Food Control 137 (2022) 108957

Contents lists available at ScienceDirect

Food Control

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

Instrument-agnostic multivariate models from normal phase liquid chromatographic fingerprinting. A case study: Authentication of olive oil

Christian H. Pérez-Beltrán, Ana M. Jiménez-Carvelo *, Sandra Martín-Torres, Fidel Ortega-Gavilán, Luis Cuadros-Rodríguez

Department of Analytical Chemistry, Faculty of Sciences, University of Granada, C/ Fuentenueva, s/n, E-18071, Granada, Spain

ARTICLE INFO

Keywords:
Max 6)
Instrument-agnostic chromatographic fingerprints
Instrument-independent multivariate models
Data mining and chemometrics
Olive oil authentication

ABSTRACT

The application of non-targeted analytical strategies such as instrumental chromatographic fingerprinting is commonly applied in the field of food authentication/food quality. Although the multivariate methods developed to date are able to solve any authenticity problem, they remain dependent on the instrument state where the signals were acquired, which difficult their transfer to other laboratories. The aim of this research is to develop multivariate models independent of both instrument state and time at which the signals were acquired. For this, chromatograms obtained from the polar fraction of different olive oil samples analysed by (NP)UHPLC-UV/Vis are transformed to instrument-agnostic fingerprints. Instrument independence is achieved by transferring the chromatographic behaviour of an 'ad-hoc' external standards mixture solution analysed throughout an analysis sequence to the remaining analysed samples.

The SIMCA models developed from the chromatographic fingerprint matrix before and after instrument-agnostizing showed significant differences in the number of samples classified as "inconclusive", with the after model showing the best results. Furthermore, the PLS-DA and SVM models obtained before and after signal instrument-agnostizing showed similar outcomes. The main conclusion of the work has been to verify that the instrument-agnostizing methodology could allow the building of multivariate classification models which could be transferred among different laboratories as they are not influenced by the signal acquisition time.

1. Introduction

The untargeted approach is an emergent approach which is increasingly used in the field of food authentication/food quality. Untargeted methodology is focused on the study of unspecific instrumental signals without taking on any previous knowledge of relevant/irrelevant food components and it is mainly represented by fingerprinting methodology (Muñoz Olivas, 2004; Creydt & Fischer, 2020). In this sense, the instrumental fingerprint of a foodstuff can be defined as a non-specific signal that contains sufficient information about the chemical composition or structure of a food product or a food commodity to be able to unequivocally characterise and/or differentiate it from others similar foodstuffs (Cuadros Rodríguez et al., 2016a).

The application of instrumental fingerprinting methodology involves resorting to advanced mathematical data processing methods to extract useful information which is not obvious and not explicitly shown, such as data mining/chemometric methods. Usually, the recorded analytical signal is subjected to a previous pre-processing in order to clean it before being used for the development of a multivariate model. The most commonly used pre-processing techniques are, autoscaling, mean centring, noise filtering, baseline correction and normalization (Jiménez Carvelo et al., 2020). In the case of chromatographic signals, it is also necessary to carry out a peak alignment. This last pre-processing step is probably the most important since retention times (RT) are often shifted among chromatographic analyses. There are different algorithms for peak alignment being COW (Tomasi et al., 2004) and icoshift (Tomasi et al., 2011) the most commonly used in chromatography.

A large amount of literature is available on analytical methods using different analytical techniques together with data mining/chemometric methods in the food science field, which are focused on solving almost any authentication or quality problems (Boccard & Rudaz, 2020; Jiménez Carvelo & Cuadros Rodríguez, 2021; Oliveri et al., 2020; Tahir et al., 2022). Despite of, there is an important challenge still to be solved: reposted multivariate methods are based on the instrumental

* Corresponding author.
E-mail address: amariajc@ugr.es (A.M. Jiménez-Carvelo).

https://doi.org/10.1016/j.foodcont.2022.108957
Received 24 November 2021; Received in revised form 24 February 2022; Accepted 5 March 2022
Available online 8 March 2022

fingerprints acquired by a single analytical instrument, at a specific time and under particular conditions; this are instrument-sensitive fingerprints. This leads to multivariate models which are dependent of the analytical laboratory where the data have been acquired. The performance of those models with samples analysed by different manufacturer instrument or by the same one but in different time periods is largely unknown and by experience unsuccessful. Focussing towards a global or universal model, more fundamental work is required.

Despite of different attempts to create universal linear retention indexes (LRI) in liquid chromatography (Rigano et al., 2021), such activity aims to identify the compounds of interest in the sample and not to standardize the instrumental fingerprint. In fact, a common occurrence is to find some non-negligible variations in retention times or even in peak intensities when carrying out replicate chromatographic analysis. These scrolling on the axes of the chromatographic intensity/time signals (or chromatograms) make it difficult to use the recorded signals to create a representative database capable of being used for reliable comparisons/identifications or for multivariate classification or quantitation model building.

Some proposal regarding the standardization of spectroscopic signals, usually NIR and Raman for quality control purposes in the medical and pharmaceutical fields, have been reported (Fornasaro et al., 2020; Gou et al., 2018; Zhang et al., 2019). This methodological practice has been called 'instrument cloning' and is applied in order to obtain a 'transfer model', mainly calibration models for analytical quantitation, was proposed by Wang et al. (1991). This procedure is based on the statement that the position coordinates of a spectrum are characteristic of each instrument and that they remain practically invariant over time. Generally, calibration transfer is implemented as follows: the spectrum of a sample obtained by the NIR or Raman equipment from which a particular multivariate calibration model has been developed (Master instrument) must have a similar spectral profile to the spectrum obtained by the equipment to which the model is to be transferred (Satellite instrument) (Folch Fortuny et al., 2017). Despite the recent industrial and technological progress, the spectra obtained by different NIRS instruments differ for various reasons, among which the instrument configuration and optics are the most common. Thus, each transfer model is only applicable to pairs of instruments.

However, no equivalent strategy was suggested specifically for chromatographic signals. In this context, Cuadros Rodríguez et al. (2021a, 2021b) have recently proposed an innovative methodology to be followed in order to obtain standardized instrumental fingerprints when the gas and liquid chromatography are employed; this methodology has been termed by the authors as instrument-agnostizing (Cuadros Rodríguez et al., 2021a, 2021b). It was proof to be able to standardize conventional chromatograms so that the new instrument-agnostic signal (fingerprint) is independent of the chromatographic state or the date of analysis, so that chromatographic fingerprints acquired from different instruments states (two or more) should have a high degree of similarity. For this purpose, both internal and external chemical standards series are used as instrumental references. Briefly, this methodology is summarised below: firstly, it is performed a stage for setting up an invariant set of standard retention scores (SRS) from the external standards, which is only applied once; then, the agnostizing step is performed in which both intensities and retention times of the signal is standardized using the previously established SRS. Note that this is the first methodology that attempt to obtain a database of EVOO instrumental fingerprints and, thus it can be employed as potential tool to achieve multivariate 'instrument-agnostic' models.

Olive oil is one of the main vegetable oils chosen by consumers due to its nutritional characteristics and health benefits, being one of most regulated and controlled foodstuffs in the European Union (EU). It should be noted that EU legislation allows the blending of olive oil with other vegetable oils, however, some European producer countries, such as Spain or Italy, have specific legislation which forbids the blending of olive oil with other vegetable oils. There are three different European marketing quality categories of edible olive oil: (i) extra virgin olive oil (EVOO), (ii) virgin olive oil (VOO) and (iii) olive oil (OO), the latter being a blend of chemically refined olive oil and EVOO/VOO. These oils vary in price and quality, due to their organoleptic and physico-chemical properties. In fact, EVOO and VOO achieve much higher prices on international markets than any other type of vegetable oil, which makes it potentially considered to be adulterated with lower quality edible vegetal oils, such as seed oils (e.g., sunflower oil), refined olive (ROO) oil and/or olive-pomace oil (OPO), in order to obtain a higher illicit profit. Currently, the European official method of analysis used to detect adulteration of EVOO/VOO with ROO or OPO involves carrying out several chemical analyses in order to determine specific analytical parameters such as ECN42 and to quantify some particular compounds (chemical markers) or family of compounds such as triterpene dialcohols or waxes, among others, using different sample treatments and/or analytical procedures for each one (Commission Regulation (EEC) No. 2568/91). As an example, the triterpene dialcohols such as erythrodiol and uvaol are separated from the unsaponifiable matter by thin-layer chromatography on a basic silica gel plate. The fractions recovered from the silica gel are derivatised into trimethylsilyl ethers and then analysed by gas chromatography. Thus, this method is highly time-consuming and entails a large consumption of chemicals.

