Journal of Chromatography B, 910 (2012) 71–77

Contents lists available at SciVerse ScienceDirect

Journal of Chromatography B

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

Quantification of blending of olive oils and edible vegetable oils by triacylglycerol fingerprint gas chromatography and chemometric tools☆

Cristina Ruiz-Samblás[a,*], Federico Marini[b], Luis Cuadros-Rodríguez[a], Antonio González-Casado[a]

[a] Department of Analytical Chemistry, University of Granada, c/Fuentenueva s.n., E-18071 Granada, Spain
[b] Department of Chemistry, University of Rome "La Sapienza", P. le A. Moro 5, I-00185 Rome, Italy

ARTICLE INFO

Article history:
Received 27 October 2011
Accepted 23 January 2012
Available online 9 February 2012

Keywords:
Olive oil
Vegetable oil
Blends
Genetic algorithm
PLS
GC–MS

ABSTRACT

A reliable procedure for the identification and quantification of the adulteration of olive oils in terms of blending with other vegetable oils (sunflower, corn, seeds, sesame and soya) has been developed. From the analytical viewpoint, the whole procedure relies only on the results of the determination of the triacylglycerol profile of the oils by high temperature gas chromatography–mass spectrometry. The chromatographic profiles were pre-treated (baseline correction, peak alignment using iCoshift algorithm and mean centering) before building the models. At first, a class-modeling approach, Soft Independent Modeling of Class Analogy (SIMCA) was used to identify the vegetable oil used blending. Successively, a separate calibration model for each kind of blending was built using Partial Least Square (PLS). The correlation coefficients of actual versus predicted concentrations resulting from multivariate calibration models were between 0.95 and 0.99. In addition, Genetic algorithms (GA–PLS), were used, as variable selection method, to improve the models which yielded R^2 values higher than 0.90 for calibration set. This model had a better predictive ability than the PLS without feature selection. The results obtained showed the potential of this method and allowed quantification of blends of olive oil in the vegetable oils tested containing at least 10% of olive oil.

1. Introduction

Edible vegetable oils are a valuable component of a fully mature seed. They are mainly the mixtures of triacylglycerols (TAGs), with different concentration levels. The remaining nonglyceridic fraction consists of different compound classes such as hydrocarbons, tocopherols, phytosterols and sterol esters [1]. The vegetable seed or fruit from which the oil is extracted determine most of its characteristics and composition that also depends on several factors such as soil, climate, processing, harvesting and chemical process occurring during storage [2].

Among edible oils, olive oil (OO) shows important and outstanding characteristics due to its differentiated sensorial qualities (taste and flavor) and higher nutritional value which have been acknowledged internationally. Several health benefits associated with its consumption were initially observed among Mediterranean people and its dietary consumption is nowadays considered to provide many benefits to human health [3].

TAGs represent up to 95–98% (weight to weight – w/w) of vegetable oil composition and show a characteristic distribution. As a consequence, the addition of other edible vegetable oils to olive oils modifies TAG distribution and because of that, they are considered to be good fingerprints for adulteration detection purposes [4].

Companies have been taking advantage of selling OO blends at the same price as pure OO, obtaining important economic benefits. The adulterants used in blends are the ones with similar physical and chemical properties and usually they are cheaper and easy to obtain. In the case of OO this usually implies the dilution with less expensive oils or other inferior quality olive oils [5,6]. Moreover, a lot of methods and limits were introduced into the International Olive Oil Council (IOOC) trade standard, into EC Regulation 2568/91 and into the Codex Alimentarius Standard for controlling product authenticity and quality. In addition, In the EU, requirements has being established in Regulation (EC) No. 29/2012, concerning commercialization and labeling of products which contain olive oil, blends of olive and other edible vegetable oils. The presence of olive oil higher than 50% has to be indicated on the label, but if the percentage is lower than 50% the name of olive oil cannot be used in the label [7].

However, the reasons for mixing olive oil with others are not only economical, but also nutritional. It is clear from the composition of vegetable oils, that no single oil, even olive oil, meet all the

☆ This paper belongs to the Special Issue Chemometrics in Chromatography, Edited by Pedro Araujo and Bjørn Grung.

* Corresponding author. Tel.: +34 958240797; fax: +34 958243328.
E-mail address: crsamblas@ugr.es (C. Ruiz-Samblás).

1570-0232/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2012.01.026

oil nutritional requirements of essential fatty acids and vitamins [8].

The interest of researchers in the authentication of vegetable oils has led to an improvement in the control of adulteration and to the development of analytical methods to establish compositional differences in olive oils blends [9]. An extensive literature, discussing the suitability of a wide assortment of analytical methods aimed at evaluating the authenticity and the presence of adulterants in OO, has been published [10,11]. When such methods are applied in conjunction of chemometric tools, spectroscopic analytical techniques as NIR, MIR, Raman, NMR or MS, and sensor-based analytical techniques as electronic nose, have been frequently used. This techniques share as common feature that yield low-selective instrumental signals (instrumental fingerprints) which are very suitable for developing valid chemometric models for pattern recognition. However, there are not many studies about vegetable oils authentication, which use directly the raw analytical signal which come from the chromatographic instrument (chromatographic fingerprint) with multivariate statistical methods [12,13]. In most cases, the chromatographic applications use derived information from the raw analytical signal provided by the instrument, such as peak areas or concentration profiles for the classification of edible vegetable oils and detection of their adulterations. Chemometric tools have been commonly applied for matching and discrimination, classification and prediction in assessing authenticity of vegetable oils [14–16]. Thus, GC and LC methods in combination with multivariate statistical techniques such as principal component analysis (PCA), discriminant analysis (DA), cluster analysis (CA), K-nearest neighbor, genetic algorithm (GA), partial least squares (PLS) [17] and artificial neural networks (ANN) have been applied successfully to classify and discriminate the oils [18,19].

