Assessment of sesame oil fatty acid and sterol composition with FT-NIR spectroscopy and chemometrics

İbrahim Sani ÖZDEMİR, Öznur KARAOĞLU, Çağdaş DAĞ, Somer BEKİROĞLU

1. Introduction

As an important industrial crop, sesame (Sesamum indicum L.) plant is cultivated globally in tropical and subtropical regions for its oily seed. Apart being used for oil production, sesame seeds are also used as a condiment, either roasted or not, in dishes and patisserie products of different cuisines.

In Turkey, sesame seed is mainly used in production of tahini halva, biscuits, patisseries, and simit (a circular bread product). Several sesame varieties and ecotypes, which have been cultivated for centuries, are distributed in various ecological regions of Turkey (Baydar et al., 1999). According to the Turkish Statistical Institute, in 2017, 18,417 t of sesame was produced in Turkey, 51.7% of which was from the provinces of Antalya, Muğla, and Manisa (TÜİK, 2017).

Sesame seed contains approximately 40%–60% oil, which is mainly composed of mono- and polyunsaturated fatty acids accounting for almost 85% of the total fatty acids (Ünal and Yağcı, 2008). Sesame is also a rich source of protein (25%) and several health-promoting compounds such as phytosterols, tocopherols, and lignans (Miraj and Kiani, 2016). This peculiar biochemical composition also makes sesame oil one of the most resistant vegetable oils against oxidation (Tan et al., 2002). Therefore, it is of prime importance to analyze these bioconstituents as a part of quality control during procurement, storage, and processing and for the selection of varieties during breeding studies. However, the classical analysis techniques used for the quantification of these constituents are time-consuming, and expensive laboratory equipment and skilled personnel are often required, which makes them inappropriate for routine quality checks.

In this essence, Fourier transform-near infrared (FT-NIR) spectroscopy coupled with appropriate chemometric techniques can be considered as an alternative technique for the rapid, simple, and simultaneous analysis of the major lipid constituents of sesame oil, which can be used effectively in quality control and breeding studies of sesame.

Key words: Chemometrics, fatty acids, near infrared spectroscopy, sesame oil, sterols

Abstract: In this study, the Fourier transform-near infrared (FT-NIR) spectroscopy technique was used to predict the fatty acid and sterol content of sesame. For this purpose, partial least square regression (PLS-R)-based prediction models were developed, which relate the FT-NIR spectra to reference GC measurements. In total, 39 sesame samples were collected from local producers around Muğla Province. Sesame oil was extracted from the seeds by using a screw press and extracted oil samples were analyzed without any refining. The results showed that among different fatty acids found in sesame oil, oleic and linoleic acid contents (which account for 85% of the total fatty acids) can be precisely predicted with corresponding PLS-R models having $R^2 = 0.991$, $RMSECV = 0.092\%$, $RPD = 10.7$ and $R^2 = 0.988$, $RMSECV = 0.118$, $RPD = 9.01$, respectively. Similarly, good model performances were obtained for the fatty acids grouped according to their saturation degree, namely saturated, monounsaturated, and polyunsaturated fatty acids with $R^2$, $RMSECV$, and $RPD$ values in the ranges of 0.865–0.976, 0.148%–0.148%, and 2.72–6.41, respectively. In addition, models with moderate quality for the β-sitosterol and Δ5-avenasterol content of the sesame oils could be established with $R^2 = 0.756$, $RMSECV = 0.651\%$, $RPD = 2.04$ for β-sitosterol and $R^2 = 0.823$, $RMSECV = 0.343\%$, $RPD = 2.38$ for Δ5-avenasterol content models. In conclusion, FT-NIR spectroscopy was proved to be a valuable analytical technique that enables rapid and simultaneous measurement of the major lipid constituents of sesame oil, which can be used effectively in quality control and breeding studies of sesame.

