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

Rosa María Alonso-Salces, Luis Ángel Berrueta, Beatriz Quintanilla-Casas, Stefania Vichi, Alba Tres, María Isabel Collado, Carlos Asensio-Regalado, Gabriela Elena Viacava, Aimará Ayelen Poliero, Enrico Valli, Alessandra Bendini, Tullia Gallina Toschi, José Manuel Martínez-Rivas, Wenceslao Moreda, Blanca Gallo

Food Chemistry 336 : (2022)

DOI:10.1016/j.foodchem.2021.130588

This document is the Accepted Manuscript version of a Published Work that appeared in final form in Food Chemistry, Volume 366, 1 January 2022, 130588 https://doi.org/10.1016/j.foodchem.2021.130588

Stepwise strategy based on ^{1}H-NMR fingerprinting in combination with chemometrics to determine the content of vegetable oils mixtures. Part II: Blends with the 'olive oil' category

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], Enrico Valli[6], Alessandra Bendini[6], Tullia Gallina Toschi[6], José Manuel Martínez-Rivas[7], Wenceslao Moreda[8], 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, S.ta Coloma de Gramenet, 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] Department of Agricultural and Food Sciences, Alma Mater Studiorum University of Bologna (UNIBO), Piazza Goidanich, 60, I-47521, Cesena (FC), Bologna, Italy.

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

[8] 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). ^{1}H-NMR spectral data of oils of the 'olive oil' category and their mixtures with the VOs most commonly used to make blends, i.e. sunflower, high oleic sunflower, corn, refined avocado, refined hazelnut, refined palm olein and desterolized high oleic sunflower oils, is analysed by pattern recognition techniques. Partial least squares (PLS) discriminant analysis provides stable and robust binary classification models to identify the VO, and PLS regression affords models with excellent precisions and acceptable accuracies to determine the percentage of VO in the mixture. The performance of this approach is tested with blind samples: the satisfactory results achieved confirm its potential to support regulations and control bodies.

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

1. Introduction

Olive oil adulteration for the purpose of financial gain has become one of the biggest sources of agricultural fraud in the European Union, as pointed out by the European Parliament (EC, 2020; European Parliament, 2014). Both the EU being the world's largest olive oil producer and consumer, accounting for 70% of global production (IOC, 2019), as well as the enlarged competitiveness, highlight the need to update and harmonize analytical methods for quality and authenticity control of olive oil. In the long-term, a lack of trust in the quality and authenticity of olive oil has the potential to damage the reputation and competitiveness of the European olive oil sector. In this context, the European Commission supported the so called OLEUM Project 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).

According to the Reg. (EU) 29/2012, olive oils, of any edible category, can be mixed with vegetable oils (VOs) and marketed highlighting the presence of olive oil on the labelling outside of the list of ingredients, by words, images or graphic representations, only in the case that it accounts for at least 50% of the blend (EC, 2012). Amendments to this regulation established that Member States have the possibility to allow or prohibit the production in their territory of blends of olive oil and other vegetable oils for internal consumption, the marketing in their territory of such blends coming from other countries, and/or the production in their territory of such blends for marketing in another Member State or for exportation. In this context, the terms of 'legal' and 'illegal' blends arise alluding to admixtures that comply or not with the EU regulation, as well as those adopted by each Member State. It is noteworthy that the 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. Hence, the need of analytical methods in order to implement the established regulations is obvious and extremely important (Conte, Bendini, Valli, Lucci, Moret, Maquet, et al., 2019). Indeed, in literature, very few works deals 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, i.e.

blends with percentages of olive oil close to the 50%. Among them, Gomez-Coca et al. (2020) successfully proposed the combination of four purity parameters, some described in legislation (Commission regulation (EEC) 2568/91), organized in decisional trees to discern olive oil concentrations, using sunflower oil as a model seed oil to blend with olive oil (Gómez-Coca, Pérez-Camino, Martínez-Rivas, Bendini, Gallina Toschi, & Moreda, 2020). The potential of the fatty acid composition of the oil determined using the official method and multivariate data analysis was also proved for blends with sunflower oil (Monfreda, Gobbi, & Grippa, 2012) and other seed oils (Monfreda, Gobbi, & Grippa, 2014). Spectroscopic techniques as TD-NMR and FTIR combined with chemometrics were investigated for the same goal (De la Mata, Dominguez-Vidal, Bosque-Sendra, Ruiz-Medina, Cuadros-Rodríguez, & Ayora-Cañada, 2012; Santos, Kock, Santos, Lobo, Carvalho, & Colnago, 2017). Few research works studied the mixtures of olive oil (OO), i.e. blends of virgin and refined olive oil, with other VOs (De La Mata-Espinosa, Bosque-Sendra, Bro, & Cuadros-Rodríguez, 2011; De la Mata et al., 2012; Monfreda et al., 2012, 2014), or the challenged adulteration of refined olive oil with refined hazelnut oil (Agiomyrgianaki, Petrakis, & Dais, 2010; Mannina, D'Imperio, Capitani, Rezzi, Guillou, Mavromoustakos, et al., 2009).

Chemical methods traditionally used for quality and authenticity control of olive oil are laborious, time-consuming, involves expensive and toxic chemicals with high environmental impact, require sample preparation and skilled operators. New and complementary analytical techniques will overcome some of these drawbacks and/or even support currently used methods to accomplish the complex task of the detection and quantification of olive oil mixtures with other oils. In this sense, metabolomic approaches, which allow rapid determination of several classes of chemical components using efficient advanced instrumental techniques coupled to chemometrics, are given great attention. MS-based methodologies and NMR spectroscopy are widely used in nowadays research in food analysis for quality control and traceability (Lioupi, Nenadis, & Theodoridis, 2020). Different NMR techniques, i.e. ^{1}H-NMR, ^{13}C-NMR, ^{31}P-NMR and/or ^{19}F-NMR have been used to characterise olive oils with authenticity and traceability purposes (Alonso-Salces,

Segebarth, Garmón-Lobato, Holland, Moreno-Rojas, Fernández-Pierna, et al., 2015; Guillén & Ruiz, 2001; Jiang, Li, Chen, & Weng, 2018; Vigli, Philippidis, Spyros, & Dais, 2003). Most of the NMR approaches developed for olive oil authentication, detection of olive oil adulteration and/or determination of olive oil blends with VOs were based on measuring signals of the NMR spectrum that give quantitative information of certain compounds or are used to calculate some parameters and ratios (i.e. profiling) (Agiomyrgianaki et al., 2010; García-González, Mannina, D'Imperio, Segre, & Aparicio, 2004; Jiang et al., 2018; Mannina et al., 2009; Popescu, Costinel, Dinca, Marinescu, Stefanescu, & Ionete, 2015; Vigli et al., 2003; Vlahov, 2009; Zamora, Alba, & Hidalgo, 2001). Instead, NMR fingerprinting approaches were 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 for the study of mixtures of olive oil with other VOs for the first time in the framework of OLEUM research project. The aim of the study was to develop an analytical strategy based on ^{1}H-NMR fingerprinting together with multivariate classification and regression models organised in a decision tree scheme in order 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 present article describes the second part of the stepwise strategy, which allows to identify the VO and determine the percentage of VO in a blend with OO, once the presence of oil of the 'olive oil' category has been confirmed by the classification model in the first step of the decision tree. Furthermore, the performance of the complete stepwise strategy is evaluated by the prediction results obtained on an external set of blind oil samples and commercial oils. Moreover, it is worth to be noticed that this analytical strategy addresses some issues not considered in previous approaches, such as the discrimination between (*i*) oil samples containing oil of the 'virgin olive oil' category (VOO) and 'olive oil' category (OO) and (*ii*) pure and blended oils, and the study of (*iii*) a large sample set with pure oils and blends of the most common VOs used for olive oil adulteration, and

(*iv*) a wide range of percentages of the VOs in the blend (including those percentages for the verification of the labelling compliance of Reg. (EU) 29/2012).

2. Material and methods

2.1. Samples

Genuine samples of virgin (VOO) and extra virgin olive oils (EVOO) (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 blended samples). Samples were obtained in the framework of Autenfood project (ACCIÓ Programa Operatiu FEDER Catalunya 2014–2020) and OLEUM project (EC H2020 Programme 2014–2020). Oils from the sample banks of both projects were produced during two consecutive harvest years (2016/17 and 2017/18). In addition, eight commercial oil samples collected in the Swedish market were analysed. The label of these commercial oils described that they were mixtures of VOO or OO, and other VO such as rapeseed oil, sunflower oil, or a 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

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

179 PCA, PLS-DA and PLS-R were applied to the autoscaled or centered data matrix of ^{1}H-NMR
180 spectra (542 variables) of the oil samples. The presence of outliers in the dataset was analysed by
181 PCA. In PLS-DA and PLS-R, the optimal number of PLS-components are estimated by cross-
182 validation by plotting the PRESS or RMSEP against the number of PLS-components. Sometimes
183 there are several almost equivalent local minima on the curve; the first one should be preferred to
184 avoid overfitting (according to the principle of parsimony). The model with the smallest number of
185 features should be accepted from among equivalent models on the training set. In PLS-DA, once the
186 number of PLS-components is optimised, the predictions in the training-test set are represented in a
187 box and whisker plot in order to define the half of the distance between the quartiles as the
188 boundary. The regression coefficients (B) of the optimal number of PLS-components denote the
189 importance of the NMR variables on the model: the larger the B-coefficient, the higher the
190 influence of the variable on the PLS-DA model or PLS-R model. A large B-coefficient may also
191 indicate a variable with small absolute values but large relative differences (Esbensen, Guyot,
192 Westad, & Houmøller, 2002). Classification and regression models achieved by PLS-DA and PLS-
193 R respectively were validated by 3-fold cross-validation or leave-one out cross-validation for
194 parameter optimization, and by external validation when an external set of samples was available.
195 Binary classification models can lead to artefacts if they are not used and validated properly
196 (Kjeldahl & Bro, 2010). The reliability of the classification models developed was studied in terms
197 of recognition ability (percentage of the samples in the training set correctly classified during the
198 modelling step), prediction ability in the cross-validation (percentage of the samples in the test set
199 correctly classified by using the model developed in the training step), and prediction ability in the
200 external validation (percentage of the samples in the external set correctly classified by using the
201 optimised model) (Berrueta et al., 2007). The goodness of the regression model fit was evaluated by
202 means of the prediction error, which is an expression of the error expected when using the

calibration model to predict; the correlation coefficient between predicted and measured values in calibration and validation (R-cal and R-val); the coefficient of determination in calibration and validation (R^2-cal and R^2-val), which indicates the percentage of the variance in the dependent variable that the independent variables explain collectively; and the evaluation of the residuals, which show how well each individual object is modelled and predicted. The RMSEP (root mean square error in the prediction) expresses the average error to be expected associated with future predictions, i.e. the estimated precision. Thus, the RMSEP is the practical average prediction error as estimated by the validation set, and therefore an empirical error estimate, which is expressed 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 %, i.e. RMSEP divided by the measured data, and is comparable to the analytical accuracy (% of relative standard deviation) (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 adulteration or making legal blends, i.e. refined conventional sunflower oil, refined high oleic sunflower oil, desterolized and deodorized high oleic sunflower oil, refined hazelnut oil, virgin hazelnut oil, refined soybean oil, virgin avocado oil, refined avocado oil, refined palm olein oil and refined corn oil, 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 ^{1}H signals and their assignments to protons of the different functional groups are shown in Table S1 in the 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 (Alonso-Salces, Berrueta, Quintanilla-Casas, Vichi, Tres, Collado, et al., 2020).

