UNIVERSITÀ DEGLI STUDI DI MILANO LA STATALE

Faculty of Agricultural and Food Sciences

Department of Food, Environmental and Nutritional Sciences

PhD in

Food Systems

XXXIV Cycle

Feasibility studies and engineering of optical simplified and stand-alone devices for agri-food applications

Thesis of

Alessio Tugnolo

R12341

Supervisor: Prof. Ing. Riccardo GUIDETTI

PhD Coordinator: Prof. Diego MORA

Academic Year 2020-2021

To Stefania

Contents

1. ABSTRACT

The food industry needs to comply with more strict rules (from regulatory agencies) and meet customers' demands for higher quality. Emerging technologies for quality and safety inspection are becoming fundamental and needful to fulfil these purposes.

The optical techniques have been using in different fields (agriculture, food, chemistry etc.). During harvest, post-harvest and food processing these techniques are well applied to predict crucial quality parameters. Spectra of intact food samples can be measured in few seconds without any sample preparation and expert personnel trained. Compared to this technology, chemical techniques are time-consuming, require sample preparation and use of chemical reagents which are often not sustainable for the environment.

This PhD project regards different applications of non-destructive optical techniques to evaluate the quality of agri-food products as well as the development of customized optical devices to fulfil the needs of agri-food chain which is going toward a concept of industry 4.0.

This thesis starts with a focus on optical sensing in terms of concept, data management and general applications in agri-food chains highlighting the attitude to be a green technology. Then, in two chapters were set the stage for models' development using commercial benchtop and portable optical devices to enhance the advantages of this technique in terms of performance (compared to other analytical instruments) and versatility in coffee industry and in the olive supply chain. Moreover, thanks to the need of the winemaking industry to improve the production of high-quality wines, another chapter was developed to show the latest frontiers in terms of data collection for maturation control of wine grapes. Therefore, optical hand-held and stand-alone prototypes were designed, built, and tested in order to shift the current paradigm of grape maturation monitoring (based on lab analysis) with a new one that allows a low-cost non-destructive real-time monitoring providing information with temporal and spatial resolution.

Finally, a last chapter has been introduced as an initial step for future developments in the field of hyperspectral imaging sensors. Therefore, a cost-effective hyperspectral device was developed to drastically reduce the cost of these instruments comparing it with those present on the market. To clarify, for these instruments the cost limitations are not strictly related to the device itself but for the specific applications. Indeed, even though the hyperspectral imaging technique can collect a large amount of data, the application of only one device (in some cases) is not enough to cover all the critical points the industry has to face. The production process in a firm or the monitoring in field require distributed systems to collect data and provide information. In these circumstances, considering the application of several hyperspectral devices, the costs become prohibitive for the majority of the companies and the research is

going toward the development of hyperspectral sensors taking into account a considerable cost reduction.

To conclude, this PhD project has proved advantages and frontiers of optical sensing as one of the most efficient and advanced tools for safety and quality evaluation in the food industry throughout the entire production process.

Keywords: spectroscopy, hyperspectral imaging, chemometric, food quality, green technology, sensors

2. INTRODUCTION AND LITERATURE REVIEW

2.1 Fundaments of optical sensing

Optical sensing is defined as the science of acquiring, processing, and interpreting data that record interactions between electromagnetic radiation (composed by photons) and a specific target (Borràs *et al.*, 2015). These interactions involve reflection, absorption, and transmission of a flux of photons by the target matter, in addition to emission of radiation and fluorescence (Corti *et al.*, 2018).

Optical techniques, based on infrared (IR), ultra-violet (UV), visible (vis), fluorescence, are widely used in agriculture and for food fingerprinting. The light sent to the sample interacts with it in different ways, a part is absorbed, a part emitted, others scattered and transmitted (figure 1). Most of these techniques offer the possibility of analyze a sample (or sample extract) in a non-destructive way allowing a simultaneous determination of several properties in the sample (Tugnolo *et al.*, 2019) .

In the IR region, near-infrared (NIR, 4000–14,286 cm^{-1}; 700–2500 nm) and mid-infrared (MIR, 400–4000 cm^{-1}; 2500–25,000 nm) spectroscopy involve the absorbance of radiation at molecular vibrational frequencies occurring for the O-H, N-H, and C-H groups and for the C-C, C-O, C-N, and N-O groups in organic materials (Malegori *et al.*, 2018; Tugnolo *et al.*, 2020). Instead, electronic transitions absorb in the vis region (14,286–25,000 cm^{-1}; 400–700 nm) and in the UV region (25,000–40,000 cm^{-1}; 250–400 nm) (Buratti et al., 2017; Gomez-Caravaca et al., 2016).

Moreover, the region from 0.75 to 100 μm is used as a non-destructive technique (thermography) to monitor the temperature evaluating the IR radiation emitted by an object. The thermography provides key facts on the dimension, heat distribution as well as structural analysis (Fernández-Cuevas *et al.*, 2015). The main factor which influences the amount of radiation is the emissivity (energy ratio emitted from a sample at the surface temperature). The emissivity depends on the wavelength of the infrared region, temperature as well as the surface of the sample (Mohd Ali *et al.*, 2020).

The resulting information derived from the interaction between photons and the target can be handled in terms of punctual information (spectroscopy) or image information (thermal imaging and hyperspectral/multispectral imaging). The advantaging of imaging techniques is characterized by the presence of the spatial resolution (Sx and Sy) which measures the geometric relationship between the image pixels. In thermography, the information retrieved from the thermal image could be used to describe the thermal distribution without exerting any energy to the sample. The technique employs the thermal distribution of an image that produces a thermal map or thermogram of the object (Still *et al.*, 2019). In addition to the spatial

resolution, the hyperspectral and multispectral imaging techniques also handle the spectral resolution (Sλ) which measure the variations in illumination within the image pixels as a function of wavelength (Khan *et al.*, 2018). These data are represented in the form of a 3-Dimensional hyperspectral data cube where each slice of this data cube along Sλ, represents a specific band from the electromagnetic spectrum (Stuart, McGonigle and Willmott, 2019; Amigo, 2020).

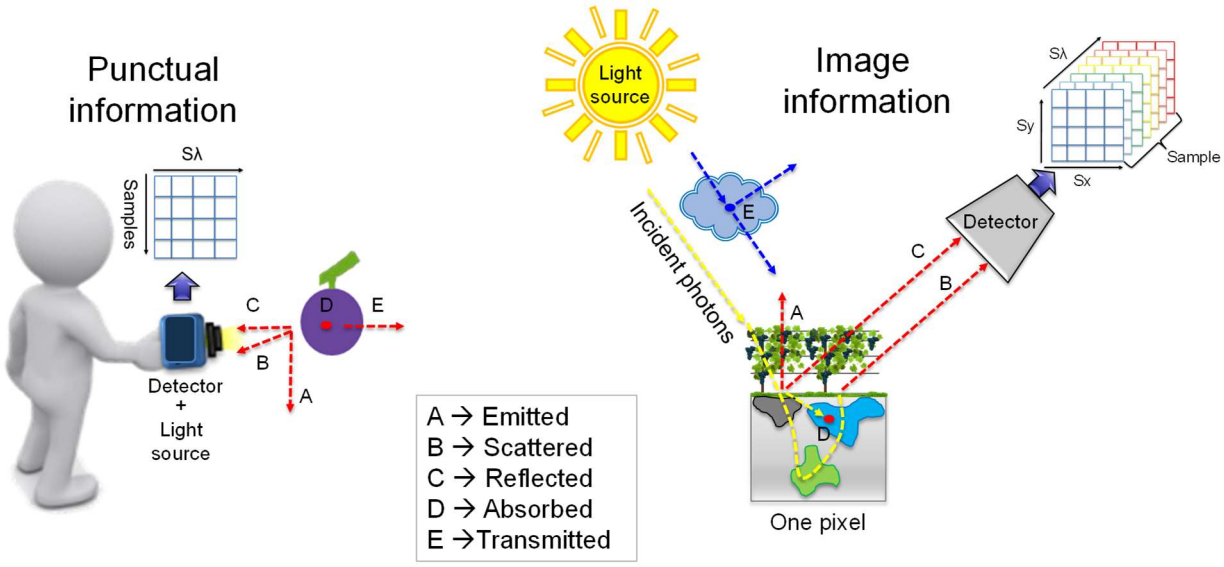


Figure 2.1 Logic scheme to highlight differences between the acquisition of punctual information and image information.

2.2 Use of the optical data

Nowadays, optical instruments provide rapidly a large amount of data. However, this large amount of information (dataset) is hard to handle and needs to be pre-treated for a correct evaluation (Cortés *et al.*, 2019). For this purpose, chemometrics is one of the most powerful and widely applied technique capable to manage these chemical data in a multivariate way ensuring that experimental data contain the maximum information (Bro and Smilde, 2014).

Chemometric techniques are normally applied to data structures represented by a bidimensional (second order tensor) or three-dimensional (third order tensor) matrix (Figure 2.1). On the data matrix, before carrying out any type of processing, mathematical pre-treatments, normalization and/or scaling are needed. Then, the process continues with: (i) exploratory data analysis (which summarize the main information contained into the data), (ii) model calibration and validation and (ii) model transfer (Cortés *et al.*, 2019).

The main objective of spectra pre-treatment is to reduce the noise associated with the data, minimize artifacts, and amplify the information. Many smoothing methods (for example, Gaussian filter, moving average, median filter, and Savitzky-Golay smoothing) are used as a filter to reduce spectral noise. Additive and multiplicative effects are very common in optical data influencing the global intensity (typically arising from unwanted light scattering) and/or producing baseline vertical shifts (offsets). For these reasons, treatments like Standard Normal Variate (SNV) transform or the Multiplicative Scatter Correction (MSC) are generally applied. Derivation methods (usually first and second derivative) are also used to enhance the resolution and minimize the spectra offset and drifts. Finally, different normalization and/or scaling treatments (like mean centering, autoscaling, range scaling, pareto scaling etc.) become fundamentals to homogenize the data in order to perform a correct explorative and modelling phase (Boulet and Roger, 2012; Biancolillo and Marini, 2018; Oliveri *et al.*, 2019). Moreover, different combinations of these methods applied simultaneously can also be used for signal processing (Rinnan, 2014).

Then, with the aim of (i) extracting useful information, (ii) correlating the variables, (iii) eliminating anomalous data and (iv) hypothesizing the subsequent work procedures an explorative unsupervised Principal Component Analysis (PCA) is essential (Bro and Smilde, 2014; Tefas and Pitas, 2016). Afterwards, supervised regression and classification techniques can be applied according to the specific application (Todeschini, 1998). A multivariate regression/classification method provides the best relationship between a set of independent variables (optical outputs) that describe the objects (predictors) studied and a set of measured response for the same objects.

Regression methods are focused on predicting some of the properties of the sample (e.g., soluble solids, acidity, anthocyanins, polyphenols, water potential etc.). For these purpose methods like Multiple Linear Regression (MLR), Principal Component Regression (PCR), Partial Least Square regression (PLS), Support Vector machine Regression (SVR) or Artificial Neural Network (ANN) are broadly used (Liakos *et al.*, 2018).

The regression models' accuracy is usually evaluated with Root Mean Square Error (RMSE), as well as bias and R^2 (coefficient of determination); the lower the error, the nearer to zero is the bias and the higher the R^2 (as maximum equal to 1), the better the model performances. Besides, RPD (ratio between the standard deviation of the response variable and RMSE) gives an immediate description of the prediction accuracy (RPD between 1.5 and 2 means that the model can discriminate low from high values of the response variable; a value between 2 and 2.5 indicates that coarse quantitative predictions are possible, and a value between 2.5 and 3 or above corresponds to good and excellent prediction accuracy) (Nicolaï et al., 2007).

The whole set of parameters for the evaluation of model goodness have to be calculated not only in calibration but also in validation (cross-validation and/or test-set), allowing to choose the best number of latent variables for maximizing model reliability, balancing good predictions and overfitting.

Differently from regression, a multivariate classification model is created with a set of samples belonging to known categories (class), and subsequently (as for regression) the model is evaluated by cross-validation (internal validation) or test set (external validation). Many methods like Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), Soft Independent Modelling by Class Analogy (SIMCA), Support Vector Machine (SVM) and Partial Least Square regression – Discriminant analysis (PLS-DA) are used to define membership of each sample to its appropriate class. To evaluate the classification performance, many indexes commonly derived from the confusion matrix (matrix where the entries on the main diagonal represent the number of correct class assignations, while off-diagonal entries represent classification errors) are used (error rate, non-error rate, class specificity and sensitivity etc.) (Ballabio and Consonni, 2013; Oliveri and Downey, 2013; Zhong *et al.*, 2018).

Finally, clustering is another unsupervised method typically used to find natural groupings (clusters). It is the process of grouping the data into clusters, so that the objects within the same cluster have higher degree of similarity in comparison to one another but are very much dissimilar to the objects in different clusters. There are several clustering techniques available and those are organized into the following categories as partitioning methods, hierarchical methods, density-based methods, grid-based methods, model-based methods, methods for high-dimensional data and constraint-based clustering (Bonilla, De Toda and Martínez-Casasnovas, 2015; Liakos *et al.*, 2018).

2.2.1 Multivariate analytical methods

As mentioned above, chemometrics is a powerful tool able to study at best the information coming from a large amount of analytical variables. In this chapter, a more detailed description of the two most important methods of data exploration and regression has been done in order to give an initial overview of the main statistical approaches used in this PhD thesis.

Principal Component Analysis (PCA)

PCA is an exploratory data analysis which is at the base of the multivariate approach. The starting point of the multivariate data analysis is a data matrix (a data table) denoted by “X”. The “n” rows in the table are defined “objects” and “k” columns are the “variables” related to the measurements performed on the objects (S. Wold 1995).

Currently, large datasets are increasingly widespread in many disciplines. In order to understand these datasets, methods which drastically reduce their dimensionality in an interpretable way and preserve the information contained into the data are needed. One of them is PCA which reduces the dimensionality of the dataset while preserving as much 'variability' (i.e. statistical information) as possible. This translates into finding new variables, the principal components (PCs), that are linear functions of those in the original dataset, that successively maximize variance and that are uncorrelated with each other (Jolliffe & Cadima, 2016).

From a geometrical point of view, this method consists of a rotation of the original data matrix, carried out in such a way that the first new axis is oriented in the direction of maximum variance of the data (PC1), the second (PC2) is perpendicular to the first and is in the direction of the next maximum variance direction and so on up to a number of PCs equal to the number of original variables. Technically, the maximum PCs amount is equal to the number of original variables, but practically, only the first components include the useful information, as variables co-vary, especially in spectroscopic data, and further PCs only contain random noise. Since the main components are the axes relative to the direction of maximum variance in descending order, as the number of PCs increases, the percentage of variance which they explain decreases.

In particular, each PC is composed by the product of two vectors: the scores vector (t_i), and the loadings vector (p_i), where i represents the number of components s expressed in the equation below

Equation $$\boldsymbol{X} = \boldsymbol{t_1 p_1^T} + \boldsymbol{t_2 p_2^T} + \cdots + \boldsymbol{t_i p_i^T} + \boldsymbol{E} = \boldsymbol{T P^T} + \mathrm{E}$$

X is the original data matrix with observations and variables, whereas **E** is the residual matrix, that contains the unmodelled variation and has the same dimensions of **X**. The scores matrix **T** contains information on how each sample relates to each other, whereas the loading matrix **P** expresses the influence of the measured variables on the scores. PCs number **i** defines the amount of variation in the data, i.e. the independent phenomena. In figure 2.2 it is shown the PCA decomposition scheme.

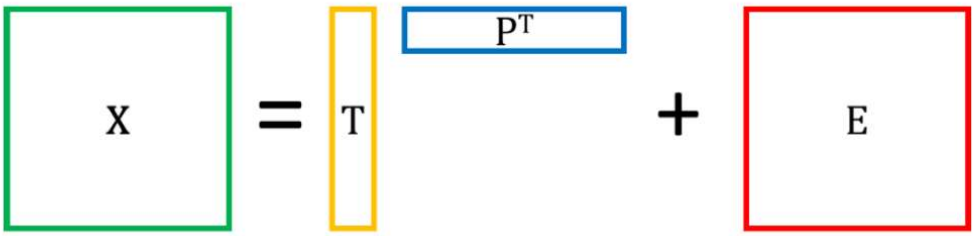


Figure 2.2 PCA decomposition of X.

Mathematically, components are calculated by an eigenvalue decomposition of the covariance matrix:

- Equation $\boldsymbol{cov(X)= X^T X/(m-1)}$
- Equation $\boldsymbol{cov(X)p_i= p_i\lambda_i}$

where $\boldsymbol{\lambda_i}$ are the eigenvalues related to the eigenvectors $\mathbf{p_i}$. The scores vectors $\mathbf{t_i}$ are the projections of the data matrix **X** onto $\mathbf{p_i}$. T^2 and Q parameters are used to evaluate PCA models. T^2 statistic represents the distance of a sample in the model space (i.e. in the space of significant PCs), whereas Q statistic represents the distance of a sample from the model space, meaning that the PCA model cannot describe efficiently their variability (Figure 13). These two parameters are extremely useful for outliers detection.

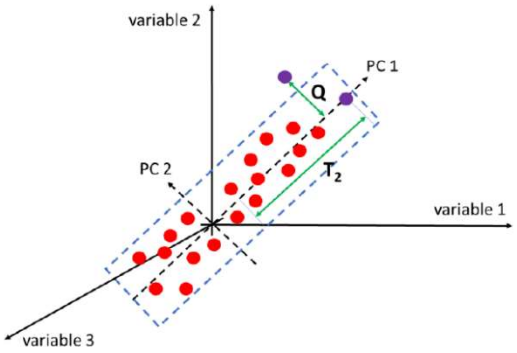


Figure 2.3 Graphical representation of T2 and Q statistics.

Through this technique, it is possible to: evaluate the correlation between variables and their importance, view the samples (identify outliers, classes, etc.), summarize the data description (elimination of the error or redundant information), reduce the size of the data and search of the main properties (Todeschini, 1998)

Finally, in the study of multivariate problems, an aspect of great importance concerns the possibility to graphically visualize the data. With PCA is possible to represent the original data matrix as the product of two matrices: scores and loadings. The loadings visualize the role of each variable to describe the PCs and also identify any correlations. In the graph, the values are between 0 (origin of the plane) and 1/-1 (end of the axis). Thus, the elements very close to the origin are not informative and the variables far from the origin will have greater variability and therefore greater relevance to describe the PC. Instead, the scores plot is obtained by combining the original variables and the loadings. The values of the scores represent the new coordinates of the objects in the score plot thus It shows the behaviour of the original objects (samples) in the space of the PCs (Todeschini, 1998).

