Chemometric Discrimination of the Varietal Origin of Extra Virgin Olive Oils: Usefulness of 13C Distortionless Enhancement by Polarization Transfer Pulse Sequence and 1H Nuclear Magnetic Resonance Data and Effectiveness of Fusion with Mid-Infrared Spectroscopy Data

Astrid Maléchaux, Raquel Garcia, Yveline Le Dréau, Arona Pires, Nathalie Dupuy, Maria Joao Cabrita

Chemometric Discrimination of the Varietal Origin of Extra Virgin Olive Oils: Usefulness of ^{13}C Distortionless Enhancement by Polarization Transfer Pulse Sequence and ^{1}H Nuclear Magnetic Resonance Data and Effectiveness of Fusion with Mid-Infrared Spectroscopy Data

Astrid Maléchaux, Raquel Garcia, Yveline Le Dréau,* Arona Pires, Nathalie Dupuy, and Maria Joao Cabrita

ABSTRACT: The label authentication of monovarietal extra virgin olives is of great relevance from a socio-economical point of view. This work aims to gain insights into the prediction of the varietal origin of extra virgin olive oil (EVOO) samples obtained from single olive cultivars, French cultivars Olivière, Salonenque, and Tanche and Portuguese cultivars Blanqueta, Carrasquenha, and Galega Vulgar, collected in 2016−2017 and 2017−2018 harvest seasons. To pursue this study, spectroscopic approaches based on one-dimensional nuclear magnetic resonance (1D NMR) spectroscopy, namely, ^{1}H and ^{13}C NMR distortionless enhancement by polarization transfer (DEPT) 45 pulse sequence, and Fourier transform mid-infrared spectroscopy (FT-MIR) are used in combination with partial least square discriminant analysis (PLS1-DA). The results obtained by PLS1-DA models using ^{1}H and ^{13}C NMR DEPT 45 data are compared to those of PLS1-DA models using MIR data. The application of a control chart method allows for the optimization of the interpretation of the PLS1-DA results, and an efficient two-step strategy is proposed to improve the discrimination of the six studied cultivars. Then, NMR and MIR data are combined by either a mid- or high-level data fusion approach to further improve the discrimination. The models are also tested on samples from other cultivars to check their ability to reject varieties that were not considered in the calibration process.

KEYWORDS: *NMR, MIR, PLS-DA, control chart, data fusion, olive oil cultivar*

INTRODUCTION

Food authentication is a process that verifies that a food is in compliance with its label description, including the geographic and varietal origin, the production method, and processing technologies.[1] This topic is of great interest from both commercial and legal points of view. In fact, authentication of the origin of food products has attracted much research efforts, with special focus on the assessment of certified origin and the differentiation of varieties/cultivars. This is particularly true where olive oil is concerned. Consumers, aware of the health benefits associated with olive oil use in the diet,[2] are paying attention to authenticity issues. Because extra virgin olive oil (EVOO) has a high added value, it is also prone to adulteration practices. Thus, the development of analytical techniques for authenticity assessment is of pivotal relevance.[3]

In the field of food authentication, targeted methods, focusing on the analysis of a specific metabolite or group of metabolites, and untargeted methods, aiming for the discrimination of patterns of metabolites that may change according to several stimuli, can be used. Thus, metabolomic studies comprise two main approaches, metabolomic profiling and metabolomic fingerprint,[4] that aim to find differences between samples and to create statistical models to predict class memberships.[5] However, as a result of the great complexity and dynamic range of the metabolome, there is no single analytical instrument able to analyze the whole metabolome at once. In the case of olive oil authentication, in some studies, DNA-based approaches are considered complementary to analytical chemistry methodologies.[6] For example, single nucleotide polymorphisms and the microsatellites or single sequence repeats (SSRs) turned out to be markers of choice for olive oil traceability purposes.[7−9] However, methodologies requiring minimum sample manipulation are mandatory when dealing with authenticity testing. Therefore, the high-resolution melting (HRM) technology, which allows for the genotyping of varieties because it requires previous identification of the variants[10] might have a limited use in such an aim, even if the SSR-HRM methodology was proposed as a powerful approach to authenticate monovarietal olive oils.[11] For olive oil varietal authentication, spectroscopic techniques, which do not require any preparation of samples or simple

dilution in a suitable solvent, such as mid-infrared spectroscopy (MIR) and nuclear magnetic resonance (NMR), are among the most employed.[12] These qualitative techniques lead to multiple non-specific signals used as a fingerprint of the samples apportioned in defined classes and, consequently, require a multivariate classification approach also referred to as non-target analysis.[13]

The potential of MIR spectroscopy combined with chemometric modeling to estimate quality parameters, to detect adulterations, and to discriminate the geographical or varietal origins of olive oils, the olive oils obtained from either whole or stoned olive pastes, or also, for instance, the fresh olives from different cultivars has already been established.[14−17] With regard to NMR spectroscopy, ^{1}H NMR is a preferential method for the study of olive oil as a result of its higher sensitivity and shorter relaxation times of proton nuclei compared to less sensitive ^{13}C NMR.[18] However, the ^{13}C NMR spectrum of an olive oil shows a larger number of signals spread over a wide range of chemical shifts, thus, although it appears to be more complicated, it gives much more information than the ^{1}H NMR spectrum.[19] Moreover, the sensitivity can be partially improved using the distortionless enhancement by polarization transfer (DEPT) pulse sequence, which transfers the polarization from the highly sensitive ^{1}H nuclei to the less sensitive ^{13}C nuclei, thus enhancing signal intensities.[20] As is the case with MIR data, NMR data can be subjected to a large panoply of chemometric analyses, with the most common analyses being unsupervised principal component analysis (PCA) for data exploration seeking groups of samples and possibly highlighting specific markers[21,22] and supervised partial least square discriminant analysis (PLS-DA) for classification models discriminating the varietal or geographical origin of the oils.[18,23−25] However, these studies are mostly focused on ^{1}H NMR data, so that classification and prediction according to varietal origin have not yet been performed on ^{13}C NMR data of olive oils. Only Vlahov et al. used intensity data of triacylglycerol resonances obtained in ^{13}C NMR spectroscopy to correctly classify by linear discriminant analysis 173 olive oils from three Italian protected designation of origin.[26] Moreover, Merchak et al. demonstrated that ^{13}C insensitive nuclei enhancement by polarization transfer (INEPT) has a higher potential than ^{1}H NMR in the classification of olive oils according to the altitude of the olive plantations.[27] Recently, the varietal origin of olive oils using a NMR-based approach was successfully demonstrated for Portuguese varietal olive oils using ^{1}H and ^{13}C NMR DEPT 45 spectroscopy, combined with a multivariate statistical analysis.[28]

Furthermore, the fusion of data from different analytical techniques is expected to improve the discrimination of origin through a synergetic effect of complementary information, using multiblock predictor and multiblock response. In the literature, several strategies have been proposed and classified into low-level (or variable level), mid-level (or factor level), and high-level (or decision level) data fusion.[29] The simplest method (low level) consists in putting the descriptor blocks into the same matrix and then applying PLS analysis on the concatenated matrix. Sometimes this methodology gives good results, but the difficulty is to scale the individual blocks to obtain interpretable results. For instance, in a study conducted by Monakhova and co-workers, the fusion of ^{1}H NMR and isotope ratio mass spectrometry (IRMS) data enhanced the prediction of geographical origin and vintage year of wines but not the recognition of grape variety.[30] With regard to the fusion with infrared spectroscopy, ^{1}H NMR has been successfully combined with MIR and IRMS to differentiate between organically and conventionally grown tomatoes by linear discriminant analysis.[31] Another way is to consider each block independently at the beginning to conduct the fusion at the factor level. For instance, PCA or PLS can be applied on each block, and then the scores obtained from each block are collected to form a super matrix. This mid-level method presented by Tenenhaus and Vinzi was called H-PLS.[32] A few applications could be found in the literature. Among them, Casale et al. worked on three analytical methods and two chemometric strategies.[33] Another mid-level fusion of ^{1}H NMR with excitation−emission multidimensional fluorescence (EEM) and high-performance liquid chromatography (HPLC) data improved the discrimination of wines from three varietal origins by PLS-DA.[34] Mid-level fusion of ^{1}H NMR, MIR, NIR, and EEM data also improved the discrimination of wine vinegars from three protected designations of origin.[35] For olive oil authentication, multiblock PLS-DA models were developed from gas chromatography (GC) and MIR data sets, with and without weighting the block scores, to evaluate their performance against those of the PLS-DA models applied separately to each data set.[36] However, very few studies have combined ^{13}C NMR and MIR data. In a recent article, the prediction of crude oil properties was improved by mid-level fusion of ^{13}C and ^{1}H NMR and MIR data using PLS scores but not by low- or mid-level fusion using PCA scores.[37] Finally, few high-level data fusion approaches have been applied. For instance, Bayesian probabilistic rules were successfully used to classify musts of grapes according to their variety with MIR, ultraviolet (UV), and aroma sensor data[38] or to identify the botanical origin of honeys based on NIR, Raman, and proton transfer reaction−mass spectrometry.[39] Other studies used high-level fusion with the majority vote to improve the detection of banned dyes in culinary spices based on ^{1}H NMR and ultraviolet−visible (UV−vis) data[40] or to enhance the prediction of olive oil varietal origin with NIR and MIR data.[41] Nevertheless, to our knowledge, high-level fusion has not yet been applied to combine NMR and MIR data for the authentication of food products.

This work comprises the analysis of monovarietal EVOO from three French cultivars (Olivière, Salonenque, and Tanche) and three Portuguese cultivars (Blanqueta, Carrasquenha, and Galega Vulgar) collected during 2016−2017 and 2017−2018 harvest periods. The samples were analyzed by ^{1}H and ^{13}C NMR DEPT 45 pulse sequence and MIR spectroscopy to explore the usefulness of these approaches to discriminate the olive oils according to their varietal origin. Classification of EVOO according to the cultivar was achieved by applying the PLS1-DA algorithm, associated with the control chart for acceptability limits. A mid-level fusion strategy using PLS scores and a high-level fusion strategy using the majority vote was also applied to combine the NMR and MIR data, with the aim to improve the discrimination of varietal origin of the EVOO.