The official method of the International Olive Council (IOC) is based on the use of the reverse phase-liquid chromatography with a refractive index detector (HPLC-RID), to establish the difference between actual and theoretical content of TAGs with Equivalent Carbon Number 42 (ECN42) [20]. In the literature, the contribution in chromatography to the authentication of vegetable oils by quantifying different major and minor compounds have been reviewed by Aparicio and Aparicio-Ruiz [21]. Detection of adulterated oils based on TAG compositions by high temperature-gas chromatography (HT)-GC was studied previously by Park and Lee [1].

This study focuses on the quantification of olive oil in blends with vegetable oils using multivariate calibration. The TAGs profiles, measured by HTGC–MS systems, have been applied for the quantification of olive oils in blends with vegetable oils (sunflower, corn, seeds, sesame and soya) and considering the different categories of olive oil (extra virgin olive oil, virgin, olive oil and pomace) and varieties (picual, hojiblanca and arbequina) at several percentages (10–90%). Multivariate statistical analyses, such as SIMCA, PLS and GA–PLS were applied to achieve this purpose.

2. Materials and methods

2.1. Samples

The olive oil samples, to build the blends, were fourteen, including four categories [22]: extra virgin (EVOO), virgin (VOO), olive oil (OO, blend of virgin and refined) and pomace oil (POO), and three Spanish olive fruit varieties named “arbequina” (ARB), “hojiblanca” (HOJ) and “picual” (PIC).

In addition, eleven vegetable oils samples were used: two sunflowers oils (SUN), one high-oleic sunflower oil (OSUN), two corn oils (COR), one sesame oil (SES), three soya oils (SOY), and two vegetable seeds oils (SEE), where the label did not specified what kind of seeds were used.

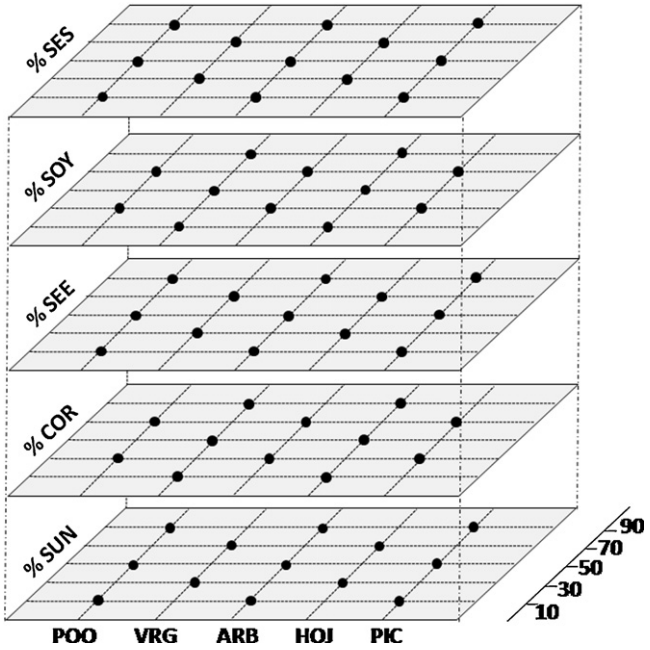

Fig. 1. Calibration Building Blocks with the percentage of each kind of olive oil and each vegetable oil in the blends for the training set (Note: acronym VRG includes the three categories, extra virgin (EVOO), virgin (VOO) and olive oil (OO), which is a blend of virgin and refined olive oil.).

All the vegetable oils samples were purchased in Spain and France and were stored in dark bottles, at −4 °C until their analysis.

A pure sample of each edible vegetable oil was also analyzed. Seventy-eight blend samples were prepared by mixing one olive oil with one vegetable oil in different percentages obtaining five different levels of concentration from 10 to 90% (w/w). The prepared blends were used as calibration or prediction samples, as needed and they covered all the possible combinations of vegetable oils with the different categories and varieties of olive oils The *Calibration Building Block*, for the training set of samples, is shown in Fig. 1. This methodology of preparation was followed in order to cover all the possibilities of blending and to avoid any lack of information. Table 1 shows, more specifically, the composition of each blend which were used as calibration set.

For analysis, olive oil was dissolved in chloroform (99%, reagent grade) to a final concentration of 0.2% (w/w). The diluted olive oil sample was directly injected into the system, without any preliminary chemical derivatization or purification step prior to chromatographic analysis.

2.2. Chromatographic conditions

All separations were performed with a VARIAN GC 3800 gas chromatograph (PA, USA) equipped with a split/splitless injector coupled to a VARIAN 4000 ion trap mass spectrometer (PA, USA) equipped with an electron impact source. A split injection with a ratio of 1:50 was used. The samples were introduced using a robotized autosampler module (Combipal, CTC ANALYTICS, Switzerland). Scan control, data acquisition, and processing were performed by a MS Workstation software (VARIAN, PA, USA) data system. The sample volume injected was 2 μL.

