



# Comparison of different classification methods for analyzing fluorescence spectra to characterize type and freshness of olive oils

A. Dankowska<sup>1</sup> · W. Kowalewski<sup>2</sup>

Received: 25 June 2018 / Revised: 27 October 2018 / Accepted: 31 October 2018 / Published online: 10 November 2018  
© The Author(s) 2018

## Abstract

The potential of synchronous fluorescence spectroscopic measurements for the classification of olive oil samples with respect to type (extra virgin/refined) and shelf-life condition (expired/non-expired) was examined. Data multidimensionality was reduced by application of the PCA method. Six classification methods, namely, LDA, QDA, RDA, KNN, SVM, and RF, were used to analyze fluorescence data. Classification accuracy was employed to evaluate the performance of these methods for the classification of olive oil samples. The results indicate that fresh and expired olive oils produced with different methods, extra virgin (cold pressed) and refined, are characterized by different fluorescence spectra. KNN and SVM outperform the other classification methods in distinguishing olive oils. The lowest classification error rates were obtained with these two classification methods for measurements at the wavelength interval of 30 nm, and equaled 5.4, 5.7 and 2.9, 3.5% for the four and three classes' models, respectively.

**Keywords** Food adulteration · Food storage · Olive oil authenticity · Synchronous fluorescence spectroscopy · Chemometric analysis

## Abbreviations

E	Expired olive oil	NIR	Near-infrared spectroscopy
EEV	Expired extra virgin olive oil	NMR	Nuclear magnetic resonance
ER	Expired refined olive oil	PCA	Principal component analysis
FTIR	Fourier-transform infrared spectroscopy	QDA	Quadratic discriminant analysis
GC	Gas chromatography	RDA	Regularized discriminant analysis
HPLC	High-performance liquid chromatography	RF	Random forest
KNN	<i>k</i> -nearest neighbors	MCE 80:20	Mean classification error (80–20%)
LDA	Linear discriminant analysis	SF	Synchronous fluorescence
LIBS	Laser-induced breakdown spectroscopy	SVM	Support vector machine
MIR	Mid-infrared spectroscopy	UV-IMS	Ultra violet-ion mobility spectrometry
NEEV	Non-expired extra virgin olive oil	Vis	Visible spectroscopy
NER	Non-expired refined olive oil		

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00217-018-3196-z>) contains supplementary material, which is available to authorized users.

✉ A. Dankowska  
anna.dankowska@ue.poznan.pl

<sup>1</sup> Department of Food Commodity Science, Poznań University of Economics and Business, Al. Niepodległości 10, 61-875 Poznań, Poland

<sup>2</sup> Department of Geoinformation, Adam Mickiewicz University, Bogumiła Krygowskiego 10, 61-680 Poznań, Poland

## Introduction

Extra virgin olive oil is the healthiest and most sought-after of all the categories of olive oil, and it fetches higher prices than other kinds of seed and nut oils. Extra virgin olive oil is a virgin olive oil which has free acidity, expressed as oleic acid, of no more than 0.8 g per 100 g, and whose other characteristics correspond to those fixed for the category in the IOC standard. Extra virgin olive oil does not undergo any refining process. Refined olive oil is the olive oil obtained from virgin olive oils by refining methods which do not lead to alterations in the initial glyceridic structure. It has free

acidity, expressed as oleic acid, of no more than 0.3 g per 100 g and its other characteristics correspond to those fixed for this category in the IOC standard. This kind of olive oil may only be sold directly to the consumer if permitted in a country of retail sale [1].

Adulteration of extra virgin olive oil undermines the confidence of consumers and decreases the profits of honest producers. Therefore, it is one of the main issues for the olive oil industry and thus arises a need for analytical techniques to control the quality and authenticity of virgin olive oils. Several instrumental techniques to classify, authenticate, and control the quality of olive oils have been developed, for example, GC [2, 3], HPLC [4, 5], NMR [6, 7], NIR [9], MIR [9], FTIR spectroscopy [10, 11], Raman spectroscopy [11], fluorescence spectroscopy [12–14], UV-IMS [15], LIBS [16], microwave reflectometry [17], and Vis [18]. Most of these instrumental techniques require harmful reagents or expensive equipment with large operational and maintenance costs. In this context, fluorescence spectrometry is a simpler and less costly alternative. This technique has been successfully applied for the classification of honeys [19], coffees [20] oils [21], wines [22], beers [23], cereal products [24] as well for freshness estimation of such food products as fish, meat, eggs, olive oil, and rapeseed oil [25, 26].