MATERIALS AND METHODS

EVOO Samples. A first group of 119 samples from six monovarietal extra virgin olive oils produced over two harvest years (2016−2017 and 2017−2018) was used for this study. The samples came from three French cultivars, Olivière (OL, n = 24), Salonenque (SA, n = 24), and Tanche (TA, n = 23), and three Portuguese

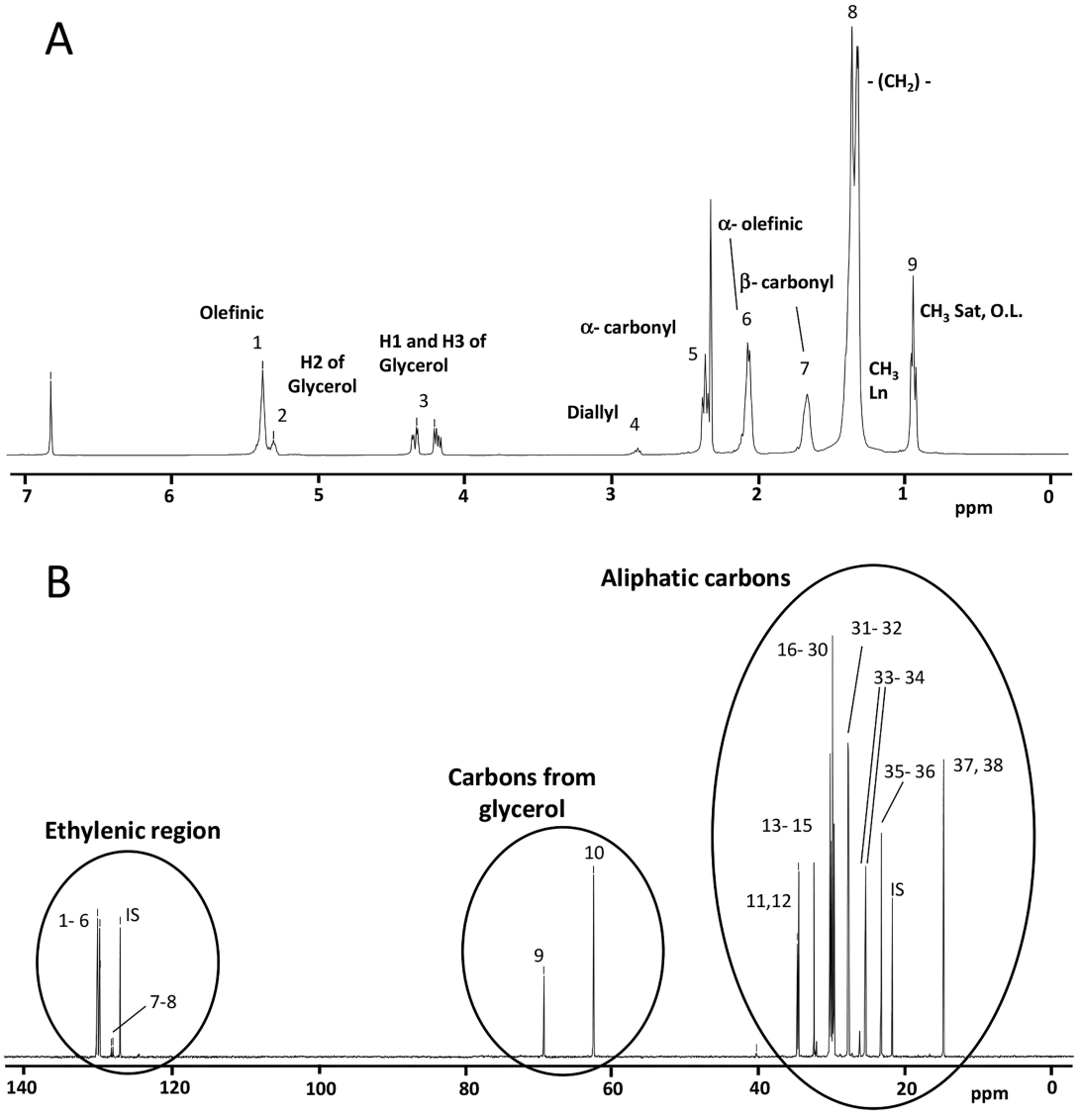

Figure 1. (A) ^{1}H NMR spectrum (400 MHz in $CDCl_3$) and (B) ^{13}C NMR DEPT 45 spectrum (100.13 MHz in $CDCl_3$) of the Portuguese olive oil Galega Vulgar cultivar (IS = internal standard).

cultivars, Blanqueta (BL, *n* = 14), Carrasquenha (CR, *n* = 14), and Galega Vulgar (GA, *n* = 20). Moreover, the prediction models for these six cultivars were then tested on a second group of 75 samples from other cultivars that were not used in the calibration and validation process (Aglandau, AG, *n* = 37; Cailletier, CA, *n* = 22; and Cobrançosa, CB, *n* = 16).

Some olive oil samples were obtained in an Abencor system immediately after harvesting. Fruits were crushed with a hammer mill; the olive paste was malaxed at 25 °C, room temperature, for 30 min in an olive paste mixer; and finally, the olive oil was separated by centrifugation. Other olive oil samples were processed by commercial olive oil mills. French olive oil samples were obtained from the producers and from the AFIDOL organization (French Interprofessional Association of Olive, Aix en Provence, France) that certify the varietal origin of samples. All orchards were located in a fairly broad geographic area with a latitude from 43° 05′ N to 44° 27′ N, a longitude from 2° 14′ E to 7° 20′ E, and an altitude from 20 to 550 m, with specific locations, cultivar by cultivar, in accordance with the geographical areas defined by the protected designation of origin (PDO) to which they contribute. The often-dry limestone soils, with basement with a variable water holding capacity, constitute plateaus or moderately high massifs from south to north; the annual sunshine is around 2700 h; and the average annual rainfall is 560 mm. Portuguese olive oil samples were obtained from the Portuguese collection of olive cultivars established at experimental olive orchard Herdade do Reguengo (INIAV, Elvas, Portugal) or from certified olive plantations. All orchards were located in a geographic area with a latitude from 37° 94′ N to 39° 06′ N, a longitude from −7° 15′ W to −8° 16′ W, and an altitude from 100 to 391 m. Soils are mostly of schist and limestone origin; the annual rainfall around 530 mm; and the annual sunshine is around 2700 h. Samples were collected immediately after processing to avoid possible undeclared mixtures with oils from other cultivars and geographical origins before bottling. All samples were stored in dark-brown glass bottles at 4 °C for up to a maximum of 6 months and then brought back at 20 °C in the dark before analyses.

Sample Preparation and NMR Experiments. Samples were analyzed according to Garcia et al. using 1D multinuclear NMR spectroscopy (^{1}H and ^{13}C NMR DEPT 45 pulse sequence), on a Bruker Advance III 400 MHz spectrometer, equipped with a wide band (BBO) observation probe at a temperature of 303 K and using TopSpin software 3.2 pl 6 for file handling.[42]

Briefly, for the sample preparation, 100 μL of olive oil and 10 μL of mesithylene (internal standard, Sigma-Aldrich) were dissolved in 500 μL of deuterated chloroform (Cambridge Isotope Laboratories, Inc.)

and were placed in a 5 mm diameter NMR tube. Each sample was subjected to two 1D NMR experiments that included ^{1}H and ^{13}C NMR DEPT 45 pulse sequence. These experiments were installed in the Bruker TopSpin 3.2 pl 6 suit, and the analysis of the samples was facilitated by the application of the ICON-NMR user interface installed within the same software suit. The free induction decay (FID) acquisition parameters for the standard single pulse test were as follows: (1) (zg30) in ^{1}H NMR, spectral width (SW) = 20.64 ppm, dummy scans (DS) = 2, number of scans (NS) = 16, acquisition time (AQ) = 4.089 s, and received gain (RG) = 10, giving a total run time of 1 min and 32 s, and (2) (zgpg30) at ^{13}C NMR DEPT 45, angle value = 45°, spectral width (SW) = 238.89 ppm, dummy scans (DS) = 4, number scans (NS) = 256, acquisition time (AQ) = 1.3631 s, and receiver gain (RG) = 2050, giving a total running time of 14 min and 46 s. Four repetitions were obtained and averaged for each sample. Peak intensities were normalized against the internal standard peak (one signal at δ 126.91 ppm). An example of ^{1}H and ^{13}C NMR DEPT 45 spectra of an olive oil obtained for this study is depicted in panels A and B of Figure 1, respectively, and the chemical shifts and functional group assignments are presented in Tables 1 and 2.

Table 1. Chemical Shifts and Proton Assignments of a ^{1}H NMR Spectrum of an Olive Oil Sample

signal	chemical shift (ppm)	functional group
H-1	5.26−5.40	−C**H**=C**H**−, all unsaturated fatty acids
H-2	5.20−5.26	>C**H**OCOR, glycerol (triacylglycerols)
H-3	4.10−4.32	−C**H**$_2$OCOR, glycerol (triacylglycerols)
H-4	2.70−2.84	=CH−C**H**$_2$−CH=, linoleyl and linolenyl
H-5	2.23−2.36	−OCO−C**H**$_2$−, all acyl chains
H-6	1.94−2.14	−C**H**$_2$−CH=CH−, all unsaturated fatty acids
H-7	1.52−1.70	−OCO−CH$_2$−C**H**$_2$−, all acyl chains
H-8	1.22−1.42	−(C**H**$_2$)$_n$, all acyl chains
H-9	0.83−0.93	−C**H**$_3$, all acids except the linolenyl group

MIR Experiments. Fourier transform mid-infrared (FT-MIR) spectra were recorded in a temperature-controlled room at 21 °C, using a Nicolet Avatar spectrometer (Thermo Scientific, Waltham, MA, U.S.A.) equipped with a Golden Gate attenuated total reflectance (ATR) accessory (Specac, Orpington, U.K.), an Ever-Glo source, a KBr/Ge beam splitter, and a nitrogen-cooled HgCdTe detector. Before each spectrum acquisition, the ATR plate was cleaned with ethanol and air was taken as a background reference. A drop of EVOO was then poured on the diamond crystal of the ATR, and its spectrum was recorded between 4000 and 700 cm^{-1} by the accumulation of 64 scans with a resolution of 4 cm^{-1} and a data spacing of 1.926 cm^{-1}. Three repetitions were obtained and averaged for each sample. A typical MIR spectrum with band assignments according to a previous paper[36] is shown in Figure 2.

Chemometrics. ^{1}H and ^{13}C DEPT 45 NMR data, resulting from the integration of the 47 peaks (i.e., the 38 ^{13}C signals plus the 9 ^{1}H signals), were normalized according to their respective maximum peaks (i.e., H-8 and C-30), and each variable was scaled by its standard deviation prior to chemometric analyses. The MIR range between 4000 and 1800 cm^{-1}, a noisy area containing non-informational or redundant absorbances, was not included in the models. Thus, only the remaining MIR range between 1800 and 700 cm^{-1} (i.e., 571 variables), containing most of the useful information, was used for this study so as not to add noise. The spectra were corrected using the standard normal variate (SNV) pretreatment before chemometric analyses, namely, exploratory analysis by PCA and development of models predicting the varietal origin of samples by PLS1-DA.

