Preprints.org

Discrimination/Classification of Edible Vegetable Oils From Raman Spatially Solved Fingerprints Obtained on a Portable Instrumentation

Guillermo Jiménez-Hernández , Fidel Ortega-Gavilán [*] , M.Gracia Bagur-González [*] ,
Antonio González-Casado

Article

Discrimination/Classification of Edible Vegetable Oils from Raman Spatially Solved Fingerprints Obtained on a Portable Instrumentation

Guillermo Jiménez Hernández [1], Fidel Ortega Gavilán [1,2,*], M. Gracia Bagur González [1,*] and Antonio González Casado [1]

[1] Department of Analytical Chemistry, Faculty of Science, University of Granada, C/ Fuentenueva w/n, E-18071, Granada (Spain)

[2] Department of Chemical Analysis of Residues, Animal Health Central Laboratory (LCSA), Ministry of Agriculture, Fisheries and Food, Camino del Jau w/n, E-18320, Santa Fe (Granada)

* Correspondence: bec_labgranada2@mapa.es or fog@ugr.es (F.O.G.); mgbagur@ugr.es (M.G.B.G.); Tel.: +34 958243327 (M.G.B.G.)

Abstract: Nowadays, the combination of fingerprinting methodology with friendly-environmental and economical analytical instrumentation are becoming increasingly relevant in the food sector. In this study, a highly versatile portable analyser based on Spatially Offset Raman Spectroscopy (SORS) to obtain the edible vegetable oils (sunflower and olive oils) fingerprints was used to evaluate the capability of such fingerprints, obtained quickly, reliable and without any sample treatment, to discriminate/classify the analysed samples. After data treatment, not only HCA and PCA as unsupervised pattern recognition techniques but also SVM, kNN and SIMCA as supervised pattern recognition techniques, showed that the main effect over the discrimination/classification was associated to those regions of RAMAN fingerprint related to the free fatty acids content, especially oleic and linoleic acid. These facts allowed the discrimination attending to the original raw material used in the oil's elaboration. In all the model established, reliable qualimetric parameters were obtained.

Keywords: RAMAN fingerprints; edible vegetable oils; sunflower oil; olive oil; pattern recognition techniques; portable analyser; SORS

1. Introduction

Sunflower oil is produced in several countries, including Russia, Ukraine and China from the seeds of *Helianthus Annuus L.*, [1]. In the European Union, represents the second most important source of seed oils after rapeseed oil, and the area under sunflower crops will increase up to 1.0 million ha by the end of 2031 [2]. The plant grows best in dry, temperate climates (among 20-25ºC) with high solar radiation, low humidity and deep soils to spread its roots in search of nutrients and water. Their deshelled seeds are responsible for 80% of the total fruit weight and has a high oil content up to 55% w/w [3,4]. Thus, these oils have a high lipids content, where most of them are triglycerides composed mainly of long-chain unsaturated fatty acids with different unsaturation, being linoleic acid (59%, polyunsaturated omega 6) the most important [5,6]. In addition, it should be noted the presence of other fatty acids as oleic acid (30%, monounsaturated omega 9), stearic acid (6%, saturated) or palmitic acid (5%, saturated).

Like Olive Oil (OO), there are several commercial categories, the main ones being: i) Refined Sunflower Oil (SFO), characterized by a high linoleic acid (LA) content. ii) High Oleic Sunflower Oil (HOSFO), with an oleic acid (OA) content of at least 75%; measured as a percentage of the total fatty acid content, and iii) Medium Oleic Sunflower Oil (MOSFO), with a seed OA content ranged among 50% and 75%. The last two commercial categories come from seeds genetically modified to increase

naturally not only the ratio of oleic to linoleic acid but also other monounsaturated fatty acids or vitamins such as vitamin E. These facts confer to these oils, in especial to HOSFO, an overall composition with remarkable similarities to OO, and as an extra factor, greater resistance to oxidation and possibilities of use [7,8].

Currently, the main analytical methods used to check the authenticity and evaluate possible adulteration of edible vegetable oils, especially OO, are chromatography-based analysing the presence of triglycerides and fatty acids [9,10]. The most used are Thin-Layer Chromatography (TLC), High-Pressure Liquid Chromatography (HPLC) and Gas Chromatography (GC). The former, TLC is used both, qualitative and quantitatively in official methods such as the analysis of sterols and stanols fractionation in oils [11,12]. HPLC as, among other uses, an International Olive Council reference method commonly used in routine laboratories [13,14]. The later, GC uses not only as GC-FID (Flame Ionization Detector) official method specified in IOCs guides, but also uses for characterising the triglyceride profile in vegetable oils, or the separation and quantification of fatty acids according to their esterified fraction (FAME), by means of a prior hydrolysis and methylation of fatty acids step [15–17].

As an alternative to chromatographic techniques, the spectroscopic ones (Raman, MIR, FIR, NMR, etc.) provide simple, fast and reliable results, with the advantage of using directly on the sample without the need for any sample pre-treatment stage. Moreover, they appear to be not only economical but also environmentally sustainable analytical methods [17–19]. NMR provides a quick way of measuring the oil content of oil seeds such as sunflower seeds, as well as assessing their resistance to high temperatures by evaluating the degradation of their constituents, as their main domestic use *i.e.* food frying [20]. Some of these techniques, such as Raman or NIR spectroscopy, can be easily adapted to portable devices, allowing the acquisition of analytical data "in situ"/in real time, although in many cases these techniques have proven to be incapable of measuring through packaging. To overcome this inconvenient, some instrumental improvements has been made to Raman spectroscopy, as in the case of Spatially offset Raman spectroscopy (SORS) technique [21]. Despite of these advantages, they have a lower resolution capability than chromatographic techniques, as they do not provide information on each individual compound present, but rather on the bonds that occurs in the components of a sample as a whole.

Considering that the final output of an analytical device (bench top, handheld or portable devices) is an analytical signal, the use of this to obtain information about the properties of a material related to its chemical composition e.g. fingerprinting methodology, has shown to be a powerful tool to identify, discriminate and authenticate edible oils among other commodities. In this sense, these signals may relate to the identity of a food, to its physico-chemical or other natural properties, as well as to the presence or quantity of compounds in the chemical composition of the food [22,23]. Despite of this, instrumental fingerprints contain hidden information about it, which normally requires the use of chemometric tools in particular or machine learning methods in general [24–26].

In the food sector, the fingerprinting methodology has been widely used to solve different problems such as the authentication and discrimination of olive oils, margarines, and fat spread or the evaluation of the quality and production process of spirits among others foodstuff [27–31] Regarding the use of handheld or portable devices based on SORS applications few studies have been carried out despite the great potential of the technique [Error! Bookmark not defined.]. Recent examples include the detection of possible adulteration in alcoholic beverages [32], the study of the evolution of the alcoholic fermentation process in white wine [33], the authenticity assessment of the animal milk used origin of the in the production of cheese, the characterisation of several commercial categories of cheese (Cheddar, Manchego and Pecorino Romano), the analysis of packaged margarines and fat spreads [34–36] or the study of adulteration of extra virgin olive with other edible vegetable oils as in the case of Varnasseri et al. [37].

Thus, the aim of this work is to propose a more environmentally friendly and sustainable analytical method that combine the Raman fingerprints obtained from a portable instrument with chemometrics/machine learning to build classification models to reliably differentiate sunflower oils from olive oils of different commercial categories.

