View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Queen's University Research Portal

Identification of vegetable oil botanical speciation in refined vegetable oil blends using an innovative combination of chromatographic and spectroscopic techniques

Osorio, M. T., Haughey, S. A., Elliott, C. T., & Koidis, A. (2015). Identification of vegetable oil botanical speciation in refined vegetable oil blends using an innovative combination of chromatographic and spectroscopic techniques. Food Chemistry, 189(SI), 67-73. DOI: 10.1016/j.foodchem.2014.11.164

Published in:
Food Chemistry

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Publisher rights
© 2015, Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International http://creativecommons.org/licenses/by-nc-nd/4.0/ which permits distribution and reproduction for non-commercial purposes, provided the author and source are cited.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Download date:15. Feb. 2017

Identification of vegetable oil botanical speciation in refined vegetable oil blends using an innovative combination of chromatographic and spectroscopic techniques.

Maria Teresa Osorio, Simon A. Haughey, Christopher T. Elliott, Anastasios Koidis*

Institute for Global Food Security, Queen's University Belfast, Northern Ireland, UK

Running title: vegetable oil speciation in refined oil blends

Corresponding author:

Dr Anastasios Koidis

Institute for Global Food Security

Queen's University Belfast

18-30 Malone Road

Belfast, BT9 5BN

Northern Ireland, UK

Tel: +44 28 90975569, Fax: +44 28 90976513

email: t.koidis@qub.ac.uk;

Abstract

European Regulation 1169/2011 requires producers of foods that contain refined vegetable oils to label the oil types. A novel rapid and staged methodology has been developed for the first time to identify common oil species in oil blends. The qualitative method consists of a combination of a Fourier Transform Infrared (FTIR) spectroscopy to profile the oils and fatty acid chromatographic analysis to confirm the composition of the oils when required. Calibration models and specific classification criteria were developed and all data were fused into a simple decision-making system. The single lab validation of the method demonstrated the very good performance (96% correct classification, 100% specificity, 4% false positive rate). Only a small fraction of the samples needed to be confirmed with the majority of oils identified rapidly using only the spectroscopic procedure. The results demonstrate the huge potential of the methodology for a wide range of oil authenticity work.

keywords: authentication, labelling, vegetable oil, blend, palm oil, spectroscopy.

1. Introduction

Almost all processed foods, such as confectionary, pastry and other ready-to-eat products, contain substantial amounts of refined vegetable oil. According to current European legislation requirements it is labelled as simply ‘vegetable oil’ although it can be pure oil or blends of different oil botanical species. The most common oils used in food manufacturing are refined palm, rapeseed, sunflower oil, soybean oil and to a lesser extent, cottonseed and coconut oil. The EU Regulation 1169 on the Provision of Food Information to Consumers which take effect from 13 December 2014 across the EU, introduces a new requirement that changes the way these products are labelled. In order to provide additional information to the consumer, the food manufacturers will need to include the type of oil used (i.e. to list all the oil botanical species) in the ingredients’ list on the label. It will not be necessary to indicate the proportion in which oils are used where there is a mixture, and the label may indicate that the oils are used in different proportions (to allow for seasonal and market fluctuations).

From the vegetable oils listed, palm oil will be the most abundant oil and is used extensively in food manufacturing. Palm oil has emerged as the preferred oil source due to its naturally low trans-fatty acid content, unique flavour, cost and desirable physical properties (texture and melting point). Today the majority of the processed foods contain palm oil in some form (palm oil or its derivatives, palm olein and palm stearin). According to the food industry, this might have some impact on food choices that consumers make. The reason is that palm oil production and agricultural practises have generated global interest with regards to sustainability and fair trade and concerns about damage to biodiversity in some tropical areas with palm oil plantations. Currently, many leading food companies are members of the Roundtable on Sustainable Palm Oil (RSPO), an organisation that promotes and certifies palm oil. Even if only a small portion of palm oil production is currently certified, sustainable palm oil is in great demand in food manufacturing especially in Europe.

From an analytical point of view, testing an unknown vegetable oil to identify its origin and composition is a very difficult task, one that has confounded researchers, public chemists and legislation authorities for years, especially with regards to premium oil authenticity. One could undertake a battery of chemical tests, such as fatty acid analysis, determination of sterol and hydrocarbon fractions, tocopherols, pigments and still remain uncertain about the oil's authenticity. Premium vegetable oils such as extra virgin olive oil have been researched extensively in order to identify adulteration with other oils such as refined olive oil, deodorised olive oil, seed oils, etc. (Gurdeniz & Ozen, 2009; Baeten, Ferrnández, Dardenne, Meurens, García-González & Aparicio-Ruiz, 2005; De Luca et al., 2011). Developing such methodology for refined oils, where most of the polar fraction has been significantly reduced or eliminated during the refinement process, creates an additional challenge (Koidis & Osorio-Argüello, 2013). One has to mostly rely on the unique chemical information retained in the triacyglycerols and the fatty acids as major components of the oil, rather than focusing on the minor constituents that remain after the refinement process such as sterols as the majority of the methods in the literature have exercised. In a comprehensive literature review (Osorio, Haughey, Elliott & Koidis, 2014), all potential analytical methods ranging from DNA methods to stable isotope analysis including spectroscopic and chromatographic methods were critically discussed and evaluated for this particular analytical problem. Unsurprisingly, the majority of the sources cite analytical techniques applied to crude vegetable oils such as virgin olive oil rather than refined oils. This was expected as the analytical need for identifying refined vegetable oil species didn't exist before the introduction of the new EU legislation. Based on critical analysis of a) the literature review results on vegetable oil authenticity, b) the chemical composition of the three particular vegetable oils (palm, sunflower and rapeseed oil) and c) the impact of refining process on their constituents, it was found that only a small number of specific chromatographic and spectroscopic fingerprinting methods coupled with chemometrics appear applicable to this complex challenge. The latter techniques are mainly FTIR and Raman spectroscopy in untargeted mode and triacylglycerol and fatty acid analysis in targeted mode. Combining these two techniques in the area of oil analysis has been suggested (Aparicio & Aparicio-Ruiz, 2000) but has not been applied before according to the literature. There are many uncertainties as to the extent that these methods will work, how different results of different nature (targeted and untargeted) are going to be combined and how the two methods, spectroscopic and chromatographic, can be used mutually and collectively to strengthen the accuracy of the result. It was therefore clear that major experimental work has to be performed to test these hypotheses.