PCA was conducted with the Unscrambler X version 10.4 software (Camo Analytics). PCA is an unsupervised modeling method that allows for exploratory data analysis as it extracts information from data set and removes noise. It allows for classification of samples, by investigating similarities and differences between them. PCA transforms correlated variables into new variables, called "principal components" (PCs), uncorrelated with each other. PCA models lead to score plots and loading plots. Scores describe the variation in the samples compared to the data set, while loadings describe the correlations among the variables. PCs describe, in decreasing order, the higher variations among the samples. The first PC (PC1) contains the most information, followed by PC2, PC3, etc. Because PCs are calculated to be orthogonal to others, each PC can be interpreted independently. That allows for the visualization of the repartition of the samples and the correlations between variables.[43]

Models predicting the varietal origin of the samples by PLS1-DA were then developed using MATLAB, version R2014b (MathWorks). For this purpose, two-thirds of the samples from each cultivar (n = 80) were randomly selected and used as a calibration set (i.e., training

Table 2. Chemical Shifts and Functional Groups Assigned from a ^{13}C DEPT 45 Spectrum of an Olive Oil Sample[a]

signal	chemical shift (ppm)	functional group	signal	chemical shift (ppm)	functional group
C1	130.20	L13 $\alpha\beta$	C20	29.57	O14 $\alpha\beta$
C2	130.01	O 10 $\alpha\beta$	C21	29.52	S6 α
C3	129.93	L9 α	C22	29.41	S15 α
C4	129.83	L9 β	C23	29.36	L15 $\alpha\beta$
C5	129.71	O9 α	C24	29.31	O15,13 $\alpha\beta$
C6	129.69	O9 β	C25	29.23	S5 α
C7	128.10	L10 $\alpha\beta$	C26	29.21	O, L5 β
C8	127.92	L12 $\alpha\beta$	C27	29.14	O, L5 α
C9	68.91	Gl β	C28	29.12	S4 α-O, L6 α, β-O, L4 α
C10	62.10	Gl α	C29	29.08	O, L4 β
C11	34.20	O, L 2 β	C30	29.02	unknown
C12	34.04	S2 α	C31	27.25	O11 $\alpha\beta$
C13	31.95	S16 α/O16 $\alpha\beta$	C32	27.20	L8 $\alpha\beta$, O8 $\alpha\beta$
C14	31.82	unknown	C33	25.65	L11 $\alpha\beta$
C15	31.56	L16 $\alpha\beta$	C34	24.87	O, L3 $\alpha\beta$, S3 α
C16	29.80	O12 $\alpha\beta$	C35	22.72	S17 α, O17 $\alpha\beta$
C17	29.74	unknown	C36	22.61	L17 $\alpha\beta$
C18	29.70	unknown	C37	14.13	S18 α, O18 $\alpha\beta$
C19	29.66	unknown	C38	14.09	L18 $\alpha\beta$

[a]Abbreviations: S, stearoyl; O, oleoyl; L, linoleoyl; Ln, linoleolenyl; and Gl, glycerol.

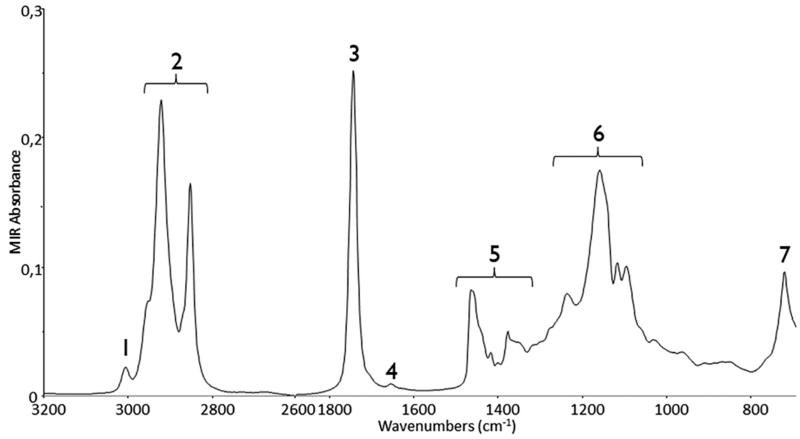

Figure 2. Example of a MIR spectrum from EVOO with identification of the bands: 1, ═C−H *cis* stretching; 2, C−H stretching; 3, C═O stretching; 4, C═C *cis* stretching; 5, C−H bending; 6, C−O and C−C bending; and 7, C−H bending (long chains).

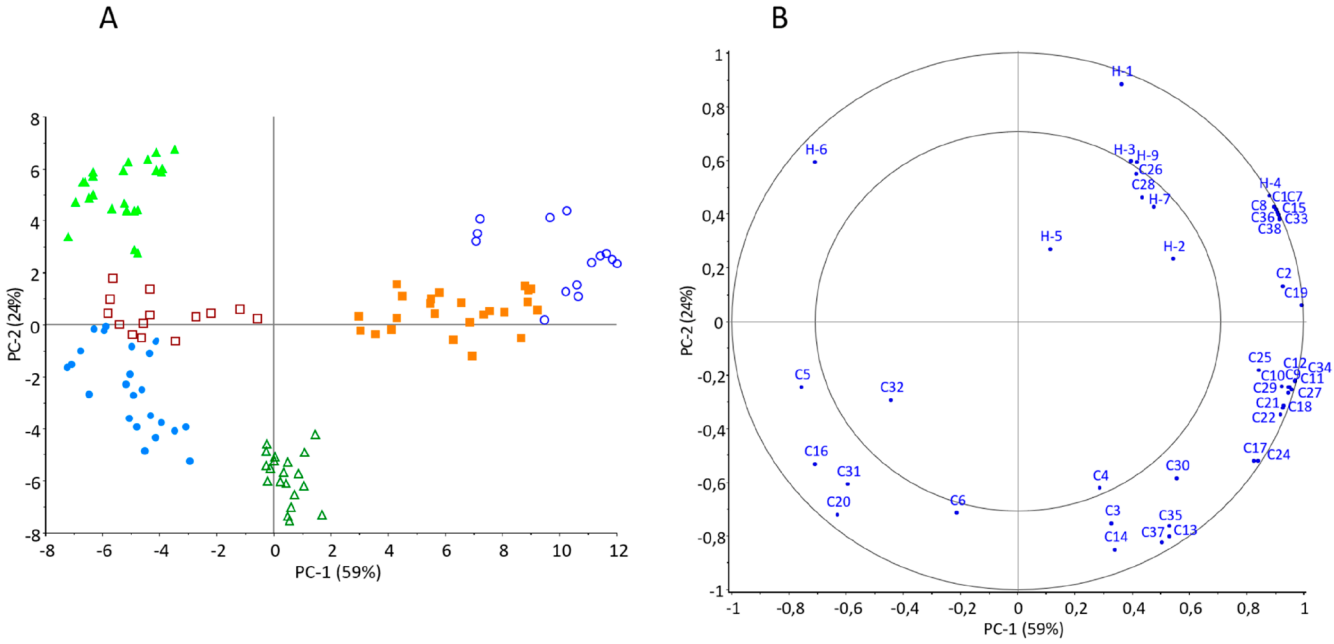

Figure 3. Plots of the (A) scores and (B) loadings for the first two PCs of the PCA analysis on the NMR data of monovarietal olive oil samples (●, Olivière; ■, Salonenque; ▲, Tanche; ○, Blanqueta; □, Carrasquenha; and △, Galega Vulgar).

samples) to build the models, and the remaining third (n = 39) served as a prediction set (i.e., test set) to test the performance of the models. PLS-DA is a versatile algorithm that can be used for the classification task. It is a supervised method that has been shown that it often outperforms a class-modeling method (for example, SIMCA) in the correct classification rates.[44] The first step in PLS-DA modeling is recoding the categorical variables (i.e., ordinal or nominal) into continuous variables (i.e., numerical) handleable by the PLS algorithm, historically used to regression tasks. Typically, 0 and 1 are used to indicate "out-class" and "in-class", respectively. These dummy y variables are employed as output variables by the PLS1-DA algorithm (that models one class at a time) that associates them to the input (X) data (i.e., spectral data) to construct the PLS latent variables (LVs, i.e., new axes). By maximization of the covariance between X and y, the weight vector (w) is estimated and then X scores (t), X loading (p), and Y loading (q). After that, the resulting w, p, and q are used to estimate the regression coefficient (b). Therefore, the first PLS LV is established. Then, the residuals X (res$X$) and y (res$y$) of the first PLS LV become the input data (X) and output data (y), respectively, for constructing the second PLS LV. The procedures are repeated n times if n PLS LVs are required to construct the desired prediction model. The training samples are used for the construction of the n PLS LVs and a regression coefficient matrix, B. For prediction, the test set (X test) is reduced into the new dimensions (i.e., n LVs) via B to produce the predicted values. The perfect predicted values are supposed to be "1" to indicate if samples belong or "0" if they do not belong to the modeled cultivar, but because the studied samples were subject to annual variations resulting from uncontrollable weather and farming conditions, the predicted values take on any values between 0 and 1. It is thus important to define the thresholds of acceptance of predicted values according to the variability of the samples within their varietal origin class.[45] Therefore, thresholds were built as a control chart, and warning limits and control limits were established as confidence intervals at 95 and 99%, respectively, around the mean calibration scores, independent for each modeled cultivar.[46] Following this rule, samples were accepted as belonging to the modeled cultivar if their predicted value was inside the warning limits, rejected if their predicted value was outside the control limits, and suspect if their predicted value was between the warning and control limits. Four parameters were calculated to evaluate the performance of the prediction models, namely, positive predictive value (PPV), negative predictive value (NPV), and

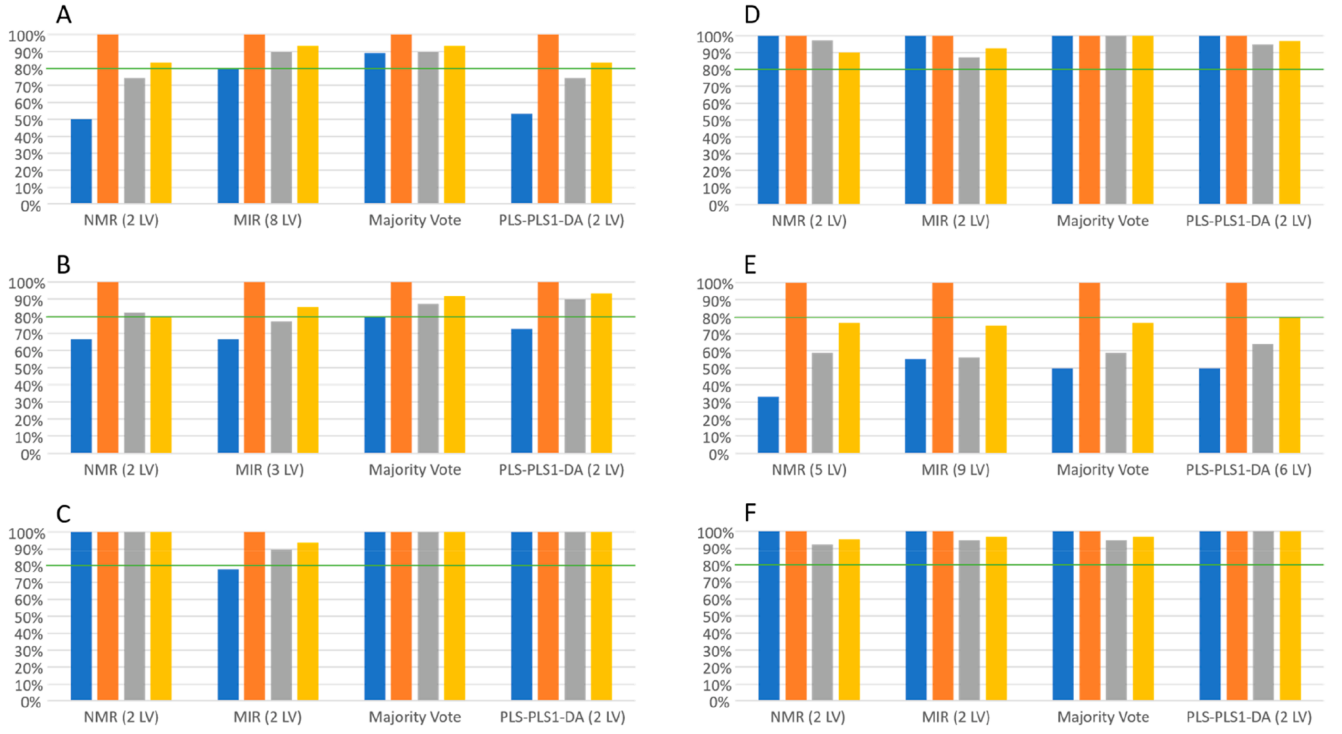

Figure 4. Performance parameters (blue bar, positive predictive value; orange bar, negative predictive value; gray bar, efficiency; and yellow bar, balanced accuracy) of the PLS1-DA models using either NMR data, MIR data, data fusion by majority vote, or data fusion by PLS−PLS1-DA to predict the varietal origin of French and Portuguese olive oil samples using six cultivars (A, Olivière; B, Salonenque; C, Tanche; D, Blanqueta; E, Carrasquenha; and F, Galega Vulgar).

