International Food Research Journal 21(2): 723-727 (2014)
Journal homepage: *http://www.ifrj.upm.edu.my*

Prediction of oil content in olive fruit using Fourier transformed infrared spectroscopy FT-IR coupled with partial least squares regression

Hirri, A., Bassbasi, M., Kzaiber, F. and *Oussama, A.

Laboratory of applied and environmental Spectrochemistry, Faculty of Science and Technology of Beni Mellal, University Moulay soulymane, Morocco

Article history

Received: 10 October 2013
Received in revised form:
25 November 2013
Accepted: 27 November 2013

Keywords

Olives
Mesocarp
FT-IR
PLS
Oil content

Abstract

A rapid Fourier transformed infrared (FT-IR) attenuated total reflectance (ATR) spectroscopic is applied to predict the quantity of oil in fresh olive in tree. The analytical method is evaluated by use of validation samples with nearly quantitative oil. 80 samples of olives in oil content, which varies between 8 and 21%, picked up in the Moroccan region, were subjected to infrared analysis. Analytical data were collected, by Fourier transform infrared spectroscopy (FT-IR) applied to the mesocarp of the fresh olives without any preliminary treatment. The objective of this study is to develop a calibration model for prediction of oil content in olive fruit by using FT-IR spectroscopy before harvest time. The transmission spectra of olive fruit were obtained in the wavelength range from 4000 to 600 cm^{-1}. The prediction models were developed by partial least square regression (PLS). The values obtained for correlation coefficient for oil content and root mean square errors of prediction (RMSEP) are 0.99 and 0.076 respectively. This show the capability of FTIR and the important role of chemometric in developing accurate models to predict oil content in olive fruit.

Introduction

Olive (from *Olea europaea*) is considered one of the most widely grown fruit crop in the countries of the Mediterranean basin (Amane *et al*., 1999). The olive products, such as olive oil, table olives and olive pastes are the basic constitutes of the Mediterranean diet due to their benefits for human health, besides other applications such as in cosmetics field. Olive shows high genetic variability with more than 2600 cultivars described (Rugini *et al*., 1992), although many of them might be synonyms, homonyms (Barranco *et al*., 2000), ecotypes or the result of crosses between neighboring individuals. A few of these cultivars are found over large areas of cultivation but most have a very local distribution (Besnard *et al*., 2001a). Nowadays, one of the major problems in the agricultural–food industry is to set down objective tools in order to determine the oil yield in olive fruit before extraction. Most of analyses used for characterization of cultivars are time-consuming, expensive and involve a considerable amount of manual work. Very often, complex chemical treatment of the sample and the use of sophisticated instruments are required (Bassbasi *et al*., 2014). Recently, Fourier transform infrared (FT-IR) spectroscopy has become an emerging well-accepted analytical technique, due to its simplicity with advantages in terms of cost per sample. It achieves high analysis speed and requires little or no sample preparation. FT-IR spectroscopy has been widely used as an analytical tool in many laboratories and industrial sectors such as food agricultural (De Luca *et al*., 2011; Aouidi *et al*., 2012), petrochemical (Roman *et al*., 2008; Bassbasi *et al*., 2013), textile (Langeron *et al*., 2007) and pharmaceutical (Wu *et al*., 2008). FT-IR data have been often combined with chemometric techniques to develop methods of classification and characterization. This approach has been found to be very useful in many applications, due to the ability of these methods in achieving the spectral resolution of the FTIR signals. Up to now, a lot of studies have been published on the utilization of near and mid FT-IR for authentication, identification or classification of many agro-foods, notably olive oils (Sinelli *et al*., 2010; Sinelli *et al*., 2010) and table olives (Casale *et al*., 2010) by multivariate statistical analysis of spectral data. FT-MIR has been applied to olive leaves to discriminate between five Tunisian cultivars (Aouidi *et al*., 2012) and to predict nutritive composition (Fernandez-Cabanas *et al*., 2008).

The aim of this study is to develop, by FT-MIR spectroscopy associated to chemometric treatment, a direct and rapid test method that quantified oil content before harvest of olive fruit. This approach can be considered a fast, clean and affordable methodology that allows producers to improve fruit harvesting and storage.