Moreover, a wide number of different procedures for the adulteration detection of EVOO/VOO with different edible oils (sunflower, soybean, peanut, corn, rapeseed, hazelnut oils, among the most common) have been proposed (Zhang et al., 2021; Meenu et al., 2019). Basically, two methodologies outstand for this purpose: i) the use of nontargeted spectroscopic approaches, such as Raman (Duraipandian et al., 2019) and Fourier transform infrared (FTIR) (Abdallah et al., 2016; Karunathilaka et al., 2016); and ii) the use of high-performance liquid chromatography (HPLC) or gas chromatography (GC) for quantitative analysis of peculiar marker compounds (Mingchih et al., 2015; Jabeur et al., 2017).

Nonetheless, the adulteration of EVOO/VOO with refined olive oil and/or olive-pomace oil by means of HPLC and chemometrics has been addressed to a lesser extent. In fact, it was possible to find only five research studies involving one or both of these topics which mainly mass spectrometry (MS) as detection system and targeted approach are employed on minor polar compounds (Carranco et al., 2018; Drira et al., 2020; Li et al., 2021; Navratilova et al., 2022; Tata et al., 2022).It should be noted that all these studies were performed with non-standardized signals for the development of the multivariate models, what limit their implementation to routine analytical laboratory, being applicable only under the conditions of measurement under which they were carried out.

To date, the development of a single multivariate instrument-agnostic model has not been proposed in food authentication field using chromatographic signals. In this context, the current study proposes a multivariate analytical method for the detection of olive oil adulteration with ROO or OPO using agnostic-instrument chromatographic fingerprints for the first time. In this sense, this study proposes the use of the chromatographic fingerprints from the polar compounds fraction of the olive oils of different quality categories, acquired using normal phase ultra-high-performance liquid chromatographic coupled to an ultraviolet–visible molecular absorption detector ((NP)UHPLC-UV/Vis), as a source of analytical information to set up instrument-agnostic multivariate classification models. The discrimination results from each method and strategy were compared and ranked using several classification performance metrics, such as sensitivity, specificity, precision, efficiency (or accuracy), area under the receiver operative curve (AUC), among others. More details on the specific features of the classification strategies and the meaning of the classification metrics can be found in the tutorial published by Cuadros Rodríguez et al. (2016b).

2. Materials and methods

2.1. Chemicals

HPLC-grade solvents, such as n-hexane, 2-propanol and ethanol were employed within the study. *N*-Hexane was purchased from Panreac Quimica S.L.U. (Barcelona, Spain), 2-propanol from Honeywell (Deutschland, Germany) and ethanol from VWR (Darmstadt, Germany). Deionized water was obtained using a Milli-Q system (Millipore, Bedford, MA).

Chemical standards, such as 1,2,3-trimethyl benzene (TMB) provided by Sigma-Aldrich (St. Louis, USA), propiophenone (PROP) provided by AlfaAesar (Kandel, Germany), 2,5-dimetylphenol (2,5-DP) provided by Sigma-Aldrich (St. Louis, USA), 2-naftol (2-NAF) provided by ACROS (Geel, Belgium) and ethyl paraben (EPB) provided by Fluka Chemika (Buch, Germany) were employed to create the external standard mix (ESM) solution. Each of the chemicals were added into the mix at 12, 100, 16, 4 and 40 mg/L, respectively, using n-hexane/2-propanol 99/1 (v/v) as solvent.

2.2. Samples

A total of 88 vegetables oils samples were analysed: 35 extra-virgin olive oil samples (EVOO) of different regions from Spain, 4 virgin olive oils (VOO), 4 olive oil (OO), 5 refined olive oil (ROO), 4 olive-pomace oil (OPO), and 36 blends (BLE) of EVOO or VOO with ROO or OPO. These blends represented adulterated olive oils with other olive oils of poorer quality in 20, 40 and 60%.

2.3. Sample preparation

1 g of oil was placed in a 10 Ml tube and 4 Ml of n-hexane were added into the tube for further agitation with vortex for 10 s. Then, 1 Ml of ethanol/water 87/13 (v/v) mixture was added and vortexed for another 10 s. The polar fraction at the bottom of the tube was extracted and this step was repeated twice. Finally, the polar fraction was centrifuged for 3 min at 1500 g and further filtered with 0.22 μm nylon filters. The polar fraction solutions were frozen (-4 °C) and kept in the dark until analysis.

2.4. Chromatography

(NP)UHPLC-UV/Vis analysis was performed with a Dionex Ultimate 3000 UHPLC + Focused chromatography system (Thermo Scientific, Waltham, MA, USA) equipped with a RS autosampler and column compartment. Detection was performed with an RS variable wavelength detector. Chromeleon™ version 7.0 software was used to visualize and export data. A silica stationary phase column (ZORBAX RX-SIL, 150 × 4.6 mm i.d, 5 μm) coupled to a pre-column with the same diameter (12.5 × 4.6 mm) were used through all the analysis. Both pre-column and column were kept at 35 °C during the experimental work.

The chromatographic analysis of the ESM was performed 32 times along 6 days in order to considerer in the calculation of the SRS as much variability as possible. Additionally, the ESM was analysed at the beginning and at the end of each chromatographic run for further calculation of the SRS, and as quality control of the behaviour of the equipment. For this purpose, 1.5 Ml of the ESM were placed in a chromatographic glass vial for its corresponding analysis with a flow rate of 0.8 Ml/min during the entire operation. The gradient mode of the mobile phase was the following: the ESM was injected at time 0 and was eluted with hexane for 15 min. Then, solvent was changed to hexane-isopropanol 90/10 (v/v) for 2 min. Finally, from minute 17 to minute 21, the chromatographic system came back to the initial conditions of hexane 100%. Note that during the second day, ESM from the first day was analysed together in the same batch with ESM of day two; for day three, ESM from day two was analysed with ESM from day three; the same process was followed for the remaining days.

Just before the chromatographic analysis, 750 mL of the polar fraction solution, previously thawed, were added into a 2 Ml chromatographic glass vial, and then 180 Ml of TMB solution (100 mg/L in n-hexane/2-propanol 99/1, v/v) were added as an internal control standard. The vial was sealed and vortexed for 20 s and 5 mL of this solution were injected in the LC equipment. A flow rate of 1.2 Ml/min was kept during the entire operation. The gradient mode of the mobile phase was the following: the samples were injected at time 0 and were eluted with hexane for 1 min, at a flow rate of 1.2 mL/min. Then, solvent was changed to hexane-isopropanol 80/20 (v/v) for 3 min. Afterwards, the solvent was changed again to hexane/isopropanol 60/40 (v/v) for 4 min, going back to 80/20 (v/v) after 2 min. Finally, from min 10 to minute 13, the system came back to the initial conditions of hexane 100%.

2.5. Methodology: development of a multivariate model from instrument-agnostic chromatographic fingerprints

In order to be able to have multivariate models for common use, the chromatographic signals must be standardized, and then the multivariable models are developed. In this regard, the following steps were needed to obtain the instrument-agnostic chromatograms. All data (88 samples × 1950 variables) used in this study were exported from the instrument software to an Excel environment (.csv, comma separated values), and then converted to Matlab environment (.mat). In this way, each chromatogram was firstly turned into a data vector.

