217-004-0996-0.

Gómez-Coca, R. B., Pérez-Camino, M. d. C., Martínez-Rivas, J. M., Bendini, A., Gallina Toschi, T., & Moreda, W. (2020). Olive oil mixtures. Part one: Decisional trees or how to verify the olive oil percentage in declared blends. *Food Chemistry, 315*, Article 126235. https://doi.org/https://doi.org/10.1016/j.foodchem.2020.126235.

Green, H. S., & Wang, S. C. (2020). First report on quality and purity evaluations of avocado oil sold in the US. *Food Control, 116*, Article 107328. https://doi.org/https://doi.org/10.1016/j.foodcont.2020.107328.

Grob, K., Biedermann, M., Bronz, M., & Giuffré, A. M. (1994). The Detection of Adulteration with Desterolized Oils. *Lipid / Fett, 96*(9), 341-345. https://doi.org/10.1002/lipi.19940960905.

Guillén, M. D., & Ruiz, A. (2003). Rapid simultaneous determination by proton NMR of unsaturation and composition of acyl groups in vegetable oils. *European Journal of Lipid Science and Technology, 105*(11), 688-696. https://doi.org/10.1002/ejlt.200300866.

Guyader, S., Thomas, F., Portaluri, V., Jamin, E., Akoka, S., Silvestre, V., et al. (2018). Authentication of edible fats and oils by non-targeted $^{13}C$ INEPT NMR spectroscopy. *Food Control, 91*, 216-224. https://doi.org/10.1016/j.foodcont.2018.03.046.

Jabeur, H., Zribi, A., & Bouaziz, M. (2016). Extra-virgin olive oil and cheap vegetable ols: Distinction and dtection of aulteration as dtermined by GC and chemometrics. *Food Analytical Methods, 9*(3), 712-723. https://doi.org/10.1007/s12161-015-0249-9.

Jabeur, H., Zribi, A., Makni, J., Rebai, A., Abdelhedi, R., & Bouaziz, M. (2014). Detection of chemlali extra-virgin olive oil adulteration mixed with soybean oil, corn oil, and sunflower oil by using GC and HPLC. *Journal of Agricultural and Food Chemistry, 62*(21), 4893-4904. https://doi.org/10.1021/jf500571n.

Jiang, X. Y., Li, C., Chen, Q. Q., & Weng, X. C. (2018). Comparison of $^{19}F$ and $^{1}H$ NMR spectroscopy with conventional methods for the detection of extra virgin olive oil adulteration. *Grasas y Aceites, 69*(2), Article e249. https://doi.org/10.3989/gya.1221172.

Jović, O., Smolić, T., Primožič, I., & Hrenar, T. (2016). Spectroscopic and chemometric analysis of binary and ternary edible oil mixtures: Qualitative and quantitative study. *Analytical chemistry, 88*(8), 4516-4524. https://doi.org/10.1021/acs.analchem.6b00505.

Kjeldahl, K., & Bro, R. (2010). Some common misunderstandings in chemometrics. *Journal of Chemometrics, 24*(7-8), 558-564. https://doi.org/10.1002/cem.1346.

Lioupi, A., Nenadis, N., & Theodoridis, G. (2020). Virgin olive oil metabolomics: A review. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences, 1150*, Article 122161. https://doi.org/10.1016/j.jchromb.2020.122161.

Mannina, L., D'Imperio, M., Capitani, D., Rezzi, S., Guillou, C., Mavromoustakos, T., et al. (2009). $^{1}H$ NMR-based protocol for the detection of adulterations of refined olive oil with refined hazelnut oil. *Journal of Agricultural and Food Chemistry, 57*(24), 11550-11556. https://doi.org/10.1021/jf902426b.

683 Monfreda, M., Gobbi, L., & Grippa, A. (2012). Blends of olive oil and sunflower oil:
684 Characterisation and olive oil quantification using fatty acid composition and chemometric
685 tools. *Food Chemistry, 134*(4), 2283-2290. https://doi.org/10.1016/j.foodchem.2012.03.122.
686 Montoya, C., Cochard, B., Flori, A., Cros, D., Lopes, R., Cuellar, T., et al. (2014). Genetic
687 architecture of palm oil fatty acid composition in cultivated oil palm (*Elaeis guineensis*
688 Jacq.) compared to its wild relative *E. oleifera* (H.B.K) Cortés. *PLoS ONE, 9*(5), Article
689 e95412. https://doi.org/10.1371/journal.pone.0095412.
690 OLEUM Project. (2016). About OLEUM: Aims and Objectives. http://www.oleumproject.eu/about-
691 oleum/aims-and-objectives, Accessed date: December 1, 2020.
692 Parcerisa, J., Casals, I., Boatella, J., Codony, R., & Rafecas, M. (2000). Analysis of olive and
693 hazelnut oil mixtures by high-performance liquid chromatography-atmospheric pressure
694 chemical ionisation mass spectrometry of triacylglycerols and gas-liquid chromatography of
695 non-saponifiable compounds (tocopherols and sterols). *Journal of Chromatography A,*
696 *881*(1-2), 149-158. https://doi.org/10.1016/S0021-9673(00)00352-6.
697 Parker, T., Limer, E., Watson, A. D., Defernez, M., Williamson, D., & Kemsley, E. K. (2014).
698 60MHz $^{1}$H NMR spectroscopy for the analysis of edible oils. *TrAC - Trends in Analytical*
699 *Chemistry, 57*, 147-158. https://doi.org/10.1016/j.trac.2014.02.006.
700 Popescu, R., Costinel, D., Dinca, O. R., Marinescu, A., Stefanescu, I., & Ionete, R. E. (2015).
701 Discrimination of vegetable oils using NMR spectroscopy and chemometrics. *Food Control,*
702 *48*, 84-90. https://doi.org/10.1016/j.foodcont.2014.04.046.
703 Priego Capote, F., Ruiz Jiménez, J., & Luque De Castro, M. D. (2007). Sequential (step-by-step)
704 detection, identification and quantitation of extra virgin olive oil adulteration by
705 chemometric treatment of chromatographic profiles. *Analytical and Bioanalytical*
706 *Chemistry, 388*(8), 1859-1865. https://doi.org/10.1007/s00216-007-1422-9.
707 Ranade, S., & Thiagarajan, P. (2015). A review on *Persea Americana* Mill. (Avocado) - Its fruit
708 and oil. *International Journal of PharmTech Research, 8*(6), 72-77.

Santos, P. M., Kock, F. V. C., Santos, M. S., Lobo, C. M. S., Carvalho, A. S., & Colnago, L. A. (2017). Non-invasive detection of adulterated olive oil in full bottles using time-domain NMR relaxometry. *Journal of the Brazilian Chemical Society, 28*(2), 385-390. https://doi.org/10.5935/0103-5053.20160188.

Vigli, G., Philippidis, A., Spyros, A., & Dais, P. (2003). Classification of edible oils by employing $^{31}P$ and $^{1}H$ NMR spectroscopy in combination with multivariate statistical analysis. A proposal for the detection of seed oil adulteration in virgin olive oils. *Journal of Agricultural and Food Chemistry, 51*(19), 5715-5722. https://doi.org/10.1021/jf030100z.

Wang, X., Wang, G., Hou, X., & Nie, S. (2020). A rapid screening approach for authentication of olive oil and classification of binary blends of olive oils using low-field nuclear magnetic resonance spectra and support vector machine. *Food Analytical Methods, 13*, 1894–1905. https://doi.org/10.1007/s12161-020-01799-z.

Yang, Y., Ferro, M. D., Cavaco, I., & Liang, Y. (2013). Detection and identification of extra virgin olive oil adulteration by GC-MS combined with chemometrics. *Journal of Agricultural and Food Chemistry, 61*(15), 3693-3702. https://doi.org/10.1021/jf4000538.

Zamora, R., Alba, V., & Hidalgo, F. J. (2001). Use of high-resolution $^{13}C$ nuclear magnetic resonance spectroscopy for the screening of virgin olive oils. *Journal of the American Oil Chemists' Society, 78*(1), 89-94. https://doi.org/10.1007/s11746-001-0225-z.

## Figure captions

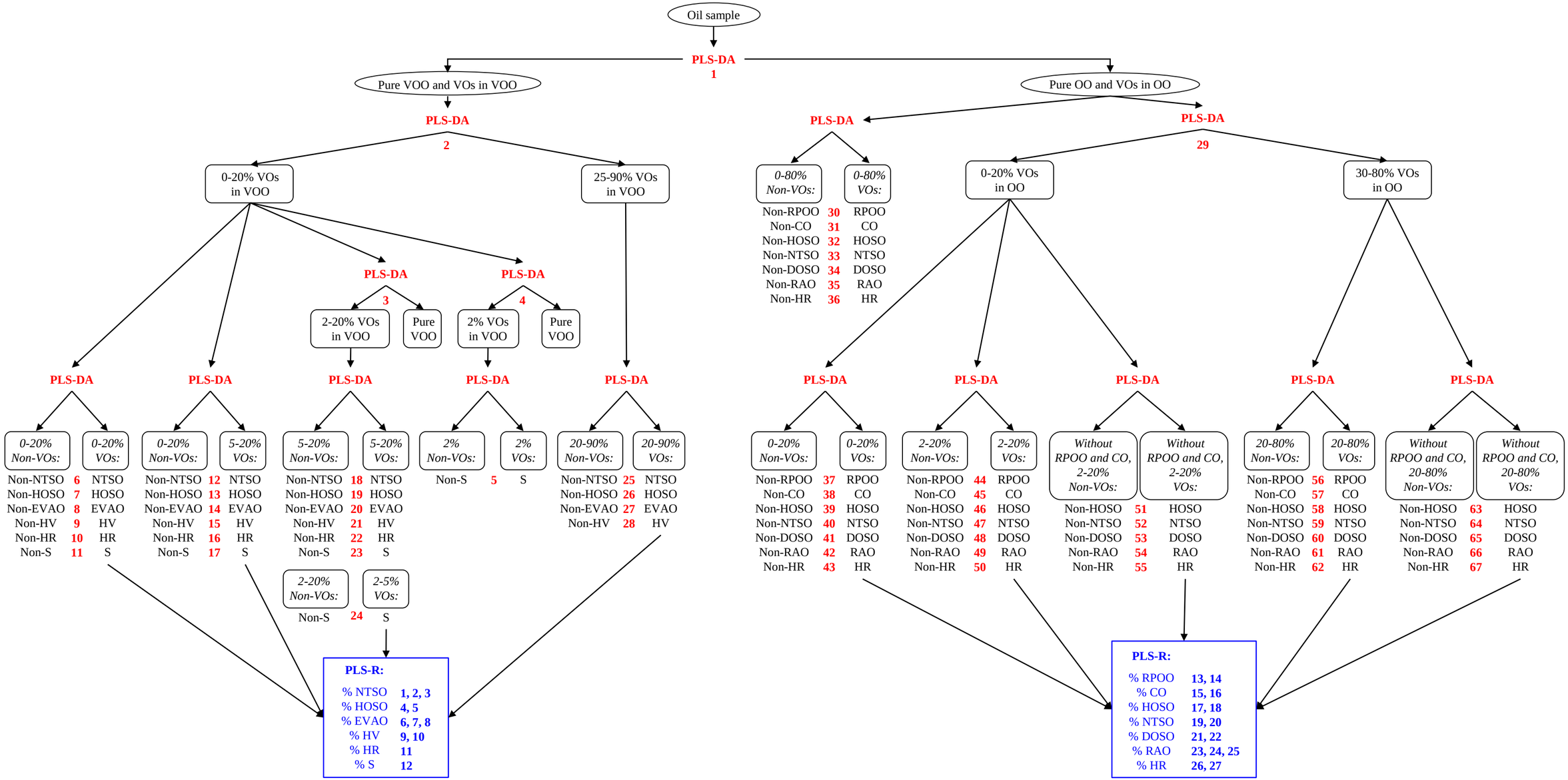
**Figure 1.** Decision tree constituted of PLS-DA classification and PLS-R regression models to determine the composition of binary mixtures of oils of the 'virgin olive oil' or 'olive oil' categories and other vegetable oils. Abbreviations: VOO, virgin olive oil; OO, olive oil; VO, vegetable oil; NTSO, refined conventional sunflower oil (normal type sunflower oil); HOSO, refined high oleic sunflower oil; DOSO, desterolized and deodorized high oleic sunflower oil; HR, refined hazelnut oil; HV, virgin hazelnut oil; S, refined soybean oil; EVAO, virgin avocado oil; RAO, refined avocado oil; RPOO, refined palm olein oil; CO, refined corn oil.

