Discrimination of edible olive oils by means of synchronous fluorescence spectroscopy with multivariate data analysis

By A. Dankowska*, M. Małecka, and W. Kowalewski

*Faculty of Commodity Science, Poznań University of Economics, Poznań, Poland
Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Poznań, Poland
Corresponding author: anna.dankowska@ue.poznan.pl

SUMMARY
Discrimination of edible olive oils by means of synchronous fluorescence spectroscopy with multivariate data analysis

The potential of fluorescence spectroscopy for the classification of olive oils was investigated. Synchronous fluorescence spectra were collected in the region of 240-700 nm with the wavelength intervals of 10, 30, 60 and 80 nm. Successive projection algorithm (SPA) was applied for the determination of representative wavelengths while the linear discriminant analysis (LDA) method was used to classify olive oils. The classification error of the method was low (0.9-6.4%) for measurements collected at all wavelength intervals. The best classification accuracy was obtained for the measurements collected at 10 selected wavelengths with the wavelength interval equal to 10 nm.


1. INTRODUCTION
According to the International Olive Council, the basic grades of olive oils suitable for consumption are: virgin olive oil, olive oil and olive pomace oil (IOC, 2011). Virgin olive oil is obtained from the fruit of the olive tree solely by mechanical or physical means under conditions that do not lead to alterations in the oil. Only minimal treatment such as washing, pressing, decantation, centrifugation and filtration is acceptable. The acidity of extra virgin olive oil can not exceed 0.8% (w/w) of free fatty acid content calculated as oleic acid and the oil must possess perfect sensory characteristics. Oils labeled as olive oil are usually a blend of refined olive oil and small amounts of virgin olive oil. Olive-pomace oil, the lowest grade of olive-based oils, is a blend of refined pomace olive oil and a small amount of virgin oil fit for consumption (IOC, 2011).

Extra virgin olive oil is the most valuable and health promoting of all the categories of olive oils and is more expensive than the major kinds of seed and nut oils. Adulteration of the extra virgin olive oil undermines the confidence of consumers and also decreases the profit of scrupulous producers. Therefore, it is one of the main issues for the olive oil industry. Thus, the need for analytical techniques to control the quality and authenticity of the virgin olive oils arises. Several instrumental techniques to classify, authenticate and control the quality of olive oil samples have been developed, for example: GC (Dourtoglou et al., 2003; Al-Ismail et al., 2010; Fernandes-Silva et al., 2013), HPLC (Cichelli and Pertesana, 2004; Fasciotti and Pereira Netto, 2010; Arslan and Özcan, 2011), NMR (Zamora et al., 1994; Vlahov et al., 2003; Rezzia et al., 2005, NIR (Mannina and Segre, 2002), MIR (Sinelli et al., 2014), FTIR spectroscopy (Maggio et al., 2010), fluorescence spectroscopy (Poulli et al., 2005, 2006; Guimet et al., 2006), and PCR/CE-SSCP (Wu et al., 2011). Most of those techniques require expensive reagents and are very time consuming. Synchronous fluorescence spectroscopy is an alternative, as it is a simpler, less costly and quicker method than the major techniques widely used (Sikorska et al., 2005, Poulli et al., 2006, Dankowska et al., 2009).
between excitation and emission wavelength maintained. Narrowing the spectral bands, synchronous fluorescence permits simplification of the spectra and reduction of the spectral overlaps. Synchronous fluorescence can accomplish greater selectivity than the conventional fluorescence technique. Moreover, synchronous fluorescence can be carried out using conventional fluorometers with little or no modification to the instrument (Patra and Mishra, 2002).