A capillary column coated with 65% diphenyl-35% dimethylpolysiloxane stationary phase (Restek Rtx-65TG; 30 m × 0.32 mm i.d. × 0.1 μm; maximum temperature 370 °C; Restek Corp., Bellefonte, PA, USA) was used. The GC oven

Table 1
Percentage of the olive oil and vegetable oil in the oil blend samples for the calibration set.

Sample	Composition	Sample	Composition
1	POO 10% + SUN 90%	32	ARB 90% + SEE 10%
2	POO 50% + SUN 50%	33	HOJ 30% + SEE 70%
3	POO 90% + SUN 10%	34	HOJ 70% + SEE 30%
4	EVOO 30% + SUN 70%	35	PIC 10% + SEE 90%
5	ARB 50% + SUN 50%	36	PIC 50% + SEE 50%
6	ARB 90% + SUN 10%	37	PIC 90% + SEE 10%
7	HOJ 30% + SUN 70%	38	POO 10% + SES 90%
8	HOJ 70% + SUN 30%	39	POO 50% + SES 50%
9	PIC 10% + SUN 90%	40	POO 90% + SES 10%
10	PIC 50% + SUN 50%	41	EVOO 30% + SES 70%
11	PIC 90% + SUN 10%	42	VOO 70% + SES 30%
12	EVOO 70% + SUN 30%	43	ARB 10% + SES 90%
13	POO 30% + COR 70%	44	ARB 50% + SES 50%
14	POO 70% + COR 30%	45	ARB 90% + SES 10%
15	EVOO 10% + COR 90%	46	HOJ 30% + SES 70%
16	VOO 50% + COR 50%	47	HOJ 70% + SES 30%
17	EVOO 90% + COR 10%	48	PIC 10% + SES 90%
18	ARB 30% + COR 70%	49	PIC 50% + SES 50%
19	ARB 70% + COR 30%	50	PIC 90% + SES 10%
20	HOJ 10% + COR 90%	51	POO 30% + SOY 70%
21	HOJ 50% + COR 50%	52	POO 70% + SOY 30%
22	HOJ 90% + COR 10%	53	EVOO 10% + SOY 90%
23	PIC 30% + COR 70%	54	VOO 50% + SOY 50%
24	PIC 70% + COR 30%	55	EVOO 90% + SOY 10%
25	POO 10% + SEE 90%	56	ARB 30% + SOY 70%
26	POO 50% + SEE 50%	57	ARB 70% + SOY 30%
27	POO90% + SEE 10%	58	HOJ 10% + SOY 90%
28	VOO 30% + SEE 70%	59	HOJ 50% + SOY 50%
29	EVOO 70% + SEE 30%	60	HOJ 90% + SOY 10%
30	ARB 10% + SEE 90%	61	PIC 30% + SOY 70%
31	ARB 50% + SEE 50%	62	PIC 70% + SOY 30%

temperature was programmed from 315 to 350 °C at 1 °C/min. The injection port was held isothermally at 370 °C. Helium (99.995%) was used as the carrier gas and its flow rate was 1.5 mL/min. The mass spectrometric conditions were as follows. The ion source temperature was held at 250 °C during the GC/MS runs [23].

The transfer-line temperature was maintained at 350 °C throughout the analyses. The electron energy was 70 eV and the emission current 10 μA. Chromatograms were recorded in full-scan mode. Average spectra were acquired in the m/z range of 200–1000 m/z and were recorded at a scan speed of 1.20 s. Scan control, data acquisition, and processing were performed by a MS Workstation software (VARIAN, PA, USA) data system.

2.3. Chemometrics

The analytical data were arranged in two matrixes to perform the statistical analysis. The calibration data set (62 samples) was made of as many rows as samples analysed and as many columns (1724 elements) as the entire chromatogram data points recorded during the acquisition time. Obviously, the signal maxima are the heights from the different chromatographic peaks. The validation set was composed of 16 samples.

All chemometric treatments were performed by using the PLS Toolbox (Eigenvector Research Inc., Wenatchee, WA), for Matlab® software (Mathworks Inc., Natick, MA, USA).

2.3.1. Pre-processing of the data

Preprocessing is, in general, the ensemble of mathematical treatments performed on the data before the model building step. Aim of preprocessing is to make the data suitable for statistical analysis, by removing the extraneous sources of variation (variance) which are not related to the information sought. These sources of variance can increase the difficulty in modeling. Interfering variance appears in almost all real data because of systematic errors present in experiments. This unwanted variance will require the model to work harder to isolate the variance of interest from the interfering one. Because of this the whole data set of samples was preprocessed before building any model [24].

First of all, a baseline correction algorithm (Penalized Asymmetric Least Squares) was applied for automatically removing baseline contributions from the data. It subtracts a baseline from a chromatogram (or a signal in general) using an iterative asymmetric least squares procedure. Points with residuals <0 are up-weighted at each iteration of the least squares fitting. This results in a robust "non-negative" residual fit when residuals of significant amplitude (e.g. signals on a background) are present [25]. Secondly, after baseline correction, peak shifting was corrected with interval correlation optimized shifting, iCoshift [26], which splits the chromatogram into intervals and "coshifts" each vector to get the maximum correlation toward a target signal in that interval (when no reference chromatogram is available, average or maximum signals can be used).

Finally, the chromatograms were mean centered by subtracting from each signal the mean chromatogram, in order to remove the variability related to this overall offset term.

