Microchemical Journal 178 (2022) 107378

Contents lists available at ScienceDirect

Microchemical Journal

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

Rapid and non-destructive spatially offset Raman spectroscopic analysis of packaged margarines and fat-spread products

Ana M. Jiménez-Carvelo [a], Alejandra Arroyo-Cerezo [a,*], Sanae Bikrani [b], Wenyang Jia [c], Anastasios Koidis [c], Luis Cuadros-Rodríguez [a]

[a] Department of Analytical Chemistry, University of Granada, C/ Fuentenueva s/n, E-18071, Granada, Spain
[b] Department of Chemistry, Faculty of Sciences, University of 'Abdelmalek Essaàdi', Av. Sebta, Mhannech II, 93002, Tetouan, Morocco
[c] Institute for Global Food Security, Queen's University, 18-30 Malone Road, Belfast BT9 5BN, Northern Ireland, United Kingdom

ARTICLE INFO

Keywords:
Spatially offset Raman spectroscopy (SORS)
Non-destructive analytical techniques
Chemometrics and data mining
In-pack measurement
Food quality and authenticity
Margarines and fat-spreads

ABSTRACT

Spatially offset Raman spectroscopy (SORS) is a novel technique capable of measuring samples through the original packaging and recovering the spectra without the contribution of surface layers. Here, a portable SORS equipment was used to measure 62 samples of margarines and fat spreads through the original plastic container. Chemometric tools were used to analyse the data obtained. A total of 25 classification models were developed based on: (i) geographical origin, (ii) vegetable oils and (iii) some significant minor constituents present in the samples. Partial least squares-discriminant analysis (PLS-DA), support vector machine (SVM) and soft independent modelling of class analogy (SIMCA) were used for model classification. Quantitative analysis using the partial least squares regression (PLSR) method was also performed to determine the total fat content. In parallel, a benchtop conventional Raman spectrometer was used to analyse the same samples, develop the models with the same training and validation sets in order to compare the results. The calculated classification performance metrics showed better classification models from SORS data than conventional Raman spectroscopy (CRS), highlighting the one-class SIMCA models for margarines containing phytosterols, olive oil or linseed oil. These models exhibited very high predictability (performance parameters with values equal to or higuer than 0.8, 0.9 and 1, respectively). The quantitation model developed from SORS exhibited a higher R^2 than from CRS data, and prediction errors below 5% from SORS versus errors between 5 and 13% from CRS data.

These results reveal the ability of SORS to avoid the influence of fluorescence, a major drawback when analysing Raman spectra, but also the potential of the technique as a fast, non-destructive and non-invasive analytical technique in the field of food analysis. In conclusion, the tandem 'SORS-chemometrics' has been shown to be a potential tool in the food quality and food authentication fields. Thus, it is necessary to perform further investigations in this field in order to advance the knowledge of this technique and to be able to develop new methods of rapid analysis.

1. Introduction

Advanced Raman spectroscopy techniques such as spatially offset Raman spectroscopy (SORS) have demonstrated the ability to overcome some disadvantages of conventional Raman spectroscopy (CRS). When SORS is applied, Raman signal is acquired at a certain distance from the laser incidence (some millimetres) and the collected signal provides spectral information of both the outer and inner layers of the measured material due to the spatially shift. This shift allows the deep layers photons to be emitted from a laterally shifted point of the incidence region while the surface photons are emitted from the same incidence point. The pathway followed by the interior photons is randomised, so it is more likely to retrieve this interior emission from a shifted point than from the same point of incidence [1]. For this, SORS becomes even more relevant by being able to collect photon emission from the inside of diffusely scattering materials, since most samples are either opaque or contained in opaque materials, while allowing measurements to be performed without damaging the sample, making SORS a remote, non-invasive and non-destructive technique [2,3].

Ensuring the stability of the sample during measurements as well as

* Corresponding author.
E-mail address: arroyoc@ugr.es (A. Arroyo-Cerezo).

https://doi.org/10.1016/j.microc.2022.107378
Received 14 December 2021; Received in revised form 7 March 2022; Accepted 8 March 2022
Available online 17 March 2022
0026-265X/© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

properly selecting the optimal offset are important parameters for measurement success. The optimal offset is characteristic for each type of sample as the intensity of spectral bands and the depth range to be reached depends on it [1,4]. Conventional instruments to perform SORS measurements are custom design builds, primarily composed of a laser source, a charge-coupled device (CCD) camera as a sensor and a fibre bundle or fibre optic. Moreover, filters are used to supress the fluorescence radiation emitted by the measured sample that could interfere with the low intensity Raman signals. Usually, wavelengths excitation in SORS measurements are 785, 830 and 1064 nm. The latter has demonstrated a greater ability to remove fluorescence [2].

However, since 2017, it is possible to acquire Raman signals with branded portable and handheld SORS equipment [5]. These instruments were manufactured for particular applications, making easier fast and effective analysis, e.g., for hazardous substances security screening or raw materials control in pharmaceutical industry. Commercial equipment still has some drawbacks inherent to the Raman spectroscopy, such as the inability to perform measurements through fully opaque package materials as metal materials, including tetrabrik. Minor sensitivity and lower signal-to-noise ratio in collected Raman signals have also been reported using commercial instruments [2].

Note that when Raman spectroscopy is employed as an analytical technique to evaluate the authenticity or quality of foodstuffs, the signal obtained is considered as non-specific, in which all the chemical and structural information of the sample is collected. In other words, this signal is an instrumental fingerprint and therefore it is necessary to apply chemometric/data mining tools in order to extract the information of interest which is not shown in an evident form [6]. The tandem composed by fingerprinting methodology and chemometrics is focused on the development of multivariate (qualitative or quantitative) models using proper mining/machine learning algorithms [7], so-called pattern recognition methods. The aim is to establish the belonging to one class or another of a set of samples with a characteristic feature (e.g., origin, ingredients, manufacturing, differentiated quality claims, etc.) or to carry out the quantitation of one or more feature-related parameters [8]. Chemometric pattern recognition methods are usually divided into two categories: unsupervised and supervised methods. Unsupervised methods show the intrinsic data pattern and are typically used to exploratory purposes. The most common unsupervised approaches are hierarchical cluster analysis (HCA) and principal components analysis (PCA). Supervised methods consider the belonging class of the sample and the development of model involves a training step followed by a validation step which can be performed using new samples different from those used in the training step (external validation) or using the same samples (cross-validation). Some of the supervised methods include k-nearest neighbours (kNN), soft independent modelling by class analogy (SIMCA), partial least squares-discriminant analysis (PLS-DA), or support vector machine (SVM). Furthermore, as far as quantitative analysis is concerned, the most widely used is by partial least squares regression (PLSR) [9].

SORS measurements in combination with chemometric tools have been used for multiple applications in different fields. To date, the most explored with industrial application is pharmaceutical industry for the identification of raw materials, drug detection through packaging or control the adulteration of drugs [2]. Other fields of application are the non-invasive analysis of artworks [10], the detection of explosives in liquids in the context of the security [11], and biomedical applications [12] or some in the food and beverage sector [2,4]. However, despite the potential of this analytical technique to be used in the food quality control, food safety or food authentication fields among others, applications of SORS-chemometrics are still limited in the scientific literature and as far as is known, they are not yet applied at the industry level.

Margarines and related fat-spread products could be an appropriate target for SORS research. The container for these products is usually made of plastic, a material that is permeable to the laser allowing acquisition of Raman spectra of the sample contained inside. Historically, margarine was created to replace butter as a lower cost option. It is a water-in-oil solid emulsion composed mainly of vegetable fats (such as sunflower, rapeseed, palm and olive oil) and rarely animal fats up to a maximum of 3% [13]. Unlike butter which must derive only from milk, other ingredients are also permitted in margarine manufacturing such as phytosterols, vitamins, minerals or sugars as well as additives, including colouring agents, emulsifiers, stabilisers or antioxidants [14]. Fat-spreads are classified according to the total fat percentage by Codex Alimentarius. Commonly, for a product to be properly called margarine, it must have at least 80% fat, otherwise it is called fat-spread (with less than 80%) [15]. However, the EU legislation establishes more categories: (i) margarine, if fat content is between 80% and 90%, (ii) three-quarter-fat margarine, if fat content is between 60% and 62%, (iii) half-fat margarine if fat content is between 39% and 41% and (iv) fat spreads X% for the rest of fat content percentages [16]. The legislation of others counties may be different. For instance, Moroccan law provides that margarine is any edible fat other than butter and lard. This legislation does not specify a minimum fat percentage required to call the product margarine, and also allows for the addition of up to 10% milk fat, either from milk or whipped cream [17].

Most traditional analytical techniques, such as high performance liquid chromatography (HPLC) or gas chromatography (GC), employed to analyse this type of product involve several sample pre-treatments, such as previous extraction or isolation of the required compound or compounds family as well as need long time to perform the analysis [18]. However, vibrational spectroscopy techniques such as Fourier transform-near infrared (FT-NIR) provide the possibility to conduct rapid, simple and non-destructive analysis of margarines and fat-spreads in terms to assess proper quality control and food authentication [19]. Among these techniques, SORS has already been applied to detect the presence of margarine in butter, i.e., to detect butter adulteration [20]. Thus, a food such as margarine could be a good product to be analysed with the SORS technique in a fast, non-invasive and non-destructive form, for authenticating this food product.

The present paper aims the use of SORS to extract Raman spectra of a set of margarines and fat-spread products measured through the original packaging using a recently commercialised handheld instrument. Different multivariate models have been developed using unsupervised (PCA) and supervised (PLS-DA, SVM, SIMCA) methods for prior screening and qualitative classification, and PLSR to quantify the fat contents of the samples. To further improve the reliability of the measurements performed, a comparison with CRS has been also carried out.