2. Materials and methods

2.1. Oil samples

A sample bank consisting of 145 samples from different types of edible vegetable oils, purchased at local supermarkets, was used for this study, separating these oils into two main groups: sunflower and olive oil. The first group was constituted by 48 samples where 7 of them were labelled as high oleic acid content. The second group was composed by a total of 97 samples from which: 65 were extra virgin olive oil (EVOO), 22 were virgin olive oil (VOO) and 10 were pomace olive oil (POO).

2.2. Spectroscopic analysis

A Vaya Raman portable spectrometer (Agilent Technologies, Santa Clara, CA, USA) was used in this study. It was equipped with a 3B class diode laser operating at 830 nm and a maximum power of 450 mW. The exposure time of the sample to the laser varied between 0.5 and 2 s. The spectral resolution of the device was 12-20 cm^{-1} with the wavenumber range among 735-1540 cm^{-1} using a cooled CCD (charge coupled devices) as detection system. The specific offset length from the incidence point was fixed by equipment at 0.6 cm. For measurement (with a total time of the measurement ranged between 30 s and 2 min), edible oil samples were introduced into 4mL silicate boron of 1mm thickness.

After carrying out the measurement, the equipment software performs: i) a correction from the information obtained to eliminate any possible influence of the container, ii) a baseline adjustment and finally ii) a normalization of the intensity values. Therefore, the result of each analysed sample is a Raman spectrum from 350 to 2000 cm^{-1} in which the intensity values are between 0 and 1. Thus, the final spectrum was a normalized spatial resolved RAMAN spectrum (NSR RAMAN Spectrum).

2.3. Data treatment

Each NSR Raman spectrum was collected in .CSV format (comma separated value) and exported to .mat format for the subsequent elaboration of the fingerprints matrix within the MATLAB environment (version 9.3, Mathworks Inc., Natick, MA, USA). As a result, a comprehensive 145×1651 data matrix consisting of 145 NSR RAMAN fingerprints (rows), each of which is composed of 1651 normalized intensity values (columns), was obtained as raw data matrix and once pre-processed, used for the different chemometric studies.

The pre-processing stage included: a selection of interval of interest for each fingerprint, a filtering smoothing of the signals using a Savitzky-Golay filter (1st derivate, 2nd order polynomial and a filter width of 21 points) and a mean center, obtaining a reduced-pre-processed NSR RAMAN fingerprints matrix of 145 samples × 902 variables. Figure 1 show the NSR RAMAN fingerprint edible vegetable olive oils analysed for this study before and after pre-processing step. Unsupervised and supervised pattern recognition techniques were explored using PLS_Toolbox, version 8.6.1 (Eigenvector Research Inc., Manson, WA, USA) [38].
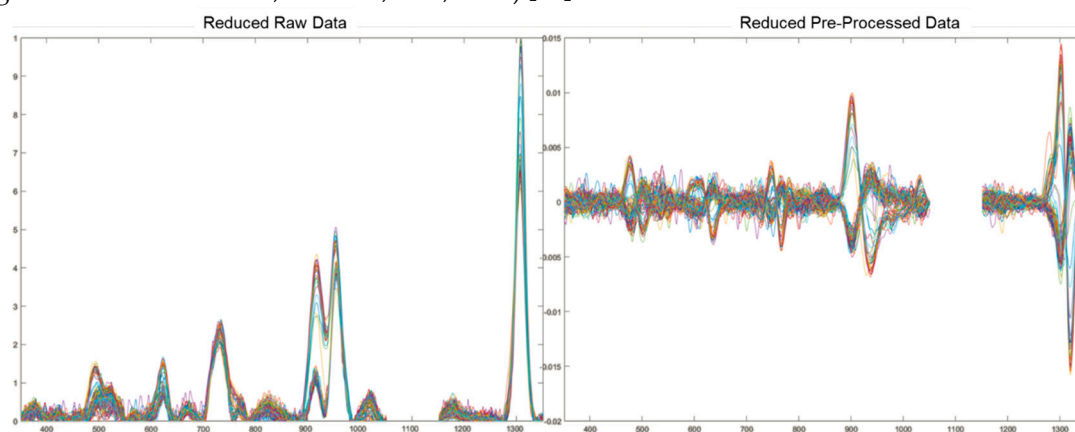

Figure 1. Reduced RAMAN fingerprint of vegetable edible oils before and after data pre-processing.

3. Results and discussion

3.1. Unsupervised Pattern Recognition Methods

The analysis of the natural grouping trends could allow establishing a correlation between the data in the NSR RAMAN fingerprints and their impact on vegetable origin of the edible oils analysed. This would lead to evaluate the use of a green analytical technology as a 'Vaya Raman' portable spectrometer to distinguish among vegetable oils obtained from different types of raw materials such as olive and sunflower oils.

3.1.1. Hierarchical Cluster Analysis of Raman Edible Vegetable Oils Fingerprints

First of all, an HCA was performed using the data matrix that had been previously defined. In this analysis, Ward's method and Manhattan distance were used as the linkage criterion and the measure of distance between pairs of observations respectively. To select the number of clusters, a $D_{linkage} = 2/3$ of the D_{max} was used as internal criterion. It could be observed that the edible oils clustered naturally not only according to the raw material employed, but also by the oleic acid contents. Thus, two main clusters can be observed: i) group I constituted by those oils obtained from sunflower seeds and ii) group II in which all the olive oils (extra virgin, virgin and pomace) and high oleic sunflower oils are clustered (Figure 2).