The aim of the current study, therefore, was to develop a novel analytical procedure based on both spectroscopic and chromatographic techniques for the identification of oil blends of same or different species of refined vegetable oils.

2. Materials and methods

2.1. Sourcing of refined authentic vegetable oils

Refined vegetable oils (whole palm oil, palm stearin and palm olein, palm kernel oil, sunflower and rapeseed oil, n=23) were obtained as reference authentic samples and were used to build in-house admixtures (Section 2.2) and develop the methodology. In addition, 23 authentic samples of extra virgin olive oil (n=19) and refined hazelnut oil (n=4) were obtained to aid in method validation as ‘negative examples’. All 46 samples were pure oils (100%) purchased from reliable and reputable sources within major food companies, oil processing industry and directly from oil producers when possible. The refined oils had global origin and represent the European and global oils supply market. Due to confidentiality issues the origin of some of all the samples sourced is not provided. However, most of the palm oils samples originate from Indonesia, Malaysia, Papua New Guinea and South America.

The sample dataset (n=47 pure oils, Figure 1) was divided into 3 independent sets, the calibration dataset, the prediction set and the validation set (Section 2.2). Due to the lack of authentic samples, some oils were used for both calibration and prediction set although validation samples were completely independent at all cases. The sample distribution is presented in Figure 1. The samples from the calibration set were used to calibrate the chemometric models for every class. The prediction set utilisation was two-fold: it was used firstly to benchmark the prediction efficiency of the calibration models (intra-validation) and provide evidence on the best spectroscopic and chemometric technique and secondly, as the basis for the development of the confirmation chromatographic analysis criteria. The validation samples (n=23) were comprised of an independent group of refined authentic vegetable oils (and their admixtures, see Section 2.2. and Figure 1) and a group of extra virgin olive oil and refined hazelnut pure oils, also referred as 'negative samples'. 'Negative samples' (n=23) are meant to confirm if the method returns any false positives. The validation dataset (n=46) was used to test the entire methodology, both screening and confirmation stage.

2.2. Preparation of in-house oil mixtures

Oil binary admixtures, derived from authentic oils (palm oil, palm kernel, palm stearin, palm olein, sunflower and rapeseed oils), were created in-house (n=213, Figure 1). After consultation with the industry and law enforcing bodies it was determined which were the most relevant oil binary blends used in food manufacturing. These binary admixtures were:

palm stearin + palm oil (PS-PO): 24 admixtures; palm olein + sunflower oil (POL-SO): 24 admixtures; sunflower oil + palm oil (SO-PO): 28 admixtures; rapeseed oil + palm kernel oil (RO-PKO): 19 admixtures; sunflower oil + palm kernel oil (SO-PKO): 19 admixtures; palm oil + palm kernel oil (PO-PKO): 24 admixtures; rapeseed oil + sunflower oil (RO-SO): 24 admixtures; rapeseed oil + palm oil (RO-PO): 28 admixtures. All binary admixtures (e.g. A:B) contained various concentrations of A and B in 4% intervals from 4 to 96% (for PS-PO, POL-SO, PO-PKO, RO-SO), in 4 and 2% intervals from 6 to 96% (for SO-PO, RO-PO) and in 4 and 6% intervals from 4 to 94% (for RO-PKO, SO-PKO). However, in order to improve the model performance, the oil admixtures with extreme analogies were not included in the calibration set. The optimal admixture analogies contained between 15:85 of each oil (n=115 samples). Limited ternary admixtures were also created but were not used in the study, as it is uncommon for 3 different species to be used in one product.

In the preparation of every admixture, oils from different sources and geographical origin were used in order to capture compositional variability. All oil samples were stored individually in 125 ml amber glass vials in the dark at -18°C with a headspace of <5% to avoid auto-oxidation and photo-oxidation.

2.3. Spectral Data Acquisition with FTIR and Raman spectroscopy

For FTIR, samples were kept at 50°C prior to analysis and immediately placed in the ATR sample area of a Thermo Nicolet iS5 spectrometer (Thermo Fisher Scientific, Dublin, Ireland) equipped with ATR iD5 diamond and DTGS KBr detector. A few drops of oil were used and each spectrum was acquired in the 550 - 4000 cm^{-1} range. The acquisition parameters were: number of sample scans: 32; collection length: 51.1 s; resolution: 4.000; levels of zero filling: 2, number of scan points: 12415; number of FFT points: 65536; laser frequency: 15798.0 cm^{-1}; interferogram peak position: 6100; apodization: N-B Strong; phase correction: mertz; number of background scans: 32. The acquisition was repeated 3 times.

Raman spectra acquisition was performed in an Advantage 1064 Raman Spectrometer (DeltaNu Inc., Wyoming, USA). Three hundred microlitres of the oils were pipetted into glass vials, with a pathlength of 10 mm and shortly kept at 50°C prior to the analysis. Acquisition was performed for all samples at 10 cm^{-1} resolution across the spectral range 200 - 2000 cm^{-1}. Using the NuSpec software, the following parameters were inserted: number of points: 6950; data spacing: 0.482117; integration time: 10 sec. The acquisition was repeated twice.

All spectra were pre-processed according to a suitable standardized treatment which includes three spectral filters, standard normal variate (SNV), first order derivative and Savitsky-Golay smoothing, applied in a sequential order (Graham, Haughey, Ervin, Cancouët, Bell, & Elliott, 2012). For FTIR, 3781 variables were selected in the range intervals (654.2 to 1875.4 cm^{-1}) and (2520.0 to 3120.7 cm^{-1}). The Raman interval used for data analysis was 800.3 to 1800.2 cm^{-1} resulting in 1038 variables.