---

*Corresponding author.
Email: *elmahfoud.ass@gmail.com*

Materials and Methods

Olive fruits sampling

In this work, 80 samples of olive have been quantitatively investigated. All these samples of olives are analyzed four times per sample by FT-IR in a range from 4000 and 600 cm^{-1}. The olive fruit samples are picked from trees in Beni Mellal area (center of Morocco) and grown in the same agronomical and climatic conditions. The investigated concentration of oil content in olives varied from 8 to 21%. At the beginning, 60 samples are prepared. 50 are randomly selected for calibration (Cal) and 10 for external validation (Val 1). In separate experiments, two new sets (Val 2 and Val 3) of 20 samples (10 samples per validation), are prepared to assess the model reproducibility.

FT-IR analysis

FTIR spectra are obtained using a Vector 22 Bruker FTIR Spectrophotometer equipped with an attenuated total reflectance accessory (ATR single-reflexion, Diamond, incident angle 45^{o}), DTGS detector, Globar (MIR) source and KBr Germanium separator, with a resolution of 4 cm^{-1} at 98 scans. Spectra are scanned in the absorbance mode from 4000 to 600 cm^{-1} and the data are handled with OPUS logiciel. Analyses are carried out at room temperature. The background is collected before each sample was measured.

Extraction of olive oil

Olive oil is contained in tiny pockets of the olive cells, called vacuoles. In order to recover this oil, the wall of the pockets and therefore the first cells of olives must be broken. This operation is called grinding. In most cases, the olives are completely crushed, i.e. with their core. This gives us a paste having more or less liquid consistency, depending on the varieties of olives and the time of picking. As the grinding is not sufficient to break all of the vacuoles in order to release the maximum amount of oil, mixing is applied to the dough. Thence, combination of grinding and mixing leads to a dough with solid materials (debris of nuclei, of epidermis, of cell walls...) and fluids (oil and vegetation water, i.e. the water in the cells of the olive). After that, comes the phase of separation, which consist first in separating the solid part (called grignon) from the fluid part (called margine) and last in separating the oil from the vegetation water (the settling) where the oil content is obtained in fruits using 250 grams of fresh mesocarp.

Partial least square regression (PLS) theory

PLS is a supervised analysis which is based on the relation between the signal intensity and the characteristics of the sample (Beebe *et al.*, 1987). Interference and overlapping of the spectral information may be overcome by using powerful multicomponent analysis such as PLS regression. La PLS (Fuller *et al.*, 1978) allows a sophisticated statistical approach using spectral region rather than unique and isolated analytical bands. The first step is to perform a calibration model. This involves collecting a set of reference calibration samples, which should contain all chemical and physical variations to be expected in the unknown samples, which will be predicted later. The model was built by full cross-validation methods during the calibration development. The optimal number of PLS Latentes variables (LVs) were found according to the full cross-validation procedure. The second step is to test the model using a prediction set (different to the calibration one), i.e. to compare the values obtained by the model to the values obtained by the reference method. The evaluation of the errors in the calibration is estimated by computing the standard error of calibration (SEC) after comparing the real concentration with the computed one for each component. The formula for the standard error of calibration is:

$$SEC = \sqrt{\left(\sum_{i=1}^{N} \frac{(C_i - C_i')^2}{N-1-p}\right)}$$

where C_i is the known value, C_i' is the calculated value, N the number of samples and p is the number of independent variables in the regression optimized by cross validation.

The standard error of prediction (SEP) gives an estimation of the prediction performance during the step of validation of the calibration equation:

$$SEP = \sqrt{\left(\sum_{i=1}^{M} \frac{(C_i - C_i')^2}{M}\right)}$$

Where C_i is the known value, C_i' is the value calculated by the calibration equation, and M is the number of samples in the prediction set. Another useful parameter is the relative error of prediction (REP %) that shows the predictive ability of the model. This is calculated from the equation:

$$\text{REP\%} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i)^2}} \text{ x100}$$

Quality assessment of the obtained results is discussed by comparison of predicted values versus measured values, both for calibration and for validation data sets. All chemometrics calculations were performed using the Unscramble x software version 10.2 from CAMO (Computer Aided Modeling, Trondheim, Norway) was used for chemometric treatments of FTIR-ATR data spectra.