## Supplementary material

**Figure S1.** Decision tree constituted of PLS-DA classification and PLS-R regression models for a case-study: Discrimination between 'legal' (containing NTSO or HOSO) and 'illegal' (not containing NTSO or HOSO) blends, and determination of % NTSO or HOSO in binary mixtures with oils of the 'virgin olive oil' or 'olive oil' categories. Abbreviations: VOO, virgin olive oil; OO, olive oil; VO, vegetable oil; NTSO, refined conventional sunflower oil (normal type sunflower oil); HOSO, refined high oleic sunflower oil.

Figure 1

Click here to access/download;Figure(s);RMAS_Figure-1.pptx



# Tables

## Table 1

PLS-DA models to discriminate between pure and blended oils containing oils of the 'virgin olive oil' or 'olive oil' categories and vegetable oils, and binary mixtures with different proportions of vegetable oil in olive oil.[1]

| PLS-DA model | Data | PLS-comp | Boundary | Class[2] | Class code | n | p | %R | %P |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Pure & blend VOO/OO | 4 | 0.4079 | VOO | 0 | 838 | 0.70 | 97 | 97 |
| | | | | OO | 1 | 356 | 0.30 | 98 | 98 |
| 2 | Pure & blend VOO | 6 | 0.3283 | 0–20% VOs in VOO (low) | 0 | 704 | 0.84 | 98 | 98 |
| | | | | 25–90% VOs in VOO (high) | 1 | 132 | 0.16 | 97 | 97 |
| 3 | 0–20% VOs in VOO | 5 | 0.2230 | 2–20% VOs in VOO | 0 | 549 | 0.78 | 90 | 89 |
| | | | | Pure VOO | 1 | 155 | 0.22 | 86 | 86 |
| 4 | 0–2% VOs in VOO | 14 | 0.4264 | 2% VOs in VOO | 0 | 204 | 0.57 | 90 | 90 |
| | | | | Pure VOO | 1 | 155 | 0.43 | 93 | 92 |
| 5 | 2% VOs in VOO | 19 | 0.4265 | non-S | 0 | 159 | 0.78 | 99 | 97 |
| | | | | S | 1 | 45 | 0.22 | 100 | 100 |
| 29 | Pure & blend OO | 16 | 0.4388 | 0–20% VOs in OO (low) | 0 | 184 | 0.52 | 97 | 96 |
| | | | | 30–80% VOs in OO (high) | 1 | 171 | 0.48 | 95 | 94 |

[1] Abbreviations: n, number of samples; centered data; PLS-comp, number of PLS components; p, prior probability; %R, % of recognition ability; %P, % of prediction ability in cross-validation; VOO, virgin olive oil; OO, olive oil; VO, vegetable oil; NTSO, refined conventional sunflower oil (normal type sunflower oil); HOSO, refined high oleic sunflower oil; DOSO, desterolized and deodorized high oleic sunflower oil; HR, refined hazelnut oil; HV, virgin hazelnut oil; S, refined soybean oil; EVAO, virgin avocado oil; RAO, refined avocado oil; RPOO, refined palm olein oil; CO, refined corn oil.

[2] Samples contained in each class: VOO, pure VOOs and blends of VOO with VOs (NTSO, HOSO, EVAO, HV, HR or S); OO, pure OOs and blends of OO with VOs (RPOO, CO, HOSO, NTSO, DOSO, RAO or HR); 0–20% VOs in VOO, pure VOOs and blends of VOO with 2–20% VOs (NTSO, HOSO, EVAO, HV, HR or S); 25–90% VOs in VOO, blends of VOO with 25–90% VOs (NTSO, HOSO, EVAO, HV, HR or S); 2–20% VOs in VOO, blends of VOO with 2–20% VOs (NTSO, HOSO, EVAO, HV, HR or S); Pure VOO, pure VOOs; 2% VOs in VOO, blends of VOO with 2% VOs (NTSO, HOSO, EVAO, HV, HR or S); non-S, blends of VOO with 2% VOs (NTSO, HOSO, EVAO, HV or HR); S, blends of VOO with 2% S; 0–20% VOs in OO, pure OOs and blends of OO with 2–20% VOs (RPOO, CO, HOSO, NTSO, DOSO, RAO or HR); 30–80% VOs in OO, blends of OO with 30–80% VOs (RPOO, CO, HOSO, NTSO, DOSO, RAO or HR).

**Table 2**

PLS-DA models to detect the presence of a certain vegetable oil in a binary mixture of 2–20% vegetable oil in olive oil.[1]

| PLS-DA model | Data | PLS-comp | Boundary | Class[2,3,4] | Class code | n | p | %R | %P |
|---|---|---|---|---|---|---|---|---|---|
| 18 | 5–20% non-NTSO in VOO | 7 | 0.3029 | non-NTSO | 0 | 267 | 0.77 | 93 | 91 |
| | 5–20% NTSO in VOO | | | NTSO | 1 | 78 | 0.23 | 94 | 90 |
| 19 | 5–20% non-HOSO in VOO | 16 | 0.4039 | non-HOSO | 0 | 243 | 0.70 | 88 | 85 |
| | 5–20% HOSO in VOO | | | HOSO | 1 | 102 | 0.30 | 92 | 88 |
| 20 | 5–20% non-EVAO in VOO | 11 | 0.3002 | non-EVAO | 0 | 330 | 0.96 | 98 | 98 |
| | 5–20% EVAO in VOO | | | EVAO | 1 | 15 | 0.04 | 93 | 93 |
| 21 | 5–20% non-HV in VOO | 13 | 0.2335 | non-HV | 0 | 300 | 0.87 | 91 | 83 |
| | 5–20% HV in VOO | | | HV | 1 | 45 | 0.13 | 91 | 87 |
| 22 | 5–20% non-HR in VOO | 20 | 0.3291 | non-HR | 0 | 285 | 0.83 | 90 | 83 |
| | 5–20% HR in VOO | | | HR | 1 | 60 | 0.17 | 93 | 88 |
| 23 | 5–20% non-S in VOO | 7 | 0.3715 | non-S | 0 | 300 | 0.87 | 98 | 97 |
| | 5% S in VOO | | | S | 1 | 45 | 0.13 | 98 | 98 |
| 24 | 2–20% non-S in VOO | 13 | 0.4514 | non-S | 0 | 166 | 0.65 | 99 | 97 |
| | 2–5% S in VOO | | | S | 1 | 90 | 0.35 | 98 | 97 |
| 44 | 2–20% VOs in OO | 2 | 0.2604 | non-RPOO | 0 | 130 | 0.86 | 98 | 97 |
| | | | | RPOO | 1 | 21 | 0.14 | 95 | 95 |
| 45 | 2–20% VOs in OO | 7 | 0.3987 | non-CO | 0 | 132 | 0.87 | 96 | 96 |
| | | | | CO | 1 | 20 | 0.13 | 100 | 100 |
| 46 | 2–20% VOs in OO | 3 | 0.3359 | non-HOSO | 0 | 140 | 0.92 | 98 | 98 |
| | | | | HOSO | 1 | 12 | 0.08 | 100 | 100 |
| 47 | 2–20% VOs in OO | 12 | 0.3176 | non-NTSO | 0 | 114 | 0.75 | 96 | 89 |
| | | | | NTSO | 1 | 38 | 0.25 | 97 | 89 |
| 48 | 2–20% VOs in OO | 8 | 0.2189 | non-DOSO | 0 | 131 | 0.87 | 92 | 85 |
| | | | | DOSO | 1 | 20 | 0.13 | 95 | 95 |
| 49 | 2–20% VOs in OO | 6 | 0.2633 | non-RAO | 0 | 131 | 0.86 | 83 | 82 |
| | | | | RAO | 1 | 21 | 0.14 | 90 | 90 |
| 50 | 2–20% VOs in OO | 14 | 0.3408 | non-HR | 0 | 131 | 0.87 | 97 | 92 |
| | | | | HR | 1 | 19 | 0.13 | 100 | 95 |

[1] See abbreviations in Table 1.