Received: 27.02.2018 ● Accepted/Published Online: 30.05.2018 ● Final Version: 11.12.2018

This work is licensed under a Creative Commons Attribution 4.0 International License.
general, a typical FT-NIR spectrum is composed of large absorption peaks, which cannot be assigned to specific molecules and exhibit only small differences at first sight. However, the information captured by the spectra can be extracted by applying appropriate statistical analysis techniques and related to reference values that enable constructing regression models. If reliable calibration models are established, FT-NIR spectroscopy allows predicting the concerned constituents rapidly without any (or little) need for sample preparation. In addition, FT-NIR spectroscopy instruments are generally operated with user-friendly software that can be used even by unskilled personnel.

It was reported that the peroxide value, moisture content, oil content, and major fatty acids (oleic and linoleic acid) could be predicted with partial least square regression (PLS-R) models developed using FT-NIR spectra of both intact sesame seeds and sesame oil (Ha et al., 1998; Sato et al., 2004; Xu et al., 2017). However, to the best of our knowledge, no research has been carried out regarding the application of FT-NIR spectroscopy for the quantification of the sterol content of sesame oil.

Therefore, the aim of this study was to assess the potential of FT-NIR spectroscopy in combination with appropriate chemometric analysis for the quantification of the fatty acids and sterol contents of sesame oil obtained from sesame seeds grown in Muğla Province, Turkey.

2. Materials and methods

2.1. Chemicals

All chemical reagents were obtained from Sigma-Aldrich-Fluka Co. Ltd. (Steinheim, Germany), unless otherwise stated. Potassium hydroxide and anhydrous sodium sulfate were purchased from Carlo Erba (Milano, Italy). Trimethyl chlorosilane, 5α-cholastene-3-β-ol, and fatty acid methyl esters (FAMEs) standard mixture were purchased from Merck (Darmstadt, Germany) and Supelco (Bellefonte, USA).

2.2. Sesame samples

In total, 39 sesame seed samples cultivated in Muğla Province were supplied from the Ula Directorate of Provincial Food, Agriculture, and Livestock. The provenance of the sesame samples is given in Figure 1. The pedoclimatic conditions of these cultivation locations are characterized by alluvial/sandy soil and a temperate Mediterranean climate, which is favorable for sesame cultivation (Şahin, 2014). Sesame samples came from 7 different cultivars, namely Yerli (22 samples), Beyaz susam (4 samples), Fethiye Sarısı (3 samples), Ortaca Sarısı (3 samples), Tanas (3 samples), Kocasusam (2 samples), and Sarus (2 samples). All samples were from the 2016 crop season.

Extraction of sesame oil was carried out with a pilot-scale screw press extractor equipped with a 3-kW variable-speed electric motor (KMS cold press oil, İzmir, Turkey). The screw rotation speed was 60 rpm. Each oil sample was prepared from 1 kg of sesame seed. The obtained sesame oil samples were frozen with liquid nitrogen and stored at −40 °C. Samples were used for analyses within 2 days after extraction.

2.3. Spectral measurements

Sesame oil samples were placed in glass vials with 8-mm path length and held for temperature equilibration at 40 °C for 20 min in a climatic cabinet. Then the vials were placed into the sample port of the FT-NIR spectrometer (MPA, Bruker Optics, Ettlingen, Germany), which was equipped with an InGaAs detector and thermostated at 40 °C. The FT-NIR spectra were acquired in absorbance mode from 12,000 (833 nm) to 4000 (2500 nm) cm⁻¹ with resolution of 8 cm⁻¹ and scanner velocity of 10 kHz. The spectra were the average of 64 scans. Three spectra from each oil sample were acquired and averaged.

2.4. Determination of fatty acids

The ISO 12966-2:2011 (ISO, 2011) method was applied for the determination of FAMEs of the samples. Determination of FAME composition and quantification of individual FAMEs was performed with a PerkinElmer gas chromatography (GC) system (Auto system GLX; Waltham, MA, USA) equipped with a flame ionization detector (FID). Chromatographic separation of FAMEs was achieved by a fused silica capillary column (SP-2380; 100 m length × 0.25 mm with 0.25 µm film thickness) acquired from Supelco (Bellefonte, PA, USA). The working conditions for GC were as follows: carrier gas, helium; flow rate, 0.5 mL min⁻¹; injector temperature, 280 °C; detector temperature, 260 °C; oven temperature program, initial temperature of 120 °C for 2 min, increased at 5 °C min⁻¹ to 220 °C, held for 10 min. Individual FAMEs were identified by comparison with the retention times of FAME standards and quantified using the calibration curves established for individual FAMEs. TotalChrom Navigator software was used for the analysis of the chromatographic data. The results were expressed as percent concentration.

