Classification of olive oils according to their cultivars based on second-order data using LC-DAD

Ana M. JIMÉNEZ-CARVELO✉[1], Carlos M. CRUZ[2], Alejandro C. OLIVIERI[3], ANTONIO GONZÁLEZ-CASADO[1], Luis CUADROS-RODRÍGUEZ[1]

[1] Department of Analytical Chemistry, [2] Department of Organic Chemistry, Faculty of Sciences, University of Granada, C/ Fuentenueva s/n, E-18071, Granada, Spain

[3] Instituto de Química Rosario (IQUIR–CONICET), Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531, S2002LRK, Rosario, Argentina

Abstract

Second-order data acquired using liquid chromatography coupled to a diode array detector were used to classify extra virgin olive oils samples according to their cultivars. The chromatographic fingerprints from the epoxidised fraction were obtained using normal-phase liquid chromatography. To reduce the data matrices two strategies were employed: (1) multivariate curve resolution-alternating least squares (MCR-ALS) and (2) a new strategy proposed in this work based on the fusion of the mean data profiles in both spectral and time domains. Several conventional chemometric tools were then applied to both raw and reduced data: principal component analysis (PCA), partial least-squares-discriminant analysis (PLS-DA), soft independent modelling of class analogies (SIMCA) and n-way partial least-squares-discriminant analysis (NPLS-DA). Furthermore, an emergent multivariate classification method known as random forest (RF) has been first applied to second-order data. It was shown that RF is more efficient than conventional tools. Indeed, the obtained sensibility, specificity and accuracy are 1.00, 0.92 and 0.95 respectively; these performance metrics are significantly better than the values found for the other methods.

Keywords

Olive oil authentication; Liquid chromatography; Three-way data classification method; Multivariate curve resolution; Random forest

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

1. Introduction

Extra-virgin olive oil (EVOO) is a food which contains valuable bioactive compounds as tocopherols and tocotrienols (vitamin E), β-carotenes, sterols, or phenols, which confer cardioprotective, antioxidant and anti-inflammatory properties over the health of consumers [**1**,**2**]. Furthermore, it is mainly composed by triacylglycerols (more than 90%) having a high proportion of monounsaturated fatty acids, especially oleic acid (around 70%). Its chemical composition could vary depending on many factors such as cultivar, agronomic conditions, extraction process, and ripeness, among others [**3**]. EVOO have thus characteristic organoleptic properties [**4**] due to the presence of many different flavouring organic compounds.

This essential food represents a treasure into the Mediterranean diet, giving unique flavour and aroma to the dishes where it is employed. In the last few years, EVOO is gaining ground in high-quality cuisine, due to its broad spectrum in terms of organoleptic properties which allow choosing each cultivar for a specific flavour. Around 1700 olive varieties are being cultivated nowadays according to the World Catalogue of Olive Varieties of the International Olive Council (IOC). Nevertheless, only a handful of them are mostly used to produce olive oil [**5**].

In the last years, the main producers of olive oil have shown a special interest in the marketing of monovarietal olive oils as a way to improve the competitiveness, and to try to deal with the effects of the globalization process in the olive oil sector. The aim is to market high-quality olive oil with specific organoleptic characteristics, which reflect the effect of the cultivar and the geographical origin where it has been grown. A good strategy to take a prominent position over the competitors is to take advantage of the difference in chemical composition, organoleptic characteristics or the kind of cultivar of each EVOO, bearing a recognised quality-differentiated food seal as the 'Protected Designation Origin' (PDO) or 'Protected Geographical Indication' (PGI) according to European regulations [**6**], and also labelling the oil as monovarietal EVOO, *i.e.*, an extra-virgin olive oil obtained from a single kind of olive fruit botanical variety. These credentials, enforced within the EU and being gradually expanded internationally via bilateral agreements between the EU and non-EU countries, add value to the final product and bring exclusivity to the consumer.

The 'arbequina' cultivar is a commonly botanical variety in Spain since the XVII century. The monovarietal olive oil obtained from arbequina olive fruits shows special organoleptic properties in comparison with other olive varieties, characterized for their freshly and fruity aroma and for showing a slight pungency or even none. These particular organoleptic properties make this olive oil an appreciated product for a wide spectrum of consumers. In

Spain there are some PDO concerns to Arbequina cultivars as 'Estepa' (South of Spain) [**7**], 'Les Garrigues' [**8**] or 'Siurana' [**9**] (North of Spain).

In this sense, proper analytical methods which enable to distinguish quickly and reliably cultivar olive oils are currently demanded. There are some works reporting the classification of EVOO according to its cultivar using spectroscopic techniques [**10**], liquid chromatography [**11**,**12**,**13**,**14**] or gas chromatography [**15**,**16**,**17**]. Nevertheless, all these works are based on the quantification of specific compounds or on the study of the profile of a family of components such as chlorophylls, sterols, fatty acids and phenolic compounds.

On the other hand, it is possible to develop a global method for the classification of EVOO according to its cultivar by applying the chromatographic fingerprinting methodology [**18**] which combines second-order data with chemometric tools. Conventionally, second-order data have been used for the quantification of compounds due to what is known as 'the second-order advantage', *i.e.*, 'the analytes can be quantitated in the presence of uncalibrated interfering substances'. Therefore, only small sets of pure compounds are required for building the calibration model, instead of large calibration sets containing all possible interfering substances. The main algorithms employed to process these data are: (i) parallel factor analysis (PARAFAC) [**19**], (ii) multivariate curve resolution-alternating least squares (MCR-ALS) [**20**] and (iii) unfolded or multidimensional partial least-squares with residual bilinearization (UPLS-RBL or NPLS-RBL) [**21**].

Nevertheless, the application of this kind of data to build multivariate classification models for authentication of olive oils has not been extensively explored. The literature reports some studies applying PARAFAC together with unfolded principal component analysis (UPCA) to discriminate between commercial samples of virgin and pure olive oils [**22**], to detect adulterations in EVOO samples from the PDO [**23**], or PARAFAC with unfolded partial least-squares-discriminant analysis (UPLS-DA) to detect adulteration of olive oils with other vegetable oils and to quantify the proportion in binary blends [**24**].In all these studies, fluorescence spectroscopy was mainly employed. As far as we know, no studies have been reported where these algorithms are combined with chromatographic data and traditional supervised pattern recognition methods such as partial least-squares discriminant analysis (PLS-DA) and soft independent modelling of class analogies (SIMCA), or with recently introduced classification methods such as random forest (RF). Only few applications are known in the food field with second-order data to authenticate the cultivar of extra-virgin olive oils.