2.4. Chromatographic determination of fatty acid methyl esters

Fatty acid methyl esters were prepared according to British Standard BS EN ISO 12966-2:2011 using a Varian CP3800 Gas chromatograph fitted with Flame Ionisation Detector (JVA Analytical, Dublin, Ireland) running on a Agilent CP-88-SIL (100m x 0.25mm id, 0.2µm film thickness) analytical column. Briefly, oil blends were heated to 60°C to ensure complete melting of the solid fat component before being thoroughly mixed prior to sampling. Subsamples (300 mg) were taken in duplicate and dissolved in 10 ml of hexane. An aliquot of the fatty acid methyl esters in hexane was transferred to a vial prior to analysis by gas chromatography. Individual fatty acid methyl esters were detected by flame ionisation detection, identified by comparison with external fatty acid methyl ester standards and quantified by the use of methyl tridecanoate (Sigma-Aldrich, Dorset, UK) as internal standard. Blanks were included within each batch of samples to establish base line stability and instrument readiness. The internal standard was added to each sample prior to preparation and determination of the fatty acid methyl esters. All analyses were carried out in duplicate. Final results are expressed both as mg fatty acid g^{-1} of sample and as percentage of total fatty acids in the oil.

2.5. Calibration modelling and prediction

Multivariate data exploration (Principal Component Analysis) was performed using Umetrics SIMCA 13.0 (Umea, Sweden). Calibration of specific model classes was performed using two independent supervising classification techniques, Partial Least Square Discriminant Analysis (PLS-DA) and Soft Independent Model Class Analogy (SIMCA). Independently of the technique selected, cross validation in SIMCA 13 is carried out automatically as follows: The data are divided into 7 parts and each $1/7^{th}$ in turn is removed. A model is built on the $6/7^{th}$ data left in and the left out data are predicted from the new model. This is repeated with each $1/7^{th}$ of the data until all the data have been predicted. The predicted data are then compared with the original data and the sum of squared errors calculated for the whole dataset. This is then called the Predicted Residual Sum of Squares (PRESS). The better the predictability of the model the lower this value will be. For convenience, SIMCA 13 converts PRESS into Q2 to resemble the scale of the R2. R2 is a measure of variation of the training set explained by the model and is a measure of fit, i.e. how well the model fits the data. Q2 indicates how well the model predicts new data. Good predictions will have high Q2. After calibration, prediction and external validation sets were used independently in the developed models and their prediction parameters, R2 and the Q2, along with Distance-to-Plot scores were calculated.

3. Results and discussion

The proposed method to identify oil botanical species in vegetable oil blends consists of a screening stage based on a spectroscopic technique operating in untargeting mode and a confirmation stage based on a chromatographic targeted analysis.

During the development of the staged method, it was important to establish a) the specific spectroscopic technique (FTIR *vs.* Raman) that is most suitable for screening, b) the exact multivariate classification technique (SIMCA *vs.* PLS-DA) and c) the actual model classes, i.e. the oil types included in every class. In addition, although the chromatographic method of the confirmation stage was early identified (fatty acid analysis using GC/FAME) specific criteria for individual fatty acids had to be developed. These criteria had to be quantitative and based on the final model classes in order to confirm the nature of an unknown sample.

3.1. Choice of screening spectral technique, classification algorithm and model classes

Both FTIR and Raman spectroscopy were used and compared as screening techniques in order to create a database of spectroscopic data from vegetable oil samples (pure and admixtures) and use it as the basis for building the calibration models. Recorded spectra of some pure oils (4 palm kernel oils, 5 palm oils, 2 palm stearins, 1 palm olein, 4 rapeseed oils and 4 sunflower oils) can be seen in Figure 2A for FTIR. Substantial differences were observed amongst the six different types of pure oils when all spectra were superimposed, which was an early indication that there was sufficient signal differences between the oils at the molecular level (stretching and bending vibrations induced by infrared absorption). Similar information was observed in the superimposed Raman spectra of pure oils. Preprocessing of spectral data removed undesired systematic variation in the data (i.e. baseline drift and wavenumber regions of low information content) and therefore enhanced the predictive power of multivariate calibration models (Eriksson, Johansson, Kettaneh-Wold, Trygg, Wikstrom & Wold, 2006). FTIR data exploration with Principal Component Analysis (PCA), an unsupervised technique, showed that initial spectral differences correspond to a very good separation in the scores plot using 2 or 3 principal components (PCs) (Figure 2B, 2C). Loadings plot revealed that the most discriminative wavelengths in the FTIR spectra were those within the range of a) 1117-1142 cm^{-1} corresponding to stretching vibration of the C-O ester group, b) 1732-1747 cm^{-1} accounting for the ester carbonyl functional group of the triglycerides and c) 2845-2925 cm^{-1} relating to the asymmetrical and symmetrical stretching vibration of methylene (-CH2) group (Guillen & Cabo, 1997; Lerma-Garcia, Ramis-Ramos, Herrero-Martinez & Simo-Alfonso, 2010; Rohman & Che Man, 2010). Palm kernel oil (PKO) samples have very distinctive spectral characteristics and can be considered as a class of their own in the PCA score plot (Figure 2B, 2C). Palm olein (POL), palm stearin (PS), whole palm oil (PO) and the admixture PS-PO are grouped together due to their very similar chemical composition and origin (e.g. palm stearin + olein = whole palm oil) and were therefore considered as one class (P class) instead of 3 different classes. The same applies to sunflower oil (SO), rapeseed oil (RO) and the admixtures comprise of those two seed oils (RO-SO) that are clustered together (RS class) due to their similar polyunsaturated character. The remaining classes, POL-SO, SO-PO, RO-PKO, SO-PKO, PO-PKO and RO-PO were clustered in three groups and therefore considered as RSPKO (RO-PKO, SO-PKO), RSPO (SO-PO, RO-PO, POL-SO) and PPKO (PO-PKO) classes. These three new classes were clustered, as expected, in the virtual space between the 3 initial new classes (PKO, P and RS) and they accommodate all remaining oil admixture samples (Figure 2C). The Raman spectral data (not shown) also support the 6-

class argument although the class separation is clearer when using FTIR data. Several other iterations have been attempted including a trial model with 18 independent classes but the best prediction performance was obtained with the 6-class model design selected which is parsimonious and has a chemical composition rationale.