As reported for VOO mixtures with VOs (Alonso-Salces et al., 2020), the proposed approach to detect blends of OO with other VOs and quantify the percentage of VO in the blend, is based in the use of the ^{1}H-NMR fingerprint of the oil and a set of multivariate classification and regression models organized in the decision tree scheme (Figure 1). The first step of this strategy is to determine whether the oil sample contains VOO or OO using the classification model (PLS-DA model 1 in Table 1), already described in the first part of the study (Alonso-Salces et al., 2020). Once the presence of OO is confirmed in the oil sample, binary classification models developed by PLS-DA are used to (*i*) detect a certain VO in a blend with OO, (*ii*) determine in which proportion (low or high) the VO is present, and (*iii*) differentiate between 'legal' (containing NTSO or HOSO) and 'illegal' (containing other VOs) blends with OO, using sunflower oil as a seed oil model (Gómez-Coca et al., 2020; Monfreda et al., 2012). The PLS-DA models achieved for blends of VOs and OO (PLS-DA models 29–67 and 70–71) are shown in Tables 1–4 and Tables S2–S5 (supplementary material). Finally, a regression model built by PLS-R determines the percentage of VO in the blend with OO. These PLS-DA and PLS-R models and their chemical interpretation are described in the next sections.

3.1.1. PLS-DA models to discriminate blends of olive oils with vegetable oils

Satisfactory PLS-DA models for all the VOs (RPOO, CO, HOSO, NTSO, DOSO, RAO and HR) were achieved using the whole percentage range of VO in the OO mixture, i.e. 0–80% VO in OO (PLS-DA models 30–36 in Table S2 in the supplementary material). Prediction abilities of the binary classification models developed to discriminate between OO blends with RPOO, CO or HOSO were 95–100% for both categories; with NTSO, DOSO or RAO, 84–87% for the OO blends with the specific VO, and 91–97% for the OO blends that did not contain the specific VO; and with HR, 97% for HR-OO category and 89% for the non-HR category.

The main ^{1}H signals responsible for the identification of OO blends containing RPOO were the methylene protons of saturated fatty acids (#9a), which presented significantly higher intensities in

RPOO-OO blends, even though were not completely discriminant; whereas the methylene (#9c) and allylic (#12b) protons of linoleic acid showed lower intensities in the RPOO category. Palm oil is, among the VOs studied, the one that contains the highest amounts of saturated fatty acids (Vigli et al., 2003). Indeed, palmitic acid is the major saturated fatty acid in palm oil, being present 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 other VOs (Montoya, Cochard, Flori, Cros, Lopes, Cuellar, et al., 2014).

The blends of CO in OO were distinguished from the other VOs in OO as a result of its characteristic profiles of fatty acids and triacylglycerides (Christopoulou, Lazaraki, Komaitis, & Kaselimis, 2004; Gómez-Ariza, Arias-Borrego, García-Barrera, & Beltran, 2006; Jabeur, Zribi, Makni, Rebai, Abdelhedi, & Bouaziz, 2014; Yang, Ferro, Cavaco, & Liang, 2013). The intensity of the most influential signals, i.e. the methyl protons of the acyl groups of linoleic acid (#7c) and saturated fatty acids (#7a), the bis-allylic protons of linolenic acid (#15b) and the glyceryl protons at 4.30–4.32 ppm of triacylglycerides (#18), were higher in the blends containing CO; while the glyceryl protons at 4.28–4.30 ppm of triacylglycerides (#18) showed the opposite trend. In fact, corn oil present similar linoleic contents to sunflower oil, and significantly higher ones than refined avocado, refined hazelnut, palm and olive oil; slightly higher amounts of linolenic acid than the other oils studied; and a level of saturated fatty acids lower than palm oil but similar or slightly higher than the rest of the oils considered in the model (Guillén & Ruiz, 2003; Jiang et al., 2018; Monfreda et al., 2012; Vigli et al., 2003).

The major contribution to the binary classification models built to determine the presence of HOSO in OO was due to the methylene protons oleic acid (#9b), the allylic protons of linoleic acid (#12b) and the vinylic protons of unsaturated fatty acids (#24), which exhibited higher intensities in the blends with HOSO, partially overlapping with the non-HOSO category. These observations were in accordance with the fact that 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 amounts in HOSO than in palm, olive, hazelnut and avocado oils and lower than in sunflower and corn oils. The allylic protons of linolenic acid (#12c) of HOSO-OO blends displayed intensities overlapping with the 1st and 2nd quartiles of the non-HOSO category. Actually, linolenic contents of HOSO are similar to those of HR and slightly lower than the other studied oils. The methylene protons of saturated fatty acids (#9a) were also influent in the model, displaying intermediate-high values for HOSO-OO blends overlapping the 2nd, 3rd and 4th quartiles of the non-HOSO category, but far from RPOO-OO blends, which exhibit the largest contents. These observations are supported by previously descriptions of the composition of the pure oils (Green & Wang, 2020; Guillén et al., 2003; Jović, Smolić, Primožič, & Hrenar, 2016; Vigli et al., 2003).

Among the most influential variables in the binary classification models achieved for the detection of NTSO in OO, the methyl (#7c) and bis-allylic (#15a) protons of linoleic acid, and the vinylic protons (#24), the methyl protons at 1.00−1.02 ppm (#7) and the methylene protons at 1.32−1.34 ppm (#9) of unsaturated fatty acids, displayed higher intensities in the OO blends with NTSO. The opposite trends were observed for the allylic (#12a) and methyl (#7b) protons of oleic acid. This behaviour agreed with the composition of sunflower oil, which is characterised by the largest contents of linoleic acid and unsaturated fatty acids, and the lowest contents of oleic acid respect to the other oils studied (Christopoulou et al., 2004; Guillén et al., 2003; Jabeur et al., 2014; Jović et al., 2016; Monfreda et al., 2012; Yang et al., 2013).

The PLS-DA model built to distinguish OO blends containing DOSO presented the highest absolute B coefficients for the allylic protons of oleic (#12a) and linoleic (#12b) acids and the methylene protons of oleic acid (#9b). During the desterolization process, the dehydration of sterols and the elimination of the acid group of sterol esters take place by bleaching, producing olefinic degradation products and di-steryl ethers; however the profiles of triacylglycerides and fatty acids are practically unaltered (Grob, Biedermann, Bronz, & Giuffré, 1994). Instead, 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 & Harwood, 2013). Thus, DOSO is characterized by relatively high contents of oleic and linoleic acids as HOSO. Nevertheless, the intensities of the signal due to oleic acid (#12a and #9b) in DOSO-OO blends presented a large variability and were significantly higher than in the non-DOSO category, even though not completely discriminant between both categories. In contrast, linoleic acid signal (#12b) intensities of DOSO-OO blends exhibited less variability and overlapped with the 1st and 2nd quartiles of the non-DOSO category. The methyl protons of linoleic acid (#7c) in DOSO-OO blends exhibited a narrow range of intensities close to the 2nd quartiles of the non-DOSO category. The vinylic protons of unsaturated fatty acids (#24) displayed lower intensities and high variability in the blends with DOSO overlapping the 1st, 2nd and 3rd quartiles of the non-DOSO category.

Refined avocado oil presents intermediate compositions regarding its fatty acid profiles compared to the other oils studied (Guillén et al., 2003; Jabeur et al., 2014; Jović et al., 2016; Vigli et al., 2003; Yang et al., 2013), as well as its sterol contents, in particular β-sitosterol (Al-Ismail, Alsaed, Ahmad, & Al-Dabbas, 2010; Fernandes, Gómez-Coca, Pérez-Camino, Moreda, & Barrera-Arellano, 2017; Green et al., 2020; Parcerisa, Casals, Boatella, Codony, & Rafecas, 2000). In fact, the intermediate composition of RAO was reflected in the NMR spectra, and in particular in the most important variables in the binary classification model to detect RAO in OO. Thus, the methyl (#7c), allylic (#12b) and α-methylene (#13c) protons of linoleic acid and the methyl protons of β-sitosterol (#4) exhibited signals for the RAO-OO blends with intensity ranges overlapping with 2nd, 3rd and 4th quartiles of the non-RAO category. Whereas, the α-methylene protons of linolenic acid (#13d) and the methylene protons at 1.32−1.34 ppm of unsaturated fatty acids (#9) of RAO-OO blends presented intensity ranges overlapping with 1st, 2nd and 3rd quartiles of the non-RAO category, and a median intensity value lower for RAO-OO mixtures than for the other types of mixtures. The signal intensities of the methylene protons at 1.20−1.22 ppm of the acyl group of saturated fatty acids (#9) in RAO-OO blend overlapped the 2nd and 3rd quartiles of the non-RAO category.