2.3.2. Soft Independent Modeling of Class Analogies (SIMCA)

Firstly, a class-modeling approach using SIMCA was applied to identify the nature of blending. In class-modeling, a separate model space is built for each of the investigated categories: samples falling within the model space are accepted by that category, while samples falling outside are considered as outliers for the specific class [27]. In particular, SIMCA describes the similarities among the samples of a category using a principal component approach, so that the distance from the class models is a combination of the distance within the PC space (T^2 or leverage) and the residuals (Q). Accordingly, class space is defined by imposing a threshold to this distance value: samples for which the distance to the model is less than the threshold are accepted, while others are rejected as outliers. The optimal number of PC for each category has to be optimized during the calibration phase. In this study, a separate model for each category of blend (different categories and varieties of olive oils with SUN, COR, SEE, SES and SOY) was built.

2.3.3. Partial Least Square (PLS)

Once the nature of blending was established by SIMCA classification, quantification of the percentage of OO was carried out by building separate calibration models for each kind of blending. PLS was used to build the regression models relating the chromatographic profiles to the quantity of OO. PLS is a well-known bilinear multivariate method for building regression models when the number of variables exceeds the number of samples and/or when the variables themselves are highly correlated. PLS modeling includes the dependent and independent variables in the data compression and decomposition operations, i.e. both y and x data are actively used in the data analysis. In particular, data points are projected onto a subspace whose axes have maximum covariance with the dependent vector. The optimum dimensionality of this subspace, i.e. the optimum number of latent vectors used to describe the data should be selected during the training phase: choice of the proper dimensionality is crucial as including too few components can result in poor modeling ability while overfitting can occur when too many latent vectors are retained. Because the model has to be used for the prediction of new samples, all possible sources of variation that can be encountered later should be included in the training set. Usually, a second independent set of samples (validation or test set) has to be used to check the generalization ability of the optimized model and to compute an error score for prediction.

In order to predict the percentage of the different OOs in the blends with other vegetable oils, a specific PLS model was prepared

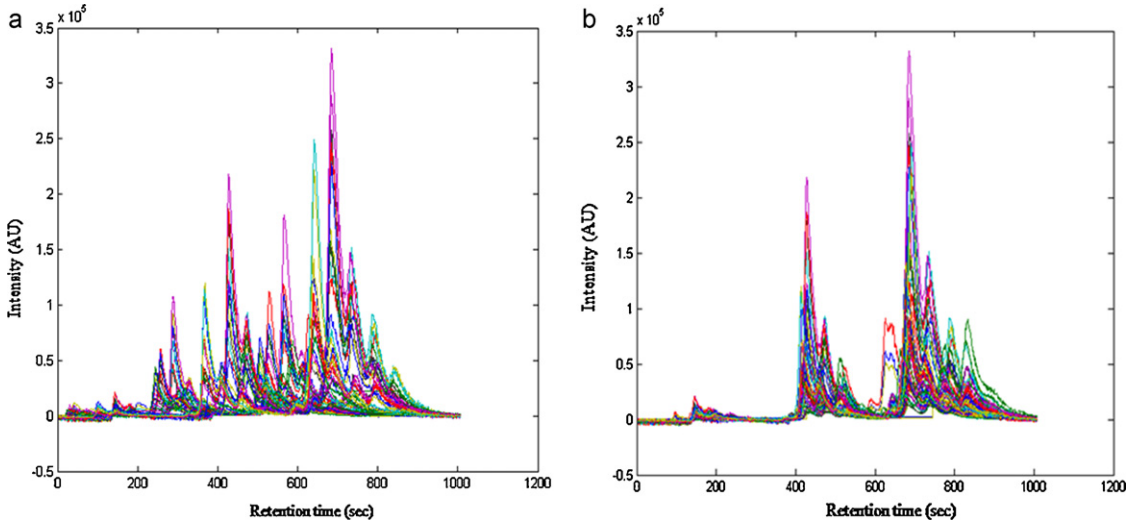

Fig. 2. GC–MS chromatograms of vegetable oil blends samples (all type of oils analyzed in the study) before peak shifting pretreatment (iCoshift) (a) and after (b).

for each class of vegetable oil. Due to the limited amount of training samples, the optimal complexity of the PLS models was chosen on the basis of the minimum error in leave-one-out cross validation, a procedure where each sample is in turn removed from the calibration set, the model is constructed using the remaining $n-1$ samples and used to predict the y value for the sample left out. This process is repeated n times, until each one of n calibration samples has been left out once. In a successive stage, a PLS model for all the blends together was built for the sake of comparison. In this case, selection of the optimal complexity was made based on a seven-fold cross-validation procedure: training data were divided into 7 cancellation groups and in turn each of them was used as internal validation set. All models were also externally validated with a set of 16 samples which were not used to build the calibration model.

2.3.4. Variable selection: genetic algorithm (GA)

The selection of variables for multivariate calibration can be considered an optimization problem. GA is an optimization method based on the principles of genetics and natural selection in the theory of evolution. The algorithm starts with a randomly selected population. Each individual of the population, represented by a chromosome of binary values, represents a subset of descriptors. The number of the genes at each chromosome is equal to the number of the descriptors. A gene is given the value of one, if its corresponding descriptor is included in the subset; otherwise, it is given the value of zero. Each chromosome is evaluated for its performance through an objective function called fitness function. A high fitness value of a chromosome corresponds to a higher chance to be selected for the next generation. Then the genetic information is exchanged between chromosomes by crossover and perturbed by mutation. The result is a new generation with better fitness. This process is repeated until the stopping criterion is reached [28].