[2] Samples contained in each class for PLS-DA models 18–23: non-NTSO, blends of VOO with 5–20% VOs (HOSO, EVAO, HV, HR or S); NTSO, blends of VOO with 5–20% NTSO; non-HOSO, blends of VOO with 5–20% VOs (NTSO, EVAO, HV, HR or S); HOSO, blends of VOO with 5–20% HOSO; non-EVAO, blends of VOO with 5–20% VOs (NTSO, HOSO, HV, HR or S); EVAO, blends of VOO with 5–20% EVAO; non-HV, blends of VOO with 5–20% VOs (NTSO, HOSO, EVAO, HR or S); HV, blends of VOO with 5–20%HV; non-HR, blends of VOO with 5–20% VOs (NTSO, HOSO, EVAO, HV or S); HR, blends of VOO with 5–10% HR; non-S, blends of VOO with 5–20% VOs (NTSO, HOSO, EVAO, HV or HR); S, blends of VOO with 5% S.

[3] Samples contained in each class for PLS-DA models 24: non-S, blends of VOO with 2-20% VOs (NTSO, HOSO, EVAO, HV or HR); S, blends of VOO with 2–5% S.

[4] Samples contained in each class for PLS-DA models 44–50: non-RPOO, blends of OO with 2–20% VOs (CO, HOSO, NTSO, DOSO, RAO or HR); RPOO, blends of OO with 2–20% RPOO; non-CO, blends of OO with 2–20% VOs (RPOO, HOSO, NTSO, DOSO, RAO or HR); CO, blends of OO with 2–20% CO; non-HOSO, blends of OO with 2–20% VOs (RPOO, CO, NTSO, DOSO, RAO or HR); HOSO, blends of OO with 2–20% HOSO; non-NTSO, blends of OO with 2–20% VOs (RPOO, CO, HOSO, DOSO, RAO or HR); NTSO, blends of OO with 2–20% NTSO; non-DOSO, blends of OO with 2–20% VOs (RPOO, CO, HOSO, NTSO, RAO or HR); DOSO, blends of OO with 2–20% DOSO; non-RAO, blends of OO with 2–20% VOs (RPOO, CO, HOSO, NTSO, DOSO or HR); RAO, blends of OO with 2–20% RAO; non-HR, blends of OO with 2–20% VOs (RPOO, CO, HOSO, NTSO, DOSO or RAO); HR, blends of OO with 2–20% HR.

**Table 3**

PLS-DA models to detect the presence of a certain vegetable oil in a binary mixture of 20–90% vegetable oil in olive oil.[1]

| PLS-DA model | Data | PLS-comp | Boundary | Class[2,3] | Class code | n | p | %R | %P |
|---|---|---|---|---|---|---|---|---|---|
| 25 | 20–80% non-NTSO in VOO | 4 | 0.4955 | non-NTSO | 0 | 73 | 0.47 | 100 | 100 |
| | 20–90% NTSO in VOO | | | NTSO | 1 | 83 | 0.53 | 100 | 100 |
| 26 | 20–90% non-HOSO in VOO | 4 | 0.4120 | non-HOSO | 0 | 130 | 0.83 | 100 | 100 |
| | 20–80% HOSO in VOO | | | HOSO | 1 | 26 | 0.17 | 92 | 92 |
| 27 | 20–90% non-EVAO in VOO | 4 | 0.3985 | non-EVAO | 0 | 131 | 0.84 | 100 | 99 |
| | 20–80% EVAO in VOO | | | EVAO | 1 | 25 | 0.16 | 100 | 100 |
| 28 | 20–90% non-HV in VOO | 3 | 0.3563 | non-HV | 0 | 134 | 0.86 | 100 | 100 |
| | 20–80% HV in VOO | | | HV | 1 | 22 | 0.14 | 100 | 100 |
| 56 | 20–80% VOs in OO | 1 | 0.3445 | non-RPOO | 0 | 185 | 0.88 | 100 | 100 |
| | | | | RPOO | 1 | 25 | 0.12 | 100 | 100 |
| 57 | 20–80% VOs in OO | 7 | 0.4410 | non-CO | 0 | 178 | 0.85 | 100 | 100 |
| | | | | CO | 1 | 31 | 0.15 | 100 | 100 |
| 58 | 20–80% VOs in OO | 5 | 0.4063 | non-HOSO | 0 | 182 | 0.87 | 99 | 99 |
| | | | | HOSO | 1 | 28 | 0.13 | 86 | 86 |
| 59 | 20–80% VOs in OO | 6 | 0.3650 | non-NTSO | 0 | 151 | 0.72 | 100 | 99 |
| | | | | NTSO | 1 | 59 | 0.28 | 93 | 92 |
| 60 | 20–80% VOs in OO | 4 | 0.3127 | non-DOSO | 0 | 188 | 0.90 | 100 | 99 |
| | | | | DOSO | 1 | 20 | 0.10 | 100 | 100 |
| 61 | 20–80% VOs in OO | 5 | 0.3195 | non-RAO | 0 | 187 | 0.89 | 95 | 94 |
| | | | | RAO | 1 | 23 | 0.11 | 91 | 91 |
| 62 | 20–80% VOs in OO | 9 | 0.3083 | non-HR | 0 | 187 | 0.91 | 99 | 98 |
| | | | | HR | 1 | 19 | 0.09 | 100 | 100 |

[1] See abbreviations in Table 1.

[2] Samples contained in each class for PLS-DA models 25–28: non-NTSO, blends of VOO with 20–80% VOs (HOSO, EVAO or HV); NTSO, blends of VOO with 20–90% NTSO; non-HOSO, blends of VOO with 20–90% VOs (NTSO, EVAO or HV); HOSO, blends of VOO with 20–80% HOSO; non-EVAO, blends of VOO with 20–90% VOs (NTSO, HOSO or HV); EVAO, blends of VOO with 20–80% EVAO; non-HV, blends of VOO with 20–90% VOs (NTSO, HOSO or EVAO); HV, blends of VOO with 20–80% HV.

[3] Samples contained in each class for PLS-DA models 56–62: non-RPOO, blends of OO with 20–80% VOs (CO, HOSO, NTSO, DOSO, RAO or HR); RPOO, blends of OO with 20–80% RPOO; non-CO, blends of OO with 20–80% VOs (RPOO, HOSO, NTSO, DOSO, RAO or HR); CO, blends of OO with 20–80% CO; non-HOSO, blends of OO with 20–80% VOs (RPOO, CO, NTSO, DOSO, RAO or HR); HOSO, blends of OO with 20–80% HOSO; non-NTSO, blends of OO with 20–80% VOs (RPOO, CO, HOSO, DOSO, RAO or HR); NTSO, blends of OO with 20–80% NTSO; non-DOSO, blends of OO with 20–80% VOs (RPOO, CO, HOSO, NTSO, RAO or HR); DOSO, blends of OO with 20–80% DOSO; non-RAO, blends of OO with 20–80% VOs (RPOO, CO, HOSO, NTSO, DOSO or HR); RAO, blends of OO with 20–80% RAO; non-HR, blends of OO with 20–80% VOs (RPOO, CO, HOSO, NTSO, DOSO or RAO); HR, blends of OO with 20–80% HR.

**Table 4**

PLS-R models to determine the percentage of a certain vegetable oil in a binary mixture with olive oil.[1]

| PLS-R model | Data[2] | n | PLS-comp | R-cal | R-val | $R^2$-val | RMSEP (% VO) |
|---|---|---|---|---|---|---|---|
| 1 | 2–10% NTSO in VOO[3] | 113 | 6 | 0.86 | 0.83 | 0.68 | 1.2 |
| 2 | 10–20% NTSO in VOO[3] | 24 | 6 | 0.9995 | 0.9946 | 0.989 | 0.49 |
| 3 | 20–90% NTSO in VOO[3] | 76 | 1 | 0.9990 | 0.9989 | 0.998 | 0.96 |
| 4 | 2–20% HOSO in VOO[3] | 100 | 7 | 0.75 | 0.71 | 0.50 | 3.4 |
| 5 | 20–80% HOSO in VOO[4] | 21 | 5 | 0.998 | 0.994 | 0.987 | 1.9 |
| 6 | 2–20% EVAO in VOO[4] | 20 | 6 | 0.998 | 0.988 | 0.98 | 1.0 |
| 7 | 20–45% EVAO in VOO[4] | 14 | 3 | 0.995 | 0.987 | 0.97 | 1.7 |
| 8 | 45–80% EVAO in VOO[4] | 12 | 3 | 0.998 | 0.996 | 0.992 | 1.3 |
| 9 | 10–30% HV in VOO[4] | 25 | 7 | 0.995 | 0.986 | 0.97 | 1.3 |
| 10 | 30–80% HV in VOO[4] | 16 | 1 | 0.994 | 0.993 | 0.986 | 2.3 |
| 11 | 2–10% HR in VOO[3] | 84 | 3 | 0.58 | 0.55 | 0.30 | 2.8 |
| 12 | 2–5% S in VOO[3] | 86 | 9 | 0.87 | 0.78 | 0.61 | 0.95 |
| 13 | 2–20% RPOO in OO[4] | 20 | 4 | 0.9997 | 0.9993 | 0.9986 | 0.25 |
| 14 | 20–80% RPOO in OO[3] | 25 | 1 | 0.9993 | 0.9992 | 0.998 | 0.80 |
| 15 | 2–10% CO in OO[4] | 12 | 1 | 0.997 | 0.996 | 0.992 | 0.32 |
| 16 | 10–80% CO in OO[3] | 32 | 1 | 0.99992 | 0.99990 | 0.9998 | 0.32 |
| 17 | 2–20% HOSO in OO[4] | 10 | 2 | 0.994 | 0.983 | 0.97 | 1.0 |
| 18 | 10–80% HOSO in OO[3] | 25 | 3 | 0.9994 | 0.9992 | 0.998 | 0.80 |
| 19 | 2–20% NTSO in OO[3] | 34 | 4 | 0.9989 | 0.9978 | 0.996 | 0.45 |
| 20 | 20–80% NTSO in OO[3] | 54 | 1 | 0.997 | 0.994 | 0.989 | 1.4 |
| 21 | 2–20% DOSO in OO[4] | 19 | 6 | 0.998 | 0.994 | 0.987 | 0.78 |
| 22 | 20–80% DOSO in OO[4] | 18 | 2 | 0.997 | 0.996 | 0.991 | 2.0 |
| 23 | 2–10% RAO in OO[4] | 11 | 5 | 0.997 | 0.963 | 0.93 | 0.76 |
| 24 | 2–20% RAO in OO[4] | 17 | 9 | 0.9994 | 0.9812 | 0.963 | 1.3 |
| 25 | 20–80% RAO in OO[4] | 17 | 4 | 0.9991 | 0.9974 | 0.995 | 1.5 |
| 26 | 2–20% HR in OO[4] | 14 | 3 | 0.9988 | 0.9977 | 0.995 | 0.49 |
| 27 | 20–80% HR in OO[3] | 21 | 3 | 0.9997 | 0.9995 | 0.9990 | 0.64 |

[1] Abbreviations: n, number of samples; centered data; PLS-comp, number of PLS components; R-cal, correlation coefficient in calibration; R-val, correlation coefficient in validation; $R^2$-val, coefficient of determination in validation; RMSEP, root mean square error in the prediction (% VO).