The most contributing variables for the identification of HR in OO were the methyl (#7b) and allylic (#12a) protons of oleic acid, presenting higher intensities in the HR-OO blends; and the methyl (#7d), bis-allylic (#15b) and allylic (#12c) protons of linolenic acid, the vinylic protons of unsaturated fatty acids (#24), the β-methylene (#10a) and methyl (#7a) protons of saturated fatty acids and the methyl protons of terpenic alcohols or sterols (#2), showing lower intensities in the mixtures of HR with OO than in the other VO-OO blends. Refined hazelnut oil contains the highest amounts of oleic acid among the VOs studied, and comparable contents to those of olive oil. HR presents the lowest linolenic acid contents similar to HOSO, therefore the linolenic signal intensities overlapped with 1st and 2nd quartiles of the non-HR category (Green et al., 2020; Guillén et al., 2003; Jović et al., 2016; Parcerisa et al., 2000; Vigli et al., 2003). Each oil type presents characteristic profiles of sterols and terpenic alcohols (Al-Ismail et al., 2010; Aparicio et al., 2013; Fernandes et al., 2017; Parcerisa et al., 2000), which in the present model contributed to the distinction of OO blends with and without HR.

In order to improve the classification results of the full percentage range models, further PLS-DA models were developed, the stepwise strategy proceed as follows. Once the oil sample is classified as containing OO in the first stage of the decision tree scheme (Figure 1), a PLS-DA model classifies the sample according to their level of VO in the OO, i.e. low (OOs with 0–20% VOs) and high (OOs with 30–80% VOs); the prediction abilities being of 96% and 94% for the low and high categories respectively (PLS-DA model 29 in Table 1). The most influential variables on the model were the methyl protons of saturated fatty acids (07a) and β-sitosterol (#4); the allylic (#12b), bis-allylic (#15a) and α-methylene (#13c) of linoleic acid; and the vinylic protons (#24), the methyl protons at 1.00–1.02 ppm (#7) and the methylene protons at 1.32–1.33 ppm (#9) of unsaturated fatty acids, which exhibited lower intensities in the low category, overlapping with the 1st and 2nd quartiles of the high category. In contrast, the ^{1}H signals of linolenic acid (#7d and #15b) and oleic acid (#12a) displayed higher intensities in the low category, overlapping with the 3rd and 4th quartiles of the high category. The chemical composition of the blends that constituted each

355 category justified these observations, since the low category contained samples with highest
356 proportions of OO, which is the oil that contains the highest concentrations of oleic acid together
357 with HR; whereas in the high category, blends with high percentages of VOs characterised by high
358 linoleic and β-sitosterol contents were included (Al-Ismail et al., 2010; Aparicio et al., 2013;
359 Fernandes et al., 2017; Green et al., 2020; Guillén et al., 2003; Jović et al., 2016; Parcerisa et al.,
360 2000; Vigli et al., 2003).

361 In next stage of the decision tree scheme (Figure 1), an oil sample, classified in the low category
362 (PLS-DA model 29 in Table 1), is predicted by binary classification models in order to identify the
363 specific VO contained in the OO blend (PLS-DA models 44–50 in Table 2 and PLS-DA models
364 37–43 in Table S3 in the supplementary material). The PLS-DA models developed including or not
365 pure OO samples in the dataset were similar. The recognition and prediction abilities achieved were
366 higher than 95% of hits in the models for detecting RPOO, CO and HOSO in OO blends; c.a. 90%
367 for NTSO, DOSO and HR in OO blends; and c.a. 80–85% for RAO in OO blends. The most
368 influential variables in the binary classification models for OO blends with ≤20% RPOO, NTSO,
369 DOSO or HR (except the saturated fatty acid signals) were the same as in the corresponding models
370 built for the full percentage range of VO in OO, but also other characteristic signals stood out. In
371 the particular case of the PLS-DA models for mixtures of CO and OO, the methyl protons of the
372 acyl group of linoleic acid (#7c) was the main responsible for the distinction of OO mixtures
373 containing CO from those without CO. The identification of low proportions of HOSO in OO were
374 mainly due to the higher intensity of the ^{1}H signals of the methylene protons of saturated fatty acids
375 (#9a) and unsaturated fatty acids with signals at 1.30–1.34 ppm (#9), which presented higher
376 intensities in the HOSO category. The most influent signals for distinguishing low percentages of
377 RAO in OO were those due to the polyunsaturated fatty acids. Thus, the intensities of the methyl
378 (#7c) and β-methylene (#10c) protons of linoleic acid displayed intensities with similar median
379 values in both categories but different variabilities. Furthermore, the signal intensities of the
380 methylene protons of linolenic acid (#9c) and the methylene protons at 1.31–1.34 ppm (#9) and the

vinylic protons (#24) of unsaturated fatty acids were lower in the OO blends containing RAO, overlapping with the 1st and 2nd quartiles of the non-RAO category.

Taking into account that with the above models, all CO-OO blends and 95% of the RPOO-OO blends were identified, as well as at least 95% of the OO blends not containing CO or RPOO (Table 2 and Table S3 in the supplementary material), further classification models were developed with 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 in Table S4 in the supplementary material) afforded better classification abilities to detect NTSO and RAO in OO-blends, and similar results to resolve the presence of HOSO, DOSO or HR in OO-blends. Hence, once it is discarded that the OO sample contains RPOO or CO using the PLS-DA models 37, 38, 44 and 45 (Table 2 and S3 in the supplementary material), the presence of HOSO, NTSO, DOSO, RAO or HR in an OO blend is predicted by PLS-DA model 51–55. A part from the ^{1}H signals that were important in the previous PLS-DA models, these models presented particular influential variables that contributed to improve their classification abilities. Thus, regarding PLS-DA model 52 for NTSO-OO blends, the methylene protons of oleic acid (#9b) presented lower intensities in the NTSO category, whereas the glyceryl protons of triacylglycerides (#18) showed higher intensities in the NTSO category. Concerning PLS-DA model 54 for RAO-OO blends, the signal intensities of the methyl protons of oleic acid (#7b) and the glyceryl protons of triacylglycerides (#23) in the OO blends containing RAO were overlapping with the 1st, 2nd and 3rd quartiles of the non-RAO category; and the α-methylene of linoleic (#13c) and linolenic acids (#13d), with the 2nd, 3rd and 4th quartiles of the non-RAO category. This observation is consistent with the characteristic intermediate composition of fatty acids that RAO presents, as referred above. No models to discriminate pure OOs from blends of VOs with OO were built due to the low number of pure OO samples available for modelling in the present study.

PLS-DA models were also developed for the blends with 20–80% VOs in OO in order to identify the VO contained in the mixture (PLS-DA models 56–62 in Table 3). The recognition and

prediction abilities of the binary classification models achieved for OO blends with RPOO, CO, DOSO or HR were 98–100% for both categories; OO blends with NTSO and RAO, ≥91% for both categories; and OO blends with HOSO, 86% for the HOSO category and 99% for the non-HOSO category. Since all blends were correctly classified by the RPOO-OO and CO-OO 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 and CO blends with OO (PLS-DA models 63–67 in Table S5 in the supplementary material). These models afforded better recognition and prediction abilities than the previous ones, except for HR-OO blend. Indeed, the models for the identification of OO mixtures with and without NTSO or HOSO allowed the correct classification of all samples of both categories; the model for RAO-OO mixtures identified all samples containing RAO and 92% of samples in the non-RAO category; and the model for DOSO-OO mixtures gave the same classification results as the model built including RPOO and CO spectral data. The model to differentiate OO mixtures with and without HR provided slightly worse classifications than previous model but over 90% of hits for both categories. In fact, the detection of the adulteration of OO with HR is one of the main challenge in fraud detection due to the close composition of both refined oils (Agiomyrgianaki et al., 2010; García-González et al., 2004; Gómez-Ariza et al., 2006; Mannina et al., 2009).

The PLS-DA model for OO blends with RPOO (PLS-DA model 56) disclosed that the signals of the methylene protons of saturated fatty acids (#9a), mainly due to palmitic acid, and the methylene protons of linoleic acid (#9c) were completely discriminant between OO blends containing ≥20% RPOO and OO blends with the other VOs in the same percentages. As a result, the measurement of just one of these two variables would be enough to establish whether an OO is mixed with RPOO in percentages ≥20%. The trends observed for these signals were the same as in the OO blends with low proportions of VOs.

With regard to the binary classification model obtained to detect OO blends with ≥20% CO (PLS-DA model 57), the most significant variables were the same as those in the full range model, and

showed similar trends; even though in the model for the high range (≥20% CO in OO), the methylene protons of oleic acid (#9b) and the methyl protons of linolenic acid (#7d) and β-sitosterol (#4) were also preponderant. Oleic acid signal presented lower intensities in the CO category and overlapped with the 1st and 2nd quartiles of the non-CO category, since corn oil together with sunflower oil are the oils with the lowest contents of this fatty acid. While the signals of linolenic acid and β-sitosterol displayed higher intensities in the CO category, in agreement with the known higher amounts of both compounds in corn oil compared to the other VOs blended with OO (Aparicio et al., 2013; Guillén et al., 2003; Monfreda et al., 2012; Vigli et al., 2003).

The models achieved to identify the presence of HOSO in OO in high proportions (PLS-DA models 58 and 63) presented the highest absolute B coefficients for the same variables than in the full range model (PLS-DA model 32) and followed the same trend. In PLS-DA model 63, additional signals due to oleic acid (#12a) and linoleic acid (#9c) contributed to the discrimination. The signal due to saturated fatty acids (#9a) presented a B coefficient three times higher than the rest of the signals in this model than in the full range model.

The binary classification models built to detect OO blends with ≥20% NTSO (PLS-DA model 59 and 64) presented the same influential variables and observed trends as the previous NTSO-OO models described above (the full percentage and low content models). However, once the presence of RPOO and CO in the OO blend was discarded by the PLS-DA models 56 and 57 respectively, the signals of the bis-allylic (#15a) and allylic (#12b) protons of linoleic acid, the methylene protons at 1.34–1.36 ppm (#9) and the vinylic protons (#24) of unsaturated fatty acids, and the methylene protons of oleic acid (#9b) were completely discriminant between both categories, thus any of this variable can be used as markers to determine whether an OO blend contains NTSO at concentration ≥20%.