GA–PLS is a sophisticated hybrid approach that combines GA as a powerful multivariate variable selection with PLS as calibration method [29]. In GA–PLS, the chromosome is corresponding to a set of variables, to be used as inputs for the PLS regression model. The

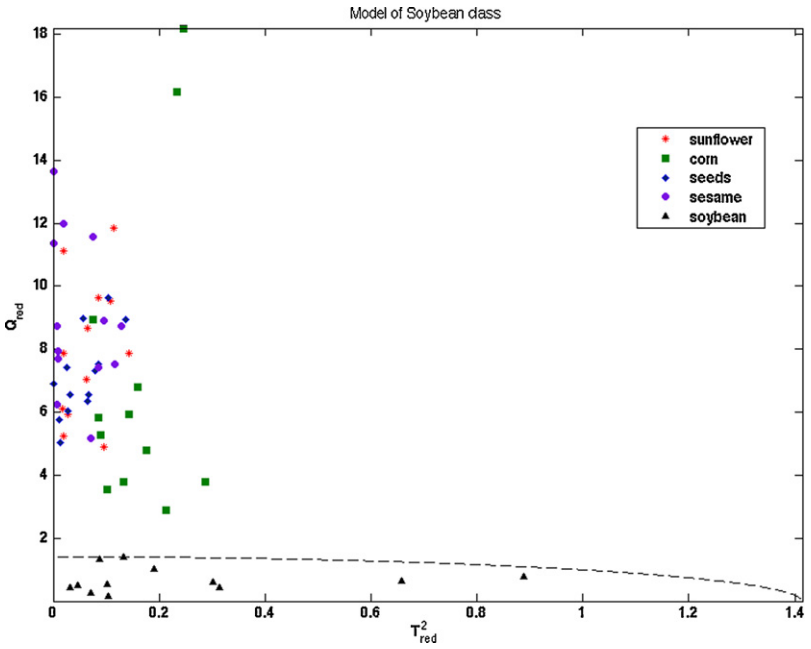

Fig. 3. SIMCA modeling. Projection of the samples (cross-validation results) onto the T^2/Q model space of soybean category.

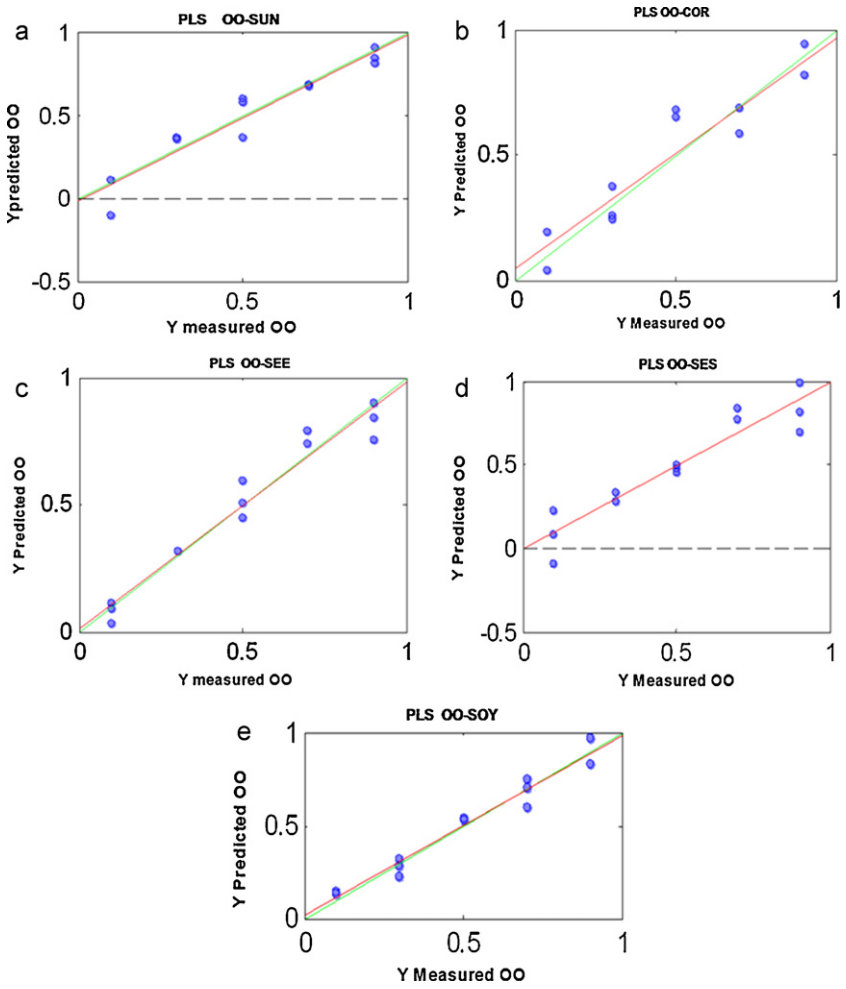

Fig. 4. Predicted percentage of olive oil in the different vegetable oil blends (a) olive oil–sunflower olive oil, (b) olive oil–corn oil, (c) olive oil–seed oil, (d) olive oil–sesame oil and (e) olive oil–soya oil. See Fig. 1 for more detail of composition. See Table 1 for the calibration parameters of the models.

coupling of GA with PLS has been successfully applied to many spectral data sets and have been shown to provide better results than full-spectrum approaches [30]. GA–PLS analysis has been carried out using Leardi's Genetic Algorithm Toolbox freely available on Internet [31]. The GA–PLS approach was adopted both for the vegetable oil-specific calibration and for building a regression model for all the blends together.