2.5. Determination of sterols

The ISO 12228:1999 (ISO, 1999) method was applied to attain the sterol compositions of oil samples. Determination of sterol composition and quantification of the individual sterol forms were performed by employing a GC system (PerkinElmer Auto system GLX) equipped with a FID. Separation of the sterols was performed using the SE-54 column (5%-phenyl-1%-vinylmethylpolysiloxane, 30 m × 0.32 mm × 0.25 µm film thickness) from Agilent J&W Scientific (Santa Clara, CA, USA). The working conditions of the GC were as follows: carrier gas, helium; flow rate, 0.8 mL min⁻¹; injector temperature, 280 °C; detector temperature, 300 °C; oven temperature program, initial temperature of 60 °C for 2 min, increased at 40 °C min⁻¹ to 220 °C, held for 1 min and then increased at 5 °C min⁻¹ to...
310 °C, held for 30 min. Individual sterols were identified and quantified using 5α-cholestan-3β-ol as an internal standard. TotalChrom Navigator software was used for the analysis of the chromatographic data. The result of each sterol was expressed as percent concentration and total amount was stated as mg kg⁻¹ of oil.

2.6. Multivariate statistical analyses
In order to visualize the grouping tendencies and possible outliers that may be present in the data, the spectral and reference chemical data were first subjected to principal component analysis (PCA) by using XLStat software (Version 2010.5.08, Addinsoft, France). The spectral range between 12,000 and 4000 cm⁻¹ was used for the analysis. The raw FT-NIR spectra were processed by standard normal variate (SNV) method in order to eliminate any noninformative variance due to additive and multiplicative scattering effects.

The spectral and reference data were then used to establish calibration models to predict the contents of the total and individual sterols and fatty acids. PLS-R models were established using OPUS software (version 7.2, Bruker Optik GmbH, Ettlingen, Germany). In order to validate the established calibration models, the leave-one-out cross-validation method was used. In this method, one sample from the dataset was taken out and the calibration model was constructed, and the remaining samples were tested with the sample held out and its error was evaluated. This procedure was repeated for all the samples and a generalization error estimate was obtained (RMSECV), which is the average of all the error terms obtained. The model performances were assessed on the basis of the coefficient of determination of the calibration (\( R^2 \)) and cross-validation (\( R^2_{\text{CV}} \)) models, root mean square errors of the calibration (RMSEE) and validation (RMSECV), and residual prediction deviation (RPD), which is the ratio of the standard deviation of the reference values to RMSEE or RMSECV.