The description of the process of standardization of signals as well as building of the multivariate model can be summarised in 6 major steps:

1. Application of the automatic Whittaker filter to correct the baseline of raw chromatograms in which values of $\lambda = 100$ y $p = 0.001$ were selected. Λ indicates the baseline curvature to allow (the smaller this value, the more curved the baseline fit will be), whilst 'p' ($0 < p < 1$) indicates the asymmetry to use in the Whittaker filter (the smaller this value, the smaller the allowed negative proportion of the result that has been adjusted) (Wise et al., 2006).
2. Selection of the data to create the training and external validation data sets. The choice was performed through the Kennard-Stone algorithm already implemented in Matlab. The proportion of samples to include in the training and validation data sets was 70 and 30%, respectively. At the end, the training data set was composed by 61 samples, whilst the external validation set by 27.
3. Intensity normalization of the chromatogram from the 88 oil samples to create a homogeneous intensity scale. All the chromatographic intensities (height) were normalized taking as a reference the maximum peak of the internal standard TMB, assigning an intensity value = 1 on the y-axis. The peak of TMB was found among variables number 224 and 250.
4. Establishment of the standard retention scores (SRS) according to the protocol given by Cuadros et al. (2021b).
5. Replacement of the retention time values for the calculated SRS in each sample signal, in order to unify the scale on the x-axis.
6. Re-sampling to fix into a same number of variables all chromatograms. The specified range for the re-sampling function was from 1 to 5.8 considering the SRS scale, obtaining a variable reduction from 1256 to 575.

Once the chromatograms were standardized, different multivariate models were developed. The corresponding information can be found in the following section.

2.6. Multivariate analysis

After performing the six major steps outlined in subsection 2.5, exploratory analysis using principal component analysis (PCA) was performed to detect possible natural grouping or outliers in the data.

Furthermore, soft independent modelling of class analogies (SIMCA), partial least squares - discriminant analysis (PLS-DA) and support vector machine (SVM) were used to create alternative classification models capable to identify EVOO and VOO, and detect blends of these olive oils adulterated with ROO and OPO). For this, PLS_Toolbox (version 8.6.1, 2019, Eigenvector Research Inc., Manson, WA, USA) was used under Matlab (version R2013b, 8.2.0.701, The Mathworks Inc. MA, USA) environment.

3. Results and discussion

The raw chromatograms obtained can be observed in Fig. 1. The chromatograms from EVOO and VOO have some similarities around minutes 4.00–4.20 and 4.80–5.40 that are attributed to their chemical composition. However, these same two chromatograms also show differences of intensity between minutes 4.30–4.70 and a lack of a small peak around minute 6.80. Such differences could be attributed to the presence of small concentrations of defective compounds of the VOO that diminish its category from EVOO to VOO.

Additionally, in Fig. 1 B it can be appreciated that OO, ROO, OPO and BLE have similar fingerprints around times 4.70, 5.20 and 5.40 min. The pattern among the first three (OO, ROO and OPO) might be due to the similar chemical composition of these oils as they have all suffered a refining process to a greater or lesser extent. Moreover, the intense peak around at 1.40–1.60 min corresponds to the internal standard TMB added to each sample. In this regard, it can be observed that EVOO and VOO have a particular feature with respect to the chromatograms of the other oils, specifically, a bit more intense peaks around minutes 4.20–4.60.

In order to be able of agnostizing the raw chromatograms, the ESM was analysed before and after each analytical run (the corresponding chromatogram is shown in Fig. 2). The first eluted compound was the internal standard (TMB) around 2.30 min, then PROP, 2,5-DP and 2-NAF at 4.37, 8.47 and 9.46 min respectively.

Then, the first step was to normalize the intensity of all chromatograms considering the most intense peak before minute 2, which was the internal standard TMB. Afterwards, an invariant reference chemical system for normalizing the retention values was established. For this purpose, the estimation of SRS values was performed (see subsection 2.5). The second step focussed on the retention time normalization remains on the transference of retention standard scores from the ESM to any intensity-normalized chromatogram and involves the transformation of the chromatographic intensity-normalized vectors from the instrumental-dependent RT domain to an instrumental-agnostic SRS domain. Fig. 3 shows the overlapped chromatograms of the same six samples plotted in Fig. 1 after intensity normalization (time domain) and after agnostizing methodology on SRS domain.

3.1. Exploratory analysis

The raw chromatograms (which were conventionally aligned using icoshift algorithm) and the instrument-agnostic fingerprints were used to perform the different steps of this multivariate study. Note that when aligned raw chromatograms are used as input data to develop multivariate models, they will be referred to as before-agnostizing models. Conversely, if instrument-agnostic fingerprints are used, the after-agnostizing model term is then employed.

First, an exploratory PCA was performed for both data sets and their

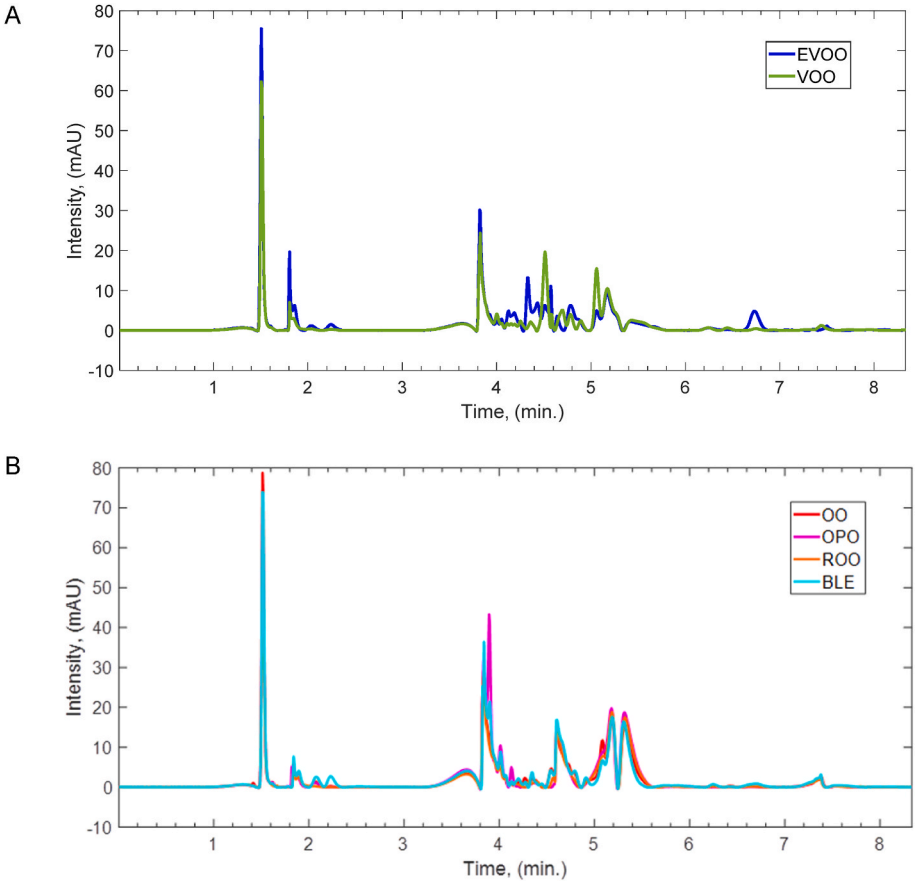

Fig. 1. Overlapped raw chromatograms of six olive oils considering one representative sample per olive oil category: A) EVOO and VOO, and B) OO, ROO, OPO and BLE. EVOO: extra virgin olive oil, VOO: virgin olive oil, OO: olive oil, ROO: refined olive oil, olive-pomace oil (OPO) and BLE: blend of EVOO or VOO with ROO or OPO.