3. Results and discussion

Fig. 2 shows the baseline corrected chromatograms of the whole data set before and after iCoshift. By inspecting the figure, one can see the large difference between the two plots and how the peak shifting was perfectly corrected with this algorithm. In a first stage, iCoshift was applied separately to the different kinds of blends and successively to the whole data set with all the samples of blends together.

The first 500 data points of each chromatogram were eliminated due to the lack of information. As described in the previous section, identification and quantification of the oil blends was carried out in two successive stages: at first, a classification approach was used to identify the vegetable oil used for blending and then separate calibration models were built for each of the different blends. In particular, classification of the oil samples according to vegetable oil used for blending was carried out using a modeling approach based on the SIMCA algorithm. Accordingly, 5 independent models, one for each of the investigated categories (OO blending with SUN, COR, SEE, SES and SOY) were built on mean centered data. Leave-one-out cross-validation was used to assess the model complexity and 2 principal components were selected for all the category models, corresponding to an explained variance of more than 94.5% for each class.

It must be stressed that none of the training samples was identified as outlying for all the category models. Moreover, all samples are nearest to their respective class, so there was no misclassification either in modeling or in cross-validation. When considering the other figures of merit, it was observed that sensitivity (percentage of samples from the modeled category that are accepted by the class model) and the specificity (the percentage of samples from other categories which are rejected by the class model) [27], were both 100% for all the five classes. These results can be observed in Fig. 3, where the projection of the samples onto the model of soybean category (chosen as example) is reported. It can be observed that all the samples belonging to the modeled class (in this case, soybean) lie within the model space delimited by the threshold, while all the samples from other origins are recognized as outlier by the category model, so that sensitivity and specificity are 100%, as already mentioned.

The next step, in order to quantify the percentage of OO in the blends, was to build separate multivariate calibration models by the PLS algorithm, using the chromatographic data of the training set in order to observe how each class behaved. Because of this, a specific PLS model was prepared for each kind of blend of vegetable oil (olive oil–sunflower olive oil, olive oil–corn oil, olive oil–seed

Table 2
Calibration parameters and statistical data of the PLS calibration models for the individual classes.

	LVs	R^2	RMSEC	RMSECV	CV Bias
OO–SUN	6	0.999478	0.00641991	0.0895547	−0.0149237
OO–COR	2	0.955166	0.0573394	0.0960388	0.00643143
OO–SEE	6	0.999075	0.00892755	0.0625305	−0.00335554
OO–SES	6	0.998651	0.0107803	0.103153	−0.00644528
OO–SOY	2	0.980184	0.0381209	0.0529116	0.00185313

oil, olive oil–sesame oil and olive oil–soya oil). The performance of models was evaluated in terms of the root mean squared error of calibration/prediction (RMSEC/P) and of the determination coefficient R^2. Leave-one-out cross validation was used to choose the optimal model complexity.

Fig. 4 shows the plot of predicted versus actual concentrations of olive oil in the vegetable oil blends. Mean centering chromatographic pretreatment was used in all the calibration models. These models were built previously for each vegetable oil blend in order to see the behavior of each model by itself. Table 2 shows the statistical results obtained for the calibration step, including statistical parameters such us PLS factors, RMSECV and R^2. As it is shown, the correlation coefficients of actual versus predicted concentrations resulting from multivariate calibration models for the different oils were between 0.95 and 0.99. The model of the blend of olive oil and sunflower oil had the highest R^2 (0.9995) and all of them have a RMSECV < 0.1; the model of olive oil and corn oil had the lowest R^2 (0.9552). According to these criteria all types of vegetable oils could be quantified with excellent results. The optimum number of factors was selected in order to avoid overfitting when using PLS.

In order to evaluate the performances of the proposed approach, involving a first classification of unknown samples and later the application of the PLS model built on the samples of the nearest class, a further external validation stage was carried out. For this purpose, 16 test samples consisting in blends of olive oil with 4 of the 5 analyzed vegetable oils (sunflower, corn, seeds and soybean) in different proportions, were processed by SIMCA and PLS. SIMCA model was able to correct assign all the validation samples to the class corresponding to their true origin. Then, for each sample, the PLS model corresponding to its predicted category was used for quantification and an RMSEP lower than 0.1 was obtained, thus confirming the validity of the proposed approach also for external samples.

However, there can be cases -for instance when blends are produced mixing EVO with more than one kind of other vegetable oils, or when an oil from a category not considered in this study is used – where a sample is recognized as outlier by all the class models in the preliminary SIMCA step. In those cases, none of the individual PLS models could be perfectly suitable for quantifying the blending ratio, or in general assessing EVO purity.