The application of LDA as well as many other chemometric methods e.g. multiple linear regression (MLR), principal component regression (PCR), partial least squares (PLS) to spectrophotometric multicomponent analysis requires spectral variable selection for building well-fitted models. LDA is a probabilistic classification method which searches for directions (canonical variables) with maximum separation among categories (Massart et al., 1998). It is a challenge and a crucial issue to select the proper analytical wavelengths from a spectrum. The successive projections algorithm (SPA) is a new approach, carried out for selecting effective wavelength variables from the full spectrum. SPA performs simple operations in a vector space to determine a subset of variables with small collinearity. SPA has been compared to the genetic algorithm, which is a popular method for variable selection in multivariate calibration. The results were proven to be in favor of SPA (Ugolino Araújo et al., 2001, Pontes et al., 2005). It was found to be less sensitive to instrumental noise than the genetic algorithm SPA-MLR models, which have been proven to be comparable to or better than the full spectrum PLS or PCR models for UV-Vis (Ugolino Araújo et al., 2001) and NIR spectra analysis (Kawakami Harrop Galvão et al., 2001). Promising results were obtained for SPA-LDA as a method for dealing with low-resolution UV-Vis spectra with strong overlapping (Pontes et al., 2005). Recently, SPA was applied for variable selection to determine the organic acid of plum vinegar (Vis/NIR) (Liu and He, 2009), classification of coffees (UV-Vis) (Polari Souto et al., 2010), and edible oils (UV-Vis) (Pontes et al., 2005).

The aim of this study was to evaluate the potential of the synchronous fluorescence spectroscopy and successive projection algorithm combined with linear discriminant analysis (SPA-LDA) for the discrimination of different types of olive oils. To the best of the authors’ knowledge, it has been the first attempt at choosing wavelengths from the synchronous fluorescence spectra of edible oils using successive projection algorithm (SPA).

2. MATERIALS AND METHODS

2.1. Chemical reagents and samples

Ten samples of extra virgin olive oil were obtained from the IOC organization (3 from Portugal, 3 from Spain, 2 from Greece, 1 from Turkey and 1 from Israel). Another 36 samples of extra virgin olive oils, 12 olive oils and 12 pomace olive oils were purchased in local shopping centers (40 of the samples were bottled in Italy, 14 in Spain, while 6 samples came from Greece). All the samples met the requirements of the International Olive Council with respect to free fatty acid content, fatty acid profile and UV absorption at 232 (K232) and 270 nm (K270) established for the respective olive oil types. The samples were stored in the dark at room temperature and analyzed shortly after their arrival to the laboratory. The experimental samples were prepared in duplicate. All reagents used in the experiment were of analytical grade.

2.2. Synchronous fluorescence spectra measurement

Fluorescence spectra were obtained on a Fluorolog 3-11 spectrofluorometer, Spex-Jobin Yvon SA where a Xenon lamp provides the source of excitation parameters of the analyses, which were set as in Sikorska et al. (2005). Excitation and emission slit widths equaled 2 nm. The acquisition interval and the integration time were maintained at 1 nm and 0.1 s, respectively. The excitation wavelength ranged from 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 acquired by simultaneous scanning of the excitation and the emission monochromators, with constant distances λλ between the excitation and emission wavelengths. Four spectra were collected for each sample, at the wavelength intervals of 10, 30, 60 and 80 nm. All analyses were carried out in duplicate, and the results were reported as mean values. Fluorescence intensities were plotted as a function of the excitation wavelength.

2.3. Statistical analysis

Principal component analysis (PCA) and linear discriminant analysis (LDA) were performed using Statistica 8.0 (StatSoft Inc, Tulsa, US). PCA was used to visualize differences in olive oils on the basis of the whole spectrum, while LDA permitted the classification of the different types of olive oils on the basis of selected data. The leave-one-out cross validation method was applied to evaluate LDA models. Successive projection algorithm (SPA) was coded in C++. SPA is a forward variable selection for multivariate calibration, which employs operations in a vector space to obtain variables with the smallest collinearity. It starts with one wavelength, then incorporates a new one at each iteration, until a specified number N of wavelength is reached. The new variable selected is the one from the remaining variables that has the maximum projection value on the orthogonal subspace of the previous selected variable. Its purpose is to select
wavelengths whose information content is minimally redundant in order to solve the co-linearity problem. The initial variable and number of variables can be given as input information or can be determined on the basis of the smallest root mean squared error of prediction in the validation set of the calibration model. A detailed description of SPA is provided by Ugolino Araújo et al. (2001).