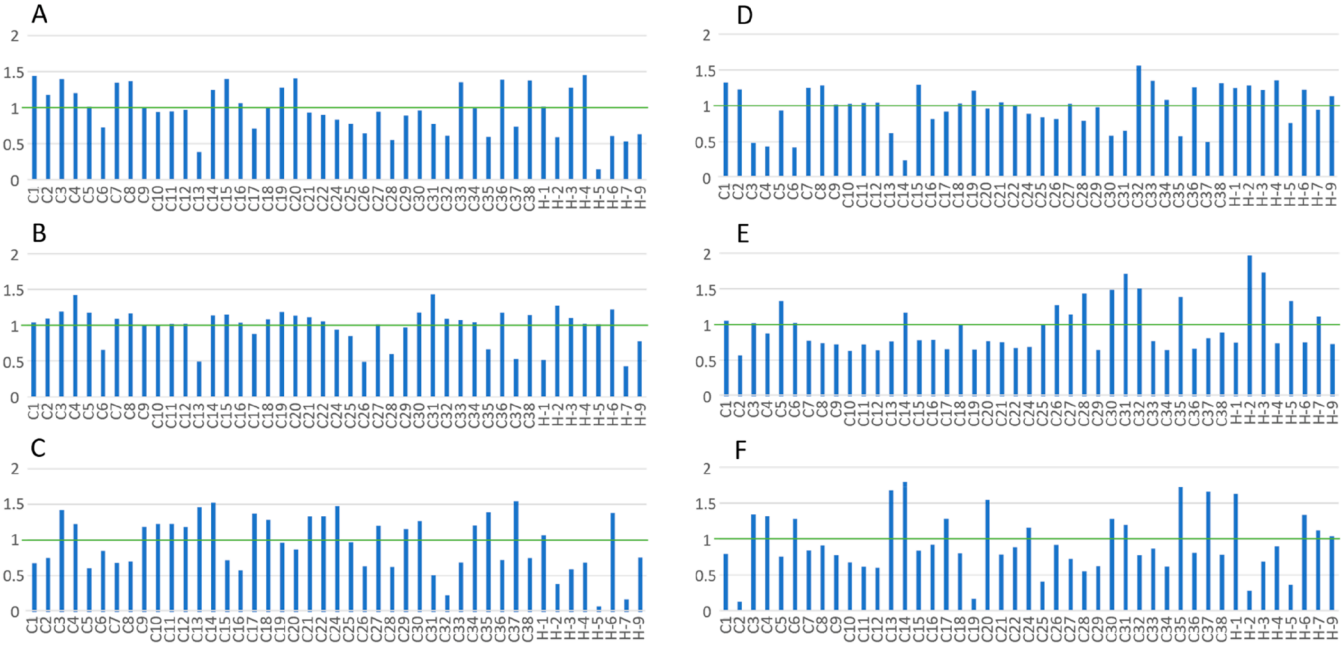

Figure 5. VIP values of the NMR variables in the PLS1-DA models predicting the varietal origin of French and Portuguese olive oil samples using six cultivars (A, Olivière; B, Salonenque; C, Tanche; D, Blanqueta; E, Carrasquenha; and F, Galega Vulgar).

efficiency (EFF) using the equations detailed by Cuadros-Rodríguez et al.[47] and balanced accuracy (BA) described by Bekkar et al.[48] also named area under the receiver operating curve.[46] The calibration included a full leave-one-out cross-validation step after which the optimal number of LVs was chosen as the lowest number of LVs, giving a BA of cross-validation greater than 80%, which led to a low number, thus avoiding overfitting. Furthermore, the contribution of the NMR and MIR variables to each discrimination model was studied by means of the variable importance in projection (VIP), which was calculated with the equation from Mehmood et al.[49] A high VIP value indicates a strong influence of the variable on the model; for instance, variables with a VIP over 1 are often considered as more relevant.

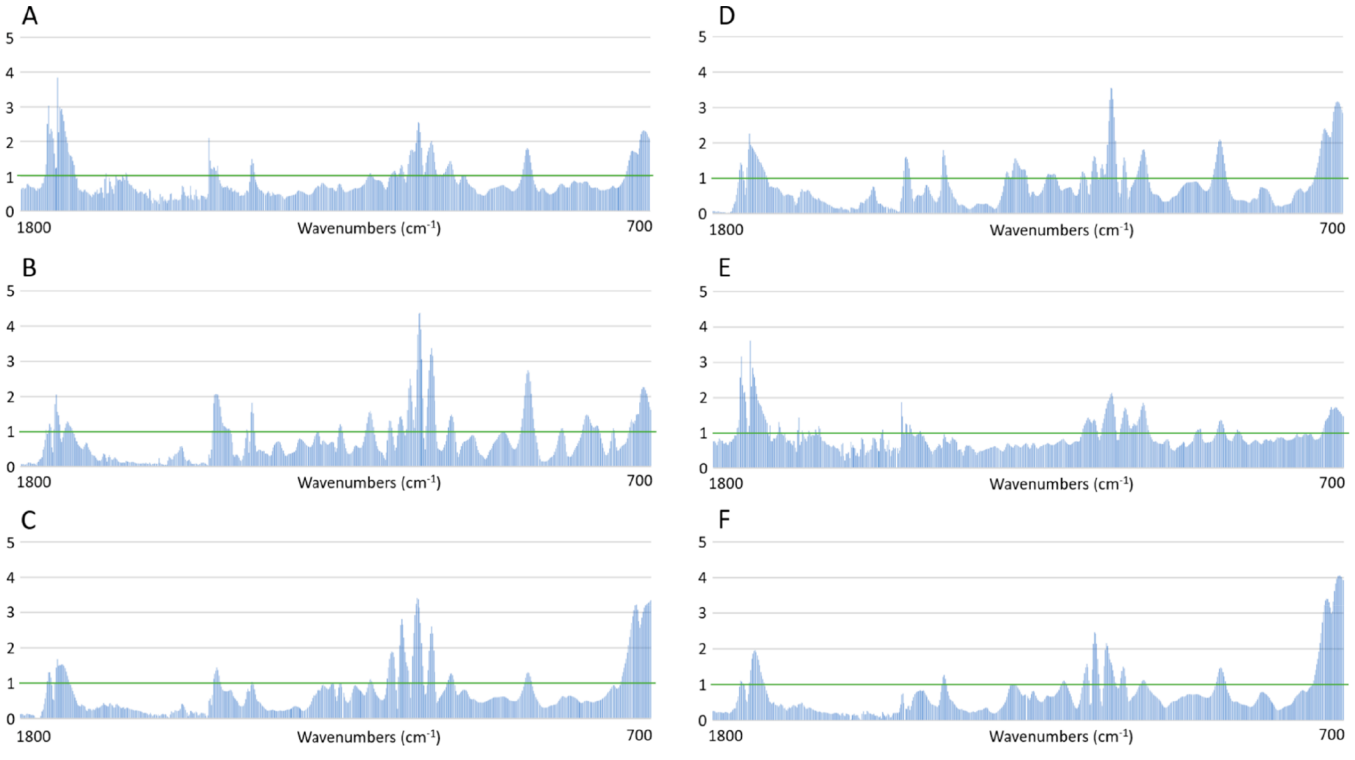

Figure 6. VIP values of the MIR variables in the PLS1-DA models predicting the varietal origin of French and Portuguese olive oil samples using six cultivars (A, Olivière; B, Salonenque; C, Tanche; D, Blanqueta; E, Carrasquenha; and F, Galega Vulgar).

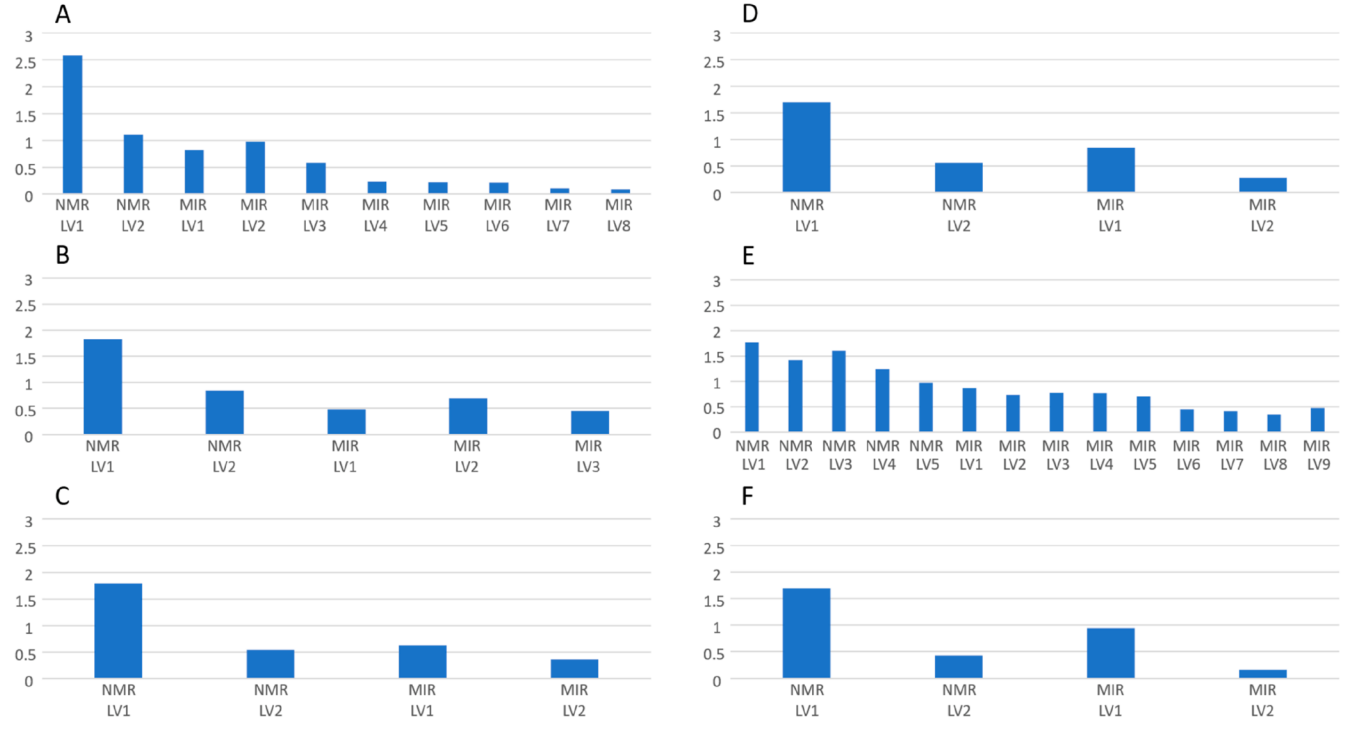

Figure 7. VIP values of the NMR and MIR latent variables in the PLS−PLS1-DA models predicting the varietal origin of French and Portuguese olive oil samples using six cultivars (A, Olivière; B, Salonenque; C, Tanche; D, Blanqueta; E, Carrasquenha; and F, Galega Vulgar).