In parallel to the model class design, the exact spectroscopic techniques (FTIR and Raman) and classification algorithms were also explored. In general, FTIR contains more high-level processed signal parameters and slightly richer information than Raman, which is essentially a low-end dispersive instrument. In fact, FTIR has shown better performance in vegetable oil botanical speciation especially with olive oil according to the literature (Osorio et al., 2014). These two techniques are based in different light optical phenomena (absorption vs. scattering) and, in theory, both of them would be useful as they can be complimentary. On the other hand, both classification techniques (SIMCA and PLS-DA) have proven useful in classifying spectroscopic data of oils (Sinelli, Cosio, Gigliotti & Casiraghi, 2007; Gurdeniz & Ozen, 2009; Rohman & Chen Man, 2010). Comparing the spectral and classification techniques was done simultaneously. The model performance in classifying oil admixtures spectroscopically using the prediction set was equally good on FTIR and Raman data (Table 2), although, in some cases, Raman achieved marginally higher model parameters Q2 and R2. In terms of prediction power, all 4 combinations produced excellent results when the calibration models were challenged with the prediction set (Table 2). The classification rates were slightly overestimated due to the presence of the sample replicates. SIMCA, however, proved more accurate when testing unknown and control oil admixtures by producing less classification errors using the prediction set. More specifically, SIMCA is not 'forced' to classify all samples to a particular class in contrast to PLS-DA (Wold, 1976; Wold & Sjostrom, 1977; Bevilacqua, Bucci, Magrì, Magrì, Nescatelli & Marini, 2013). In fact, it will return samples unclassified, i.e. not fitted in any of the model classes, if the residual distance from the model is above the statistical limit in every class. This provides great flexibility, reduces classification errors and fits very well with the purpose of the two-staged classification approach presented in this study. In addition, in supervised methods, it is important to avoid overfitting by using a relatively large validation set or with robust internal cross-validation (Berrueta, Alonso-Salces & Héberger, 2007). PLS-DA is especially prone to overfitting (Brereton, 2009) and random noise introduced as more laternt valiables are added (Zielinski, Haminiuk, Nunes, Schnitzler, van Ruth & Granato, 2014) compared to SIMCA and . FTIR in conjunction with SIMCA produced the highest overall classification rate when tested with the prediction set (Table 2). Therefore, the combination of FTIR and SIMCA classification technique was established as the most suitable screening tool that is fit-for-purpose.

3.2. Development of decision system and confirmation technique criteria

Unclassified oil samples in the screening stage were transferred to the second stage where a confirmation technique was applied. This was realised through a simple procedure based on the probabilities of the SIMCA classification algorithm during the screening stage: when an unknown oil spectra is loaded, SIMCA calculates the distance-to-model to produce a probability score for every oil sample to belong in each one of the 6 classes. Samples are then divided into 3 groups: of high probability (> 0.1) to belong in the particular class, of medium probability (0.05 to 0.1) and of low probability (< 0.05, not classified) (Figure 3). Only the unclassified samples of the latter group were transferred to the second stage due to the uncertainty of the result. A sample may be found to belong to multiple classes. In this case the class with the highest probability (the lowest residual distance to model) is chosen.

Meticulous care was exercised so that the decision system would not a) erroneously classify a sample to a different class (misclassification, false positive), b) does not refer an ambiguous sample to the confirmation stage (false negative or 'miss').

Gas chromatography for the analysis of fatty acid methyl esters was chosen as the confirmation technique for its wide applicability, accessibility and accuracy in the results (Aparicio & Aparicio-Ruiz, 2000). Fatty acid criteria based on individual key FA contents were developed to classify the samples in one of the 6 classes. Every class has unique and highly specific classification criteria as seen in Table 3. These criteria were developed analysing the fatty acid profile of the prediction set and analysing standardised compositional ranges for vegetable oils found in the Codex Alimentarius (CODEX STAN 210, 2011). The criteria were validated using the validation set. The final procedure is illustrated as a two-stage decision making system (Figure 3).

3.3. Single Lab Validation of the method using external samples

A single lab validation with external samples was performed to demonstrate the performance of the method on a new set of 46 oil samples (pure oils and oil blends) including 23 'negative samples' (Figure 1). It has to be reiterated that these oils were different from the oils used in the calibration modelling and prediction sets. The proposed method flowchart (Figure 3) was followed to assess the assignment success of the external samples in the 6 modelled classes. FTIR spectra were recorded and pre-processed for all external samples (see 2.3, 2.5). This set was tested against the SIMCA calibration models and a probability score was assigned to each sample according to the classification algorithm. A total of 18 oil samples were classified as follows: 6 in P class; 4 in RS class, 4 in RSPKO class and 4 in RSPO class. The rest of the samples (n=28) were referred to the confirmation stage due to their low probability score. These samples were analysed chromatographically to determine their fatty acid (FA) profile and individual contents (mg fatty acid g^{-1} oil blend or % of total FA) were calculated. The following classification results were obtained when the FA criteria (Table 3) were applied: 1 in PKO class; 2 in P class, 2 in RSPO class, 1 in PPKO class and 23 samples remained unidentified. The 23 unidentified oil samples were the 'negative examples' and were correctly rejected by the method (initially rejected by the SIMCA algorithm due to the large residual distance from all modelled classes and subsequently failed to comply with the FA criteria). These samples represent the 'true negatives' of the test. At the end of the procedure, 45 out of 46 samples were correctly classified (97.8%). The incorrect sample (palm kernel oil) was erroneously classified in the spectroscopic stage as palm oil (P class) and was considered a 'false positive'.