The aim of this study is to discriminate between arbequina extra-virgin olive oil from extra-virgin olive oils from other cultivars, using three-way data to develop multivariate

105 classification methods. For this purpose, we have developed a quick analytical method using
106 high performance liquid chromatography coupled to a UV absorption diode array detector
107 (HPLC-DAD). The second-order data were processed with PLS-DA, SIMCA and RF, in their
108 original format or by first reducing them using MCR and a newly proposed approach. In
109 addition, a set of quality metrics: (i) sensitivity, (ii) specificity, (iii) positive (or precision) and
110 negative predictive values, (iv) Youden index, (v) positive and negative likelihood ratios, (vi)
111 classification odds ratio; (vii) F-measure (or F-score), (viii) discriminant power, (ix) efficiency
112 (or accuracy), (x) AUC (area under the receiver operating curve), (xi) G-mean; (xii) Matthews
113 correlation coefficient and (xiii) Kappa coefficient, were used to assess the performance of
114 the classifications.

115

116 **2. Materials and methods**

117

118 **2.1 Chemicals and reagents**

119 HPLC-grade solvents (n-hexane, isopropanol, methanol and *tert*-butyl methyl ether (TBME))
120 were purchased from VWR International Eurolab, S.L. (Barcelona, Spain).

121 Other reagents, sodium methoxide (MeONa), citric acid monohydrate, and anhydride sodium
122 sulphate were provided by Merck (Darmstadt, Germany), sodium sulphate anhydrous was
123 provided by Panreac, S.L (Barcelona, Spain) and 3-chloroperbenzoic acid was purchased
124 from Sigma-Aldrich (Missouri, USA).

125

126 **2.2 Samples**

127 Sixty-four single-variety extra virgin olive oil samples (EVOO) of different regions from Spain
128 and olive fruit varieties were analysed. The samples were obtained directly from local
129 providers. More specifically, 20 samples were from 'arbequina' fruit variety and 44 samples
130 were from different fruit varieties which include: 'picual', 'hojiblanca', 'cornicabra', 'frantoio',
131 'koroneiki', 'picudo', 'royal', 'loaime', 'lechin', 'lucio', 'arbosana' and 'manzanilla'. Table 1
132 summarizes the different EVOO and the number of samples analysed.

133

TABLE 1

134

135 **2.3 Sample preparation**

First a transesterification reaction was applied to the EVOO samples. This reaction is a modification of the original procedure described by Bierdemann et al. [**25**]. For further information regarding to this modification see references [**26**,**27**]. Then, the methyl-transesterification fraction of the EVOO samples was epoxidised as follows: 1000 µL of the transesterified fraction were added to a 10 mL tube and mixture with 1000 µL of a solution of 5% (m/v) 3-chloroperbenzoic acid in TBME. The tube was stirred for 20 s and then allowed to stand for 10 min. Next, 4 mL n-hexane and 1 mL 20% sodium sulphate anhydrous in water were added, and the mixture was shaken. The aqueous phase was removed with a Pasteur pipet and finally the organic fraction was filtered using a syringe filter of polytetrafluoroethylene (PTFE) membrane with a 0.22 µm pore diameter. The solution was stored in cold until analysis.

For chromatographic analysis, 200 µL of the stored solutions was transferred to a 2 mL HPLC vial. The epoxidisation step was carried out to enhance the difference between arbequina EVOOs and the ones from other cultivars.

2.4 Instrumentation

The chromatographic analysis was carried out with an Agilent 1100 series liquid chromatography (Santa Clara, CA) equipped with a G1316A column thermostat, G1311A quaternary pump, a G1379A degasser and a G1313A autosampler. Detection was performed with a G1315B diode-array detector (DAD). Agilent ChemStation software (rev.A.09.03 [1417]) for HPLC systems was used.

2.5 Chromatographic analysis

The chromatographic fingerprint from the epoxidised fraction was obtained by HPLC-DAD using a column Lichrospher® 100 CN (250×4 mm, i.d, 4 µm) provided by Merck (Darmstadt, Germany). During the analysis the column temperature was constant at 30 ºC. Isocratic chromatographic conditions were employed using a mixture of n-hexane/isopropanol (96:4, v/v) as mobile phase at a flow rate of 1.2 mL min^{-1}. The injection volume was 20 µL and the run time was only 8 min. The DAD collected spectra every 2 s in the range 190-400 nm, each 1 nm.

2.6 Chemometrics

The raw data files from each chromatogram were exported in 'comma separated value'

(CSV) format, and then converted to MATLAB format (version R2013b). The dimension of the matrix for each sample was of 1343×211 where 1343 is the number of rows corresponding to the number of elution times and 211 is the number of absorbance spectra recorded. It is important to notice that the chromatographic fingerprints from the epoxidised fraction were reproducible from sample to sample due to the short chromatographic run time (3-4 min); for this reason, it was not necessary to apply any alignment procedure.

The original dataset was randomly split into a training set, which was composed of 44 EVOO samples (14 EVOO samples from arbequina cultivar and 30 from non-arbequina cultivar) and an external validation set was made up with 20 EVOO samples (6 EVOO samples from arbequina cultivar and 14 from non-arbequina cultivar).

MCR-ALS and NPLS-DA were applied using the interface MVC2 MATLAB toolbox, freely available on the internet [**28**]. Conventional multivariate chemometrics pattern recognition such PCA, SIMCA and PLS-DA, were employed using PLS_Toolbox ver 8.5.1 (Eigenvector Research Inc., Wenatchee, WA). RF was employed using perClass ver 4.7 (Delft, Netherlands). All the interface graphics, MVC2 toolbox, PLS Toolbox and perClass were designed for MATLAB software (Mathworks Inc., Natick, MA, USA).

3. Results and discussion

A two-way data array was recorded for each EVOO sample. Figure 1 illustrates a chromatographic-spectral landscape for an EVOO sample from 'cornicabra' cultivar.

FIGURE 1

Variable reduction

Two strategies of variable reduction were employed: (i) strategy 1, named "decomposition and vector fusion" (DVF) and (ii) strategy 2, using MCR-ALS for the resolution into individual components. Figure 2 shows a flow chart of the two strategies performed.