Multivariate regression

A multivariate regression method provides the best relationship between a set of independent variables that describe the objects (predictors) studied (x_1, x_2,..., x_n) and a set of measured response (y) for the same objects.

$$y = f(x_1, x_2,..., x_n)$$

On the one hand, the form of the equation below describes how the description of the system is linked to the experimental measurement (fitting), and, on the other, the model obtained (once validated) allows to predict the future value of the objects.

$$y=\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n+\varepsilon$$

The equation above shows the general form of a multivariate calibration model where:

y=experimental response;

$x_1,x_2,\ldots x_n$=values of the independent variables experimentally acquired;

β_0, β_1, $\beta_2\ldots$ β_n=regression coefficients of the linear model;

ε=experimental noise.

The Partial Least-Squares regression (PLS) is one of the most used linear regression methods which model the covariance structure between the matrix of predictors (X) and the matrix of the response (Y). PLS operates a simultaneous decomposition of both X and Y matrices in order to explain as much as possible of the variability of X and to find the best correlation with Y. Its aim is to maximize the covariance between the two matrices, creating at the same time latent variables that describe the maximum variability of X. The algorithm performs the decomposition of X and Y as in in PCA:

- Equation $\boldsymbol{X=TP^T+E}$

- Equation $$Y=UQ^T+R$$

The maximum covariance criterion is imposed through a regression model for each component between the scores of X (t_i) and Y (u_i), obtaining their inner relation:

- Equation $$u_i=b_i t_i$$
- Equation $$b_i= u_i^T t_i/(t_i^T t_i)$$

where **b** is the regression coefficient related to the **i**th component. However, this is not the best possible strategy, as the components are calculated separately for each block, resulting in a weak relation between them. For this reason, inner relation is improved by rotating the components, which means making **t** and **u** switch places in the NIPALS algorithm. As final step, scores are orthogonalized by introducing loading weights, **W**, which are orthonormal, calculated according to the equation below:

Equation $$W= T^T X$$

Scores and loadings have the same properties described for PCA, but in PLS the latent variables (LV) explain the variability of X that most influences the responses predictions in Y. The choice of LVs number is critical: on the one hand is essential to include all the useful information to obtain better results, on the other hand LVs that contain only noise and irrelevant information should not be included. In fact, choosing to many LVs leads to the creation of a model that is perfectly suited to explain the variability of the data used to create it, even managing to represent its noise, but it would be hardly adaptable for the prediction of unknown samples, thus resulting not functional. The most used method for the determination of the optimal number of LVs is the cross-validation. Once the model is created, it must be validated, i.e. subjected to a test that demonstrate its reliability. For this purpose, it is strongly recommended to use samples that were not used in the calculation of the model itself, in order to avoid an overly favourable estimate of the error. The previously generated prediction algorithm is applied on this new set of samples (test set), whose characteristics are known, in order to evaluate the predictive capability in terms of error on the real data (Todeschini, 1998).

2.3 Proximal sensing applications overview

The increasing importance of vis/NIR spectroscopy in pre and postharvest technology is obvious from the increase in numbers of publications, as well as from the fact that many manufacturers of on-line grading lines have now implemented vis/NIR systems to measure various quality attributes. The objective the paper below (Paper 1) is to give a comprehensive overview of vis/NIR spectroscopy for measuring quality attributes of agricultural products.

PAPER 1: Visible/near infrared spectroscopy for horticulture: case studies from pre-harvest to post-harvest

R. Beghi[a], V. Giovenzana, A. Tugnolo and R. Guidetti

Department of Agricultural and Environmental Sciences (DiSAA), Università degli Studi di Milano, via Celoria 2, 20133 Milano, Italy

[a] roberto.beghi@unimi.it

Abstract

The Italian agri-food system is mostly characterized by small and medium-sized enterprises (SME). They are characterized by the lack of crucial information for the management and control of their processes. The main parameters related to important information to better lead the decision-making stages of the production process (e.g. monitoring the shelf life and post-harvest life, etc.) need to be measured in a simple, non-destructive and quick way. From this point of view, the optical techniques based on visible near-infrared spectroscopy (Vis/NIR) are an established, simple and rapid application for the determination of many parameters related to productions quality representing a valid support to the various supply chains. Nevertheless, the available optical systems are complex and expensive devices and their real use in SME is still very limited. The aim of the research was to test commercial hand-held devices and prototypes of low-costs and user-friendly systems, useful for different applications from pre-harvest to post-harvest. This work proposed a collection of case studies regarding the application of Vis/NIR spectroscopy on different matrices with different goals: on (i) grape and (ii) blueberry for ripening evaluation, on (iii) cultivated mushrooms to control quality at harvest, and (iv) for senescence monitoring of fresh-cut baby leaf salad during shelf life. A Light Emitting Diode (LED) based prototype at four specific wavelengths (630, 690, 750 and 850 nm) was proposed and tested on grape for estimating ripeness analysis. Moreover, a commercial hand-held Vis/NIR device (400-1000 nm) was used for the evaluation of blueberries ripeness, for the postharvest monitoring of mushrooms, and for the senescence evaluation of fresh-cut baby leaf salad. The use of these devices could be helpful for the monitoring of quality standards, to provide a high quality product to the consumers.

Keywords: handheld device, fruits and vegetables, ripeness, chemometrics

Acta Hortic. 1311. ISHS 2021. DOI 10.17660/ActaHortic.2021.1311.16
Proc. VI Int. Symp. on Applications of Modelling as an Innovative Technology in the Horticultural Supply Chain - Model-IT 2019
Eds.: M.L. Amodio and G. Colelli

INTRODUCTION

The Italian agri-food system is mostly characterized by small and medium-sized enterprises (SME). Due to their size and the small possibility of significant investments in analytical instrumentation, they are characterized by the lack of crucial information for the management and control of their processes. The main parameters related to important information to better lead the decision-making stages of the production process (e.g. decide the harvest time, monitoring fermentations, monitoring shelf life and post-harvest life, etc.) need to be measured in a simple, non-destructive and quick way. By sorting harvested products according to their maturity, it is possible to split immature-green, partially mature, and fully mature products, in order to obtain the uniformity of lots at destination. Thus, the assessment of the maturity stage is of great importance for determining the optimal postharvest strategy for product handling and marketing (Slaughter, 2009). Nowadays, ripening stage and harvest time are estimated mainly by visual assessment based on growers' experience, which can result in unsatisfactory homogeneity of the fruit lots and large variability from farm to farm. Operators need simple and portable devices capable to provide fast ripeness assessment of a large set of samples in the field to help them to determine the optimum harvest time.

Consumers' interest in berry fruits has recently increased together with the growing demand for health-producing foods, considering the berries' bioactive and health-protecting properties. Due to these aspects, blueberries, blackberries, and raspberries are considered a rich source of antioxidant compounds, particularly of flavonoids and anthocyanin (López et al., 2010). Moreover, the growing demand for these fruits may represent an important income generating opportunity for marginal rural areas, because their cultivation is particularly suited for small farms and mountain regions.

The non-destructive techniques, and in particular the optical analysis in the region of visible-near infrared (Vis/NIR) and near-infrared (NIR) and, have been developed considerably over the last 20 years (Guidetti et al. 2012; Nicolai et al. 2007). Until now, many studies reported in the literature have evaluated the applicability of Vis/NIR and NIR spectroscopy to obtain an effective system to perform a wide check on fruits. For example, several applications have been developed to estimate ripeness parameters of different fruit species, especially on grapes (Cozzolino et al., 2006).

Ready-to-eat products are characterized by a shelf-life shorter than that of the original unprocessed raw material. These foods are often subjected to rapid loss of colour, organic acids, vitamins and other compounds that determine flavour and nutritional value. Monitoring

Acta Hortic. 1311. ISHS 2021. DOI 10.17660/ActaHortic.2021.1311.16
Proc. VI Int. Symp. on Applications of Modelling as an Innovative Technology in the Horticultural Supply Chain - Model-IT 2019
Eds.: M.L. Amodio and G. Colelli

the quality decay of fresh-cut products is necessary to control the freshness level during the entire production chain and to ensure quality product for the consumer. Hence, also fresh-cut fruit and vegetable sector could be greatly helped by new analytical methods that are accurate, rapid and could be integrated into the production chain to better managing the shelf life of minimally processed products and to meet consumer demand.

These optical approaches, however, are always related to the analysis of wide spectra (thousands of wavelengths or variables) and, therefore, require multivariate techniques for data processing to build predictive models. To explain the chemical information encoded in the spectral data, chemometric analysis is required (Cogdill and Anderson, 2005). Moreover, the limitations to NIRs adoption by the agricultural sector could be attributed to cost, technical limitations, grower resistance and supply chain weakness (Magwaza et al., 2012). Hence, to avoid these limitations, simplified handheld devices are desirable. For a simplification and greater diffusion of these non-destructive techniques, in recent years interest has shifted towards the development of portable systems that could be used in pre- and post-harvest (Beghi et al., 2017).

Portable vis/NIR instruments have been tested in controlled laboratory conditions or directly in the field. Concerning laboratory applications, Puangsombut et al. (2012) investigated the feasibility of using Vis/NIR spectroscopic devices in transmission mode to predict the soluble solids content (SSC) and total acidity of fresh-cut products. Regarding applications in uncontrolled field conditions, Larrain et al. (2008) developed a portable NIR instrument (640-1300 nm) for determining ripeness in wine grapes, Guidetti et al. (2008) tested a Vis/NIR device working in reflectance configuration to predict blueberry ripeness, Guidetti et al. (2010) evaluated grape quality parameters with a portable Vis/NIR system, finally, Beghi et al. (2013) assessed the nutraceutical properties of apples through a visible and near-infrared portable system.

All these approaches, nevertheless, rely always on wide spectral ranges (thousands of wavelengths) and thus require multivariate techniques for data processing to build prediction models. Chemometrics can be also used for the selection of a small number of relevant variables, which represent the most useful information contained in the full spectra (Xiaobo et al., 2010). In this way the spectral noise and the variables containing redundant information can be eliminated. Moreover, a reduced cost for potential miniaturized devices, realized to work at only these selected wavelengths, can be foreseen.

Acta Hortic. 1311. ISHS 2021. DOI 10.17660/ActaHortic.2021.1311.16
Proc. VI Int. Symp. on Applications of Modelling as an Innovative Technology in the Horticultural Supply Chain - Model-IT 2019
Eds.: M.L. Amodio and G. Colelli

Few examples of commercial non-destructive devices based on a small number of wavelengths are already available on the market. These applications are mainly dedicated to fruits. For example, the University of Bologna (Costa et al., 2011) patented innovative and simplified NIRs equipment, namely DA-Meter for apple and Kiwi-Meter for kiwi. These systems are used for the analysis of the ripeness level of the fruit through indices based on differences in absorbance between specific wavelengths. This type of instrument, simple and portable, can be used directly on the fruit on the trees and can help growers in taking decision regarding the best cultural management practices (such as pruning, thinning, nutrition, etc.). In this way the heterogeneity of the product can be reduced and, therefore, can be simplified the management of lots product during post-harvest.

The aim of this group of researches was to apply Vis/NIR spectroscopy and to design, build and test concepts and prototypes of low-costs and user-friendly devices useful for different case-study applications from pre-harvest to post-harvest. The authors conducted tests on grape (Civelli et al., 2015; Giovenzana et al., 2015; Giovenzana et al., 2014) and blueberry (Beghi et al., 2013) ripening evaluation, on cultivated mushrooms to control postharvest (Giovenzana et al., 2019), and on fresh-cut baby leaf salad during shelf life for senescence monitoring (Beghi et al., 2014).

MATERIALS AND METHODS

Different experimental setups were carried out for the different matrices considered.

Samplings

Evaluation of grape ripeness

The experimentation was carried out on *Vitis vinifera* L., Nebbiolo cultivar (ecotype Chiavennasca) in the Valtellina viticultural area (Sondrio, northern Italy, approx. 46.2° N, 9.9° E) during the last period of ripening of the 2012 season. Samples were drawn from 17 different zones, throughout the entire wine area of the valley, in order to represent environmental variability and monitor the entire production region of DOC (controlled denomination of origin) and DOCG (controlled and guaranteed denomination of origin) wines. A total of 68 samples of red grapes were collected on four sampling dates.

Acta Hortic. 1311. ISHS 2021. DOI 10.17660/ActaHortic.2021.1311.16
Proc. VI Int. Symp. on Applications of Modelling as an Innovative Technology in the Horticultural Supply Chain - Model-IT 2019
Eds.: M.L. Amodio and G. Colelli

The total soluble solids (TSS) content was measured using a digital pocket refractometer while the Glories method (Glories, 1984) was used to measure the total polyphenols (TP). The polyphenols quantification was based on the optical density (OD) measurement at 280 nm using a UV/Vis spectrophotometer.

Evaluation of blueberries ripeness

In this case a total of 942 Vis/NIR fruit spectra in the range of 445-970 nm were acquired for *Vaccinium corymbosum* L. ("Brigitta" cultivar) during two different growing seasons (2009 and 2010), cultivated in different farms in the area of Valtellina, Italy.

Field spectra acquisitions of berries were carried out at weekly intervals. Spectra measurements were taken in the field on individual berries along their equator region after artificial illumination with the probe tip. Just after measurement of the plants, fruits samples were picked and divided into four ripeness classes according to the criteria used by expert growers (mainly relying on size and surface colour distribution). According to commercial classification, blueberry fruits were graded in four classes, from totally unripe to fully ripe.

Postharvest control of cultivated mushrooms

Samples of *Agaricus bisporus* were randomly harvested from three different flushes. The ripening process was conducted in conditioned wide rooms, without any windows. The environmental conditions of the room were controlled regarding temperature, CO_2 and relative humidity. The mushrooms grew on bed of specific soil placed on steel structures. The mushrooms were considered ready for the harvest when they reached the specific dimensional standards required by the retailer.

Samples' variability was increased according to an accelerated shelf-life program. The first analysis (t0) was performed the day of sampling directly at the production company. After that, the samples were stored at room temperature (24-26 °C with 60-70 % of relative humidity) for three (t1), five (t2) and seven (t3) days following the harvest in order to simulate an accelerated qualitative decay process. At the end of the accelerated shelf-life program the number of samples available was 167.

The reference parameters were evaluated after the Vis/NIR analysis using the traditional destructive methods. The analysed physico-chemical parameters were: firmness, TSS and water content (WC).

Acta Hortic. 1311. ISHS 2021. DOI 10.17660/ActaHortic.2021.1311.16
Proc. VI Int. Symp. on Applications of Modelling as an Innovative Technology in the Horticultural Supply Chain - Model-IT 2019
Eds.: M.L. Amodio and G. Colelli

Senescence monitoring of fresh-cut baby leaf salad

Three storage/shelf-life temperatures were investigated: 4.0±0.5 °C, 10.0±0.5 °C and 20.0±0.5 °C. The duration of the experimentation was different for the different storage temperature. The *Valerianella locusta* L. packages preserved at 4 °C and 10 °C were sampled for 16 days, while samples stored at 20 °C were analysed only for 7 days, due to the rapid degradation of the lettuce at this temperature. The numbers of sampling points during storage monitoring for the fresh-cut leaves were therefore 10, 11 and 6 for 4 °C, 10 °C and 20 °C, respectively. The quality decay of samples was evaluated by chemical parameters (pH, moisture and total polyphenols content).

Vis/NIR spectroscopy measurements

Spectral acquisitions were performed on samples using a Vis/NIR portable system operating in the wavelength range of 400-1000 nm, spectral resolution 0.3 nm (Jaz, OceanOptics, USA). Spectra were acquired in reflectance mode: light radiation was guided from the light source to the surface of the sample through a Y-shaped, bidirectional fiber-optic probe (OceanOptics, USA).

Data analysis

Chemometric analyses were performed using The Unscrambler® 9.8 software package (CAMO ASA, Norway). The correlation between the spectral data matrix and the reference parameters were carried out using the Partial Least Square (PLS) regression algorithm. Depending on the dimension of the available datasets, leave-more-out cross-validation with different cancellation groups was used to validate the PLS regression models. To evaluate model accuracy, the statistical parameters used were the coefficient of determination in calibration (R^2_{cal}), the coefficient of determination in cross-validation (R^2_{cv}), the root mean square error of cross-validation (RMSECV). The best calibrations were selected based on minimizing the RMSECV.

In some cases, Regression Coefficient Analysis (RCA) was carried out for relevant variable selection, representing the most useful information of the full Vis/NIR spectral region (Xiaobo et al., 2010, Chong and Jun, 2005). Standardized regression coefficients of the PLS model were used for the elaboration. The standardization considered both the standard deviation of reflectance and the standard deviation of the reference data. Therefore, the selected wavelengths were employed for the elaboration of Multi Linear Regression (MLR) models. MLR is a regression method that, compared with PLS, allows to develop models using only

Acta Hortic. 1311. ISHS 2021. DOI 10.17660/ActaHortic.2021.1311.16
Proc. VI Int. Symp. on Applications of Modelling as an Innovative Technology in the Horticultural Supply Chain - Model-IT 2019
Eds.: M.L. Amodio and G. Colelli

few important variables. Verifying the prediction ability of the MLR models was performed to study the efficiency of the selected wavelengths.

RESULTS AND DISCUSSION

Results for grape ripeness evaluation

Correlations between the spectral data and technological (total soluble solids, TSS) and phenolic (polyphenols, TP) parameters were carried out using partial least square (PLS) regression. Standardized regression coefficients of the PLS model were used to select the relevant variables, representing the most useful information of the full spectral region. Ratio Performance Deviation (RPD) values were also calculate to evaluate models performance.

The three selected wavelengths were 670 nm, corresponding to the chlorophyll absorption peak, 730 nm, equal to the maximum reflectance peak, and 780 nm, representing the third overtone of OH stretching. MLR was applied on the three selected wavelengths to verify their effectiveness. A comparison between PLS derived from the full Vis/NIR spectra and MLR arising only from the three wavelengths was carried out on TSS and TP. The overall calibration and prediction results of the MLR models, both for TSS and TP prediction, were satisfactory, although the performance of the MLR models (Table 1) were slightly worse than the good PLS models. The obtained RMSECV values were similar for the PLS and MLR models. RPD value for TSS decreased from 2.26 for PLS to 2.13 for MLR and, regarding TP, RPD showed a decrement from 1.98 to 1.79.

Table 1. Statistics of the MLR based on the three selected wavelengths (670, 730, 780 nm) and of the PLS models to predict the ripening parameters of the Nebbiolo grape (Giovenzana et al., 2014).