Various statistical methods can be used for classification purposes, such as linear discriminant analysis (LDA) [20, 21], quadratic discriminant analysis (QDA) [27], regularized discriminant analysis (RDA), *k*-nearest neighbors (KNN) [21], support vector machine (SVM) [28], or random forest (RF) [29]. In discriminant analysis, spectra are assigned to definite classes, so that qualitative information complements quantitative spectral data. If the samples are numerous enough, they may be divided into two sets: a training set to construct the classification method (calibration stage) and a testing set to validate it (validation stage). The purpose of the classification method is to obtain weighted combinations of data that minimize variances within classes and maximize variances between classes. Then, the classification rules are used to assign new or unknown samples to the most probable subclasses. Prior to discriminant analysis, principal component analysis is often applied to spectral data sets to reduce data set size and co-linearity. The validity of a classification method can be verified by a comparison of distances or testing.

The objective of the present study is to evaluate the potential of synchronous fluorescence spectrometry for the classification of olive oil samples with respect to type (extra virgin/refined) and shelf-life condition (expired/non-expired) and to compare the accuracy of six chemometric methods for this classification purpose. To the best of our knowledge, it is the first time that the RDA and RF methods are applied to synchronous fluorescence data. Moreover, a comparison of the most frequently used

classification methods for fluorescence spectra reported in the literature can greatly facilitate the selection of the appropriate method.