Regarding group I, it also could be observed that a previous partial clustering takes place according to the percentage of oleic acid shown on the label of commercial container, where the natural nesting distinguishes those sunflower oils with oleic acid locating them in the upper part of the group. Although a similar behaviour can be observed when group II is analysed, in this case the nesting order is the opposite to the previous one, due to the upper partial cluster groups those olive oil samples with low acidity which implies a lower content in free oleic acid. Meanwhile, those samples with higher content in this acid appear grouped in the second partial cluster (some EVOO, refined and pomace samples). Thus, the appearance of those sunflower oil samples (commercially labelled as high oleic acid content sunflower oils) in this last sub-group could be justified.
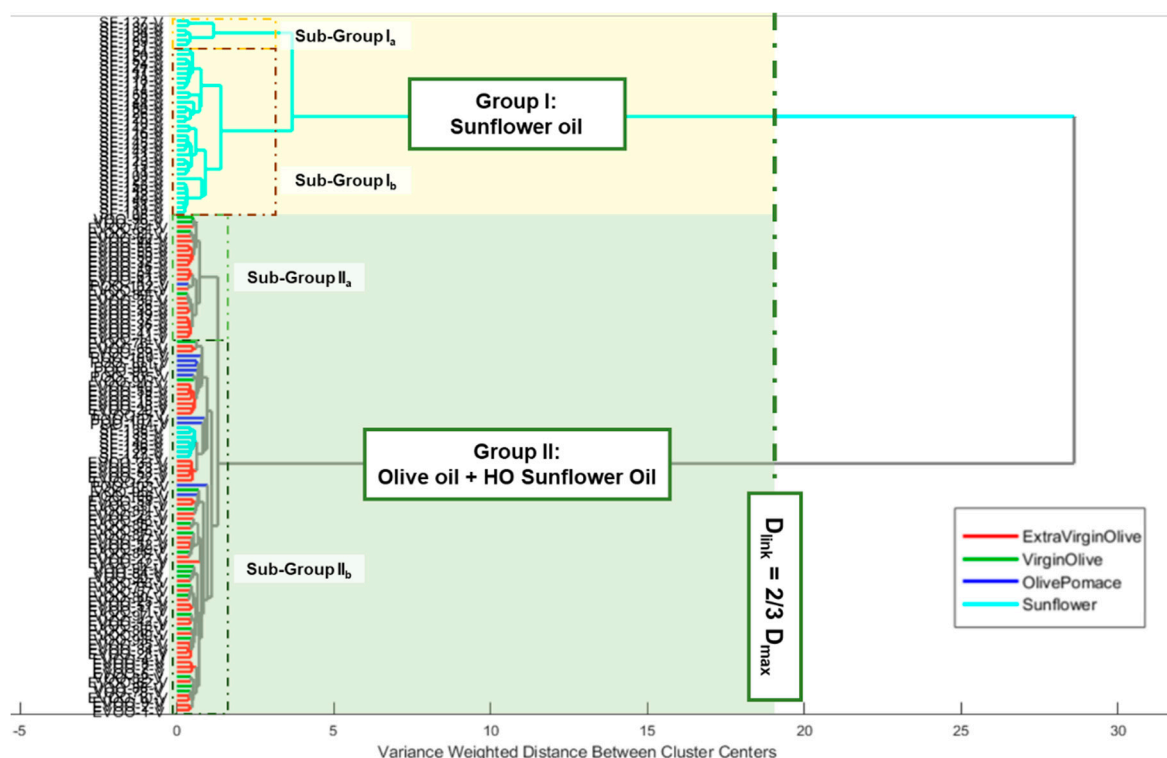

Figure 2. Dendrogram from the HCA of the analysed edible vegetable oils.

5

3.1.2. Principal Component Analysis

When PCA was applied to the data matrix, 5 principal components (PCs) were obtained which explained 92.42% of the variance of the model for the vegetable edible oils analysed. PC1 explained 88.99% of the total variance of the system, while the other four principal components explained the remaining 3.43% of the variance (PC2: 1.06%; PC3 0.94%; PC4: 0.73% and PC5: 0.70%).

Figure 3a illustrates the scores received by the analysed oil samples in the space of the first two components (PC2 vs. PC1). In this figure, it can be observed that, similarly to what occurs when HCA is applied, the vegetable edible oils are once again grouped according to the raw material used in the elaboration process. It can also be seen that most of the sunflower oils received negative scores for PC1 (Group I), while the olive oils scored positively for this component (Group II). Besides, those high oleic sunflower oils included in this group received negative scores for PC2.

Figure 3b shows the graphical representation of the scores received by the analysed oil samples in the PC5 vs PC1 space. Considering the distribution of the edible oils' score along the PC1 space, once again, it can be observed the same main grouping that attends to the raw material from which the oils are obtained (Group I and II). In the case of Group II, a new tendency is observed, which can be explained attending to the oleic acid content. Thus, the oils with an oleic acid high content, received positive scores on the PC5, while the remaining were in the negative area of this component. Furthermore, in this same group, it could be also observed that as the high oleic acid content of the oil samples, the higher PC1 positive scores are given to the samples. The new described trends could be explained attending to those fingerprint regions that have minimum variations on the total variance, i.e. minimum differences among fingerprints.

When the 3D space defined as PC5 vs PC2 vs PC1 (Figure 3c), a clearer grouping can be observed. Thus, sunflower oils received increasingly positive PC1 scores, as oleic acid content was higher. In addition, the olive pomace oils received positive values from PC1, PC2 and PC5, which it could allow to distinguish among sunflower, virgin/extra virgin olive oils and olive pomace oils.
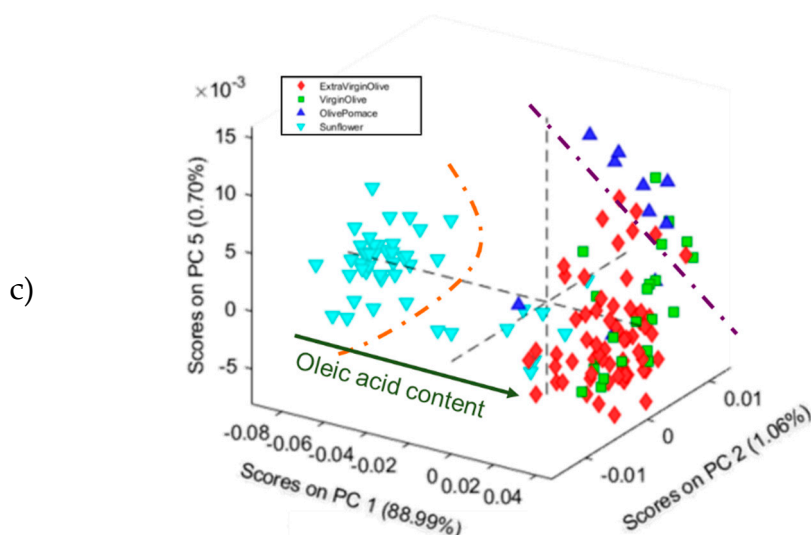
c)
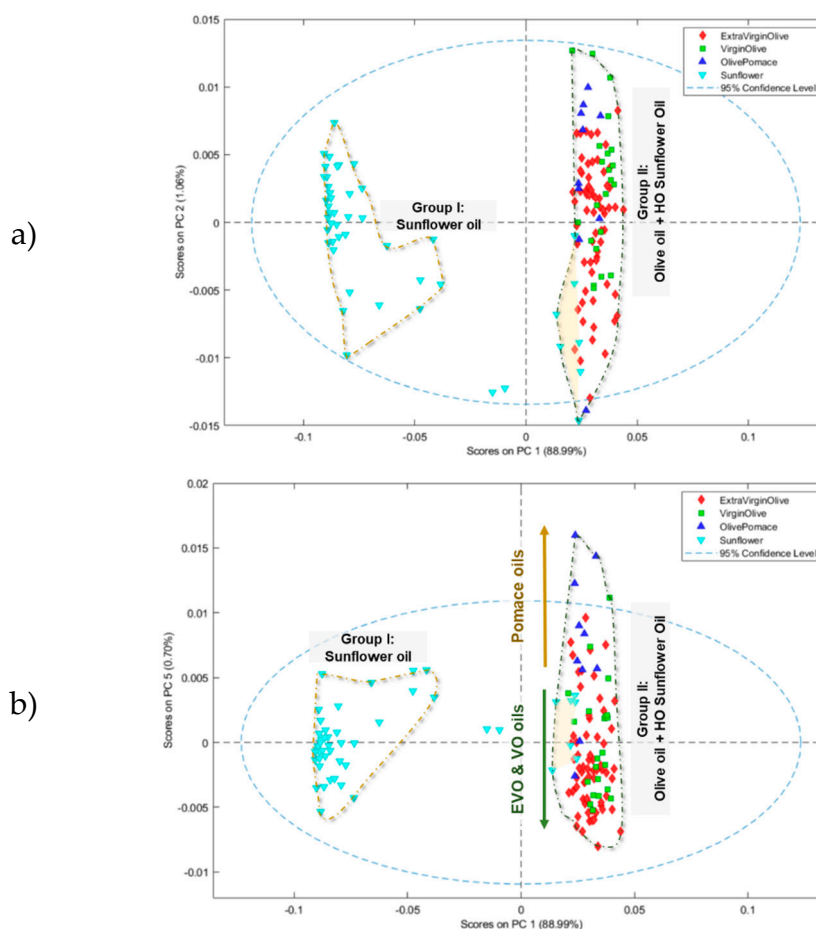
Figure 3. Score plots of the analysed edible vegetable oils in: a) the plane PC2 *vs* PC1, b) the plane PC5 *vs* PC1 and c) in the 3D space defined by PC1 *vs* PC2 *vs* PC5.