Ripening Parameters	Calibration MLR		Cross-validation MLR			Cross-validation PLS		
	R^2_{cal}	RMSEC	R^2_{cv}	RMSECV	RPD	R^2_{cv}	RMSECV	RPD
TSS (°Brix)	0.75	0.77	0.71	0.83	2.13	0.77	0.78	2.26
TP (OD_{280} nm)	0.74	3.95	0.70	4.30	1.79	0.74	3.88	1.98

TSS=total soluble solid content, TP=total polyphenols, OD=optical density, RPD=Ratio Performance Deviation

Results showed that only a small loss of information was noticeable between the PLS model calculated using 2048 wavelengths and the MLR models employing only three effective

Acta Hortic. 1311. ISHS 2021. DOI 10.17660/ActaHortic.2021.1311.16
Proc. VI Int. Symp. on Applications of Modelling as an Innovative Technology in the Horticultural Supply Chain - Model-IT 2019
Eds.: M.L. Amodio and G. Colelli

variables. The results demonstrate the feasibility of a simplified, low-cost handheld device for ripeness assessment in the field.

Results for blueberry ripeness evaluation

A PCA of blueberries spectra highlighted that two main wavebands (680 nm and 740 nm) can maximize the differences between fully ripe samples and those close to ripeness or unripe. Hence, spectral values at 680, 740, and 850 nm (the latter being an additional normalization waveband) were used to create a blueberry ripeness index (BRI) as a linear combination of two spectral ratios: S1=log(I680/I850) and S2= (I740/I850).

The definition of specific ripeness thresholds for the BRI according to different selective criteria was tested, and the ripeness classification capability was then assessed on a separate validation set of 471 berries. When applying a less selective threshold approach, the BRI correctly classified as ripe 85% of manually graded fully ripe berries, whereas 13% of close but not yet ripe validation samples were misclassified as fully ripe and ready to harvest. Comparatively, when a more demanding ripeness threshold was applied, the amount of nearly ripe berries misclassified as ripe decreased to 8%, but, the amount of fully ripe berries not identified as ripe rose to 25%. In both cases, none of the unripe samples was erroneously classified as a ripe fruit.

Also in this case the results, which were obtained with a BRI defined by spectral measurements at just three discrete wavelengths, point to the feasibility of a simple, microcontroller-based, handheld optical device able to implement the BRI to quickly assess the ripeness of sets of berries during the last and most delicate stages of the ripening process.

Results for postharvest control of cultivated mushrooms

PLS regression models were built from the spectra of cap and stipe of *Agaricus bisporus*. Table 2 shows the performance parameters of the PLS regression for the prediction of firmness, WC and SSC.

Acta Hortic. 1311. ISHS 2021. DOI 10.17660/ActaHortic.2021.1311.16
Proc. VI Int. Symp. on Applications of Modelling as an Innovative Technology in the Horticultural Supply Chain - Model-IT 2019
Eds.: M.L. Amodio and G. Colelli

Table 2. Statistics of PLS models elaborated on Vis/NIR spectra to estimate the firmness, SSC and WC of the *Agaricus bisporus* mushrooms (Giovenzana et al., 2019).

	Parameter	LV	Calibration		Validation			
			R^2	RMSEC	R^2	RMSEV	RMSE %	RPD
Cap	Firmness	8	0.83	6.36	0.78	8.81	17.28	2.03
	SSC	3	0.27	1.12	0.30	1.47	22.14	1.58
	WC	2	0.71	1.63	0.55	3.12	3.49	1.46
Stipe	Firmness	11	0.79	7.15	0.55	11.10	18.95	1.52
	SSC	6	0.66	0.92	0.65	1.32	21.25	1.70
	WC	9	0.84	1.55	0.78	2.14	2.38	2.15

LV=Latent Variables, RPD=Ratio Performance Deviation

In particular for the cap, interesting results were obtained for the prediction of the firmness with a R^2 of 0.78 and RPD of 2.03. The prediction of SSC and WC showed results below expectations with R^2 0.33 and 0.55 respectively. Probably, the colour variation during the accelerated shelf-life period increased the variability of the sample reducing the prediction capacity of the WC and SSC using a Vis/NIR device, compared to prediction performance that can be found in literature for these parameters using NIR spectroscopy.

Regarding the prediction of physico-chemical parameters of the stipe the results were instead different from those obtained from the cap. The best result was obtained for the prediction of WC (R^2 0.78 and RPD 2.15) while for firmness and SSC poor results were achieved. In this case the reduced colour changes of the stipe compared to those of the cap showed less influence on the possibility to calibrate an acceptable model for WC, also using the Vis/NIR range.

Results for senescence monitoring of fresh-cut baby leaf salad

A PLS-RCA technique was applied on *Valerianella locusta* L. spectra. The standardized regression coefficients of PLS models were used to select the relevant variables, representing the most useful information of full spectral region. In this case study, the four selected wavelengths were 520 nm, 680 nm, 710 nm and 720 nm. Statistics of the MLR models for pH, moisture and TP prediction are reported in Table 3. A comparison between PLS derived from

Acta Hortic. 1311. ISHS 2021. DOI 10.17660/ActaHortic.2021.1311.16
Proc. VI Int. Symp. on Applications of Modelling as an Innovative Technology in the Horticultural Supply Chain - Model-IT 2019
Eds.: M.L. Amodio and G. Colelli

the full Vis/NIR spectra (400-1000 nm) and MLR arising only from the four wavelengths showed that the overall calibration and prediction results of the MLR models, for all the parameters, were satisfactory, although the performance of the MLR models was slightly worse compared to PLS models, as expected. RPD value for pH decreased from 2.54 for PLS to 1.83 for MLR, regarding the MC, RPD showed a slight decrement from 2.25 to 2.08, and for TP from 3.19 to 2.48 for PLS and MLR respectively (Table 3).

Table 3. Statistics of the MLR models, based on the four selected wavelengths (520, 680, 710, 720 nm) to predict the freshness level of *Valerianella locusta* L. leaf samples, and of the PLS models (cross-validation) (Beghi et al., 2014).

Quality parameters	Calibration MLR			Cross-validation MLR			Cross-validation PLS		
	R^2	RMSEC	RPD	R^2	RMSECV	RPD	R^2	RMSECV	RPD
pH	0.82	0.13	2.54	0.70	0.18	1.83	0.86	0.13	2.54
TP (mg/100g)	0.88	12.38	3.26	0.80	16.28	2.48	0.89	12.64	3.19
MC (%)	0.87	0.09	3	0.75	0.13	2.08	0.84	0.12	2.25

RPD=Ratio Performance Deviation

CONCLUSIONS

Different feasibility studies were performed on several pre- and post-harvest matrices. Correlations between the optical data and reference parameters were carried out using PLS regression for spectra and using MLR for data from the effective wavelengths selected for the different applications.

The overall calibration and prediction results of the MLR models were satisfactory, although the performance of the MLR models was generally slightly worse than the PLS models. A small loss of information using only the selected wavelengths is noticeable. Consequently, the applicability of LED lightning system and inexpensive hardware could be envisaged to build simplified optical systems.

It is desirable the design of prototypes of pocket-size, inexpensive and easy to use optical devices for ripeness analysis, paying attention to versatility and modularity. The possibility to adjust light sources with a specific choice of wavelengths for LEDs, makes it possible a future use of the same simplified optical device for many different applications (i.e. ripeness

Acta Hortic. 1311. ISHS 2021. DOI 10.17660/ActaHortic.2021.1311.16
Proc. VI Int. Symp. on Applications of Modelling as an Innovative Technology in the Horticultural Supply Chain - Model-IT 2019
Eds.: M.L. Amodio and G. Colelli

evaluation, chemicals and physical properties prediction or shelf life analysis) and for different food matrix.

Literature cited

Beghi, R., Buratti, S., Giovenzana, V., Benedetti, S., & Guidetti, R. (2017). Electronic nose and visible-near infrared spectroscopy in fruit and vegetable monitoring. Reviews in Analytical Chemistry, 36(4) https://doi.org/10.1515/revac-2016-0016.

Beghi, R., Giovenzana, V., Civelli, R., Malegori, C., Buratti, S., and Guidetti, R. (2014). Setting-up of a simplified handheld optical device for decay detection in fresh-cut Valerianella locusta L. J. Food Eng. *127*, 10-15 https://doi.org/10.1016/j.jfoodeng.2013.11.019.

Beghi, R., Giovenzana, V., Spinardi, A., Bodria, L., Guidetti, R., and Oberti, R. (2013). Derivation of a blueberry ripeness index with a view to a low-cost, handheld optical sensing device for supporting harvest decisions. T. ASABE *56*(*4*), 1551-1559 http://dx.doi.org/10.13031/trans.56.10169.

Beghi, R., Spinardi, A., Bodria, L., Mignani, I., and Guidetti, R. (2013). Apples nutraceutic properties evaluation through a visible and near-infrared portable system. Food Bioprocess Technol. *6*(*9*), 2547-2554 https://doi.org/10.1007/s11947-012-0824-7.

Chong, I.G., and Jun, C.H. (2005). Performance of some variable selection methods when multicollinearity is present. Chemom. Intell. Lab. Syst. *78*, 103-112 https://doi.org/10.1016/j.chemolab.2004.12.011.

Civelli, R., Giovenzana, V., Beghi, R., Naldi, E., Guidetti, R., and Oberti, R. (2015). A Simplified, Light Emitting Diode (LED) Based, Modular System to be Used for the Rapid Evaluation of Fruit and Vegetable Quality: Development and Validation on Dye Solutions. Sensors *15*(*9*), 22705-22723 https://doi.org/10.3390/s150922705.

Cogdill, R.P., and Anderson, C.A. (2005). Efficient spectroscopic calibration using net analyte signal and pure component projection methods. J. Near Infrared Spectrosc. *13*(*3*), 119-132 https://doi.org/10.1255/jnirs.464.

Costa, G., Bonora, E., Fiori, G., and Noferini, M. (2011). Innovative non-destructive device for fruit quality assessment. Acta Hortic. *913*, 575-581 https://doi.org/10.17660/ActaHortic.2011.913.78.

Acta Hortic. 1311. ISHS 2021. DOI 10.17660/ActaHortic.2021.1311.16
Proc. VI Int. Symp. on Applications of Modelling as an Innovative Technology in the Horticultural Supply Chain - Model-IT 2019
Eds.: M.L. Amodio and G. Colelli

Cozzolino, D., Dambergs, R.G., Janik, L., Cynkar, W.U., and Gishen, M. (2006). Analysis of grapes and wine by near infrared spectroscopy. J. Near Infrared Spectrosc. *14*, 279-289 https://doi.org/10.1255/jnirs.679.

Giovenzana, V., Tugnolo, A., Casson, A., Guidetti, R., and Beghi, R. (2019). Application of visible-near infrared spectroscopy to evaluate the quality of button mushrooms. J. Near Infrared Spectrosc. *27*(*1*), 38-45 https://doi.org/10.1177/0967033518811921.

Giovenzana, V., Beghi, R., Malegori, C., Civelli, R., and Guidetti, R. (2014). Wavelength selection with a view to a simplified handheld optical system to estimate grape ripeness. Am. J. Enol. Vitic. *65*(*1*), 117-123 https://doi.org/10.5344/ajev.2013.13024.

Giovenzana, V., Civelli, R., Beghi, R., Oberti, R., and Guidetti, R. (2015). Testing of a simplified LED based vis/NIR system for rapid ripeness evaluation of white grape (*Vitis vinifera* L.) for Franciacorta wine. Talanta *144*, 584-591 https://doi.org/10.1016/j.talanta.2015.06.055.

Glories, Y. (1984). The colour of red wines. Connaissance de la Vigne et Du Vin *18*, 195-217.

Guidetti, R., Beghi, R., and Giovenzana, V. (2012). Chemometrics in Food Technology. In Chemometrics in Practical Applications, InTech, Kurt Varmuza eds. (Rijeka, Croatia), p.217-252.

Guidetti, R., Beghi, R., and Bodria, L. (2010). Evaluation of Grape Quality Parameters by a Simple Vis/NIR System. T. ASABE *53*(*2*), 477-484 https://doi.org/10.13031/2013.29556.

Guidetti, R., Beghi, R., Bodria, L., Spinardi, A., Mignani, I., and Folini, L. (2008). Prediction of blueberry (Vaccinium corymbosum) ripeness by a portable Vis-NIR device. Acta Hortic. *310*, 877-885 https://doi.org/10.17660/ActaHortic.2009.810.117.

Larrain, M., Guesalaga, A.R., Agosin, E. (2008). A Multipurpose Portable Instrument for Determining Ripeness in Wine Grapes Using NIR Spectroscopy. Instrumentation and Measurement, IEEE Transactions *57*(*2*), 294-302 https://doi.org/10.1109/TIM.2007.910098.

López, J., Uribe, E., Vega-Gálvez, A., Miranda, M., Vergara, J., Gonzalez, E., and Di Scala, K. (2010). Effect of Air Temperature on Drying Kinetics, Vitamin C, Antioxidant Activity, Total Phenolic Content, Non-enzymatic Browning and Firmness of Blueberries Variety O´Neil. Food Bioprocess Technol. *3*(*5*), 772-777 https://doi.org/10.1007/s11947-009-0306-8.

Magwaza, L.S., Opara, U.L., Nieuwoudt, H., Cronje, P.J., Saeys, W., and Nicolaï, B. (2012). NIR spectroscopy applications for internal and external quality analysis of citrus fruit - a review. Food Bioprocess Technol. *5*(*2*), 425-444 https://doi.org/10.1007/s11947-011-0697-1.

Acta Hortic. 1311. ISHS 2021. DOI 10.17660/ActaHortic.2021.1311.16
Proc. VI Int. Symp. on Applications of Modelling as an Innovative Technology in the Horticultural Supply Chain - Model-IT 2019
Eds.: M.L. Amodio and G. Colelli

Nicolai, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I., and Lammertyn, J. (2007). Non-destructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. Postharvest Biol. Technol. *46*, 99-118 https://doi.org/10.1016/j.postharvbio.2007.06.024.

Puangsombut, A., Pathaveerat, S., Terdwongworakul, A., and Puangsombut, K. (2012). Evaluation of internal quality of fresh-cut pomelo using vis/NIR trasmittance. J. Texture Stud. *43*, 445-452 https://doi.org/10.1111/j.1745-4603.2012.00354.

Slaughter, D. C. (2009). In Nondestructive maturity assessment methods for mango: A review of literature and identification of future research needs. Orlando, FL, USA: National Mango Board.

Xiaobo, Z., Jiewen, Z., Povey, M.J.W., Holmes, M., and Hanpin, M. (2010). Variables selection methods in near-infrared spectroscopy. Anal. Chim. Acta *667*, 14-32 https://doi.org/10.1016/j.aca.2010.03.048.

Acta Hortic. 1311. ISHS 2021. DOI 10.17660/ActaHortic.2021.1311.16
Proc. VI Int. Symp. on Applications of Modelling as an Innovative Technology in the Horticultural Supply Chain - Model-IT 2019
Eds.: M.L. Amodio and G. Colelli

2.4 Visible near-infrared spectroscopy as a green technology

Nowadays the reduction of solvents used and of energy consumption became a crucial aspect to be investigated. The non-destructive optical methods based on optical sensing (e.g. vis/NIR spectroscopy) represent a simple, rapid, and easy-to-use method to predict quality parameters and hence for objectively evaluating indices of different fruit and vegetable products (Nicolaï et al., 2007; Beghi et al., 2017). Once calibrated, a vis/NIR device can analyse samples in a non-destructive way and in few seconds, without sample processing and without expert personnel able to use complex laboratory instrumentations. Compared to vis/NIR technology, chemical techniques result time-consuming, require samples preparation and the use of chemical reagents, resulting also expensive due to the need of expert laboratory technicians (Table 2.1) (Guidetti et al., 2010).

Table 2.1. Comparison between optical sensing and chemical analysis highlighting the peculiar aspects.

	Optical sensing	**Chemical analysis**
Rapidity	x	-
More parameters analysed at the same time	x	-
Non-destructive analysis	x	-
Direct measurement	-	x
Accuracy	High number of samples required	x
Use of chemicals	-	x

For instance, the replacement of the analytical tools and reagents related to chemical analyses with one vis/NIR spectrometer could reduce the environmental impact of analyses. Thus, in order to define vis/NIR spectroscopy as green technology, it is necessary to focus the attention on sustainability aspects. The environmental impact of vis/NIR spectroscopical systems can be assessed using Life Cycle Assessment (LCA) (Marquez et al., 2005). LCA is a tool to analyse the energy loads and environmental impacts associated with the different phases of the entire life cycle of a product or service: from cradle to grave (from the extraction of raw materials to the stages of transformation, production, distribution, use and finally disposal or recycling). An LCA can be an excellent support tool for sustainable design, and its drafting is defined by the ISO 14040 and ISO 14044 standards (ISO, 2021).

Thanks to an accurate LCA, it will be possible to identify for a given product or service: the critical phases from an environmental point of view, the subjects who can mediate to change the situation and the data necessary to be able to carry out adequate environmental improvement interventions.

In this paper, the contribution of sustainability experts was needed to quantify for the first time the sustainable advantages of the proximal sensing techniques (in the olive oil production in this case). Here, the PhD candidate dealt with the LCA experts to provide data and methodologies in order to build a flow sheet of input and output quantifying the environmental impacts using a from cradle to grave approach.

PAPER 2: Environmental advantages of visible and near infrared spectroscopy for the prediction of intact olive ripeness

Andrea Casson, Roberto Beghi, Valentina Giovenzana*, Ilaria Fiorindo, Alessio Tugnolo, Riccardo Guidetti

Department of Agricultural and Environmental Sciences - Production, Landscape, Agroenergy, Università degli Studi di Milano, via Celoria 2, 20133, Milano, Italy

*Corresponding Author: valentina.giovenzana@unimi.it

Article history: Received 25 July 2019; Received in revised form 24 September 2019; Accepted 1 November 2019; Published online 22 November 2019

Abstract

Conventional ripeness analyses performed on olives require different analytical tools, chemicals, sample preparation and they are time consuming. The same analyses performed using an optical and non-destructive technology as visible and near infrared (vis/NIR) spectroscopy allow to predict ripeness parameters in a simple and quickly way. The purpose of this work is to compare the environmental impact of conventional ripeness analyses and optical one performed on olives fruits. The conventional analyses identified as reference were: moisture, oil and phenol content. The Life Cycle Assessment was applied to assess the environmental impact. The approach “from cradle to grave” considered all the inputs and outputs of each analysis, enclosing machineries, reagents and energy necessary for analyses. Furthermore, for the optical analysis were also considered the activities required for the instrument calibration. Quantifying the environmental damage, the results showed clear advantages for optical analysis allowing to define vis/NIR spectroscopy as green technology.