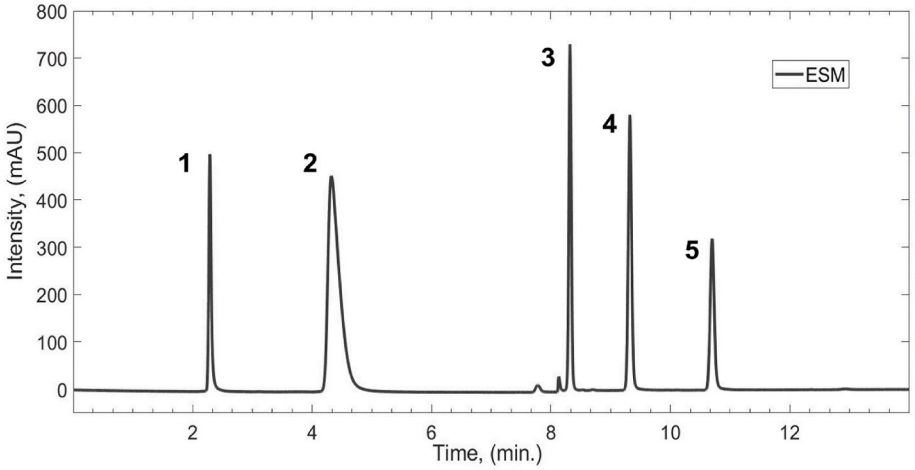

Fig. 2. Chromatogram of the external standard mix (ESM) composed by five different chemical components: 1) 1,2,3-trimethyl benzene (TMB); 2) propiophenone (PROP); 3) 2,5-dimetylphenol (2,5-DP); 4) 2-naftol (2-NAF); and 5) ethyl paraben (EPB).

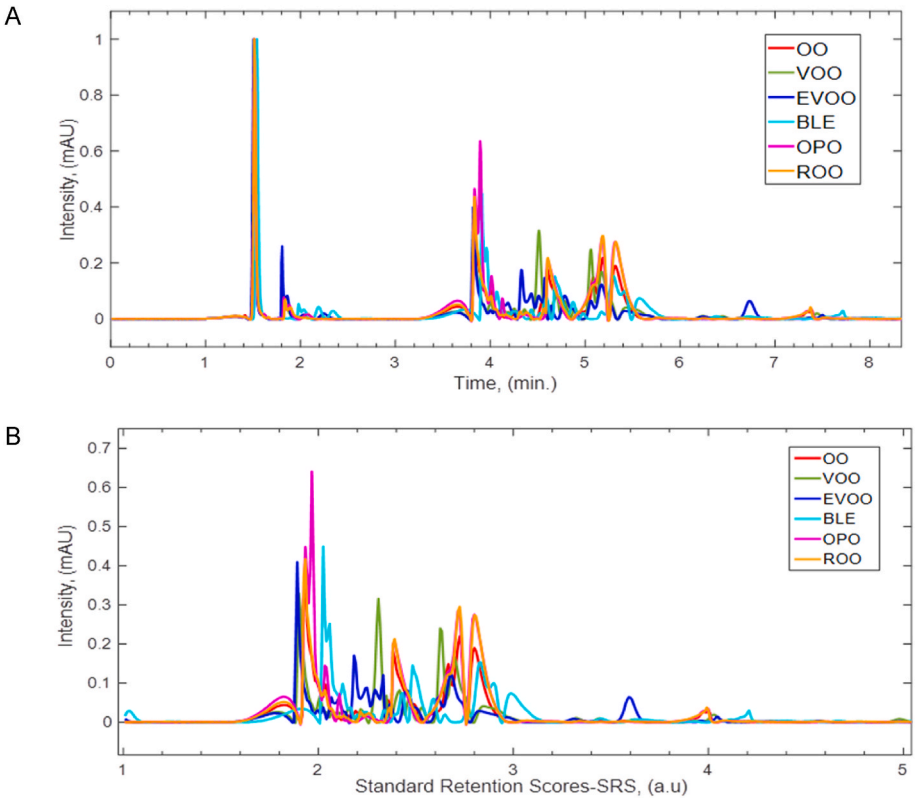

Fig. 3. Overlapped chromatograms of the same six different olive oil samples as Fig. 1 A) after intensity normalization (time domain) and B) after retention time normalization (SRS domain). EVOO: extra virgin olive oil, VOO: virgin olive oil, OO: olive oil, ROO: refined olive oil, olive-pomace oil (OPO) and BLE: blend of EVOO or VOO with ROO or OPO.

corresponding score plots can be observed in Fig. 4 A and B, respectively. The before-agnostizing PCA model was built considering 3 principal components (PCs) which explained 73% of the total variance with a root mean square error for cross validation (RMSECV) = 1.87, whilst after-agnostizing PCA model was developed with 4 PCs explaining 67% of the total variance with RMSECV = 0.03. Only PC1 vs PC2 were used to perform the score plots in both cases, since they provided the best grouping overview. The EVOO and VOO were grouped together in both PCA score plots; the same ensued for the OO, ROO and OPO samples, which could be attributed to the similar chemical composition.

However, the different pattern in the two scores plots deserves further comment. Fig. 4A clearly shows the differentiation between the two groups mentioned above, but the same is not evident in the layout shown in Fig. 4B. This is because PCA is not a classification method, but outputs groups based on the variability observed in the corresponding input signals. These results in the BLE oils being further separated into three subgroups, possibly because the agnostizing of the signals enhances the dissimilarity amount blended oils. However, this fact does not hinder the aim of the exploratory analysis, which was to show that there is a difference between the two concerned groups, which is clearly evident. As a consequence, the application of appropriate classification methods should yield good results.

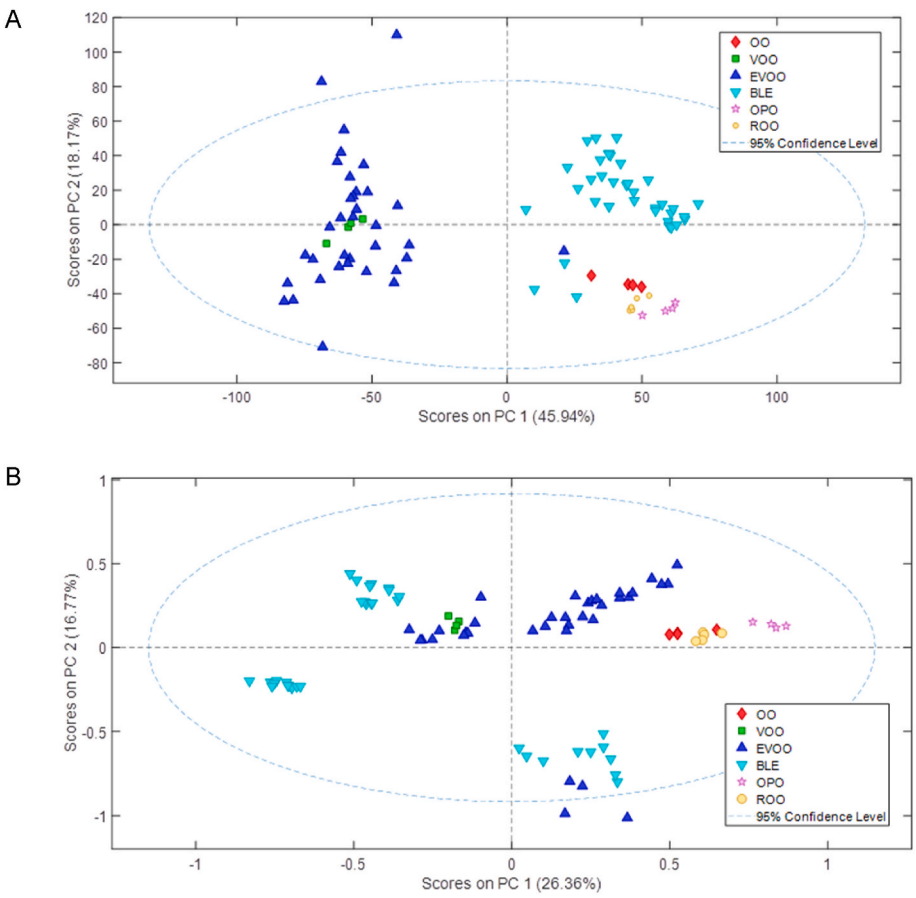

Fig. 4. Exploratory PCA score plots of 88 olive oil samples belonging to six different quality categories: A) before agnostizing, and B) after agnostizing. EVOO: extra virgin olive oil, VOO: virgin olive oil, OO: olive oil, ROO: refined olive oil, olive-pomace oil (OPO) and BLE: blend of EVOO or VOO with ROO or OPO.