3. Results and discussion
The major fatty acids found in sesame oil were oleic, linoleic, palmitic, and stearic acids in decreasing order (Table 1). The sum of mono- and polyunsaturated fatty acids accounted for 84.6% of the total fatty acids, whereas that of saturated fatty acids constituted 15.5%. The fatty acid concentration ranges of the samples used in the present study fall well within the ranges reported by Yermanos et al. (1972), who investigated the fatty acid profile of sesame seeds from the world core collection consisting of 721 accessions, 154 of which were from Turkey. When compared to the results of the studies carried out on Turkish (Uzun et al., 2008) and Mediterranean core collections (Yol et al., 2015), some discrepancies were observed. For instance, the mean oleic and stearic acid contents of the samples were higher than the values reported for the samples from the Turkish core collection, whereas similar values were reported for linoleic and palmitic acid. On the other hand, linoleic acid contents of the samples used in this study were higher whereas stearic acid was lower than the values reported for the accessions constituting the Mediterranean core collection. It is well known that agricultural practices, genotype, climate, and maturity stage substantially affect the fatty acid profiles of the sesame seeds (Baydar et al.,
It should be noted that the samples used in the present study were obtained from different local producers located in Muğla Province. Therefore, the sample set used in this study is narrower in terms of genotype than the studies carried out on core collections, which might be the reason for the slight differences in the fatty acid profiles. In this study, the mean total sterol content of the sesame seed samples was 4864 mg kg⁻¹, oil which is lower than the value (5400 mg kg⁻¹) reported for Moroccan sesame varieties (Gharby et al., 2017). Regarding the sterol profile of the sesame seeds, β-sitosterol (63.73%) was the most abundant sterol form detected, which was followed by campesterol (20.15%), Δ5-avenasterol (8.58%), stigmasterol (5.28%), and Δ7-stigmasterol (2.26%). These findings are in confirmation with the results reported for 7 Turkish varieties from the Menemen and Antalya regions (Ünal and Yağcı, 2008). Compared to sesame seeds from Morocco and Sudan (Gharby et al., 2017), except for stigmasterol, the percentage of β-sitosterol, campesterol, Δ5-avenasterol, and Δ7-stigmasterol in the oil extract of the sesame seed samples used in this study was higher. This indicates that sterol composition can exhibit important variations depending on geographical origin. Indeed, in some other vegetable oils such as olive oil, sterol composition was proved to be an effective discriminant of geographical origin (Temime et al., 2008).

The raw FT-NIR spectra acquired from the sesame oil samples from the present study are given in Figure 2. The spectral profile of the sesame oil very much resembles that of other vegetable oils such as sunflower, soy, corn, and olive oil (Azizian and Kramer, 2005), which comprise 8 major absorbance peaks located at wavenumbers of 8561, 8261, 7187, 7073, 5792, 5677, 4662, and 4591 cm⁻¹. According to the previous literature (Christy et al., 2004; Sinelli et al., 2010; Özdemir et al., 2018), these peaks can be assigned as follows: second overtones of C–H stretching vibrations of CH₃– and –CH₂– functional groups (8560 and 8260 cm⁻¹) and their combination bands (7187 and 7074 cm⁻¹), first overtones of C–H stretching vibrations of –CH₂– functional group (5791 and 5677 cm⁻¹), second overtones of C=O stretching vibrations (5260 and 5179 cm⁻¹), and combination bands of =C–H and C=C stretching vibrations of the –HC=CH– functional groups (4662 and 4595 cm⁻¹).

As a first step of the data analysis, PCA was carried out using the chemical (fatty acid and sterol composition) and spectral datasets. PCA is a widely used unsupervised explanatory statistical analysis technique that enables reducing variable dimensions in high dimensional datasets and helps to visualize the clustering tendencies of the samples. As can be seen from the score plot of the PCA carried out using spectral data (Figure 3), 93% of the total variance could be explained by the first two principal components (PCs). However, complete overlapping of the samples was observed, which indicates that NIR spectra of the sesame oil do not exhibit significant variations among samples of different geographical locations (Figure 3).

The major objective of this study was to develop PLS-R models using the FT-NIR spectra for the prediction of the content of the fatty acids and sterols found in sesame seeds. The figures of merit of the models developed are given in Tables 2 and 3. The model performances were evaluated on the basis of the coefficient of determination (R²), root mean squared error (RMSE), and residual prediction deviation (RPD) values of the calibration and cross-validation models. The R² statistic term is an indicator of the goodness of fit between the predicted and reference values for each constituent and it can have values between 0 (lack of fit) and 1 (perfect fit). The root mean squared error of calibration (RMSEE) and cross-validation (RMSECV) terms indicate the error associated to prediction and for good models.
**Figure 2.** The raw FT-NIR spectra of sesame oil samples.