Keywords: green technology; optical analysis; sustainability; agro-food sector; LCA; efficiency

INTRODUCTION

Olive oil yield and quality are influenced by different field factors defined as “primary” or “uncontrollable”. These include genetic (olive variety) and environmental aspects (soil characteristics, pest attacks and climatic condition which may vary every year) (Criado et al., 2004; Beltrán et al., 2005). Furthermore “secondary” or “controllable” factors need to be considered, as the agronomic practices which can damage the fruit physiology (irrigation, fertilization and harvesting) and technological procedures which could alter the olive oil composition (processing and storage activities) (Baccouri et al., 2008). The secondary activities can alter the olive quality, as change in weight, pulp ratio, chemical composition and oil accumulation (Jemai et al., 2009; Desouky et al., 2009). If the secondary factors can be controlled and optimized by critical control point analyses or by best practices controls, the primary factors cannot be easily modified. In this scenario, the activities aimed to monitor the olives ripening to decide the best harvesting period is a critical control and fundamental step in extra-virgin olive oil production chain (Nasini & Proietti, 2014).

Optimizing olive harvest means obtaining the highest amount of oil preserving the quality. Even if the main requirement in the olive oil sector could be the yield, in most of cases, the harvesting decision is determined by the olive oil quality. Therefore, based on suitable sensory and analytical requirements, the olive yield is overshadowed by the quality (Nasini & Proietti, 2014). To satisfy quality requirements, laboratory analyses on olives are needed to objectify the quality parameters. Most of quality analyses is standardized and optimized by Regulations (EEC/2568/91) and Association of Official Agricultural Chemists (A.O.A.C.) methods. Even if the procedures are standardized, they represent criticisms because of they are time consuming (from sampling to results, the time could vary from hours to days) and energy consuming (long execution times and type of analytical tools). Moreover, the conventional methods result dangerous due to the reagents use (A.O.A.C., 1975; European Communities, 2003; Fernández -Espinosa, 2016). Finally, most of these methods are expensive, labour-intensive, and require previous treatment of samples.

As a solution to these technical problems the visible/near-infrared (vis/NIR) spectroscopy could be identified as a suitable analytical method. It doesn’t require any treatment of the sample, it is a fast and easy analysis and it can be automatized. The vis/NIR technology is widely used in the agro-food sector and for rapid quality fruit and vegetable monitoring (Beghi et al., 2017; Magwaza et al., 2012; Nicolai et al., 2007). Agro-food production has a large impact on

environment (Notarnicola et al., 2019; Roy et al., 2019; Guarino et al., 2019; Recchia et al., 2019) and current food processes set-up are not optimised (Van der Goot et al., 2016), therefore more sustainable food products are desirable.

The vis/NIR technology represents also a rapid technique to characterise physical and chemical parameters, providing real-time results directly in field to help operators to take decision (Fernández -Espinosa, 2016). This technology is widely used in the olive oil sector to estimate different quality parameters on intact olives (Cayuela et al., 2010; Bellincontro et al., 2012; Salguero-Chaparro et al., 2012; Giovenzana et al., 2015).

Since the use of spectroscopy generally gives predictions comparable with the conventional results (Giovenzana et al., 2018; Fernández-Espinosa, 2016) and does not require sample preparation and reagents use, it could reach the qualification of green-technology. In literature, few studies based on environmental approach define and quantify the environmental damage due to optical analyses (Casson et al, 2019).

This study aims to quantify and compare the environmental impact related to the olive ripeness analyses. The Life Cycle Assessment (LCA) was applied to evaluate and compare the conventional method to the optical one. This kind of environmental evaluation defines the impact related to a service and not to a product. The results of this study could be useful to enforce the green aspect associated to non-destructive analyses.

MATERIALS AND METHODS

Life Cycle Assessment (ISO 14040, 2006) is a standardized method used to evaluate the environmental impacts throughout the study of the life cycle of a product or a service. It accounts all the input and output factors from the raw material extraction, through the production, transportation and use processes, till the final stage of the product life cycle. The following environmental analysis is developed in compliance with the international standards of series 14000 which divide the LCA method into four phases as follow (ISO 14040-14044, 2006).

Goal and scope definition

The goal of this study is to quantify the environmental profile of two methods, conventional and optical, to monitor the olives ripeness, considering the measurement of three main parameters capable to characterize olive samples (reference parameters). According to the olive mill

requirements, the two main parameters analysed are moisture and oil content. In fact, the measurement of these parameters allows to determine the payment amount to the growers and the olives destinations (acceptability or rejection for milling) (Salguero-Chapparo et al., 2013). The third one parameter considered, is the phenols content which is important to quantify the fruit ripeness and it is directly linked to sensory characteristics (Rodrigues et al., 2019).

The functional unit, especially in a comparative LCA, as far as possible should be related to a function of the product rather than to the product itself (ISO 14040, 2006). In this study the functional unit was identified with the analysis service provided by the laboratory to obtain the three parameters analysed, i.e. three chemical analyses for the conventional method and one non-destructive analysis (to provide the estimation of the three reference parameters) for the optical method.

An approach called “from cradle to grave” was used to study the system considering all the input from the extraction of raw materials, through the construction of the tools, electronic devices and laboratory materials, considering the energy and water supply and all the outputs of the analyses from dangerous material to wastewater, paying also attention to end of life of every inputs previously quantified and inserted in the analysis.

The system under study considers the services provided by four conventional laboratories located in different Italian provinces (Bari, Reggio Calabria, Sassari and Teramo) which are specialized on olives ripeness analyses and serve on average 90 olive growers each. A study performed on olive should consider a reference period of a year (Casson et al., 2019), while the ripening analyses can be performed only for three months (in Italian provinces analysed, from October to December). Considering that all the 90 olive growers request to perform 5 samplings during olive ripening, the laboratory capacity should be defined equal to 450 analyses year-1.

For the conventional analyses, the studied laboratories perform the analyses during the three months of ripening period, the same tools are not used for other tasks and in another period. To measure the reference parameters, as alternative to conventional method, spectroscopy was chosen (Giovenzana et al., 2018; Trapani, et al., 2016; Bellincontro et al., 2012).

Life Cycle inventory (LCI)

The inventory phase is based on to the Regulation EEC/2568/91 and on the AOAC methods. Even if the procedures are standardized and the quantity of solvents are precisely defined, data referred to the procedures of each analysis were obtained with surveys sent to the laboratories in order to identify the differences between the standard procedures and the real execution parameters of the analyses (times, number of samples used, type and obsolete degree of tools).

The survey helped (i) to identify the different analytical tools used, (ii) to figure out the machineries composition quantifying the main components as plastic, metal, glass and electronic components (46% steel, 32% plastic, 14% printed wiring board and 8 % of cable of various type) (Table 1), and (iii) to quantify the lifetime of each machinery and of laboratory materials used for the analyses, defined in years, ranged from 1 to 15, or in number of analyses, ranged from 1 to 6750 based on the nature of the tool used. Table 1 summarizes information obtained from survey elaboration relating to laboratory analytical tools. Starting from average capacity of the laboratory to carried out conventional and optical analyses identified as 450 analyses year-1 (concentrated in three months) and considering the laboratory works equal to 25 days month-1 (75 days in three months), the laboratory was sized considering a number of analyses equal to 6 samples day-1.

Considering the aim of the study, no attention was paid to equal activities performed in both the conventional and the optical analyses (farming, harvesting and transporting).

The inventory of every analysis was performed separately and reported below.

Table 1. Material composition, weight, electric power and number of analytical tools in the laboratory

Analytical tool	Composition	Total weight (kg)	Electric power (kW)	Number of machineries
Balance	60% Electronic components 40% Stainless steel	5	0.13	1
Centrifuge	83% Electronic components 17% Stainless steel	30	0.75	1
Computer	100% Electronic components	13	0.13	1
Vis/NIR device	93% Electronic components 7% Glass fibre	15	0.24	1
Olive mill	87.5% Steel 12.5% Electronic components	8	1.00	1
Oven	91% Steel 9% Electronic components	54	1.00	1
Spectrophotometer	90% Electronic components 100% Steel	10	0.13	1
Suction hood	36% Glass 36% Steel 28% Electronic components	55	0.44	3
Water heater	100% Stainless steel	10	1.00	5

Moisture determination LCI

The olives are crushed with a portable and miniaturized mill (crushing time 2 min) to obtain olive paste, the sample is then weighed, inserted in the oven to be dried until reaching constant weight and then weighed again (A.O.A.C., 1975). In this analysis the oven used could vary the time necessary to perform the analysis. Oven capacity allows to dry more than one sample simultaneously. Moreover, oven use is also required in the oil content analysis. Therefore, considering the oven samples capacity and the analysis typology, an allocation procedure must be performed to allocate the electricity consumption and the material composition of the oven to the functional unit.

For each input (Table 2), an allocation factor ranged from 0 to 1 identifies the allocation percentage considering the time of usage and the multiple use of the input for the different reference parameters. Moreover, as performed in a previous study (Casson et al., 2019) in order to calculate the amount per analysis (Apa) for each input and output, equation 1 was calculated:

$$Apa=(Q*Af)/Noa \quad (1)$$

Where:

Q: Quantity

Af: Allocation factor

Noa: Number of analyses

Table 2. Quantity, allocation factors, lifetime and amount per analyses for moisture determination analysis

Input	Quantity (Q)	Allocation factor (Af)	Lifetime		Amount per analysis (Apa)	Unit
			Years	Number of analyses (Noa)		
Olive mill	8	0.33	15	6750	0.00039	kg
Balance	5	0.59	10	4500	0.00066	kg
Oven	54	0.94	15	6750	0.00752	kg
Dryer	3	0.50	10	4500	0.00033	kg
Pottery dish	18	1.00	10	4500	0.004	kg
Dryer silica	50	0.50	/	1000	0.025	g
Refractory material	15	1.00	/	1000	0.015	g
Electricity	3.8	1.00	/	1	3.8	kWh
Output						
Inert waste	500	/	/	/	500	g

Oil content analysis LCI

The olives are crushed using portable and miniaturized mill (crushing time 2 min) to obtain olive paste and then dried in the oven for two hours. The obtained dried olive paste passed through different steps to extract the oil (EEC/2568/91). The extraction method implies the use of different Soxhlet extractors and five water heaters equipped by three units each working simultaneously. This method requires long execution time (6 hours) and high quantity of diethyl ether (750 ml) per analysis. As for the water heaters also for the glass extractor the method requires twelve extractors to satisfy the 6 samples per day capacity, considering two replicates per sample.

The amount of diethyl ether necessary for one analysis is 750 ml but in most of the laboratories under study this reagent is majorly recovered (around 95%) through distillation procedure. Moreover, the water demanded by this analysis is around 600 litres. This large water amount is required to condense the diethyl ether which is subsequently used to extract the oil from the pomace. Input and output data related to this analysis are reported in Table 3.

Table 3. Quantity, allocation factors, lifetime and amount per analyses for oil content analysis

Input	Quantity (Q)	Allocation factor (Af)	Lifetime		Amount per analysis (Apa)	Unit
			Years	Number of analyses (Noa)		
Olive mill	8	0.33	15	6750	0.00039	kg
Balance	5	0.21	10	4500	0.00023	kg
Oven	54	0.06	15	6750	0.00048	kg
Suction hood	110	0.75	15	6750	0.01222	kg
Water heater	50	1.00	15	6750	0.00740	kg
Textile woven	0.9	1.00	/	1	0.9	g
Glass	90	1.00	10	4500	0.02	kg
Carbon filters	10	0.75	/	2000	0.00375	kg
Dryer silica	3	0.50	10	4500	0.00033	kg
Dryer silica	50	0.50	10	1000	0.025	g
Diethyl ether	750	1.00	/	20	37.5	g
Tap water	600	1.00	/	1	600	kg
Electricity	6.023	1.00	/	1	6.023	kWh
Output						
Inert waste	0.9	/	/	/	0.9	g
Hazardous waste	37.5	/	/	/	37.5	g
Wastewater	600	/	/	/	600	l

Phenols determination LCI

The olive samples, after crushing using the portable and miniaturized mill, are added of a quantity of hexane and then undergone into an extraction phase of the fat fraction content using the centrifuge. The extracted fraction is than filtered using a mono-use syringe and nylon filter. Folin-reactive and sodium carbonate are added to the filtered solution and conditioned. The solution is analysed using the spectrophotometer at a defined wavelength to calculate the phenols content (A.O.A.C., 1975). This analysis is less time-consuming respect to the others (1.5 hour). However, it requires different chemicals and analytical tools to be performed. Input and output data related to this analysis are reported in Table 4.

Table 4. Quantity, allocation factors, lifetime and amount per analyses for phenols determination analysis

Input	Quantity (Q)	Allocation factor (Af)	Lifetime		Amount per analysis (Apa)	Unit
			Years	Number of analyses (Noa)		
Olive mill	8	0.33	15	6750	0.00039	kg
Centrifuge	30	1.00	10	4500	0.00667	kg
Balance	5	0.20	10	4500	0.00022	kg
Spectrophotometer	10	1.00	10	4500	0.00222	kg
Computer	13	1.00	15	6750	0.00193	kg
Suction hood	55	0.50	15	6750	0.00407	kg
Packaging glass	12	1.00	1	1	12	g
Carbon filter	600	1.00	10	4500	0.13333	kg
Nylon filters	12	1.00	/	1	12	g
PVC cuvette	5	0.50	/	2000	0.00125	g
Hexane	3	1.00	/	1	3	ml
Methanol	12	1.00	/	1	12	ml
Folin-reactive	0.375	1.00	/	1	0.375	ml
Sodium carbonate	3.75	1.00	/	1	3.75	mg
Electricity	1.04	1.00	/	1	1.04	kWh
Output						
Solvent mixture	20	/	/	/	20	g
Plastic	12	/	/	/	12	g

Vis/NIR spectroscopy LCI

The intact olive samples are analysed using light emission deriving from vis/NIR spectrophotometer, without any sample preparation. The spectra obtained from the light reflection by olive is then analysed using software for model calibration installed on a computer. Multiple information can be deducted from the same spectral dataset. This technology implies the use of only two analytical tools: the vis/NIR device and the computer. No chemicals and

no expertise of laboratory procedure are necessary, the samples could be analysed directly in field, on intact olives, reducing any waste product.

Also, in this case an allocation procedure based on the usage time of the tools was considered to calculate the environmental impact of vis/NIR spectroscopy analysis. Input data related to this analysis are reported in Table 5. No outputs were produced.

Table 5. Quantity, allocation factors, lifetime and amount per analyses for vis/NIR spectroscopy analysis

Input	Quantity (Q)	Allocation factor (Af)	Lifetime		Amount per analysis (Apa)	Unit
			Years	Number of analyses (Noa)		
Vis/NIR	15	1.00	15	6750	0.00222	kg
Computer	13	1.00	15	6750	0.00192	kg
Calibration	2100	0.015	15	6750	0.31	*n*
Electricity	0.12	1.00	/	1	0.12	kWh

The calibration of the vis/NIR device is a fundamental phase to obtain reliable results. The vis/NIR spectroscopy analysis gives not a measure of the parameters but, basing on a model built with the calibration, gives an estimation of them. The calibration is intended as the activity which requests to perform 700 conventional analyses, per parameter, correlating with the respective 700 optical analyses, performed on the same samples. Most of the 700 calibration analyses (500) were performed at the beginning of the implementation of the optical method (initial instrument training), while the others (200) were used as validation set. In this scenario, three models for moisture, oil and phenols content, were allocated to the 6750 analyses performed in 15 years.

Life Cycle Impact Assessment (LCIA)

According to Wolf et al. (2012), ILCD 2011 (International Life Cycle Data System) midpoint method was used to provide indicators to compare environmental impact in a more detailed level (16 categories) respect to the endpoint level (4 aggregated categories). Table 6 reports the impact categories used to evaluate the environmental impact related to the two methods to measure the three reference parameters to characterise olive samples, the conventional laboratory analyses and the optical system. The study was conducted according to the

reference standards for LCA ISO 14040-14044:2006 (ISO 14040, 2006), using the software of analysis SimaPro (version 9) (PRé Sustainability, Amersfoort, The Netherlands).

Table 6. Impact categories, acronyms and unit of ILCD 2011 midpoint method

Impact category	Acronyms	Unit
Climate change	CC	kg CO_2 eq
Ozone depletion	OD	kg CFC-11 eq
Human toxicity, non-cancer effects	HT-NC	CTUh
Human toxicity, cancer effects	HT-C	CTUh
Particulate matter	PM	kg PM2.5 eq
Ionizing radiation HH	IRHH	kBq U235 eq
Ionizing radiation E (interim)	IRE	CTUe
Photochemical ozone formation	POF	kg NMVOC eq
Acidification	ACID	molc H^+ eq
Terrestrial eutrophication	TEU	molc N eq
Freshwater eutrophication	FEU	kg P eq
Marine eutrophication	MEU	kg N eq
Freshwater ecotoxicity	FECO	CTUe
Land use	LU	kg C deficit
Water resource depletion	WRD	m^3 water eq
Mineral, fossil & ren resource depletion	RRD	kg Sb eq

RESULTS AND DISCUSSION

According to the purpose of the study, the two analyses methods were firstly analysed separately and then compared. While for the single study of the different analyses methods the hotspot identification criterion was used, for the comparison a numerical and quantitative rationale is needed in order to move from a single percentage consideration. The first method helps to identify the hotspots which are, in this paper, thus defined as activities or factors throughout the life cycle of the different analyses that are associated with higher environmental impacts.

Conventional methods

Figure 1 shows the environmental impact in percentage values, identifying the different hotspots obtained from the elaboration of the inputs and outputs determined for conventional analyses.

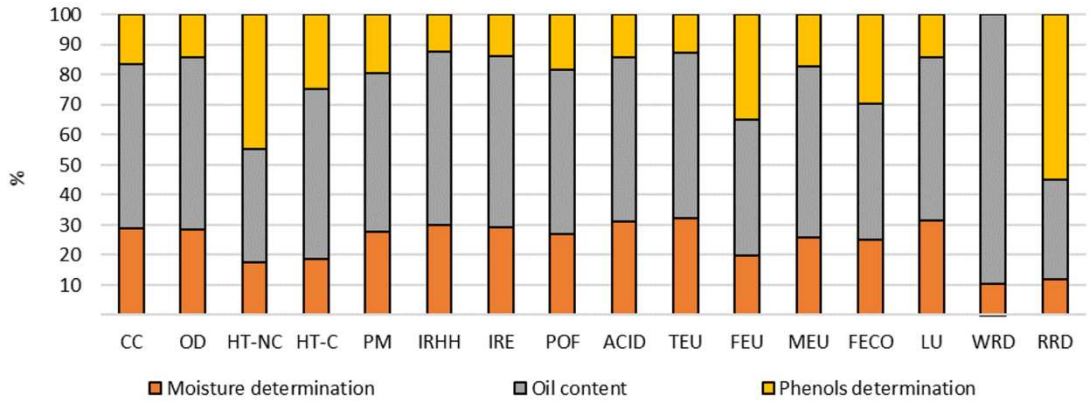


Figure 1. Hotspots for conventional analyses based on the reference parameters. Impact categories and acronyms are defined in table 6.