The most discriminant variables in the binary classification models for OO blends with ≥20% DOSO (PLS-DA model 60 and 65) were the same as in the corresponding model achieved for the full percentage range. With respect to the PLS-DA models for the detection of ≥20% RAO in OO

(PLS-DA models 61 and 66), fewer ^{1}H signals were influent in these models compared to the full percentage range and the low content models. Among them, the signals of the methylene protons of oleic acid (#9b) and the allylic protons of linoleic acid (#12b) in the RAO category overlapped with the 3rd and 4th quartiles of the non-RAO category, as a result of RAO characteristic intermediate composition of these fatty acids among the VOs studied. Meanwhile, the signal intensities of methylene protons of linolenic acids (#9c) and the vinylic protons of unsaturated fatty acids (#24) showed lower intensities in the RAO-OO blends, overlapping with the 1st quartile of the non-RAO category.

As for the full percentage range and the low range models built to detect HR in OO, the most contributing variables to PLS-DA models 62 and 67 for blends with ≥20% HR in OO were, not only the allylic protons of linoleic (#12b) and oleic (#12a) acids and the methyl protons of oleic acid (#7b), but also the methylene protons of oleic acid (#9b); all of them presenting higher intensities in the HR-OO blends. The trends of the signal intensities of the methyl protons of saturated fatty acids (#7a) and linolenic acid (#7d), the vinylic protons of unsaturated fatty acids (#24) were similar to those observed in previous models, presenting lower intensities in the HR-OO blends. The allylic (#12c) and α- methylene (#13d) protons of linolenic acids were also significant for the differentiation of the OO blends with and without HR, and followed the same trend as its methyl protons (#7d).

3.1.2. PLS-DA models to discriminate between 'legal' and 'illegal' blends of olive oil and vegetable oils

Regarding the aforementioned possibilities that Reg. (EU) 29/2012 and its amendments offer to the Member States, the potential of the present multivariate approach to implement those regulations was demonstrated with the following case study. The most common seed oils used to blend with olive oil are refined conventional (normal type) sunflower oil and refined high oleic sunflower oil, thus sunflower oil was considered as a model seed oil of 'legal' blends. While the blends of olive oil with the other VOs studied would not be permitted in some or all of the situations specified in

485 the regulation, and therefore, were considered as ‘illegal’ blends. Binary classification models were
486 developed to first distinguish between ‘legal’ and ‘illegal’ blends, and then differentiate which of
487 the two types of sunflower oils, i.e. NTSO or HOSO, is in a ‘legal’ blend with OO (Figure S2 in the
488 supplementary material). Afterwards, the percentage of NTSO or HOSO in the blend can be
489 determined by the regression models that are reported in section 3.1.4 in order to verify the
490 percentage of NTSO or HOSO in the declared blend or determine the level of adulteration of the
491 OO.

492 The binary classification model discriminating between ‘legal’ blends (OO with NTSO or HOSO)
493 and ‘illegal’ blends (OO with RPOO, CO, DOSO, RAO or HR) provided prediction abilities of
494 86% and 98% respectively (PLS-DA model 70 in Table 4). As previously reported for ‘legal’ and
495 ‘illegal’ blends with VOO (Alonso-Salces et al., 2020), the most influential variables in the
496 classification model were the bis-allylic (#15b) and methyl (#7d) protons of linolenic acid, which
497 presented higher intensities and/or variability in the ‘illegal’ category. Whereas, the methyl (#7c)
498 and bis-allylic (#15a) protons of linoleic acid and the vinylic protons of unsaturated fatty acids
499 (#24) showed higher intensities and/or variability in the ‘legal’ category. Other relevant variables in
500 the model for OO blends were the allylic (#12a) and methyl (#7b) protons of oleic acid, which
501 exhibit lower intensities and high variability in the ‘legal’ category; and the α-methylene protons of
502 linoleic acid (#13c), the methyl protons of β-sitosterol (#4) and the methyl protons of terpenic
503 alcohols or sterols (#2), with higher intensities in the ‘legal’ blends. The well-known differences in
504 the composition of fatty acids, sterols and terpenic alcohols of NTSO and HOSO respect to the VOs
505 included in the ‘illegal’ category and OO supported these observations.

506 In order to differentiate ‘legal’ OO blends containing NTSO from those with HOSO, binary
507 classification models were constructed affording prediction abilities of 97% for both categories
508 (PLS-DA models 71 in Table 4). Regarding the most important variables in the model, the most
509 discriminant variables were the methylene protons of oleic acid (#9b) and the glyceryl protons of
510 triacylglycerides (#18), which presented higher intensities and lower variability in HOSO category;

and the vinylic protons at 5.30–5.32 ppm of unsaturated fatty acids (#24), which exhibited lower intensities and variabilities in HOSO category. Furthermore, the α-methylene protons of oleic acid (#13b) and the vinylic protons at 5.32–5.34 ppm of unsaturated fatty acids (#24) showed higher signal intensities in HOSO blends than in NTSO blends; in contrast with the α-methylene (#13d) and allylic (#12c) of linolenic acid, the methylene protons at 1.32–1.36 ppm of unsaturated fatty acids (#9) and the methyl protons of linoleic (#7c) that exhibited higher intensities and variability in NTSO blends. Indeed, HOSO contains higher amounts of oleic acid and lower concentrations of linoleic and linolenic acids (polyunsaturated fatty acids) than NTSO (Jović et al., 2016).

3.1.3. PLS-DA models to discriminate between different blends of vegetable oils and olive oil

In the case that more than one binary PLS-DA model of the decision tree scheme (Figure 1) classifies an oil sample as containing the corresponding VO, further binary classification models can be built using datasets containing only the information related to those specific VOs. For instance, PLS-DA models shown in Table 4 were built to distinguish 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).

The prediction abilities of the model to differentiate between mixtures with DOSO and HR were 84% for the DOSO-OO blends and 95% for the HR-OO blends. The major contributing signals to this model were the allylic protons of oleic acid (#12a) and the methyl protons of saturated fatty acids (#9a) and oleic acid (#9b), which presented lower intensities in the HR-OO blends overlapping with the 1st, 2nd and 3rd quartiles of the DOSO category; and the allylic protons (#12b), bis-allylic (#15a), methyl (#7c) and methylene (#9c) protons of linoleic acid, which showed lower intensities in the DOSO-OO blends. The allylic and bis-allylic signal intensity ranges of the DOSO category partially overlapped with the 1st and 2nd quartiles of the HR category, whereas the methyl signal of the DOSO category completely overlapped with the 1st quartile, and the methylene signal, with the 1st, 2nd and 3rd quartiles of the HR category.

Concerning the distinction of RAO-OO and HR-OO blends, the prediction abilities achieved by the corresponding model were 82% for RAO category and 84% for HR category. The most influential variables in the model were the oleic acid signals due to the methylene (#9b), methyl (#7b) and allylic (#12a) protons of oleic acid and the methylene protons of linoleic acid (#9c), which presented higher intensities in HR-OO blends, overlapping with 3rd and 4th quartiles of the RAO category. In contrast, the intensity range of the methylene protons of saturated fatty acids (#9a) in the RAO category overlapped with the 3rd and 4th quartiles of the HR category. Refined hazelnut oils are known to present higher contents of oleic acid, similar concentrations of linoleic acid and lower amounts of saturated fatty acids than refined avocado oil (Green et al., 2020; Parcerisa et al., 2000).

With respect to the model to differentiate RAO-OO and DOSO-OO blends, the prediction abilities afforded by the model were 95% for the RAO category and 97% for the DOSO category. Among the most dominant variables in the model, the linoleic signals due to the methyl (#7c) and allylic (#12b) protons, and the methyl proton signal of squalene (#11) displayed higher intensities in RAO-OO mixtures; whereas the oleic signals due to the allylic (#12a), methylene (#9b) and methyl (#7b) protons and the signals of the methylene (#9c) and β-methylene (#10c) protons of linolenic acid showed higher intensities in DOSO-OO mixtures.

The binary classification model for the differentiation of DOSO-OO and HOSO-OO mixtures achieved prediction abilities of 95% and 100% respectively. The main signals responsible for the discrimination were the methylene protons of saturated fatty acids (#9a), which presented similar intensity values for both categories but higher variability in the HOSO-OO blends; the allylic (#12b) and methyl (#7c) protons of linoleic acid, and the vinylic protons at 5.32–5.34 ppm of unsaturated fatty acids (#24), which exhibited higher intensities in the OO blends with HOSO; and the allylic (#12a) and methylene (#9b) protons of oleic acid, and the vinylic protons at 5.35–5.38 ppm of unsaturated fatty acids (#24), which presented higher intensities and variabilities in OO blends with DOSO.

562 Taking into account the trends observed for the most influential variables in the models achieved
563 for the discrimination of DOSO-OO blends from other VO-OO mixtures, DOSO-OO blends
564 presented higher concentrations of oleic acid than OO blends with HR, RAO and HOSO, which are
565 the VOs that presented the highest contents according to literature (Green et al., 2020; Guillén et al.,
566 2003; Jović et al., 2016; Parcerisa et al., 2000); and contained lower amounts of linoleic acid than
567 the OO blends with these VOs. DOSO, which is an oil obtained from the desterolization and
568 deodorization of high oleic sunflower oil was expected to present higher contents of linoleic acid,
569 close to HOSO. However, DOSO blends contained even lower amounts of linoleic acid than the
570 blends with HR and RAO. These results evidenced that during the desodorization and/or
571 desterolization process the fatty acid profile of the oil was altered resulting in lower linoleic acid
572 contents and higher oleic acid contents. It is already known that during this raffination processes
573 drastic conditions are used that leads to olefinic degradation of sterols, the isomerization of
574 squalene and linoleic and linolenic acid, among other changes in the chemical composition of the
575 oil (Aparicio et al., 2013; Grob et al., 1994).

