Characterisation of olive fruit for the milling process by using visible/near infrared spectroscopy

Roberto Beghi,1 Valentina Giovenzana,1 Raffaele Civelli,1 Enrico Cini,2 Riccardo Guidetti1
1Department of Agricultural and Environmental Sciences – Production, Landscape, Agroenergy, University of Milan; 2Department of Economics, Engineering, Science and Technology Agriculture and Forestry, University of Florence, Italy

Abstract

Increasing consumption of olive oil and table olives has recently determined an expansion of olive tree cultivation in the world. This trend is supported by the documented nutritional value of the Mediterranean diet. The aim of this work was to test a portable visible/near infrared (vis/NIR) system (400-1000 nm) for the analysis of physical-chemical parameters, such as olive soluble solid content (SSC) and texture before the olive oil extraction process. The final goal is to provide the sector with post-harvest methods and sorting systems for a quick evaluation of important properties of olive fruit. In the present study, a total of 109 olives for oil production were analysed. Olive spectra registered with the optical device and values obtained with destructive analysis in the laboratory were analysed. Specific statistical models were elaborated to study correlations between optical and laboratory analysis, and to evaluate predictions of reference parameters obtained through the analysis of the visible-near infrared range. Statistical models were processed using chemometric techniques to extract maximum data information. Principal component analysis (PCA) was performed on vis/NIR spectra to examine sample groupings and identify outliers, while partial least square (PLS) regression algorithm was used to correlate samples spectra and physical-chemical properties. Results are encouraging. PCA showed a significant sample grouping among different ranges of SSC and texture. PLS models gave fairly good predictive capabilities in validation for SSC ($R^2=0.67$ and RMSECV%=7.5%) and texture ($R^2=0.68$ and RMSECV%=8.2%).

Introduction

Growing consumption of olive oil and table olives has recently determined an expansion of olive tree (Olea europaea L.) cultivation in many countries throughout the world.

There is increasing evidence to suggest that monounsaturated fatty acids as a nutrient, olive oil as a food, and the Mediterranean diet as a food pattern are associated with a decreased risk of cardiovascular disease, obesity, metabolic syndrome, type 2 diabetes and hypertension (Lopez-Miranda et al., 2010).

Ripening process control is essential. In fact, during olive oil fruits ripening, biochemical processes occur: sugar content decreases with time, while the oil accumulation increases (Cherubini et al., 2009; Salas et al., 2002). Moreover, olives with a high sugar content may present oils with defects because of sugar fermentation during the production process. Therefore, sugar concentration may be considered an index capable of defining an appropriate level of olive ripening for processing (Cherubini et al., 2009).

Furthermore, degradation occurs during the processing and shelf-life of extra virgin olive oil and success on the market may also depend on the product’s stability. Degradation may result in variations in the nutritional quality of the product, since antioxidant content decreases and free radical content increases, variations in sensory descriptors may reduce appreciation of the product, since aroma, colour, taste and flavour attributes change and some unpleasant sensory phenomena may occur (Zanoni et al., 2005).

During maturation, fruit weight increases. The flesh texture, related to the dry matter content, is a quality parameter for table olive fruits (Beltra et al., 2004). Yousfi et al. (2006) studied changes in quality and phenolic compounds of virgin olive oils during fruit maturation. They confirmed that firmness allowed better discrimination at the initial maturity stages than the other methods tested (harvest date, amount of chlorophylls and carotenoids in the oil).

Studies consistently support the concept that the level of olive ripening may affect oil quality and this also holds true for the quality of table olives. Marsillo et al. (2001) carried out an experimental investigation on olive fruit cultivars to assess free sugar and polyphenolic compositions and their changes during ripening and processing. Patumi et al. (2002) established olive and olive oil quality after intensive monocone olive growing in different irrigation regimes. Borzillo et al. (2000) evaluated quality of Ointoria table olives during ripening and process-
ing by biomolecular components.

Diaz et al. (2004) compared three algorithms to classify table olives in four quality categories using computer vision. Classification of table olives according to their quality was carried out after the fermentation process.

Established methods for olive and oil quality assessment are generally based on either colourimetric or chromatography techniques such as high performance liquid chromatography. However, the difficult preparation of samples for this analysis requires a well-equipped laboratory and results are only available after 8-10 h. A limited number of laboratories and the lack of readily available data mean that oil mills must begin the process of oil production without having the necessary information, thus reducing their chances of diversifying production and obtaining products with the desired characteristics.