To overcome these drawbacks, based on the promising results of the modeling performed on each category separately, in a further stage the possibility of building an overall model, able to quantify the percentage of EVO in the samples, irrespectively of the kind of oil used for blending, was studied. Accordingly, a calibration set was built on all the 62 training samples and therefore including a wider variability (four categories of olive oil, EVOO, VO, OO and pomace, from three varieties of olive fruit, arbequina, picual and hojiblanca and five different vegetable oils sunflower, corn, seed, sesame and soya oil). The calibration set for the quantification of OO was constituted of 62 samples, including mixtures containing 10%, 30%, 50%, 70% and 90% of olive oil in the different vegetable oils tested. The optimal complexity of the model was assessed by sevenfold cross-validation. Successively, in order to evaluate the predictive ability of the model on unknown samples, it was applied to an external test set composed of 16 samples which were not employed in the calibration phase. This validation data set was also baseline corrected and aligned. Validation results were evaluated in terms of root-mean-squared error of prediction (RMSEP) and R^2.

In order to check whether it was possible to obtain accurate result with a reduced set of variables, a GA–PLS was adopted on the same set of samples. As the number of original variables (i.e. all the data points in the chromatogram) exceeded 1000, GA was applied in two stages in order to avoid overfitting, at first considering the mean of 5 adjacent points and then on the intervals selected in the previous stage. Eventually, 18 variables were selected as input for the final PLS model (in this case autoscaling was chosen as pretreatment).

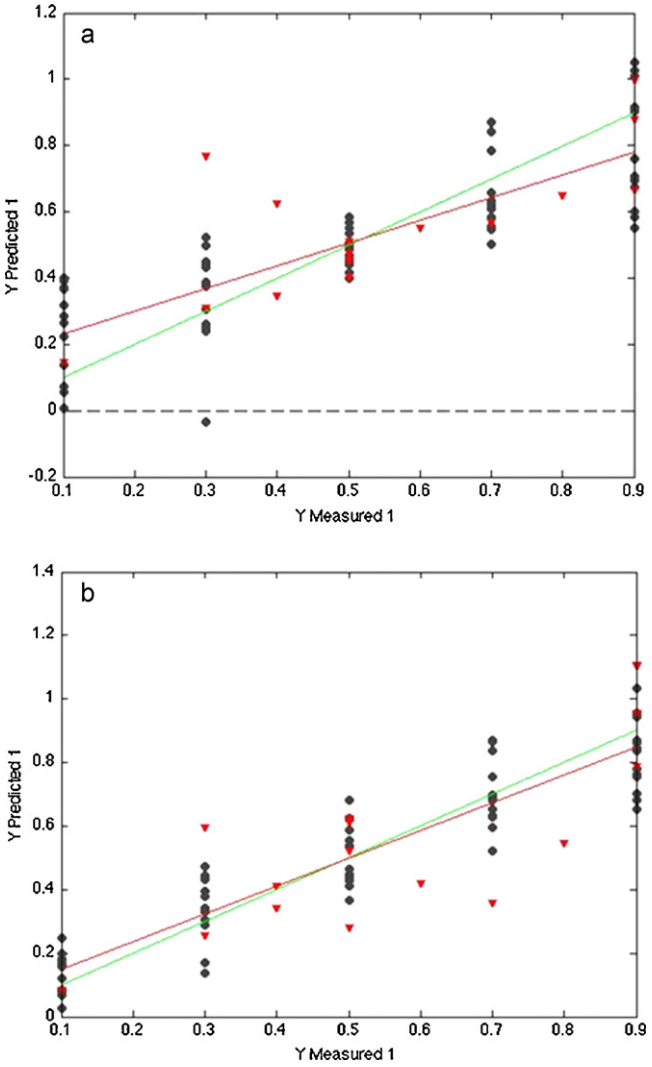

Fig. 5. PLS model using all the blends of oils together as well as external validation (●, training set; ▼, validation set). (a) All the chromatographic data were used to built the model. (b) Variable selection was applied with genetic algorithm.

Table 3
Calibration and prediction parameters of the PLS model for all the olive oil and vegetable oil blends.

	LVs	R^2	RMSEC	RMSECV	RMSEP	CV Bias	PredBias
PLS	2	0.695	0.1562	0.1642	0.1675	0.0014	−0.0114
GA–PLS	2	0.863	0.1046	0.1139	0.1737	0.0012	−0.0129

Fig. 5 compares the PLS model without variable reduction (a) with the GA–PLS model (b), where the genetic algorithms approach was used to reduce the big amount of variables. The results obtained in terms of calibration and prediction abilities for both models are summarized in Table 3. As it is shown, reduction in the number of variables does not lead to a significant worsening of the modeling and predictive ability. Alternatively, on one hand the complexity of the models in terms of the number of latent variables remains the same (2 LVs), on the other, the error in modeling and cross-validation is significantly lowered (0.1046 and 0.1139, respectively, with GAPLS versus 0.1562 and 0.1642 without variable reduction), while the RMSE in prediction is only slightly worse (0.1737 versus 0.1675).

As a result, variable reduction results in a model having a lower complexity in terms of retained experimental variables and comparable predictive ability.

In conclusion the high temperature gas chromatographic proposed method and the chemometric class-modeling techniques such SIMCA and quantification techniques PLS and GA–PLS with feature selection, appear to be appropriate tools to verify the percentage of olive oil in blends with vegetable oils and could become an important instrument to verify the labeling compliance and for quality control in the detection of adulterations. Indeed, the reliability of the proposed qualification model is very high as the kind of vegetable oils used for blending was correctly identified for all samples. Moreover, reliable quantification models were built for each of the different kinds of blending. Lastly, when the possibility of quantifying the purity of oil samples irrespectively of the adulterating vegetable oil, promising results were obtained by applying PLS (on the entire chromatogram or with GA variable selection) to the whole data set without preliminary classification of the oils. Thus, this study is especially interesting since olive oils are frequently adulterated with other vegetable oils.