[2] Samples used to build each model.

[3] 3-fold cross-validation.

[4] Leave-one-out cross-validation.

**Table 5**

PLS-DA models to discriminate between 'legal' and 'illegal' blends of olive oil and vegetable oils, 'legal' blends of VOO or OO with NTSO and HOSO, VOO blends with 2% S and 5% S, VOO blends with 2–5% HR and 10% HR, OO blends with DOSO and HR, OO blends with RAO and HR, OO blends with RAO and DOSO, and OO blends of with DOSO and HOSO.[1]

| PLS-DA model | Data | PLS-comp | Boundary | Class[2,3,4,5] | Class code | n | p | %R | %P |
|---|---|---|---|---|---|---|---|---|---|
| 68 | 2–90% VOs in VOO | 10 | 0.5290 | 'Illegal' blend | 0 | 302 | 0.44 | 78 | 77 |
| | | | | 'Legal' blend | 1 | 381 | 0.56 | 81 | 77 |
| 69 | 2–90% NTSO in VOO | 9 | 0.5543 | NTSO | 0 | 207 | 0.54 | 85 | 83 |
| | 2–80% HOSO in VOO | | | HOSO | 1 | 174 | 0.46 | 88 | 85 |
| 70 | 2–80% VOs in OO | 13 | 0.3960 | 'Illegal' blend | 0 | 199 | 0.61 | 99 | 98 |
| | | | | 'Legal' blend | 1 | 125 | 0.39 | 87 | 86 |
| 71 | 2–80% NTSO in OO | 5 | 0.3979 | NTSO | 0 | 88 | 0.70 | 98 | 97 |
| | 2–80% HOSO in OO | | | HOSO | 1 | 37 | 0.30 | 97 | 97 |
| 72 | 2–5% S in VOO | 9 | 0.4643 | 2% S | 0 | 44 | 0.50 | 95 | 93 |
| | | | | 5% S | 1 | 44 | 0.50 | 93 | 93 |
| 73 | 2–10% HR in VOO | 6 | 0.4429 | 2–5% HR | 0 | 59 | 0.66 | 83 | 80 |
| | | | | 10% HR | 1 | 30 | 0.34 | 80 | 77 |
| 74 | 2–80% DOSO in OO | 3 | 0.4805 | DOSO | 0 | 37 | 0.50 | 86 | 84 |
| | 2–80% HR in OO | | | HR | 1 | 37 | 0.50 | 97 | 95 |
| 75 | 2–80% RAO in OO | 3 | 0.5011 | RAO | 0 | 38 | 0.51 | 79 | 82 |
| | 2–80% HR in OO | | | HR | 1 | 37 | 0.49 | 86 | 84 |
| 76 | 2–80% RAO in OO | 6 | 0.4723 | RAO | 0 | 38 | 0.51 | 95 | 95 |
| | 2–80% DOSO in OO | | | DOSO | 1 | 37 | 0.49 | 100 | 97 |
| 77 | 2–80% DOSO in OO | 3 | 0.4280 | DOSO | 0 | 37 | 0.50 | 95 | 95 |
| | 2–80% HOSO in OO | | | HOSO | 1 | 37 | 0.50 | 100 | 100 |

[1] See abbreviations in Table 1.

[2] Samples contained in each class for PLS-DA models 68–69: 'Illegal' blend, blends of VOO with 2–80% VOs (EVAO, HV, HR or S); 'Legal' blend, blends of VOO with 2-90% VOs (NTSO or HOSO); NTSO, blends of VOO with 2–90% NTSO; HOSO, blends of VOO with 2-80% HOSO.

[3] Samples contained in each class for PLS-DA models 70–71: 'Illegal' blends, blends of OO with 2–80% VOs (RPOO, CO, DOSO, RAO or HR); 'Legal' blends, blends of OO with 2–80% VOs (HOSO or NTSO); NTSO, blends of OO with 2–80% NTSO; HOSO, blends of OO with 2–80% HOSO.

[4] Samples contained in each class PLS-DA models 72–73: 2% S in VOO, blends of VOO with 2% S; 5% S in VOO, blends of VOO with 5% S; 2–5% HR in VOO, blends of VOO with 2–5% HR; 10% HR in VOO, blends of VOO with 10% HR.

[5] Samples contained in each class PLS-DA models 74–77: DOSO, blends of OO with 2–80% DOSO; HR, blends of OO with 2–80% HR; RAO, blends of OO with 2–80% RAO; HOSO, blends of OO with 2–80% HOSO.

Supplementary Material - Figure
Click here to access/download;Supplementary Material;RMAS_Figures-Supplementary-material.pdf

# Stepwise strategy based on $^{1}$H-NMR fingerprinting in combination with chemometrics to determine the content of vegetable oils in olive oil mixtures

R. M. Alonso-Salces[1,*], L. A. Berrueta[2], B. Quintanilla-Casas[3], S. Vichi[3], A. Tres[3], M. I. Collado[4], C. Asensio-Regalado[2], G. E. Viacava[5], A. A. Poliero[6], E. Valli[7], A. Bendini[7], T. Gallina Toschi[7], J. M. Martínez-Rivas[8], W. Moreda[9], B. Gallo[2]

## Supplementary material: Figures

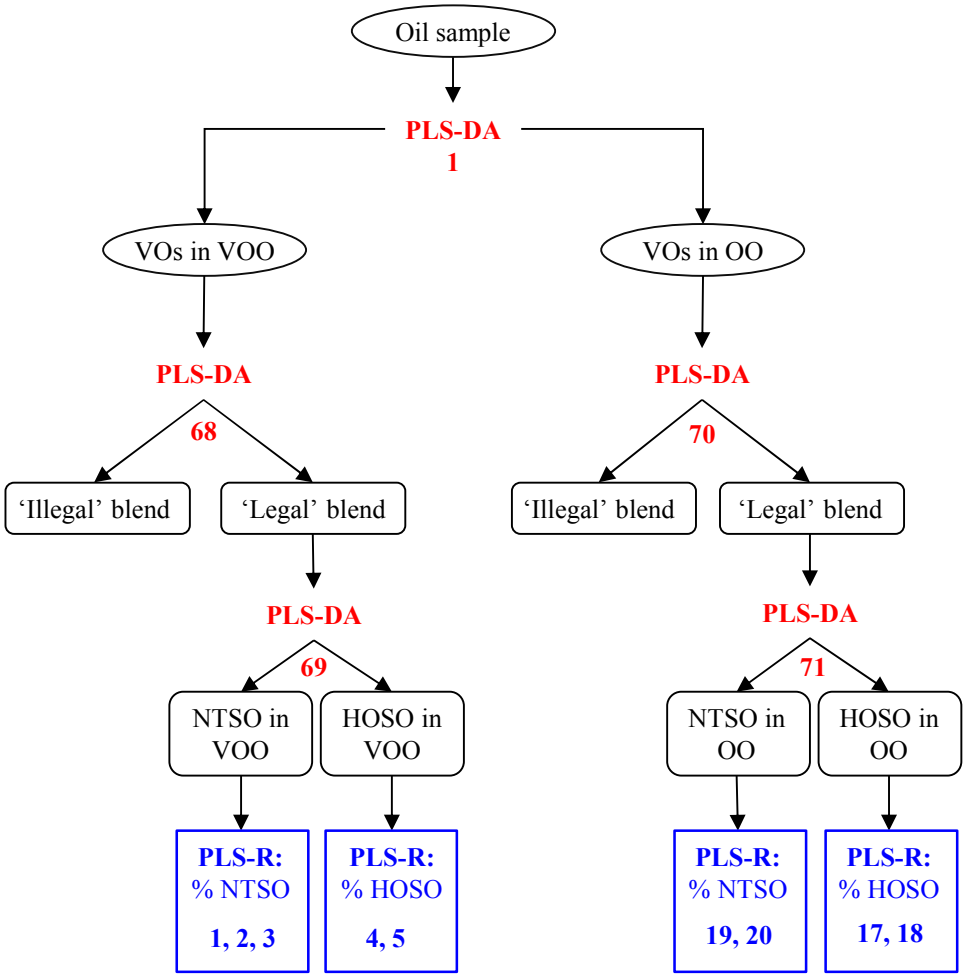



**Figure S1.** Decision tree constituted of PLS-DA classification and PLS-R regression models for a case-study: Discrimination between 'legal' (containing NTSO or HOSO) and 'illegal' (not containing NTSO or HOSO) blends, and determination of % NTSO or HOSO in binary mixtures with oils of the 'virgin olive oil' or 'olive oil' categories. Abbreviations: VOO, virgin olive oil; OO, olive oil; VO, vegetable oil; NTSO, refined conventional sunflower oil (normal type sunflower oil); HOSO, refined high oleic sunflower oil.