576 ***3.1.4. PLS-R models to determine the percentage of vegetable oil in a blend with olive oil***

577 After the VO contained in the blend with OO is identified using the classification models in the
578 decision tree scheme (Figure 1 and Figure S2 in the supplementary material), a PLS regression
579 model determines the percentage of VO contained in the mixture with OO. Thus, PLS-R models for
580 binary mixtures of RPOO, CO, HOSO, NTSO, DOSO, RAO or HR with OO were successfully
581 built (PLS-R models 13–27 in Table 5). The full range of % VO in the OO blend was divided in
582 smaller sub-ranges and a PLS-R model was developed for each one. These regression models
583 provided more accurate predictions than the PLS-R models constructed for the full range of % VO.
584 The most influential variables on the regression models for each VO-OO blend corresponded to
585 those observed for the classification models for that blend, and therefore, were explained by the
586 characteristic composition in fatty acids, triacylglycerides and squalene of the oils present in the

587 blend. The coherence in the chemical interpretation of the B regression coefficients supported the
588 robustness of the models attained.

589 All regression models disclosed for OO blends presented excellent precisions; most of the models
590 yielded R^2 values >0.99, except for the low % range models of HOSO and RAO with R^2 values of
591 0.97 and 0.93–0.96 respectively. The models permitted the quantification of the % VO in OO with
592 accuracies under 5% R-RMSEP for contents of ≥5% RPOO, ≥6% CO, ≥10% HR, ≥16% DOSO,
593 ≥16% HOSO, ≥9% NTSO and ≥31% RAO; 5–15% R-RMSEP for contents of 2–5% RPOO, 2–6%
594 CO, 3–10% HR, 5–16% DOSO, 7–16% HOSO, 3–9% NTSO and 5–31% RAO; and 15–20% R-
595 RMSEP for contents of 2–3% HR, 4–5% DOSO, 5–7% HOSO, 2–3% NTSO and 4–5% RAO.

596 As reported in section 3.1.1, the classification abilities of the PLS-DA models to identify RAO-OO
597 blends at low RAO percentages became better when pure OO samples were excluded of the dataset
598 in the modelling step. This fact indicated that the samples of 2% RAO in OO and pure OO were
599 close to the boundary and therefore misclassified. Thus, the limit of detection was established in the
600 range between 2% and 4% RAO in OO. Indeed, RAO and OO contains similar amounts of saturated
601 fatty acids, linoleic and linolenic acids, and RAO presents relative high contents of oleic acid
602 respect to most of the other VOs blended with OO. The detection limits of the models to verify the
603 presence of RPOO, CO, HOSO, NTSO, DOSO and HR in OO were under 2% VO in the blend.
604 These results are alike or outperform those reported in literature using NMR or other analytical
605 techniques. In previous NMR studies, high field ^{1}H-NMR detected the adulteration of refined
606 hazelnut oil in olive oil at a proportion of 10% using linear discriminant analysis (Mannina et al.,
607 2009), and even at 8% using ^{1}H and ^{13}C-NMR and artificial neural networks (García-González et
608 al., 2004), or as low as 1% using ^{1}H and ^{31}P-NMR and canonical discriminant analysis or
609 classification binary trees (Agiomyrgianaki et al., 2010). ^{13}C-NMR and discriminant data analysis
610 distinguished several VOs such as sunflower, avocado, soybean and hazelnut oils at 5% in OO
611 (Guyader, Thomas, Portaluri, Jamin, Akoka, Silvestre, et al., 2018). Other approaches that detected
612 contents of 1–4% sunflower, soybean and corn oils in olive oil were based on voltammetric

fingerprinting (Tsopelas, Konstantopoulos, & Kakoulidou, 2018), mass spectrometry fingerprinting (Sánchez-Hernández, Nozal, Marina, & Crego, 2012), fluorescence spectroscopy (Tan, Li, Jiang, Shi, Xiao, Jia, et al., 2018), Raman spectroscopy (Philippidis, Poulakis, Papadaki, & Velegrakis, 2017), or the determination of the composition of fatty acids, sterols, triglycerides and different chemical parameters and ratios (Contiñas, Martínez, Carballo, & Franco, 2008; Christopoulou et al., 2004; Jabeur et al., 2014; Monfreda et al., 2012; Yang et al., 2013). The analysis of the volatile profile managed to disclose hazelnut oil in olive oil in level as low as 5% (Mildner-Szkudlarz & Jeleń, 2008). FTIR or triacylglyceride fingerprints allowed the detection of high oleic sunflower oil in levels as low as 10% (De La Mata-Espinosa et al., 2011; Jović et al., 2016). The analysis of sterols and their degradation products were proposed to determine the adulteration of edible oils with desterolized sunflower oil, but the limit of detection of this approach was not determined (Biedermann, Grob, Mariani, & Schmidt, 1996; Grob et al., 1994). To the authors' knowledge, no data was found in literature regarding the adulteration of OO with avocado oil, neither virgin nor refined.

3.2. Prediction of blends of oil of the 'virgin olive oil' or 'olive oil' categories and other vegetable oil

The composition of thirty six blind oil samples provided within the OLEUM project and eight commercial oils was predicted by the binary classification models and regression models following the complete decision tree scheme (Figure S1 and S2 in the supplementary material). Table S6 (supplementary material) gathers for each sample: *i*) the PLS-DA models used, and the corresponding classification predictions regarding the category of the olive oil, the VO and the low/high level of VO in the blend; *ii*) % VO in the blend determined by the PLS-R model built for the type of blend (VO-VOO or VO-OO) previously predicted by the classification models; and *iii*) the predictions of the complementary PLS-DA models, such as PLS-DA models 74–77 (Table 4), which are specific binary classification models to distinguish mixtures of two particular VOs.

638 Most of the samples were predicted satisfactory according to the description provided. Indeed, the
639 category of olive oil, i.e. VOO or OO, the particular VO and the percentage of VO in the oil sample
640 were accurately determined. All 'legal' mixtures of VOO or OO with 40–60% NTSO or HOSO, all
641 the blends of RPOO-OO and HV-VOO (5–30% VO), and the blends of EVAO-VOO and HR-OO
642 with ≥10% VO were correctly identified and the percentage of VO properly figured out. Only blind
643 samples 16, 17 and 19 were predicted to present slightly higher % VO in VOO, and blind sample
644 26, scarcely lower % HR in OO than those given in the description. The DOSO-OO blends were
645 satisfactory determined by the classification and regression models achieved for these mixtures; just
646 for blind sample 36, the % DOSO in OO was barely lower than predicted. The blend of 10% DOSO
647 in OO (blind sample 34) was confused with mixtures of 2–11% of HOSO in OO. For the blend of
648 5% EVAO in VOO (blind sample 13), the VO contained was not recognised by any of the
649 classification models, but the percentage of VO was within the calibration range of the regression
650 models made for EVAO-VOO, HOSO-VOO and HR-VOO blends. Indeed, the EVAO-VOO model
651 predicted correctly the percentage of EVAO in the mixture. The blend of 5% HR in OO (blind
652 sample 25) was not detected by any of the HR-OO classification models but it was by the RAO-OO
653 models. The RAO-OO blends at different proportions (blind samples 21–24) were identified by the
654 classification models built for both RAO-OO and DOSO-OO, and the % VO in OO determined by
655 the regression models constructed for both mixtures. However, the PLS-DA model 76 (Table 4),
656 which distinguishes these two mixtures, predicted satisfactorily that the blends contained RAO,
657 except for the mixture of 10% RAO in OO. Despite this, the RAO-OO regression model calculated
658 accurately the content of RAO in the blends of 10%, as well as 20% RAO in OO. The error in the
659 prediction of % RAO for the blend of 30% RAO in OO was 15.7%, higher than expected. The
660 blends of 5% RAO in OO was close to the detection limit, therefore the corresponding regression
661 model was not able to predict it properly.

662 Regarding the commercial oils analysed, they were declared to be mixtures of vegetable oils or
663 NTSO with EVOO or VOO. Samples 37 and 38 were confirmed to contain VOO, whereas sample

44 was classified as an OO blend. Furthermore, these 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. From the predictions of the classification and regression models of the decision tree scheme, it could be infer that most of these samples (39, 41–43) presented close composition to blends of 50% CO in OO or to pure HR oil. Taking into account the content of oleic acid, the main fatty acid in rapeseed, CO and HR oils, which is around 60%, 30% and 80% respectively, according to literature (Guillén et al., 2003; Jiang et al., 2018; Vigli et al., 2003; Yang et al., 2013), it may be deduced that those samples, if containing rapeseed oil, have about 25% rapeseed oil in OO, as specified in the label. Sample 40 was identified by all classification models as NTSO-OO blends, and predicted to have 93% NTSO. In order to predict blends of rapeseed oil with VOO and OO, these type of blends must be included in the dataset used to build the models of the decision tree scheme, as well as any other VO that could be of interest so as to certify the percentage of VO in VOO or OO blends or detect olive oil adulterations.

4. Conclusion

^{1}H-NMR fingerprinting of olive oils coupled to multivariate data analysis provides chemical tools to detect blends between oils of the ‘olive oil’ category and vegetable oils, and quantify the percentage of each oil in the mixture, as reported for oil blends with the ‘virgin olive oil’ category in the first part of the present study (Alonso-Salces et al., 2020). ^{1}H-NMR spectral data of oils of the ‘olive oil’ category and their mixtures with refined palm olein oil, corn oil, sunflower oil, high oleic sunflower oil, desterolized high oleic sunflower oil, refined avocado oil and refined hazelnut oil, was used to optimized and validated binary classification and regression models by PLS-DA and PLS-R respectively. These PLS-DA and PLS-R models were arranged in a decision tree scheme in order to determine the composition of an oil sample. Satisfactory, robust and stable

690 classification models were achieved, and excellent precisions and acceptable accuracies were
691 afforded by the regression models developed for the determination of the percentage of VO in the
692 OO blends. Moreover, the reliability of both classification and regression models was supported by
693 the chemical interpretation of the most influential variables in the models. The percentage of the
694 VO in the OO blend is determined with accuracies under the 20% of R-RMSEP for contents as low
695 as 2% RPOO, CO, NTSO or HR, 4% DOSO or RAO and 5% HOSO in OO. The limits of the
696 detection were under 2% RPOO, CO, HOSO, NTSO, DOSO or HR and 2–4% RAO in OO.