The mathematical formulas that describe method validation metrics as precision, accuracy, robustness etc., are linked with quantitative methods, (AOAC, 1995; Boque, Maroto, Rui & Rius, 2002) and cannot be applied in qualitative analysis. Ellison and Fearn (2005) argue that it is necessary to rethink the conventional metrology so that qualitative methods are also factored. Although there are no universally accepted validation standards in qualitative analysis, the reliability indexes presented in García-González, Viera, Tena and Aparicio (2007) and Cárdenas and Valcárel (2005) have been acknowledged as an accepted evaluation of the performance of such methods. The reliability indexes therefore are: False Negative rate (FNr): 0%, False Positive (FPr): 4%, sensitivity: 100%, specificity: 100% and efficiency: 98%. On the other hand, if a confusion matrix is used, a common classification technique in machine learning that factors in the individual class success (Kohavi & Provost, 1998), the following parameters are calculated: average accuracy 85.7%, average reliability: 78.5%, overall accuracy: 97.8%.

It is therefore confirmed that the decision making process and especially the criteria of the chromatographic confirmatory analysis are rigorous if challenged with external samples and the ~4% classification error can be attributed to the calibration models that need further optimisation. This applies especially to the PKO model (Q2 cumulative 0.249) which had a low prediction power and may be the reason for the misclassification of the external palm kernel oil. This, however, should not undermine the excellent overall method performance and the significant advantages of the two-staged procedure (only 21% of the 'true' validation samples required confirmation) in terms of speed of analysis and low cost benefits of a spectroscopic measurement if the confirmatory chromatographic analysis is omitted.

4. Conclusions

In the current study, an innovative staged method has been developed through the unique merging of spectroscopic and chromatographic analysis for the botanical species identification of vegetable oils. The combination of FTIR spectroscopy technique and SIMCA classification technique was established as the most suitable screening tool for the purpose of this work. SIMCA class-models achieved high levels of correct classification when FTIR spectral data were used and strongly suggest the utility of this combined approach in vegetable oil screening. PLS-DA discriminant models also performed very well but the risk of misclassified samples is higher. Fatty acid analysis performed by GC-FID proved to be powerful in identifying samples that could not be assigned to a class by the SIMCA models. In general, this qualitative method produced very good results in the single laboratory validation. The sample size used for building the calibration models was relative small although representative of the global vegetable oil supply and this limits a true assessment of model performance. The current results have gone some way to proving the concept of this novel and highly sensitive two-staged approach for identifying the kind of oils present in oil blends and indicate the need of a larger study for a more robust and representative method in both plain oil blends as well as in processed foods containing refined oil blends.

This study also highlights the numerous analytical challenges that legislation and enforcement authorities are facing with the current analytical methods to monitor compliance of EU legislation of food labels in processed foods and oil authenticity in general.

Acknowledgements

This research was funded by the Department of Environment, Food and Rural Affairs (DEFRA), FAO117 project entitled ‘Identification of oil species in refined vegetable oil mixtures’. We are grateful to our oil industry and food companies contacts for providing authentic vegetable oil samples. The authors also thank Dr Jesus Martinez del-Rincon for his constructive comments and suggestions regarding method validation.

References

AOAC INTERNATIONAL (1995). Guidelines for Collaborative Study Procedures to Validate Characteristics of a Method of Analysis *Journal of the Association of Official Analytical Chemists International 78(5),* 143A-160A.

Aparicio, R., & Aparicio-Ruíz, R. (2000). Authentication of vegetable oils by chromatographic techniques. *Journal of Chromatography A, 881(1),* 93-104.

Baeten, V., Fernández, J., Dardenne, P., Meurens, M., García-González D. L., & Aparicio-Ruiz, R. (2005). Detection of the presence of hazelnut oil in olive oil by FT-Raman and FT-MIR spectroscopy. *Journal of Agricultural and Food Chemistry, 53*, 6201-6206.

Berrueta, L. A., Alonso-Salces, R. M., & Héberger, K. (2007). Supervised pattern recognition in food analysis. *Journal of Chromatography A, 1158(1)*, 196-214.

Bevilacqua, M., Bucci, M. R., Magrì A. D., Magrì A. L., Nescatelli, R., & Marini, F. (2013). Classification and class-modelling. In F. Marini (Eds.), *Chemometrics in food chemistry* (pp. 171-232). Elsevier.

Boqué, R., Maroto, A., Riu, J., & Rius, X. (2002). Validation of Analytical Methods. *Grasas Aceites, 53*, 128-143.

Brereton, R. (2009). Chemometrics for pattern recognition. validation and optimisation (pp. 311-390). Wiley, Chichester, United Kingdom. ISBN: 978-0-470-98725-4, pp. 522.

BS EN ISO 12966-2:2011. Animal and vegetable fats and oils. Gas chromatography of fatty acid methyl esters. Preparation of methyl esters of fatty acids (British Standard). The British Standards Institution 2014: London.

Cárdenas S. & Valcárcel, M. (2005). Modern Qualitative Analysis. *Trends in Analytical Chemistry, 24(6)*, 477-487.

CODEX STAN 210-1999, Codex Standard for Named Vegetable Oils (Adopted 1999. Revisions 2001, 2003, 2009. Amendment 2005, 2011), Codex Alimentarius, Rome.

De Luca, M., Terouzi, W., Ioele, G., Kzaiber, F., Oussama, A., Oliverio, F., Tauler, R., & Ragno, G. (2011). Derivative FTIR spectroscopy for cluster analysis and classification of morocco olive oils. *Food Chemistry, 124*(3), 1113-1118.

Ellison, S. L. R., & Fearn, T. (2005). Characterising the performance of qualitative analytical methods: Statistics and terminology. *Trends on Analytical Chemistry, 24(6)*, 468-476.

Eriksson, L., Johansson, E., Kettaneh-Wold, N., Trygg, J., Wikstrom, C., & Wold, S. (2006). Basic principles and applications. In Multi- and megavariate data analysis (Part I). Umetrics AB, Umea, Sweden.