In the same regard as the results presented in this study, Drira et al. (2020) could identify grouping trends applying PCA among the EVOO and the EVOO/OPO adulterated samples using the profile from the phenolic compounds, sterolic composition and antioxidants. Nonetheless, authors used only nine samples in total within the study, which is a very low number of samples to ensure that the used information of the chromatographic profiles is sufficient enough to discriminate between the different vegetable oils. Navratilova et al. (2022) could not observe clear groups of EVOO and EVOO/ROO analysing the polar fingerprints with PCA. Finally, Carranco et al., 2018 also used chromatographic polar fingerprints of different EVOO samples, and EVOO samples adulterated with ROO and sunflower oil as analytical signal for PCA. As a result, it was possible to find a tendency of the olive oils against the other vegetable oils samples, but authors do not mention if there was some pattern of EVOO against adulterated EVOO with ROO.

On the contrary to these studies, the current study included a wide number of samples, the PCA displayed a better grouping of all samples and it was possible to distinguish EVOO and VOO from OO, OPO, ROO and adulterated samples with PCA. This demonstrates that the instrumental fingerprint of the polar fraction contains the information of interest to authenticate the olive oil as discussed above.

3.2. Authentication of olive oil – discrimination multivariate models

For the sake of clarity, only the characteristics of each after-agnostizing model developed with instrument-agnostic fingerprints, as well as the classification plots, classification contingencies and classification performance metrics tables for each one is presented here. In order to compare these outcomes with the results of the before-agnostizing multivariate models, classification plots, contingencies and metrics tables can be found in supplementary material.

Several multivariate classification models were built using both data sets (before and after agnostizing) employing SIMCA, PLS-DA and SVM as data mining methods to find the best multivariate method capable to differentiate among EVOO, VOO or OO from ROO, OPO or BLE. For this, two classes were considered to build the models: class 1 (EVOO/VOO/OO) and class 2 (ROO/OPO/BLE). The training was set of 61 samples (24 EVOO, 2 VOO, 3 OO, 4 ROO, 3 OPO and 25 BLE) and further validated with a data set of 27 samples (11 EVOO, 2 VOO, 1 OO, 1 ROO, 1 OPO and 11 BLE), as outlined in subsection 2.5.

Firstly, SIMCA was employed. It is a multivariate classification method that builds models based on PCA and considers the classes independent from each other. For this particular case, the model was performed using 3 PCs for class 1 and 4 PCs for class 2, which explained 62.34% and 88.70% of the total variance, respectively. The classification outcomes can be evaluated using the Cooman's plot which is showed in Fig. 5. Coomans' plot is a visual representation of the separation between two classes, in which the two axes represent the normalized orthogonal distances of all the samples respect to each individual model. Optimally, the validation samples should be classified in the class 1 or class 2. In real conditions, some validation samples could be assigned to both classes simultaneously, in this case these samples are considered as inconclusive ones. In addition, some samples can be not recognized as belonging to any class.

It can be observed in Fig. 5 that there are three validation samples placed in the 'inconclusive' quadrant (bottom-left quadrant) which are samples of OO, ROO and blend of VOO (80%) with OPO (20%), respectively. The classification contingency is shown in Fig. 6.

When comparing these classification outcomes with those obtained from the after-agnostizing SIMCA model, it is shown that the agnostizing

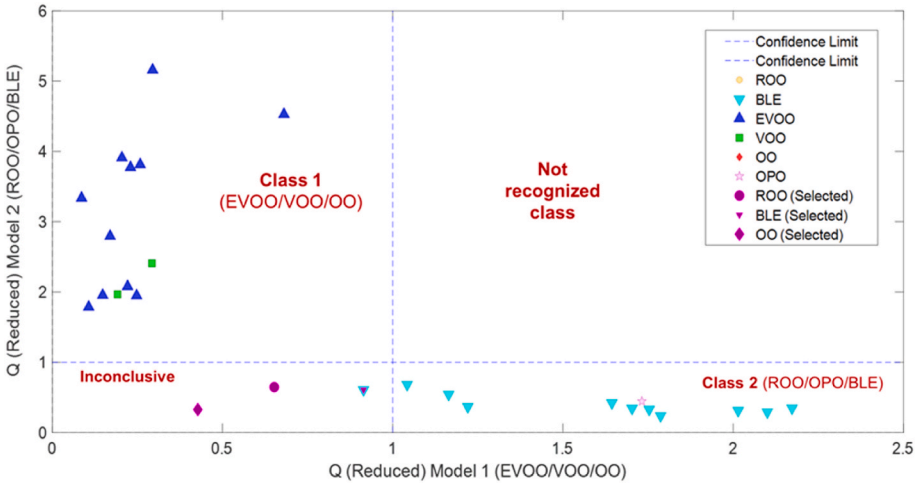

Fig. 5. Cooman's classification plot of the validation set samples from the after-agnostizing SIMCA model. EVOO: extra virgin olive oil, VOO: virgin olive oil, OO: olive oil, ROO: refined olive oil, olive-pomace oil (OPO) and BLE: blend of EVOO or VOO with ROO or OPO. *(Left upper quadrant 'Class 1' includes EVOO, VOO, and OO samples; right upper quadrant is for not recognized samples; left bottom quadrant is for inconclusive samples; right bottom quadrant 'Class 2' includes ROO, OPO and BLE samples; 3 samples were classified as 'inconclusive').*

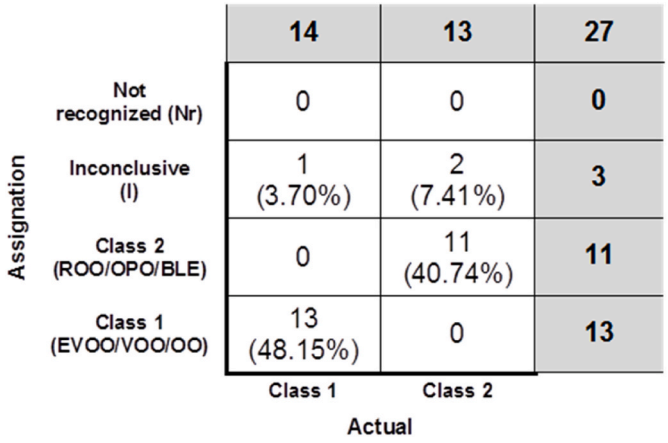

Fig. 6. Validation contingencies from the after-agnostizing SIMCA classification model. Class 1 (target class): EVOO: extra virgin olive oil, VOO: virgin olive oil, OO: olive oil) – Class 2 (non-target class): ROO: refined olive oil, OPO: olive-pomace oil, and BLE: blend of EVOO/VOO with ROO/OPO oils.

step improved the results using the standard icoshift alignment, since the before-agnostizing SIMCA model placed in the 'inconclusive' quadrant 10 samples, providing worse classification results (see supplementary material, Figs. S1 and S2).