Supplementary Material - Tables
Click here to access/download;Supplementary Material;RMAS_Tables-Supplementary-material.pdf

# Stepwise strategy based on $^{1}$H-NMR fingerprinting in combination with chemometrics to determine the content of vegetable oils in olive oil mixtures

R. M. Alonso-Salces[1,*], L. A. Berrueta[2], B. Quintanilla-Casas[3], S. Vichi[3], A. Tres[3], M. I. Collado[4], C. Asensio-Regalado[2], G. E. Viacava[5], A. A. Poliero[6], E. Valli[7], A. Bendini[7], T. Gallina Toschi[7], J. M. Martínez-Rivas[8], W. Moreda[9], B. Gallo[2]

## Supplementary material: Tables

### Table S1

Chemical shift assignments of the $^{1}$H-NMR signals of the main components in olive oil.

| # | Chemical shift (ppm) | Multiplicity[a] | Functional group | Attribution |
|---|---|---|---|---|
| 1 | 0.318 | d | -C***H***$_2$- (cyclopropanic ring) | cycloartenol |
| 2 | 0.527 | s | -C***H***$_2$- | alcohol, sterol |
| 3 | 0.543 | d | -C***H***$_2$- (cyclopropanic ring) | cycloartenol |
| 4 | 0.669 | s | -C***H***$_3$ (C18-steroid group) | β-sitosterol |
| 5 | 0.687 | s | -C***H***$_3$ (C18-steroid group) | stigmasterol |
| 6 | 0.740 | t | -C***H***$_3$ ($^{13}$C satellite of signal at 0.87 ppm, acyl group) | |
| 7 | 0.80-1.04 | t | -C***H***$_3$ (acyl group) | |
| 7a | 0.83 | t | -C***H***$_3$ (acyl group) | saturated |
| 7b | 0.866 | t | -C***H***$_3$ (acyl group) | oleic (or ω-9) |
| 7c | 0.89 | t | -C***H***$_3$ (acyl group) | linoleic (or ω-6) |
| 7d | 0.960 | t | -C***H***$_3$ (acyl group) | linolenic (or ω-3) |
| 8 | 0.987 | t | -C***H***$_3$ ($^{13}$C satellite of signal at 0.87 ppm, acyl group) | |
| 9 | 1.19-1.44 | | -(C***H***$_2$)$_n$- (acyl group) | |
| 9a | 1.243 | | -(C***H***$_2$)$_n$- (acyl group) | saturated |
| 9b | 1.256 | | -(C***H***$_2$)$_n$- (acyl group) | oleic (or ω-9) |
| 9c | 1.288 | | -(C***H***$_2$)$_n$- (acyl group) | linoleic (or ω-6) and linolenic (or ω-3) |
| 10 | 1.51-1.65 | | -OCO-CH$_2$-C***H***$_2$- (acyl group) | |
| 10a | 1.57 | | -OCO-CH$_2$-C***H***$_2$- (acyl group) | saturated |
| 10b | 1.58 | | -OCO-CH$_2$-C***H***$_2$- (acyl group) | oleic (or ω-9) |
| 10c | 1.59 | | -OCO-CH$_2$-C***H***$_2$- (acyl group) | linoleic (or ω-6) and linolenic (or ω-3) |
| 11 | 1.662 | s | -C***H***$_3$ | squalene |
| 12 | 1.96-2.07 | | -C***H***$_2$-CH=CH- (acyl group) | |
| 12a | 1.97 | | -C***H***$_2$-CH=CH- (acyl group) | oleic (or ω-9) |
| 12b | 2.01-2.03 | | -C***H***$_2$-CH=CH- (acyl group) | linoleic (or ω-6) and linolenic (or ω-3) |
| 12c | 2.05-2.07 | | -C***H***$_2$-CH=CH- (acyl group) | linolenic (or ω-3) |
| 13 | 2.22-2.32 | m | -OCO-C***H***$_2$- (acyl group) | |
| 13a | 2.24 | m | -OCO-C***H***$_2$- (acyl group) | saturated |
| 13b | 2.25 | m | -OCO-C***H***$_2$- (acyl group) | oleic (or ω-9) |
| 13c | 2.27 | m | -OCO-C***H***$_2$- (acyl group) | linoleic (or ω-6) |
| 13d | 2.31 | m | -OCO-C***H***$_2$- (acyl group) | linolenic (or ω-3) |
| 14 | 2.40-2.45 | m | -OCO-C***H***$_2$- ($^{13}$C satellite of signal at 2.26-2.32 ppm, acyl group) | |

| # | Chemical shift (ppm) | Multiplicity[a] | Functional group | Attribution |
|---|---|---|---|---|
| 15 | 2.72-2.82 | | =CH-C***H***$_2$-CH= (acyl group) | |
| 15a | 2.754 | t | =CH-C***H***$_2$-CH= (acyl group) | linoleic (or ω-6) |
| 15b | 2.789 | t | =CH-C***H***$_2$-CH= (acyl group) | linolenic (or ω-3) |
| 16 | 3.69-3.73 | d | -C***H***$_2$OH (glyceryl group) | *sn*-1,2-diacylglycerides |
| 17 | 4.05-4.09 | q | >C***H***-OH (glyceryl group) | *sn*-1,3-diacylglycerides |
| 18 | 4.09-4.32 | | -C***H***$_2$OCOR (glyceryl group) | triacylglycerides |
| 19 | 4.571 | d | | terpene |
| 20 | 4.648 | s | | terpene |
| 21 | 4.699 | s | | terpene |
| 22 | 5.05-5.15 | m | >C***H***OCOR (glyceryl group) | *sn*-1,2-diacylglycerides |
| 23 | 5.22-5.28 | m | >C***H***OCOR (glyceryl group) | triacylglycerides |
| 24 | 5.28-5.38 | m | -C***H***=C***H***- (acyl group) | |
| 25 | 5.52-5.43 | m | -C***H***=C***H***- ($^{13}$C satellite of signal at 5.28-5.38 ppm, acyl group) | |
| 26 | 5.72-5.76 | dt | =C***H***- (phenolic ring) | phenolic compounds |
| 27 | 5.986 | | =C***H***- (phenolic ring) | phenolic compounds |
| 28 | 6.551 | dt | =C***H***- (phenolic ring) | phenolic compounds |
| 29 | 6.607 | dd | =C***H***- (C8'; phenolic ring) | dialdehyde of oleuropein lacking a carboxymethyl group aldehydic form of oleuropein |
| 30 | 6.79-6.73 | d | =C***H***- (C5', C7'; phenolic ring) | dialdehyde of secoiridoids (oleuropein, ligstroside) lacking a carboxymethyl group aldehydic form of secoiridoid (oleuropein, ligstroside) |
| 31 | 7.05-7.00 | dt | =C***H***- (C4', C8'; phenolic ring) | dialdehyde of ligstroside lacking a carboxymethyl group aldehydic form of ligstroside |
| 32 | 7.562 | s | =C***H***-O- (C3) | aldehydic form of secoiridoid (oleuropein, ligstroside) |
| 33 | 8.14-8.06 | | >C(O***H***)OR | volatile compounds |
| 34 | 9.215 | d | -C***H***O (C1) | dialdehyde of secoiridoids (oleuropein, ligstroside) lacking a carboxymethyl group |
| 35 | 9.51 | d | -C***H***O | *E*-2-alkenals (*E*-2-hexenal) |
| 36 | 9.626 | dd | -C***H***O (C3) | dialdehyde of secoiridoids (oleuropein, ligstroside) lacking a carboxymethyl group |
| | | dd | -C***H***O (C1) | aldehydic form of secoiridoids (oleuropein, ligstroside) |

**Table S2**

PLS-DA models to detect the presence of a certain vegetable oil in a binary mixture of 2–20% vegetable oil in virgin olive oil.[1]

| PLS-DA model | Data | PLS-comp | Boundary | Class[2] | Class code | n | p | %R | %P | %P-EV |
|---|---|---|---|---|---|---|---|---|---|---|
| 6 | 0–20% non-NTSO in VOO | 12 | 0.3623 | non-NTSO | 0 | 238 | 0.64 | 86 | 85 | 79 |
| | 2–20% NTSO in VOO | | | NTSO | 1 | 132 | 0.36 | 90 | 86 | - |
| 7 | 0–20% non-HOSO in VOO | 14 | 0.4713 | non-HOSO | 0 | 245 | 0.62 | 83 | 79 | 82 |
| | 2–20% HOSO in VOO | | | HOSO | 1 | 152 | 0.38 | 83 | 79 | - |
| 8 | 0–20% non-EVAO in VOO | 6 | 0.3791 | non-EVAO | 0 | 81 | 0.80 | 94 | 93 | 97 |
| | 2–20% EVAO in VOO | | | EVAO | 1 | 20 | 0.20 | 90 | 90 | - |
| 9 | 0–20% non-HV in VOO | 6 | 0.3815 | non-HV | 0 | 137 | 0.68 | 78 | 75 | 73 |
| | 2–20% HV in VOO | | | HV | 1 | 65 | 0.32 | 82 | 75 | - |
| 10 | 0–20% non-HR in VOO | 5 | 0.4011 | non-HR | 0 | 195 | 0.68 | 77 | 75 | 58 |
| | 2–20% HR in VOO | | | HR | 1 | 90 | 0.32 | 78 | 76 | - |
| 11 | 0–20% non-S in VOO | 11 | 0.4248 | non-S | 0 | 208 | 0.70 | 98 | 96 | 95 |
| | 2–20% S in VOO | | | S | 1 | 90 | 0.30 | 97 | 96 | - |

[1] Abbreviations: n, number of samples; centered data; PLS-comp, number of PLS components; p, prior probability; %R, % of recognition ability; %P, % of prediction ability in cross-validation; %P-EV, % of prediction ability in external validation; VOO, virgin olive oil; VO, vegetable oil; NTSO, refined conventional sunflower oil (normal type sunflower oil); HOSO, refined high oleic sunflower oil; HR, refined hazelnut oil; HV, virgin hazelnut oil; S, refined soybean oil; EVAO, virgin avocado oil.

[2] Samples contained in each class: non-NTSO, pure VOOs and blends of VOO with 2−20% VOs (HOSO, EVAO, HV, HR or S); NTSO, blends of VOO with 2−20% NTSO; non-HOSO, pure VOOs and blends of VOO with 2−20% VOs (NTSO, EVAO, HV, HR or S); HOSO, blends of VOO with 2−20% HOSO; non-EVAO, pure VOOs and blends of VOO with 2−20% VOs (NTSO, HOSO, HV, HR or S); EVAO, blends of VOO with 2−20% EVAO; non-HV, pure VOOs and blends of VOO with 2−20% VOs (NTSO, HOSO, EVAO, HR or S); HV, blends of VOO with 2−20% HV; non-HR, pure VOOs and blends of VOO with 2−20% VOs (NTSO, HOSO, EVAO, HV or S); HR, blends of VOO with 2−10% HR; non-S, pure VOOs and blends of VOO with 2−20% VOs (NTSO, HOSO, EVAO, HV or HR); S, blends of VOO with 2−5% S.