The predictive ability of the model should also be expressed by the bias and the square of correlation coefficient (R^2) also called determination coefficient, usually called Q^2 in prediction. The regression coefficients are the numerical coefficients which express the link between the predictor variations and the response variations. The bias is systematic difference between predicted and measured values. The bias is computed as average value of the residuals. The residual is the measure of the variation which is not taken into account by the model. The residual for a given sample and a given variable is computed as the difference between observed value and fitted (projected or predicted) value of the variable on the sample.

Results and Discussion

Typical FTIR spectra of olive fruit

Fourier transform infrared (FT-IR) spectra obtained for 80 samples of olive fruit are simulated by visual inspection. One spectrum is the average of 4 scans of the same sample of olive on FT-IR. In this case, 80 samples of olives are made from 80 different crop samples. The average spectra of all considered samples are presented in Figure 1.

We have used the entire spectral range 4000-600 cm^{-1} as a starting point for the investigation. According to the literature (Terouzi *et al*., 2011), the region between 2400 and 2300 cm^{-1} was discarded before the chemometric elaboration because the signal/noise ratio is low and the signal variation resulted to be independent of the sample composition. The region between 4000 and 3033 cm^{-1} is also removed due to the water which is instrumental noise and brings useless information (Dupuy *et al.*, 2010). All the spectra are dominated by two peaks at 2860 and 2850 cm^{-1}, due to bands arising from asymmetrical and symmetrical stretching vibrations of methylene ($–CH_2$) groups and strong bands between 1750 and 1650 cm^{-1} arising from the stretching vibration of the ester carbonyl functional groups of the triglycerides. The bands from 1480 to 1450 cm^{-1} is assigned to the bending vibrations of the $–CH_2$ and $–CH_3$ aliphatic groups (Shiroma *et al.*, 2009). The visual examinations of the spectral variations don't permit to apprehend clearly the difference between chemical structure and chemical species concentration in olive fruits. Chemometric treatments are, often, applied in order to extract information from the spectral data set.

Table 1. Explained variances (%) of PCs used in the PLS model

Explained	factor 0	factor 1	factor 2	factor 3	factor 4	factor 5	factor 6	factor 7
Calibration	0	86,35	96,96	98,53	99,45	99,85	99,91	99,93
Validation	0	84,59	96,68	98,12	99,08	99,69	99,83	99,87

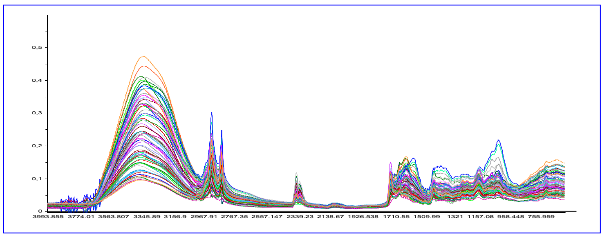

Figure 1. Mean FT-IR spectra of olives fruits

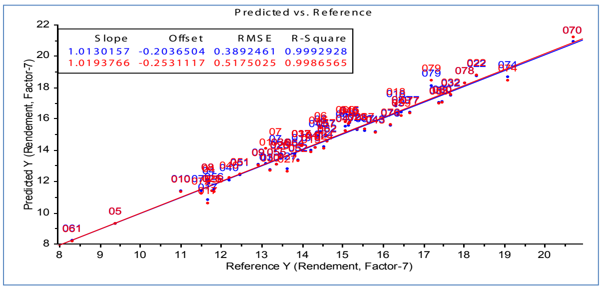

Figure 2. Plot of the predicted and measured values for olives fruits yield oil (%), obtained from the final PLS model developed from the MIR spectra

PLS modeling

The PLS model is built by considering the two frequency intervals 3000–2400 cm^{-1} and 2300–600 cm^{-1} with X as variable and the Y variables is associated to the oil yield. The PLSR models are evaluated using coefficient of determination (R^2) in calibration, root-mean-square error of calibration (RMSEC) and cross validation (RMSECV). The performance of the PLSR models on the independent validation set is assessed using R^2, RMSEP and the residual prediction deviation (RPD).