Food Information Regulation No 1169, European Commission (2011) on the provision of food information to consumers. Official Journal of the European Union, L 304/18, Luxembourgh.

García-González, D. L., Viera, M., Tena, N., & Aparicio, R. (2007). Evaluation of the methods based on triglycerides and sterols for the detection of hazelnut oil in olive oil. *Grasas y Aceites*, *58(4)*, 344-350.

Graham, S. F., Haughey, S. A., Ervin, R. M., Cancouët, E., Bell, S., & Elliott, C. T. (2012). The application of near-infrared (NIR) and Raman spectroscopy to detect adulteration of oil used in animal feed production. *Food Chemistry*, *132(3)*, 1614-1619.

Guillen, M. D., & Cabo, N. (1997). Characterization of edible oils and lard by Fourier transform infrared spectroscopy. Relationships between composition and frequency of concrete bands in the fingerprint region. *Journal of the American Oil Chemists' Society, 74*, 1281–1286.

Gurdeniz, G., & Ozen, B. (2009). Detection of adulteration of extra-virgin olive oil by chemometric analysis of mid-infrared spectral data. *Food Chemistry*, *116*, 519–525.

Kohavi, R., & Provost, F. (1998). Glossary of terms. *Machine Learning*, 30, 271-274.

Koidis, A. & Osorio-Argüello, M. T. (2013). Identification of oil mixtures in extracted and refined vegetable oils. *Lipid Technology*, *25*, 247–250.

Lerma-Garcia, M. J., Ramis-Ramos, G., Herrero-Martinez, J. M., & Simo-Alfonso, E. F. (2010). Authentication of extra virgin olive oils by Fourier transform infrared spectroscopy. *Food Chemistry, 118,* 78–83.

Osorio, M. T., Haughey, S., Elliott, C., & Koidis, A. (2014). Evaluation of methodologies to determine vegetable oil species present in oil mixtures: Proposition of an approach to meet the EU legislation demands for correct vegetable oils labelling. *Food Research International, 60,* 66–75.

Rohman, A., & Che Man, Y. B. (2010). Fourier transform infrared (FTIR) spectroscopy for analysis of extra virgin olive oil adulterated with palm oil. *Food Research International, 43(3),* 886-892.

Sinelli, N., Cosio, M. S., Gigliotti, C., & Casiraghi, E. (2007). Preliminary study on application of mid infrared spectroscopy for the evaluation of the virgin olive oil "freshness". *Analytica Chimica Acta, 598(1),* 128-134.

Wold, S. & Sjostrom, M. (1977). SIMCA: A method for analyzing chemical data in terms of similarity and analogy. In B.R. Kowalski (Eds.), *Chemometrics theory and application* (pp. 243-282). American Chemical Society Symposium Series 52, Wash., D.C., American Chemical Society.

Wold, S. (1976). Pattern recognition by means of disjoint principal components models, *Pattern Recognition*, *8*, 127-139.

Zielinski, A., Haminiuk, C., Nunes, C., Schnitzler, E., van Ruth, S. M., & Granato, D. (2014). Chemical composition, sensory properties, provenance, and bioactivity of fruit juices as assessed by chemometrics: A critical review and guideline. *Comprehensive Reviews in Food Science and Food Safety, 13,* 300–316.

FIGURES AND TABLES

Table 1. SIMCA and PLS-DA model characteristics on calibration dataset using Raman and FTIR variables on all oil samples.

	Class	R2X * (cumulative)		Q2 ** (cumulative)	
		RAMAN	FTIR	RAMAN	FTIR
SIMCA	P	0.978	0.815	0.966	0.712
	PKO	0.841	0.617	0.789	0.239
	RS	0.960	0.778	0.934	0.761
	PPKO	0.917	0.926	0.883	0.906
	RSPO	0.983	0.937	0.978	0.919
	RSPKO	0.985	0.926	0.979	0.917
PLS-DA	All classes	0.991	0.971	0.592	0.739

Table 2. SIMCA and PLS-DA model performance on prediction dataset using Raman and FTIR (84 samples for Raman and 126 samples for FTIR including replications).

	Target Group	**Correctly classified target samples**		**Correctly classified non-target samples**		**Sensitivity (%)**		**Specificity (%)**		**Overall correct classification rate (%)**	
	Technique	Raman	FTIR	Raman	FTIR	Raman	FTIR	Raman	FTIR	Raman	FTIR
SIMCA	P	8/8	12/12	76/76	114/114	100	100	100	100	100	100
	PKO	1/2	2/3	82/82	123/123	50	66.7	100	100	98.8	99.2
	RS	11/14	19/21	70/70	105/105	78.6	90.5	100	100	96.4	98.4
	PPKO	10/10	11/15	74/74	111/111	100	73.3	100	100	100	96.8
	RSPO	25/34	43/51	50/50	75/75	73.5	84.3	100	100	89.3	93.7
	RSPKO	16/16	22/24	68/68	102/102	100	91.7	100	100	100	98.4
	TOTAL (%)	71/84 85%	109/126 87%							97%	**98%**
PLS-DA	P	7/8	10/12							87.5	83.3
	PKO	2/2	3/3							100	100
	RS	14/14	20/21							100	95.2
	PPKO	10/10	15/15							100	100
	RSPO	33/34	51/51							97.1	100
	RSPKO	14/16	24/24							87.5	100
	TOTAL (%)	80/84 95%	123/126 97%							95%	96%

Table 3. Classification criteria for classification of vegetable oils in 6 classes according to their fatty acid content. Upper number corresponds to the % FA area per total FA; lower number correspond to the absolute FA value expressed as mg fatty acid g^{-1} of sample

Fatty Acids (% total FA and mg fatty acid g^{-1})	**VEGETABLE OIL CLASSES**					
	P	**PKO**	**RS**	**PPKO**	**RSPO**	**RSPKO**
C8:0		> 3.0		> 0.3		> 0.25
Caprylic acid		> 20		> 0.25		> 2.5
C12:0	> 0.13	>48	< 0.01			
Lauric acid	> 0.99	> 300	< 0.1			
C14:0	1.00 – 1.45		< 0.09			
Myristic acid	7.8 – 10.0		< 0.7			
C16:0	43 – 69			≥ 10	8.0 – 41.5	4.9 – 8.5
Palmitic acid	315 – 490			≥ 70	58 – 330	35 – 70
C18:1					≥ 25	
Oleic acid					≥ 195	
C18:2 cn6	5 – 12		18 – 67	3 – 10	9.5 – 56	3 – 60
Linoleic acid	43 – 80		135 – 550	25 – 75	70 – 425	24 – 450
P:S ratio[1]	< 0.25	< 0.04	> 4.0	≤ 0.3	≥ 0.325	

[1] P:S (Polyunsaturated/Saturated) ratio is an index of the polyunsaturated character of the oil and it is calculated using the ratio between C18 polyunsaturated fatty acids and C4-C24 saturated fatty acids.