2. Material and methods

2.1. Margarines and spreads samples

A total of 62 samples of margarines and fat-spreads from different geographic origins of production were analysed. Table 1 shows the different geographic origins of manufacture and the number of samples from each of them. The samples were purchased in different local grocery shops and supermarkets in Spain, France, United Kingdom and Morocco. As for the samples whose manufacturing origin is different

Table 1
Geographical origin of the margarine samples included in the study.

Geographical origin		Number of samples
Europe	Spain	19
	France	13
	United Kingdom	12
	Belgium	4
	Germany	1
	The Netherlands	1
Morocco		12
Total number of samples		62

(Belgium, Germany and Holland), they were purchased in retail in Spain. In addition to the geographic origins of manufacture, the samples differ in composition (ingredients) and amount of macronutrients (carbohydrates, fats and proteins). After purchase, the samples were stored under conditions similar to those of retail sale, i.e., refrigerated at 4 °C.

For CRS measurements, a small portion of each sample (approximately 10 g) was transferred to a vial for analysis. For the SORS measurements, this was not necessary as they were measured through the original container.

2.2. SORS measurements

The commercial SORS equipment used for measuring margarines and fat-spreads products through the original packaging was the Vaya Raman (Agilent Technologies, Santa Clara, CA, USA). This device uses a laser with an excitation wavelength of 830 nm, which allows fluorescence to be suppressed. The spectral range was 350–2000 cm^{-1}, while the maximum power of the laser was 450 mW (100% power), which is user adjustable.

The equipment performs two measurements: with zero offset (equivalent to the CRS spectrum) and with spatial offset, namely with an offset of 0.7 mm from the point of incidence of the laser to the collection point. After internal signal processing, the SORS spectrum is obtained: a Raman spectrum of the sample without the influence of the container layers. Fig. 1 shows an example of the spectra obtained. Fig. 1A shows the raw spectra of the container (spectrum without offset, light green line) and of one of the samples without offset (dark green line) and with offset (blue line). Note that the contribution of the container seen in the spectrum (dark green line), coinciding with the spectrum of the container, is completely subtracted from SORS measurement. Fig. 1B shows the final spectrum of the same sample pre-processed and normalised. The 'final spectrum' obtained from SORS were the data used to carry out the multivariate data treatment.

Measurements were taken directly from the original packaging of the 62 margarine or fat-spread samples. The measurement time for each sample was between 30 s and 2 min, while the exposure time of the samples to the laser was 0.5 to 2 s.

2.3. CRS measurements

The measurements were performed with the IDRAMAN Reader (Ocean Optics, Oxford, UK), equipped with laser wavelength at 785 nm with 100 mW laser power. The scattered light was collected on a 2048-element NIR-enhanced CCD array with thermoelectric cooling to −10 °C. The samples were heated to 40 °C in a water bath until fully mixed, and a 2 mL liquid was placed into a sealed glass vial. For the spectra acquisition, 2 mL vial analysed sample was used and each sample was obtained three times. The spectral range was 200–3200 cm^{-1} with an integration time of 20 s, and all Raman measurements were conducted at room temperature. Fig. 1C shows an example of the obtained spectra of one of the samples using CRS.

2.4. Multivariate data treatment

Raman and SORS raw data were exported from CSV format (comma-separated vectors) to.mat using MATLAB (Mathworks, Massachusetts, USA, version R2013a).

All multivariate data treatment was carried out using PLS_Toolbox (Eigenvector Research Inc. MA, USA, version 7.5.0) working under the MATLAB framework. Chemometric tools applied were (i) PCA as a non-supervised pattern exploratory method, (ii) PLS-DA, SVM and SIMCA as supervised pattern recognition methods to build different classification models and (iii) PLSR to develop quantitation models to estimate fat percentage. All of the above methods were applied to both Raman and SORS data and full comparisons are presented in the results section.

For the classification models, a plot with training and validation set will be presented in the next section. The discrimination threshold was established at 0.5 for all models and a range of ± 0.1 was assigned to designate an area of inconclusive results. In addition, a cut-off was also set at ± 1 (i.e., 1.5 above and −0.5 below) and samples falling in these areas, both at the upper and lower limits, were assigned to the inconclusive results group when calculating quality parameters.

The most appropriate pre-processing was chosen according to the multivariate method to be used. Different pre-processing methods were tested and the best results were obtained by applying mean center or autoscale, as discussed in the next section.

A total of 50 classifications models were developed according to different characteristics of the samples. Firstly, 8 models related to the different geographical origin were developed, 4 for each of the techniques used (CRS and SORS), differentiating between samples manufactured in Morocco, Spain, France and United Kingdom. Then, 42 models (21 with the data obtained for each technique) based on the different ingredients that constitute the margarines analysed in order to differentiate between those that have a particular ingredient from those that do not. The ingredients consider as 'target class' were sunflower oil, olive oil, linseed oil, palm oil or fat, buttermilk, phytosterols and lecithin. Different classification performance metrics, such as sensitivity, specificity, positive and negative predictive value or efficiency, among others, were determined according to the tutorial published by Cuadros *et al.* [21]. For the four quantitation models developed, the root mean square error of validation (RMSEV), the mean absolute error of validation (MAEV), the median absolute error of validation (MdAEV), the standard deviation of validation residuals (SDV) [22] and the coefficient of determination (R^2) were used to evaluate prediction accuracy. A comprehensive layout is shown in Fig. 2.

Note that for classification purposes, samples were divided into two general classes. For the geographical origin models, one class was the country in which the samples were manufactured and the other class consisted of the rest of the samples manufactured in another country (e.g. for 'Spain / no Spain' model, the samples of the 'no Spain' class were the samples manufactured in the UK, France, Belgium, Germany and Netherlands). And for the ingredients models, one class is the 'target class' (samples containing the ingredient in its composition) and the other class is composed of the samples that do not have that ingredient.

3. Results and discussion

The Raman spectra peak at 1750 cm^{-1} (see Fig. 1B) corresponds to the ester bonds related to the presence of triglycerides [23]. The band observed at 3050–3090 cm^{-1} (see Fig. 1C) may be attributable to lipids, corresponding to the peaks observed in different vegetable oils in the study of Dyminska *et al.* [24]. Peaks around 875 cm^{-1} may be attributed to phosphodiester symmetric stretching of the phospholipids present in the margarine samples [25]. These peaks only showed in either SORS or CRS (Table 2), therefore, these three bands were excluded for the model development. The rest of the peaks correspond to margarine samples can be found below. The band around 1654 cm^{-1}, belongs to the aromatic ring stretch, were related to oil content [26,27]. The peak around 1440 cm^{-1} is related to CH_2 deformations [28], while the peak observed around 1260 cm^{-1} is assigned to the (C–H) bending vibration at the cis double bond in R–HC = CH–R [29]. And the band at 1061 cm^{-1} originate from the (C–C) stretching [30].

3.1. Exploratory analysis

PCA was employed to study the natural grouping of the 62 margarine samples. The pre-processing chosen was mean centring for both CRS and SORS data. Five and eight principal components (PCs) explained 86.91% and 99.96% of the cumulative variance (CV) for the SORS and CRS data, respectively. Fig. 3 shows the scores obtained for each sample in the first two PCs of the SORS and CRS data. It is evident that with both techniques is possible to establish a grouping of the samples according to