**Figure 3.** The score plot of the PCA carried out on the FT-NIR spectra of the sesame oil samples.
Table 2. PLS model performance parameters (LV: number of latent variables, $R^2_c$: coefficient of determination of calibration model, $R^2_{cv}$: coefficient of determination of cross-validation model, RMSEE: root mean square error of calibration, RMSECV: root mean square error of cross-validation, RPD: residual prediction deviation) and informative spectral regions and preprocessing methods (FD: first derivative, SNV: standard normal variate, MSC: multiplicative scatter correction, SLS: straight line subtraction, COE: constant offset elimination, SD: second derivative) used for the calibration and validation of the prediction models for fatty acids and sterols.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Preprocessing</th>
<th>Spectral region</th>
<th>Calibration</th>
<th>Cross-validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>LV</td>
<td>$R^2_c$</td>
</tr>
<tr>
<td>Fatty acids</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C16:0 Palmitic</td>
<td>MMN</td>
<td>9403–8370</td>
<td>9</td>
<td>0.831</td>
</tr>
<tr>
<td>C16:1 Palmitoleic</td>
<td>FD + SNV</td>
<td>9403–8887</td>
<td>1</td>
<td>0.379</td>
</tr>
<tr>
<td>C17:0 Heptadecanoic</td>
<td>FD + MSC</td>
<td>9403–7340, 6827–6306</td>
<td>2</td>
<td>0.349</td>
</tr>
<tr>
<td>C18:0 Stearic</td>
<td>FD + SLS</td>
<td>6827–6306</td>
<td>10</td>
<td>0.800</td>
</tr>
<tr>
<td>C18:1 Oleic</td>
<td>FD + SNV</td>
<td>8891–8370, 6827–6306</td>
<td>9</td>
<td>0.991</td>
</tr>
<tr>
<td>C18:2 Linoleic</td>
<td>FD + MSC</td>
<td>8891–8370, 7344–6823</td>
<td>9</td>
<td>0.988</td>
</tr>
<tr>
<td>C20:0 Arachidic</td>
<td>FD + SLS</td>
<td>6827–6306</td>
<td>2</td>
<td>0.563</td>
</tr>
<tr>
<td>C20:1 Eicosenoic</td>
<td>MSC</td>
<td>8891–8370, 7344–5276</td>
<td>10</td>
<td>0.818</td>
</tr>
<tr>
<td>C18:3 Linolenic</td>
<td>FD + SNV</td>
<td>9403–7853, 6310–4760</td>
<td>7</td>
<td>0.665</td>
</tr>
<tr>
<td>C22:0 Behenic</td>
<td>SNV</td>
<td>7340–6823, 5280–4760</td>
<td>4</td>
<td>0.316</td>
</tr>
<tr>
<td>SFA</td>
<td>FD + SLS</td>
<td>8374–7853, 5797–5277</td>
<td>9</td>
<td>0.942</td>
</tr>
<tr>
<td>MUFA</td>
<td>FD + SLS</td>
<td>8891–8370, 6827–6306</td>
<td>10</td>
<td>0.990</td>
</tr>
<tr>
<td>PUFA</td>
<td>FD + MSC</td>
<td>8891–8370, 7857–6306</td>
<td>9</td>
<td>0.990</td>
</tr>
<tr>
<td>Sterols</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Campesterol</td>
<td>FD + MSC</td>
<td>6310–5277, 4766–4247</td>
<td>1</td>
<td>0.284</td>
</tr>
<tr>
<td>Stigmasterol</td>
<td>COE</td>
<td>8374–7340</td>
<td>1</td>
<td>0.200</td>
</tr>
<tr>
<td>β-Sitosterol</td>
<td>MSC</td>
<td>9403–8887, 7344–5794</td>
<td>9</td>
<td>0.930</td>
</tr>
<tr>
<td>Δ5-Avenasterol</td>
<td>MMN</td>
<td>6310–5794</td>
<td>7</td>
<td>0.932</td>
</tr>
<tr>
<td>Total sterols</td>
<td>SD</td>
<td>7857–7340, 6827–4760</td>
<td>10</td>
<td>0.834</td>
</tr>
</tbody>
</table>