Therefore, there is a strong need in the modern oil industry for a simple, rapid, and easy-to-use method for objectively evaluating the level of olive ripening. A tool enabling real-time analysis at the receiving station would allow preliminary decision-making about olives during consignment thanks to the rapid analysis of ripening parameters (i.e. soluble solid content, SSC and texture) simultaneously.

Near infrared (NIR) spectroscopy has been shown to be one of the most efficient and advanced tools to monitor process and control product quality in the food industry. It is widely used for rapid quality control of several products (Guidetti et al., 2012).

During fruit ripening, chlorophyll degradation is responsible for the degreening of the ground colour, which is a well-established ripeness indicator for several species. In completely red-pigmented cultivars of fruits such as apples and peaches this process is not visible, being masked by the anthocyanins in the skin. Optical systems were developed to assess the chlorophyll content in these fruits in a non-destructive manner, to estimate ripeness, and to optimise harvesting and post-harvest management (Bodria et al., 2004).

Studies available in the literature discuss quality evaluation of olives and olive oil using optical analysis. Salguero-Chaparro et al. (2013), Kavdir et al. (2009), and Conte et al. (2003) considered the application of NIR analysis to intact olives and olive oil production quality control. In 2004, Mailer studied rapid evaluation of olive oil quality by NIR reflectance spectroscopy. Marquez et al. (2005) used optical NIR sensor for on-line virgin olive oil characterisation. Bendini et al. (2007) presented a preliminary evaluation of the application of Fourier transform infrared spectroscopy to control the geographical origin and quality of virgin olive oils. Bellincontro et al. (2012) studied the application of a portable NIR for on-field prediction of phenolic compounds during the ripening of olives. Mailer (2004) calibrated chemical factors, including free fatty acids (FA), induction time, polyphenol content, and FA profiles, for NIR analysis. The results provided evidence of the ability of NIR analysis to measure most olive oil compounds rapidly and accurately. Morales-Sillero et al. (2011) studied the feasibility of NIR spectroscopy for non-destructive characterisation of table olives.

The application of vis/NIR technology, in order to monitor ripening and estimate quality parameters, has already been carried out on different fruits. Nicolai et al. (2007) presented a review regarding non-destructive measurements of fruit and vegetable quality by means of NIR spectroscopy. Furthermore, this acquisition technique was proved to be suitable for a direct use to monitor quality parameters, SSC in particular, and good correlations were obtained (Beghi et al., 2013). Using vis/NIR technique, it was possible to estimate changes in the firmness and SSC of stored Red Delicious apples undergoing no detectable change in skin colour (Bodria et al., 2004).

NIR and vis/NIR instrumentation must always be complemented with chemometric analysis to extract useful information present in the spectra (Guidetti et al., 2012). The most used chemometric techniques are the principal component analysis (PCA) as a qualitative analysis of the data and partial last square (PLS) regression analysis to obtain quantitative prediction of the desired parameters (Cen and He, 2007). Therefore, we studied the capability of a portable and non-destructive optical system (vis/NIR spectrophotometer) in combination with multivariate analysis to investigate two parameters (SSC and texture) for the characterisation of olive fruits entering the processing mill. Chemometric tools were used, such as PCA and PLS regression methods. For both parameters (SSC and texture) dedicated chemometric models were created.

### Materials and methods

#### Sampling

The experiment was carried out on 109 healthy olives harvested in November and December 2011 on the Montepaldi experimental farm in Florence (Tuscany, Italy). Olive fruits used in this study were Moraiolo and Frantoio (approx. 50% of each) cultivated in the Province of Florence; these varieties are typical of the Tuscan hills.

Picked fruits taken at random from the bin were measured. Harvested and classified berries were analysed in the laboratory to determine parameters, indicating stage of ripening. Two spectral measurements were taken on individual berries along their equator region. Subsequent to the spectral acquisition, analyses of texture and of SSC of each olive were carried out. For SSC analysis, fruits were destoned and the flesh (pulp) was crushed and measured using a portable refractometer (Pocket Refractometer PAL-1 by ATAGO, Iabashi-ku, Tokyo, Japan). Hardness was assessed using a portable penetrometer (AGROSTA®100 by Agro-Technologie, Forges les Eaux, France) where a spring is compressed onto the fruit and a tip (25 mm²) is displaced (Barreiro et al., 2004). Hardness is expressed as AGROSTA®100 units (0-100).