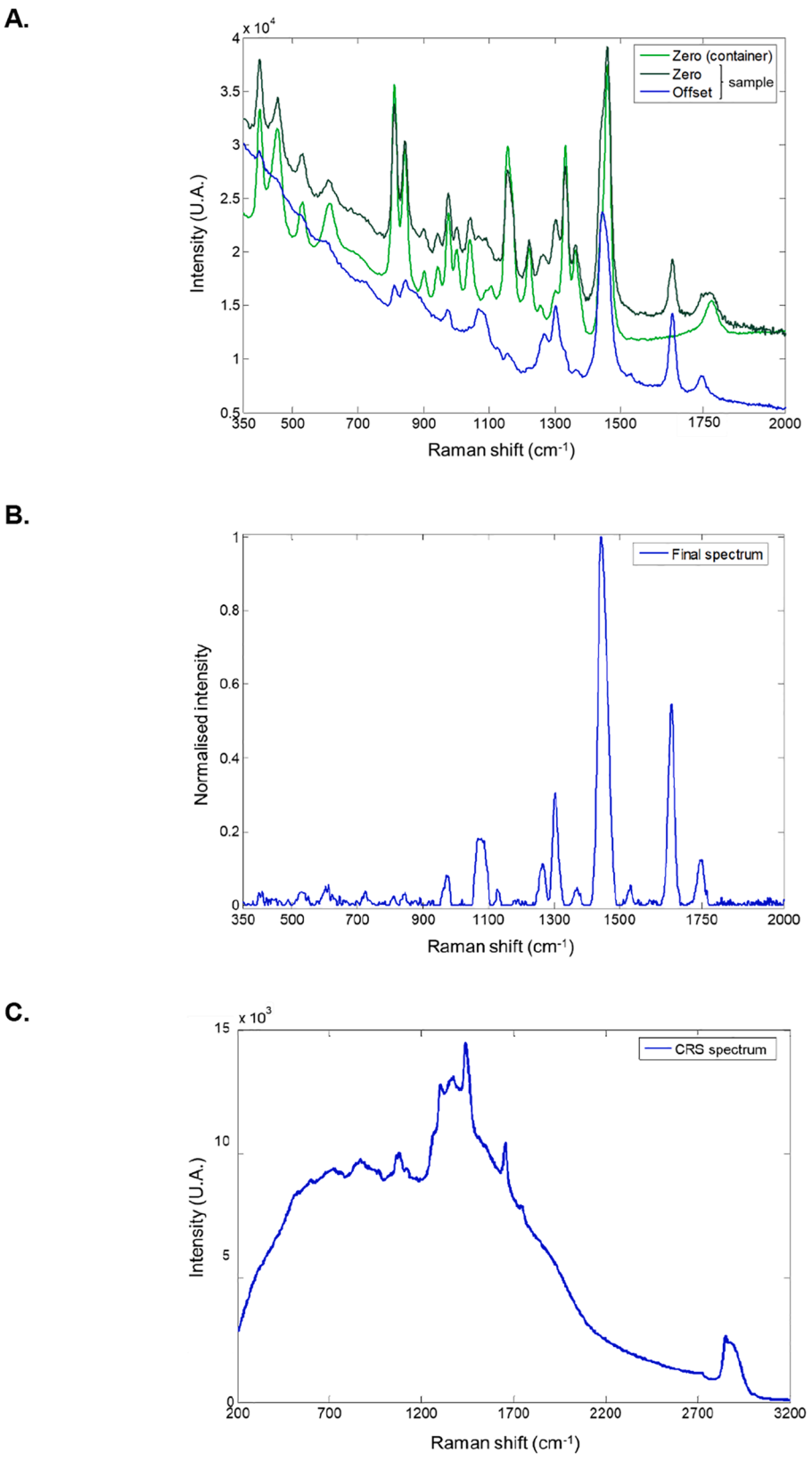

Fig. 1. Raw Raman spectra of the container (zero offset) and of one of the samples measured through packaging without (zero) and with offset using spatially offset Raman spectroscopy (SORS) (A), final Raman spectrum of the same sample after pre-processing and normalise using SORS (B), Raman spectrum of the same sample using conventional Raman spectroscopy (CRS) (C).

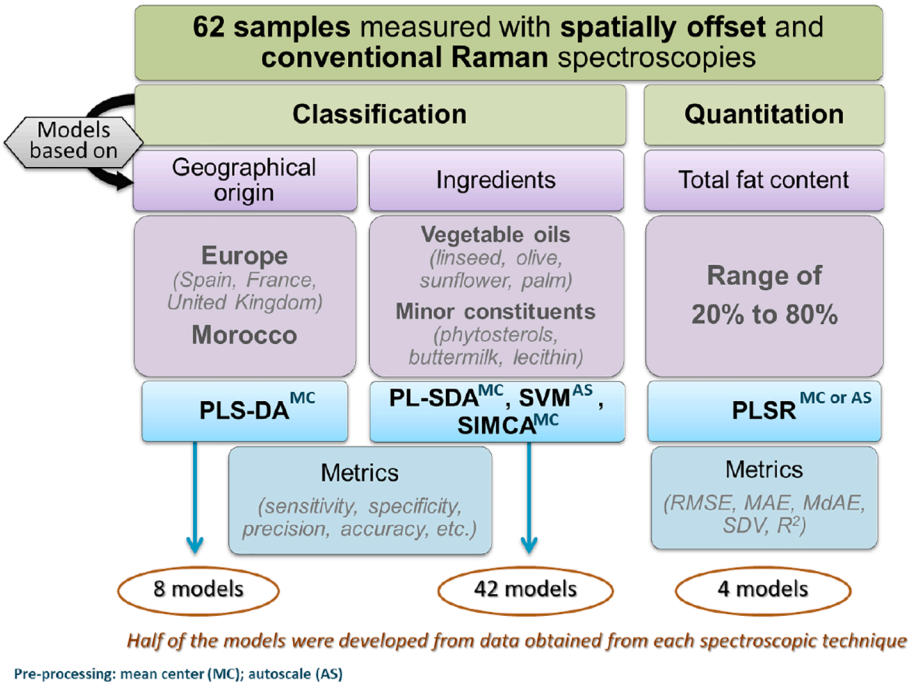

Fig. 2. Comprehensive layout of the strategy used to perform the classification and quantitation models.

Table 2
Band assignment of the SORS and CRS spectra. Table adaptated from [23].

Raman shift from SORS (cm^{-1})	Raman shift from CRS (cm^{-1})	Molecule	Group	Vibration
1750	—	RC = OOR	C = O	stretching
1656	1654	*cis* RCH = CHR	C = C	stretching
1442	1440	– CH_2	C–H	deformation
1264	1252	*cis* RCH = CHR	=C–H	deformation
1067	1061	– $(CH_2)_n$ –	C–C	stretching
—	875	– $(CH_2)_n$ –	C–C	stretching

"—": peak not shown in the spectrum.

their geographical origin of production. Fig. 3A and 3B illustrate a clear separation between European or Moroccan origin, and in Fig. 3C and 3D, slight differences and natural groupings can be seen, differentiating those of European origin between Spain, France, United Kingdom and other countries.

For the data obtained from CRS measurements, the positive part of PC1 is related to the samples coming from Morocco and the negative contribution of the same PC to the European samples. However, regarding the SORS data, the difference between the two origins is related to the negative contribution of PC2 for the Morocco origin and the positive contribution of the same PC for the Europe origin.

With respect to the different countries of origin of the European samples, more separation between the groups from Spain and United Kingdom can be observed when comparing the SORS (Fig. 3C) with respect to the CRS data (Fig. 3D), where a greater overlap between the groups is evident.

3.2. Classification by geographical origin

According to the results of the exploratory analysis, different classification models were developed to classify the samples on the basis of geographic origin namely: 'Europe / Moroco' model (50 / 12 samples, respectively), 'Spain / no Spain' with membership in the 'Europe' class (19 / 31 samples, respectively), 'France / no France' with membership in the 'no Spain' class (13 / 18 samples, respectively) and 'United Kingdom / no United Kingdom' model with membership in the 'no France' class (12 / 6 samples, respectively). Pre-processing was mean center for both SORS and CRS data.

PLS-DA method was used to perform the classification models according to geographical origin. Fig. 4 shows an example of the classification plot obtained for 'Europe / Morocco' model for both SORS data (Fig. 4A) and CRS data (Fig. 4B). The models were built with a training set composed of 43 samples and an external validation set composed of 19 samples. Five and six latent variables (LVs) explaining 85.53% and 99.95% of the CV for SORS and CRS data respectively, were selected for the development of the models.

Table 3 shows the results of the four classification models: (i) 'Europe / Morocco', (ii) 'Spain / no Spain' (seven LVs explaining 85.73% CV for SORS data and three LVs explaining 99.53% CV for CRS data), (iii) 'France / no France' (six LVs explaining 82.68% CV for SORS data and two LVs explaining 99.63% CV for CRS data) and (iv) 'United Kingdom / no United Kingdom' (six LVs explaining 90.21% CV for SORS data and four LVs explaining 99.56% CV for CRS data). The corresponding classification plots for the last three models can be found in the supplementary material (Figures S1-S3). The training sets were composed of 35, 23 and 12 samples and the external validation sets of 15, 8 and 6 samples for the 'Spain / no Spain' model, the 'France / no France' model and the 'United Kingdom / no United Kingdom' model, respectively.

Some of the most relevant classification performance metrics calculated for the geographical origin classification models are shown in Table 4. The values correspond to the average calculated after estimating the parameters for the two classes that constitute each model. It should be noted that the best results were obtained for both 'Europe / Morocco' and 'France / no France' models, and especially with the data obtained with the SORS. This may be related to the influence of fluorescence on the spectra obtained with the CRS or with the applied preprocessing. Both aspects are discussed in section 3.6.

3.3. Classifications by vegetable oil types

As mentioned above, legislation allows margarines and fat spreads to include different types of vegetable oils or fats in their composition. This

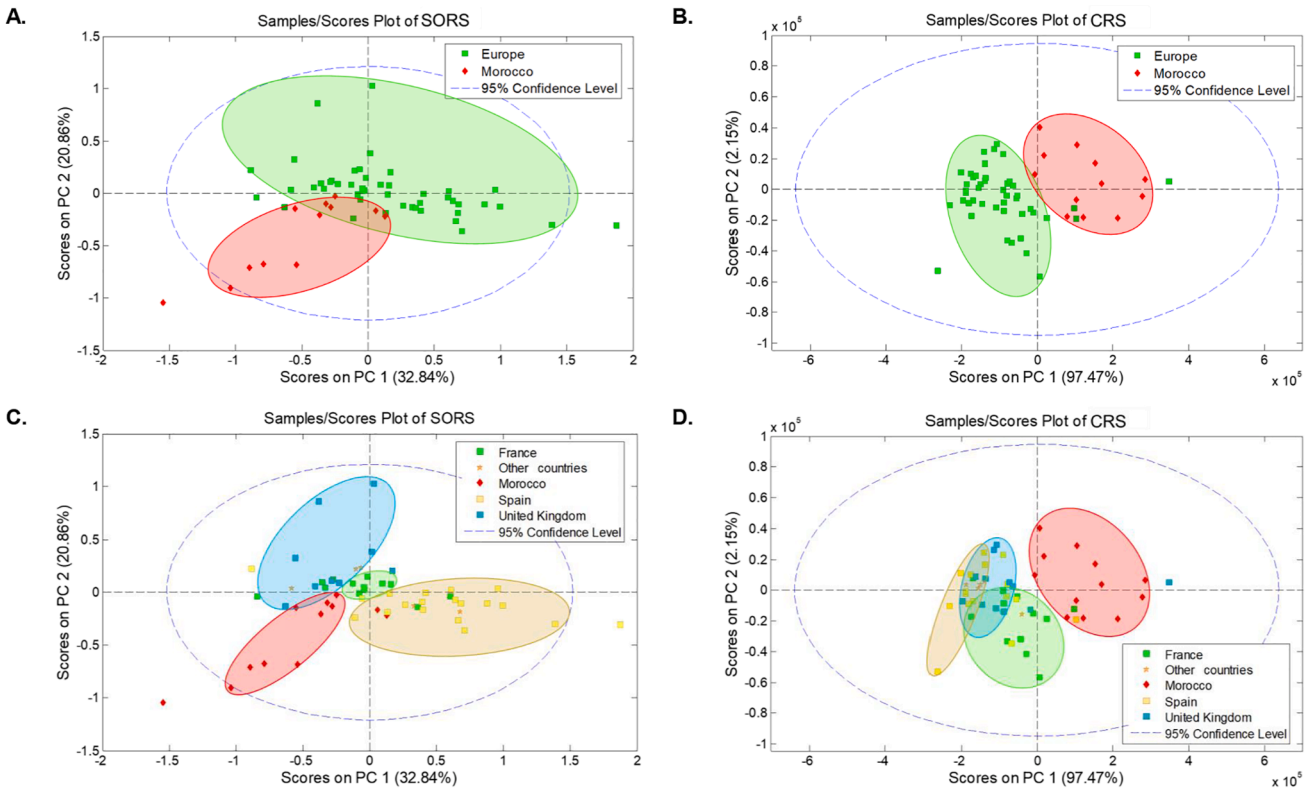

Fig. 3. PCA scores plot of Raman spectra of margarines samples with data obtained from SORS data (A,C) and CRS (B,D). The samples are differentiated according to their geographical origin of fabrication following the legend.