Table 3. PLS model performance parameters (LV: number of latent variables, $R^2_c$: coefficient of determination of calibration model, $R^2_{cv}$: coefficient of determination of external test validation model, RMSECV: root mean square error of prediction, RPD: residual prediction deviation) and informative spectral regions and preprocessing methods (FD: first derivative, SNV: standard normal variate, MSC: multiplicative scatter correction, MMN: min–max normalization) used for the calibration and external test set validation of the prediction models for some selected fatty acids and sterols.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Preprocessing</th>
<th>Spectral region</th>
<th>Calibration</th>
<th>Test set validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>LV</td>
<td>$R^2_c$</td>
</tr>
<tr>
<td>Fatty acids</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C18:1 Oleic</td>
<td>FD + SNV</td>
<td>8891–8370, 6827–6306</td>
<td>10</td>
<td>0.993</td>
</tr>
<tr>
<td>C18:2 Linoleic</td>
<td>FD + MSC</td>
<td>8891–8370, 7344–6823</td>
<td>9</td>
<td>0.990</td>
</tr>
<tr>
<td>Sterols</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β-Sitosterol</td>
<td>MSC</td>
<td>9403–8887, 7344–5794</td>
<td>9</td>
<td>0.876</td>
</tr>
<tr>
<td>Δ5-Avenasterol</td>
<td>MMN</td>
<td>6310–5794</td>
<td>8</td>
<td>0.751</td>
</tr>
</tbody>
</table>
these terms are expected to be as low as possible (Uncu and Ozen, 2015). In addition, the difference between the RMSEE and RMSECV is also expected to be close to 0 for a good model. When the difference between these two error terms is high, the calibration model is considered to be overfitted and thus unreliable in predicting future samples. The residual prediction deviation (RPD) is defined as the ratio of the natural variation in the reference dataset to the size of probable errors occurring during the prediction. RPDs between 1.5 and 2 are considered good for discriminating low from high values of the response variable; values between 2 and 2.5 indicate that the model can be used for rough quantitative predictions, and values between 2.5 and 3 or above indicate good and excellent prediction accuracy, respectively (Nicolaï et al., 2007).

In this respect, among the prediction models developed for different fatty acid forms the best model performances were obtained for the oleic ($R^2 = 0.991$, RMSEE = 0.092%, RPD = 10.7) and linoleic ($R^2 = 0.988$, RMSEE = 0.118%, RPD = 9.01) acid models. For oleic acid the best result was obtained by using first derivative (Savitsky-Golay 17-point smoothing) and SNV spectral pretreatment methods, whereas for linoleic acid first derivative (Savitsky-Golay 17-point smoothing) and multiplicative spectral correction (MSC) pretreatment methods were applied. In PLS-R model development using FT-NIR spectra, the choice of pretreatment method depends on the nature of the scattering effects observed on the spectra (Qu et al., 2015). The spectral pretreatment methods including 1st or 2nd derivatives are useful in dealing with both additive and multiplicative scattering effects. On the other hand, SNV is effective only in removing additive scattering effects such as baseline offsets that are not wavelength-dependent. MSC, however, is more effective in eliminating multiplicative scattering effects that are wavelength-dependent. It should be noted that the software used for the model development in the present study makes use of an optimization algorithm that searches for the best combination of pretreatment method, number of latent variables, and informative wavelength regions resulting in the least root mean square error of prediction (RMSEP) from the list of commonly used preprocessing methods and various spectral regions. Therefore, the choice of pretreatment method was an outcome of optimization study.

In general, if the sample number is lower than 50 the prediction ability of the model is tested by using the leave-one-out cross-validation method (Uncu and Ozen, 2015). Nevertheless, we also carried out external test validation by using 1/3 of the sample set, which was not used in calibration model development (Table 3; Figure 4). As seen from Table 2, $R^2$, RMSECV, and RPD values of the