697 The complete stepwise strategy based on ^{1}H-NMR fingerprinting of an oil sample in combination
698 with chemometrics and proposed to determine the content of mixtures of oils of the 'virgin olive
699 oil' or 'olive oil' categories and vegetable oils allow to (*i*) confirm the presence of VOO or OO in
700 an oil sample; (*ii*) discriminate between pure olive oils and their blends with VOs to a certain
701 extent, given by the detection limit disclosed for each VO; (*iii*) identify the VO in the blend with
702 VOO or OO; (*iv*) distinguish between blends made with different VOs and VOO or OO, or (*v*) with
703 the same VO at different concentrations; and (*vi*) determine the percentage of VO blended with
704 VOO or OO. The performance and effectiveness of the proposed strategy was tested predicting
705 blind samples, which confirmed its feasibility to support Reg. (EU) 29/2012. Further studies should
706 be carried out with larger balanced sample sets that cover the variability of olive oils of both
707 categories (VOO and OO) and the vegetable oils of interest. The different possible sources of
708 variability, i.e. varieties of each botanical species, agronomical and climatic conditions, harvests
709 and geographical origins among others, should be considered. The implementation of this approach
710 requires the development of a databank of ^{1}H-NMR fingerprints of oils legally blended or submitted
711 to adulteration, as well as of the adulterants and their blends, representative of the natural oil
712 variability and compositional differences, in order to guarantee robust models for both
713 authentication and fraud detection. It is worth to note that this requirement is feasible in practice
714 since among the objectives of the OLEUM project are to develop the OLEUM databank, an online
715 integrated quality assurance database of olive oil analytical methods and chemical data, as well as

716 the OLEUM Network of a worldwide community of proficient analytical laboratories involved in
717 olive oil analysis, which can also contribute to feeding and updating the databank over time.

718 **Acknowledgments**

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

728 **References**

729 Agiomyrgianaki, A., Petrakis, P. V., & Dais, P. (2010). Detection of refined olive oil adulteration
730 with refined hazelnut oil by employing NMR spectroscopy and multivariate statistical
731 analysis. *Talanta, 80*(5), 2165-2171. https://doi.org/10.1016/j.talanta.2009.11.024.
732 Al-Ismail, K. M., Alsaed, A. K., Ahmad, R., & Al-Dabbas, M. (2010). Detection of olive oil
733 adulteration with some plant oils by GLC analysis of sterols using polar column. *Food*
734 *Chemistry, 121*(4), 1255-1259. https://doi.org/10.1016/j.foodchem.2010.01.016.
735 Alonso-Salces, R. M., Berrueta, L. A., Quintanilla-Casas, B., Vichi, S., Tres, A., Collado, M. I., et
736 al. (2020). Stepwise strategy based on ^{1}H-NMR fingerprinting in combination with
737 chemometrics to determine the content of vegetable oil mixtures. Part I: Blends with oils of
738 the "virgin olive oil" category. *Food Chemistry*, submitted for publication.
739 Alonso-Salces, R. M., Segebarth, N., Garmón-Lobato, S., Holland, M. V., Moreno-Rojas, J. M.,
740 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. *European Journal of Lipid Science and Technology*, *117*(12), 1991-2006. https://doi.org/10.1002/ejlt.201400243.

Aparicio, R., & Harwood, J. (2013). *Handbook of olive oil: Analysis and properties* (2nd ed.). USA: Springer US.

Berrueta, L. A., Alonso-Salces, R. M., & Héberger, K. (2007). Supervised pattern recognition in food analysis. *Journal of Chromatography A*, *1158*(1-2), 196-214. https://doi.org/10.1016/j.chroma.2007.05.024.

Biedermann, M., Grob, K., Mariani, C., & Schmidt, J. P. (1996). Detection of desterolized sunflower oil in olive oil through isomerized Δ7-sterols. *Zeitschrift fur Lebensmittel - Untersuchung und -Forschung*, *202*(3), 199-204. https://doi.org/10.1007/BF01263540.

Conte, L., Bendini, A., Valli, E., Lucci, P., Moret, S., Maquet, A., et al. (2019). Olive oil quality and authenticity: A review of current EU legislation, standards, relevant methods of analyses, their drawbacks and recommendations for the future. *Trends in Food Science & Technology*, in press. https://doi.org/https://doi.org/10.1016/j.tifs.2019.02.025.

Contiñas, A., Martínez, S., Carballo, J., & Franco, I. (2008). Detection of contaminations and/or adulterations of the extra virgin olive oil with seeds oils (sunflower and soybean) and olive pomace oil. *Grasas y Aceites*, *59*(2), 97-103. https://doi.org/10.3989/gya.2008.v59.i2.496.

Christopoulou, E., Lazaraki, M., Komaitis, M., & Kaselimis, K. (2004). Effectiveness of determinations of fatty acids and triglycerides for the detection of adulteration of olive oils with vegetable oils. *Food Chemistry*, *84*(3), 463-474. https://doi.org/10.1016/s0308-8146(03)00273-5.

De La Mata-Espinosa, P., Bosque-Sendra, J. M., Bro, R., & Cuadros-Rodríguez, L. (2011). Olive oil quantification of edible vegetable oil blends using triacylglycerols chromatographic fingerprints and chemometric tools. *Talanta*, *85*(1), 177-182. https://doi.org/10.1016/j.talanta.2011.03.049.

De la Mata, P., Dominguez-Vidal, A., Bosque-Sendra, J. M., Ruiz-Medina, A., Cuadros-Rodríguez, L., & Ayora-Cañada, M. J. (2012). Olive oil assessment in edible oil blends by means of ATR-FTIR and chemometrics. *Food Control, 23*(2), 449-455.

EC. (2012). European Commission Regulation 29/2012 on marketing standards for olive oil. *Official Journal of the European Union, January 13, 2012*, 12-21.

EC. (2020). European Commission website. Food Fraud. https://ec.europa.eu/knowledge4policy/food-fraud-quality/food-fraud-data-bases_en, 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)).

Fernandes, G. D., Gómez-Coca, R. B., Pérez-Camino, M. d. C., Moreda, W., & Barrera-Arellano, D. (2017). Chemical Characterization of Major and Minor Compounds of Nut Oils: Almond, Hazelnut, and Pecan Nut. *Journal of Chemistry, 2017,* Article 2609549. https://doi.org/10.1155/2017/2609549.

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-Ariza, J. L., Arias-Borrego, A., García-Barrera, T., & Beltran, R. (2006). Comparative study of electrospray and photospray ionization sources coupled to quadrupole time-of-flight mass spectrometer for olive oil authentication. *Talanta, 70*(4), 859-869. https://doi.org/10.1016/j.talanta.2006.02.019.

793 Gómez-Coca, R. B., Pérez-Camino, M. d. C., Martínez-Rivas, J. M., Bendini, A., Gallina Toschi,
794 T., & Moreda, W. (2020). Olive oil mixtures. Part one: Decisional trees or how to verify the
795 olive oil percentage in declared blends. *Food Chemistry, 315,* Article 126235.
796 https://doi.org/https://doi.org/10.1016/j.foodchem.2020.126235.
797 Green, H. S., & Wang, S. C. (2020). First report on quality and purity evaluations of avocado oil
798 sold in the US. *Food Control, 116,* Article 107328.
799 https://doi.org/https://doi.org/10.1016/j.foodcont.2020.107328.
800 Grob, K., Biedermann, M., Bronz, M., & Giuffré, A. M. (1994). The Detection of Adulteration with
801 Desterolized Oils. *Lipid / Fett, 96*(9), 341-345. https://doi.org/10.1002/lipi.19940960905.
802 Guillén, M. D., & Ruiz, A. (2001). High resolution ^{1}H nuclear magnetic resonance in the study of
803 edible oils and fats. *Trends in Food Science and Technology*, *12*(9), 328-338.
804 https://doi.org/10.1016/s0924-2244(01)00101-7.
805 Guillén, M. D., & Ruiz, A. (2003). Rapid simultaneous determination by proton NMR of
806 unsaturation and composition of acyl groups in vegetable oils. *European Journal of Lipid*
807 *Science and Technology*, *105*(11), 688-696. https://doi.org/10.1002/ejlt.200300866.
808 Guyader, S., Thomas, F., Portaluri, V., Jamin, E., Akoka, S., Silvestre, V., et al. (2018).
809 Authentication of edible fats and oils by non-targeted ^{13}C INEPT NMR spectroscopy. *Food*
810 *Control, 91*, 216-224. https://doi.org/10.1016/j.foodcont.2018.03.046.
811 IOC. (2019). International Olive Council. Economic affairs & promotion unit. World olive oil
812 figures. https://www.internationaloliveoil.org/what-we-do/economic-affairs-promotion-
813 unit/#figures, Accessed data: December 1, 2020.
814 Jabeur, H., Zribi, A., Makni, J., Rebai, A., Abdelhedi, R., & Bouaziz, M. (2014). Detection of
815 chemlali extra-virgin olive oil adulteration mixed with soybean oil, corn oil, and sunflower
816 oil by using GC and HPLC. *Journal of Agricultural and Food Chemistry, 62*(21), 4893-
817 4904. https://doi.org/10.1021/jf500571n.