Finally, two data fusion strategies described in a previous paper[41] were applied: mid-level data fusion strategy: hierarchical PLS−PLS1-DA, involving a first step of dimension reduction by PLS1-DA on the separate NMR and MIR data set, followed by a second PLS1-DA modeling on the concatenated scores obtained with the optimal number of LV in the first step.

RESULTS AND DISCUSSION

NMR and MIR Spectra Interpretation. As commonly observed, the ^{1}H NMR spectrum shows nine resonance signals that are attributed to the fatty acyl chain and the glyceryl protons of the triacylglycerol (TAG) component. Because the

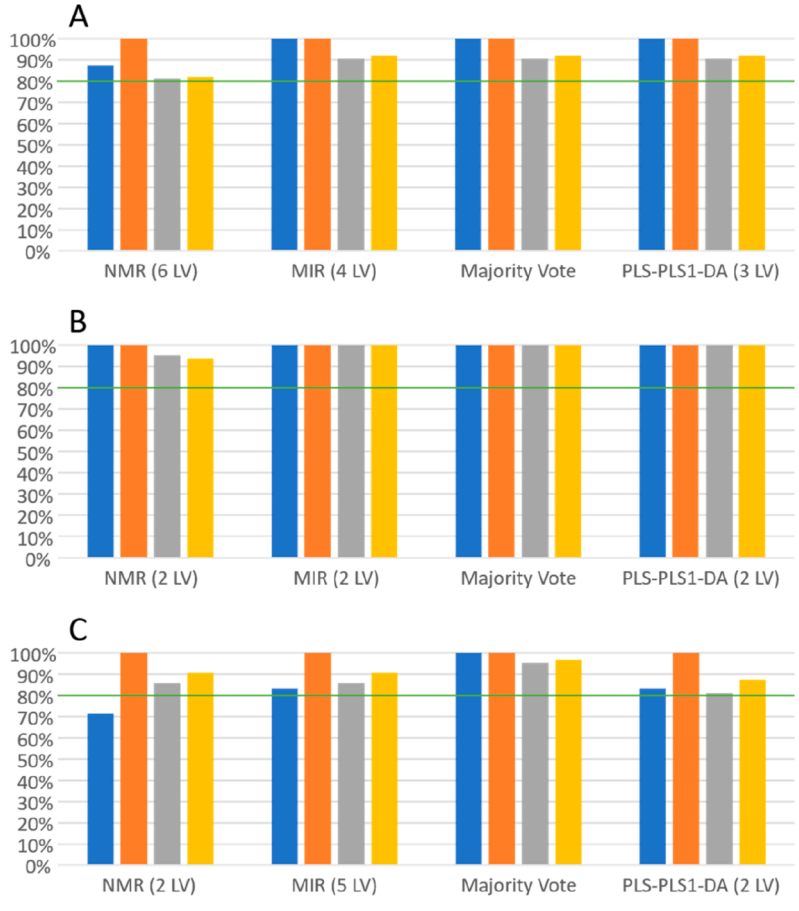

Figure 8. Performance parameters (blue bar, positive predictive value; orange bar, negative predictive value; gray bar, efficiency; and yellow bar, balanced accuracy) of the PLS1-DA models using either NMR data, MIR data, data fusion by the majority vote, or data fusion by PLS−PLS1-DA to predict the varietal origin of French and Portuguese olive oil samples using three cultivars (A, Olivière; B, Salonenque; and C, Carrasquenha).

olive oil fatty acids are similar, in free form or as glyceride esters, overlapping of the ^{1}H NMR signals occurs, hampering the differentiation between several components.[50] As commonly observed, the olive oil ^{1}H NMR spectrum (Figure 1A) shows nine resonance massifs. In fact, ^{1}H NMR spectrum contains several overlapping peaks as a result of the presence of different multiplet patterns that arise from the spin coupling of different protons, which are condensed into a very narrow spectral window (~15 ppm). Additionally, because the olive oil fatty acids are similar, in free form or as glyceride esters, in the ^{1}H NMR spectrum, it is not possible to attribute the positional distribution of the fatty acids in the glycerol backbone. Nevertheless, the nine resonance massifs are attributed to the fatty acyl chain and the glyceryl protons of the TAG component and, once integrated, lead to nine significant cumulative variables. As depicted in Figure 1B, the ^{13}C NMR DEPT 45 spectrum presents 38 characteristics resonances corresponding to CH, CH_2, and CH_3 groups (all protonated carbons). However, the signals of the quaternary carbons, including the signals of the deuterated chloroform solvent, are not detected or observed. The ^{13}C NMR DEPT 45 tool has been used in this work as a result of its strong advantage compared to the broadband ^{13}C NMR spectrum. Indeed, within this carbon-13 editing technique, significant structural and compositional information is obtained more rapidly and with better sensitivity than with broadband ^{13}C NMR, owing to the polarization transfer.

MIR spectra (Figure 2) are almost similar for all samples, and no manifest difference in band intensity is visible without chemometric pretreatments and modeling.

PCA Analysis. The scores plot (Figure 3A) shows that samples can be grouped according to their varietal origin based on their ^{1}H and ^{13}C DEPT 45 NMR data, even though some cultivars appear to be close to each other. The first component (PC1) represents 59% of the variability. On this component, SA and BL samples have positive scores and are well-separated from OL, TA, and CR samples, which have negative scores, while GA samples are in the middle. The second component represents 24% of variability. It separates TA and BL samples with positive scores from GA and OL samples with negative scores, while SA and CR samples are in the middle.

The loadings plot (Figure 3B) indicates which variables are correlated or anti-correlated and can be used to identify variables that are typical from each cultivar. OL samples, being situated in the bottom left corner of the plot, should present higher values for the C5, C16, C20, and C31 variables and lower values for some of the variables in the opposite corner. On the contrary, BL samples in the top right corner should have opposite characteristics. TA samples in the top left corner could have higher values for the H-6 variable and lower values for some of the variables in the opposite corner. The position of GA samples at the bottom of the plot could mean that they present higher values for the variables C3, C6, and C14 and/or lower values for H-1. SA samples in the middle of the right side

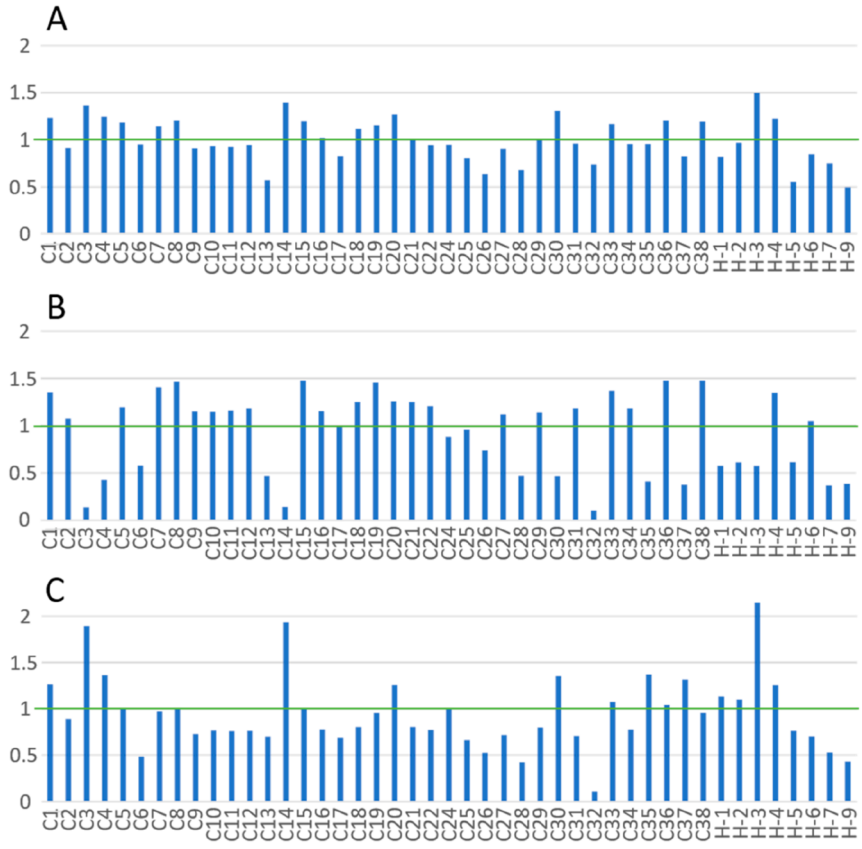

Figure 9. VIP values of the NMR variables in the PLS1-DA models predicting the varietal origin of French and Portuguese olive oil samples using three cultivars (A, Olivière; B, Salonenque; and C, Carrasquenha).

could be due to higher values of C2 and C19 and/or high values of some variables at both the top and bottom right corners, whereas CR samples in the middle of the left side should have opposite characteristics.

PLS1-DA Analysis. The results of the different prediction models built with the six cultivars are summarized in Figure 4. More detailed confusion matrices can be found in Tables S1 and S2 of the Supporting Information.

NMR Data. All of the PLS1-DA model built with only NMR data results in rather satisfying BA of prediction with values over 75%, and no sample is incorrectly rejected from a modeled cultivar as indicated by the 100% NPV. However, EFF is not very good for the CR and OL cultivars, reaching only 59 and 74%, respectively (Figure 4). This is due to the fact that the warning and control limits are lower for these two models. Some samples from other cultivars are thus mistakenly accepted as belonging to the CR or OL cultivar, which is reflected by the low PPV (33 and 50%, respectively). This issue is also present for the SA model, but with a lesser influence because the PPV reaches 67%. The BL and GA models reach very satisfying EFF values of 97 and 92%, respectively, with PPV and NPV of 100% indicating that only a few samples are predicted as suspect by these models. Finally, the TA model gives a perfect prediction for all of the studied samples.