The next multivariate method was PLS-DA, which was performed with 3 latent variables (LVs) that could explain 77.99% of the total variance. The classification results can be observed in Fig. 7 in which only one OO sample belonging to class 1 was classified in class 2, since it did not trespass the threshold of 0.5, associating it to be more similar to ROO and OPO BLE samples. The classification contingency can be observed in Fig. 8. Note that the performance of both before-agnostizing and after-agnostizing PLS-DA models was the same (see supplementary material, Figs. S3 and S4).

The after-agnostizing SVM classification model was performed considering the same two classes. The Kernel type algorithm radial basis function (RBF) with gamma and cost values, established by default in the PLS_Toolbox software, was applied. As observed in Fig. 9, all samples were classified within their corresponding classes. In this case, the OO sample belonging to the class 1, previously misclassified by SIMCA and PLS-DA classification models, was classified correctly. The classification 'contingency chart' can be observed in Fig. 10.

The same classification performance results were found using raw chromatograms (before-agnostizing) and instrument-agnostic

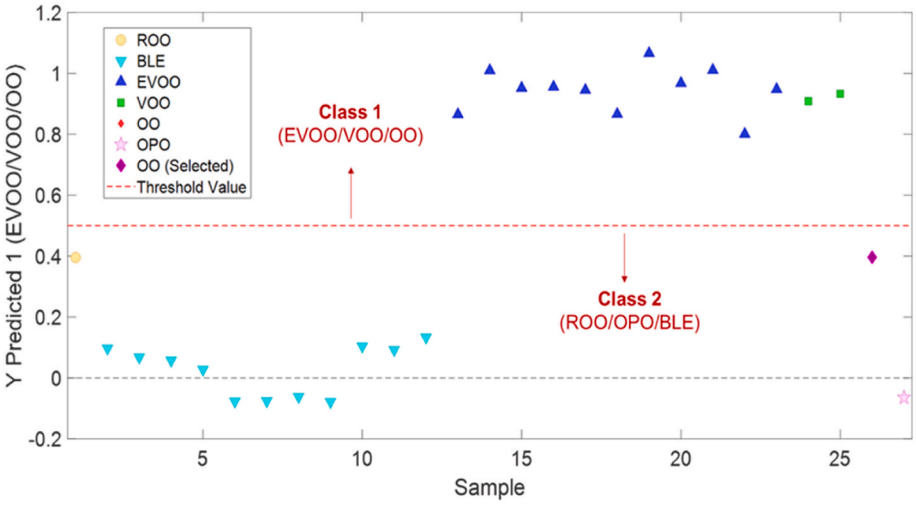

Fig. 7. Classification plot of the validation set samples from the after-agnostizing PLS-DA model. EVOO: extra virgin olive oil; VOO: virgin olive oil, OO: olive oil; ROO: refined olive oils; OPO: olive-pomace oil and BLE: blend of EVOO/VOO with ROO/OPO oils. *(The solid line signifies the threshold decision of 0.5; the circled sample from class 1 is the only misclassified in class 2).*

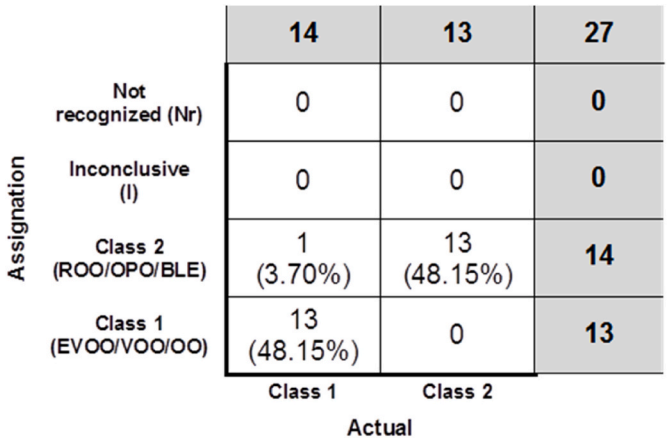

Fig. 8. Validation contingencies from the after-agnostizing PLS-DA classification model. Class 1 (target class): EVOO: extra virgin olive oil, VOO: virgin olive oil, OO: olive oil) – Class 2 (non-target class): ROO: refined olive oil, OPO: olive-pomace oil, and BLE: blend of EVOO/VOO with ROO/OPO oils.

fingerprints (after-agnostizing) (see supplementary material, Figs. S5 and S6). This finding suggests that SVM is a data mining/chemometric classification method suitable to be applied for the authentication of olive oil evidencing the huge difference of EVOO and VOO from ROO of any other kind. In addition it is reaffirmed that the instrument-agnostizing methodology for the standardization of raw chromatograms yields equal or even better results than the conventional icoshift alignment but with the advantage that it results in single multivariate models without the need to repeat the data alignment step each time new samples are analysed.

The contingency results were used to further calculate the classification performance metrics, presented in Table 1. The model SVM obtained the best results for sensitivity (SENS), specificity (SPEC), positive predictive value (PPV) and negative predictive value (NPV). For SIMCA and PLS-DA models, both of them obtained a SENS of 0.93, which indicates the ratio of agreement of the class 1; SIMCA model obtained a SPEC of 0.85 and PLS-DA of 1.00 what indicates that the latter shows a better ratio of agreement of class 2. The PPV in both models was 1, indicating that the models are capable to correctly classify in all the cases the samples belonging to class 1, and that SIMCA model is better classifying the samples of class 2, since it obtained a NPV of 1.00 and PLS-DA of 0.93.

In addition, Bayes' conditional probabilities 1/1 and 2/2, which report good quality of the model and/good probability of classification, are equal or close to one.

Note that, the parameters that indicate bad quality/probability of misclassification for the SVM model, like FPR, FNR, MR and Bayes' conditional probabilities 1/2 and 2/1, are equal to zero. In this regard, the SVM is capable to avoid wrong assignations with an MR value of 0, whilst PLS-DA and SIMCA classification models can perform wrong assignations with 0.04 and 0.11, respectively. Another example can be observed in PROB (1/2) which indicates that the SVM will not classify a sample from class 1 to class 2. On the contrary, PLS-DA and SIMCA models can make that misclassification with 0.07 and 0.08, respectively. Such results reveal that PLS-DA and SVM are better in classifying the samples used within this study, providing good results among data before and after agnosticism.

In this adulteration context, a similar study was performed by Tata et al. (2022) in which EVOO chromatographic polar fingerprints were analysed with PLS-DA and SVM to detect adulterations in EVOO with soft-refined olive oil. In this study, PLS-DA was used mainly as an exploratory technique, since authors declared to observe a good separation between the different kinds of vegetable oils. Afterwards, authors performed a SVM model with values reported on the training set of SENS, SPEC and EFFIC of 0.94, 0.93 and 0.95, respectively. However, such model was further validated with only six correctly classified samples. It is important to note that, the SVM model developed within the present study performs better than the model developed by Tata

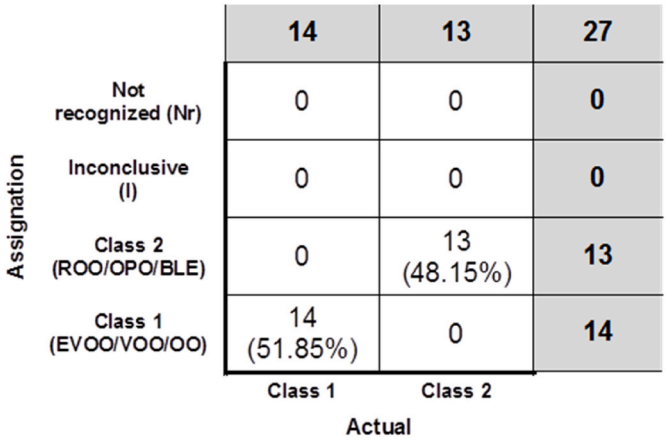

Fig. 10. Validation contingencies of the after-agnostizing SVM classification model. Class 1 (target class): EVOO: extra virgin olive oil, VOO: virgin olive oil, OO: olive oil) – Class 2 (non-target class): ROO: refined olive oil, OPO: olive-pomace oil, and BLE: blend of EVOO/VOO with ROO/OPO oils.