If the loading plot of each component (Figure 4a-c) are analysed, based on Figure 4a, it can be observed that the PC1 loadings would allow to identify those variables that explain the natural grouping according to the raw material. The variables (wave numbers) associated to the bands in "conventional" Raman spectra associated to free fatty acid (mainly oleic and linoleic acids) [39]: i) v(C=C) vibration of *cis*-alkene ($\approx$1650 cm^{-1}); ii) δ(C–H) twisting of CH$_2$ ($\approx$1300 cm^{-1}); iii) δ(=C–H) scissoring ($\approx$1250 cm^{-1}); iv) v(C–C) and δ(C=C) twisting of *trans*-alkene ($\approx$1000 cm^{-1}) and v) v(C–H) symmetric and/or asymmetric stretching ($\approx$800 cm^{-1}). Regarding Figure 4b, it can be observed that for the PC2 loadings the wavenumbers in which appear the free fatty acid information are only: i) v(C=C) vibration of *cis*-alkene ($\approx$1650 cm^{-1}) and ii) v(C–H) symmetric and/or asymmetric stretching ($\approx$800 cm^{-1}), being the first one the most influential in the natural grouping observed. When the PC5 loadings are considered (Figure 4c), the natural grouping could be once again associated to the same wavenumbers in PC2. The main difference between PC2 and PC5 loadings arising from the v(C–H) symmetric and/or asymmetric stretching ($\approx$800 cm^{-1}). Thus, slightly variations on these Raman modes provoke the separation among the edible olive oils analysed.
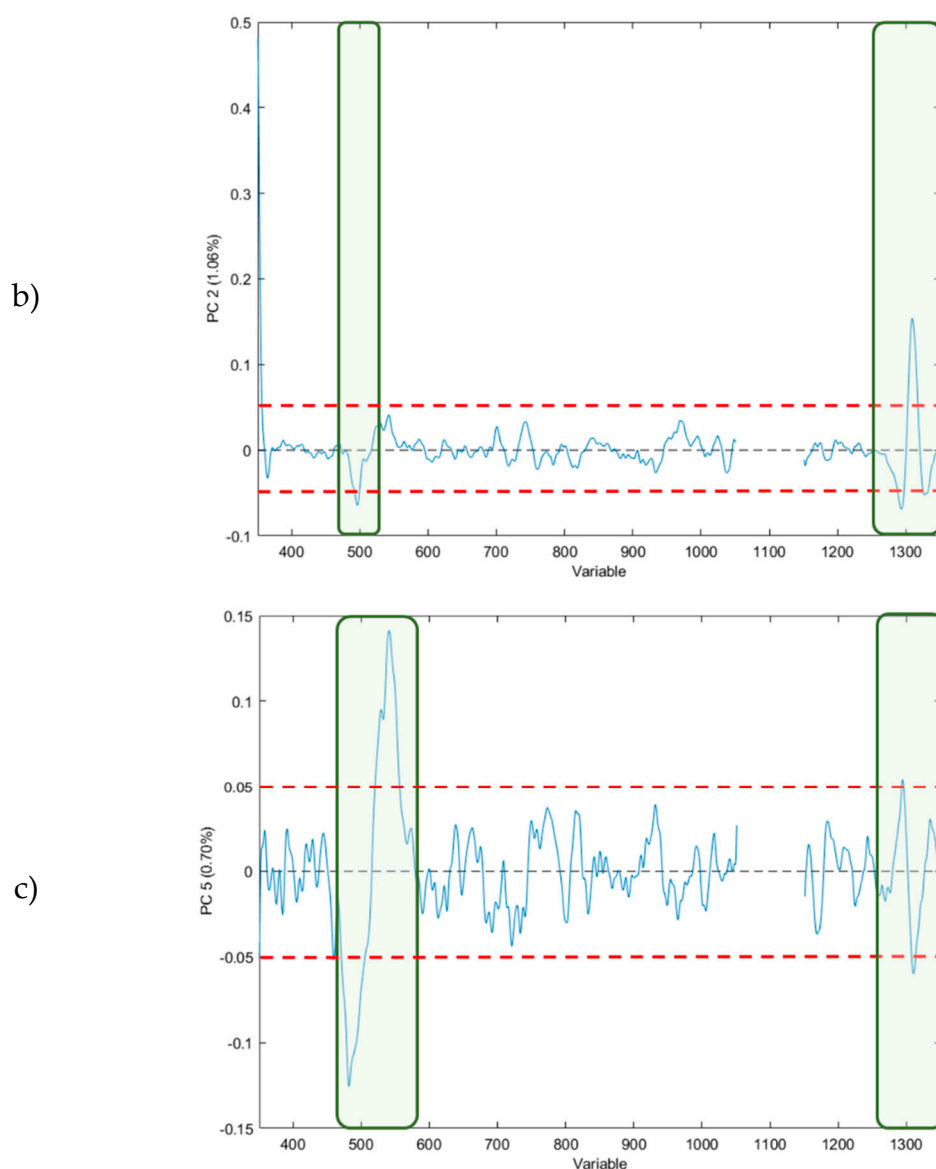

a)

b)

c)
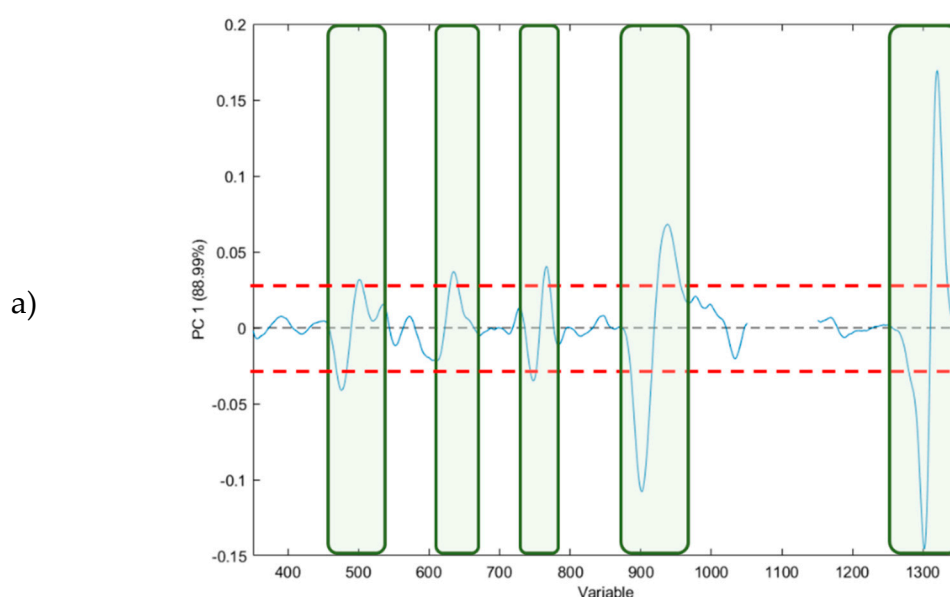
Figure 4. Loading plots from the analysed edible vegetable oils in: a) PC1, b) PC2 and c) PC5.