The VIP values presented in Figure 5 show that each model uses several important variables that differ for each cultivar. Moreover, even the weakly participative variables in the first two components of the PCA appear to play an important part in the PLS1-DA models.

The model predicting the OL cultivar is mostly influenced by the variables H-4, C1, and C20, which have VIP values over 1.4, followed seven other variables with VIP values over 1.3 (C15, C3, C36, C38, C8, C33, and C7). The SA model has many variables with VIP values between 1 and 1.2 but only two variables with a VIP value over 1.4, namely, C31 and C4. The TA model is mostly influenced by C37 and C14 with VIP values over 1.5, followed by C24, C13, and C3 with VIP values over 1.4 and five other variables with VIP values over 1.3 (C35, H-6, C17, C22, and C21). The BL model only has one variable with a VIP value over 1.5, C32, followed by H-4, C33, C1, and C38 with VIP values over 1.3. The CR model is strongly influenced by H-2, H-3, C31, and C32 with VIP values over 1.5, followed by C30 and C28 with VIP values over 1.4 and C35, C5, and H-5 with VIP values over 1.3. Finally, the GA model is mostly influenced by six variables, C14, C35, C13, C37, H-1, and C20, with VIP values over 1.5 and then C3, H-6, and C4 with VIP values over 1.3.

Comparing these results to the PCA plots indicates the specificities of each cultivar: OL is characterized by lower values of C1, C7, and C8 from ethylenic carbons of some TAG with linoleoyls (L13 $\alpha\beta$, L10 $\alpha\beta$, and L12 $\alpha\beta$), C15, C33, C36, and C38 from aliphatic carbons of some TAG with linoleoyls (L16 $\alpha\beta$, L11 $\alpha\beta$, L17 $\alpha\beta$, and L18 $\alpha\beta$), and H-4 from H on Csp3 of diene from linoleoyl and linolenoyl but higher values of C3 from ethylenic carbons of other TAG with linoleoyl (L9 α) and C20 from aliphatic carbons of TAG with oleoyls (O14 $\alpha\beta$). SA has medium values for most of the variables. TA is characterized by lower values of C3 from ethylenic carbons of some TAG with linoleoyl (L9 α) and C13, C14, C17, C21,

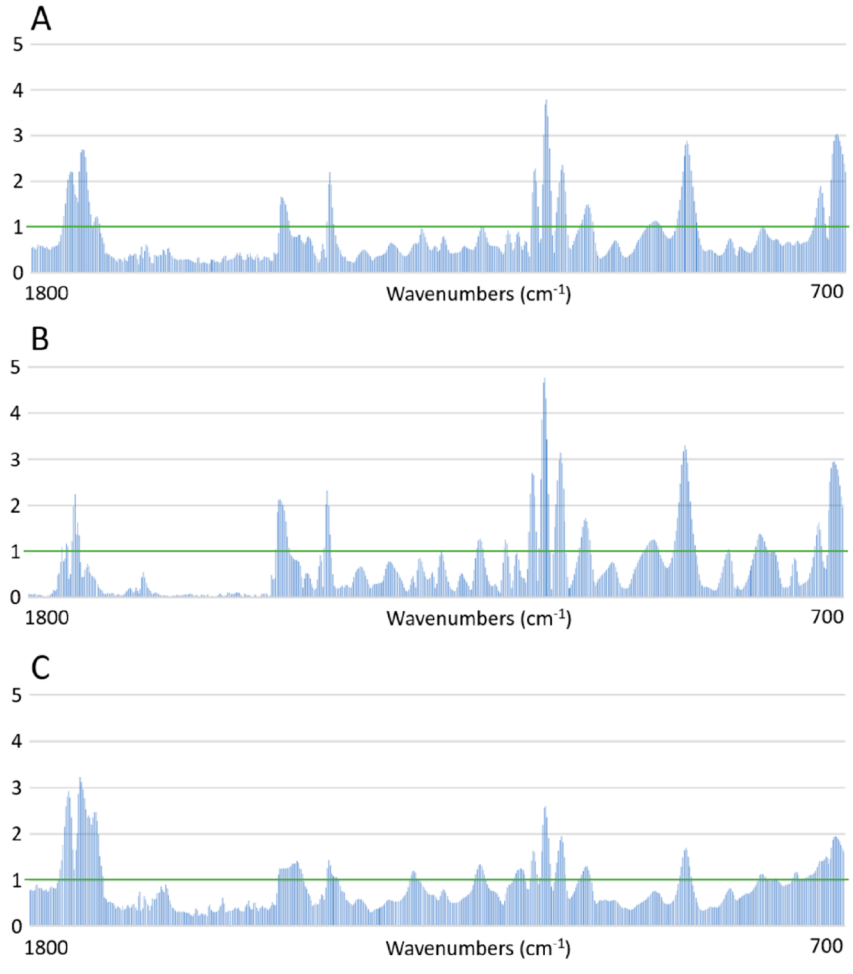

Figure 10. VIP values of the MIR variables in the PLS1-DA models predicting the varietal origin of French and Portuguese olive oil samples using three cultivars (A, Olivière; B, Salonenque; and C, Carrasquenha).

Table 3. Prediction Results of the PLS1-DA Models with the Two-Step Procedure Applied to EVOO Samples from Other Cultivars Not Used in the Calibration and Validation Process

NMR, step 1 (n = 75)	TA	not TA	suspect	BL	not BL	suspect	GA	not GA	suspect
	1	68	6	1	72	2	0	75	0
NMR, step 2 (n = 73)	OL	not OL	suspect	SA	not SA	suspect	CR	not CR	suspect
	8	49	16	2	68	3	26	25	22
MIR, step 1 (n = 75)	TA	not TA	suspect	BL	not BL	suspect	GA	not GA	suspect
	2	66	7	4	52	19	4	63	8
MIR, step 2 (n = 65)	OL	not OL	suspect	SA	not SA	suspect	CR	not CR	suspect
	9	44	12	3	49	13	25	23	17
PLS−PLS1-DA, step 1 (n = 75)	TA	not TA	suspect	BL	not BL	suspect	GA	not GA	suspect
	1	67	7	3	71	1	0	75	0
PLS−PLS1-DA, step 2 (n = 71)	OL	not OL	suspect	SA	not SA	suspect	CR	not CR	suspect
	7	43	21	1	65	5	30	23	18
majority vote, step 1 (n = 75)	TA	not TA	suspect	BL	not BL	suspect	GA	not GA	suspect
	2	68	5	2	71	2	0	74	1
majority vote, step 2 (n = 71)	OL	not OL	suspect	SA	not SA	suspect	CR	not CR	suspect
	1	54	16	3	65	3	20	33	18

C22, C24, C35, and C37 from aliphatic carbons of TAG with stearoyl and oleoyl (including S16 α/O16 $\alpha\beta$, S6 α, S15 α, O15/13 $\alpha\beta$, S17 α/O17 $\alpha\beta$, and S18 α/O18 $\alpha\beta$) but higher values of H-6 from H on C adjacent to unsaturation of unsaturated fatty acids. This confirms the results of the comparative study related to the fatty acid and triacylglycerol compositions of the main French cultivars, which showed that the TA cultivar has low linoleic acid and palmitic acid contents.[51] BL is characterized by lower values of C32 from aliphatic carbons of some TAG with linoleoyls and oleoyls (L8 $\alpha\beta$/O8 $\alpha\beta$) but higher values of C1 from ethylenic carbons of other TAG with linoleoyls (L13 $\alpha\beta$), C33 and C38 from aliphatic carbons of other TAG with linoleoyls (L11 $\alpha\beta$ and L18 $\alpha\beta$), and H-4 from H on Csp3 of diene from linoleoyl and

linolenoyl. CR is difficult to characterize because several of its important variables are not very well-represented on the PCA (C28 and C32 from aliphatic carbons of TAG with stearoyl, linoleoyl, and oleoyl S4 α-O/L6 α/β-O/L4 α and L8 $\alpha\beta$/O8 $\alpha\beta$, H-2 from H on C2 of glycerol, and H-5 from H of acyl chains), and it also seems to have medium values. GA is characterized by lower values of H-1 and H-6 from H on C adjacent to unsaturation of unsaturated fatty acids but higher values of C3 and C4 from ethylenic carbons of TAG with linoleoyl (L9 α and L9 β) and C13, C14, C20, C35, and C37 from aliphatic carbons of TAG with stearoyl and oleoyl (including S16 α/O16 $\alpha\beta$, O14 $\alpha\beta$, S17 α/O17 $\alpha\beta$, and S18 α/O18 $\alpha\beta$).

MIR Data. The models using only MIR data gave similar results for most of the cultivars, with excellent NPV but perfectible PPV (Figure 4). The global parameter of efficiency EFF for the PLS1-DA models using only MIR data ranges from a mediocre 56% for CR to a very good 95% for GA, and the BA of prediction is satisfying for all of the cultivars, between 75% for CR and 97% for GA.

The VIP values presented in Figure 6 show that all of the models are mostly influenced by the same spectral areas: 1760−1700 cm^{-1} attributed to C═O stretching, 1465−1450 cm^{-1} and around 1395 cm^{-1} related to C−H bending, around 1190 cm^{-1} and several bands between 1155 and 1040 cm^{-1} that are associated with C−O and C−C bending, 925−900 cm^{-1} attributed to C−H bending of unsaturations, and finally 745−700 cm^{-1} attributed to C−H bending of long chains.[52,53] However, the respective importance of the spectral bands varies between the predicted cultivars. For instance, the region between 1760 and 1700 cm^{-1} is the most important in the OL and CR models, whereas the importance of the 1155−1040 cm^{-1} area is higher in the SA, TA, and BL models. The SA model also has higher VIP values than the others in the 1465−1450 and 925−900 cm^{-1} areas. The TA and BL models also have higher VIP values in the 745−700 cm^{-1} range, and for the GA model, this last area is the most important.

The main differences in the prediction results compared to NMR models are the better performance of the MIR model to predict the OL cultivar (80% PPV and 90% EFF with MIR versus 50% PPV and 74% EFF with NMR) but the poorer performance of the MIR model to predict the TA cultivar (78% PPV and 90% EFF with MIR versus 100% PPV and 100% EFF with NMR). These differences suggest a complementarity of the two data sets, which could be exploited with the data fusion strategies to improve the discrimination of varietal origin.