aspect has been used to perform different classification models according to some of the types of oil present in the analysed samples. PLS-DA, SVM and SIMCA models based on different samples parameters were built with data obtained both from SORS and CRS measurements. Linseed oil, olive oil, sunflower oil and palm oil were the four selected vegetable oil types to perform the classification models. PLS-DA and SIMCA classification models were developed applying mean center as pre-processing, while SVM classification models applying autoscale. This was applicable for the data obtained by both techniques (SORS and CRS) in order to be compared later.

SIMCA models were developed following the one-class strategy (OC-SIMCA). This approach has been highlighted several times in the literature for food authentication and it is especially relevant when classifying foods that have a particularity (in this study, an ingredient) from the rest that do not [31]. For this purpose, the target input class with which the model is trained are only samples that include this ingredient in their composition. The plots to be presented for the models developed with OC-SIMCA represent Hotteling's T^2 versus Q-residuals of the target class at a confidence level of 95%. To consider a sample as a target class, i.e. 'within the model', the values of these two statistics (T^2 and Q) must be below 1, within the bounded square in the corresponding plot. The OC-SIMCA plots shown represent the values obtained only for the validation set and are zoomed to see the confidence area.

'Linseed / no linseed' models are shown in Fig. 5. PLS-DA models were built with eight LVs each, explaining 79.53% and 99.99% of the CV, and OC-SIMCA models were built with seven and two PCs explaining 88.27% and 99.81% of the CV, respectively for SORS and CRS data in both cases. PLS-DA and SVM models were developed with a training set of 43 samples and 19 samples for the external validation set. The OC-SIMCA models were performed with a training set of 18 samples of the target class and the validation set including all samples (62 in total).

Best results were obtained with PLS-DA models compared to SVM, especially for SORS data. Furthermore, with respect to the OC-SIMCA model, notice that better results were obtained with SORS data compared to CRS despite that the same samples were analised. This can be seen in Table 5 which gives the quality parameters of the six models built to differentiate the samples containing linseed oil. For this case, the OC-SIMCA model developed from the SORS data is the one that best classifies the samples according to this parameter, all the classification performance metrics are above 0.8.

Note that negative value of the Matthews correlation coefficient of the OC–SIMCA model built from CRS data indicate a negative correlation. However, negative value of the Kappa coefficient indicates the model's prediction to be worse than a random prediction. This occurs in a few more models throughout the text.

The corresponding classification plots for the others vegetable oils models (olive, sunflower and palm) can be found in the supplementary information (Figures S4-S6), as well as the parameters selected to carry out the PLS-DA and OC-SIMCA models, explained in the figure caption. The quality parameters of these models are shown in Tables S2-S4.

For the 'olive / no olive' models developed with SVM the quality parameters were not calculated (Table S2), because these models did not provide good results, not being able to distinguish samples with and without olive oil (Figure S4C and S4D). This may be due to the limited number of samples containing olive oil in their composition (6 samples out of 62). The same occurs with the PLS-DA models, which were not able to correctly discriminate the samples of the validation set containing olive oil. However, it is again remarkable the results obtained with the OC-SIMCA model with the SORS data, for which all quality parameters are above 0.9.

Regarding the other models ('sunflower / no sunflower' and 'palm / no palm'), the results were generally not as satisfactory, perhaps because the amount of these ingredients in the margarines is relatively small, or because the spectra do not offer sufficiently relevant chemical information to detect these ingredients compared to other components. Even so, the quality parameters of the 'sunflower / no sunflower' SVM model developed from the data obtained with CRS (Figure S5C, Table S3) and the PLS-DA 'palm / no palm' model from the SORS data

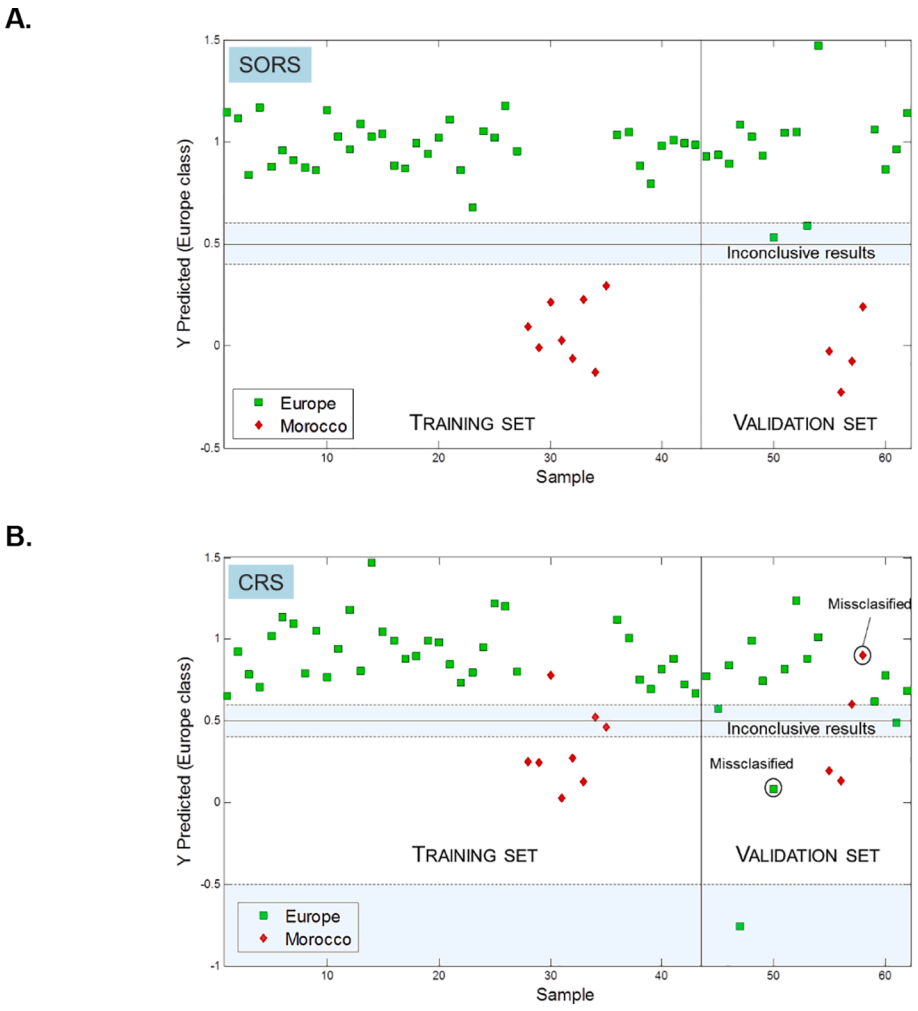

Fig. 4. Classification plots obtained for the 'Europe / Morocco' models developed by applying PLS-DA to the SORS data (A) and CRS data (B). The green squares (■) and the red rhombuses (◆) represent the European and Moroccan origin of the margarines, respectively.

Table 3
Results of the assigned class to the validation sets of the classification models developed according to the geographical origin of the samples, both for SORS and CRS data.

			SORS			CRS		
			Reference class			Reference class		
Europe / Morocco			Europe	Morocco	TNA	Europe	Morocco	TNA
	Assigned class	Europe	13	0	13	11	1	12
		Morocco	0	4	4	1	2	3
		Inconclusive	2	0		3	1	
	TNB		15	4		15	4	
Spain / no Spain			Spain	no Spain	TNA	Spain	no Spain	TNA
	Assigned class	Spain	4	0	4	2	2	4
		No Spain	0	8	8	2	7	9
		Inconclusive	2	1		2	0	
	TNB		6	9		6	9	
France / no France			France	no France	TNA	France	no France	TNA
	Assigned class	France	3	0	3	2	0	2
		No France	0	4	4	1	3	4
		Inconclusive	0	1		0	2	
	TNB		3	5		3	5	
UK / no UK			UK	no UK	TNA	UK	no UK	TNA
	Assigned class	UK	2	0	2	2	1	3
		No UK	0	2	2	0	0	0
		Inconclusive	2	0		2	1	
	TNB		4	2		4	2	

TNA: total number of samples assigned to the class; TNB: total number of samples belonging to the class; UK: United Kingdom.

Table 4
Classification performance metrics for PLS-DA geographical origin models developed from both types of Raman techniques.

	Europe / Morocco		Spain / no Spain		France / no France		UK / no UK	
	SORS	CRS	SORS	CRS	SORS	CRS	SORS	CRS
Sensitivity	0.89	0.68	0.80	0.60	0.88	0.63	0.67	0.33
Specificity	0.97	0.55	0.76	0.51	0.93	0.64	0.83	0.17
Positive predictive value (Precision)	1.00	0.86	1.00	0.67	1.00	0.84	1.00	—
Negative predictive value	1.00	0.72	1.00	0.61	1.00	0.91	1.00	—
Efficiency (Accuracy)	0.89	0.68	0.80	0.60	0.88	0.63	0.67	0.33
AUC (Correctly classified rate)	0.93	0.62	0.78	0.56	0.90	0.63	0.75	0.25
Matthews correlation coefficient	0.93	0.45	0.77	0.23	0.89	0.55	0.71	—
Kappa coefficient	0.75	0.33	0.65	0.25	0.77	0.37	0.50	0.00

"—": the parameter could not be determined because the number of samples assigned to the class was 0.