3.2. Supervised Pattern Recognition Methods

In order to study the capability of the NSR RAMAN fingerprint to discriminate/classify among the edible vegetable oils analysed, several one-input class classification/discrimination model, as SVM, kNN and SIMCA, were developed. For these purposes, the target class was assigned to sunflower oils considering Not-SF (olive oils) as non-target class. The two necessary sets (calibration and prediction) were selected using the Kennard-Stone algorithm **[Error! Bookmark not defined.–40]**, keeping the 66.6% of original samples for calibration/cross-validation purposes and the remaining 33.4% for external prediction purposes.

The SVM model was obtained without reduction of the data dimensionality (uncompressed) during model development. The Kernel algorithm was applied as a radial basis function (RBF), using the PLS_Toolbox™ default values for the gamma and cost parameters.

It can be seen in Figure 5 how only one sample labelled as sunflower oil was classified as Not-SF. In addition, other samples labelled as Sunflower appear closer to the threshold which should be considered as bad classified spite of appearing in the sunflower prediction probability area. Attending the information included in the commercial label of analysed sample, all of them appear described as high oleic sunflower oils with different percentages content of this acid which explain the behaviour observed in this figure.
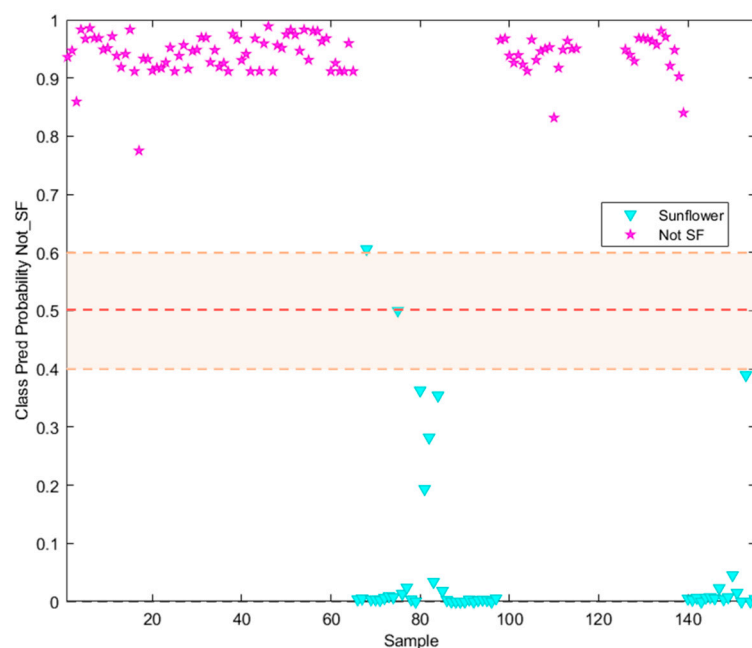
Figure 5. Probability prediction plot obtained from SVM classification model.

When the discrimination/classification power of the NSR RAMAN fingerprint was evaluated through neighbour distances, a kNN was used being k = 7 the best value to decide the neighbour distance in the model. The sunflower class was defined by a class predicted probability value equal to 1, while the non-target class (olive oil) was defined by a probability of 0. Considering the two hard modelling techniques used, a similar behaviour can be observed being the main difference the probability prediction value belonging to each class. All those sunflower samples labelled without high oleic acid content were well classified (probability = 1), with exception of those samples labelled as containing oleic acid (7 samples) which had an assigned probability lower than 1 (Figure 6). That is why these samples were classified by the model as Not-SF. When the non-targeted class was analysed, only 10 samples had an assigned probability value higher than 0. These samples correspond mostly to pomace olive oil which NSR RAMAN fingerprint should be different from the remaining olive oil fingerprints.
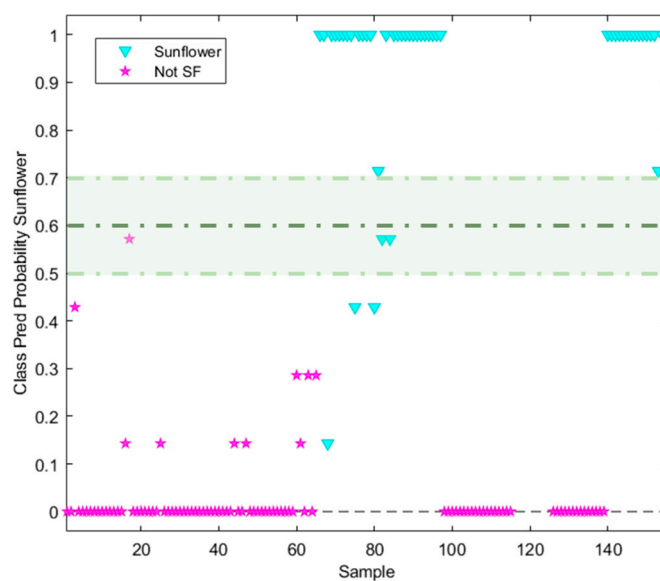

Figure 6. Probability prediction plot obtained from kNN (k=7) classification model.

Finally, a SIMCA model as soft modelling technique was performed. Thus, 3PCs were selected for target and non-target classes which explain the 93.4% and 30.5 % of the cumulative variance respectively.

Figure 7 shows the Cooman's classification plot for the analysed samples. It can be seen that some samples appeared as not-conclusive (SFO samples labelled as containing a high oleic acid content), as outliers (tree SFO samples and most of the POO samples) and miss classified (a HOSFO sample assigned to the non-target class). Comparing with non-supervised patter recognition techniques, HCA and PCA, these samples appear grouping in the olive oil samples cluster or region.
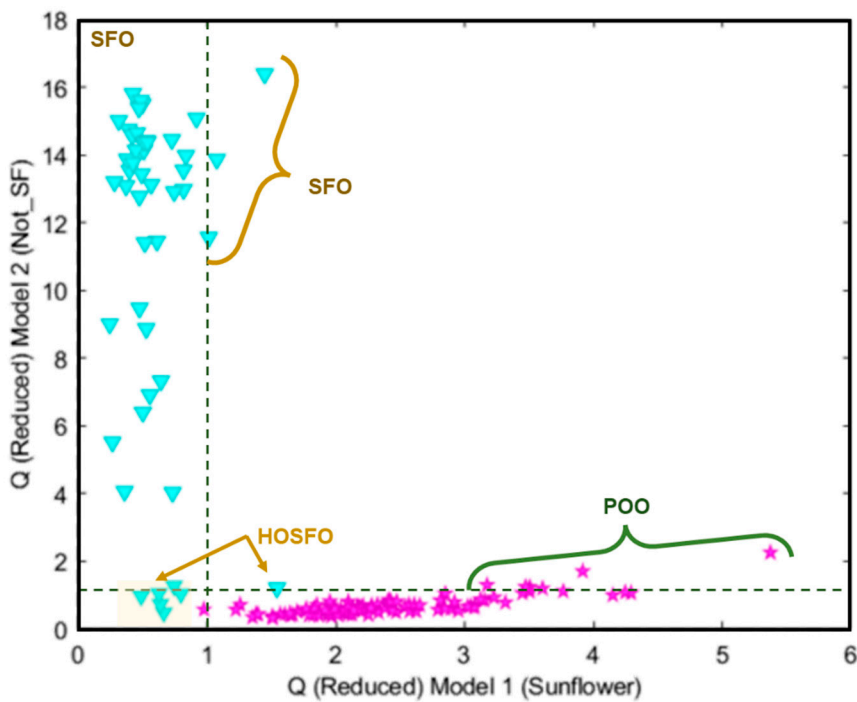

Figure 7. Cooman's classification plots obtained for SIMCA model.