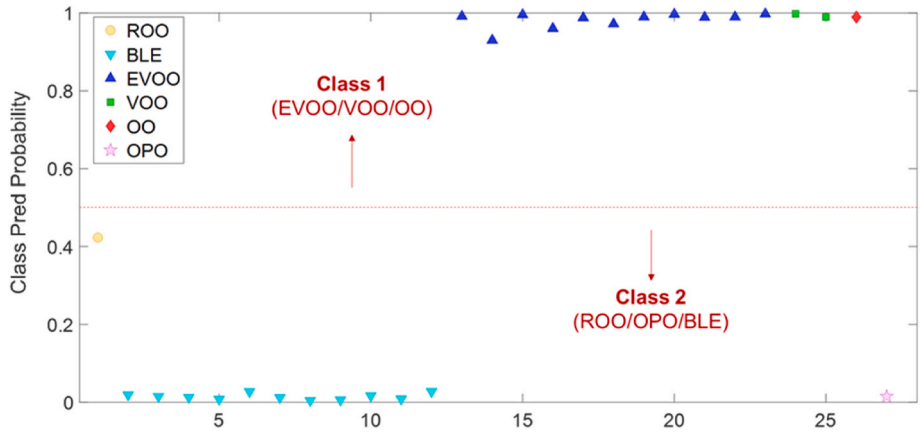

Fig. 9. Classification plot of the validation set samples from the after-agnostizing SVM model. EVOO: extra virgin olive oil; VOO: virgin olive oil, OO: olive oil; ROO: refined olive oils; OPO: olive-pomace oil and BLE: blend of EVOO/VOO with ROO/OPO oils. (*The solid line signifies the threshold decision value of 0.5; all validation samples were rightly classified*).

Table 1
Summary of classification performance metrics of after-agnostizing SIMCA, PLS-DA and SVM models.

Classification performance metrics	SIMCA	PLS-DA	SVM
	Class 1 (EVOO/VOO/OO)		
Inconclusive rate (IR)	0.04	0.00	0.00
Sensitivity (SENS)	0.93	0.93	1.00
Specificity (SPEC)	0.85	1.00	1.00
False positive rate (FPR)	0.15	0.00	0.00
False negative rate (FNR)	0.07	0.07	0.00
Positive predictive value (precision) (PPV)	1.00	1.00	1.00
Negative predictive value (NPV)	1.00	0.93	1.00
Youden index (YOUD)	0.77	0.93	1.00
Positive likelihood rate (LR (+))	6.04	–	–
Negative likelihood rate (LR (−))	0.08	0.07	0.00
Classification odds ratio (COR)	71.50	–	–
F-measure (F)	0.96	0.96	1.00
Discriminant power (DP)	1.02	–	–
Efficiency (or accuracy) (EFFIC)	0.89	0.96	1.00
Misclassification rate (MR)	0.11	0.04	0.00
AUC (correctly classified rate)	0.89	0.96	1.00
Gini coefficient (Gini)	0.77	0.93	1.00
G-mean (GM)	0.89	0.96	1.00
Matthew's correlation coefficient (MCC)	0.89	0.93	1.00
Chance agreement rate (CAR)	0.45	0.50	0.50
Chance error rate (CER)	0.50	0.50	0.50
Kappa coefficient (KAPPA)	0.80	0.93	1.00
PROB (1/1)	0.87	1.00	1.00
PROB (2/2)	0.92	0.93	1.00
PROB (1/2)	0.08	0.07	0.00
PROB (2/1)	0.13	0.00	0.00

The hyphen "–" signifies that the performance feature cannot be determined since it involves a division between zero.

et al. (2022), since it provides improved quality performance metrics considering a larger number of samples. Additionally, it is worth noting that such results were obtained with instrument-agnostic fingerprints of EVOO what leads to an important interlaboratory application as well as to the expansion and implementation of multivariate methods for the control of olive oil authenticity.

4. Conclusions and future perspectives

The main problem with the expansion of multivariate models is that they are very dependent on the alignment of the chromatographic signals and the application of alignment algorithms such as icoshift implies repeating the process of applying this algorithm each time a new chromatographic signal is obtained. This involves that multivariate classification models must be retrained and validated with all chromatograms again. In this regard, due to the application of the instrument-agnostizing methodology this is not essential since the instrument dependence has been minimise and, once the model has been trained using instrument-agnostic fingerprints, it is not necessary to do it again. Therefore this model could be transferred to another laboratory for its application. In fact, the authors are currently carrying out more experiments in collaboration with other laboratories in order to transfer and to implement a unique model.

In this study, the advantage of the instrument-agnostizing methodology over the application of a conventional chromatographic signal alignment procedure applying the icoshift algorithm has been demonstrated. Thus, we can conclude that the methodology for standardizing raw chromatograms have allowed to obtain instrument-agnostic fingerprints of olive oil independent from the chromatographic state or date of chromatographic analysis. This offers an important advance in knowledge as it provides the opportunity to establish the first universal database of olive oil chromatographic fingerprints, generating from these instrument-agnostic fingerprints single multivariate models that could be universally implemented in routine laboratories in order to easily authenticate olive oil.

CRediT authorship contribution statement

Christian H. Pérez-Beltrán: Investigation, Methodology, Software, Validation, Writing – original draft. **Ana M. Jiménez-Carvelo:** Investigation, Methodology, Resources, Writing – review & editing, Supervision, Project administration. **Sandra Martín-Torres:** Methodology, Software, Writing – review & editing. **Fidel Ortega-Gavilán:** Methodology, Software, Writing – review & editing. **Luis Cuadros-Rodríguez:** Resources, Writing – review & editing, Funding acquisition, Supervision, Project administration.

Declaration of interests

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

Acknowledgments

Funding for open access charge: University of Granada / CBUA.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodcont.2022.108957.

References

Abdallah, M., Vegara-Barberán, M., Lerma-García, M. J., Herrero-Martinez, J. M., Simó-Alfonso, & E., & Guerfel, M. (2016). *European Journal of Lipid Science and Technology, 118*, 1236–1242. https://doi.org/10.1002/ejlt.201500041

Boccard, J., & Rudaz, S. (2020). Analysis of metabolomics data – a chemometric perspective. In S. Brown, T. Roma, & B. Walczak (Eds.), *Comprehensive chemometrics – chemical and biochemical data analysis* (pp. 483–505). Amsterdam: Elsevier.

Carranco, N., Farrés-Cebrián, M., Saurina, J., & Núñez, O. (2018). Authentication and quantification of fraud in extra virgin olive oils based on HPLC-UV fingerprinting and multivariate calibration. *Foods, 7*, 44. https://doi.org/10.3390/foods7040044

Commission Regulation (EEC) No 2568/91. Characteristics of olive oil and olive-residue oil and the relevant methods of analysis. Official Journal of European Union, 031.005, 1-128. https://www.fao.org/faolex/results/details/en/c/LEX-FAOC040621/.