FIGURE 2

(i) Strategy 1 for variable reduction: DVF

For each sample, the corresponding mean vectors in both time and spectral domains were

obtained. In this way, two individual vectors per sample were computed, a mean vector of size 1343×1 (time domain) and another mean vector of size 211×1 (spectral domain). These two vectors were then fused, so that the resulting fused vector was composed of 1544 variables. Finally, the fused vectors for all samples were grouped in a single matrix of dimension 64×1544 (64 samples and 1544 variables). Figure 3 displays the mean vectors in the time and spectral domain for an EVOO sample from 'cornicabra' cultivar, respectively. Figure 4 shows the overlay of the fused mean vectors from the 64 EVOO samples.

FIGURE 3

FIGURE 4

(ii) Strategy 2 for variable reduction: MCR-ALS

The successful application of this algorithm requires that enough selectivity exists in the spectral domain. If the samples show similar spectra, they cannot be resolved into individual components using MCR-ALS. In these cases, if the chromatograms are reproducible, matrix augmentation can be performed along the spectral domain before MCR-ALS is applied [**29**].

MCR-ALS was applied to the row-wise augmented matrix (i.e., along the spectral domain). The number of components was estimated using principal component analysis of the augmented data matrix under a series of constraints: non-negativity in both domain (time and spectral) and none unimodality. According to the PCA results, 8 components were selected, which explained 99.94% of the data variance. After MCR-ALS decomposition, the chromatographic fingerprint information was arranged into a matrix of dimension 64×8 (64 samples and 8 components), which was subsequently processed with PCA, PLS-DA, SIMCA and RF for classification purposes.

Exploratory analysis

Two PCA models were built using each of the matrices computed by strategies 1 and 2, to test if there was some natural grouping in the data set. Both PCA models were built with four principal components (PCs) (98.96% and 98.08% of explained variance for each strategy, respectively). They grouped the samples in a similar way.

Figure 5a and 5b show the scores score-score plot on the PC4 *vs* PC1 plane and 3D plot with PC1-PC2-PC3, respectevely. PC4 and PC1 explained 4.08% and 68.18% of the

variance, respectively. Two groups of EVOO cultivars are distinguished: the positive region of PC4 mainly groups the EVOOs of the arbequina cultivar, while the positive region of PC1 clusters the EVOOs of the other cultivars.

Figure 5 (a)

Figure 5 (b)

Conventional multivariate classification methods

As mentioned in the Introduction, multivariate classification methods using second-order data for the authentication of the cultivar kind of EVOOs are scarce, and the most commonly applied chemometric in these cases is linear discriminant analysis (LDA).

In the present report, three classification models were developed: (i) two well-known classification methods (PLS-DA and SIMCA) using the resulting matrices obtained from the application of both strategies 1 and 2 (see above) , and (ii) NPLS-DA over the raw three-way data array . The main aim was to test whether there was significant difference between the classification methods usually applied to first-order (PLS-DA and SIMCA) and second-order data (NPLS-DA) when the chromatographic fingerprinting methodology is applied.

For classification purposes with PLS-DA, the class "arbequina" was indexed with the value 1 and the class "non-arbequina" with the value 0. The classification threshold was established by the software around of the value 0.6 for the arbequina class.

The classification of the samples with SIMCA was carried out from both Q-reduced (Q) and Hotelling T^2-reduced values. The classification region for the arbequina class was established according to Q and T^2 values equal to 1, meaning that a sample must take values lower than 1 to be classified in the arbequina class.

Table 2 shows the specifications of the PLS-DA and SIMCA models.

Table 2

The raw three-way data array was analysed using NPLS-DA . The estimated number of latent variables (LVs) was 15 according to the leave-one-out cross-validation method. As in the PLS-DA model, the "arbequina" class was denoted using the number 1 and the "non-arbequina" class using the number 0.

The prediction results from both strategies were the same. Table 3 shows the results of PLS-DA, SIMCA and NPLS-DA models and Table 4 presents the classification quality metrics calculated from the prediction results on the external validation set.

Table 3

Table 4

In Table 3, the wrongly classified samples have been highlighted. As can be seen, PLS-DA and NPLS-DA classification models are more efficient than SIMCA. Furthermore, the former two models misclassified the same samples, which make sense since PLS-DA and NPLS-DA work similarly.

Emergent multivariate classification methods

Random forest (RF) was first employed to process the second-order data. This algorithm is a combination of several prediction trees, which then selects the best split at each node among a random selection of predictor variables. RF shows significant advantages about other more applied classification methods such as high capability in handling mixed or badly unbalanced datasets, flexibility with no formal assumption on data structure, and the ability to deal address complex non-linear systems, and therefore it is able to build a more robust classification model than other conventional algorithms. Moreover, RF readily handles larger numbers of predictors and the cross-validation is unnecessary because it generates an internal unbiased estimate of the generalization error (test error) as the forest building progresses. The potential of RF for modelling linear and nonlinear multivariate calibration allows to be used for feature selection too, with two different objectives: (i) to find the subset of features with the minimum possible generalization error, or (ii) to select the smallest possible subset with a given discrimination capability [**30**].

Both classification models using the reduced data sets by strategies 1 and 2 achieved the same results. In both cases, 20 trees were combined to perform the prediction of the classes of the EVOO samples. Table 5 shows the obtained classification contingency table on the external validation data set, and table 6 displays the prediction results and the different classification quality metrics for the RF models, respectively.

Table 5

Table 6

The RF results are significantly better than the obtained ones from the previously applied conventional classification methods. The sensibility, specificity and efficiency from PLS-DA and NPLS-DA were 0.67, 0.92 and 0.84, respectively, while the same performances featured by the RF model were 1.00, 0.92 and 0.95, respectively. This suggests that the analysis of second-order data with to a powerful algorithm such as RF is a promising methodology to authenticate cultivars of EVOO samples.

4. Conclusions

The potential of second-order (or) fingerprint data obtained using LC-DAD to identify and discriminate extra-virgin olive oils from 'arbequina' botanical variety in respect of other varieties of milled olive fruits has been proved. A new fast-methodology has been proposal for the quality control of the extra virgin olive oil from arbequina cultivar using three multivariate classification algorithms, including two widely-recognised methods (partial least-squares-discriminant analysis, PLS-DA, and soft independent modelling of class analogies, SIMCA) and a third one (random forest, RF) which is much less known and has been first used on second-order data. Surprisingly RF has shown itself to be the more efficient one in validation, yielding values of sensibility, specificity and accuracy of 1.00, 0.92 and 0.95, respectively, which are significantly better than the values found for the other methods.