Here, the criteria of classifying RPD values (Mouazen *et al*., 2006) is adopted as follows: an RPD value below 1.5 indicates that the calibration is not usable; an RPD value between 1.5 and 2.0 indicates the possibility of differentiating between high and low values; an RPD value between 2.0 and 2.5 makes possible approximate quantitative predictions. For RPD value between 2.5 and 3.0 and beyond 3.0, the prediction is classified as good and excellent, respectively. Generally, a good model should have high values of R2 and RPD, and low values of RMSEC, RMSECV and RMSEP.

The resulting model seems to be able to determine the oil yield for 50 samples of olives fruits. As can be seen from Figure 2. The PLS model is validated by full cross validation. The obtained

Table 2. Prediction of the oil content of external validation

val 1				val 2				val 3			
Samples	**Predi,**	**Dev**	**Ref,**	**Samples**	**Predi,**	**Dev**	**Ref,**	**Samples**	**Predi,**	**Dev**	**Ref,**
3	17,12	0,35	17,10	**21**	18,06	0,43	18,07	**12**	15,50	0,34	15,51
11	14,19	0,36	14,16	**28**	17,03	0,45	17,03	**14**	16,18	0,33	16,18
16	10,24	0,55	10,30	**31**	10,88	0,65	10,88	**20**	19,52	0,41	19,53
33	12,94	0,70	12,90	**35**	14,82	0,79	14,82	**29**	14,80	0,26	14,80
34	10,52	0,39	10,50	**39**	14,09	0,29	14,05	**42**	9,00	0,43	9,02
38	14,93	0,30	14,86	**48**	15,23	0,39	15,24	**44**	16,81	0,31	16,82
46	12,46	0,45	12,35	**50**	14,18	0,34	14,20	**60**	16,96	0,56	16,98
54	13,60	0,57	13,60	**58**	14,65	0,51	14,68	**65**	12,60	0,53	12,63
63	17,27	0,35	17,24	**68**	18,83	0,38	18,86	**69**	11,51	1,12	11,52
75	18,91	0,38	18,85	**71**	12,41	0,37	12,43	**73**	18,30	0,61	18,32

Predi: prediction; Dev: Deviation; Ref: reference

Table 3. Statistical parameters carried out by external validation on PLS

External validation	LVs	Rp^2	RMSEP	Bias	RPD	REP (%)
Val 1	7	0.998	0.0536	0.0318	22.36	0.3718
Val 2	7	0.999	0.0149	-0.0122	31.62	0.0969
Val 3	7	0.996	0.0209	-0.0071	15.81	2.1558

the statistical parameters RMSEC, RMSECV and R^2 are summarized in figure 2. The coefficient of determination (R^2) of 0.99, RMSEC of 0.38 and RMSECV of 0.51, could be considered satisfactory. Seven VLs are necessary to have a good PLS performance. Table 1 lists the explained variances from the developed model.

Determination of oil content in the new samples

For application, the proposed analytical method is applied to determine the oil yield in 30 new samples, devised in 3 groups of external validation (10 samples for each group), and collected from different regions of Beni Mellal. For each sample, the target parameter is carefully estimated using the traditional method of oil extraction and the optimized PLS one. The obtained results are given in tables 2 and 3.

As clearly indicated in Table 3, the FTIR-PLS method is an effective method for estimating oil yield in olive fruits before extraction. The results indicate that, there is no significant difference between the reference methods and the proposed one. The PLS model for the FT-IR data treatment appears to be appropriate therefore.

Conclusion

For this study, our ATR-FTIR-PLS based strategy is developed for determination of content oil in fresh olive. FT-IR spectroscopy coupled to chemometrics techniques is reported as an adequate method for determination of content oil in olive fruit, without any previous sample pretreatment and sample destructive manipulation. The analysis is directly performed on a section of the olive mesocarp in the tree to give a complete fingerprint of olive without the need sample preparation. The results shows that MIR spectroscopic technique, associated with chemometric elaboration, can be a reliable way for predicting the content oil in olive fruits in advance. Therefore, the proposed spectroscopic method furnishes a nightly convenient alternative in terms of time and solvent saving for routine analysis of large number of olive samples. This approach can be considered fast, clean and affordable methodology that allows producers to improve fruits harvesting and storage.

Reference

Amane, M., Lumaret, R., Hany, V., Ouazzani, N., Debain, C., Vivier, G. and Deguilloux, M.F. 1999. Chloroplast-DNA variation in cultivated and wild olive (*Olea europaea* L.). Theoretical and Applied Genetics 99: 133-139.