In Figure 1, the overall impact is divided in the three conventional analyses performed to obtain three reference parameters, allowing to quantify environmental damaging. The higher percentage of impact comes from the measurement of the oil content. The high demand of water necessary for the condensation of the hexane (600 L analysis-1) lets the oil content analysis reach 89% in WRD impact category. For the other impact categories, the responsibilities percentage of oil content analysis is balanced around 54%, apart for the impact category RRD (33%), in which the oil content analysis has lower level of responsibilities due to the limited number of analytical tools respect to the moisture and phenol determination analyses.

Figure 1 identify the most environmental impactful analysis but is not capable to highlight the motivation. Therefore, the three conventional analyses need to be considered as a single conventional analysis (Figure 2). To consider the set of three analyses as one allows to subdivide the inputs and outputs in 6 factors: analytical tools, energy, chemicals, waste treatment, laboratory materials and water and allows to identify the hotspots.

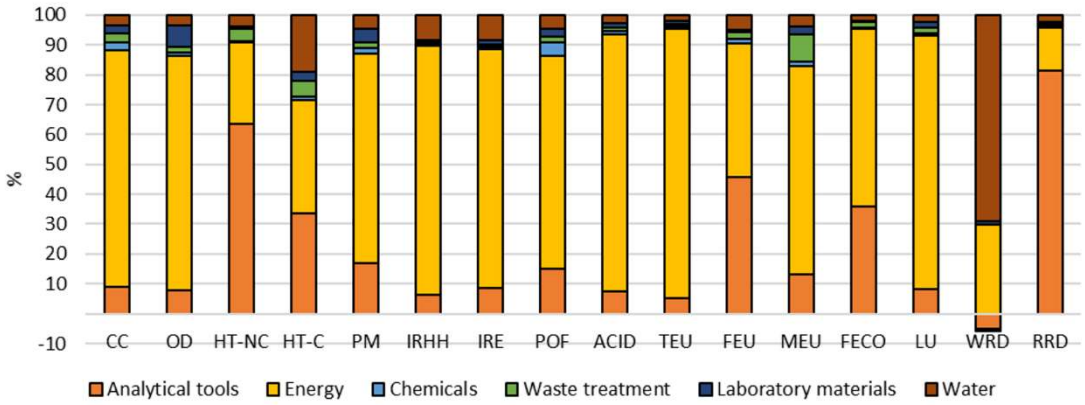


Figure 2. Hotspots for conventional analyses based on a factor subdivision. Impact categories and acronyms are defined in table 6.

For most of the impact categories reported in Figure 2, the higher contribution derives from the energy due to the demand coming from the analytical tools (Table 1). The energy power absorption and the usage time of each electronic tool let the environmental damage increase due to the energy contribution with a medium percentage of 76%. The impact categories related to the human health as for HT-NC and HT-C shows different behaviours regarding the energy hotspot, in HT-NC the major impact (63%) comes from the analytical tools. The same consideration can be reached regarding the RRD. The extraction of metals and polymers for the construction of the machineries justify also the 81% related to the analytical tools.

Even if the impacts of the analytical tools are positive in most of the impact categories, regarding the WRD impact category the construction of the machineries shows a negative percentage value (-5%) which, related to the environment, represent a positive aspect. This negative value finds justifications in the extraction companies behaviour which recycle or use recycled water to reduce water loss in this process which requires large amount of water.

As shown in Figure 1, WRD impact category was mainly allocated to the oil content analysis (89%). Matching this consideration whit results in Figure 2, the high level of impact can be justified by the water used in this analysis 69%.

The impact related to the chemicals showed lower values respect all the other factors (lower than 5% in all the impact categories).

Vis/NIR spectroscopy method

The vis/NIR method does not require the same inputs and outputs of the conventional one, the subdivision in inputs and outputs criteria showed the necessity to identify only 3 factors: analytical tools, calibration, and energy.

Figure 3 reports the environmental impact of the vis/NIR method. The LCA analysis not only considers the execution of the optical analysis, but also include all the background activities relating to the spectrophotometer calibration. Among the impact categories, the major hotspot is the calibration due to the large amount of conventional analyses to calibrate the predictive model to estimate reference parameters for olive characterization.

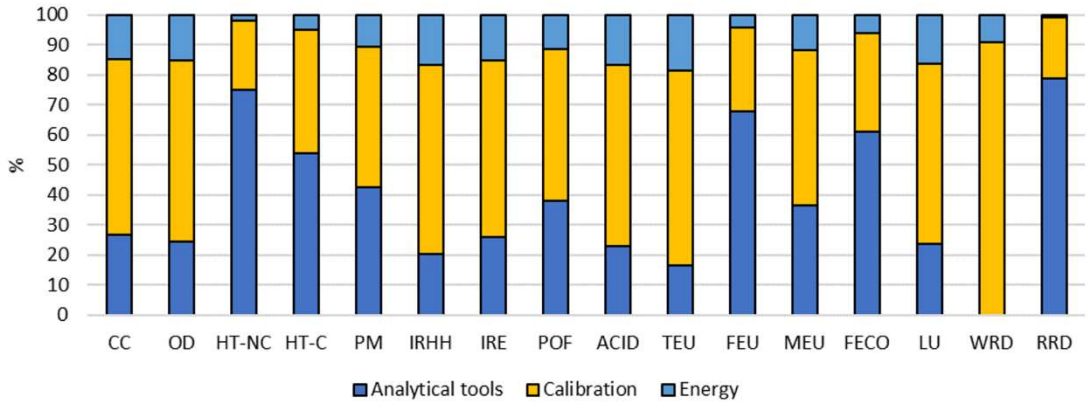


Figure 3. Hotspots for optical analysis based on a factor subdivision. Impact categories and acronyms are defined in table 6.

The second hotspot, especially for the impact categories directly correlated to the metal and to the electronics industries, for example HT-C, HT-NC and RRD, is the analytical tools factor. The minerals extraction and the electronic devices construction lead the impact in higher percentages of human health impact categories (HT-C 54% and HT-NC 74%) and non-renewable impact categories as for RRD 78%, respect to the other factors.

To compare the environmental impact of conventional and optical analyses, quantitative values are needed. According to the functional unit defined in section 2.1, the comparison results showed the impact related to the two method of analyses for each impact category (Table 7).

Table 7. Impact values related to the execution of the two methods (conventional and optical) to measure the three reference parameters to characterize olive.

Impact category	Unit	Conventional method	Optical method	Ratio level
Climate change	kg CO_2 eq	5.95	0.35214	16.88
Ozone depletion	kg CFC-11 eq	6.56×10^{-7}	3.76×10^{-8}	17.44
Human toxicity, non-cancer effects	CTUh	3.93×10^{-6}	5.95×10^{-7}	6.61
Human toxicity, cancer effects	CTUh	5.30×10^{-7}	4.47×10^{-8}	11.84
Particulate matter	kg PM2.5 eq	0.00323	0.00023	13.53
Ionizing radiation HH	kBq U235 eq	9.15×10^{-1}	0.05036	18.16
Ionizing radiation E (interim)	CTUe	2.39×10^{-6}	1.40×10^{-7}	17.08
Photochemical ozone formation	kg NMVOC eq	0.01521	0.00104	14.58
Acidification	molc H^+ eq	0.05733	0.00328	17.48
Terrestrial eutrophication	molc N eq	0.16818	0.00899	18.69
Freshwater eutrophication	kg P eq	0.003464	0.00042	8.17
Marine eutrophication	kg N eq	0.00630	0.00042	15.00
Freshwater ecotoxicity	CTUe	180.9074	19.14027	9.45
Land use	kg C deficit	12.7167	0.73403	17.32
Water resource depletion	m^3 water eq	0.13356	0.00367	36.37
Mineral, fossil & ren resource depletion	kg Sb eq	0.00065	0.00011	5.89

As showed in table 7, the conventional method implies higher impacts. They accounted for human health in HT-NC for 3,93x10-06 and HT-C for 5,30 x10-07 Comparative toxic unit (CTUh), while for the vis/NIR method the impact is respectively 5,95x10-07 and 4,47x10-08

CTUh. These two impact categories with FEU and RRD showed ratio values lower than 10, due to the analytical tools, as noticeable in Figures 2 and 3. Regarding all the other impact categories, the ratio level ranged 17-20 (WRD reaches 36 ratio level), this means that for most of the impact categories the vis/NIR method present lower environmental impact respect to the conventional method.

Considering the obtained results for the studied parameters, the optical, non-destructive and rapid method can be defined 15 times greener respect to the conventional methods.

This value cannot be assumed as final ratio value, in the conventional methods, the higher criticism is linked to the possibility to perform the same analysis in different ways, using different analytical tools. In order to evaluate the change of impact due to a change in procedure, the conventional moisture content analysis was replaced with an alternative method using the thermobalance and a sensitivity analysis was performed.

Sensitivity analysis

The thermobalance is an analytical tool which can substitute the conventional oven, an energy consuming analytical tool. The thermobalance can analyse few grams of sample per time but in less time and with lower energy respect to the method using the conventional oven. Considering the possibility to substitute the conventional moisture content analysis performed using oven with the thermobalance (weight 8 kg; electric power equal to 0.23 kW), the aim of the sensitivity analysis is to assess the variation in term of allocation for those analytical tools used in the three conventional analyses and evaluate the impact assessment that derived from changing one analytical tool.

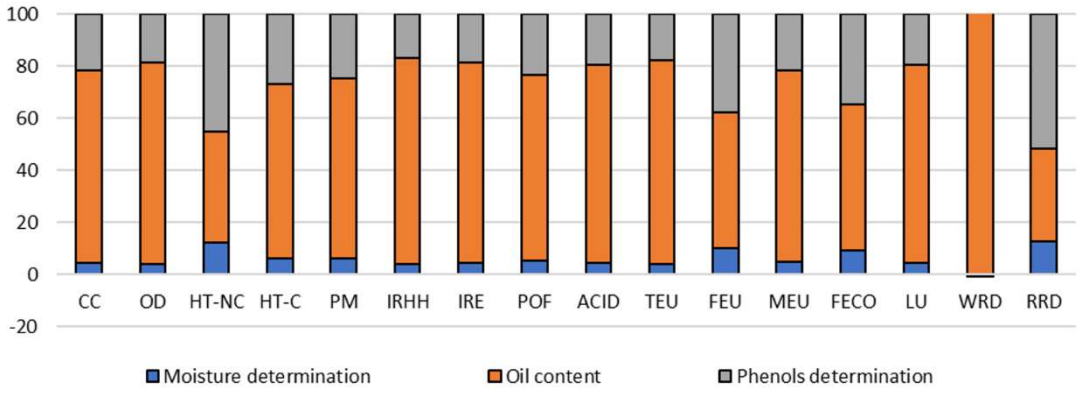


Figure 4. Hotspots for conventional analyses based on the reference parameters, replacing oven with thermobalance for moisture determination analysis. Impact categories and acronyms are defined in table 6.

As showed in Figure 4, considering the moisture determination analysis, significant reduction of the impact from the change in analytical tools can be seen in most of the impact categories (medium decrease of 24%). The decrease of the environmental impact due to introduction of thermobalance, cause on the contrary an increase of the percentage responsibilities related to oil and phenols content analyses. The complete allocation of the oven on the oil content analysis lets this method increase the percentage responsibility in all the impact category (+19% CC, +10% HT-C, +11% WRD).

Table 8 shows the impact assessment using oven or thermobalance for moisture determination analysis among conventional analyses. Despite the increase of impact related to the oil content and phenols determination analysis and the decrease in the moisture content, the overall impact of the conventional method decreases of about 18%. The only one impact category which shows a negative result is the RRD with an increase of impact equal to +6.2%. This increase, linked also to the impact categories HT-C and HT-NC which show lower environmental benefits, is due to the use of thermobalance which not replace the oven that results still necessary for the oil content analysis. Moreover, although the use of thermobalance imply an increase of environmental impact due to an additional analytical tool, it results in a decrease of energy consumption compare to the use of the oven only. Considering the new environmental impact due to the use of thermobalance instead of oven, the overall ratio gap between the conventional and the optical methods decrease from 15.28 to 12.24.

Table 8. Impact values related to the execution of the two methods (conventional using thermobalance instead oven for moisture determination analysis and optical) to measure the three reference parameters to characterize olive

Impact category	Unit	Conventional method using thermobalance	Reduction respect to oven	Optical method	New ratio
Climate change	kg CO_2 eq	4.47	-24.8%	0.3521	12.69
Ozone depletion	kg CFC-11 eq	4.94×10^{-7}	-24.7%	3.76×10^{-8}	13.13
Human toxicity, non-cancer effects	CTUh	3.87×10^{-6}	-1.6%	5.95×10^{-7}	6.50
Human toxicity, cancer effects	CTUh	4.83×10^{-7}	-8.9%	4.47×10^{-8}	10.78
Particulate matter	kg PM2.5 eq	0.0025	-21.5%	0.0002	10.62
Ionizing radiation HH	kBq U235 eq	0.6721	-26.5%	0.0503	13.35
Ionizing radiation E (interim)	CTUe	1.79×10^{-6}	-25.2%	1.40×10^{-7}	12.78
Photochemical ozone formation	kg NMVOC eq	0.0119	-21.5%	0.0010	11.44
Acidification	molc H^+ eq	0.0417	-27.3%	0.0033	12.71
Terrestrial eutrophication	molc N eq	0.1194	-28.9%	0.0089	13.28
Freshwater eutrophication	kg P eq	0.0032	-7.6%	0.0004	7.55
Marine eutrophication	kg N eq	0.0049	-21.1%	0.0004	11.83
Freshwater ecotoxicity	CTUe	153.5268	-15.1%	19.1403	8.02
Land use	kg C deficit	9.2266	-27.4%	0.7340	12.57
Water resource depletion	m^3 water eq	0.1185	-11.2%	0.0037	32.28
Mineral, fossil & ren resource depletion	kg Sb eq	0.0007	+6.2%	0.0001	6.26

Overall, the use of optical analysis in order to characterize olive imply a reduction in kg of CO_2 released. The use of the oven for conventional method releases 2677.5 kg of CO_2; using the thermobalance the overall emission of CO_2 decreased (-666 kg of CO_2), but remain high (2011.5 kg CO_2). Considering replacing the conventional method with the vis/NIR spectrophotometer, the reduction in CO_2 reaches 157.5 kg CO_2, avoiding 2520 kg of CO_2 emitted every year.

If 14.1 kg CO_2 eq is considered as the mean value of carbon footprint of one litre of packaged olive oil (Pattara et al., 2016), 5.95 kg CO_2 eq is the carbon footprint of the conventional method and 0.35 kg CO_2 eq is related to the optical one. The environmental impact of the analysis compared to the impact of the entire production process depend on the sampling frequency. The conventional analysis method weights 21 % of the total impact when analysis frequency is every 10 dm3 of oil sample, 7 % in case of analyses every 100 dm3 of oil sample and 0.2 % for analyses performed every 1000 dm3 of oil sample. These values decrease if the conventional analyses are replaced by the optical one, the percentage obtained are respectively equal to 1.2 % (10 dm3), 0.12 % (100 dm3) and 0.012 % (1000 dm3).

CONCLUSIONS

In this study the environmental impact of two methods of analyses, conventional and optical, to characterize and evaluate the olive ripening process, were analysed and compared. The LCA study highlighted how the chemicals used and the waste management are not the main responsible of the environmental damages.

The achieved results in terms of energy and analytical tools are the main drivers of the environmental impact in the conventional method, while the calibration for the optical one. More in details, due to the electric power absorption of the analytical tools and the relative construction, the conventional method shows a considerably higher impact respect to the vis/NIR spectroscopy (15 times more). Moreover, the sensitivity analysis identified highly variability of results due to the type of conventional analyses studied. Finally, future research activities should consider an expansion of the system boundary to consider the implementation of different parameters to analyse using vis/NIR devices.

Overall, quantifying the environmental damage, the results showed clear advantages for optical analysis defining spectroscopy as green technology. Moreover, this technology is also time-saving, non-destructive and user-friendly (an expert of laboratory procedure is not

required). Considering that food production has a large impact on environment, for the future, adopting spectroscopy for routine analyses could be helpful to reach more sustainable food processes.

Acknowledgments

This work was supported by AGER 2 project, grant number 2016, 0105.

References

A.O.A.C., 1975. Association of Official Agricultural Chemists. Official Methods of Analysis, 12th ed., P. O. Box 450, Benjamin Franklin station, Washington, D.C., pp: 832.

Baccouri, O., Guerfel, M., Baccouri, B., Cerretani, L., Bendini, A., Lercker, G., ... & Miled, D. D. B. (2008). Chemical composition and oxidative stability of Tunisian monovarietal virgin olive oils with regard to fruit ripening. Food chemistry, 109(4), 743-754.

Beghi, R., Buratti, S., Giovenzana, V., Benedetti S., & Guidetti R. (2017). Electronic nose and visible-near infrared spectroscopy in fruit and vegetable monitoring. Reviews in Analytical Chemistry, 36(4).

Bellincontro, A., Taticchi, A., Servili, M., Esposto, S., Farinelli, D., & Mencarelli, F. (2012). Feasible application of a portable NIR-AOTF tool for on-field prediction of phenolic compounds during the ripening of olives for oil production. Journal of agricultural and food chemistry, 60(10), 2665-2673.

Beltrán, G., Aguilera, M. P., Del Rio, C., Sanchez, S., & Martinez, L. (2005). Influence of fruit ripening process on the natural antioxidant content of Hojiblanca virgin olive oils. Food Chemistry, 89(2), 207-215.

Casson, A., Beghi, R., Giovenzana, V., Fiorindo, I., Tugnolo, A., & Guidetti, R. (2019). Visible Near Infrared Spectroscopy as a Green Technology: An Environmental Impact Comparative Study on Olive Oil Analyses. Sustainability, 11(9), 2611.

Cayuela, J. A., & Camino, M. D. C. P. (2010). Prediction of quality of intact olives by near infrared spectroscopy. European journal of lipid science and technology, 112(11), 1209-1217.

Criado, M. N., Morelló, J. R., Motilva, M. J., & Romero, M. P. (2004). Effect of growing area on pigment and phenolic fractions of virgin olive oils of the Arbequina variety in Spain. Journal of the American Oil Chemists' Society, 81(7), 633.

Desouky, I. M., Haggag, L. F., El-Migeed, M. A., & El-Hady, E. S. (2009). Changes in some physical and chemical properties of fruit and oil in some olive oil cultivars during harvesting stage. World Journal of Agricultural Sciences, 5, 760-5.

European Communities. Regulation 1989/2003 of 6 November 2003 amending Regulation (EEC) No. 2568/91 on the Characteristics of Olive Oil and Olive- pomace Oil and on the Relevant Methods of Analysis.

Fernández-Espinosa, A. J. (2016). Combining PLS regression with portable NIR spectroscopy to on-line monitor quality parameters in intact olives for determining optimal harvesting time. Talanta, 148, 216-228.