Before building multivariate classification models, the raw three-way data matrices have been reduced by applying two strategies: (1) multivariate curve resolution-alternating least squares (MCR-ALS), and (2) a new strategy named "decomposition and vector fusion" (DVF) which has been proposed in this work and based on the fusion of the mean vector obtained from the signal profiles in both spectral and time domains. No differences on the performance classification are found when both strategies are applied.

318 **Reference**

[1] M.I. Covas, Olive oil and the cardiovascular system, Pharmacol. Res. 55 (2007) 175-186. https://doi.org/10.1016/j.phrs.2007.01.010

[2] R.W. Owen, A. Giacosa, W.E. Hull, R. Haubner, B. Spiegelhalder, H. Bartsch, The antioxidant/anticancer potential of phenolic compounds isolated from olive oil, Eur. J. Cancer 36 (2000) 1235-1247. https://doi.org/10.1016/S0959-8049(00)00103-9

[3] R. Aparicio, G. Luna, Characterisation of monovarietal virgin olive oils, Eur. J. Lipid Sci. Technol. 104 (2002) 614-627. https://doi.org/10.1002/1438-9312(200210)104:9/10<614::AID-EJLT614>3.0.CO;2-L

[4] C. Campestre, G. Angelini, C. Gasbarri, F. Angerosa, The compounds responsible for the sensory profile in monovarietal virgin olive oils, Molecules 22 (2017) 1833. https://doi.org/10.3390/molecules22111833

[5] G. Bertolini, G. Prevost, C. Messeri, G. Carignani, Olive germplasm: cultivars and world-wide collections, FAO, Rome, 1998.

[6] Regulation (EU) No 1151/2012 of the European Parliament and of the Council of 21 November 2012 on quality schemes for agricultural products and foodstuffs. Official Journal of the European Union. (2012) L343/1-29.

[7] Commission Implementing Regulation (EU) Nº 2017/597 of 15 March 2017 approving non-minor amendments to the specification for a name entered in the register of protected designations of origin and protected geographical indications (Estepa (PDO)). Official Journal of the European Union. (2017) L81/15.

[8] Commission Regulation (EU) Nº 1902/2004 of 29 October 2004 amending the specification of a name appearing in the Annex to Regulation (EC) No 1107/96 on the registration of geographical indications and designations of origin (Les Garrigues). Official Journal of the European Union. (2004) L328/1-73.

[9] Commission Regulation (EU) Nº 2156/2005 of 23 December 2005 amending the specification of a protected designation of origin listed in the Annex to Regulation (EC) No 1107/96 (Siurana) (PDO). Official Journal of the European Union. (2005) L342/1-54.

[10] N. Sinelli, M. Casale, V. Di Egidio, P. Oliveri, D. Bassi, D. Tura, E. Casiraghi, Varietal discrimination of extra virgin olive oils by near and mid infrared spectroscopy, Food Res. Int. 43 (2010) 2126-2131. https://doi.org/10.1016/j.foodres.2010.07.019

[11] A. Cichelli, G.P. Pertesana, High-performance liquid chromatography analysis of chlorophylls, pheophytins and carotenoids in virgin olive oils: chemometric approach to variety classification, J. Chromatogr. A 1046 (2004) 141-146. https://doi.org/10.1016/j.chroma.2004.06.093

[12] A. Bajoub, S. Medina-Rodríguez, M. Gómez-Romero, E.A. Ajal, M.G. Bagur-González, A. Fernández-Gutiérrez, A. Carrasco-Pancorbo, Assesing the varietal origin of extra-virgin olive oil using liquid chromatography fingerprints of phenolic compound, data fusion and chemometrics, Food Chem. 215 (2017) 245-255. https://doi.org/10.1016/j.foodchem.2016.07.140

[13] N.P. Kalogiouri, R. Aalizadeh, N.S. Thomaidis, Application of an advanced and wide scope non-target screeninworkflow with LC-ESI-QTOF-MS and chemometrics for the classification of the Greek olive oil varieties, Food Chem. 256 (2018) 53-61. https://doi.org/10.1016/j.foodchem.2018.02.101

[14] M.B. Mohamed, F. Gusami, S.B. Ali, F. Radhouani, J. Faghim, T. Triki, N.G. Kammoun, C. Baffi, L. Lucini, C. Benincasa, The LC-MS/MS characterization of phenolic

compounds in leaves allows classifying olive cultivars grown in South Tunisia, Biochem. Syst. Ecol. 78 (2018) 84-90. https://doi.org/10.1016/j.bse.2018.04.005

[15] M.R. Alves, S.C. Cunha, J.S. Amaral, J.A. Pereira, M.B Oliveira, Classification of PDO olive oils on the basis of their sterol composition by multivariate analysis, Anal. Chim. Acta 549 (2005) 166-178. https://doi.org/10.1016/j.aca.2005.06.033

[16] G. Gurdeniz, B. Ozen, F. Tokatli, Classificaton of Turkish olive oils with respect to cultivar, geographic origin and harvest year, using fatty acid profile and mid-IR spectroscopy, Eur. Food Res. Technol. 227 (2008) 1275-1281. https://doi.org/10.1007/s00217-008-0845-7

[17] A. Kritioti, G. Menexes, C. Drouza, Chemometric characterization of virgin olive oils of the two major Cypriot cultivars based on their fatty acid composition, Food Res. Int. 103 (2018) 426-437. https://doi.org/10.1016/j.foodres.2017.10.064

[18] L. Cuadros-Rodríguez, C. Ruiz-Samblás, L. Valverde-Som, E. Pérez-Castaño, A. González-Casado. Chromatographic fingerprinting: An innovative approach for food 'identitation' and food authentication – A tutorial, Talanta. 909 (2016) 9-23. https://doi.org/10.1016/j.aca.2015.12.042.