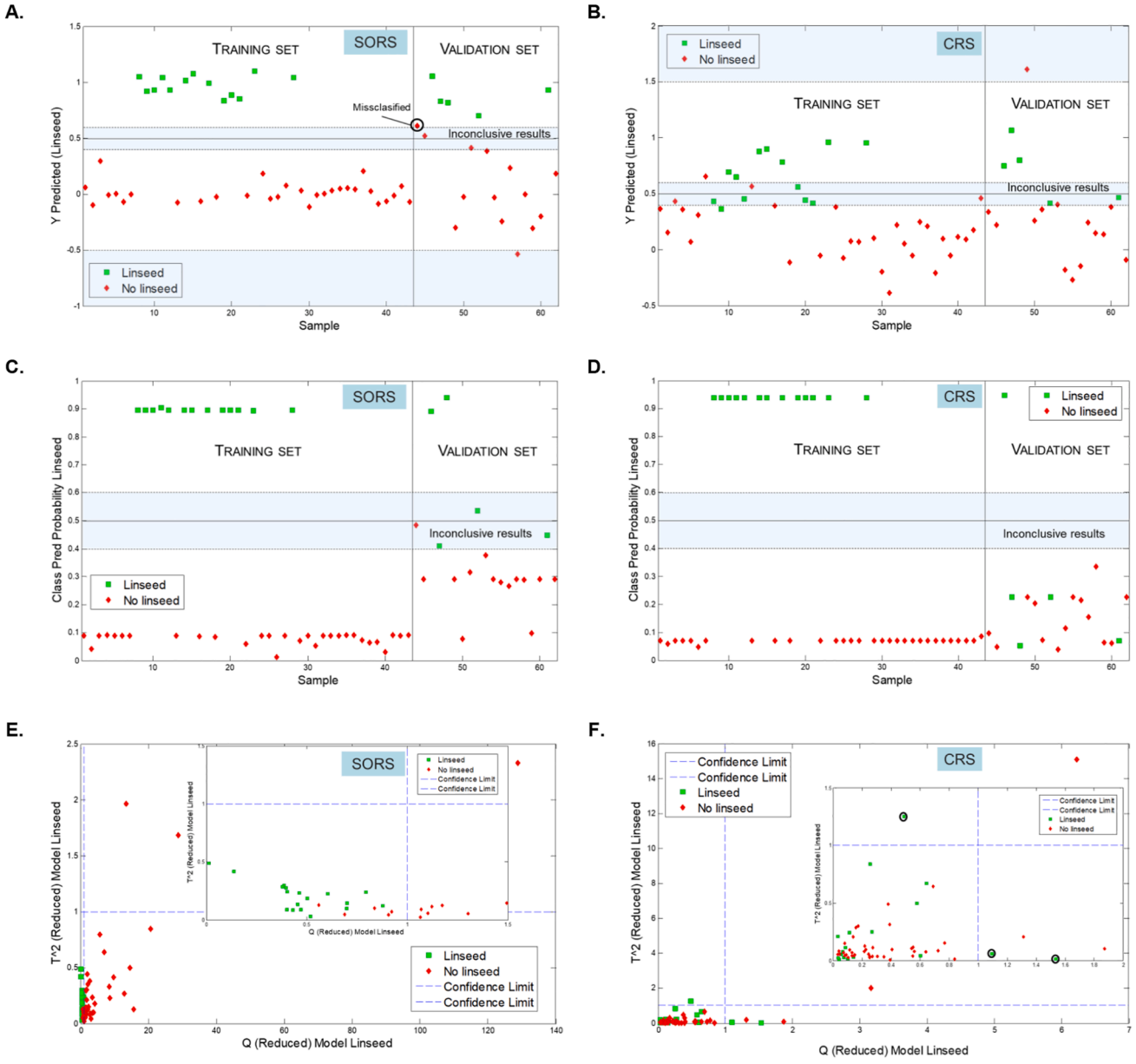

Fig. 5. Classification plots obtained for the 'linseed / no linseed' models developed by applying PLS-DA to the SORS data (A) and CRS data (B), SVM-C to the SORS data (C) and CRS data (D), and OC-SIMCA to the SORS data (E) and CRS data (F). The green squares (■) and the red rhombuses (◆) represent the margarines including and non-including linseed oil, respectively.

Table 5
Classification performance metrics for PLS-DA, SVM and OC-SIMCA 'linseed / no linseed' models developed from both types of Raman techniques.

	PLS-DA		SVM		OC-SIMCA	
	SORS	CRS	SORS	CRS	SORS	CRS
Sensitivity	0.79	0.79	0.79	0.79	0.92	0.31
Specificity	0.92	0.67	0.54	0.41	0.97	0.62
Positive predictive value (Precision)	0.96	1.00	1.00	0.84	0.94	0.48
Negative predictive value	0.88	1.00	1.00	0.94	0.85	0.36
Efficiency (Accuracy)	0.79	0.79	0.79	0.79	0.92	0.31
AUC (Correctly classified rate)	0.86	0.73	0.66	0.60	0.94	0.46
Matthews correlation coefficient	0.77	0.72	0.61	0.39	0.83	−0.11
Kappa coefficient	0.60	0.57	0.55	0.27	0.82	−0.05

(Figure S6B, Table S4) stand out.

3.4. Classification by other ingredients

In addition to the models based on the type of vegetable oil or fat included in some margarines, classification models based on other minor constituents were also developed: buttermilk, sterol esters (called phytosterols because they are of vegetable origin) and lecithin (some margarines contain soybean lecithin, others sunflower lecithin and others do not specify the type of lecithin, so they were divided into two groups according to whether or not they contain lecithin without specifying the origin). The same data pre-processing strategy as in the previous section was used. The results obtained for the 'phytosterols / no phytosterols' classificatory models are presented below, while the rest can be found in the supplementary information (Figures S7 y S8).

Fig. 6 shows the results for PLS-DA, SVM and OC-SIMCA

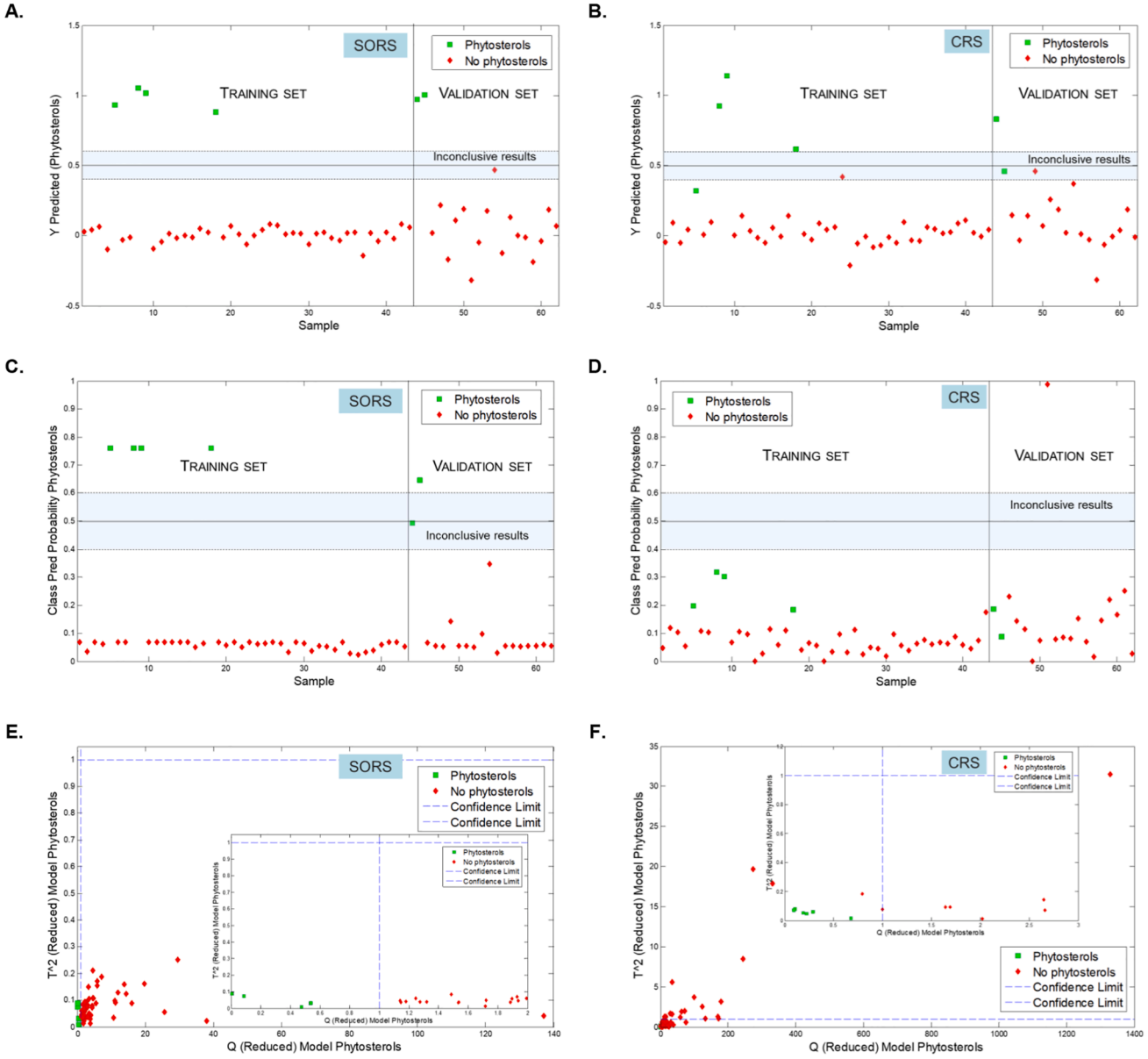

Fig. 6. Classification plots obtained for the 'phytosterols / no phytosterols' models developed by applying PLS-DA to the SORS data (A) and CRS data (B), SVM-C to the SORS data (C) and CRS data (D), and OC-SIMCA to the SORS data (E) and CRS data (F). The green squares (■) and the red rhombuses (◆) represent the margarines including and non-including phytosterols, respectively.