Giovenzana, V., Beghi, R., Romaniello, R., Tamborrino, A., Guidetti, R., & Leone, A. (2018). Use of visible and near infrared spectroscopy with a view to on-line evaluation of oil content during olive processing. Biosystems Engineering, 172, 102-109.

Giovenzana, V., Beghi, R., Civelli, R., Marai, S., & Guidetti, R. (2015). Postharvest characterization of olive oil fruits texture by NIR and vis/NIR spectroscopy. Chemical Engineering Transactions, 44. 61-66.

Guarino, F., Falcone, G., Stillitano, T., De Luca, A. I., Gulisano, G., Mistretta, M., & Strano, A. (2019). Life cycle assessment of olive oil: A case study in southern Italy. Journal of environmental management, 238, 396-407.

ISO 2006. ISO 14040 Environmental management—Life cycle assessment—Principles and framework. Geneva: International Organisation for Standardisation.

ISO 2006. ISO 14044 Environmental management—Life cycle assessment—requirements and guidelines. Geneva: International Organisation for Standardisation.

Jemai, H., Bouaziz, M., & Sayadi, S. (2009). Phenolic composition. sugar contents and antioxidant activity of Tunisian sweet olive cultivar with regard to fruit ripening. Journal of agricultural and food chemistry, 57(7), 2961-2968.

Magwaza, L. S., Opara, U. L., Nieuwoudt, H., Cronje, P. J., Saeys, W., & Nicolaï, B. (2012). NIR spectroscopy applications for internal and external quality analysis of citrus fruit—a review. Food and Bioprocess Technology, 5(2), 425-444.

Nasini L. & Proietti P. (2014). Olive harvesting. In John Wiley & Sons (Eds), The extra-virgin olive oil handbook (pp. 88-89). Italy.

Nicolai, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. Postharvest biology and technology, 46(2), 99-118.

Notarnicola, B., Sala, S., Anton, A., McLaren, S. J., Saouter, E., & Sonesson, U. (2017). The role of life cycle assessment in supporting sustainable agri-food systems: A review of the challenges. Journal of Cleaner Production, 140, 399-409.

Pattara, C., Russo, C., Antrodicchia, V., & Cichelli, A. (2017). Carbon footprint as an instrument for enhancing food quality: Overview of the wine, olive oil and cereals sectors. Journal of the Science of Food and Agriculture, 97(2), 396-410.

Pattara, C., Salomone, R., & Cichelli, A. (2016). Carbon footprint of extra virgin olive oil: A comparative and driver analysis of different production processes in Centre Italy. Journal of Cleaner Production, 127, 533-547.

Recchia, L., Cappelli, A., Cini, E., Garbati Pegna, F., & Boncinelli, P. (2019). Environmental Sustainability of Pasta Production Chains: An Integrated Approach for Comparing Local and Global Chains. Resources, 8(1), 56.

Rodrigues, N., Casal, S., Pinho, T., Peres, A. M., Bento, A., Baptista, P., & Pereira, J. A. (2019). Ancient olive trees as a source of olive oils rich in phenolic compounds. Food chemistry, 276, 231-239.

Roy, P., Nei, D., Orikasa, T., Xu, Q., Okadome, H., Nakamura, N., & Shiina, T. (2009). A review of life cycle assessment (LCA) on some food products. Journal of food engineering, 90(1), 1-10.

Salguero-Chaparro, L., Baeten, V., Abbas, O., & Peña-Rodríguez, F. (2012). On-line analysis of intact olive fruits by vis–NIR spectroscopy: optimisation of the acquisition parameters. Journal of Food Engineering, 112(3), 152-157.

Salguero-Chaparro, L., Palagos, B., Peña-Rodríguez, F., Roger, J. M., (2013). Calibration transfer of intact olive NIR spectra between a pre-dispersive instrument and a portable spectrometer. Computers and Electronics Agriculture, 96, 202-208.

Trapani, S., Migliorini, M., Cecchi, L., Giovenzana, V., Beghi, R., Canuti, V., ... & Zanoni, B. (2016). Feasibility of filter-based NIR spectroscopy for the routine measurement of olive oil fruit ripening indices. European Journal of Lipid Science and Technology, 119(6), 1600239.

Van der Goot, A. J., Pelgrom, P. J., Berghout, J. A., Geerts, M. E., Jankowiak, L., Hardt, N. A., ... & Boom, R.M. (2016). Concepts for further sustainable production of foods. Journal of Food Engineering, 168, 42-51.

Wolf, M. A., Pant, R., Chomkhamsri, K., Sala, S., & Pennington, D. (2012). The International Reference Life Cycle Data System (ILCD) Handbook -Towards more Sustainable Production and Consumption for a Resource-Efficient Europe; Publications Office of the European Union (Eds.). Luxembourg.

3. AIMS and OBJECTIVES

This PhD project regards different applications of non-destructive optical techniques to evaluate the quality of agri-food products as well as the development of customized optical devices to fulfil the needs of the agri-food chain which is going toward a concept of industry 4.0.

In the agri-food sector, the limited adoption of spectroscopy can be attributed to the poor knowledge of the method of analysis and costs. Therefore, in order to support the companies, this PhD thesis gives a complete overview about what concern:

1. Models development;
2. Methods comparison;
3. Variables selection;
4. Sensors design.

Concerning models development, methods comparison and variable selection (both for benchtop/process and portable commercial devices), different fields of applications have been explored (in this case, in the coffee and olive supply chain) in order to give a wider overview of the potentialities of this technique to predict qualitative features of intact food products revolutionising the quality monitoring systems and going toward an industry 4.0 approach.

Moreover, optical hand-held and stand-alone prototypes were designed, built, and tested in order to shift the current paradigm of grape maturation monitoring (based on lab analysis) with a new one that allows a cost-effective non-destructive real-time monitoring providing information with temporal and spatial resolution thanks to an infrastructure IoT.

Finally, given the evidence that this proposed proximal sensing technology has a considerable impact on many fields, a further step has been moved (during the experience abroad) in terms of cost reduction for a more complex optical technique which can gather even more information, the hyperspectral imaging. Indeed, thanks to the presence of the spatial information provided by an hyperspectral imaging approach it could solve different issues in terms of data accuracy and application in real operative conditions.

4. RESULTS

In spectroscopy, three main types of devices can be identified: benchtop instruments for applications in research centres or in industry laboratories, in-line or on-line process devices for application directly at the process line, portable devices (hand-held) for the use also in the field. Moreover, thanks to the technological evolution in terms of cost and miniaturization, the application of stan-alone vis/NIR optical devices is starting to spread in agriculture for the user directly in the field for real-time monitoring. Table 4.1 shows the main differences between the three types of NIRs devices (Beghi et al., 2017).

Table 4.1. Characteristics of the four main categories of vis/NIR devices.

Instrument type	**Application area**	**Flexibility of use**	**Applicability**	**Measurement accuracy and reproducibility**	**Cost**
Laboratory devices	Research/Industry	Adaptable to different matrices	Fixed system	Optimal	High
Process devices	Industry	Specific categories of products	Fixed system	Fair	Average/high
Hand-held devices	Also in field	Dedicated for individual products	Portable	Fair	Average/low
Stan-alone devices	Also in field	Dedicated for individual products	Fixed and portable	Fair	Low

In this chapter, a complete overview of studies has been done about the use and the development of vis/NIR devices in different supply chains with a high added value on the final product (coffee, olive, olive oil and grape for wine production).

4.1 Benchtop and process analytical instruments

PAPER 3: A reliable tool based on near-infrared spectroscopy for the monitoring of moisture content in roasted and ground coffee: a comparative study with thermogravimetric analysis

Alessio Tugnolo[a], Valentina Giovenzana[a], Cristina Malegori[b]*, Paolo Oliveri[b], Andrea Casson[a], Matteo Curatitoli[a], Riccardo Guidetti[a], and Roberto Beghi[a]

[a]Department of Agricultural and Environmental Sciences (DiSAA), Università degli Studi di Milano, via G. Celoria 2, 20133 Milan, Italy.

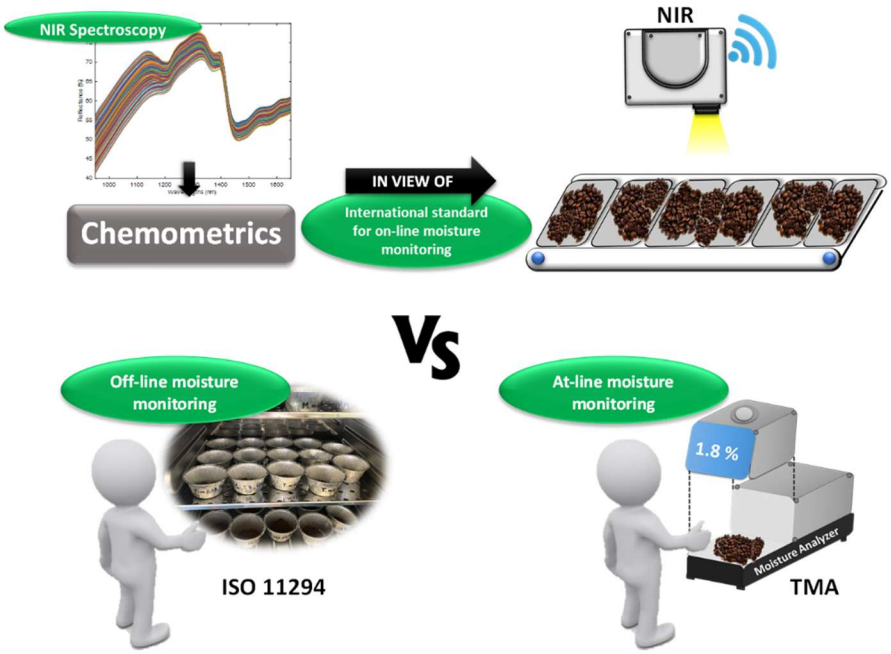
[b]Department of Pharmacy (DIFAR), Università degli Studi di Genova, viale Cembrano 4, 16148 Genoa, Italy.

* Corresponding author: malegori@difar.unige.it

Abstract

Moisture content (MC) is one of the main issues that the coffee industry has to monitor. The international standards (ISO 11817 and ISO 11294) for MC evaluation are incompatible with the rhythm of a modern productive chain. As an alternative, thermogravimetric moisture analyzers (TMA) are widely applied, but they cannot provide a real-time monitoring and management of MC. The present research aims at evaluating the performance of a near-infrared (NIR) spectrophotometer in measuring the MC of roasted beans and ground coffee, in comparison with a TMA. PLS regression models provided high predictive performances (R^2_{Pred}=0.95 and RMSEP= 0.15%, and R^2_{Pred}=0.97 and RMSEC= 0.13% for roasted and ground coffee, respectively). Passing-Bablok regression was performed to compare NIR and TMA measurements, but no significant differences were highlighted. The residual dispersion index (RDI%) was proposed, showing the higher predictive accuracy of the NIR spectrophotometer, envisaging this technology as a routine standard method for coffee MC evaluation.

Graphical abstract



Keywords:

coffee, moisture, thermogravimetric moisture analyzer (TMA), near-infrared (NIR) spectroscopy, method comparison, partial least squares regression (PLS)

INTRODUCTION

Millions of people around the world enjoy coffee every day. Currently, this hot beverage is consumed especially in Europe where, in 2019-2020, the coffee demand has exceeded 55 million bags (of 60 kg each), with an increment, since 2016-2017, of the Compound Annual Growth Rate (CAGR) by 2.1% (International Coffee Organization, 2020). The leading outcome is a highly competitive market, thus the quality of the product becomes even a more crucial aspect and is strictly related to the chemical composition of the roasted beans, which is affected by the composition of the green beans and by postharvest processing conditions (Alessandrini *et al.*, 2008).

In particular, the moisture content (MC) is a critical parameter for the whole coffee production chain. In fact, it is constantly monitored from the harvest of the berries (MC of more than 60%), passing from the drying phase, leading to the green coffee, up to the roasting and grinding phases (Adnan *et al.*, 2017). MC in green coffee, indeed, governs fermentation and mold growth during storage and transport, which could lead either to the development of off-flavors at the cup level and/or to the formation of mycotoxins (MC 8–13% enables safe transportation and storage). Besides, from an economical point of view, the coffee is paid by weight and buyers are thus more interested in buying solid coffee materials than water which, additionally, affects the bean shelf life (Reh *et al.*, 2006; Mendonça, Franca and Oliveira, 2007). As for green coffee, an inadequate MC after the roasting phase reduces the shelf life of the final product. Moreover, it produces unwanted biochemical transformations like lipid oxidation, which causes a loss of taste and body at the cup level (Caporaso *et al.*, 2018). This is due to the fact that the grinding of coffee beans with a high moisture content (higher than 6 g/100 g) leads to unacceptable particle size distribution. For these reasons, and considering also economical aspects of the ratio weight/cost, the legal limit for MC in roasted coffee is fixed at 5 g/100 g (Baggenstoss, Perren and Escher, 2008). In light of these considerations, it is possible to state that the drying and roasting phases are crucial for MC and, consequently, for the whole coffee production chain.

Regarding the analytical determination of MC, for roasted and ground coffee, two international standards are available: the ISO 11817, based on the Karl Fischer Titration (KFT), and the oven-based drying, governed by ISO 11294. The first one is basically a reaction of water and sulfur dioxide with iodine in an alcoholic solution (De Caro, Aichert and Walter, 2001), while the oven-based drying is performed by the determination of the loss in mass at 103 °C till

constant weight (Martín, Pablos and González, 1999). Due to their laboriousness and consequent lack of speed, these standards are practically incompatible with a productive chain of the modern industries, where technology has reduced the roasting time down to almost 2-4 minutes (Catelani *et al.*, 2018) followed by few minutes of grinding. Thus, to overcome this drawback, the thermogravimetric analysis (TGA) using thermogravimetric moisture analyzers (TMA) are widely utilized in the companies to obtain a fast analysis of MC. This procedure involves the measurement of 5 g of ground roasted coffee and it takes around 2–5 minutes for reaching a stable result. Despite the advantages of being solvent-free and rapid, the TMA is a benchtop equipment that has to be used off-line in a laboratory environment and usually needs a sample preparation (grinding the roasted beans). For these reasons, an off-line system like TMA may lead to undesired downtime of the production process. Currently, it is becoming essential to develop specialized, robust and low-maintenance moisture measurement systems that could allow the analysts to retrieve in real-time the MC (Cataldo *et al.*, 2017). Moreover, it is important to employ systems that could be integrated with the production chain, able to extensively analyze the whole coffee mass and whose measurement output can be made readily available to the process control systems (in this way, direct process interventions could be handled remotely, through tele-control). In this context, near-infrared (NIR) spectroscopy can be an adequate solution, thanks to its well-recognized efficiency as a process analytical technology (Bakeev, 2010).

NIR spectroscopy, combined with chemometric techniques, was already tested for the determination of MC (and other qualitative parameters) in several dry food products. In more detail, Ferreira et al. (2014) compared the promising prediction capability of NIR and MIR (mid-infrared) techniques to determine MC, proteins, lipids and ash in soybeans; several works are also available on maize, and Zhang & Guo (2020) realized an accurate detection of MC using the hyperspectral imaging in the visible/near-infrared (Vis/NIR) and NIR regions; on wheat, Miralbe & La Meta (2003) showed the high attitude of this tool to predict the main chemical parameters, including MC.

In the present work, NIR spectroscopy is proposed as an innovative technique to support coffee industry operators by providing an efficient and fast tool for the control and management of MC immediately after coffee roasting and grinding. To reach this goal, a multivariate predictive function that models the covariance structure between a block of predictor, the NIR

spectra, and a chemical response, MC, is developed thanks to the application of the partial least squares (PLS) regression method.

The reference analyses of MC were performed following the oven-based drying, governed by ISO 11294 (ISO); moreover, all the samples were analyzed by TMA. The collected dataset permitted to perform, for the first time, a direct comparison between NIR spectroscopy and TMA regarding the accuracy in determining MC in roasted and ground coffee.

An important return of this study is the possibility to propose the NIR-based approach as an international standard, with the advantage of implementing it directly on-line, in a view of a continuous process monitoring.

MATERIALS AND METHODS

Sampling

The experimental activity took place from May 2019 to January 2020, analyzing 287 samples of roasted coffee beans (200 g each one) from six blends of Arabica and Robusta varieties. The six blends were produced according to different roasting protocols that vary depending on the time-temperature conditions. Thus, the sampling scheme was defined in order to cover the widest humidity range and, consequently, to include the maximum MC variability. The samples were collected directly in the production line, after the roasting phase, and then stored at 0–4°C under vacuum in polylaminate bags equipped with a valve, which ensured an appropriate degassing of CO_2, and suitable for food contact. The bags were then shipped to the Department of Agricultural and Environmental Sciences – Production, Landscape, Agroenergy (DISAA), Università degli Studi di Milano, Milano, Italy for the analyses.

Before lab analyses, the samples of roasted coffee beans were equilibrated at room temperature (22±2°C) and then, from each bag, three aliquots of 20 g each were prepared, as outlined in Figure 1. Subsequently, one aliquot was divided into three sub-aliquots of 5.0±0.2 g each, while the remaining two aliquots were milled using a laboratory mill (MIGNON, Eureka, Italy) and, then, separated into six samples of 5.0±0.2 g each. The coffee beans sub-aliquots were analyzed firstly with the NIR spectrophotometer and, then, following the ISO MC determination. For the ground coffee samples, three of them were analyzed using the benchtop TMA (which cannot be applied on intact coffee beans), while other three sub-aliquots were analyzed with the NIR spectrophotometer and, then, following the ISO MC determination. Therefore, from 287 samples of roasted coffee beans (200 g each one), 861 aliquots (5.0±0.2

g each) of coffee samples were obtained: 287 of coffee beans and 574 of ground coffee, respectively.