Data Fusion. Indeed, the models using data fusion result in similar or better prediction performances (Figure 4). As in the PLS−PLS1-DA algorithm, individual NMR and MIR data sets are subjected to a first step of dimension reduction by PLS1-DA and then the scores obtained with the optimal number of LV for each individual model are combined and used to develop the final PLS−PLS1-DA model; the imbalanced number of variables from each data block does not have influence on the final result. Thus, in the present study, variables from NMR and MIR have not been weighted, although their numbers are very different to preserve the potential benefit from the combination of the two sources of information (i.e., NMR data containing information on the major compounds of olive oil and MIR data representative of all of the major and minor compounds). Moreover, in a previous study, it was shown that the scaling to compensate for the much larger number of variables in one block than in another strongly reduces the influence of data from the block containing the greater number of variables and, therefore, the interest of the combination.[36] The PLS−PLS1-DA strategy gives perfect predictions for the TA and GA models and very good results for the BL model (100% PPV, 100% NPV, 95% EFF, and 97% BA). The performances are also improved for the SA model, reaching 73% PPV and both EFF and BA over 90%. The model predicting CR gives better results as well, with only 50% PPV but 64% EFF and 79% BA. Only the OL model is not improved, with results similar to those obtained with NMR data alone, because the PLS−PLS1-DA models mostly use information from the NMR data, as seen in the VIP values shown in Figure 7.

Moreover, the majority vote strategy also gives equal or better results for all of the cultivars, with perfect predictions for the TA and BL models and very good results for the GA model (100% PPV, 100% NPV, 95% EFF, and 97% BA). The performances are also improved for the OL model, reaching 89% PPV and both EFF and BA over 90%, and for the SA model, with 80% PPV, 87% EFF, and 92% BA. Only the CR model still has results similar to those obtained with NMR data alone but with a higher PPV of 50%.

Two-Step Procedure. Because the discrimination of cultivars CR, OL, and SA was not completely satisfactory when the six cultivars were used, a second set of models using only the samples of these three cultivars was built to improve their discrimination. Thus, a two-step procedure is proposed: First, each new sample of unknown origin should be tested with the TA, BL, and GA models built with the six cultivars. Then, if it did not recognize as any of these, it should be tested with the OL, SA, and CR models built with only three cultivars.

The performances of the new prediction models built with the second step of this procedure are shown in Figure 8, and more detailed results can be found in Tables S3 and S4 of the Supporting Information. The new models already give very satisfying results when using NMR or MIR data alone, with slightly better predictions for MIR.

With NMR data only, all of the models result in rather good BA of prediction, with values over 80%, and no sample is incorrectly rejected from a modeled cultivar, as indicated by the 100% NPV. The efficiency EFF is improved for the three cultivars, reaching 81% for OL, 86% for CR, and an excellent 95% for SA. A few samples from other cultivars are still mistakenly accepted as belonging to the modeled cultivar in the CR and OL models, but the PPV is significantly improved to the acceptable values of 71% for CR and 88% for OL. Finally, the PPV of 100% for SA confirms the good performance of this model.

Discriminating only these three cultivars modifies the VIP values, as seen in Figure 9. Indeed, the model predicting the OL cultivar is now mostly influenced by H-3 with a VIP value over 1.4, followed by C14, C3, and C30 with VIP values over 1.3. The model predicting the SA cultivar has six variables with VIP values over 1.4, C36, C38, C15, C8, C19, and C7, and then three variables with VIP values over 1.3 (C33, C1, and H-4). The model predicting CR is strongly influenced by the same variables as the OL model: H-3, C14, and C3 have VIP values over 1.5, and then C35, C4, C30, and C37 have VIP values over 1.3. The SA cultivar is the easiest to characterize because it has higher values than OL and CR for all of its important variables (including ethylenic carbons from TAG

with linoleoyls L13 $\alpha\beta$, L10 $\alpha\beta$, and L12 $\alpha\beta$, aliphatic carbons from TAG with linoleoyls L16 $\alpha\beta$, L11 $\alpha\beta$, L17 $\alpha\beta$, and L18 $\alpha\beta$, and hydrogens from linoleoyl and linolenoyl). This is in agreement with the study from Ollivier et al., which found that oils from the SA cultivar contained high amounts of linoleic and palmitoleic acids.[50] The OL cultivar has lower values of H-3 from glycerol but higher values of C3 from ethylenic carbons of TAG with linoleoyl (L9 α) and C14 and C30 from unknown aliphatic carbons of TAG compared to CR. CR also has lower values of C4 from ethylenic carbons of TAG with linoleoyl (L9 β) and C35 and C37 from aliphatic carbons of TAG with stearoyl and oleoyl (S17 α/O17 $\alpha\beta$ and S18 α/O18 $\alpha\beta$).

The models built with MIR data are only able to perfectly discriminate the SA cultivar and also have very good performances for OL (100% PPV and NPV and EFF and BA over 90%) and CR (83% PPV, 100% NPV, 86% EFF, and 91% BA), as shown in Figure 8.

The VIP values present the same major areas of influence as in the previous models using MIR data (Figure 10). The new SA and CR models show little difference compared to their previous version; however, the new OL model gives more importance to the 1155−1040, 925−900, and 745−700 cm^{-1} areas.

The data fusion strategies bring little additional improvement to these excellent performances (Figure 8). Only the majority vote is able to enhance the results for the CR cultivar, reaching 100% PPV and NPV and both EFF and BA over 95%.

Prediction of Unknown Samples. The different PLS1-DA models were applied to the NMR and MIR data of EVOO from other cultivars using the two-step procedure to verify their ability to reject these unknown samples. The results are presented in Table 3.

The models using only NMR data give very good performances in the first step, because the model predicting the GA cultivar is able to reject all of the unknown samples, only one sample is mistakenly accepted as TA, and one other sample is wrongly recognized as BL. The second step also gives good results for SA, with only two wrongly accepted samples; however, eight other samples are recognized as OL, and the CR model mistakenly accepts 26 samples, which is not satisfying.

The models using only MIR data had good prediction performances for the samples from the six cultivars used for their calibration, but they seem somewhat less robust than those using NMR when other cultivars are considered. Indeed, in the first step, two samples are wrongly recognized as TA, four samples are wrongly recognized as BL, and four samples are wrongly recognized as GA, while in the second step, three samples are mistakenly accepted as SA, nine samples are mistakenly accepted as OL, and 25 samples are mistakenly accepted as CR. Moreover, the number of suspect samples that are not clearly rejected is higher than with NMR data for most of the models.

The fusion of NMR and MIR data with the PLS−PLS1-DA approach gives results close to those obtained with NMR alone: in the first step, all of the samples are correctly rejected from the GA model, one sample is wrongly accepted as TA and three samples are wrongly accepted as BL, while in the second step, one sample is wrongly accepted as SA, seven samples are wrongly accepted as OL, and 30 samples are wrongly accepted as CR. Data fusion with the majority vote approach slightly improves the results, especially in the second step. In the first step, there are still zero samples mistakenly accepted as GA but two samples mistakenly accepted as TA and two samples mistakenly accepted as BL. In the second step, only one sample is wrongly recognized as OL, three samples are wrongly recognized as SA, and the number of samples recognized as CR is reduced to 20.

Finally, this work has assessed the potential of ^{1}H and ^{13}C NMR with a polarization transfer technique combined with chemometric models to discriminate the cultivar origin of French and Portuguese extra virgin olive oils. In fact, NMR resonances are mainly related to fatty acid chains of EVOO. In addition to ^{1}H data, ^{13}C NMR DEPT 45 data prove to be valuable for prediction purposes, enabling the clear discrimination of EVOO samples studied in this work according to the olive cultivar using an approach that includes a chemometric-based tool. Thus, the application of PLS1-DA modeling to ^{13}C NMR DEPT 45 data has been successfully validated, showing promising results for the varietal origin discrimination of EVOOs. The interest of combining ^{1}H and ^{13}C DEPT 45 NMR data with MIR data has also been demonstrated. Indeed, the performances of the PLS1-DA chemometric models were improved by the data fusion strategies, especially with the high-level fusion using the majority vote. Moreover, the application of the control chart method to optimize the interpretation of the PLS1-DA results is further validated by this study, and a two-step strategy is proposed to improve the discrimination of the six studied cultivars and even the rejection of new samples belonging to other cultivars. Only the model predicting the variety Carrasquenha had poorer results and should be improved by taking more representative samples into account during its calibration.

REFERENCES

(1) Ortea, I.; O'Connor, G.; Maquet, A. Review on proteomics for food authentication. *J. Proteomics* **2016**, *147*, 212−225.

(2) Bendini, A.; Cerretani, L.; Carrasco-Pancorbo, A.; Gómez-Caravaca, A. M.; Segura-Carretero, A.; Fernández-Gutiérrez, A.; Lercker, G. Phenolic molecules in virgin olive oils: A survey of their sensory properties, health effects, antioxidant activity and analytical methods. An overview of the last decade. *Molecules* **2007**, *12*, 1679−1719.

(3) Danezis, G. P.; Tsagkaris, A. S.; Camin, F.; Brusic, V.; Georgiou, C. A. Food authentication: Techniques, trends & emerging approaches. *TrAC, Trends Anal. Chem.* **2016**, *85*, 123−132.

(4) Cubero-Leon, E.; Penalver, R.; Maquet, A. Review on metabolomics for food authentication. *Food Res. Int.* **2014**, *60*, 95−107.

(5) Medina, S.; Perestrelo, R.; Silva, P.; Pereira, J. A. M.; Câmara, J. S. Current trends and recent advances on food authenticity technologies and chemometric approaches. *Trends Food Sci. Technol.* **2019**, *85*, 163−176.

(6) Chedid, E.; Rizou, M.; Kalaitzis, P. Application of high-resolution melting combined with DNA-based markers for quantitative analysis of olive oil authenticity and adulteration. *Food Chem.: X* **2020**, *6*, 100082.

(7) Bazakos, C.; Khanfir, E.; Aoun, M.; Spano, T.; Zein, Z. E.; Chalak, L.; El Riachy, M.; Abou-Sleymane, G.; Ben Ali, S.; Grati Kammoun, N.; Kalaitzis, P. The potential of SNP-based PCR-RFLP capillary electrophoresis analysis to authenticate and detect admixtures of Mediterranean olive oils. *Electrophoresis* **2016**, *37*, 1881−1890.

(8) Montemurro, C.; Miazzi, M. M.; Pasqualone, A.; Fanelli, V.; Sabetta, W.; di Rienzo, V. Traceability of PDO Olive Oil "Terra di Bari" Using High Resolution Melting. *J. Chem.* **2015**, *2015*, 496986.

(9) Pasqualone, A.; Montemurro, C.; di Rienzo, V.; Summo, C.; Paradiso, V. M.; Caponio, F. Evolution and perspectives of cultivar identification and traceability from tree to oil and table olives by means of DNA markers. *J. Sci. Food Agric.* **2016**, *96*, 3642−3657.

(10) Druml, B.; Cichna-Markl, M. High resolution melting (HRM) analysis of DNA-Its role and potential in food analysis. *Food Chem.* **2014**, *158*, 245−254.

(11) Gomes, S.; Breia, R.; Carvalho, T.; Carnide, V.; Martins-Lopes, P. Microsatellite high-resolution melting (SSR-HRM) to track olive genotypes: From field to olive oil. *J. Food Sci.* **2018**, *83*, 2415−2423.

(12) Maléchaux, A.; Le Dréau, Y.; Artaud, J.; Dupuy, N. Exploring the Scientific Interest for Olive Oil Origin: A Bibliometric Study from 1991 to 2018. *Foods* **2020**, *9*, 556.