Table 6
Classification performance metrics for PLS-DA, SVM and OC-SIMCA 'phytosterols / no phytosterols' models developed from both types of Raman techniques.

	PLS-DA		SVM		OC-SIMCA	
	SORS	CRS	SORS	CRS	SORS	CRS
Sensitivity	0.95	0.89	0.95	0.84	1.00	0.98
Specificity	0.99	0.55	0.55	0.10	1.00	1.00
Positive predictive value (Precision)	1.00	1.00	1.00	0.80	1.00	0.99
Negative predictive value	1.00	1.00	1.00	0.09	1.00	0.87
Efficiency (Accuracy)	0.95	0.89	0.95	0.84	1.00	0.98
AUC (Correctly classified rate)	0.97	0.72	0.75	0.47	1.00	0.99
Matthews correlation coefficient	0.97	0.69	0.71	−0.08	1.00	0.92
Kappa coefficient	0.78	0.56	0.73	−0.08	1.00	0.91

'phytosterols / no phytosterols' models. PLS-DA models were built with nine and eight LVs explaining 83.09% and 99.99% of CV, and OC-SIMCA models were built with three PCs each, explaining 89.26% and 99.83% of CV, respectively for SORS and CRS data in both cases.

Despite the scarce number of samples including phytosterols in their composition (6 samples out of 62), the developed models demonstrated excellent results for classifying margarines according to this parameter. As shown in Table 6, classification performance metrics show a perfect OC-SIMCA model developed with SORS data (all parameters with a value of 1.00), the same as the OC-SIMCA model built from CRS data, for which all quality parameters are above 0.87.

When differentiating samples containing buttermilk or lecithin (classification performance metrics for these models can be found in supplementary material), the results obtained were not as good as in the case of phytosterols. This is probably due to the same reasons stated in the previous section for the 'sunflower / no sunflower' and 'palm / no palm' models. In fact, some quality parameters were not calculated like those of the 'buttermilk / no buttermilk' model developed applying SVM from SORS data (Table S5) or the 'lecithin / no lecithin' model developed with SVM (Table S6), because these models were not able to distinguish between the two classes.

3.5. Quantitation of fat content

Once the potential of the SORS technique to perform qualitative analysis was established, a quantitative analysis to determine the total fat content was performed. This quantitative method is just proposed to be applied as a pre-screening method to verify the fat content stated on the label declared by the manufacturer.

In line with the EU regulation [36,37], the samples were divided into different groups, namely: margarine (80% fat), fat spread 70%, three-quarter-fat margarine (60%), fat spread 50%, half-fat-margarine (40%) and fat spread 32%. It should be noted that the percentage is an average of the total number of samples included in that group (see Table 7). Also note that this percentage refers to the total fat content in grams per 100 g of product. The samples were then split into a training set and an external validation set, with the exception of six samples whose total fat content was not stated and which were reserved for a final step as test data in the model development.

As means of a fair comparison between both Raman techniques, two PLSR quantitation models for each dataset (SORS and CRS) were built applying both mean center and autoscale as pre-processing. The models performed from SORS data were built by choosing ten and eight LVs, explaining 89.28% and 54.83% of the CV, after using the mean center and autoscale as pre-processing, respectively. The CRS models were built with eight LVs each, explaining 99.99% and 99.85% of the CV, respectively. The internal validation for all the PLSR models was venetian blinds with 6 splits and 1 samples per split.

Fig. 7 shows the plots obtained by representing the total fat content predicted by the different models against the total fat content declared in the labelling of the samples. These plots correspond to the training set. The coefficient of determination (R^2) and the corresponding root mean square error of cross validation (RMSECV) are indicated in each figure. It is worth noting that the best results when building the models were obtained by using autoscaling as pre-processing for the data obtained by both techniques (Fig. 7B and 7D). However, when comparing the two techniques, the R^2 values from SORS data were much higher with both types of pre-processing.

The predicted values of the total fat content of the samples from the validation set obtained with the different models developed by PLSR (see Table S7) were compared with the original total fat content declared on the labelling of the samples to calculate the quantitation performance metrics shown in Table 8. Based on these results, it can be inferred that the best models were obtained using autoscale as pre-processing for the data obtained from CRS data and mean center for the SORS data. The results from the CRS showed quantitation performance metrics below 10%, except for the SDV with a value of 10.6%. However, with the spectral data obtained after measuring the samples through their original packaging, i.e. from the SORS, errors were even lower, below 5%. The R^2 confirms that the best predicting model for total fat content is the developed from SORS data using medium centering as a pre-processing.

Finally, evaluation of the models developed was carried out with six of the samples analysed that did not declare the total fat content on their labeling, so their content was unknown. Table 9 shows the results predicted by each model developed for these samples, while the obtained plots can be seen in supplementary material (Figure S9). Different results can be observed between the values obtained with the models developed from both SORS and CRS darta.

The values predicted of the total fat content from the CRS measurements are inconsistents because they are always largely higher than 100%. However, the predicted values from SORS data could potentially make sense, as they are ranging between 60 and 80%. This is the case with both models developed from SORS data, i.e., with the two pre-processing methods that were applied (mean center and autoscale).

3.6. Comparison between CRS and SORS

The results shown in the previous sections comparing the data obtained by the two techniques, both for qualitative and quantitative

Table 7
Fat content of each analysed sample and groups established to carry out the quantitation models.

Group	Fat % range	Fat % average	Number of samples	Samples in training set	Samples in validation set
Margarine*	90 – 80	80	2	2	0
Fat spread 70%	79 – 63	70	12	8	4
Half-fat margarine*	62 – 60	60	13	9	4
Fat spread 50%	59 – 42	50	19	13	6
Three-quarter-fat margarine*	41 – 39	40	6	4	2
Fat spread 32%	less than 38	32	4	3	1
Unknown	—	—	6	—	—

*Designation according to legislation.

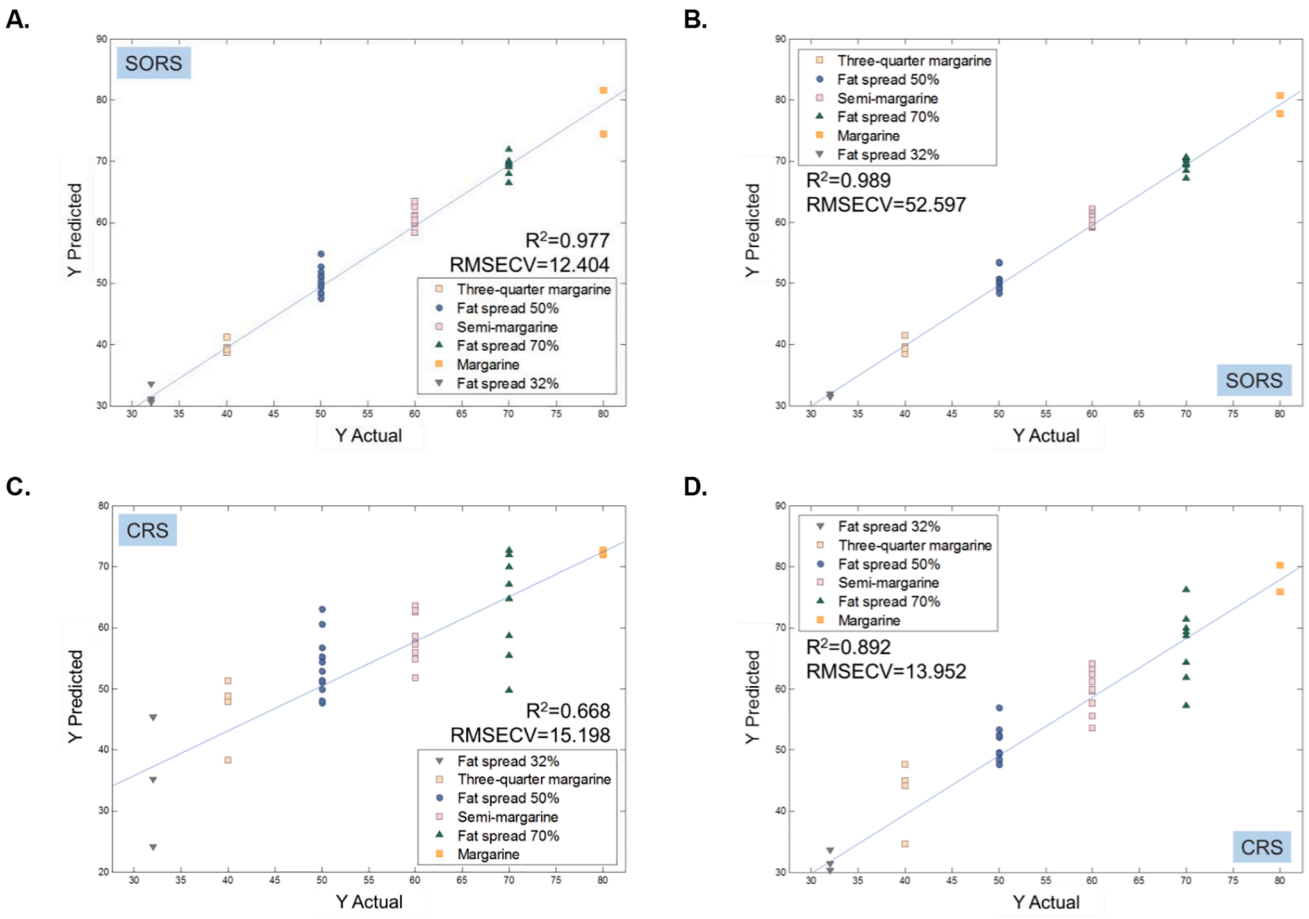

Fig. 7. Total fat content of margarine samples predicted by the PLSR model for the training set using data obtained by SORS with mean center (A) and autoscale as pre-processing (B) and CRS data with mean center (C) and autoscale as pre-processing (D) versus total fat content reported.