The quality parameters of these three models are shown in Tables 1–3. In general terms, it was observed that the best result was for the SVM model followed by kNN and finally SIMCA. Thus, it can be considered that the hard-models discriminate correctly the samples attending to the raw material. Nevertheless, the SIMCA soft-model shows a better capacity to classify those sunflower oil samples with high oleic acid contend considering that appear not only as sunflower but also as not-sunflower i.e. olive oil.

Table 1. Summary of discrimination/classification performance metrics obtained for the SVM-DA one-input class model.

TARGET Class (TC): Sunflower oil

Features:

*X Block: [Reduced RAMAN instrumental fingerprints]

*Y Block: [TC (Sunflower oil, SFO); NTC (Non Sunflower oil, Not SFO)]

Pre-processing: 1st Derivative (order 2, window;21 pt, tails:polyinterp) + Mean center

Training Set: [97 × 902]

Prediction Set: [48 × 902] See Confusion Table below

Classification performance metrics	TC (SFO)	NTC (Not SFO)
Sensitivity (SENS -prediction stage)	0.98	1.00
Specificity (SPEC-prediction stage)	1.00	0.98
False positive rate (FPR)	0.00	0.02
False negative rate (FNR)	0.02	0.00
Positive predictive value (precision) (PPV)	1.00	0.99
Negative predictive value (NPV)	0.99	1.00
Youden index (YOUD)	0.98	0.98
F-measure (F)	0.99	0.99
Discriminant power (DP)	−	−
Efficiency (or accuracy) (EFFIC)	0.99	0.99
Misclassification rate (MR)	0.01	0.01
AUC (correctly classified rate)	0.99	0.99
Gini coefficient (Gini)	0.98	0.98
G-mean (GM)	0.99	0.99
Matthews correlation coefficient (MCC)	0.98	0.98
Chance agreement rate (CAR)	0.56	0.56
Chance error rate (CER)	0.44	0.44
Kappa coefficient (KAPPA)	0.98	0.98

Confusion Table:

Assigned class		48	97	145
	Not SF	1	97	98
	Sunflower	47	0	47
		Sunflower	Not SF	T
		Reference class		

Note. The hyphen "−" refers to metrics that cannot be determined since a division by zero is involved.

Table 2. Summary of discrimination/classification performance metrics obtained for the kNN-DA one-input class model.

TARGET Class (TC): Sunflower oil

Features:

*X Block: [Reduced RAMAN instrumental fingerprints]

*Y Block: [TC (Sunflower oil, SFO); NTC (Non Sunflower oil, Not SFO)]

Pre-processing: 1st Derivative (order 2, window;21 pt, tails:polyinterp) + Mean center

Training Set: [97 × 902]

Prediction Set: [48 × 902] See Confusion Table below

Classification performance metrics	TC (SFO)	NTC (Not SFO)
Sensitivity (SENS -prediction stage)	0.90	1.00
Specificity (SPEC-prediction stage)	1.00	0.90
False positive rate (FPR)	0.00	0.10
False negative rate (FNR)	0.10	0.00
Positive predictive value (precision) (PPV)	1.00	0.95
Negative predictive value (NPV)	0.95	1.00
Youden index (YOUD)	0.90	0.90
F-measure (F)	0.95	0.97
Discriminant power (DP)	–	–
Efficiency (or accuracy) (EFFIC)	0.97	0.97
Misclassification rate (MR)	0.03	0.03
AUC (correctly classified rate)	0.95	0.95
Gini coefficient (Gini)	0.90	0.90
G-mean (GM)	0.95	0.95
Matthews correlation coefficient (MCC)	0.92	0.92
Chance agreement rate (CAR)	0.57	0.57
Chance error rate (CER)	0.44	0.44
Kappa coefficient (KAPPA)	0.92	0.92

Confusion Table:

Assigned class		97	48	145
	Sunflower	0	43	**43**
	Not SF	97	5	**102**
		Not SF	**Sunflower**	**T**
		Reference class		

Note. The hyphen "−" refers to metrics that cannot be determined since a division by zero is involved.

Table 3. Summary of discrimination/classification performance metrics obtained for the SIMCA one-input class model.

TARGET Class (TC): Sunflower oil

Features:

*X Block: [Reduced RAMAN instrumental fingerprints]

*Y Block: [TC (Sunflower oil, SFO); NTC (Non Sunflower oil, Not SFO)]

Pre-processing: 1st Derivative (order 2, window;21 pt, tails:polyinterp) + Mean center

Training Set: [97 × 902]

Prediction Set: [48 × 902] See Confusion Table below

Classification performance metrics	TC (SFO)	NTC (Not SFO)
Sensitivity (SENS -prediction stage)	0.93	0.81
Specificity (SPEC-prediction stage)	0.81	0.93
False positive rate (FPR)	0.19	0.07
False negative rate (FNR)	0.07	0.19
Positive predictive value (precision) (PPV)	0.99	1.00
Negative predictive value (NPV)	1.00	0.99
Youden index (YOUD)	0.74	0.74
F-measure (F)	0.96	0.90
Discriminant power (DP)	0.96	0.96
Efficiency (or accuracy) (EFFIC)	0.89	0.89
Misclassification rate (MR)	0.11	0.11
AUC (correctly classified rate)	0.87	0.87
Gini coefficient (Gini)	0.74	0.74
G-mean (GM)	0.87	0.87

Matthews correlation coefficient (MCC)	0.86	0.86
Chance agreement rate (CAR)	0.51	0.51
Chance error rate (CER)	0.44	0.44
Kappa coefficient (KAPPA)	0.78	0.78

Confusion Table:

Assigned class		Not SF	Sunflower	T
		97	48	145
	Nr	1	5	6
	I	6	3	9
	Sunflower	0	39	39
	Not SF	90	1	91
		Not SF	**Sunflower**	**T**
		Reference class		

Note. The hyphen "–" refers to metrics that cannot be determined since a division by zero is involved. Nr: Not recognized; I: Inconclusive.

Conclusions

The chemometric/machine learning study of the spectroscopic instrumental fingerprints of edible vegetable oils (sunflower and olive oils) obtained from a highly versatile portable analyser based on Spatially Offset Raman Spectroscopy (SORS) showed that they are capable to distinguish the analysed samples attending to the original raw material used in the oil's elaboration.