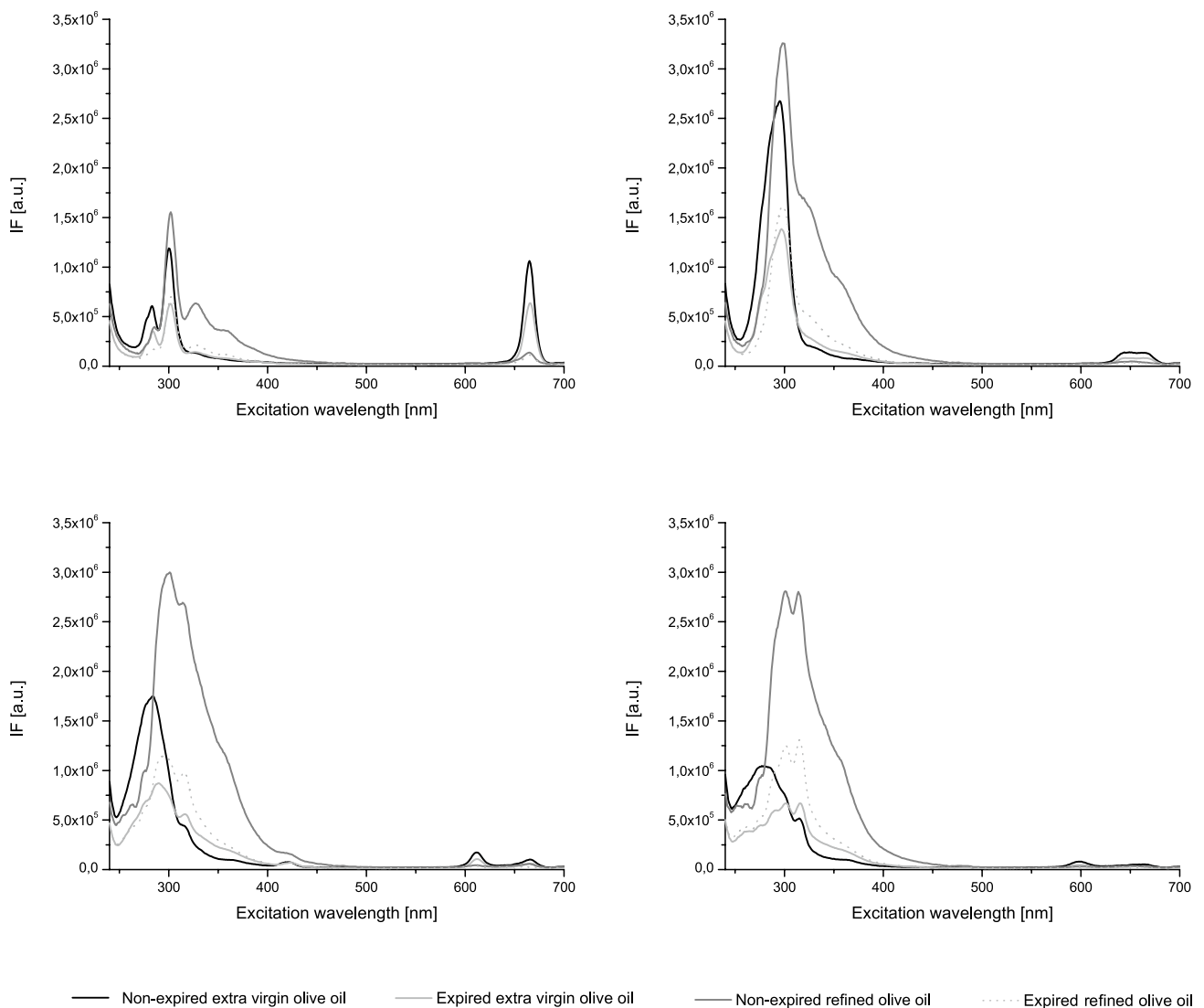
## Experimental

### Characteristics of experimental olive oil samples

A total of 82 samples of extra virgin and refined olive oils (labeled as olive oils and pomace olive oils) were acquired in the local supermarkets. It was verified that fresh samples satisfied the requirements of the IOC organization as to the free fatty acid content, fatty acids profile, and UV absorption at 232 ( $K_{232}$ ) and 270 nm ( $K_{270}$ ) [30]. The samples assigned as fresh samples were kept at room temperature in the dark and analyzed within a short time from their arrival at the laboratory. Samples intended to be analyzed as terminates had been stored in the original commercial container without strict environmental control for 18–24 months to pass the expiration date. Hereafter, these samples will be termed ‘expired’. For the four classes classification study, the following classes were considered: non-expired extra virgin olive oil (NEEV), non-expired refined olive oil (NER), expired extra virgin olive oil (EEV), and expired refined olive oils (ER) with 24, 20, 34, and 14 samples, respectively. In the three classes’ classification study, expired extra virgin and expired refined olive oils were analyzed together as expired olive oils. All reagents used in the experiment were of analytical grade.

### Synchronous fluorescence spectra measurement of experimental olive oil samples

Fluorescence spectra were gathered on a Fluorolog 3–11 spectrofluorometer Spex-Jobin Yvon SA, with a xenon lamp as a source of excitation. The excitation and emission slits were 2 nm wide each. The acquisition interval was 1 nm and the integration time was 0.1 s. The excitation wavelength covered the range of 240–700 nm (461 data points). The oil samples, diluted in *n*-hexane (1% v/v) in a 10 mm fused-quartz cuvette, were examined by means of the right-angle geometry. Synchronous fluorescence spectra were lifted by simultaneously scanning excitation and emission monochromators, with constant wavelength intervals  $\Delta\lambda$  between the excitation and emission wavelengths. For each sample, four spectra were collected, at wavelength intervals of 10, 30, 60, and 80 nm. All measurements were performed in duplicate and reported as mean values. Plots of fluorescence intensities as a function of the excitation wavelength were made.