818 Jiang, X. Y., Li, C., Chen, Q. Q., & Weng, X. C. (2018). Comparison of ^{19}F and ^{1}H NMR
819 spectroscopy with conventional methods for the detection of extra virgin olive oil
820 adulteration. *Grasas y Aceites, 69*(2), Article e249. https://doi.org/10.3989/gya.1221172.
821 Jović, O., Smolić, T., Primožič, I., & Hrenar, T. (2016). Spectroscopic and chemometric analysis of
822 binary and ternary edible oil mixtures: Qualitative and quantitative study. *Analytical*
823 *chemistry, 88*(8), 4516-4524. https://doi.org/10.1021/acs.analchem.6b00505.
824 Kjeldahl, K., & Bro, R. (2010). Some common misunderstandings in chemometrics. *Journal of*
825 *Chemometrics*, *24*(7-8), 558-564. https://doi.org/10.1002/cem.1346.
826 Lioupi, A., Nenadis, N., & Theodoridis, G. (2020). Virgin olive oil metabolomics: A review.
827 *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life*
828 *Sciences, 1150,* Article 122161. https://doi.org/10.1016/j.jchromb.2020.122161.
829 Mannina, L., D'Imperio, M., Capitani, D., Rezzi, S., Guillou, C., Mavromoustakos, T., et al. (2009).
830 ^{1}H NMR-based protocol for the detection of adulterations of refined olive oil with refined
831 hazelnut oil. *Journal of Agricultural and Food Chemistry, 57*(24), 11550-11556.
832 https://doi.org/10.1021/jf902426b.
833 Mildner-Szkudlarz, S., & Jeleń, H. H. (2008). The potential of different techniques for volatile
834 compounds analysis coupled with PCA for the detection of the adulteration of olive oil with
835 hazelnut oil. *Food Chemistry, 110*(3), 751-761.
836 https://doi.org/10.1016/j.foodchem.2008.02.053.
837 Monfreda, M., Gobbi, L., & Grippa, A. (2012). Blends of olive oil and sunflower oil:
838 Characterisation and olive oil quantification using fatty acid composition and chemometric
839 tools. *Food Chemistry, 134*(4), 2283-2290. https://doi.org/10.1016/j.foodchem.2012.03.122.
840 Monfreda, M., Gobbi, L., & Grippa, A. (2014). Blends of olive oil and seeds oils: Characterisation
841 and olive oil quantification using fatty acids composition and chemometric tools. Part II.
842 *Food Chemistry, 145*, 584-592. https://doi.org/10.1016/j.foodchem.2013.07.141.

Montoya, C., Cochard, B., Flori, A., Cros, D., Lopes, R., Cuellar, T., et al. (2014). Genetic architecture of palm oil fatty acid composition in cultivated oil palm (*Elaeis guineensis* Jacq.) compared to its wild relative *E. oleifera* (H.B.K) Cortés. *PLoS ONE, 9*(5), Article e95412. https://doi.org/10.1371/journal.pone.0095412.

OLEUM Project. (2016). About OLEUM: Aims and Objectives. http://www.oleumproject.eu/about-oleum/aims-and-objectives, Accessed date: December 1, 2020.

Parcerisa, J., Casals, I., Boatella, J., Codony, R., & Rafecas, M. (2000). Analysis of olive and hazelnut oil mixtures by high-performance liquid chromatography-atmospheric pressure chemical ionisation mass spectrometry of triacylglycerols and gas-liquid chromatography of non-saponifiable compounds (tocopherols and sterols). *Journal of Chromatography A, 881*(1-2), 149-158. https://doi.org/10.1016/S0021-9673(00)00352-6.

Parker, T., Limer, E., Watson, A. D., Defernez, M., Williamson, D., & Kemsley, E. K. (2014). 60MHz ^{1}H NMR spectroscopy for the analysis of edible oils. *TrAC - Trends in Analytical Chemistry, 57*, 147-158. https://doi.org/10.1016/j.trac.2014.02.006.

Philippidis, A., Poulakis, E., Papadaki, A., & Velegrakis, M. (2017). Comparative Study using Raman and Visible Spectroscopy of Cretan Extra Virgin Olive Oil Adulteration with Sunflower Oil. *Analytical Letters, 50*(7), 1182-1195. https://doi.org/10.1080/00032719.2016.1208212.

Popescu, R., Costinel, D., Dinca, O. R., Marinescu, A., Stefanescu, I., & Ionete, R. E. (2015). Discrimination of vegetable oils using NMR spectroscopy and chemometrics. *Food Control, 48*, 84-90. https://doi.org/10.1016/j.foodcont.2014.04.046.

Sánchez-Hernández, L., Nozal, L., Marina, M. L., & Crego, A. L. (2012). Determination of nonprotein amino acids and betaines in vegetable oils by flow injection triple-quadrupole tandem mass spectrometry: A screening method for the detection of adulterations of olive oils. *Journal of Agricultural and Food Chemistry, 60*(4), 896-903. https://doi.org/10.1021/jf202857y.

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.

Tan, J., Li, R., Jiang, Z. T., Shi, M., Xiao, Y. Q., Jia, B., et al. (2018). Detection of extra virgin olive oil adulteration with edible oils using front-face fluorescence and visible spectroscopies. *Journal of the American Oil Chemists' Society, 95*(5), 535-546. https://doi.org/10.1002/aocs.12071.

Tsopelas, F., Konstantopoulos, D., & Kakoulidou, A. T. (2018). Voltammetric fingerprinting of oils and its combination with chemometrics for the detection of extra virgin olive oil adulteration. *Analytica Chimica Acta, 1015*, 8-19. https://doi.org/10.1016/j.aca.2018.02.042.

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.

Vlahov, G. (2009). ^{13}C nuclear magnetic resonance spectroscopic detection of the adulteration of extra virgin olive oils extracted from different cultivars with cold-pressed hazelnut oil. *Journal of AOAC International, 92*(6), 1747-1754. https://doi.org/10.1093/jaoac/92.6.1747.

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 scheme constituted of PLS-DA classification and PLS-R regression models to determine the composition of binary mixtures of oils of the 'olive oil' category 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; RAO, refined avocado oil; RPOO, refined palm olein oil; CO, refined corn oil.

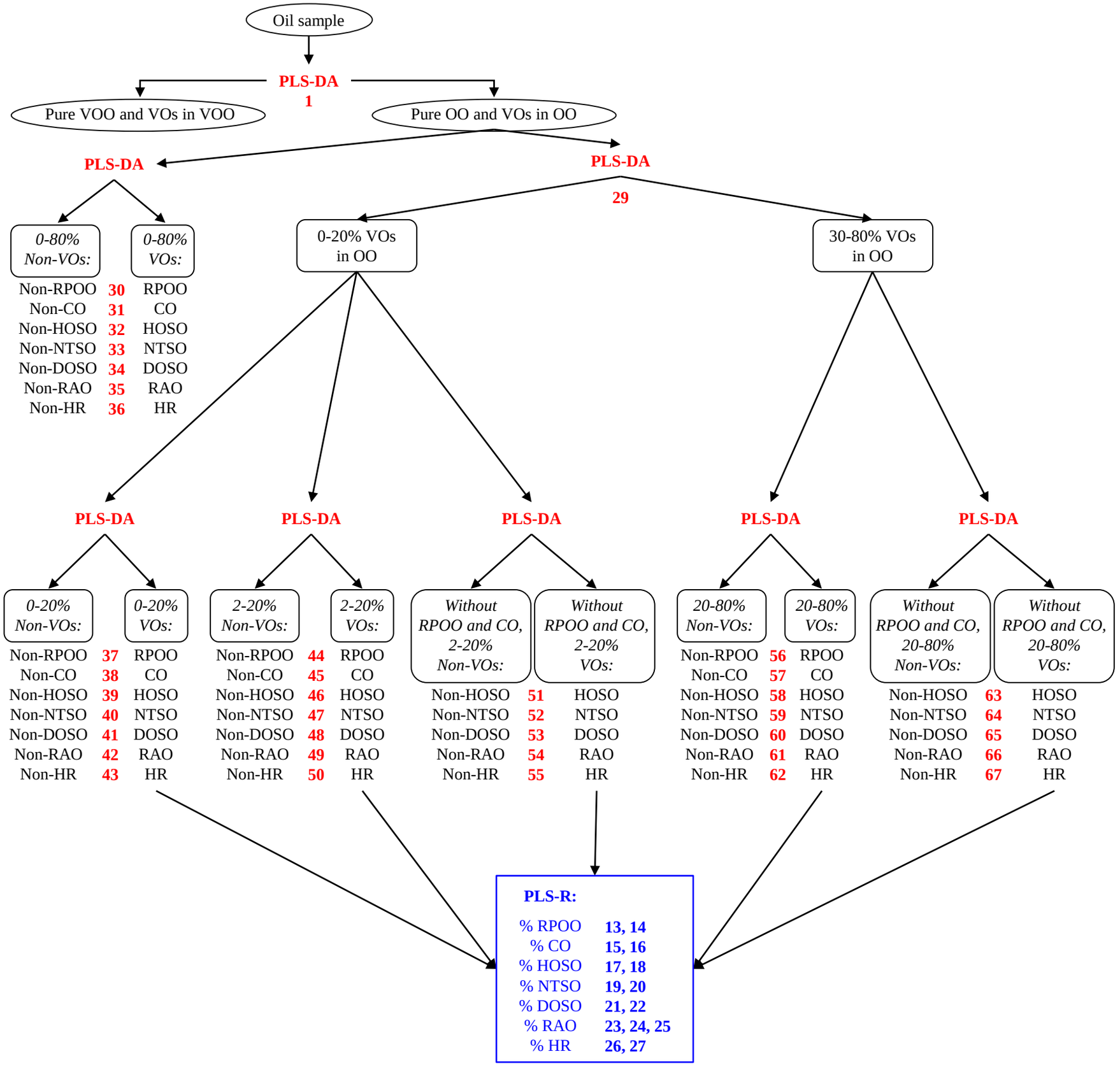
Supplementary material

Figure S1. Decision tree scheme 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.

Figure S2. Decision tree scheme 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(s)

Click here to access/download;Figure(s);RMAS_PART2_Figure-1.pptx



Tables

Table 1

PLS-DA models to discriminate between pure and blended oils containing oils of the 'olive oil' or 'virgin 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
29	Pure & blend OO	16	0.4388	0–20% VOs in OO	0	184	0.52	97	96
				30–80% VOs in OO	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 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]	Class code	n	p	%R	%P
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: 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 20–80% vegetable oil in olive oil.[1]

PLS-DA model	Data	PLS-comp	Boundary	Class[2]	Class code	n	p	%R	%P
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: 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-DA models to discriminate between 'legal' and 'illegal' blends of olive oil and vegetable oils, between 'legal' blends of OO with NTSO and HOSO, between OO blends with DOSO and HR, between OO blends with RAO and HR, between OO blends with RAO and DOSO, and between OO blends of with DOSO and HOSO.[1]

PLS-DA model	Data	PLS-comp	Boundary	Class	Class code	n	p	%R	%P
70[2]	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]	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
74[3]	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[3]	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[3]	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[3]	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: '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.