[19] R. Bro, Parafac. Tutorial and applications, Chemom. Intell. Lab. Syst. 38 (1997) 149-171. https://doi.org/10.1016/S0169-7439(97)00032-4

[20] R. Tauler, Multivariate curve resolution applied to second order data, Chemom. Intell. Lab. Syst. 30 (1995) 133-146. https://doi.org/10.1016/0169-7439(95)00047-X

[21] A.C. Olivieri, On a versatile second-order multivariate calibration method based on partial least-squares and residual bilinearization: second-order advantage and precision properties, J. Chemometrics. 19 (2005) 253-265. https://doi.org/10.1002/cem.927

[22] F. Guimet, J. Ferré, R. Boqué, F.X. Rius, Application of unfold principal component analysis and parallel factor analysis to the exploratory analysis of olive oils by means of excitacion-emission matrix fluorescence spectroscopy, Anal. Chim. Acta. 515 (2004) 75-85. https://doi.org/10.1016/j.aca.2004.01.008

[23] F. Guimet, J. Ferré, R. Boqué, Rapid detection of olive-pomace oil adulteration in extra virgin olive oils from the protected denomination of origin "Siurana" using excitation-emision fluorescence spectroscopy and three-way methods of analysis, Anal. Chim. Acta. 544 (2005) 143-152. https://doi.org/10.1016/j.aca.2005.02.013

[24] I. Durán Merás, J. Domínguez Manzano, D. Airado Rodríguez, A. Muñoz de la Peña, Detection and quantification of extra virgin olive oil adulteration by means of autofluorescence excitacion-emission profiles combined with multi-way classification, Talanta. 178 (2018) 751-762. https://doi.org/10.1016/j.talanta.2017.09.095

[25] M. Bierdermann, K. Grob, Mariani, Transesterification and on-line LC-GC for determining the sum of free and esterified in edible oils and fats, Eur. J. Lipid Sci. Technol. 95 (1993) 127-133. https://doi.org/10.1002/lipi.19930950403

[26] A.M. Jiménez-Carvelo, E. Pérez-Castaño, A. González-Casado, L. Cuadros-Rodríguez, One input-class and two input-class classifications for differentiating olive oil from other edible vegetable oils by use of the normal-phase liquid chrtomatography fingerprint of the methyl-transesterified fraction, Food Chem. 221 (2017) 1784-1791. https://doi.org/10.1016/j.foodchem.2016.10.103

[27] A.M. Jiménez-Carvelo, A. González-Casado, E. Pérez-Castaño, L. Cuadros-Rodríguez, Fast-HPLC fingerprinting to discriminate olive oil from other edible vegetable oils by multivariate classification methods, J AOAC. 100 (2) (2017) 345-350. https://doi.org/10.5740/jaoacint.16-0411

[28] A.C. Olivieri, H.L. Wu, R.Q. Yu, MVC2: a matlab graphical interface toolbox for second-order multivariate calibration, Chemom. Intell. Lab. Syst. 96 (2009) 246-251. https://doi.org/10.1016/j.chemolab.2009.02.005

[29] P. L. Pisano, M.F. Silva, A.C. Olivieri, Exploration of liquid chromatographic-diode array for Argentinean wines by extended multivariate curve resolution, Chemom. Intell. Lab. Syst. 132 (2014) 1-7. https://doi.org/10.1016/j.chemolab.2013.12.010

[30] C. Ruiz-Samblas, J.M. Cadenas, D.A. Pelta, L. Cuadros-Rodríguez, Application of data mining methods for classification and prediction of olive oil blends with other vegetable oils, Anal. Bioanal. Chem. 406 (2014) 2591-2601. https://doi.org/10.1007/s00216-014-7677-z

Tables

Table 1. Classes and olive fruit varieties of extra virgin olive oil analysed.

Class	Fruit varieties	Nº samples
'Arbequina' (20 samples)	'arbequina'	20
'Non-arbequina' (44 samples)	'picual'	10
	'hojiblanca'	4
	'cornicabra'	5
	'frantoio'	3
	'koroneiki'	3
	'picudo'	4
	'royal'	3
	'loaime'	3
	'lechin'	1
	'lucio'	3
	'arbosana'	2
	'manzanilla'	3
	Total	**64**

Table 2. Characteristics of the PLS-DA and SIMCA models

	PLS-DA		SIMCA			
	LVs	% var	PCs 'Arb-Class'	% var	PCs 'nArb-Class'	% var
(a) Strategy 1						
	4	98.83	4	99.48	5	99.60
(b) Strategy 2						
	5	97.29	6	99.92	6	99.87

328

Table 3. Prediction results of arbequina and non-arbequina classification from the external validation set using PLS-DA, SIMCA and NPLS-DA.

Class	Sample number	Class Ref	PLSDA	SIMCA	NPLS-DA
			Clas Pred	Clas Pred	Clas Pred
Arbequina (Arb)	28	1	0	0	0
	30	1	0	0	0
	67	1	1	1	1
	69	1	1	1	1
	70	1	1	1	1
	74	1	1	0	1
Non-arbequina (nArb)	1	0	0	0	0
	10	0	0	0	0
	14	0	0	0	0
	16	0	0	0	0
	19	0	0	0	0
	34	0	0	0	0
	37	0	0	0	0
	39	0	0	1	0
	48	0	0	0	0
	51	0	0	0	0
	59	0	0	0	0
	61	0	0	0	0
	73	0	1	1	1

Ref: reference; Pred: predicted

329

330

331

332

Table 4. Values of the quality metrics from the conventional multivariate classification methods.

Performance features	PLS-DA	SIMCA	NPLS-DA
Sensibility (Recall)	0.67	0.50	0.67
Specificity	0.92	0.85	0.92
Positive predictive value (Precision)	0.80	0.60	0.80
Negative predictive value	0.86	0.79	0.86
Youden index	0.59	0.35	0.59
Positive likelihood ratio	8.67	3.25	8.67
Negative likelihood ratio	0.36	0.59	0.36
Classification odds ratio	24.00	5.50	24.00
F-measure	0.73	0.55	0.73
Discriminant power	0.76	0.41	0.76
Efficiency (or Accuracy)	0.84	0.74	0.84
AUC (Correctly classified rate)	0.79	0.67	0.79
G-mean	0.78	0.65	0.78
Matthews correlation coefficient	0.62	0.36	0.62
Kappa coefficient	0.62	0.36	0.62

333

334

335

Table 5. Contingency charts for the RF classification models from the same external validation set used for the SIMCA, PLS-DA and NPLS-DA models.