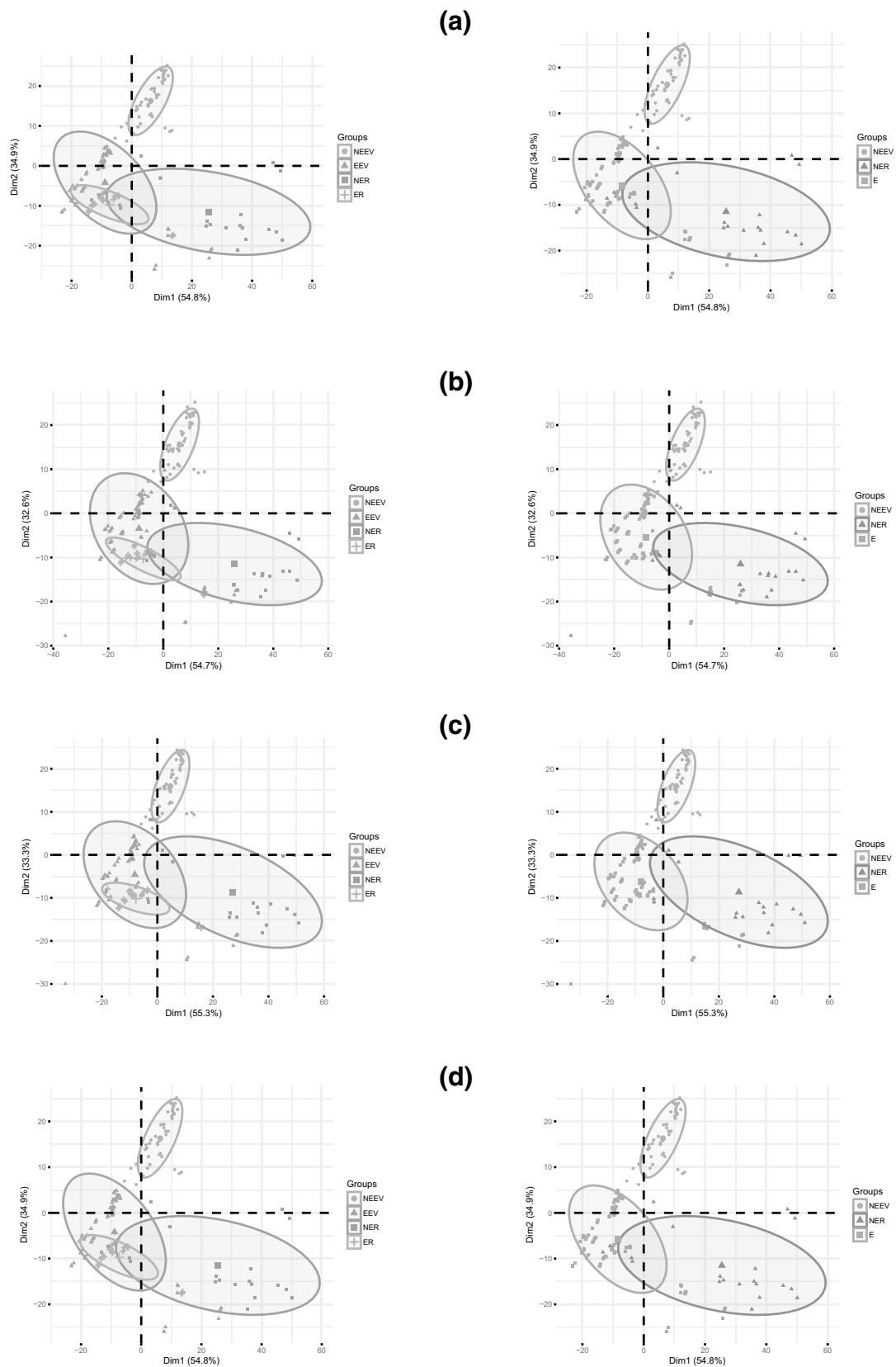
**Fig. 1** Synchronous fluorescence spectra of fresh and expired olive oils (diluted in *n*-hexane 1% v/v)

### Statistical analysis of spectroscopic measurements

Prior to classification analysis, PCA was employed to reduce the number of variables. PCA was followed by six different classification methods: LDA, QDA, RDA, KNN, SVM, and RF. Principal components (PCs) are characterized by a decreasing variance (which is a measure of their linear information capacity), so that the first principal component explains the highest percentage of total observable variable variance. The number of principal components taken for further classification analysis was chosen on the basis of the Kaiser criterion (since PCs with eigenvalues higher than one provide more information than the average single variable) [31]. In the first test, samples were classified into one of the four classes: non-expired extra virgin, non-expired refined, expired extra virgin, and expired refined, and in the second

test into three classes: non-expired extra virgin, non-expired refined, and expired olive oils. In the latter test, expired extra virgin and refined oils were combined into one group and analyzed as oils not fit for consumption.

Classification models were validated using 80–20% split validation with 100 repeats. 92 items out of all samples were 100 times randomly split into two subsets: training or calibration set (80% of all samples) and test set (20%), assuming that the content of both sets does not repeat. Each time all classification methods LDA, QDA, RDA, KNN, SVM, and RF were performed using the training set to estimate the parameters of discriminant functions, and classification error was then calculated on the basis of the test set. The process was repeated 100 times and then mean classification error (RMSEV 80:20) was calculated. There was no need to make all possible analyses with 80–20% split of 92 samples



**Fig. 2** First two PCs plots of principal component analysis obtained for four (4) and three (3) classes models for synchronous fluorescence intensities measured at **a**  $\Delta\lambda = 10$  nm, **b**  $\Delta\lambda = 30$  nm, **c**  $\Delta\lambda = 60$  nm and **d**  $\Delta\lambda = 80$  nm

**Table 1** Classification errors for different classification methods

	4 classes MCE 80:20 (%)				3-classes MCE 80:20 (%)			
	10	30	60	80	10	30	60	80
$\Delta\lambda$ (nm)	10	30	60	80	10	30	60	80
LDA	11.7	10.1	11.0	10.6	12.4	9.4	9.8	11.3
QDA	12.4	10.4	9.2	13.0	11.2	8.7	12.4	11.6
RDA	6.7	8.2	8.7	6.4	5.7	5.2	4.4	5.2
KNN	6.2	5.4	7.2	7.3	5.0	2.9	4.8	6.3
SVM	7.7	5.7	6.0	6.9	4.7	3.2	3.5	4.8
RF	11.0	9.6	12.0	11.0	4.6	4.2	4.2	4.9

as simulations showed that RMSEV 80:20 stabilizes as the number of repeats approaches 100. All statistical analyses were carried out using R, version 3.4.1 patched, a software environment for statistical computing.