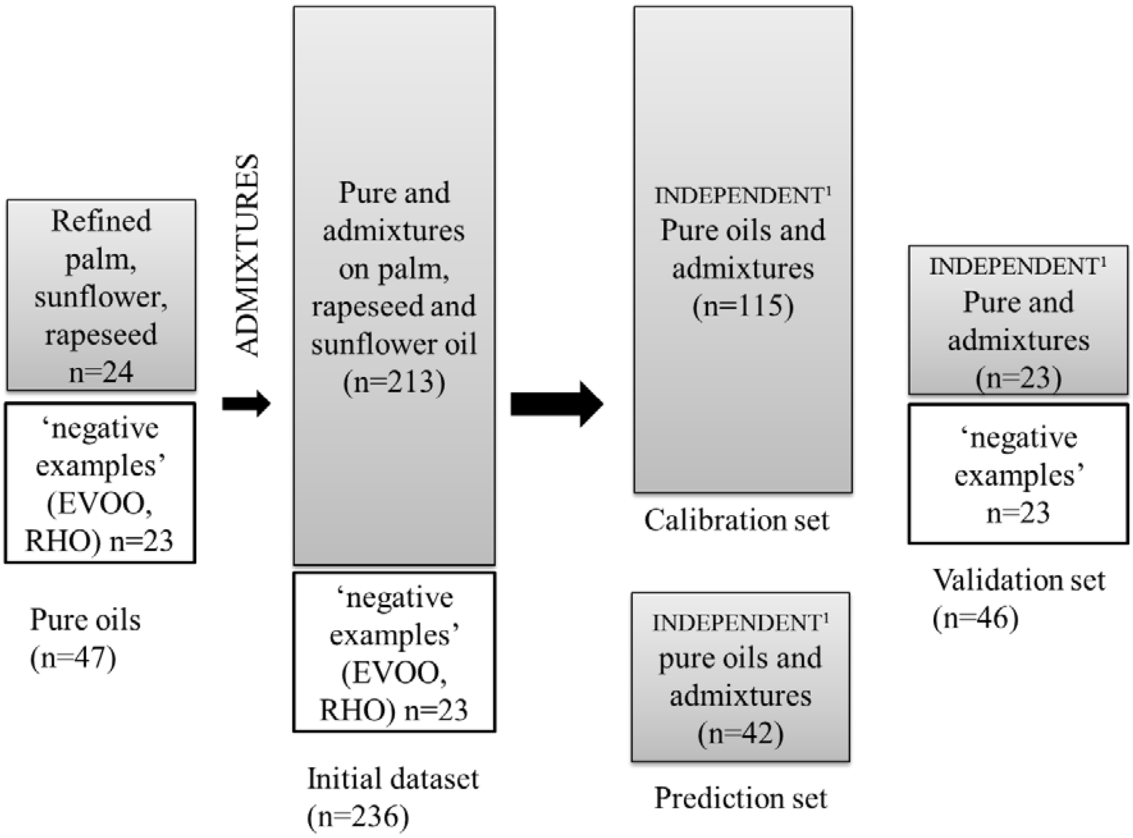

Figure 1. Graphic representation of the dataset used in this study.

[1] independent datasets means that pure and admixture samples in these datasets derive from different initial pure oils (n=23). EVOO: Extra virgin olive oil, RHO: Refined hazelnut oil.