Two distinct classifications were performed on SSC and texture (Table 1). In both cases, three arbitrary classes (a, b, and c) were created to identify different ranges for each parameter. Class definition was based on a window for the central class b equal to mean value ±3 standard error.

Spectral data and destructive reference analysis were used to elaborate chemometric predictive models.

#### Visible/near infrared system device

Spectral acquisitions were realised on samples using an optical portable system (JAZ vis/NIR spectrophotometer, OceanOptics, Inc., Dunedin, FL, USA) operating in the 400-1000 nm wavelength range. The JAZ equipment is composed of five components: i) vis/NIR lighting system; ii) fibre optic probe for reflection measurement; iii) spectrophotometer; iv) hardware for data acquisition and instrument control; v) power battery.

### Table 1. Arbitrary classes based on different ranges of soluble solid content and texture.

<table>
<thead>
<tr>
<th>Class</th>
<th>SSC</th>
<th>°Brix</th>
<th>Texture</th>
<th>Hardness units</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>&lt;17</td>
<td>a</td>
<td>&gt;80</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>17.1-19.9</td>
<td>b</td>
<td>70.1-79.9</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>&gt;20</td>
<td>c</td>
<td>&lt;70</td>
<td></td>
</tr>
</tbody>
</table>

SSC, soluble solid content.
Spectra were acquired in reflectance: light radiation was guided to the sample through a Y-shaped, bidirectional fibre optic probe (OceanOptics, USA). Y-shaped fibre allowed light from a halogen lamp to be guided to illuminate the sample while simultaneously collecting the radiation coming from the berry and guiding it back to the spectrophotometer. The tip of the optical probe was equipped with a soft plastic cap to ensure contact with sample skin during measurements, while minimising environmental light interference.

The integrated spectrophotometer was equipped with a diffractive grating for spectral measurements, optimised in the range 400-1000 nm, and a CCD sensor with a 2048 pixel matrix, corresponding to a nominal resolution of 0.3 nm.

Data processing

Chemometric analysis was performed using The Unscrambler software package (version 9.6, CAMO ASA, Oslo, Norway). Different pretreatments were applied to the vis/NIR spectra in order to maximise the accuracy of the model. Moving-averaged smoothed spectra, multiplicative scatter correction and derivatives were calculated before building the calibration models. The first and second derivatives were performed using Savitzky-Golay transformation and smoothing (15 point and second order filtering).

A qualitative analysis was carried out using a PCA tool to find possible clustering of the olive spectra (Massart et al., 1997; Naes et al., 2002). PCA was performed according to SSC and texture ranges to better highlight differences in spectra. Moreover, a quantitative analysis was performed using all samples available for the creation of a chemometric regression model for each parameter considered. The vis/NIR spectra were correlated with the ripeness parameters (SSC and texture) using the PLS regression algorithm. By this method, an orthogonal basis of latent variables is constructed one by one in such a way that they are oriented along directions of maximal covariance between spectral matrix X and response vector Y. This method ensures that the latent variables are ordered according to their relevance for predicting the Y variable. Interpretation of the relationship between the X data and the Y data (the regression model) is then simplified, as this relationship is concentrated on the smallest possible number of latent variables. The PLS method performs particularly well when the various X variables express common information, or even collinearity, which is the case for spectral variables.

The PLS method performs particularly well when the various X variables express common information, or even collinearity, which is the case for spectral variables. The main peak is detectable at 680 nm, corresponding to the absorption peak of chlorophyll. The main peak is detectable at 680 nm, corresponding to the absorption peak of chlorophyll.

Plastic cap to ensure contact with sample skin during measurements, while minimising environmental light interference.

To evaluate model accuracy, the statistics used were the coefficient of determination (R²), coefficient of determination in cross-validation (R²c), root mean square error of calibration (RMSEC), and root mean square error of cross-validation (RMSECV). Cross-validation is an internal validation method, usually used in the case of a small number of samples available for regression. With cross-validation, some samples are kept out of the calibration and used for prediction. This is repeated until all samples have been kept out once. In this case, full cross-validation was used, so only one sample at a time is kept out of the calibration.