Creydt, M., & Fischer, M. (2020). Food authentication in real life: How to link nontargeted approaches with routine analytics? *Electrophoresis, 41*, 1665–1679. https://doi.org/10.1002/elps.202000030

Cuadros Rodríguez, L., Martín Torres, S., Ortega Gavilán, F., Jiménez Carvelo, A. M., López Ruíz, R., Garrido Frenich, A., Bagur González, M. G., & González Casado, A. (2021b). Standardization of chromatographic signals – Part II: Expanding instrument-agnostic fingerprints to reverse phase liquid chromatography. *Journal of Chromatography A, 1641*, 461973. https://doi.org/10.1016/j.chroma.2021.461973

Cuadros Rodríguez, L., Ortega Gavilán, F., Martín Torres, S., Medina Rodríguez, S., Jiménez Carvelo, A. M., González Casado, A., & Bagur González, M. G. (2021a). Standardization of chromatographic signals – Part I: Towards obtaining instrument-agnostic fingerprints in gas chromatography. *Journal of Chromatography A, 1641*, 461983. https://doi.org/10.1016/j.chroma.2021.461983

Cuadros Rodríguez, L., Pérez Castaño, E., & Ruiz Samblás, C. (2016b). Quality performance metrics in multivariate classification methods for qualitative analysis. *Trends in Analytical Chemistry, 80*, 612–624. https://doi.org/10.1016/j.trac.2016.04.021

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

Drira, M., Kelebek, H., Guclu, G., Jabeur, H., Selli, S., & Bouaziz, M. (2020). Targeted analysis for detection the adulteration in extra virgin olive oil's using LC-DAD/ESI-MS/MS and combined with chemometrics tools. *European Food Research and Technology, 246*, 1661–1677. https://doi.org/10.1007/s00217-020-03522-y

Duraipandian, S., Petersen, J. C., & Lassen, M. (2019). Authenticity and concentration analysis of extra virgin olive oil using spontaneous Raman spectroscopy and multivariate data analysis. *Applied Sciences, 9*, 2433. https://doi.org/10.3390/app9122433

Folch Fortuny, A., Vitale, R., de Noord, O. E., & Ferrer, A. (2017). Calibration transfer between NIR spectrometers: New proposals and a comparative study. *Journal of Chemometrics, 31*, Article e2874. https://doi.org/10.1002/cem.2874

Fornasaro, S., et al. (2020). Surface enhanced Raman spectroscopy for quantitative analysis: Results of a large-scale European multi-instrument interlaboratory study. *Analytical Chemistry, 92*, 4053–4064. https://doi.org/10.1021/acs.analchem.9b05658

Gou, S., Heinke, R., Stöckel, S., Rösch, P., Jürgen, P., & Bocklitz, T. (2018). Model transfer for Raman-spectroscopy - based bacterial classification. *Journal of Raman Spectroscopy, 49*, 627–637. https://doi.org/10.1002/jrs.5343

Jabeur, H., Drira, M., Rebai, A., & Bouaziz, M. (2017). Putative markers of adulteration of higher-grade olive oil with expensive pomace olive oil identified by gas chromatography combined with chemometrics. *Journal of Agricultural and Food Chemistry, 65*, 5375–5383. https://doi.org/10.1021/acs.jafc.7b00687

Jiménez Carvelo, A. M., & Cuadros Rodríguez, L. (2021). Data mining/machine learning methods in foodomics. *Current Opinion in Food Science, 37*, 76–82. https://doi.org/10.1016/j.cofs.2020.09.008

Jiménez Carvelo, A. M., Martín Torres, S., Cuadros Rodríguez, L., & González Casado, A. (2020). Nontargeted fingerprinting approaches. In C. M. Galanakis (Ed.), *Food traceability and authentication* (pp. 163–193). Oxford: Academic Press/Elsevier. https://doi.org/10.1016/B978-0-12-821104-5.00010-6.

Karunathilaka, S. R., Fardin Kia, A. R., Srigley, C., Chung, J. K., & Mossoba, M. M. (2016). Nontargeted, rapid screening of extra virgin olive oil products for authenticity using near-infrared spectroscopy in combination with conformity index and multivariate statistical analyses. *Journal of Food Science, 81*, C2390–C2397. https://doi.org/10.1111/1750-3841.13432

Li, Y., Wen, S., Sun, Y., Zhang, N., Gao, Y., & Yu, X. (2021). New method based on polarity reversal for detecting adulteration of extra virgin olive oil with refined olive pomace oil. *European Journal of Lipid Science and Technology*. https://doi.org/10.1002/ejlt.202100193, 2100193.

Meenu, M., Cai, Q., & Xu, B. (2019). A critical review on analytical techniques to detect adulteration of extra virgin olive oil. *Trends in Food Science & Technology, 91*, 391–408. https://doi.org/10.1016/j.tifs.2019.07.045

Mingchih, F., Chia-Fen, T., Guan-Yan, W., Su-Hsiang, T., Hwei-Fang, C., Ching-Hao, K., Che-Lun, H., Ya Min, K., Yang Chih, D., & Yu Mei, C. (2015). Identification and quantification of Cu-clorophyll adulteration of edible oils. *Food Additives and Contaminants: Part B, 8*, 157–162. https://doi.org/10.1080/19393210.2015.1025861

Muñoz Olivas, R. (2004). Screening analysis: An overview of methods applied to environmental, clinical and food analyses. *Trends in Analytical Chemistry, 23*, 203–216. https://doi.org/10.1016/S0165-9936(04)00318-8

Navratilova, K., Hurkova, K., Hrbek, V., Uttl, L., Tomaniova, M., Valli, E., & Hajslova, J. (2022). Metabolic fingerprinting strategy: Investigation of markers for the detection of extra virgin olive oil adulteration with soft-deodorized olive oils. *Food Control, 134*, 108649. https://doi.org/10.1016/j.foodcont.2021.108649

Oliveri, P., Malegori, C., Mustorgi, E., & Casale, M. (2020). Application of chemometrics in the food sciences. In S. Brown, T. Roma, & B. Walczak (Eds.), *Comprehensive chemometrics – chemical and biochemical data analysis* (pp. 99–111). Amsterdam: Elsevier.

Rigano, F., Arigò, A., Oteri, M., La-Tella, R., Dugo, P., & Mondello, L. (2021). The retention index approach in liquid chromatography: An historical review and recent advances. *Journal of Chromatography A, 1640*, Article 461963. https://doi.org/10.1016/j.chroma.2021.461963

Tahir, H. E., Arslan, M., Mahunu, G. K., Mariod, A. A., Xiaobo, S. B. H. Z., Jiyong, S., El-Seedi, G. R., & Musa, T. H. (2022). The use of analytical techniques coupled with chemometrics for tracing the geographical origin of oils: A systematic review (2013–2020). *Food Chemistry, 366*, 130633. https://doi.org/10.1016/j.foodchem.2021.130633

Tata, A., Massaro, A., Damiani, T., Piro, R., Dall'Asta, C., & Suman, M. (2022). Detection of soft-refined oils in extra virgin olive oil using data fusion approaches for LC-MS, GC-IMS and FGC-Enose techniques: The winning synergy of GC-IMS and FGC-Enose. *Food Control, 133*, 108645. https://doi.org/10.1016/j.foodcont.2021.108645

Tomasi, G., Savorani, F., & Engelsen, S. B. (2011). Icoshift: An effective tool for the alignment of chromatographic data. *Journal of Chromatography A, 1248*, 7832–7840. https://doi.org/10.1016/j.chroma.2011.08.086

Tomasi, G., Van der Berg, F., & Andersson, C. (2004). Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *Journal of Chemometrics, 18*, 231–241. https://doi.org/10.1002/cem.859

Wang, Y., Veltkamp, D. J., & Kowalski, B. R. (1991). Multivariate instrument standardization. *Analytical Chemistry, 63*, 2750–2756. https://doi.org/10.1021/ac00023a016

Wise, B. M., Gallagher, N. B., Bro, R., Shaver, J. M., Winding, W., & Koch, R. S. (2006). *Chemometrics tutorial for PLS_toolbox and solo*. Wenatchee, WA, USA: Eigenvector Research, Inc.

Zhang, J., Guo, C., Cui, X., Cai, W., & Shao, X. (2019). A two-level strategy for standardization of near infrared spectra by multi-level simultaneous component analysis. *Analytica Chimica Acta, 1050*, 25–31. https://doi.org/10.1016/j.aca.2018.11.013

Zhang, J., Sun, H., & Lu, W. (2021). Recent advances in analytical detection of olive oil adulteration. *Food Science and Technology*, 1–10. https://doi.org/10.1021/acsfoodscitech.1c00254