## Results and discussion

### Synchronous fluorescence spectra of olive oil samples

Figure 1 shows synchronous fluorescence intensities acquired for olive oil samples as a function of excitation wavelength. The intensity of fluorescence of edible oils depends on, for example, the content of tocopherols, tocotrienols, chlorophylls and pheophytins as well as phenolic compounds [32]. Non-expired extra virgin, non-expired refined, expired extra virgin, and expired refined exhibit differences in fluorescence spectra caused by the different contents of tocopherols, polyphenols, fatty acids, and chlorophylls [33]. A band observed in the range of 270–300 nm was traced back to the emission of tocopherols and tocotrienols, while a band in the range of 660–700 nm is characteristic of chlorophylls and pheophytins *a* and *b* [34]. Refining and similar processes are supposedly responsible for changes in the content and structure of the minor compounds mentioned above as well as for the conjugation of double bonds in fatty acids that enables the distinction of extra virgin from the other categories of olive oils. Extraction methods might also be responsible for some of the differences in olive oil quality, since from the same raw material, different final products are obtained [35]. The intensity and shape of synchronous fluorescence spectra of olive oils depend on the difference between the excitation and emission wavelengths. As a result of oxidation, characteristic changes in synchronous fluorescence spectra are observed: the tocopherol and chlorophyll band intensity decreases, and simultaneously, a new emission band appears in the 320–380 nm wavelength range. These findings are in line with those presented by Sikorska et al. [36].

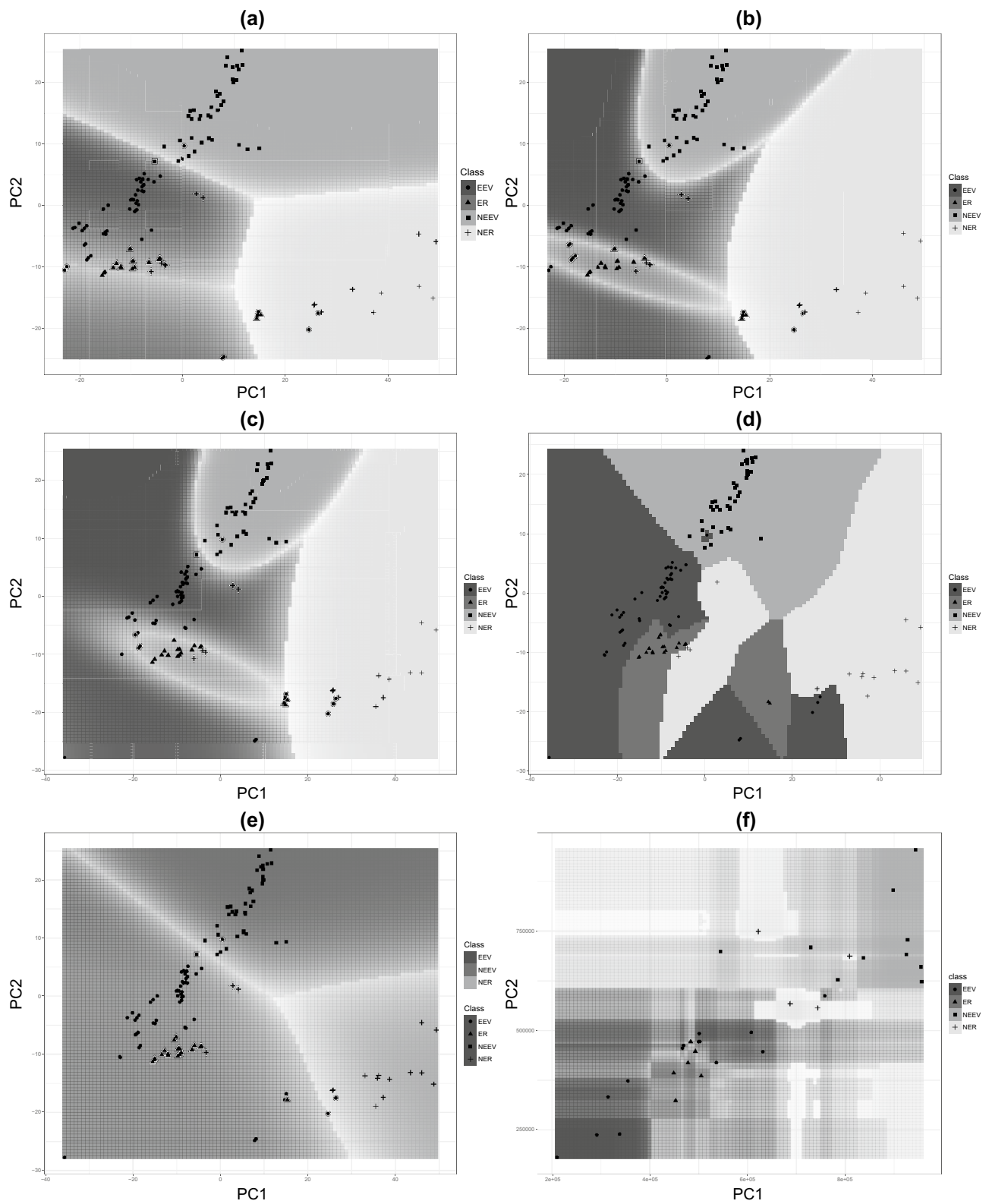
### Principal component analysis of synchronous fluorescence spectra