Coefficient of determination (R² and R²c):

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$  \hspace{1cm} (1)

RMSEC AND RMSECV:

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

where:

- $n$ is the number of validated objects, and $y_i$, $\hat{y}_i$, and $\bar{y}$ are the predicted and measured values of the $i^{th}$ observation in the calibration or validation set, respectively.

Percentage errors of cross-validation (RMSECV) were also calculated as:

$$\text{RMSECV}(\%) = \frac{\text{RMSECV}}{\text{averaged reference values of each parameter}}$$

Results and discussion

Qualitative analysis

Figures 1 and 2 show average spectra of the three analysed classes of SSC and texture, respectively. A very different trend could be noticed among classes both in the area of the visible region (400-700 nm) and in the NIR region (700-900 nm) for each reference parameter considered. The main peak is detectable at 680 nm, corresponding to the absorption peak of chlorophyll.

A correlation between reflectance in the visible band and the reference parameters considered can be observed.

As expected, the average spectrum demonstrates significant differences among the three classes, with dramatic changes in the visible spectral range, from green berries ($a$) to the completely black-pigmented olives ($c$), especially linked to anthocyanin accumulation. This leads to a decrease in reflectance in the visible band associated with the anthocyanin absorption peak centred around 540 nm. Accordingly, the green samples, with anthocyanin content near to zero, reflect more light than pigmented olives.

Firmness decreases in parallel with a decrease in reflectance absorption in the visible range, until reaching a minimum and then remain-

$y_i$ are the reference values, $\hat{y}_i$ are the values predicted by the PLS model, and $\bar{y}$ is the averaged reference value.

RMSECV (%) = RMSECV / averaged reference values of each parameter.

Figure 1. Average raw spectra of 109 olives grouped in three classes of soluble solid content. Bars indicate the standard error within each group at different wavelengths ($a<17; 17.1<19; 19.9<20$).
PCA was carried out on spectra, and grouped according to SSC and texture ranges.

Ninety-nine percent of the total data variance is explained by the first three principal components (PCs). In particular, PC1 explained 93% of the variability, PC2.5% and 1% is explained by PC3. The combination plan between PC1 and PC2 is shown in Figures 3 (SSC grouping) and 4 (texture grouping).

Concerning SSC, a fairly good sample separation in the PCs plan was obtained. PC1 allows a good separation between olives belonging to classes a and c. The PCs plan highlights the presence of a few outliers belonging to class a that are placed in class c. Less obvious is the association of samples b to a distinct group. In fact, samples belonging to class b are divided into nearly equal parts along the PC1. The 42% of samples of class b has negative values of PC1 while 58% of them have positive values. The PCA highlighted an increase in SSC in olives from high positive values to high negative values of PC1.

Regarding texture, the explorative PCA conducted on the spectra of 109 olives reveals that PC1 is largely accountable for separating class c corresponding to negative values of PC1 from classes a and b, placed, in the PC plan, at positive values of PC1.

According to Yousfi et al. (2006), firmness tends to reduce during maturation and the PCA shows the ripening trend in olives from high positive values to high negative values of PC1.

PCs loadings (Figure 5) were analysed in search of main wavelength bands ranges mostly contributing to PC1 and PC2, as candidate discriminators for the class identification. For both studied parameters, the three classes are better detected on PC1. Two main waveband ranges were identified: i) 550-650 nm, in the visible range, where relatively high positive values for PC1 and the maximum positive values of PC2 were found together; ii) 700-750 nm, in the near infrared range, where the two extrema values for both PC1 (maximum) and PC2 (minimum) were found.

Quantitative analysis

Concerning quantitative analysis, PLS regression models were created for each parameter (Figure 6).

Table 2 shows the results for the PLS regression models for the prediction of SSC and texture.

The results are encouraging. In this preliminary study, the data set available was not particularly wide, so results could be improved with more samples for the elaboration. For example, to verify the robustness of models, an external validation set is required. Regarding SSC, the performance of the regression model can be improved although there...
and classifying olives. The employment of more accurate instrumentation for the reference data, such as a laboratory texture analyser, for the reference data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data set available and, consequently, the robustness of prediction models to confirm these early results to increase the number of samples of the data st
in olive fruit characteristics and oil accumulation during ripening process. J. Sci. Food Agr. 84:1783-90.