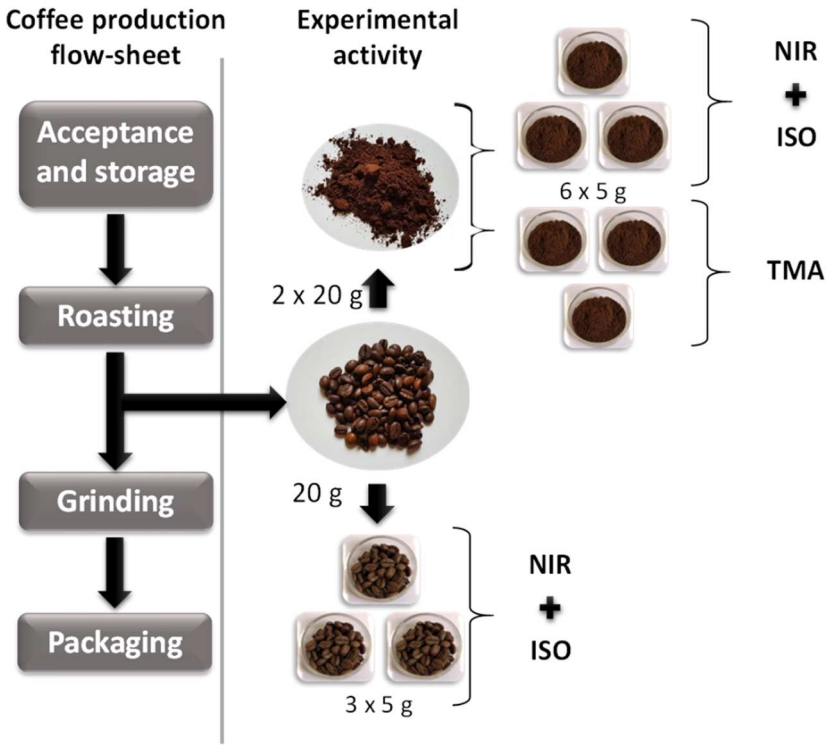


Figure 1. Description of the experimental activity performed for each coffee blend.

Oven-based drying analysis (ISO 11294)

The oven-based drying analysis, governed by ISO 11294:1994 (1994), is an International Standard which specifies a routine method for the determination of the loss in mass at 103 °C of roasted ground coffee. This method, as reported by 11294:1994 (1994), was demonstrated to give results comparable, on average, to those obtained by Karl Fischer method, reported in ISO 11817.

In the present work, according to the ISO 11294 procedure, each sub-aliquot of coffee (roasted beans and ground) was placed in a crucible and the weight was recorded using an analytical balance (LAZ 30P, Sartorius Lab Holding GmbH, Goettingen, Germany). Then, all the crucibles were placed in a lab oven (UNB400, Memmert GmbH & Co, Schwabach, Germany) set at 103°C and capable of maintaining the chamber temperature within ±1.5°C. The samples

were dried until constant weight (for around 24 h) and, then, placed in the desiccator with silica-gel, to keep the humidity constant, for 30 minutes. When the samples reached the room temperature, the crucibles were re-weighed for the MC determination:

$$MC\ (\%) = \frac{m_i - m_f}{m_i} * 100 \quad (1)$$

where m_i is the weight before the analysis and m_f is the final weight after cooling. The MC value is expressed in percentage.

Thermogravimetric moisture analysis

The thermogravimetric moisture analyzer used in the present work is the benchtop device HC103 Moisture Analyzer (Mettler Toledo®, Ohio, USA). This TMA works according to the thermogravimetric principle: the sample initial weight is recorded, then a halogen radiator dries it at 105°C, while an integrated balance continually records the sample weight. The drying process takes place through the absorption, by the sample, of infrared radiation from the halogen lamp. This heating element consists of a glass pipe filled with halogen gas, which works in combination with a gold-plated reflector enabling the distribution of the thermal radiation over the entire sample surface. The constant weight is reached after 3-5 minutes and the total loss in weight is automatically calculated. This value is interpreted as the MC and it is expressed in percentage.

NIR spectral acquisitions

Spectroscopic analyses were performed under controlled laboratory conditions, without any sample preparation, on both beans and ground coffee samples. A NIR spectrophotometer (Corona Extreme, Zeiss, Germany) for on-line measurements, equipped with a halogen light source, a diode array detector NIR module (960–1650 nm, spectral resolution 10 nm) and a plane grating spectrometer polychromator was used. To simulate the desirable final use of NIR measurements, directly on the product line, the spectrophotometer was equipped with a rotating acquisition system (TURNSTEP, Zeiss, Germany) dedicated to the analysis of large quantities of non-homogeneous samples (in the present work, 5.0±0.2 g for each sub-aliquot) within a short period of time. At the end of the acquisition phase, which takes about 10 s, the spectrum of each sample is automatically calculated as the average of 50 scans, in order to obtain a more representative spectrum and a higher signal-to-noise ratio.

Data processing

The data processing strategy can be divided in two steps: the first part was focused on developing regression functions, finding the correlation structure between the NIR data and the ISO determination of MC. The final aim of this first step was to obtain predictive models for the non-destructive and real-time quantification of MC by means of NIR spectroscopy. The second step was performed with the aim of statistically comparing the NIR-based predictions with the well-established TMA determination of MC. The whole data processing was performed on both coffee beans and ground coffee.

Outliers detection

The reference analysis (ISO 11294, in the present study) is the key element for the construction and the final performance of the calibration models. Therefore, a lack of accuracy of the reference method produces an increased error in the prediction from secondary techniques, like NIR spectroscopy (Reh *et al.*, 2006). Considering the numerous phases of the ISO analysis, even though performed by skilled personel, an evaluation of the inner variability between the three replicates (sub-aliquots of 5.0±0.2 g) was performed, with the aim of removing any possible gross error from the dataset. Therefore, following Vinutha et al. (2018), the Interquartile Range (IQR) to detect outlier for the ISO analysis was used. Moreover, a principal component analysis (PCA) was applied to detect possible outliers in the NIR spectral profiles (Croce et al., 2020).

Regression models for MC prediction

Multivariate data analysis was performed in the Matlab® environment, version 2019b (The MathWorks, Inc., Natick, MA, USA) using both the PLSToolbox package (Eigenvector Research, Inc. Manson, Washington) and in-house functions.

Different spectral preprocessing techniques were tested to remove any irrelevant information which can negatively affect the regression models. Considering the inhomogeneous physical structure of the samples (both beans and ground coffee), a correction of the baseline vertical shifts (offsets) and of the global intensity effects (typically arising from unwanted light scattering) was performed, applying the Standard Normal Variate (SNV) transform. The Savitzky and Golay second derivative (Der 2), with a third-degree polynomial order and a window size equal to 7 datapoints, was used to enhance the resolution and minimize the

spectra offset and drifts. Finally, the spectra were mean-centered column-wise, to minimize location differences between spectral variables (Biancolillo and Marini, 2018; Oliveri *et al.*, 2019).

After preprocessing, data were divided in two sub-sets by means of the Kennard and Stone duplex algorithm (Kennard and Stone, 1969): 66% of the samples was used for calibration purposes while the remaining 33% was used to perform validation.

A latent variable modelling using the PLS method, which maximizes the covariance among the NIR spectra and the ISO MC analysis, was performed (Oliveri et al., 2020). Model accuracy was evaluated using the RMSE (root mean square error), as well as bias and R^2 (coefficient of determination); the lower the error and the bias and the higher the R^2 (as maximum equal to 1), the better the model performances. Besides, the RPD (residual prediction deviation, i.e. the ratio between the standard deviation of the response variable and RMSE) was calculated. An RPD between 1.5 and 2 indicates that the model can distinguish low from high values of the response variable; a value between 2 and 2.5 indicates that rough quantitative predictions are possible, and a value between 2.5 and 3 or above corresponds to good and excellent prediction accuracy, respectively (Nicolaï *et al.*, 2007). The whole set of parameters for the evaluation of model goodness was calculated not only in calibration but also in validation, and it was used to choose the optimal number of latent variables (model complexity) for maximizing model reliability, balancing good predictions and overfitting.

Methods comparison

For a better understanding of the practical applicability of the proposed approach, the NIR-based predictions ability was compared with the TMA determinations, already accepted in the coffee industry as a valid alternative to ISO measurement. To do that, the Passing-Bablok regression method (Passing and Bablok, 2009) was applied on the MC values obtained by the different approaches (NIR, TMA and ISO) on the same set of samples (the validation set), in pairs. This regression method is particularly suitable for method comparison, since it is a symmetrical non-parametric technique, which can build regression models also when both variables (independent and dependent) have a non-negligible experimental error. For statistically evaluating the similarity/diversity between these two independent estimations, slope and intercept of the fitted line were calculated, and a significance bivariate test was conducted. The null hypothesis (H_0) was verified when the slope was not significantly different

from 1 and, simultaneously, the intercept was not significantly different from 0, meaning that there were no significant differences between the two methods, at a 95% confidence level (Mustorgi *et al.*, 2020).

So far, this statistical test did not provide any consideration regarding the magnitude of the similarity between two models considered statistically comparable. For overcoming this limitation, the information included in the residual sum of squares (RSS) of the Passing-Bablok regression was used, for the first time, as an index for quantify the similarity between two analytical methods. In more detail, a residuals dispersion index (RDI%) is proposed:

$$RDI\% = \frac{\sqrt{\frac{\sum_{i=1}^{N}(y-\hat{y})^2}{N}}}{\bar{y}} * 100 \quad (2)$$

where $\sum_{i=1}^{N}(y-\hat{y})^2$ is the RSS, N is the number of samples included in the validation set and $\bar{y}$ is the MC average value (obtained using the reference method – TB or ISO depending on the comparison that is carried out). The index is multiplied by 100 for making the interpretation even more simple: the higher the RDI% the higher the discrepancy between the two sets of data used for building the regression model. The advantage of this index is the possibility to evaluate the comparison between two methods regardless of the dataset used for performing the comparison. In the present work, it was used to evaluate comparability of NIR spectroscopy with ISO, for both coffee beans and ground coffee.

The Passing-Bablok regression function, in fact, was applied for every pair of analytical methods, analyzing the differences between the two instruments (NIR and TMA) and with the reference data (ISO) (Malegori *et al.*, 2017). This comprehensive comparison was carried out on ground coffee samples, where the TMA analysis was available, while for coffee beans only the comparison between NIR and ISO determinations was feasible.

RESULTS AND DISCUSSION

Outlier detection

Analyzing the MC data variance of each triplet of coffee sub-aliquots (from both ISO and TMA analysis), a UB value from each dataset was determined through the IQR technique (as detailed in Materials and method, section 2.5.1).

Concerning the MC determination according to the ISO procedure, the UB for roasted coffee was equal to $2.23*10^{-2}$ % while the UB for ground coffee was $3.67*10^{-2}$ %. According to the outlier detection strategy, the variance between each triplet of roasted and ground coffee was determined and, if the value was higher than the UB, the sample (all the sub-aliquots belonging to it) was considered as a potential outlier and excluded from the dataset. Starting from a total 287 samples, a total of 32 and 35 samples were excluded from the data sets of roasted and ground coffee, respectively.

Concerning the MC analysis using the TMA data, the same procedure was followed but no outliers were identified.

NIR spectra exploration and regression

Initially, a visual interpretation of the NIR spectral data, presented in Figure 2, was performed; no differences were highlighted from the spectra of roasted and ground coffee in terms of chemical assignments. However, it was possible to notice that a lower amount of radiation was reflected by the roasted beans if compared with the ground coffee (*data not shown*). This phenomenon, mainly ascribable to scattering effects, is strictly related to the physical differences (in terms of particle size and compression degree) among the coffee beans and the powder (Pizarro *et al.*, 2004).

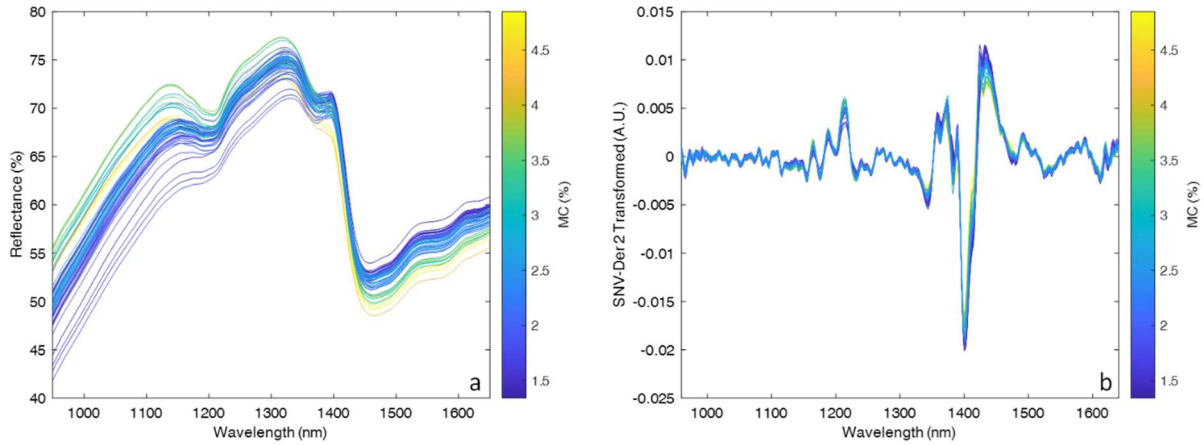


Figure 2. Raw (2a) and pretreated spectra (2b; SNV and Der 2 transformation) of ground coffee labeled according to the ISO analysis of MC (%).

As an example, Figure 2a shows the raw spectra of the ground coffee; the color bar ranges between blue (low values) and yellow (high values) of MC (ISO determination). As expected,

the predominant band of the NIR profile is approximately at 1450 nm, dominated by the water absorption bands due to the first overtone of O–H stretching (Barbin *et al.*, 2014). In the region between 1100 and 1250 nm, it was possible to notice the characteristic bands (related to the second stretching overtone of C–H) mainly linked to the absorption of lipids, carbohydrates and quinic acid, coming from the degradation of chlorogenic acids during roasting (Zhang *et al.*, 2013). Besides, due to the high content of methylxanthines, another band around 1300-1350 was identified and it is mainly relate to caffeine absorption (Osborne et al., 1993; X. Zhang et al., 2013). Lastly, another absorption was highlighted around 1550 nm, mostly related to carbohydrates compounds, amino acids and chlorogenic acids not degraded (Barbin *et al.*, 2014). Figure 2b shows the pretreated spectra, after the application of the SNV transform and second derivative; this procedure ensured a correction of the global intensity effect – due to physical scattering of the sample surface – together with the baseline drift, typically ascribable to the use of portable devices.

Figure 3 shows the outcomes of PCA performed on the pretreated spectra. The score plot (Figure 3a) shows the behavior of the samples colored according to MC (determined by ISO analysis) in the orthogonal space defined by PC1 and PC3. Almost 48% of the total variance is resumed by PC1 and it is positively related to MC; in more detail, low MC values are encountered at negative values of PC1 scores, while high MC values are detectable at positive values of PC1 scores. Adding to PC1 the contribution of PC3 (10.84 % of the total variance), six groups of samples related to the six blends of coffee used to perform this study were detected and graphically highlighted with convex hulls, which enclose samples belonging to each blend. To complete the discussion, the loading plot is reported in Figure 3b and, even if loadings are inevitably affected by data pretreatment, it can confirm the importance of the variables described above and, in particular, the water absorption band. The joint interpretation of PCA outcomes, in fact, allows to confirm that moisture, easily detectable by the NIR spectrometer, is the most informative factor that differentiates between the samples used in this study.

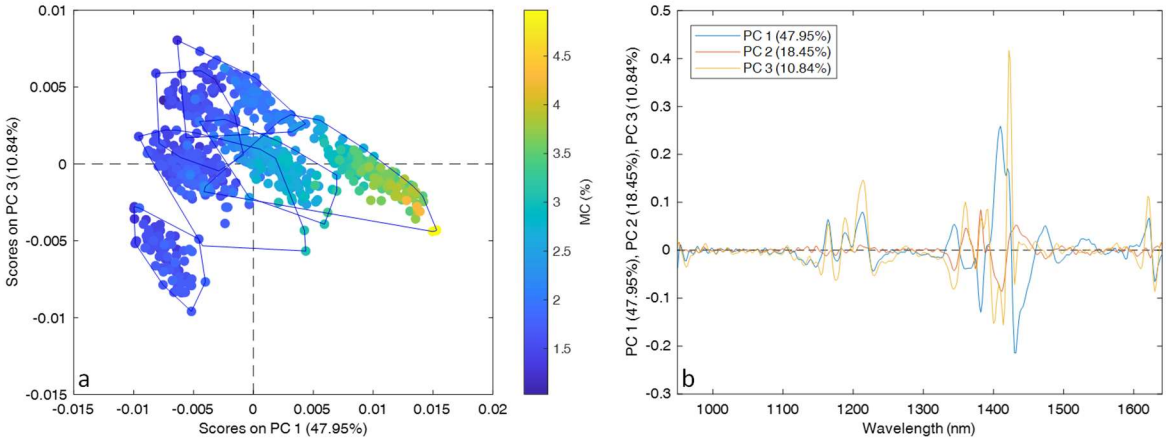


Figure 3. PCA outcomes: 3a) score plot, colored according to MC (determined by ISO analysis, expressed in %) and grouped according to the six coffee blends; 3b) loading plot of the 3 lower order components.

Thanks to the encouraging results of this exploratory step of data processing, regression models for the prediction of MC starting from NIR data were implemented, by means of the PLS method. The parameters used for evaluating the model goodness are presented in Table 1, for both roasted and ground coffee. In the same table, some descriptive statistics useful to an in-depth understanding of the dataset are detailed.

Table 1. MC descriptive statistics and figures of merit of the PLS models for roasted and ground coffee.

Matrix	Mean	SD	MC (%) Range	LVs	Cal. samples	Pred. samples	R^2_{Cal}	RMSEC (%)	R^2_{Pred}	RMSEP (%)	Pred Bias	RPD (%)
Roasted coffee	1.84	0.77	[0.79, 4.04]	4	505	260	0.96	0.17	0.95	0.15	-0.02	5.13
Ground coffee	2.21	0.73	[1.03, 4.97]	4	499	257	0.97	0.13	0.97	0.13	0.02	5.61

SD = standard deviation, LVs = latent variables, Cal = calibration, Pred = prediction.

Regarding R^2 and RMSE, it is interesting to underline the minimal differences between the performance of the models in calibration (R^2_{Cal}=0.96 and RMSEC= 0.17% for roasted coffee and R^2_{Cal}=0.97 and RMSEC= 0.13% for ground coffee) and in prediction (R^2_{Pred}=0.95 and RMSEP= 0.15% for roasted coffee and R^2_{Pred}=0.97 and RMSEC= 0.13% for ground coffee),

confirming the reliability of the model in predicting MC of unknown samples. Besides, the high RPD values around 5 (5.13 and 5.61 for roasted and ground coffee, respectively) confirm that the models are adequate for an industrial application, as a routine method.

To complete the information presented in Table 1, Figure 4 shows the reference *vs*. predicted graph, in which experimental values of MC (ISO measurement) are reported in a scatter plot versus the respective values predicted by the NIR-based model; samples represented in the graphs belong to the validation set of ground (Figure 4a) and roasted (Figure 4b) coffee. The target line, that is the bisector of the quadrant, is represented as the reference ideal line (dotted red line). As expected, the cloud of samples is more disperse for coffee beans (roasted samples) – indicating a less accurate prediction – in respect to the ground coffee, due to a more intense light scattering effect on NIR spectra.