Figure 2 presents the first two PCs score plot resulting from the application of PCA to the synchronous fluorescence intensities. As can be seen, there is an overlap between the classes, especially regarding the expired extra virgin and expired refined olive oil samples. Principal component analysis (PCA) was employed for exploratory spectral analysis, and subsequently, LDA, QDA, RDA, KNN, SVM, and RF were performed. All selected PCs cumulatively account for over 96% of the total variance for fluorescence intensities measured at each wavelength interval. The number of principal components chosen for further statistical analysis according to the Kaiser [31] criterion equaled six PCs for all wavelength intervals.

### Comparison of different classification analyses of fluorescence data

Six classification methods were applied to the principal components obtained previously by PCA. Classification analyses were carried out separately for synchronous fluorescence data acquired at each wavelength interval (10, 30, 60, and 80 nm). The visualizations of the olive oil samples in the relative coordinated systems show classification performance of all six classification methods of measured samples in four clusters: non-expired extra virgin, non-expired refined, expired extra virgin, and expired refined samples acquired for synchronous fluorescence at  $\Delta\lambda = 30$  nm (Fig. 3). It is interesting to note that for the group formed by expired extra virgin and expired refined olive oils, the samples are very close to each other. Perhaps, this is so because these samples contain some characteristic factor, e.g., products of oxidation. The results presented in Fig. 3 were confirmed by a comparison of classification error rates in three and four classes models obtained by six different methods, as shown in Table 1.

The best discrimination ability among the classification models, except for the ones obtained with the RDA method, was observed for the synchronous fluorescence measurements obtained at  $\Delta\lambda = 30$  nm with classification errors in



**Fig. 3** Classification plots of synchronous fluorescence intensities acquired at  $\Delta\lambda = 30$  nm, for **a** LDA, **b** QDA, **c** RDA, **d** KNN, **e** SVM, and **f** RF

the range from 5.4 to 10.4 and from 2.9 to 9.4% for four classes and three classes models, respectively. The results obtained with the RDA method were slightly better for measurements obtained at  $\Delta\lambda = 80$  nm and 60 nm for four classes' and three classes' models, respectively.

The errors obtained for the three classes models in most cases were lower than for the four classes models. Perhaps, oxidation products formed during storage are the reason why, in the group formed by expired extra virgin and expired refined olive oils, and samples are very similar to each other and should be analyzed jointly as expired oils.

Comparing the results between classification methods, it was found that the best classification results among all methods were obtained with the KNN and SVM methods with synchronous fluorescence measurements acquired at wavelength intervals of 30 nm, and classification errors of 5.4, 5.7% (four classes) and 2.9, 3.2% (three classes) for the KNN and SVM method, respectively. LDA and QDA analyses did not allow for very good classification performance with classification errors not lower than 9.4 and 8.7%, respectively. The results obtained with the RDA method were satisfactory with the lowest classification errors equaling 6.4 and 4.4% for the four classes and three classes models, respectively.

The results obtained in this study are in agreement with the findings obtained by Wu et al. [27], who established that RDA analysis always gives results equivalent to or better than LDA and QDA, and Zheng et al. [28] who achieved a comparable and very satisfying accuracy for classification of food products while using the KNN and LS-SVM methods. However, the authors have established that KNN, as opposed to the latter method, is very sensitive to the nature of spectral data. The KNN method performed rather poorly on the fruit data set.