Table 8
Quantitation performance metrics for the predicted total fat content results of the validation set from the different PLSR models developed.

	SORS		CRS	
	Mean center	Autoscale	Mean center	Autoscale
Root Mean Square Error (RMSE, %)	5.0	10.4	12.3	10.0
Mean Absolute Error (MAE, %)	4.1	6.0	6.2	5.6
Median Absolute Error (MdAE, %)	4.0	7.6	9.8	7.3
Standard Deviation of Validation Residuals (SDV)	5.0	10.1	12.3	10.6
Coefficient of determination (R^2)	0.7	0.2	0.1	0.3

analysis of margarine and fat spreads samples, highlight the great potential of SORS to perform the authentication of these products, considering different characteristics (geographical origin, different oil types, different minor constituents and total fat content). Evidence of this is the high performance parameters calculated for the models developed.

Table 9
Predicted results of the total fat content of margarines, whose value was unknown, by the different PLSR quantitation models developed.

Sample	Predicted total fat content (g/100 g product)			
	SORS		CRS	
	Mean center	Autoscale	Mean center	Autoscale
1	67.4	64.5	152.6	122.1
2	60.0	67.9	167.0	138.8
3	79.0	81.5	155.6	130.1
4	68.3	73.8	149.3	122.6
5	66.8	72.2	134.9	122.4
6	70.2	77.8	142.1	125.4

When comparing the Raman spectra without any pre-processing method of the same sample in the same spectral range (350–2000 cm^{-1}) with both techniques, it can be seen that the chemical information provided is the same, as shown in Fig. 8 in the regions marked in green and orange. In this figure it is also observed that the influence of the fluorescence is higher in the CRS spectrum (area marked in blue), while the non-preprocessed SORS spectrum exhibits a lower influence, due to the fact that the equipment is already programmed for this purpose. In addition, the figure shows the processed spectrum obtained with the same equipment, where the influence of fluorescence is totally eliminated, obtaining a completely clean spectrum but with the same relevant chemical information. This could be the reason why the models seem to perform better from SORS data than from CRS data. Here, it is demonstrated that despite the use of a portable SORS, the spectral resolution does not appear to be lower than that obtained from a benchtop CRS.

It should therefore be noted that with the SORS technique it is

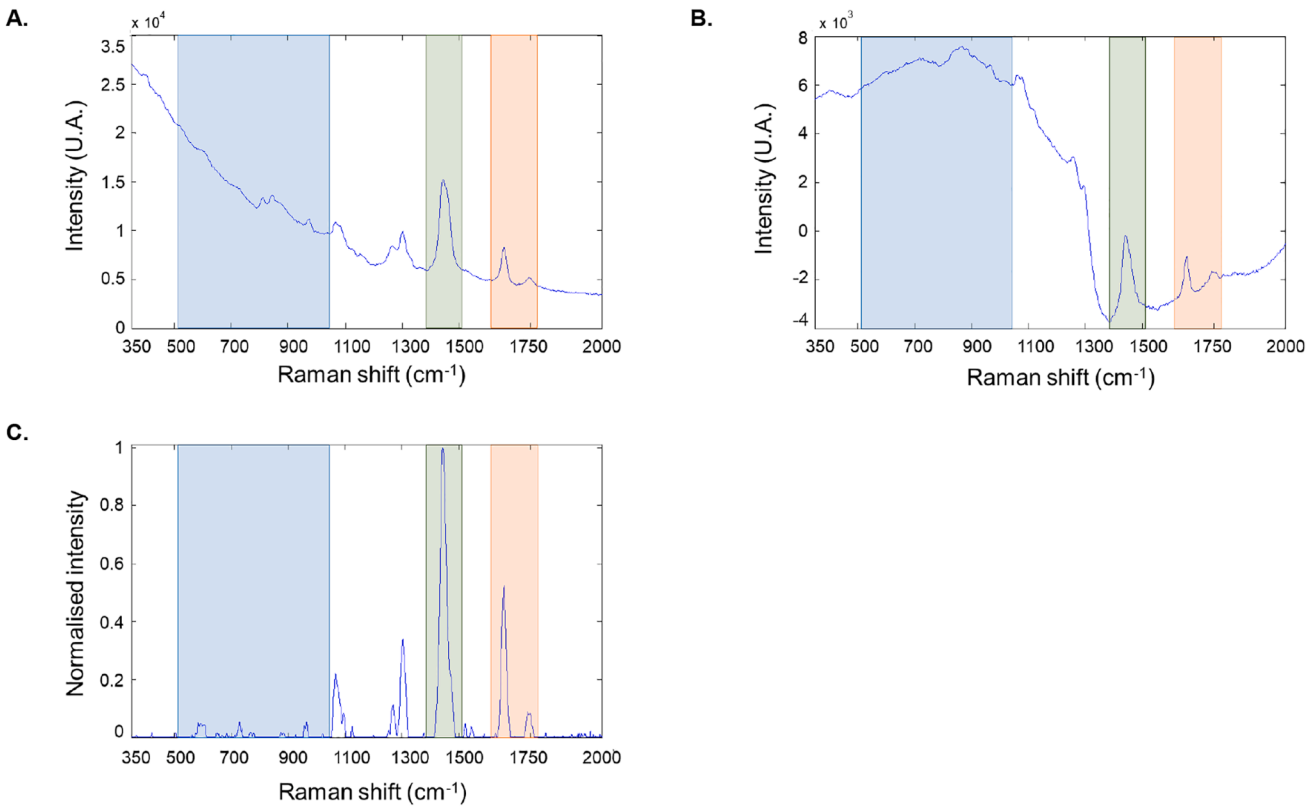

Fig. 8. Comparison of the non-preprocessed spectrum of sample 1 from SORS (A), non-preprocessed spectrum of sample 1 from CRS (B) and pre-processed spectrum of sample 1 from SORS (C). The spectrum from CRS has been cut to be in the same range as the SORS spectrum for comparison (350–2000 cm^{-1}).

possible to obtain the spectrum of the margarines by measuring through their original packaging, and also with a higher resolution, i.e. enriched structural information, of the spectrum. This can also be seen in the intensity scale, which is an order of magnitude higher in the spectrum obtained using SORS.

Loadings plots of all developed models were also examinated. Most relevant ones are included in supplementary material as examples. According to the band assignment (see Table 2) the peak around 1654 cm^{-1} stands out, which is related to oil content [25] and more specifically to sunflower oil content [32]. The information from this peak seems to be relevant for the development of the PLSDA 'sunflower / no sunflower' and PLSR models from SORS (see Figures S10-S12). Similar for CRS the peaks highlighted in the loading plot of the PLSR model are related to the presence of unsaturated and saturated fatty acids respectively [30]. However, the loading plot of the 'linseed / no linseed' model seems to be influenced by the fluorescence perturbation, which may explain why the results of this model were not so satisfactory.

In short, the chemical information provided by the data from SORS, especially related to oils and fatty acids, is responsible for developing classification and quantitation models with better results than those from CRS.

4. Conclusions

The results presented in this paper demonstrate the ability of SORS to recover the Raman spectra of margarines and fat spread products in a rapid, non-invasive and non-destructive way after measuring them through the original packaging. The chemometric tools allowed the extraction of the relevant chemical information from these spectra for qualitative and quantitative analyses. High-quality multivariate classification models were developed to distinguish samples according to their geographical origin of production and according to some of the relevant ingredients of their composition, namely linseed oil, olive oil and phytosterols. It is worth highlighting the case of both 'phytosterols / no phytosterols' and 'olive / no olive' models, which have obtained very acceptable classification results, as precision values above 0.8 in these case, and most around 1. Despite not finding significant differences between the spectra of samples containing this ingredient and those that do not, the use of chemometric tools to treat the data as an instrumental fingerprint characteristic of each product has proven to be able to extract this relevant and non-evident information. Furthermore, the quantitation model performed with PLSR allowed predicting the total fat content of the samples with errors below 5%, as well as predicting the fat content of unknown samples, with values within the limit of what can be expected. The proposed method could be therefore applied as a pre-screening method to straightforwardly verify some of the claims stated on the label declared by the manufacturer.

Comparison with CRS measurements revealed the potential of SORS to avoid the main problem associated with the use of this technique, namely the influence of fluorescence. Moreover, a lower spectral resolution was not observed compared to CRS and, in fact, the results of the classification and quantitation models were better from the SORS data. These facts are consistent with the literature on the use of the 'SORS-chemometrics' tandem and open the door to further research into the use of SORS in the area of analytical chemistry for food quality and authentication control.

CRediT authorship contribution statement

Ana M. Jimenez-Carvelo: Conceptualization, Methodology, Investigation, Resources, Data curation, Writing – review & editing. **Alejandra Arroyo-Cerezo:** Conceptualization, Methodology, Software, Investigation, Writing – original draft, Visualization. **Sanae Bikrani:** Investigation, Resources. **Wenyang Jia:** Methodology, Software, Investigation, Writing – review & editing. **Anastasios Koidis:** Conceptualization, Methodology, Investigation, Resources. **Luis Cuadros-**

Rodríguez: Conceptualization, Methodology, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

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

This work was partially supported by University of Granada (Spain) within the framemork of the funding corresponding to program 'pre-competitive research projects for young researchers'. Funding for open access charge: University of Granada / CBUA.