In this experiment, the number of wavelengths to be selected (6 and 10) and the first wavelength - 666 nm, (characteristic for chlorophyll) were given as the input information for all wavelength intervals ($\Delta \lambda = 10, 30, 60$ and 80 nm). Selection of the first wavelength has chemical support, because the compounds from the chlorophyll group reach their maximum peak at around 665-666 nm. Because the number of wavelengths to be selected was determined, it was possible to compare separation abilities of the data obtained at different wavelength intervals.

3. RESULTS AND DISCUSSION

3.1. Synchronous fluorescence spectra of olive oil samples

The intensity of fluorescence of edible oils depends on, for example, the content of tocopherols, tocotrienols, chlorophylls and pheophytins as well as phenolic compounds (Sayago et al., 2004). The band observed in the range of 270-300 nm has been ascribed to the emission of tocopherols and tocotrienols, while 660-700 nm is characteristic of the fluorescence of chlorophylls and pheophytins a and b (Sikorska et al., 2004). The occurrence of chlorophyll derivatives is characteristic of virgin olive oil and can be used as a marker for the detection of virgin olive oil in e.g. argan oil (Querci et al., 2010). Processes such as refining are responsible for changes in the content and structure of the minor compounds mentioned as well as for the conjugation of double bonds in fatty acids enabling the distinction between extra virgin and other categories of olive oil. Extraction methods might be also responsible for some of the differences observed in olive oil quality, as for the same raw material different final products are obtained (Vaz Freire et al., 2008).

The data presented in Figure 1 indicate the potential of synchronous fluorescence for the discrimination among different categories of olive oils. As Figure 1 presents, olive oils show differences in fluorescence intensities, which makes it possible to distinguish among them on the basis of the fluorescence spectra. In order not to make Figure 1 illegible, the results for the mean spectra of each class of olive oil acquired at different wavelength intervals have been presented.

The intensity and shape of synchronous fluorescence spectra of olive oils depends on the difference between the excitation and emission wavelength. For the wavelength interval $\Delta \lambda = 10$ nm we can observe a narrow band around 300 nm related to the content of tocopherols. For the higher wavelength intervals this band shifts to the blue and broadens. The spectra measured for $\Delta \lambda = 10$ nm exhibit an intense band in the synchronous fluorescence spectrum with a maximum peak at around 665 nm, related to the content of compounds from the chlorophyll group. At higher $\Delta \lambda$ this band becomes less intense and at $\Delta \lambda = 60$ and 80 nm the two bands of relatively low intensities can be noticed at around 610, 670 nm and 600, 670 nm, respectively.

3.2. Principal component analysis

Figure 2 presents the PCA plots (PC1*PC2) of synchronous fluorescence spectra obtained for all the olive oil samples at different wavelength intervals. PCA discrimination ability was similar regardless of the wavelength interval ($\Delta \lambda$) applied. Pomace olive oils formed a cluster, essentially distant from the extra virgin olive oils; regardless of the wavelength interval used, the olive oils formed two groups, one close to the pomace olive oil cluster. The second group significantly overlaps with the group of extra virgin olive oils. PCA analysis of olive oils’ spectra indicates the potential of synchronous fluorescence for the discrimination among different categories of olive oils. According to the results obtained by Poulli et al. (2005) synchronous fluorescence spectroscopy enables distinction between virgin olive oils and lampante olive oils.

3.3. Wavelength selection by successive projection algorithm

In this experiment, successive projection algorithm was applied to retain the most informative wavelengths from the spectra for further chemometric analysis. Former attempts to choose analytical wavelengths from the synchronous fluorescence spectra of edible oils had exclusively had chemical support (Sikorska et al., 2005; Poulli et al., 2005, 2006).