## Conclusion

The study has shown that non-expired extra virgin, non-expired refined, expired extra virgin, and expired refined olive oil samples exhibit significant differences in their synchronous fluorescence spectral patterns. The KNN and SVM methods clearly outperformed the LDA, QDA, RDA, and RF methods. Moreover, the best classification accuracy was obtained for the fluorescence intensities measured at 30 nm wavelength interval. The lowest classification rates were obtained with the KNN and SVM classification methods for measurements acquired at wavelength interval of 30 nm and equaled 5.4, 5.7% and 2.9, 3.5% for four and three classes' models, respectively. The findings provide a technical tool for consumer protection against profit-driven deceitful practices in the food market by contributing to a better freshness control and more efficient detection of olive oil fraud.

## Compliance with ethical standards

**Conflict of interest** Anna Dankowska declares that she has no conflict of interest. Wojciech Kowalewski declares that he has no conflict of interest.

**Compliance with ethics requirements** This article does not contain any studies with human or animal subjects.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. IOC (2015) Trade standard applying to olive oils and olive-pomace oils. International Olive Council, Geneva
2. Troya F, Lerma-García MJ, Herrero-Martínez JM, Simó-Alfonso EF (2015) Classification of vegetable oils according to their botanical origin using n-alkane profiles established by GC–MS. *Food Chem* 167:36–39
3. Li X, Kong W, Shi W, Shen Q (2016) A combination of chemometrics methods and GC–MS for the classification of edible vegetable oils. *Chemometr Intell Lab* 155:145–150
4. Fasciotti M, Pereira Netto AD (2010) Optimization and application of methods of triacylglycerol evaluation for characterization of olive oil adulteration by soybean oil with HPLC–APCI–MS–MS. *Talanta* 81:1116–1125
5. Arslan D, Özcan MM (2011) Phenolic profile and antioxidant activity of olive fruits of the Turkish variety “Sariulak” from different locations. *Grasas Aceites* 62:453–461
6. Vlahov G, Del Re P, Simone N (2003) Determination of geographical origin of olive oils using <sup>13</sup>C nuclear magnetic resonance spectroscopy. I—Classification of olive oils of the puglia region with denomination protected origin. *J Agric Food Chem* 51:5612–5615
7. Rezzia S, Axelson DE, Héberger K, Renieroa F, Marianid C, Guilloua C (2005) Classification of olive oils using high throughput flow <sup>1</sup>H NMR fingerprinting with principal component analysis, linear discriminant analysis and probabilistic neural networks. *Anal Chim Acta* 552:13–24
8. Mannina L, Segre A (2002) High resolution nuclear magnetic resonance: from chemical structure to food authenticity. *Grasas Aceites* 53:22–33
9. Sinelli N, Cerretani L, di Egidio V, Bendini A, Casiraghi E (2010) Application of near (NIR) infrared and mid (MIR) infrared spectroscopy as a rapid tool to classify extra virgin olive oil on the basis of fruity attribute intensity. *Food Res Int* 43:369–375
10. Maggio RM, Cerretani L, Chiavaro E, Kaufman TS, Bendin A (2010) A novel chemometric strategy for the estimation of extra virgin olive oil adulteration with edible oils. *Food Control* 21:890–895
11. Jiménez-Carvelo AM, Osorio MT, Koidis A, González-Casado A, Cuadros-Rodríguez L (2017) Chemometric classification and quantification of olive oil in blends with any edible vegetable oils using FTIR-ATR and Raman spectroscopy. *LWT Food Sci Technol* 86:174–184
12. Guimet F, Boqué R, Ferré J (2006) Application of nonnegative matrix factorization combined with Fisher's linear discriminant