Aouidi, F., Dupuy, N., Artaud, J., Roussos, S., Msallem, M., Perraud-Gaime, I. and Hamdi, M. 2012. Discrimination of five Tunisian cultivars by Mid InfraRed spectroscopy combined with chemometric analyses of olive. Food Chemistry 131 (1): 360-366.

Bassbasi, M., Oussama, A., Hafid, A., Platikanov, S. and Tauler, R. 2013. Study of motor oil adulteration by infrared spectroscopy and chemometrics methods. Fuel 104: 798-804.

Bassbasi, M., Platikanov, S., Tauler, R. and Oussama, A. 2014. Determination of Solid Non Fat (SNF) in raw milk using PLS and SVM chemometric methods. Food Chemistry 146: 250-254.

Beebe, K.R and Kowalski, B.R. 1987. An introduction to multivariate calibration and analysis. Analytical Chemestry 59: 1007-1017.

Casale, M., Zunin, P., M. Elisabetta Cosulich, Pistarino, E., Perego, P. and Lanter, S. 2010. Characterization of table olive cultivar by NIR spectroscopy. Food Chemistry 122: 1261-1265.

De Luca, M., Terouzi, W., Ioele, G., Kzaiber, F., Oussama, A., Oliverio, F., Tauler, T. and Ragno, G. 2011. Derivative FTIR spectroscopy for cluster analysis and classification of morocco olive oils. Food Chemistry 124: 1113–1118.

Dupuy, N., Galtier, O., Ollivier, D., Vanloot, P. and Artaud, J. 2010. Comparison between NIR, MIR, concatenated NIR and MIR analysis and hierarchical PLS model. Application to virgin olive oil analysis. Analytica Chimica Acta 666 (1-2): 23–31.

Fernandez-Cabanás, V.M., Garrido-Varo, A., Delgado-Pertiñez, M. and Gómez-Cabrera, A. 2008. Nutritive evaluation of olive tree leaves by near-infrared spectroscopy: effect of soil contamination and correction with spectral pretreatments. Appl Spectrosc 62 (1): 51-58.

Fuller, M.P. and Griffiths, P.R. 1978. Diffuse reflectance measurements by infrared fourier transform spectrometry. Analytical Chemistry 50: 1906–1910.

Langeron, Y., Doussot, M., Hewson, D.J. and Duchêne, J. 2007. Classifying NIR spectra of textile products with kernel methods. Engineering Applications of Artificial

Intelligence 20 (3): 415–427.

Mouazen, A.M., Baerdemaeker, J.D. and Ramon, H. 2006. Effect of wavelength range on the measurement accuracy of some selected soil constituents using visual-near infrared spectroscopy. Journal of Near Infrared Spectroscopy 14 (3): 189-99.

Roman, M.B. and Ravilya, Z.S. 2008. Gasoline classification by source and type based on near infrared (NIR) spectroscopy data. Fuel 8: 1096–1101.

Shiroma, C. and Rodriguez-Saona, L. 2009. Application of NIR and MIR spectroscopy in quality control of potato chips. Journal of Food Composition and Analysis 22 (6): 596–605.

Sinelli, N, Cerretani, L., Di Egidio, V., Bendini, A. and 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 Research International 43 (1): 369-375.

Sinelli, N., Casale, M., Di Egidio, V., Oliveri, P., Bassi, D.,Tura, D. and Casiraghi, E. 2010. Varietal discrimination of extra virgin olive oils by near and mid infrared spectroscopy. Food Research International 43 (8): 2126-2131.

Terouzi, W., De Luca, M., Bolli, A., Oussama, A., Patumi, M., Ioele, G. and Ragno, G. 2011. A discriminant method for classification of Moroccan olive varieties by using direct FT-IR analysis of the mesocarp section. Vibrational Spectroscopy 56 (2): 123–128.

Wu, Y.W., Sun, S.Q., Zhou, Q. and Leung, H.W. 2008. Fourier transforms Mid Infrared (MIR) and Near Infrared (NIR) spectroscopy for rapid quality assessment of Chinese medicine preparation Honghua Oil. Journal of Pharmaceutical and Biomedical Analysis 46 (3): 498–504.