**Table S3**

PLS-DA models to detect the presence of a certain vegetable oil in a binary mixture of 5–20% vegetable oil in virgin olive oil.[1]

| PLS-DA model | Data | PLS-comp | Boundary | Class[2] | Class code | n | p | %R | %P | %P-EV |
|---|---|---|---|---|---|---|---|---|---|---|
| 12 | 0–20% non-NTSO in VOO | 11 | 0.3508 | non-NTSO | 0 | 238 | 0.73 | 96 | 95 | 92 |
| | 5–20% NTSO in VOO | | | NTSO | 1 | 87 | 0.27 | 94 | 90 | - |
| 13 | 0–20% non-HOSO in VOO | 17 | 0.4098 | non-HOSO | 0 | 245 | 0.71 | 87 | 85 | 85 |
| | 5–20% HOSO in VOO | | | HOSO | 1 | 102 | 0.29 | 90 | 86 | - |
| 14 | 0–20% non-EVAO in VOO | 10 | 0.3805 | non-EVAO | 0 | 80 | 0.84 | 94 | 93 | 97 |
| | 5–20% EVAO in VOO | | | EVAO | 1 | 15 | 0.16 | 100 | 93 | - |
| 15 | 0–20% non-HV in VOO | 10 | 0.3675 | non-HV | 0 | 137 | 0.75 | 85 | 82 | 81 |
| | 5–20% HV in VOO | | | HV | 1 | 45 | 0.25 | 80 | 78 | - |
| 16 | 0–20% non-HR in VOO | 14 | 0.3808 | non-HR | 0 | 195 | 0.76 | 85 | 79 | 72 |
| | 5–20% HR in VOO | | | HR | 1 | 60 | 0.24 | 85 | 85 | - |
| 17 | 0–20% non-S in VOO | 7 | 0.4156 | non-S | 0 | 208 | 0.82 | 98 | 98 | 97 |
| | 5–20% S in VOO | | | S | 1 | 45 | 0.18 | 98 | 98 | - |

[1] See abbreviations in Table S2.

[2] Samples contained in each class: non-NTSO, pure VOOs and blends of VOO with 2−20% VOs (HOSO, EVAO, HV, HR or S); NTSO, blends of VOO with 5−20% NTSO; non-HOSO, pure VOOs and blends of VOO with 2−20% VOs (NTSO, EVAO, HV, HR or S); HOSO, blends of VOO with 5−20% HOSO; non-EVAO, pure VOOs and blends of VOO with 2−20% VOs (NTSO, HOSO, HV, HR or S); EVAO, blends of VOO with 5−20% EVAO; non-HV, pure VOOs and blends of VOO with 2−20% VOs (NTSO, HOSO, EVAO, HR or S); HV, blends of VOO with 5−20%HV; non-HR, pure VOOs and blends of VOO with 2−20% VOs (NTSO, HOSO, EVAO, HV or S); HR, blends of VOO with 5−10% HR; non-S, pure VOOs and blends of VOO with 2−20% VOs (NTSO, HOSO, EVAO, HV or HR); S, blends of VOO with 5% S.

**Table S4**

PLS-DA models to detect the presence of a certain vegetable oil in a binary mixture of 2–80% vegetable oil in olive oil.[1]

| PLS-DA model | Data | PLS-comp | Boundary | Class[2] | Class code | n | p | %R | %P |
|---|---|---|---|---|---|---|---|---|---|
| 30 | 0–80% VOs in OO | 2 | 0.1815 | non-RPOO | 0 | 315 | 0.88 | 100 | 100 |
| | | | | RPOO | 1 | 41 | 0.12 | 95 | 95 |
| 31 | 0–80% VOs in OO | 7 | 0.3545 | non-CO | 0 | 310 | 0.87 | 96 | 95 |
| | | | | CO | 1 | 46 | 0.13 | 100 | 100 |
| 32 | 0–80% VOs in OO | 7 | 0.3662 | non-HOSO | 0 | 319 | 0.90 | 98 | 97 |
| | | | | HOSO | 1 | 37 | 0.10 | 95 | 95 |
| 33 | 0–80% VOs in OO | 12 | 0.2809 | non-NTSO | 0 | 268 | 0.75 | 98 | 97 |
| | | | | NTSO | 1 | 88 | 0.25 | 85 | 85 |
| 34 | 0–80% VOs in OO | 5 | 0.1652 | non-DOSO | 0 | 319 | 0.90 | 91 | 91 |
| | | | | DOSO | 1 | 37 | 0.10 | 84 | 84 |
| 35 | 0–80% VOs in OO | 11 | 0.2354 | non-RAO | 0 | 318 | 0.89 | 96 | 92 |
| | | | | RAO | 1 | 38 | 0.11 | 95 | 87 |
| 36 | 0–80% VOs in OO | 15 | 0.2270 | non-HR | 0 | 319 | 0.90 | 93 | 89 |
| | | | | HR | 1 | 37 | 0.10 | 100 | 97 |

[1] Abbreviations: See abbreviations in Table S2; OO, olive oil; DOSO, desterolized and deodorized high oleic sunflower oil; RAO, refined avocado oil; RPOO, refined palm olein oil; CO, refined corn oil.

[2] Samples contained in each class: non-RPOO, pure OOs and blends of OO with 2−80% VOs (CO, HOSO, NTSO, DOSO, RAO or HR); RPOO, blends of OO with 2−80% RPOO; non-CO, pure OOs and blends of OO with 2−80% VOs (RPOO, HOSO, NTSO, DOSO, RAO or HR); CO, blends of OO with 2−80% CO; non-HOSO, pure OOs and blends of OO with 2−80% VOs (RPOO, CO, NTSO, DOSO, RAO or HR); HOSO, blends of OO with 2−80% HOSO; non-NTSO, pure OOs and blends of OO with 2−80% VOs (RPOO, CO, HOSO, DOSO, RAO or HR); NTSO, blends of OO with 2−80% NTSO; non-DOSO, pure OOs and blends of OO with 2−80% VOs (RPOO, CO, HOSO, NTSO, RAO or HR); DOSO, blends of OO with 2−80% DOSO; non-RAO, pure OOs and blends of OO with 2−80% VOs (RPOO, CO, HOSO, NTSO, DOSO or HR); RAO, blends of OO with 2−80% RAO; non-HR, pure OOs and blends of OO with 2−80% VOs (RPOO, CO, HOSO, NTSO, DOSO or RAO); HR, blends of OO with 2−80% HR.

**Table S5**

PLS-DA models to detect the presence of a certain vegetable oil in a binary mixture of 2–20% vegetable oil in olive oil.[1]

| PLS-DA model | Data | PLS-comp | Boundary | Class[2] | Class code | n | p | %R | %P |
|---|---|---|---|---|---|---|---|---|---|
| 37 | 0–20% VOs in OO | 2 | 0.2399 | non-RPOO | 0 | 162 | 0.89 | 98 | 98 |
| | | | | RPOO | 1 | 21 | 0.11 | 95 | 95 |
| 38 | 0–20% VOs in OO | 12 | 0.3522 | non-CO | 0 | 164 | 0.89 | 97 | 95 |
| | | | | CO | 1 | 20 | 0.11 | 100 | 100 |
| 39 | 0–20% VOs in OO | 4 | 0.3039 | non-HOSO | 0 | 172 | 0.93 | 96 | 96 |
| | | | | HOSO | 1 | 12 | 0.07 | 100 | 100 |
| 40 | 0–20% VOs in OO | 11 | 0.2770 | non-NTSO | 0 | 143 | 0.79 | 93 | 90 |
| | | | | NTSO | 1 | 38 | 0.21 | 97 | 89 |
| 41 | 0–20% VOs in OO | 8 | 0.1904 | non-DOSO | 0 | 164 | 0.89 | 88 | 89 |
| | | | | DOSO | 1 | 20 | 0.11 | 95 | 90 |
| 42 | 0–20% VOs in OO | 7 | 0.2110 | non-RAO | 0 | 163 | 0.89 | 82 | 80 |
| | | | | RAO | 1 | 21 | 0.11 | 90 | 81 |
| 43 | 0–20% VOs in OO | 14 | 0.2809 | non-HR | 0 | 162 | 0.90 | 94 | 90 |
| | | | | HR | 1 | 19 | 0.10 | 95 | 95 |

[1] See abbreviations in Table S2 and S4.

[2] Samples contained in each class: non-RPOO, pure OOs and blends of OO with 2−20% VOs (CO, HOSO, NTSO, DOSO, RAO or HR); RPOO, blends of OO with 2−20% RPOO; non-CO, pure OOs and blends of OO with 2−20% VOs (RPOO, HOSO, NTSO, DOSO, RAO or HR); CO, blends of OO with 2−20% CO; non-HOSO, pure OOs and blends of OO with 2−20% VOs (RPOO, CO, NTSO, DOSO, RAO or HR); HOSO, blends of OO with 2−20% HOSO; non-NTSO, pure OOs and blends of OO with 2−20% VOs (RPOO, CO, HOSO, DOSO, RAO or HR); NTSO, blends of OO with 2−20% NTSO; non-DOSO, pure OOs and blends of OO with 2−20% VOs (RPOO, CO, HOSO, NTSO, RAO or HR); DOSO, blends of OO with 2−20% DOSO; non-RAO, pure OOs and blends of OO with 2−20% VOs (RPOO, CO, HOSO, NTSO, DOSO or HR); RAO, blends of OO with 2−20% RAO; non-HR, pure OOs and blends of OO with 2−20% VOs (RPOO, CO, HOSO, NTSO, DOSO or RAO); HR, blends of OO with 2−20% HR.

**Table S6**

PLS-DA models to detect the presence of a certain vegetable oil in a binary mixture of 2–20% vegetable oil in olive oil, once the presence of RPOO or CO is discarded.[1]

| PLS-DA model | Data | PLS-comp | Boundary | Class[2] | Class code | n | p | %R | %P |
|---|---|---|---|---|---|---|---|---|---|
| 51 | 2–20% VOs in OO | 2 | 0.3689 | non-HOSO | 0 | 98 | 0.89 | 98 | 97 |
| | *without RPOO and CO data* | | | HOSO | 1 | 12 | 0.11 | 100 | 100 |
| 52 | 2–20% VOs in OO | 7 | 0.3706 | non-NTSO | 0 | 72 | 0.65 | 100 | 99 |
| | *without RPOO and CO data* | | | NTSO | 1 | 38 | 0.35 | 95 | 92 |
| 53 | 2–20% VOs in OO | 8 | 0.2569 | non-DOSO | 0 | 89 | 0.82 | 91 | 85 |
| | *without RPOO and CO data* | | | DOSO | 1 | 20 | 0.18 | 100 | 95 |
| 54 | 2–20% VOs in OO | 10 | 0.3905 | non-RAO | 0 | 87 | 0.81 | 91 | 87 |
| | *without RPOO and CO data* | | | RAO | 1 | 20 | 0.19 | 100 | 95 |
| 55 | 2–20% VOs in OO | 15 | 0.3948 | non-HR | 0 | 89 | 0.82 | 97 | 92 |
| | *without RPOO and CO data* | | | HR | 1 | 19 | 0.18 | 100 | 95 |

[1] See abbreviations in Table S2 and S4.