- analysis for classification of olive oil excitation–emission fluorescence spectra. *Chemom Intell Lab Sys* 81:94–106
13. Poulli KI, Mousdis GA, Georgiou CA (2005) Classification of edible and lampante virgin olive oil based on synchronous fluorescence and total luminescence spectroscopy. *Anal Chim Acta* 542:151–156
  14. da Silva CET, Filardi VL, Pepe IM, Chaves MA, Santos CMS (2015) Classification of food vegetable oils by fluorimetry and artificial neural networks. *Food Control* 47:86–91
  15. Garrido-Delgado R, Muñoz-Pérez ME, Arce L (2018) Detection of adulteration in extra virgin olive oils by using UV-IMS and chemometric analysis. *Food Control* 85:292–299
  16. Mbesse Kongbonga YG, Ghalila H, Onana MB, Ben Lakhdar Z (2014) Classification of vegetable oils based on their concentration of saturated fatty acids using laser induced breakdown spectroscopy (LIBS). *Food Chem* 147:327–331
  17. Cataldo A, Piuze E, Cannazza G, De Benedetto E (2012) Classification and adulteration control of vegetable oils based on micro-wave reflectometry analysis. *J Food Eng* 112:338–345
  18. Pizarro C, Rodríguez-Tecedor S, Pérez-del-Notario N, Esteban-Díez I, González-Sáiz JM (2013) Classification of Spanish extra virgin olive oils by data fusion of visible spectroscopic fingerprints and chemical descriptors. *Food Chem* 138:915–922
  19. Karoui R, Dufour E, Bosset JO, De Baerdemaeker J (2007) The use of front face fluorescence spectroscopy to classify the botanical origin of honey samples produced in Switzerland. *Food Chem* 101:314–323
  20. Dankowska A, Domagała A, Kowalewski W (2017) Quantification of *Coffea arabica* and *Coffea canephora* var. *robusta* concentration in blends by means of synchronous fluorescence and UV–Vis spectroscopies. *Talanta* 172:215–220
  21. Sikorska E, Górecki T, Khmelinskii IV, Sikorki M, Kozioł J (2005) Classification of edible oils using synchronous scanning fluorescence spectroscopy. *Food Chem* 89:217–225
  22. Dufour E, Letort A, Laguet A, Lebecque A, Serra JN (2006) Investigation of variety, typicality and vintage of French and German wines using front-face fluorescence spectroscopy. *Anal Chim Acta* 563:292–299
  23. Tan J, Li R, Jiang ZT (2015) Chemometric classification of Chinese lager beers according to manufacturer based on data fusion of fluorescence, UV and visible spectroscopies. *Food Chem* 184:30–36
  24. Karoui R, Cartaud G, Dufour E (2006) Front-face fluorescence spectroscopy as a rapid and non-destructive tool for differentiating various cereal products: a preliminary investigation. *J Agric Food Chem* 54:2027–2034
  25. El Masry G, Nagai H, Moria K, Nakazawa N, Tsuta M, Sugiyama J, Okazaki E, Nakauchi S (2015) Freshness estimation of intact frozen fish using fluorescence spectroscopy and chemometrics of excitation-emission matrix. *Talanta* 143:145–156
  26. Dankowska A (2016) Advances in fluorescence emission spectroscopy for food authenticity testing. In: Downey G (eds) *Advances in food authenticity testing. Improving quality throughout the food chain*. Elsevier, New York
  27. Wu W, Mallet Y, Walczak B, Penninckx W, Massart DL, Heuerding S, Erni F (1996) Comparison of regularized discriminant analysis linear discriminant analysis and quadratic discriminant analysis applied to NIR data. *Anal Chim Acta* 329:257–265
  28. Zheng W, Fu X, Ying Y (2014) Spectroscopy-based food classification with extreme learning machine. *Chemometr Intell Lab* 139:42–47
  29. Ai F, Bin J, Zhang Z, Huan J, Yang Z (2014) Application of random forests to select premium quality vegetable oils by their fatty acid composition. *Food Chem* 143:472–478
  30. Commission Regulation (EEC) 2568/91 (1991) On the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis
  31. Kaiser HF (1960) The application of electronic computers to factors analysis. *Educ Psychol Meas* 141–151
  32. Sayago A, Morales MT, Aparicio R (2004) Detection of hazelnut oil in virgin olive oil by a spectrofluorimetric method. *Eur Food Res Technol* 218:480–483
  33. Christensen J, Nørgaard L, Bro R, Engelsen SB (2006) Multivariate autofluorescence of intact food systems. *Chem Rev* 106:1979–1994
  34. Sikorska E, Romaniuk A, Khmelinskii IV, Herance R, Bourdelande JL, Sikorski M, Kozioł J (2004) Characterization of edible oils using total luminescence spectroscopy. *J Fluoresc* 14:25–35
  35. Vaz Freire L, Gouveia JM, Costa Freitas AM (2008) Analytical characteristics of olive oils produced by two different extraction techniques, in the Portuguese olive variety Galega Vulgar. *Grasas Aceites* 59:260–266
  36. Sikorska E, Romaniuk A, Khmelinskii I, Sikorski M, Kozioł J (2003) Characterization of edible oils using synchronous fluorescence scanning fluorescence spectroscopy. *Pol J Food Nutr Sci* 53:108–112