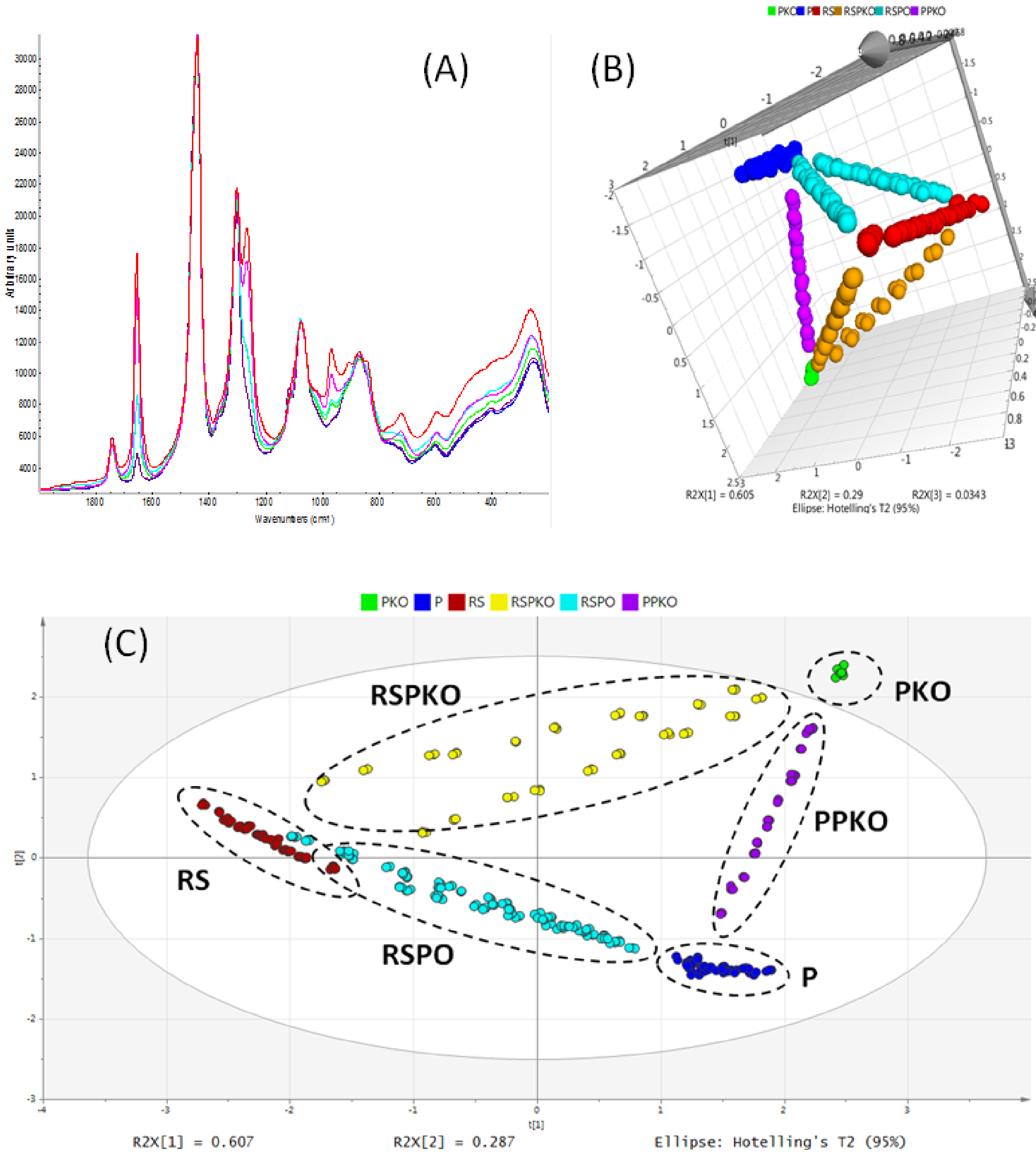

Figure 2. A) Superimposed FTIR spectra of different pure oils B) Principal component analysis scores plot of FTIR data showing the 6 clearly defined oil classes with 3 PCs, C) PCA score plots using 2 PCs . All identified oil classes are shown.

PO: palm oil; POL: palm olein; PS: palm stearin; PKO: palm kernel oil; RO: rapeseed oil; SO: sunflower oil; RS: class containing rapeseed and/or sunflower oil; ROSO: binary admixtures of sunflower and rapeseed oil, P class: class containing pure and admixtures of palm oils and its derivatives, palm olein and palm stearin; PKO class: class containing pure palm kernel oil; PPKO: binary oil admixtures containing oils from PO and PKO classes; RSPKO class: binary oil admixtures containing oils from RS and PKO classes.

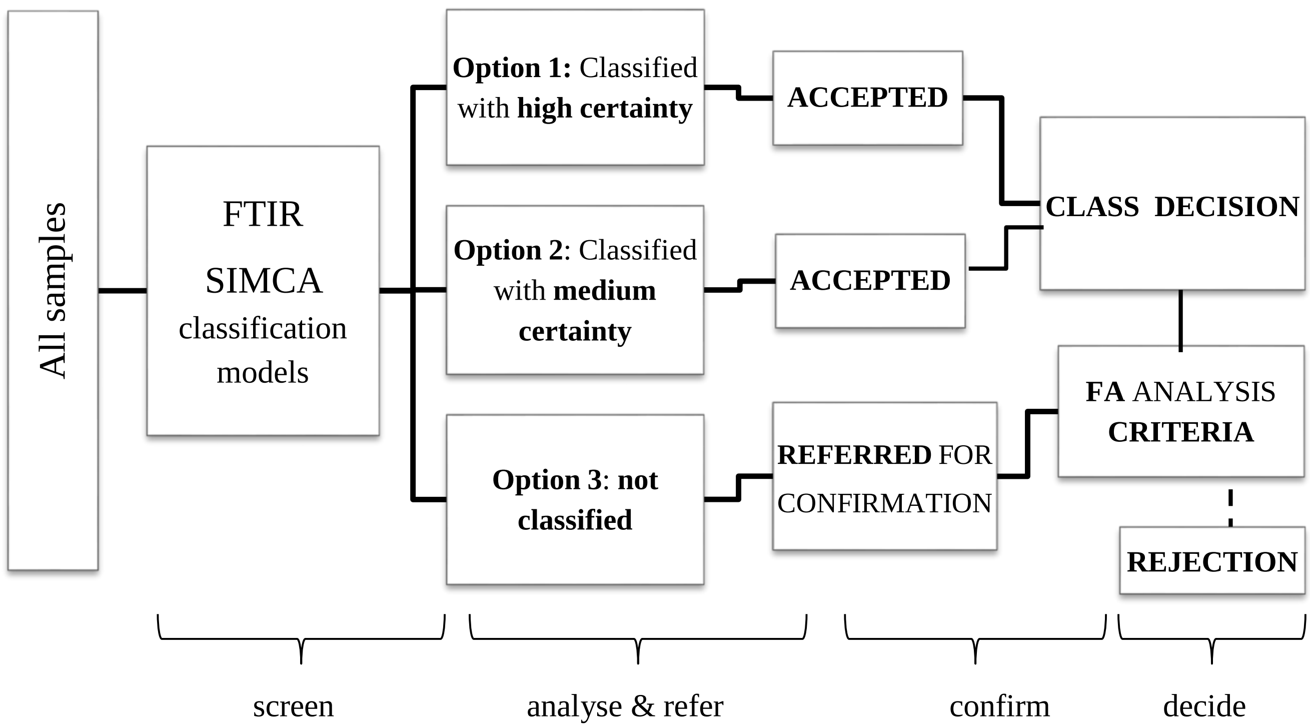

Figure 3. Classification results of the screening stage and referral to the confirmation stage.