		Decision of the classifier		
		Arb class	**nArb class**	*Total*
True class	**Arb class**	6	0	*6*
	nArb class	2	12	*14*
	Total	*8*	*12*	*20*

340

Table 6. Values of the classification quality metrics from the RF models.

Performance features	**RF (strategy 1)**	**RF (strategy 2)**
Sensibility (Recall)	1.00	1.00
Specificity	0.92	0.92
Positive predictive value (Precision)	0.76	0.76
Negative predictive value	1.00	1.00
Youden index	0.92	0.92
Positive likelihood ratio	13.00	13.00
Negative likelihood ratio	0.00	0.00
Classification odds ratio	–	–
F-measure	0.92	0.92
Discriminant power	–	–
Efficiency (or Accuracy)	0.95	0.95
AUC (Correctly classified rate)	0.96	0.96
G-mean	0.96	0.96
Matthews correlation coefficient	0.89	0.89
Kappa coefficient	0.88	0.88

The hyphen "–" indicates that the performance feature cannot be determined

341

FIGURE CAPTIONS

Figure 1. Time-wavelength chromatographic landscape of an extra virgin olive oil from 'cornicabra' cultivar.

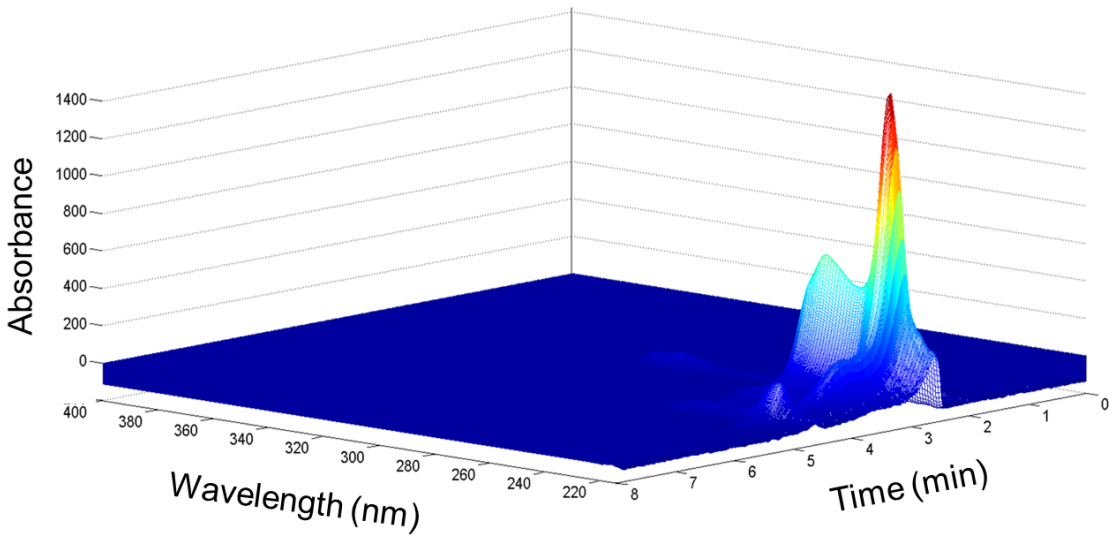
Figure 2. Flow chart showing the strategies applied for the treatment of the two-way data (N = number of objects (EVOO samples); M = number of variables in the spectral domain (wavelengths of the UV absorption spectrum); T = number of variables in the time domain (retention times of the chromatogram); L = number of latent variables (principal components)).

Figure 3. Plot of the mean vectors for an EVOO sample from 'cornicabra' cultivar: (a) mean vector in the time domain and (b) mean vector in the spectral domain.

Figure 4. Overlay of fused mean vectors of all EVOO samples.

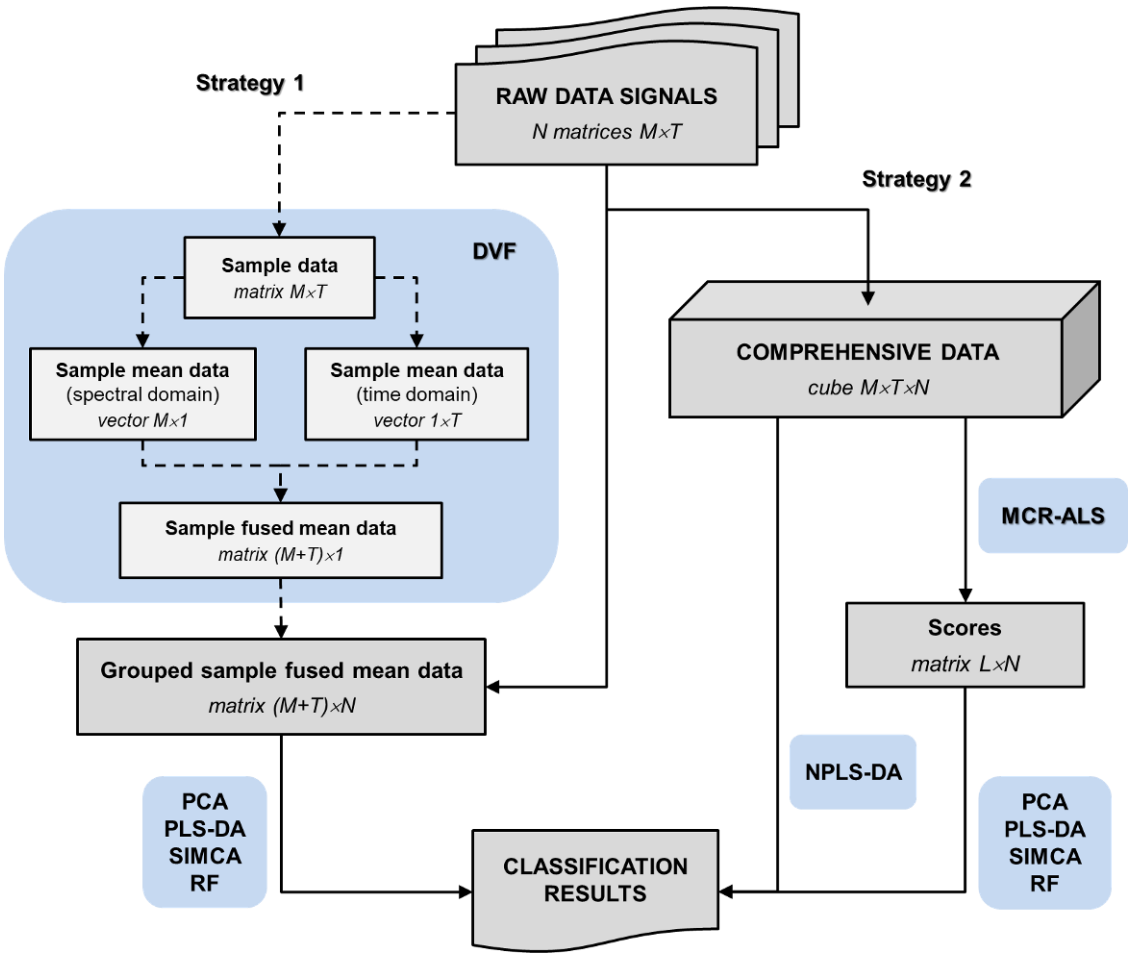
Figure 5. (a) PC1-PC2-PC3 and (b) PC4 vs PC1 plot from the matrix obtained from application of MCR of the epoxidised fraction of the 64 EVOO samples from different cultivars.