[2] Samples contained in each class: non-HOSO, blends of OO with 2−20% VOs (NTSO, DOSO, RAO or HR); HOSO, blends of OO with 2−20% HOSO; non-NTSO, blends of OO with 2−20% VOs (HOSO, DOSO, RAO or HR); NTSO, blends of OO with 2−20% NTSO; non-DOSO, blends of OO with 2−20% VOs (HOSO, NTSO, RAO or HR); DOSO, blends of OO with 2−20% DOSO; non-RAO, blends of OO with 2−20% VOs (HOSO, NTSO, DOSO or HR); RAO, blends of OO with 2−20% RAO; non-HR, blends of OO with 2−20% VOs (HOSO, NTSO, DOSO or RAO); HR, blends of OO with 2−20% HR.

**Table S7**

PLS-DA models to detect the presence of a certain vegetable oil in a binary mixture of 20−80% vegetable oil in olive oil, once the presence of RPOO or CO is discarded.[1]

| PLS-DA model | Data | PLS-comp | Boundary | Class[2] | Class code | n | p | %R | %P |
|---|---|---|---|---|---|---|---|---|---|
| 63 | 20−80% VOs in OO | 3 | 0.4447 | non-HOSO | 0 | 125 | 0.82 | 100 | 100 |
| | *without RPOO and CO data* | | | HOSO | 1 | 27 | 0.18 | 100 | 100 |
| 64 | 20−80% VOs in OO | 3 | 0.4443 | non-NTSO | 0 | 95 | 0.62 | 100 | 100 |
| | *without RPOO and CO data* | | | NTSO | 1 | 59 | 0.38 | 100 | 100 |
| 65 | 20−-80% VOs in OO | 4 | 0.2963 | non-DOSO | 0 | 131 | 0.87 | 99 | 99 |
| | *without RPOO and CO data* | | | DOSO | 1 | 20 | 0.13 | 100 | 100 |
| 66 | 20−80% VOs in OO | 2 | 0.3560 | non-RAO | 0 | 131 | 0.85 | 92 | 92 |
| | *without RPOO and CO data* | | | RAO | 1 | 23 | 0.15 | 100 | 100 |
| 67 | 20−80% VOs in OO | 8 | 0.2858 | non-HR | 0 | 132 | 0.86 | 97 | 95 |
| | *without RPOO and CO data* | | | HR | 1 | 22 | 0.14 | 91 | 91 |

[1] See abbreviations in Table S2 and S4.

[2] Samples contained in each class: non-HOSO, blends of OO with 20−80% VOs (NTSO, DOSO, RAO or HR); HOSO, blends of OO with 20−80% HOSO; non-NTSO, blends of OO with 20−80% VOs (HOSO, DOSO, RAO or HR); NTSO, blends of OO with 20−80% NTSO; non-DOSO, blends of OO with 20−80% VOs (HOSO, NTSO, RAO or HR); DOSO, blends of OO with 20−80% DOSO; non-RAO, blends of OO with 20−80% VOs (HOSO, NTSO, DOSO or HR); RAO, blends of OO with 20−80% RAO; non-HR, blends of OO with 20−80% VOs (HOSO, NTSO, DOSO or RAO); HR, blends of OO with 20−80% HR.

**Table S8**

The most influential variables on the PLS-DA models to discriminate between 'legal' and 'illegal' blends of olive oil and vegetable oils, 'legal' blends of VOO or OO with NTSO and HOSO, VOO blends with 2% S and 5% S, VOO blends with 2–5% HR and 10% HR, OO blends with DOSO and HR, OO blends with RAO and HR, OO blends with RAO and DOSO, and OO blends of with DOSO and HOSO.[1]

| PLS-DA model | Data | Class[2,3,4,5] | Most discriminant variables: *[1]H-NMR signal intensity is higher in the corresponding class* |
|---|---|---|---|
| 68 | 2–90% VOs in VOO | 'Illegal' blend | Linolenic acid (#15b, #7d) |
| | | 'Legal' blend | Linoleic (#7c, #15a), unsaturated (#24) fatty acids |
| 69 | 2–90% NTSO in VOO | NTSO | Linolenic (#13d, #12c), linoleic (#7c), unsaturated (#9 at 1.32–1.36 ppm) fatty acids |
| | 2–80% HOSO in VOO | HOSO | Oleic (#13b, #7b, #12a), unsaturated (#24 at 5.32–5.34 ppm) fatty acids |
| 70 | 2–80% VOs in OO | 'Illegal' blend | Linolenic (#15b, #7d), oleic (#12a and #7b) acids |
| | | 'Legal' blend | Linoleic (#7c, #15a, #13c), unsaturated (#24) fatty acids, β-sitosterol (#4) and terpenic alcohols or sterols (#2) |
| 71 | 2–80% NTSO in OO | NTSO | Linolenic (#13d, #12c), linoleic (#7c), unsaturated (#9 at 1.32–1.36 ppm, #24 at 5.30–5.32 ppm) fatty acids |
| | 2–80% HOSO in OO | HOSO | Oleic (#13b, #7b, #12a, #9b), unsaturated (#24 at 5.32–5.34 ppm) fatty acids, triacylglycerides (#18) |
| 72 | 2–5% S in VOO | 2% S | Oleic acid (#13b, #7b) |
| | | 5% S | Linolenic acid (#15b, #7d) |
| 73 | 2–10% HR in VOO | 2–5% HR | Linolenic acid (#10c, #12c, #15b), squalene (#11) |
| | | 10% HR | Linoleic acid (#7c) |
| 74 | 2–80% DOSO in OO | DOSO | Oleic (#12a, #9b), saturated (#9a) fatty acids |
| | 2–80% HR in OO | HR | Linoleic acid (#12b, #15a, #7c, #9c) |
| 75 | 2–80% RAO in OO | RAO | Saturated fatty acids (#9a) |
| | 2–80% HR in OO | HR | Oleic (#9b, #7b, #12a), linoleic (#9c) acids |
| 76 | 2–80% RAO in OO | RAO | Linoleic acid (#7c, #12b), squalene (#11) |
| | 2–80% DOSO in OO | DOSO | Oleic (#12a, #9b, #7b), linolenic (#9c, #10c) acids |
| 77 | 2–80% DOSO in OO | DOSO | Oleic (#12a, #9b), unsaturated (#24 at 5.35–5.38 ppm) fatty acids |
| | 2–80% HOSO in OO | HOSO | Linoleic (#12b, #7c), unsaturated (#24 at 5.32–5.34 ppm) fatty acids |

[1] See abbreviations in Table S2 and S4, and the $^{1}$H-signal assignments in Table S1.

[2] Samples contained in each class for PLS-DA models 68–69: 'Illegal' blend, blends of VOO with 2–80% VOs (EVAO, HV, HR or S); 'Legal' blend, blends of VOO with 2-90% VOs (NTSO or HOSO); NTSO, blends of VOO with 2–90% NTSO; HOSO, blends of VOO with 2-80% HOSO.

[3] Samples contained in each class for PLS-DA models 70–71: 'Illegal' blends, blends of OO with 2–80% VOs (RPOO, CO, DOSO, RAO or HR); 'Legal' blends, blends of OO with 2–80% VOs (HOSO or NTSO); NTSO, blends of OO with 2–80% NTSO; HOSO, blends of OO with 2–80% HOSO.

[4] Samples contained in each class PLS-DA models 72–73: 2% S in VOO, blends of VOO with 2% S; 5% S in VOO, blends of VOO with 5% S; 2–5% HR in VOO, blends of VOO with 2–5% HR; 10% HR in VOO, blends of VOO with 10% HR.

[5] Samples contained in each class PLS-DA models 74–77: DOSO, blends of OO with 2–80% DOSO; HR, blends of OO with 2–80% HR; RAO, blends of OO with 2–80% RAO; HOSO, blends of OO with 2–80% HOSO.

**Table S9**

Prediction of the composition of blind oil samples using the classification and regressions models in the decision trees and the complementary PLS-DA models.[1,2,3]

| | PLS-DA | | PLS-R | | | |
|---|---|---|---|---|---|---|
| **Blind sample** | **Models applied** | **Predictions** | **Predicting model** | **Blend** | **% VO** | **Description** |
| 1 | 1, 2, 25-28, 68, 69 | 'Legal' NTSO in VOO | 3 | NTSO-VOO | 39.6 ± 1.9 | *EVOO + NTSO, 60:40* |
| 2 | 1, 2, 25-28, 68, 69 | 'Legal' NTSO in VOO | 3 | NTSO-VOO | 50.8 ± 1.9 | *EVOO + NTSO, 50:50* |
| 3 | 1, 2, 25-28, 68, 69 | 'Legal' NTSO in VOO | 3 | NTSO-VOO | 61.4 ± 1.9 | *EVOO + NTSO, 40:60* |
| 4 | 1, 2, 25-28, 68, 69 | 'Legal' HOSO in VOO | 5 | HOSO-VOO | 40.0 ± 3.9 | *EVOO + HOSO, 60:40* |
| 5 | 1, 2, 25-28, 68, 69 | 'Legal' HOSO in VOO | 5 | HOSO-VOO | 50.1 ± 3.9 | *EVOO + HOSO, 50:50* |
| 6 | 1, 2, 25-28, 68, 69 | 'Legal' HOSO in VOO | 5 | HOSO-VOO | 60.3 ± 3.9 | *EVOO + HOSO, 40:60* |
| 7 | 1, 30-36, 29, 56-67, 70, 71 | 'Legal' NTSO in OO | 20 | NTSO-OO | 41.7 ± 2.8 | *OO + NTSO, 60:40* |
| 8 | 1, 30-36, 29, 56-67, 70, 71 | 'Legal' NTSO in OO | 20 | NTSO-OO | 51.2 ± 2.8 | *OO + NTSO, 50:50* |
| 9 | 1, 30-36, 29, 56-67, 70, 71 | 'Legal' NTSO in OO | 20 | NTSO-OO | 62.1 ± 2.8 | *OO + NTSO, 40:60* |
| 10 | 1, 30-36, 29, 56-67, 70, 71 | 'Legal' HOSO in OO | 18 | HOSO-OO | 39.9 ± 1.6 | *OO + HOSO, 60:40* |
| 11 | 1, 30-36, 29, 56-67, 70, 71 | 'Legal' HOSO in OO | 18 | HOSO-OO | 49.9 ± 1.6 | *OO + HOSO, 50:50* |
| 12 | 1, 30-36, 29, 56-67, 70, 71 | 'Legal' HOSO in OO | 18 | HOSO-OO | 60.3 ± 1.6 | *OO + HOSO, 40:60* |
| 13 | 1, 2, 3-24, 68, 69 | VOO; low; non-VO; 'illegal' | 6 | EVAO-VOO | 6.5 ± 2.1 | *EVOO + EVAO, 95:5* |
| | | | 4 | HOSO-VOO | 3.9 ± 6.8 | |
| | 73 | 2-5% HR in VOO | 11 | HR-VOO | 3.9 ± 5.6 | |
| 14 | 1, 2, 3-24, 68, 69 | VOO; low; EVAO; 'illegal' | 6 | EVAO-VOO | 12.9 ± 2.1 | *EVOO + EVAO, 90:10* |
| 15 | 1, 2, 3-24, 68, 69 | VOO; low; EVAO; 'illegal' | 6 | EVAO-VOO | 23.9 ± 2.1 | *EVOO + EVAO, 80:20* |
| 16 | 1, 2, 25-28, 68, 69 | VOO; high; EVAO; 'illegal' | 7 | EVAO-VOO | 42.6 ± 3.4 | *EVOO + EVAO, 70:30* |
| 17 | 1, 2, 3-24, 68, 69 | VOO; low; HV; 'illegal' | 9 | HV-VOO | 9.5 ± 2.6 | *EVOO + HV, 95:5* |
| 18 | 1, 2, 3-24, 68, 69 | VOO; low; HV; 'illegal' | 9 | HV-VOO | 10.9 ± 2.6 | *EVOO + HV, 90:10* |
| 19 | 1, 2, 3-24, 68, 69 | VOO; low; HV; 'illegal' | 9 | HV-VOO | 26.0 ± 2.6 | *EVOO + HV, 80:20* |
| 20 | 1, 2, 25-28, 68, 69 | VOO; high; HV; 'illegal' | 9 | HV-VOO | 27.4 ± 2.6 | *EVOO + HV, 70:30* |