According to the results of the successive projections algorithm analysis of the synchronous fluorescence spectra, 6 and 10 wavelengths for each of the wavelength intervals were chosen for further analysis: 282, 296, 300, 302, 303, 309, 321, 322, 326, 666 nm ($\Delta \lambda = 10$ nm); 291, 300, 301, 306, 308, 326, 327, 340, 364, 666 nm ($\Delta \lambda = 30$ nm); 294, 298, 302, 304, 311, 313, 317, 324, 326, 666 nm ($\Delta \lambda = 60$ nm); 263, 274, 277, 294, 296, 298, 302, 312, 314, 666 nm ($\Delta \lambda = 80$ nm). Wavelengths selected for the 6-variable models were comprised in the wavelengths used for the 10-variable ones, as the SPA algorithm starts with one wavelength, then incorporates a new one at each iteration, until a specified number of the most informative wavelength has been reached. Wavelengths, marked bold, were used as input
data for the 6-variable-LDA models, as well as for the 10-variable ones. Wavelengths selected by the SPA for each of the wavelength intervals are indicated by the stars in Figure 1.

3.4. Linear discriminant analysis of synchronous fluorescence intensities at selected wavelengths

Linear discriminant analysis was applied to fluorescence intensities of olive oil samples, previously selected by the successive projections algorithm. The SPA-LDA plots are presented in Figure 2.

LDA analysis was provided separately for the data acquired at each wavelength interval (10, 30, 60, 80) and for 6 or 10 variables. The plots of the first two discriminant functions (DF1*DF2) indicate good classification performance of SPA-LDA (Fig. 3). The separation of three classes of olive oils is better as compared to PCA analysis (Fig. 2). The results of LDA and PCA suggest that the discriminatory information conveyed by the full spectrum was also present in the selected variables. Table 1 presents the percentage of classification errors (the number of misclassified samples divided by the total number of samples) for SPA-LDA models for different wavelength intervals. The lowest classification errors 2.7% and 0.9% (for 6 and 10 selected variables) were obtained for the measurements acquired at the wavelength interval equal to 10 nm.

Models with 10 variables had lower or equal classification errors when compared to those with 6 variables. This suggests that an additional variable added to 10-variable models may improve their discriminatory power although Pontes et al. (2005) show that this discriminatory power may be improved only to a certain extent. At some point, adding an additional variable would cause the deterioration of the model.

4. CONCLUSIONS

Synchronous fluorescence is a simple technique, as the spectra are acquired from the samples
The results of the experiment have shown that the types of olive oil samples exhibit significant differences in fluorescent intensities. Synchronous fluorescence spectroscopy combined with successive projections algorithm and linear discriminant analysis can be successfully applied for the classification of olive oils. The lowest classification errors, 2.7% and 0.9%, were obtained for the measurements acquired at the 10 nm wavelength interval for 6 or 10 variables selected from the spectra, respectively. The findings presented here contribute to controlling the authenticity of olive oil. Synchronous fluorescence method was proven to be a useful technique for the rapid classification of olive oil categories.

Table 1

<table>
<thead>
<tr>
<th>Number of variables</th>
<th>Wavelength interval [nm]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\Delta\lambda = 10)</td>
</tr>
<tr>
<td>LDA6</td>
<td>6</td>
</tr>
<tr>
<td>Coss validation LDA6(^a)</td>
<td>2.7</td>
</tr>
<tr>
<td>LDA10</td>
<td>10</td>
</tr>
<tr>
<td>Cross validation LDA10(^a)</td>
<td>1.8</td>
</tr>
</tbody>
</table>

\(^a\) Leave one-out cross validation method was applied.
Figure 3
SPA-LDA plots for 6 (a) or 10 (b) synchronous fluorescence intensities acquired at different wavelength intervals: \( \Delta \lambda = 10 \) (A), \( \Delta \lambda = 30 \) (B), \( \Delta \lambda = 60 \) (C), \( \Delta \lambda = 80 \) (D) (\( \square \) extra virgin olive oil, + olive oil, ▲ pomace olive oil).
REFERENCES