Acknowledgments

The authors would like to thank to the Andalusia Regional Government (Consejería de Innovación, Ciencia y Empresa, Project P07-FQN-02667, and Consejería de Agricultura y Pesca) for financial assistance. This work has also been partially supported by European Regional Development Funds (ERDF). The authors are grateful to the Andalusia Regional Government (Consejería de Innovación, Ciencia y Empresa) for the personal postgraduate grant awarded to CRS.

References

[1] J.R. Park, D. Lee, Bull. Korean Chem. Soc. 24 (4) (2003) 527.
[2] J.M. Dyer, S. Stymne, A.G. Green, A.S. Carlsson, Plant J. 54 (2008) 640.
[3] F. Angerosa, C. Campestre, L. Glasante, in: D. Boskou (Ed.), Olive Oil, Chemistry and Technology, 2nd ed., AOCS Press, Champaign, IL, 2006, p. 113.
[4] R. Aparicio, R. Aparicio-Ruız, J. Chromatogr. A 881 (2000) 93.
[5] A.K. Kiritsakis, K. Kanavouras, Eur. J. Lipid Sci. Technol. 104 (2002) 628.
[6] A. Cert, W. Moreda, M.C. Pérez-Camino, J. Chromatogr. A 881 (2000) 131.
[7] Commision Regulation (EC) No. 29/2012, on marketing standards for olive oil, Off. J. Eur. Commun. L12 (2012) 14.
[8] N. Darmon, M. Darmon, E. Ferguson, J. Hum. Nutr. Diet. 19 (2006) 59.
[9] E. Christopoulou, M. Lazaraki, M. Komaitis, K. Kaselimis, Food Chem. 84 (2004) 463.
[10] I.S. Arvanitoyannis, A. Vlachos, Crit. Rev. Food Sci. Nutr. 47 (2007) 441.
[11] E.N. Frankel, J. Agric. Food Chem. 58 (2010) 5991.
[12] C. Ruiz-Samblás, L. Cuadros Rodríguez, A. González Casado, F.P. Rodríguez García, P. de la Mata Espinosa, J.M. Bosque Sendra, Anal. Bioanal. Chem. 399 (2011) 2093.
[13] P. de la Mata Espinosa, J.M. Bosque Sendra, R. Bro, L. Cuadros Rodriguez, Talanta 85 (2011) 177.
[14] V.G. Dourtoglou, Th. Dourtoglou, A. Antonopoulos, E. Stefanou, S. Lalas, C. Poulos, J. AOCS 80 (3) (2003) 203.
[15] D. Brodnjak-Voncina, Z.C. Kodbab, M. Novic, Chemometr. Intell. Lab. Syst. 75 (2005) 31.
[16] F. Peña, S. Cárdenas, M. Gallego, M. Valcarcel, J. Chromatogr. A 1074 (2005) 215.
[17] R.M. Maggio, L. Cerretani, E. Chiavaro, T.S. Kaufman, A. Bendini, Food Control 21 (2010) 890.
[18] F. Priego Capote, J. Ruiz Jiménez, M.D. Luque de Castro, Anal Bioanal. Chem. 388 (2007) 1859.
[19] B. Öztürk, A. Yalcyn, D. Özdemir, J. Near Infrared Spectrosc. 18 (2010) 191.
[20] International Olive Council, Madrid (COI/T.20/Doc. No 20/Rev. 2), 2008.
[21] R. Aparicio, R. Aparicio-Ruiz, J. Chromatogr. A 881 (2000) 93.
[22] European Community, Commission Regulation No. 1513/2001 amending Regulations No. 136/66/EEC and No. 1638/98 as regards the extension of the period of validity of the aid scheme and the quality strategy for olive oil. Off. J. Eur. Commun., L201/4.
[23] C. Ruiz-Samblás, A. González-Casado, L. Cuadros-Rodríguez, et al., Talanta 82 (2010) 255.
[24] B.M. Wise, N.B. Gallagher, R. Bro, J.M. Shaver, W. Windig, R.S. Koch, PLS Toolbox 4. 0 Manual for use with MATLABTM, Eigenvector Research, Inc., Wenatchee, WA, USA, 2006.
[25] D.L. Massart, B.G.M. Vaneginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J.J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics: Part A, Elsevier, Amsterdam, 2007.
[26] F. Savorani, G. Tomasi, S.B. Engelsen, J. Magn. Reson. 202 (2010) 190.
[27] F. Marini, A.L. Magrì, R. Bucci, F. Balestrieri, D. Marini, Chemometr. Intell. Lab. Syst. 80 (2006) 140.
[28] R. Leardi, A. Lupiáñez, Chemometr. Intell. Lab. Syst. 41 (1998) 195.
[29] R. Leardi, J. Chemometr. 14 (2000) 643.
[30] Z. Xiaobo, Z. Jiewen, M.J.W. Povey, M. Holmes, M. Hanpin, Anal. Chim. Acta 667 (2010) 14.
[31] R. Leardi, The PLS–Genetic Algorithm Toolbox for Matlab. Downloaded from: http://www.models.life.ku.dk/GAPLS.