<Figure 1>


Absorbance
1400
1200
1000
800
600
400
200
0
400
380
360
340
320
300
280
260
240
220
Wavelength (nm)
8
7
6
5
4
3
2
1
0
Time (min)

366 **<Figure 2>**

367



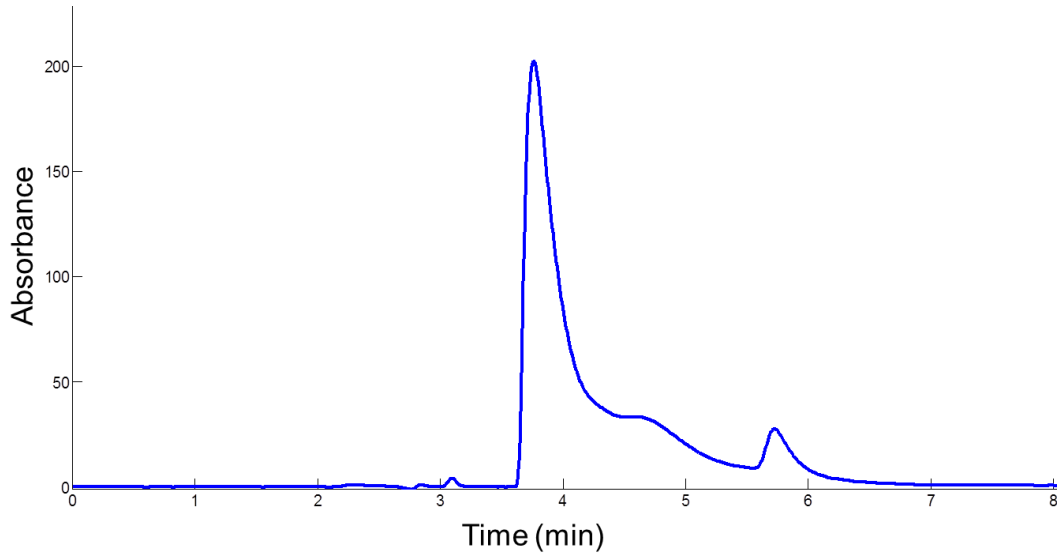
368

369

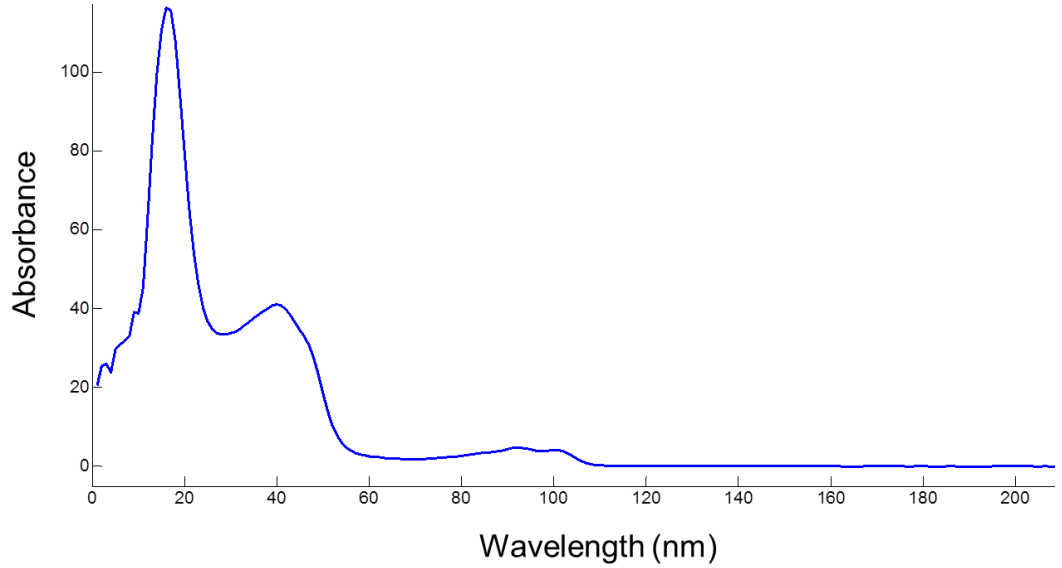
370

<Figure 3>

a)