In addition, the different zones of the normalized spatial resolved RAMAN fingerprints allowed not only the capability to group the samples attending to the raw material but also to the intra-group differentiation among types of oils. Thus, by means of the unsupervised techniques used (HCA and PCA), it was demonstrated that the most influential attribute was the raw material, e.g. sunflower or olive oil, followed by the types (commercial categories) of each type of oil, both determined by the oleic acid content and the oleic/linoleic acid ratio in the analysed samples.

By applying supervised techniques different models to discriminate (SVM and kNN) and classify (SIMCA) were obtained. Although reliable quality metrics of the models were satisfactory, as far as authors concern, the best results were obtained in SIMCA model. This soft classification model permitted not only to classify the samples attending to the oil raw material (seed or fruit) but also to classify those samples belonging to HOSFO as inconclusive due to their high oleic acid content similar to olive oil.

References

1. Long, R.; Gulya, T.; Light, S.; Bali, K.; Mathesius, K.; Meyer, R. D. Sunflower Hybrid Seed Production in California. Available on-line: https://escholarship.org/uc/item/14k450p6 (last accessed 11/11/2023)

2. European Commission, Directorate-General for Agriculture and Rural Development, EU agricultural outlook for markets, income and environment 2021-2031, Publications Office of the European Union, 2021. Available on-line: https://data.europa.eu/doi/10.2762/753688 (last accessed 11/11/2023)

3. Praveen, H. G.; Nagarathna, T. K.; Gayithri, M.; Patil, M. I. Genetic Variability for Seed Yield, Oil Content and Fatty Acid Composition in Germplasm Accessions of Sunflower (Helianthus annuus L.) and their Response to Different Seasons. *Int. J. Curr. Microbiol. App. Sci.* **2018**, *7*(6), pp. 2120-2129. https://doi.org/10.20546/ijcmas.2018.706.252

4. Salas, J. J.; Martínez-Force, E.; Harwood, J. L.; Venegas-Calerón, M.; Aznar-Moreno, J. A.; Moreno-Pérez, A. J.; Ruíz-López, N.; Serrano-Vega, M. J.; Graham, I. A.; Mullen, R. T.; Garcés, R. Biochemistry of high stearic sunflower, a new source of saturated fats. *Prog. Lipid Res.* **2014**, *55*, pp. 30-42. https://doi.org/10.1016/j.plipres.2014.05.001

5. Salas, J. J.; Bootello, M. A.; Garcés, R. Food Uses of Sunflower Oils. In *Sunflower Chemistry, Production, Processing, and Utilization*. Salas J. J.; Enrique, M. F.; Dunford, N. T. (Eds.) *AOCS Press*, Champaign, IL, **2015**, pp. 441-464. https://doi.org/10.1016/B978-1-893997-94-3.50020-9

6. Harún, M. Fatty Acid Composition of Sunflower in 31 Inbreed and 28 Hybrid. *Biomed. J. Sci. Technol. Res.* **2019**, *16*(3), pp.12032-12038. http://dx.doi.org/10.26717/BJSTR.2019.16.002851

7. Zambelli, A.; León, A.; Garcés, R. Mutagenesis in sunflower. In *Sunflower*. American Oil Chemist's Society Press.: Nueva York, United States of America, **2015**, pp. 27-52.

8. Anushree, S.; André, M.; Guillaume, D. et al. Stearic sunflower oil as a sustainable and healthy alternative to palm oil. A review. *Agron Sustain Dev.* **2017**, *37* (18), pp. 1-10. https://doi.org/10.1007/s13593-017-0426-x

9. Esteki, M.; Simal-Gandara, J.; Shahsavari, Z.; Zandbaaf, S.; Dashtaki, E.; Vander Heyden, Y. A review on the application of chromatographic methods, coupled to chemometrics, for food authentication. *Food Control*, **2018**, 93, pp. 165-182. https://doi.org/10.1016/j.foodcont.2018.06.015

10. Jiménez-Carvelo, A. M.; Pérez-Castaño, E.; González-Casado, A.; Cuadros-Rodríguez, L. One input-class and two input-class classifications for differentiating olive oil from other edible vegetable oils by use of the normal-phase liquid chromatography fingerprint of the methyl trans esterified fraction. *Food Chem.* **2017**, *221*, pp. 1784-1791. https://doi.org/10.1016/j.foodchem.2016.10.103

11. Sherma, J.; Rabel, F. A review of thin layer chromatography methods for determination of authenticity of foods and dietary supplements. *J. Liq. Chromatogr. Relat. Technol.* **2018**, *41*(10), pp. 645-657. https://doi.org/10.1080/10826076.2018.1505637

12. International Olive Council. Determination of the sterol composition and content and alcoholic compounds by capillary gas chromatography. COI/T.20/Doc. No 26/Rev. 4, **2018**.

13. Carranco, N.; Farrés-Cebrián, M.; Saurina, J.; Núñez, O. Authentication and quantitation of fraud in extra virgin olive oils based on HPLC-UV fingerprinting and multivariate calibration. *Foods*, **2018**, *7* (4), pp. 1-15. https://doi.org/10.3390/foods7040044

14. International Olive Council. Method of analysis. Difference between actual and theoretical content of triacyclglycerols with ECN 42. COI/T.20/Doc. No 20/Rev. 4, **2017.**

15. International Olive Council. Method of analysis. Determination of fatty acid methyl esters by gas chromatography. COI/T.20/Doc. No 33/Rev. 4, **2017.**

16. Ruiz-Samblás, C.; González-Casado, A.; Cuadros-Rodríguez, L. Triacylglycerols determination by high-temperature gas chromatography in the analysis of vegetable oils and foods: a review of the past 10 years. *Crit. Rev. Food Sci. Nutr.*, **2015**, *55* (11), pp.1618-1631. https://doi.org/10.1080/10408398.2012.713045

17. Gómez-Caravaca, A. M.; Maggio, R. M.; Cerretani, L. Chemometric applications to assess quality and critical parameters of virgin and extra-virgin olive oil. A review. *Anal. Chim. Acta*, **2016**, *913*, pp. 1-21. https://doi.org/10.1016/j.aca.2016.01.025

18. Casale, M.; Simonetti, R. Review: Near infrared spectroscopy for analysing olive oils. *J. Near Infrared Spectrosc.* **2014**, *22*, pp. 59-80. https://doi.org/10.1255/jnirs.1106

19. Jiménez-Sanchidrián, C.; Ruiz, J. R. Use of Raman spectroscopy for analyzing edible vegetable oils. *Appl. Spectrosc. Rev.* **2016**, *51* (5), pp. 417-430. https://doi.org/10.1080/05704928.2016.1141292

20. Chen, J.; Zhao, Y.; Wu, R.; Yin, T.; You, J.; Hu, B.; Jia, C.; Rong, J; Liu, R.; Zhang, B.; et al. Changes in the Quality of High-Oleic Sunflower Oil during the Frying of Shrimp (*Litopenaeus vannamei*) *Foods* **2023**, *12*(6), 1332, pp. 1-14. https://doi.org/10.3390/foods12061332

21. Arroyo-Cerezo, A.; Jiménez-Carvelo, A. M.; González-Casado, A.; Koidis, A.; Cuadros-Rodríguez, L. Deep (offset) non-invasive Raman spectroscopy for the evaluation of food and beverages – A review. *LWT*, **2021**, *149*, 111822, pp. 1-8. https://doi.org/10.1016/j.lwt.2021.111822