| Blind sample | PLS-DA | | PLS-R | | | Description |
|---|---|---|---|---|---|---|
| | Models applied | Predictions | Predicting model | Blend | % VO | |
| 21 | 1, 30-36, 29, 37-67, 70, 71 | OO; low; RAO, DOSO; 'illegal' | 21 | DOSO-OO | 1.4 ± 1.6 | *OO + RAO, 95:5* |
| | 76 | RAO in OO | 23 | RAO-OO | 0.0 ± 1.5 | |
| 22 | 1, 30-36, 29, 37-67, 70, 71 | OO; low; RAO, DOSO; 'illegal' | 21 | DOSO-OO | 4.4 ± 1.6 | *OO + RAO, 90:10* |
| | 76 | DOSO in OO | 23 | RAO-OO | 9.0 ± 1.5 | |
| 23 | 1, 30-36, 29, 37-67, 70, 71 | OO; low; RAO, DOSO; 'illegal' | 21 | DOSO-OO | 13.2 ± 1.6 | *OO + RAO, 80:20* |
| | 76 | RAO in OO | 24 | RAO-OO | 22.3 ± 2.7 | |
| 24 | 1, 30-36, 29, 37-67, 70, 71 | OO; low; RAO, DOSO; 'illegal' | 21 | DOSO-OO | 19.2 ± 1.6 | *OO + RAO, 70:30* |
| | 76 | RAO in OO | 24 | RAO-OO | 22.6 ± 2.7 | |
| 25 | 1, 30-36, 29, 37-55, 70, 71 | OO; low; RAO; 'illegal' | 24 | RAO-OO | 12.7 ± 2.7 | *OO + HR, 95:5* |
| 26 | 1, 30-36, 29, 37-67, 70, 71 | OO; low; HR, RAO; 'illegal' | 25 | RAO-OO | 36.2 ± 3.1 | *OO + HR, 90:10* |
| | 75 | HR in OO | 26 | HR-OO | 6.4 ± 1.0 | |
| 27 | 1, 30-36, 29, 37-55, 70, 71 | OO; low; HR; 'illegal' | 26 | HR-OO | 15.0 ± 1.0 | *OO + HR, 80:20* |
| | | | 27 | HR-OO | 20.3 ± 1.3 | |
| 28 | 1, 30-36, 29, 37-55, 70, 71 | OO; low; HR; 'illegal' | 27 | HR-OO | 28.3 ± 1.3 | *OO + HR, 70:30* |
| 29 | 1, 30-36, 29, 37-67, 70, 71 | OO; low; RPOO, RAO, DOSO; 'illegal' | 13 | RPOO-OO | 5.2 ± 0.5 | *OO + RPOO, 95:5* |
| 30 | 1, 30-36, 29, 37-67, 70, 71 | OO; low; RPOO, RAO, DOSO; 'illegal' | 13 | RPOO-OO | 10.1 ± 0.5 | *OO + RPOO, 90:10* |
| 31 | 1, 30-36, 29, 37-67, 70, 71 | OO; low; RPOO; 'illegal' | 13 | RPOO-OO | 19.8 ± 0.5 | *OO + RPOO, 80:20* |
| | | | 14 | RPOO-OO | 20.4 ± 1.6 | |
| 32 | 1, 30-36, 29, 37-67, 70, 71 | OO; low; RPOO; 'illegal' | 14 | RPOO-OO | 30.7 ± 1.6 | *OO + RPOO, 70:30* |
| 33 | 1, 30-36, 29, 37-55, 70, 71 | OO; low; DOSO; 'illegal' | 21 | DOSO-OO | 4.8 ± 1.6 | *OO + DOSO, 95:5* |
| 34 | 1, 30-36, 29, 37-55, 70, 71 | OO; low; DOSO/HOSO; legal-HOSO | 17 | HOSO-OO | 2.0 ± 2.1 | *OO + DOSO, 90:10* |
| | 77 | HOSO in OO | 18 | HOSO-OO | 11.2 ± 1.6 | |
| | | | 21 | DOSO-OO | 12.4 ± 1.6 | |
| 35 | 1, 30-36, 29, 37-55, 70, 71 | OO; low; DOSO; 'illegal' | 21 | DOSO-OO | 21.0 ± 1.6 | *OO + DOSO, 80:20* |
| | | | 22 | DOSO-OO | 20.1 ± 4.0 | |
| 36 | 1, 30-36, 29, 37-55, 70, 71 | OO; low; DOSO/HR; 'illegal' | 22 | DOSO-OO | 35.1 ± 4.0 | *OO + DOSO, 70:30* |
| | 74 | DOSO in OO | 27 | HR-OO | 29.4 ± 1.3 | |

| Blind sample | PLS-DA | | PLS-R | | | Description |
|---|---|---|---|---|---|---|
| | Models applied | Predictions | Predicting model | Blend | % VO | |
| 37 | 1, 2, 25-28, 68, 69 | VOO; high; NTSO; legal-NTSO | 3 | NTSO-VOO | 99.4* ± 1.9 | *Label: 65% NTSO + 35% EVOO*[4] |
| 38 | 1, 2, 25-28, 68, 69 | VOO; high; NTSO; legal-NTSO | 3 | NTSO-VOO | 104.9* ± 1.9 | *Label: Vegetable oil + VOO*[4] |
| 39 | 1, 30-36, 29, 37-67, 70, 71 | OO; low; CO, RAO, HR; 'illegal' | 16 | CO-OO | 56.4 ± 0.6 | *Label: Rapeseed oil + EVOO*[4,5] |
| | 75 | HR in OO | 27 | HR-OO | 107.3* ± 1.3 | |
| 40 | 1, 30-36, 29, 56-67, 70, 71 | OO; high; NTSO; legal-NTSO | 20 | NTSO-OO | 93.2* ± 2.8 | *Label: 80% Rapeseed oil + 20% VOO*[4,5] |
| 41 | 1, 30-36, 29, 37-67, 70, 71 | OO; low; CO, RAO, HR; 'illegal' | 16 | CO-OO | 52.0 ± 0.6 | *Label: 75% Rapeseed oil + 25% EVOO*[4,5] |
| | 75 | HR in OO | 27 | HR-OO | 106.9* ± 1.3 | |
| 42 | 1, 30-36, 29, 37-67, 70, 71 | OO; low; CO, RAO, HR; 'illegal' | 16 | CO-OO | 41.6 ± 0.6 | *Label: 75% Rapeseed oil + 25% EVOO*[4,5] |
| | 75 | HR in OO | 27 | HR-OO | 95.5* ± 1.3 | |
| 43 | 1, 30-36, 29, 37-67, 70, 71 | OO; low; CO, RAO, HR, DOSO; 'illegal' | 16 | CO-OO | 51.2 ± 0.6 | *Label: 80% Rapeseed oil + 20% EVOO*[4,5] |
| | 75 | HR in OO | 27 | HR-OO | 106.9* ± 1.3 | |
| 44 | 1, 30-36, 29, 56-67, 70, 71 | OO; high; NTSO; legal-NTSO | 20 | NTSO-OO | 93.3* ± 2.8 | *Label: 80% Vegetable oil + 20% VOO*[4] |

[1] See abbreviations in Table S2 and S4.

[2] Decision trees in Figures 1 and S1.

[3] Complementary PLS-DA models: PLS-DA models 72–77 in Table 5.

[4] The label did not comply with the Reg. (EU) 29/2012 and its amendments, since the commercial blend did not contain at least 50% of olive oil, and therefore, the presence of olive oil on the label is forbidden.

[5] From the predictions achieved, it could be infer that samples (**39**, **41**–**43**) did not contain NTSO or HOSO, and presented close composition to pure HR or blends of 50% CO in OO. Sample **40** was identified by all classification models as a NTSO-OO blend.

* Extrapolated results (outside the calibration range of the regression model).

**Declaration of interests**

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

| |
|---|
| |

**CRediT authorship contribution statement**

**Rosa M. Alonso-Salces:** Conceptualization, Methodology, Formal analysis, Validation, Investigation, Writing – Original Draft, Writing – Review & Editing, Visualization.

**Luis A. Berrueta:** Resources, Writing – Review & Editing.

**Beatriz Quintanilla-Casas:** Investigation.

**Stefania Vichi:** Resources, Writing – Review & Editing.

**Alba Tres:** Resources, Writing – Review & Editing.

**María Isabel Collado:** Investigation.

**Carlos Asensio-Regalado:** Investigation.

**Gabriela Elena Viacava:** Investigation.

**Aimará Ayelen Poliero:** Investigation.

**Enrico Valli:** Investigation.

**Alessandra Bendini:** Resources, Writing – Review & Editing.

**Tullia Gallina Toschi:** Resources, Writing – Review & Editing.

**José Manuel Martínez-Rivas:** Resources, Writing – Review & Editing.

**Wenceslao Moreda:** Resources, Writing – Review & Editing.

**Blanca Gallo:** Resources, Writing – Review & Editing.