![Figure 4](image-url). Reference values versus values predicted by using partial least squares (PLS) models for oleic acid, linoleic acid, β-sitosterol, and Δ5-avenasterol content of sesame oils.
Validation models for the oleic and linoleic acids were very close to those of corresponding calibration models, which indicates that the models were not overfitted and can be considered reliable for predicting the oleic and linoleic acid content of unknown samples. Similar model performances were also obtained by applying external test set validation, which indicates high model robustness (Table 3). In the literature slightly better results were reported for oleic and linoleic acid content prediction models for olive oils (Mailer, 2004; Özdemir et al., 2018). Azizian et al. (2007) reported that an increased concentration of β-sitosterol in triolein decreased and increased the observed contents of cis9-18:1 and cis11-18:1, respectively. As can be seen from Table 1, the total sterol content of the sesame oils used in this study was in the range of 4612–5353 mg kg⁻¹ oil (63.73% of which was β-sitosterol), which was 2- to 3-fold higher than the sterol content of olive oils reported in the literature (Dag et al., 2015). Therefore, it is highly probable that the high sterol content of sesame oil slightly lowered the model performances, as previously stated by Azizian et al. (2007).

The models for fatty acids other than oleic and linoleic acids performed poorly, as can be understood from their very low R² and RPD values. For olive oils better model performances were reported for palmitic, palmitoleic, and stearic acids (Mailer, 2004). As presented in Table 1, the concentration ranges of the sesame samples for palmitic, palmitoleic, and stearic acids were narrower than those reported in the above study for olive oils. As a rule of thumb, in order to attain good model performances it is desirable to use samples with wide concentration ranges. However, depending on the peculiar characteristics of the sample it is not always easy to establish such a sample set, as was the case in the present study. Indeed, the results of a compositional study carried out on a large number of sesame samples, 103 accessions, from the Mediterranean core collection (Oyol et al., 2015), showed that the concentration ranges of the major fatty acids found in sesame oil exhibited lower variation compared to olive oils (Özdemir et al., 2018). It should also be emphasized that the lower the concentration of fatty acids is, the less performant the predictive models are (Mailer, 2004).

On the other hand, when fatty acids were grouped on the basis of saturation degree as saturated (SFAs), monounsaturated (MUFA), and polyunsaturated (PUFA) fatty acids, models with good prediction abilities could be established. The R², RMSECV, and RPD values for the models constructed for SFA, MUFA, and PUFA content of the sesame oil were in the ranges of 0.865–0.976, 0.108%–0.148%, and 2.72–6.41, respectively. These findings are in confirmation with those of Uncu and Ozen (2015), who also reported that grouping fatty acids according to their saturation degrees improved the model prediction ability.

Regarding the prediction of the sterol content of the vegetable oils with FT-NIR-based PLS-R models, the literature is scarce. Only recently, Özdemir et al. (2018) reported that the total sterol content of extra virgin olive oils could be effectively predicted with FT-NIR models. However, the authors failed to establish reliable models for the individual sterol forms, namely camppesterol, stigmasterol, β-sitosterol, and Δ5-avenasterol. In the present study, models with moderate quality for the β-sitosterol, Δ5-avenasterol, and total sterol content of the sesame oils could be established. However, the results of cross-validation showed that the model constructed for the total sterol content was slightly overfitted as the RMSEP (89.1 mg kg⁻¹) was twice as high as the RMSECV (144 mg kg⁻¹). The results of the external test validation indicate that the calibration model established for the β-sitosterol content was robust enough to predict unknown samples as the RMSEP obtained was comparable to the RMSECV with similar R² and RPD values. However, regarding Δ5-avenasterol, the RMSEP obtained after external test set validation was considerably higher than the RMSECV (Table 3), indicating that the model developed for the prediction of Δ5-avenasterol was not robust enough to predict unknown future samples. It should be emphasized that the low number of samples used in the present might negatively affect the model performance and with a higher number of samples better results could be obtained.

The results of the present study showed that FT-NIR spectroscopy coupled with multivariate statistical analysis can be used for the rapid and simultaneous analysis of the major fatty acids and sterols found in sesame oil, which makes it a valuable analytical tool for quality control and breeding studies of sesame.

Acknowledgment

The authors would like to thank Süleyman Kurnaz from the Ula Directorate of Provincial Food, Agriculture, and Livestock for providing sesame seed samples.
References


Baydar H, Turgut İ, Turgut K (1999). Variation of certain characters and line selection for yield, oil, oleic and linoleic acids in the Turkish sesame (Sesamum indicum L.) populations. Turkish J Agric For 23: 431-442.