22. Cuadros-Rodríguez, L.; Ruiz-Samblás, C.; Valverde-Som, L.; Pérez-Castaño, E.; González-Casado, A. Chromatographic fingerprinting: An innovative approach for food 'identitation' and food authentication - A tutorial. *Anal. Chim. Acta*, **2016**, *909*, pp. 9-23. https://doi.org/10.1016/j.aca.2015.12.042

23. Cuadros-Rodríguez, L.; Ortega-Gavilán, F.; Martín-Torres, S.; Arroyo-Cerezo, A.; Jiménez-Carvelo, A. M. Chromatographic Fingerprinting and Food Identity/Quality: Potentials and Challenges. *J. Agric. Food Chem.* **2021**, *69*, pp. 14428–14434. https://doi.org/10.1021/acs.jafc.1c05584

24. Szymańska, E. Modern data science for analytical chemical data – a comprehensive review. *Analytica Chimica Acta* **2018**, *1028*, pp. 1–10. https://doi.org/10.1016/j.aca.2018.05.038

25. Mialon, N.; Roig, B.; Capodanno, E.; Cadiere, A. Untargeted metabolomic approaches in food authenticity: A review that showcases biomarkers. *Food Chem.* **2023**, *398*, 133856, pp. 1-12. https://doi.org/10.1016/j.foodchem.2022.133856

26. Chemometrics vs Machine Learning. Available on-line: https://ondalys.fr/en/scientific-resources/chemometrics-vs-machine-learning/ (accessed on 11/10/2023)

27. Bikrani, S.; Jiménez-Carvelo, A. M.; Nechar, M.; Bagur-González, M. G.; Souhail, B.; Cuadros-Rodríguez, L. Authentication of the geographical origin of margarines and fat-spread products from liquid chromatographic UV-absorption fingerprints and chemometrics. *Foods* **2019,** *8*(11), 588, pp. 1-12. https://doi.org/10.3390/foods8110588

28. Pérez-Castaño, E.; Medina-Rodríguez, S.; Bagur-González, M. G. Discrimination and classification of extra virgin olive oil using a chemometric approach based on TMS-4,4'-desmetylsterols GC(FID) fingerprints of edible vegetable oils. *Food Chemistry,* **2019**, *274*, pp. 518–525. https://doi.org/10.1016/j.foodchem.2018.08.128

29. Ortega-Gavilán, F.; Jiménez-Carvelo, A. M.; Cuadros-Rodríguez, L.; Bagur-González, M. G. The chromatographic similarity profile – An innovative methodology to detect fraudulent blends of virgin olive oils. *J. Chromatogr. A.* **2022**, *1679*, 463378, pp. 1-12. https://doi.org/10.1016/j.chroma.2022.463378

30. Guerrero-Chanivet, M.; Ortega-Gavilán, F. Bagur-González, M. G.; Valcárcel-Muñoz, M. J.; García-Moreno, M. V.; Guillén-Sánchez, D. A. Pattern Recognition of GC-FID Profiles of Volatile Compounds in Brandy de Jerez Using a Chemometric Approach Based on Their Instrumental Fingerprint. *Food Bioprocess Technol.* **2023**, *16*, pp. 1963-1975. https://doi.org/10.1007/s11947-023-03039-8

31. Guerrero-Chanivet, M.; Ortega-Gavilán, F. Bagur-González, M. G.; Valcárcel-Muñoz, M. J.; García-Moreno, M. V.; Guillén-Sánchez, D. A. Influence of Oak Species, Toasting Degree, and Aging Time on the Differentiation of Brandies Using a Chemometrics Approach Based on Phenolic Compound UHPLC Fingerprints. *J. Agric. Food Chem.* **2023**, XXXX, XXX, XXX-XXX. https://doi.org/10.1021/acs.jafc.3c00501

32. Ellis, D. I.; Eccles, R.; Xu, Y.; Griffen, J.; Muhamadali, H.; Matousek, P.; Goodall, I.; Goodacre, R. Through-Container, Extremely Low Concentration Detection of Multiple Chemical Markers of Counterfeit Alcohol Using a Handheld SORS Device. *Sci. Rep.* **2017**, *7* (1), pp. 1-8. https://doi.org/10.1038/s41598-017-12263-0.

33. Schorn-García, D.; Ezenarro, J.; Aceña, L.; Busto, O.; Boqué, R.; Giussani, B.; Mestres, M. Spatially Offset Raman Spectroscopic (SORS) Analysis of Wine Alcoholic Fermentation: A Preliminary Study. *Ferment.* **2023**, *9*(2), 115, pp.1-12. https://doi.org/10.3390/fermentation9020115

34. Ostovar Pour, S.; Afshari, R.; Landry, J.; Pillidge, C.; Gill, H.; Blanch, E. Spatially offset Raman spectroscopy: A convenient and rapid tool to distinguish cheese made with milks from different animal species. *J. Raman Spectrosc.* **2021**, *52*(10), pp. 1705-1711. https://doi.org/10.1002/jrs.6179

35. Arroyo-Cerezo, A.; Jiménez-Carvelo, A. M.; González-Casado, A.; Ruisánchez, I.; Cuadros-Rodríguez, L. The potential of the spatially offset Raman spectroscopy (SORS) for implementing rapid and non-invasive in-situ authentication methods of plastic-packaged commodity foods – Application to sliced cheeses. *Food Control*, **2023**, *146*, 109522, pp.1-8. https://doi.org/10.1016/j.foodcont.2022.109522

36. Jiménez-Carvelo, A. M.; Arroyo-Cerezo, A.; Bikrani, S.; Jia, W.; Koidis, A.; Cuadros-Rodríguez, L. Rapid and non-destructive spatially offset Raman spectroscopic analysis of packaged margarines and fat-spread products. *Microchem. J.* **2022**, *178*, 107378, pp.1-13. https://doi.org/10.1016/j.microc.2022.107378

37. Varnasseri, M.; Muhamadali, H.; Xu, Y.; Richardson, P. I. C.; Byrd, N.; Ellis, D. I.; Matousek, P.; Goodacre, R. Portable through Bottle SORS for the Authentication of Extra Virgin Olive Oil. *Applied. Sciences.* **2021**, *11*(18), 8347, pp.1-13. https://doi.org/10.3390/app11188347

38. Araújo, G. A.; Azcarate, S. M.; Špánik, I.; Khvalbota, L.; Goicoechea, H. C. Pattern recognition techniques in food quality and authenticity: a guide on how to process multivariate data in food analysis. *TrAC Trends Anal. Chem.* **2023**, 164, 117105, pp.1-26. https://doi.org/10.1016/j.trac.2023.117105

39. Zhang, L.; Li, P.; Sun, X.; Wang, X.; Xu, B.; Wang, X.; Ma, F.; Zhang, Q.; Ding, X. Classification and Adulteration Detection of Vegetable Oils Based on Fatty Acid Profiles. *J. Agric. Food Chem.* **2014**, *62* (34), pp. 8745–8751. https://doi.org/10.1021/jf501097c

40. Kennard, R.; Stone, L. Computer Aided Design of Experiments. *Technometrics*, **1969**, 11(1), pp. 137-148. https://doi:10.2307/1266770