AMJC wish to acknowledge the Department of Economic Transformation, Industry, Knowledge and Universities belong to Regional Andalusia Government (Spain) for the Postdoctoral fellowship (DOC_00121). In addition, AAC wants to express their sincere gratitude to the Spanish Ministry of Universities for a pre-doctoral fellowship FPU (FPU20/04711, Formación del Profesorado Universitario).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.microc.2022.107378.

References

[1] S. Mosca, P. Dey, M. Salimi, B. Gardner, F. Palombo, N. Stone, P. Matousek, Spatially offset Raman spectroscopy – How deep? Anal. Chem. 93 (2021) 6755–6762, https://doi.org/10.1021/acs.analchem.1c00490.

[2] S. Mosca, C. Conti, N. Stone, P. Matousek, Spatially offset Raman spectroscopy, Nat. Rev. Meth. Primers 1 (2021) 21, https://doi.org/10.1038/s43586-021-00019-0.

[3] P. Matousek, I.P. Clark, E.R.C. Draper, M.D. Morris, A.E. Goodship, N. Everall, M. Towrie, W.F. Finney, A.W. Parker, Subsurface probing in diffusely scattering media using spatially offset Raman spectroscopy, Appl. Spectrosc. 59 (2005) 393–400, https://doi.org/10.1366/0003702053641450.

[4] F. Martelli, T. Binzoni, A. Pifferi, L. Spinelli, A. Farina, A. Torricelli, There's plenty of light at the bottom: statistics of photon penetration depth in random media, Sci. Rep. 6 (2016) 27057, https://doi.org/10.1038/srep27057.

[5] A. Arroyo-Cerezo, A.M. Jiménez-Carvelo, A. González-Casado, A. Koidis, L. Cuadros-Rodríguez, Deep (offset) non-invasive Raman spectroscopy for the evaluation of food and beverages – A review, LWT - Food Sci. Technol. 149 (2021), 111822, https://doi.org/10.1016/j.lwt.2021.111822.

[6] Y. Xu, P. Zhong, A. Jiang, X. Shen, X. Li, Z. Xu, Y. Shen, Y. Sun, H. Lei, Raman spectroscopy coupled with chemometrics for food authentication: A review, Trends Analyt. Chem. 131 (2020), 116017, https://doi.org/10.1016/j.trac.2020.116017.

[7] A.I. Ropodi, E.Z. Panagou, G.-J.-E. Nychas, Data mining derived from food analyses using non-invasive/non-destructive analytical techniques; determination of food authenticity quality & safety in tandem with computer science disciplines, Trends Food Sci. Technol. 50 (2016) 11–25, https://doi.org/10.1016/j.tifs.2016.01.011.

[8] C. Berghian-Grosan, D.A. Magdas, Raman spectroscopy and machine-learning for edible oils evaluation, Talanta 218 (2020) 121176. DOI: 1016/j.talanta.2020.121176.

[9] A.M. Jiménez-Carvelo, A. González-Casado, M.G. Bagur-González, L. Cuadros-Rodríguez, Alternative data mining/machine learning methods for the analytical evaluation of food quality and authenticity – A review, Food Res. Int. 122 (2019) 25–39, https://doi.org/10.1016/j.foodres.2019.03.063.

[10] C. Conti, C. Colombo, M. Realini, P. Matousek, Subsurface analysis of painted sculptures and plasters using micrometre-scale spatially offset Raman spectroscopy (micro-SORS), J. Raman Spectrosc. 46 (2015) 476–482, https://doi.org/10.1002/jrs.4673.

[11] C. Eliasson, N.A. Macleod, P. Matousek, Noninvasive Detection of cocaine dissolved in beverages using displaced Raman spectroscopy, Anal. Chim. Acta 607 (2008) 50–53, https://doi.org/10.1016/j.aca.2007.11.023.

[12] M.Z. Vardaki, C.G. Atkins, H.G. Schulze, D.V. Devine, K. Serrano, M.W. Blades, R.F. B. Turner, Raman spectroscopy of stored red blood cell concentrate within sealed transfusion blood bags, Analyst 143 (24) (2018) 6006–6013.

[13] M. Arellano, I.T. Norton, P. Smith, in: Specialty Oils and Fats in Food and Nutrition, Elsevier, 2015, pp. 241–270.

[14] N.W.G. Young, P. Wassell, Margarines and spreads, in: G.L. Hasenhuettl, R.W. Hartel (Eds.), Food Emulsifiers and their Applications, third ed., Springer Nature, Cham, 2019, ch. 13, pp. 379–405. DOI: 10.1007/978-3-030-29187-7_13.

[15] Codex Stan 256-2007, Standard for fat spreads and blended spreads, Codex Alimentarius FAO/WHO, 2007.

[16] Regulation (EU) No 1308/2013 establishing a common organization of the markets in agricultural products, OJ02013R1308-EN-003.001-223 (consolidated version 01.08.2017), European Commission, Brussels, 2017.

[17] Decree No 1153-66 on the regulations implementing the Dahir of 14 October 1914 on the suppresion of fraud in the manfacture and sale of margarine, Official Gazette No 2988, p. 214, Rabat, Morocco, 1970.

[18] M.D. Guillén, M.L. Ibargoitia, P. Sopelana, Margarines: composition and analysis, in: B. Caballero, P.M. Finglas, F. Toldrá (Eds.), Encyclopedia of Food and Health, vol. III, Academic Press / Elsevier, Oxford, 2016, pp. 646–653. DOI: 10.1016/B978-0-12-384947-2.00446-3.

[19] A. Rácz, M. Fodor, K. Héberger, Development and comparison of regression models for the determination of quality parameters in margarine spread samples using NIR spectroscopy, Anal. Methods 10 (25) (2018) 3089–3099.

[20] S. Lohumi, H. Lee, M.S. Kim, J. Qin, B.K. Cho, Through-packaging análisis of butter adulteration using line-scan spatially offset Raman spectroscopy, Anal. Bioanal. Chem. 410 (2018) 5663–5673, https://doi.org/10.1007/s00216-018-1189-1.

[21] L. Cuadros-Rodríguez, E. Pérez-Castaño, C. Ruiz-Samblás, Quality performance metrics in multivariate classification methods for qualitative analysis, Trends Anal. Chem. 80 (2016) 612–624, https://doi.org/10.1016/j.trac.2016.04.021.

[22] ASTM International E2617-17. Standard practice for validation of empirically derived multivariate calibrations, 2017.

[23] V. Baeten, P. Hourant, M.T. Morales, R. Aparicio, Oil and Fat Classification by FT-Raman Spectroscopy, J. Agric. Food Chem. 46 (1998) 2638–2646, https://doi.org/10.1021/jf9707851.

[24] L. Dymińska, M. Calik, A.M.M. Albegar, A. Zając, K. Kostyń, J. Lorenc, J. Hanuza, Quantitative determination of the iodine values of unsaturated plant oils using infrared and Raman spectroscopy methods, Int. J. Food Prop. 20 (2017) 2003–2015, https://doi.org/10.1080/10942912.2016.1230744.

[25] A. Nedeljković, P. Rösch, J. Popp, J. Miočinović, M. Radovanović, P. Pudja, Raman spectroscopy as a rapid tool for quantitative analysis of butter adulterated with margarine, Food Anal. Methods 9 (2016) 1315–1320, https://doi.org/10.1007/s12161-015-0317-1.

[26] P. Bock, N. Gierlinger, Infrared and Raman spectra of lignin substructures: Coniferyl alcohol, abietin, and coniferyl aldehyde, J. Raman Spectrosc. 50 (2019) 778–792, https://doi.org/10.1002/jrs.5588.

[27] G. Yang, Q. Wang, C. Liu, X. Wang, S. Fan, W. Huang, Rapid and visual detection of the main chemical compositions in maize seeds based on Raman hyperspectral imaging, Spectrochim. Acta A Mol. Biomol. Spectrosc. 200 (2018) 186–194, https://doi.org/10.1016/j.saa.2018.04.026.

[28] F. Adar, Introduction to Interpretation of Raman Spectra Using Database Searching and Functional Group Detection and Identification, in: Spectroscopy Solutions for Materials Analysis. Molecular Spectroscopy Workbench: The 2016 Collection, 31 (2016) pp. 18-23.

[29] I.H. Boyaci, H.T. Temiz, H.E. Geniş, E.A. Soykut, N.N. Yazgan, B. Güven, R. S. Uysal, A.G. Bozkurt, K. İlaslan, O. Toruna, F.C.D. Şeker, Dispersive and FT-Raman spectroscopic methods in food analysis, RSC Adv 5 (2016) 56606–56624, https://doi.org/10.1039/c4ra12463d.

[30] J.M. Benevides, S.A. Overman, G.J. Thomas, Raman, polarized Raman and ultraviolet resonance Raman spectroscopy of nucleic acids and their complexes, J. Raman Spectrosc. 36 (2005) 279–299, https://doi.org/10.1002/jrs.1324.

[31] S. Medina, R. Perestrelo, P. Silva, J.A.M. Pereira, J.S. Câmara, Current trends and recent advances on food authenticity technologies and chemometric approaches, Trends Food Sci. Technol. 85 (2019) 163–176, https://doi.org/10.1016/j.tifs.2019.01.017.

[32] M. Varnasseri, H. Muhamadali, Y. Xu, P.I.C. Richardson, N. Byrd, D.I. Ellis, P. Matousek, R. Goodacre, Portable through Bottle SORS for the Authentication of Extra Virgin Olive Oil, Appl. Sci. 11 (2021) 8347, https://doi.org/10.3390/app11188347.