[3] Samples contained in each class: 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 5

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	R2-val	RMSEP (% VO)
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.

Stepwise strategy based on ^{1}H-NMR fingerprinting in combination with chemometrics to determine the content of vegetable oils mixtures. Part II: Blends with the "olive oil" category

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], E. Valli[6], A. Bendini[6], T. Gallina Toschi[6], J. M. Martínez-Rivas[7], W. Moreda[8], B. Gallo[2]

Supplementary material: Figures

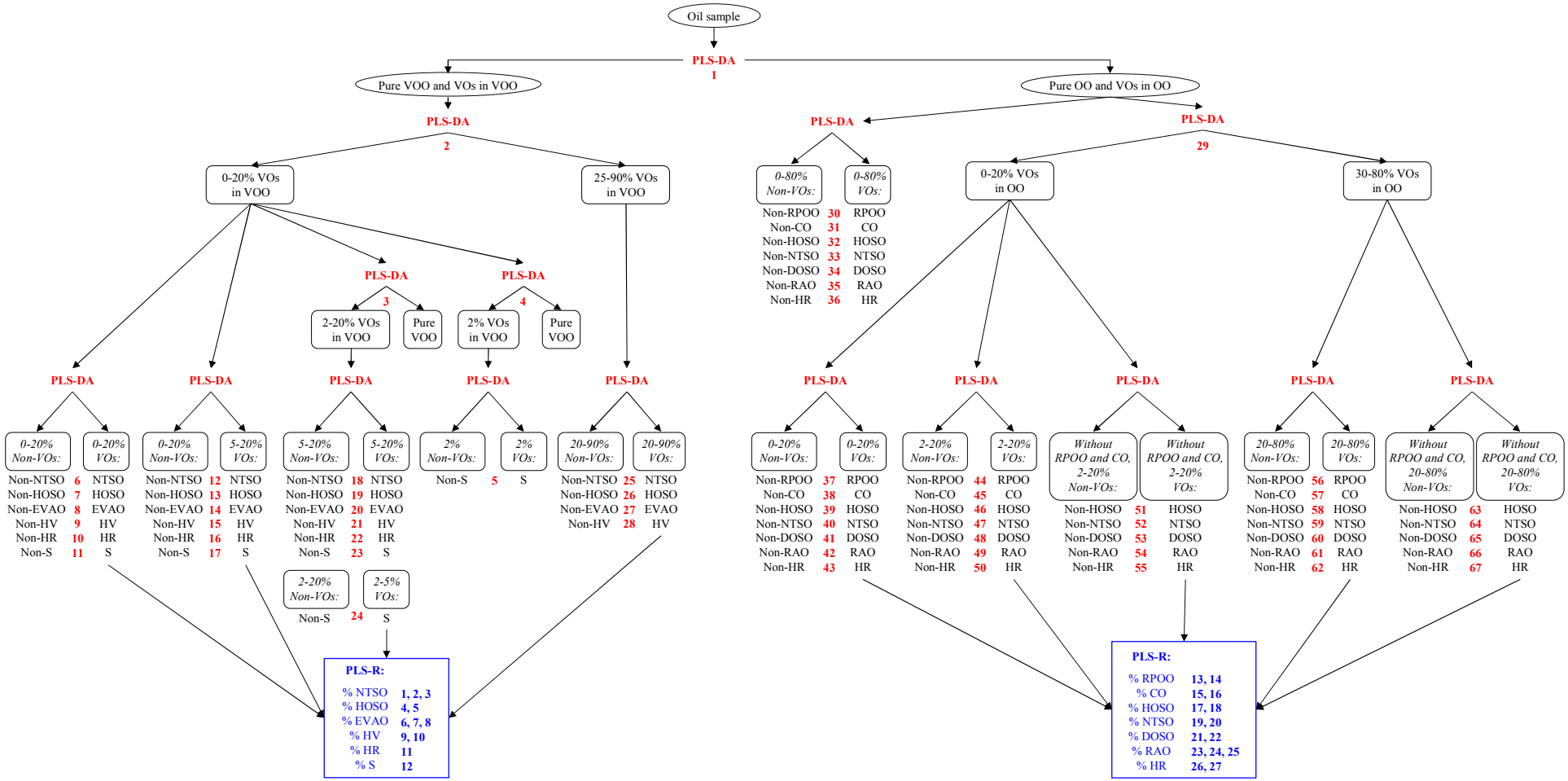


Figure S1. Decision tree scheme constituted of PLS-DA classification and PLS-R regression models to determine the composition of binary mixtures of oils of the 'virgin olive oil' or 'olive oil' categories and other vegetable oils. Abbreviations: VOO, virgin olive oil; OO, olive oil; VO, vegetable oil; 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.

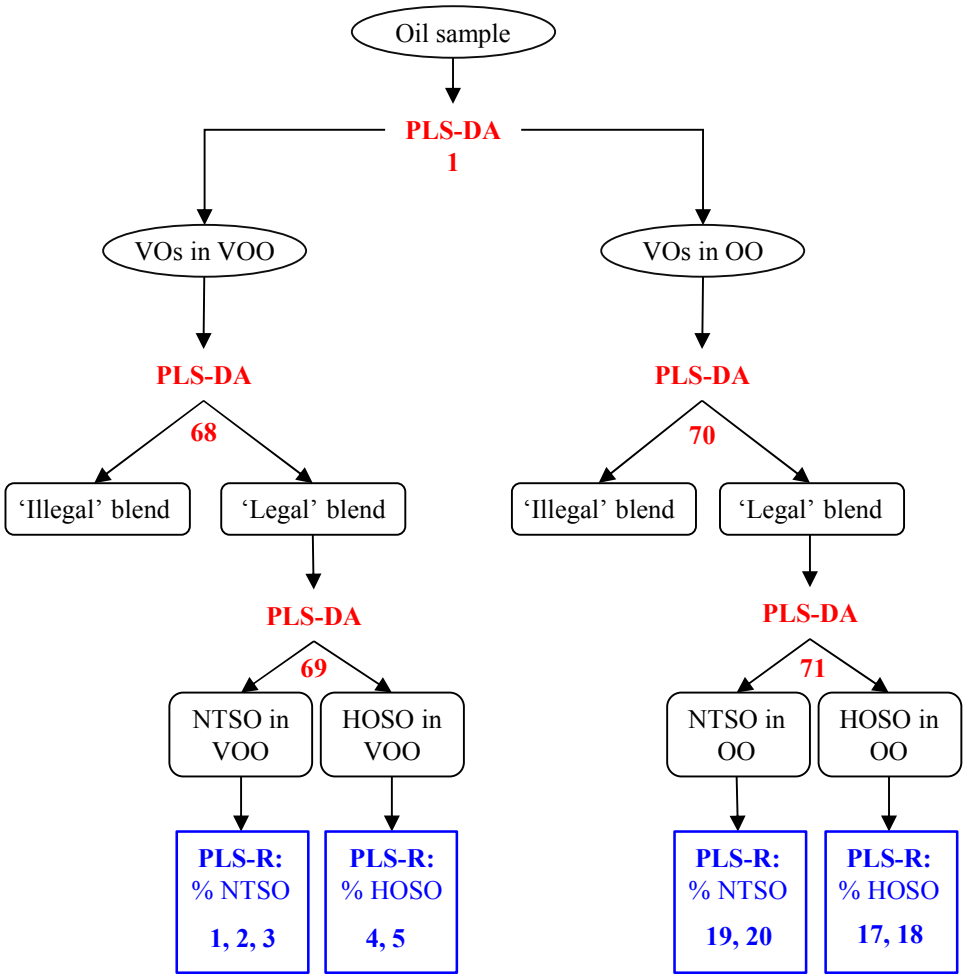


Figure S2. Decision tree scheme constituted of PLS-DA classification and PLS-R regression models for a case study: Discrimination between 'legal' (containing NTSO or HOSO) and 'illegal' (not containing NTSO or HOSO) blends, and determination of % NTSO or HOSO in binary mixtures with oils of the 'virgin 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.

Stepwise strategy based on ^{1}H-NMR fingerprinting in combination with chemometrics to determine the content of vegetable oils mixtures. Part II: Blends with the "olive oil" category

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], E. Valli[6], A. Bendini[6], T. Gallina Toschi[6], J. M. Martínez-Rivas[7], W. Moreda[8], 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 2–80% 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: 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; 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; 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 S3

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.

[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 S4

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 S1.

[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 S5

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 S1.

[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 S6

Prediction of the composition of blind oil samples using the decision tree schemes in Figures S1 and S2 in the supplementary material and the complementary PLS-DA models.[1,2]

	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'	4	HOSO-VOO	3.9 ± 6.8	*EVOO + EVAO, 95:5*
	73	2-5% HR in VOO	6	EVAO-VOO	6.5 ± 2.1	
			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			
	Models applied	**Predictions**	**Predicting model**	**Blend**	**% VO**	**Description**
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: EVOO + NTSO, 65:35*
38	1, 2, 25-28, 68, 69	VOO; high; NTSO; legal-NTSO	3	NTSO-VOO	104.9* ± 1.9	*Label: VOO + Vegetable oil*
39	1, 30-36, 29, 37-67, 70, 71	OO; low; CO, RAO, HR; 'illegal'	16	CO-OO	56.4 ± 0.6	*Label: EVOO + Rapeseed oil*
	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: VOO + Rapeseed oil, 80:20*
41	1, 30-36, 29, 37-67, 70, 71	OO; low; CO, RAO, HR; 'illegal'	16	CO-OO	52.0 ± 0.6	*Label: VOO + Rapeseed oil, 75:25*
	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: VOO + Rapeseed oil, 75:25*
	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: VOO + Rapeseed oil, 80:20*
	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: VOO + Vegetable oil, 80:20*

[1] 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.

[2] Complementary PLS-DA models: PLS-DA models 74–77 in Table 4, and PLS-DA models 72–73 in Table S1 in Alonso-Salces et al. (2020).

* Extrapolated results (outside the calibration range).