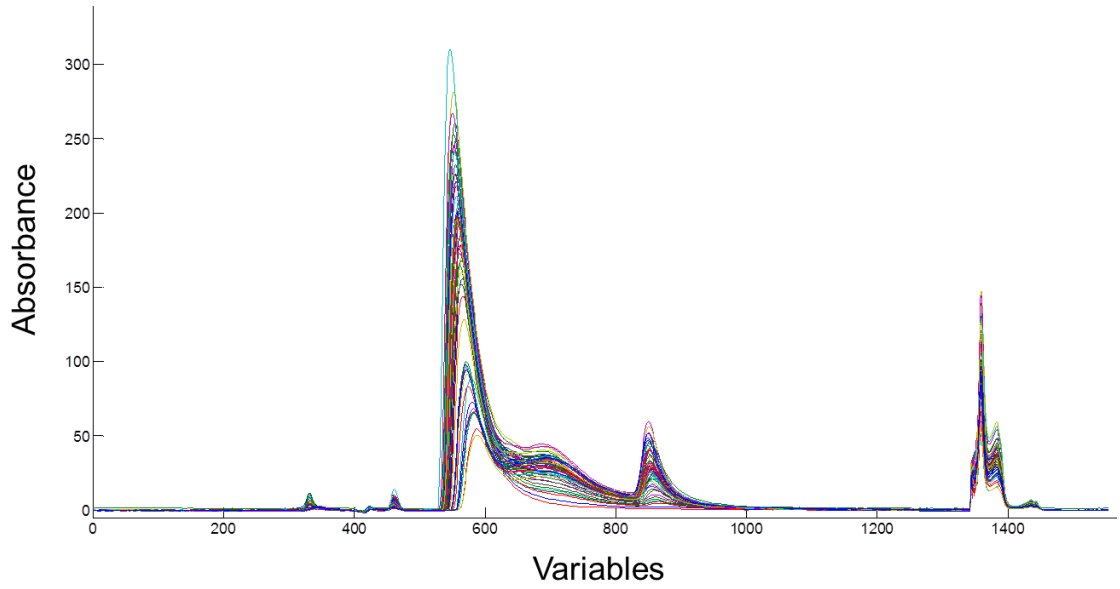

Absorbance
Time (min)
0
50
100
150
200
0
1
2
3
4
5
6
7
8

b)


Absorbance
Wavelength (nm)
0
20
40
60
80
100
0
20
40
60
80
100
120
140
160
180
200

376 **<Figure 4>**

377



378

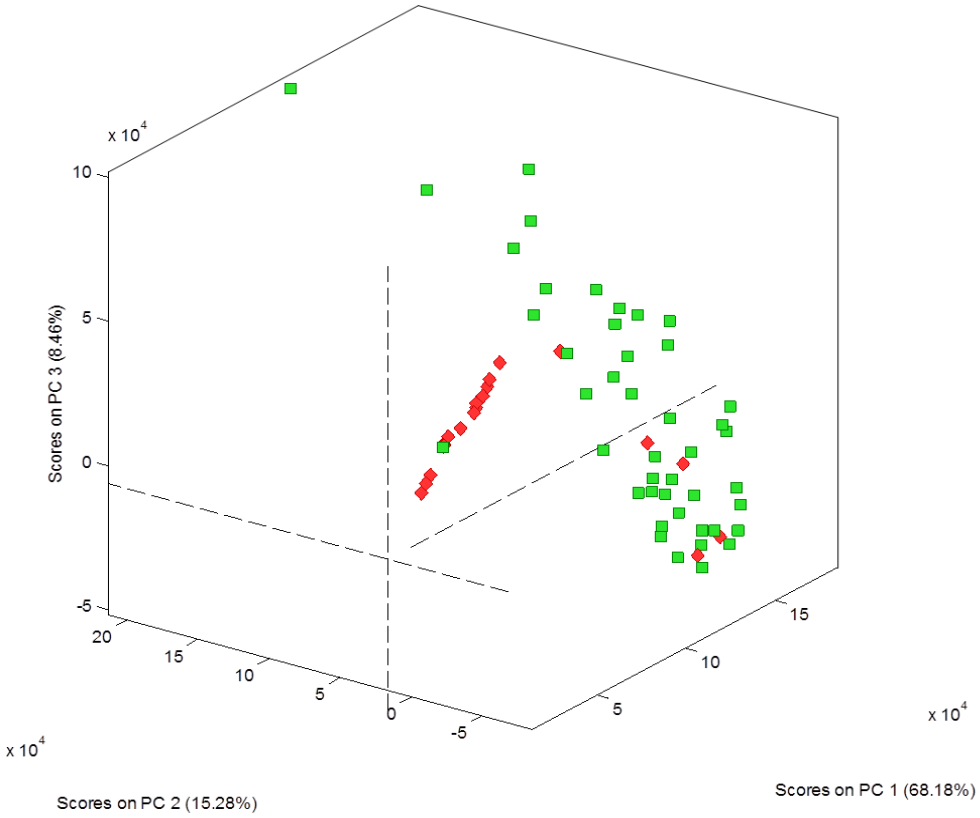
379

380

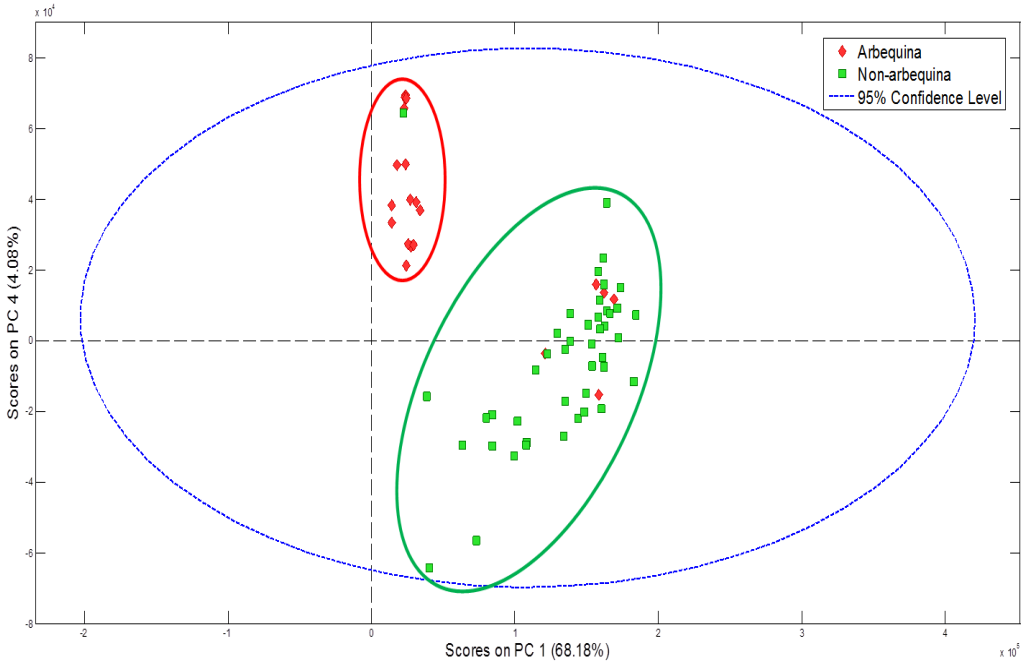
381 **<Figure 5>**

382

(a)



(b)



383

384