(13) Callao, M. P.; Ruisanchez, I. An overview of multivariate qualitative methods for food fraud detection. *Food Control* **2018**, *86*, 283−293.

(14) Valli, E.; Bendini, A.; Berardinelli, A.; Ragni, L.; Riccò, B.; Grossi, M.; Toschi, T. G. Rapid and innovative instrumental approaches for quality and authenticity of olive oils. *Eur. J. Lipid Sci. Technol.* **2016**, *118*, 1601−1619.

(15) Wang, P.; Sun, J.; Zhang, T.; Liu, W. Vibrational spectroscopic approaches for the quality evaluation and authentication of virgin olive oil. *Appl. Spectrosc. Rev.* **2016**, *51*, 763−790.

(16) De Luca, M.; Restuccia, D.; Clodoveo, M. L.; Puoci, F.; Ragno, G. Chemometric analysis for discrimination of extra virgin olive oils from whole and stoned olive pastes. *Food Chem.* **2016**, *202*, 432−437.

(17) Terouzi, W.; De Luca, M.; Bolli, A.; Oussama, A.; Patumi, M.; Ioele, G.; Ragno, G. A. discriminant method for classification of Moroccan olive varieties by using direct FT-IR analysis of the mesocarp section. *Vib. Spectrosc.* **2011**, *56*, 123−128.

(18) Dais, P.; Hatzakis, E. Quality assessment and authentication of virgin olive oil by NMR spectroscopy: A critical review. *Anal. Chim. Acta* **2013**, *765*, 1−27.

(19) Vlahov, G. Application of NMR to the study of olive oils. *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *35*, 341−357.

(20) Vlahov, G.; Shaw, A. D.; Kell, D. B. Use of 13 C nuclear magnetic resonance distortionless enhancement by polarization transfer pulse sequence and multivariate analysis to discriminate olive oil cultivars. *J. Am. Oil Chem. Soc.* **1999**, *76*, 1223−1231.

(21) D'Imperio, M.; Dugo, G.; Alfa, M.; Mannina, L.; Segre, A. L. Statistical analysis on Sicilian olive oils. *Food Chem.* **2007**, *102*, 956−965.

(22) Ün, İ.; Ok, S. Analysis of olive oil for authentication and shelf-life determination. *J. Food Sci. Technol.* **2018**, *55*, 2476−2487.

(23) Mannina, L.; Marini, F.; Gobbino, M.; Sobolev, A. P.; Capitani, D. NMR and chemometrics in tracing European olive oils: The case study of Ligurian samples. *Talanta* **2010**, *80*, 2141−2148.

(24) Girelli, C. R.; Del Coco, L.; Fanizzi, F. P. ^{1}H NMR spectroscopy and multivariate analysis as possible tool to assess cultivars, from specific geographical areas, in EVOOS. *Eur. J. Lipid Sci. Technol.* **2016**, *118*, 1380−1388.

(25) Özdemir, İ. S.; Dağ, Ç.; Makuc, D.; Ertaş, E.; Plavec, J.; Bekiroğlu, S. Characterisation of the Turkish and Slovenian extra virgin olive oils by chemometric analysis of the presaturation ^{1}H NMR spectra. *LWT - Food Science and Technology* **2018**, *92*, 10−15.

(26) Vlahov, G.; Del Re, P.; Simone, N. Determination of geographical origin of olive oils using ^{13}C nuclear magnetic resonance spectroscopy. I—Classification of olive oils of the puglia region with denomination of protected origin. *J. Agric. Food Chem.* **2003**, *51*, 5612−5615.

(27) Merchak, N.; Rizk, T.; Silvestre, V.; Remaud, G. S.; Bejjani, J.; Akoka, S. Olive oil characterization and classification by ^{13}C NMR with a polarization transfer technique: A comparison with gas chromatography and ^{1}H NMR. *Food Chem.* **2018**, *245*, 717−723.

(28) Cabrita, M. J.; Pires, A.; Burke, A.; Garcia, R. Seeking for a fast screening of varietal origin of olive oil: The usefulness of a nmr based approach. *Foods* **2021**, *10*, 399.

(29) Borràs, E.; Ferré, J.; Boqué, R.; Mestres, M.; Aceña, L.; Busto, O. Data fusion methodologies for food and beverage authentication and quality assessment—A review. *Anal. Chim. Acta* **2015**, *891*, 1−14.
(30) Monakhova, Y. B.; Godelmann, R.; Hermann, A.; Kuballa, T.; Cannet, C.; Schäfer, H.; Spraul, M.; Rutledge, D. N. Synergistic effect of the simultaneous chemometric analysis of ^{1}H NMR spectroscopic and stable isotope (SNIF-NMR, ^{18}O, ^{13}C) data: Application to wine analysis. *Anal. Chim. Acta* **2014**, *833*, 29−39.
(31) Hohmann, M.; Monakhova, Y.; Erich, S.; Christoph, N.; Wachter, H.; Holzgrabe, U. Differentiation of Organically and Conventionally Grown Tomatoes by Chemometric Analysis of Combined Data from Proton Nuclear Magnetic Resonance and Mid-Infrared Spectroscopy and Stable Isotope Analysis. *J. Agric. Food Chem.* **2015**, *63*, 9666−9675.
(32) Tenenhaus, M.; Vinzi, V. E. PLS regression, PLS path modeling and generalized Procrustean analysis: A combined approach for multiblock analysis. *J. Chemom.* **2005**, *19*, 145−153.
(33) Casale, M.; Casolino, C.; Oliveri, P.; Forina, M. The potential of coupling information using three analytical techniques for identifying the geographical origin of Liguria extra virgin olive oil. *Food Chem.* **2010**, *118*, 163−170.
(34) Silvestri, M.; Elia, A.; Bertelli, D.; Salvatore, E.; Durante, C.; Li Vigni, M.; Marchetti, A.; Cocchi, M. A mid-level data fusion strategy for the Varietal Classification of Lambrusco PDO wines. *Chemom. Intell. Lab. Syst.* **2014**, *137*, 181−189.
(35) Ríos-Reina, R.; Callejón, R. M.; Savorani, F.; Amigo, J. M.; Cocchi, M. Data fusion approaches in spectroscopic characterization and classification of PDO wine vinegars. *Talanta* **2019**, *198*, 560−572.
(36) Maléchaux, A.; Laroussi-Mezghani, S.; Le Dréau, Y.; Artaud, J.; Dupuy, N. Multiblock chemometrics for the discrimination of three extra virgin olive oil varieties. *Food Chem.* **2020**, *309*, 125588.
(37) Moro, M. K.; Neto, A. C.; Lacerda, V.; Romão, W.; Chinelatto, L. S.; Castro, E. V. R.; Filgueiras, P. R. FTIR, ^{1}H and ^{13}C NMR data fusion to predict crude oils properties. *Fuel* **2020**, *263*, 116721.
(38) Roussel, S.; Bellon-Maurel, V.; Roger, J. M.; Grenier, P. Fusion of aroma, FT-IR and UV sensor data based on the Bayesian inference. Application to the discrimination of white grape varieties. *Chemom. Intell. Lab. Syst.* **2003**, *65*, 209−219.
(39) Ballabio, D.; Robotti, E.; Grisoni, F.; Quasso, F.; Bobba, M.; Vercelli, S.; Gosetti, F.; Calabrese, G.; Sangiorgi, E.; Orlandi, M.; Marengo, E. Chemical profiling and multivariate data fusion methods for the identification of the botanical origin of honey. *Food Chem.* **2018**, *266*, 79−89.
(40) Di Anibal, C. V.; Callao, M. P.; Ruisánchez, I. ^{1}H NMR and UV-visible data fusion for determining Sudan dyes in culinary spices. *Talanta* **2011**, *84*, 829−833.
(41) Maléchaux, A.; Le Dréau, Y.; Artaud, J.; Dupuy, N. Control chart and data fusion for varietal origin discrimination: Application to olive oil. *Talanta* **2020**, *217*, 121115.
(42) Garcia, R.; Pires, A.; Martins, N.; Carvalho, T.; Burke, A. J.; Cabrita, M. J. Assessment of the triacylglycerol fraction of olive oil by 1D-NMR spectroscopy: Exploring the usefulness of DEPT tool on the peak assignments of 13 C NMR spectra. *Eur. Food Res. Technol.* **2019**, *245*, 2479−2488.
(43) Jolliffe, I. T. *Principal Component Analysis*, 2nd ed.; Springer: New York, 2002; DOI: 10.1007/b98835.
(44) Galtier, O.; Abbas, O.; Le Dréau, Y.; Rébufa, C.; Kister, J.; Artaud, J.; Dupuy, N. Comparison of PLS1-DA PLS2-DA and SIMCA for classification by origin of crude petroleum oils by MIR and virgin olive oils by NIR for different spectral regions. *Vib. Spectrosc.* **2011**, *55*, 132−140.
(45) Lee, L. C.; Liong, C.-Y.; Jemain, A. A. Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: A review of contemporary practice strategies and knowledge gaps. *Analyst* **2018**, *143*, 3526−3539.
(46) Maléchaux, A.; Le Dréau, Y.; Vanloot, P.; Artaud, J.; Dupuy, N. Discrimination of extra virgin olive oils from five French cultivars: En route to a control chart approach. *Food Control* **2019**, *106*, 106691.
(47) Cuadros-Rodríguez, L.; Pérez-Castaño, E.; Ruiz-Samblás, C. Quality performance metrics in multivariate classification methods for qualitative analysis. *TrAC, Trends Anal. Chem.* **2016**, *80*, 612−624.
(48) Bekkar, M.; Djemaa, H. K.; Alitouche, T. A. Evaluation measures for models assessment over imbalanced data sets. *J. Inf. Eng. Appl.* **2013**, *3*, 27−38.
(49) Mehmood, T.; Liland, K. H.; Snipen, L.; Sæbø, S. A review of variable selection methods in partial least squares regression. *Chemom. Intell. Lab. Syst.* **2012**, *118*, 62−69.
(50) Popescu, R.; Costinel, D.; Dinca, O. R.; Marinescu, A.; Stefanescu, I.; Ionete, R. E. Discrimination of vegetable oils using NMR spectroscopy and chemometrics. *Food Control* **2015**, *48*, 84−90.
(51) Ollivier, D.; Artaud, J.; Pinatel, C.; Durbec, J. P.; Guérère, M. Triacylglycerol and Fatty Acid Compositions of French Virgin Olive Oils. Characterization by Chemometrics. *J. Agric. Food Chem.* **2003**, *51*, 5723−5731.
(52) *Handbook of Olive Oil*; Aparicio, R., Harwood, J., Eds.; Springer: Boston, MA, 2013; DOI: 10.1007/978-1-4614-7777-8.
(53) Safar, M.; Bertrand, D.; Robert, P.; Devaux, M. F.; Genot, C. Characterization of Edible Oils, Butters and Margarines by Fourier Transform Infrared Spectroscopy with Attenuated Total Reflectance. *J. Am. Oil Chem. Soc.* **1994**, *71*, 371.