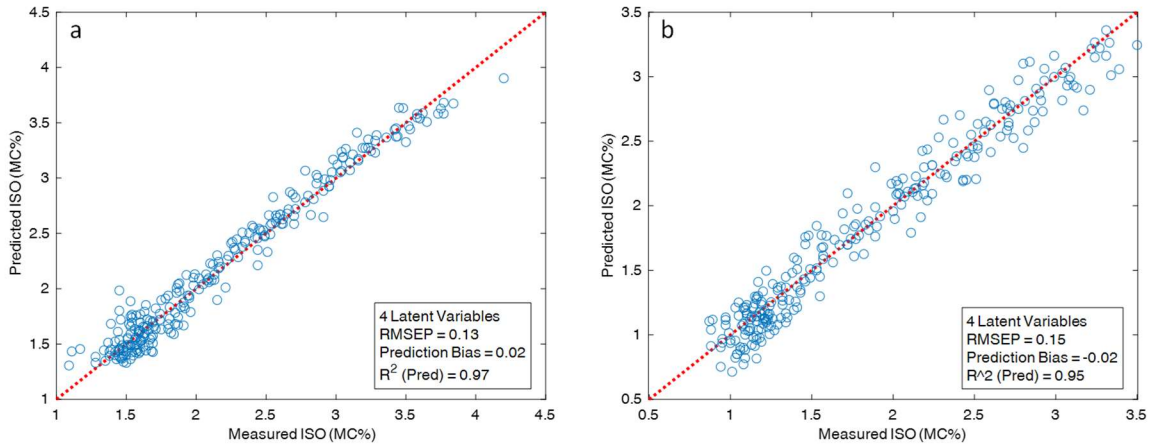


Figure 4. Reference vs. predicted graph for: 4a) ground coffee; 4b) roasted coffee.

Methods comparison

To evaluate whether significant differences of the performance between the NIR spectrophotometer and the TMA in determining the MC exist, the Passing–Bablok regression was performed on the same data used for validation of the PLS models. Using a joint test on slopes and intercepts, the three methods (NIR, TMA and ISO) were compared in pairs analyzing the differences between the two instruments (NIR and TMA), always using the ISO measurements as reference data. No statistical differences between the instruments were highlighted from the Passing–Bablok tests, at a confidence level of 95%. Therefore, the null

hypothesis (slope not significantly different from 1 and intercept not significantly different from 0) was accepted for all the paired comparisons: ISO *vs*. TMA, ISO *vs*. NIR (for ground coffee) and ISO *vs*. NIR (for roasted coffee). In Figure 5, the three Passing–Bablok regression lines are presented (solid blue lines) and the confidence interval at 95% is highlighted with dashed blue lines. The bisector of the quadrants (ideal lines) are represented, for comparison, as dotted red lines. Figure 5a reports the comparison between ISO and TMA for ground coffee, Figure 5b reports the comparison between ISO and NIR for ground coffee and Figure 5c reports the comparison between ISO and NIR for coffee beans (remembering that TMA measurements cannot be performed on coffee beans). It is possible to notice that, even if the null hypothesis of the test was accepted for all the pairs of measurements, the distribution of the samples around the regression line (solid blue line) is different, in shape and extent, in the three cases.

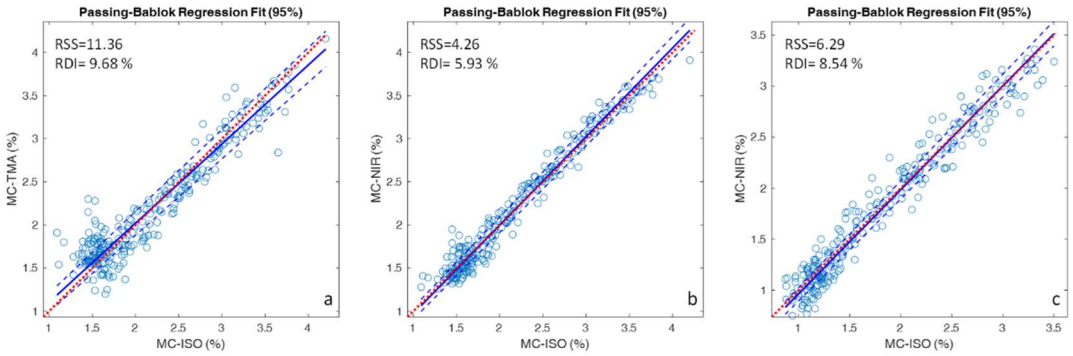


Figure 5: Passing–Bablok regression outcomes: 5a) comparison between ISO and TMA for ground coffee; 5b) comparison between ISO and NIR for ground coffee; 5c) comparison between ISO and NIR for coffee beans.

To understand this different behavior, another processing step was performed with the final aim of quantifying the magnitude of the similarity between the two instruments, NIR and TMA; therefore, the RDI% index was calculated as an informative index, which takes into account the information included into the RSS of the Passing-Bablok regression. Concerning the ground coffee, the residuals measured from ISO *vs*. TMA (Figure 5a) show an RDI equal to 9.68%, noticeably higher than the RDI coming from the comparison between ISO and NIR and equal to 5.93%. This indicates that, even if the TMA measurements and the NIR prediction can be considered statistically comparable, the NIR-based model is providing a more precise and

accurate prediction. In more detail, the TMA measurements present higher residual values along the whole range of variability of the MC, if compared with the ISO measurements (Figure 5a), especially for low MC values (between 1 and 2.5%). Conversely, the NIR-based prediction presents a lower dispersion of the samples along the entire MC range, including the low MC values.

The RDI% was also calculated for roasted coffee, comparing the data from ISO and NIR (Figure 5c). An RDI equal to 8.54% was obtained, highlighting the reliability of the NIR spectrophotometer not only after but also before the grinding process, with a predictive performance higher than TMA on ground coffee. This can represent a huge advantage, from a practical point of view, in the application of the NIR technology in the coffee industry.

CONCLUSIONS

A diode array NIR spectrophotometer (suitable for on-line measurements) was demonstrated to be a reliable tool for determining the MC in coffee samples, both roasted and ground. Thanks to PLS regression, reliable prediction models were developed, with coefficients of determination higher than 0.95 and errors in prediction lower than 0.15%. In order to understand the advantage in applying such a device in the coffee industry, in respect to the more established TMA, a comparison between the techniques was performed, using the ISO measurement as the reference values. The joint statistical test on slope and intercept, based on the Passing–Bablok regression, did not highlight significant differences between the two approaches. On the other hand, the proposed RDI% index indicated a higher accuracy and precision of the NIR-based predictions (RDI%=5.93), in respect to the TMA determinations (RDI%=9.68), along the whole MC range of variability.

The present work lays the foundation for implementing a real-scale application of the NIR technology as a routine standard method for MC evaluation of coffee, directly on-line. The final aim can be the proposal of this strategy as an international standard, paving the way for the application of NIR spectroscopy in the coffee industry as a process analytical technology.

Acknowledgements

The authors are grateful to Mattia Ardizio for the data collection and to Chiara Soncini for the technical support.

References

Adnan, A. *et al.* (2017) 'Rapid Prediction of Moisture Content in Intact Green Coffee Beans Using Near Infrared Spectroscopy', *Foods*, 6(5), p. 38. doi: 10.3390/foods6050038.

Alessandrini, L. *et al.* (2008) 'Near infrared spectroscopy: An analytical tool to predict coffee roasting degree', *Analytica Chimica Acta*, 625(1), pp. 95–102. doi: 10.1016/j.aca.2008.07.013.

Amigo, J. M. (2020) 'Hyperspectral and multispectral imaging: setting the scene', *Data Handling in Science and Technology*, 32, pp. 3–16. doi: 10.1016/B978-0-444-63977-6.00001-8.

Baggenstoss, J., Perren, R. and Escher, F. (2008) 'Water content of roasted coffee: Impact on grinding behaviour, extraction, and aroma retention', *European Food Research and Technology*, 227(5), pp. 1357–1365. doi: 10.1007/s00217-008-0852-8.

Bakeev, K. A. (2010) 'Process Analytical Technology: Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries: Second Edition', *Process Analytical Technology: Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries: Second Edition*, p. 2000139. doi: 10.1002/9780470689592.

Ballabio, D. and Consonni, V. (2013) 'Classification tools in chemistry. Part 1: Linear models. PLS-DA', *Analytical Methods*. Royal Society of Chemistry, pp. 3790–3798. doi: 10.1039/c3ay40582f.

Ballabio, D. and Consonni, V. (no date) 'Classification tools in chemistry. Part 1: linear models. PLS-DA †'. doi: 10.1039/c3ay40582f.

Barbin, D. F. *et al.* (2014) 'Application of infrared spectral techniques on quality and compositional attributes of coffee: An overview', *Food Research International*. Elsevier Ltd, 61, pp. 23–32. doi: 10.1016/j.foodres.2014.01.005.

Beghi, R. *et al.* (2017) 'Rapid evaluation of grape phytosanitary status directly at the check point station entering the winery by using visible/near infrared spectroscopy', *Journal of Food Engineering*. Elsevier Ltd, 204, pp. 46–54. doi: 10.1016/j.jfoodeng.2017.02.012.

Biancolillo, A. and Marini, F. (2018) *Chemometrics Applied to Plant Spectral Analysis*. 1st edn, *Comprehensive Analytical Chemistry*. 1st edn. Elsevier B.V. doi:

10.1016/bs.coac.2018.03.003.

Bonilla, I., De Toda, F. M. and Martínez-Casasnovas, J. A. (2015) 'Vine vigor, yield and grape quality assessment by airborne remote sensing over three years: Analysis of unexpected relationships in cv. Tempranillo', *Spanish Journal of Agricultural Research*, 13(2), pp. 1–8. doi: 10.5424/sjar/2015132-7809.

Borràs, E. *et al.* (2015) 'Data fusion methodologies for food and beverage authentication and quality assessment - A review', *Analytica Chimica Acta*, 891, pp. 1–14. doi: 10.1016/j.aca.2015.04.042.

Boulet, J. C. and Roger, J. M. (2012) 'Pretreatments by means of orthogonal projections', *Chemometrics and Intelligent Laboratory Systems*. Elsevier B.V., 117, pp. 61–69. doi: 10.1016/j.chemolab.2012.02.002.

Bro, R. and Smilde, A. K. (2014) 'Principal component analysis', *Analytical Methods*, 6(9), pp. 2812–2831. doi: 10.1039/c3ay41907j.

Buratti, S. *et al.* (2017) 'Electronic nose and visible-near infrared spectroscopy in fruit and vegetable monitoring', *Reviews in Analytical Chemistry*, 36(4), pp. 1–24. doi: 10.1515/revac-2016-0016.

Caporaso, N. *et al.* (2018) 'Rapid prediction of single green coffee bean moisture and lipid content by hyperspectral imaging', *Journal of Food Engineering*. Elsevier Ltd, 227, pp. 18–29. doi: 10.1016/j.jfoodeng.2018.01.009.

De Caro, C. A., Aichert, A. and Walter, C. M. (2001) 'Efficient, precise and fast water determination by the Karl Fischer titration', *Food Control*, 12(7), pp. 431–436. doi: 10.1016/S0956-7135(01)00020-2.

Casale, M. and Simonetti, R. (2014) 'Review: Near infrared spectroscopy for analysingolive oils', *Journal of Near Infrared Spectroscopy*. IM Publications LLP, pp. 59–80. doi: 10.1255/jnirs.1106.

Cataldo, A. *et al.* (2017) 'Moisture content monitoring of construction materials: From in-line production through on-site applications', *I2MTC 2017 - 2017 IEEE International Instrumentation and Measurement Technology Conference, Proceedings*. IEEE, pp. 1–5. doi: 10.1109/I2MTC.2017.7969762.

Catelani, T. A. *et al.* (2018) 'Real-time monitoring of a coffee roasting process with near infrared spectroscopy using multivariate statistical analysis: A feasibility study', *Talanta*, 179. doi: 10.1016/j.talanta.2017.11.010.

Cortés, V. *et al.* (2019) 'Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review', *Trends in Food Science and Technology*. Elsevier Ltd, pp. 138–148. doi: 10.1016/j.tifs.2019.01.015.

Corti, M. *et al.* (2018) 'Does remote and proximal optical sensing successfully estimate maize variables? A review', *European Journal of Agronomy*. Elsevier, 99(June), pp. 37–50. doi: 10.1016/j.eja.2018.06.008.

Croce, R. *et al.* (2020) 'Prediction of quality parameters in straw wine by means of FT-IR spectroscopy combined with multivariate data processing', *Food Chemistry*, 305. doi: 10.1016/j.foodchem.2019.125512.

Fernández-Cuevas, I. *et al.* (2015) 'Classification of factors influencing the use of infrared thermography in humans: A review', *Infrared Physics and Technology*, 71, pp. 28–55. doi: 10.1016/j.infrared.2015.02.007.

Fernández-Espinosa, A. J. (2016) 'Combining PLS regression with portable NIR spectroscopy to on-line monitor quality parameters in intact olives for determining optimal harvesting time', *Talanta*. doi: 10.1016/j.talanta.2015.10.084.

Ferreira, D. S. *et al.* (2014) 'Comparison and application of near-infrared (NIR) and mid-infrared (MIR) spectroscopy for determination of quality parameters in soybean samples', *Food Control*, 35(1), pp. 227–232. doi: 10.1016/j.foodcont.2013.07.010.

G omez-Caravaca, A. M., en Maggio, R. M. and Cerretani, L. (2016) 'Chemometric applications to assess quality and critical parameters of virgin and extra-virgin olive oil. A review'. doi: 10.1016/j.aca.2016.01.025.

Kennard, R. W. and Stone, L. A. (1969) *American Society for Quality Computer Aided Design of Experiments*. Available at: https://www.jstor.org/stable/pdf/1266770.pdf?refreqid=excelsior%3Aad741e5f06323f1f5c14558582dbd5e0 (Accessed: 9 July 2019).

Khan, M. J. *et al.* (2018) 'Modern Trends in Hyperspectral Image Analysis: A Review', *IEEE*

Access. doi: 10.1109/ACCESS.2018.2812999.

Liakos, K. G. *et al.* (2018) 'Machine learning in agriculture: A review', *Sensors (Switzerland)*, 18(8), pp. 1–29. doi: 10.3390/s18082674.

Mahmoud, S., Lotfi, A. and Langensiepen, C. (2016) 'User activities outliers detection; Integration of statistical and computational intelligence techniques', *Computational Intelligence*, 32(1), pp. 49–71. doi: 10.1111/coin.12045.

Malegori, C. *et al.* (2017) 'Comparing the analytical performances of Micro-NIR and FT-NIR spectrometers in the evaluation of acerola fruit quality, using PLS and SVM regression algorithms', *Talanta*, 165, pp. 112–116. doi: 10.1016/j.talanta.2016.12.035.

Malegori, C. *et al.* (2018) 'GlutoPeak profile analysis for wheat classification: Skipping the refinement process', *Journal of Cereal Science*. doi: 10.1016/j.jcs.2017.09.005.

Martín, M. J., Pablos, F. and González, A. G. (1999) 'Characterization of arabica and robusta roasted coffee varieties and mixture resolution according to their metal content', *Food Chemistry*, 66(3), pp. 365–370. doi: 10.1016/S0308-8146(99)00092-8.

Mendonça, J. C. F., Franca, A. S. and Oliveira, L. S. (2007) 'A comparative evaluation of methodologies for water content determination in green coffee', *LWT - Food Science and Technology*, 40(7), pp. 1300–1303. doi: 10.1016/j.lwt.2006.08.013.

Miralbe Ä, C. S. and La Meta, H. S. (2003) 'Prediction Chemical Composition and Alveograph Parameters on Wheat by Near-Infrared Transmittance Spectroscopy'. doi: 10.1021/jf034235g.

Mohd Ali, M. *et al.* (2020) 'Emerging non-destructive thermal imaging technique coupled with chemometrics on quality and safety inspection in food and agriculture', *Trends in Food Science and Technology*. Elsevier Ltd, 105(July), pp. 176–185. doi: 10.1016/j.tifs.2020.09.003.

Mustorgi, E. *et al.* (2020) 'A chemometric strategy to evaluate the comparability of PLS models obtained from quartz cuvettes and disposable glass vials in the determination of extra virgin olive oil quality parameters by NIR spectroscopy'. doi: 10.1016/j.chemolab.2020.103974.

Nicolaï, B. M. *et al.* (2007) 'Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review', *Postharvest Biology and Technology*, pp. 99–118. doi: 10.1016/j.postharvbio.2007.06.024.

Oliveri, P. *et al.* (2019) 'The impact of signal pre-processing on the final interpretation of analytical outcomes – A tutorial', *Analytica Chimica Acta*. Elsevier Ltd, 1058, pp. 9–17. doi: 10.1016/j.aca.2018.10.055.

Oliveri, P. and Downey, G. (2013) *Discriminant and class-modelling chemometric techniques for food PDO verification*. 1st edn, *Comprehensive Analytical Chemistry*. 1st edn. Copyright © 2013 Elsevier B.V. All rights reserved. doi: 10.1016/B978-0-444-59562-1.00013-X.

Osborne, B. G., Fearn, T. and Hindle, P. H. (1993) *Practical NIR spectroscopy with applications in food and beverage analysis*, *Longman Scientific and Technical*.

Passing, H. and Bablok, W. (2009) 'A New Biometrical Procedure for Testing the Equality of Measurements from Two Different Analytical Methods. Application of linear regression procedures for method comparison studies in Clinical Chemistry, Part I', *Clinical Chemistry and Laboratory Medicine*, 21(11), pp. 709–720. doi: 10.1515/cclm.1983.21.11.709.

Pizarro, C. *et al.* (2004) 'Influence of data pre-processing on the quantitative determination of the ash content and lipids in roasted coffee by near infrared spectroscopy', *Analytica Chimica Acta*, 509(2), pp. 217–227. doi: 10.1016/j.aca.2003.11.008.

Reh, C. T. *et al.* (2006) 'Water content determination in green coffee - Method comparison to study specificity and accuracy', *Food Chemistry*, 96(3), pp. 423–430. doi: 10.1016/j.foodchem.2005.02.055.

Rinnan, Å. (2014) 'Pre-processing in vibrational spectroscopy-when, why and how', *Analytical Methods*, 6(18), pp. 7124–7129. doi: 10.1039/c3ay42270d.

Still, C. *et al.* (2019) 'Thermal imaging in plant and ecosystem ecology: applications and challenges', *Ecosphere*. doi: 10.1002/ecs2.2768.

Stuart, M. B., McGonigle, A. J. S. and Willmott, J. R. (2019) 'Hyperspectral imaging in environmental monitoring: A review of recent developments and technological advances in compact field deployable systems', *Sensors (Switzerland)*, 19(14). doi: 10.3390/s19143071.

Tefas, A. and Pitas, I. (2016) 'Principal component analysis', in *Intelligent Systems*. doi: 10.1201/b17700-1.

Todeschini, R. (1998) 'Introduzione alla chemiometria', *EdiSES, Napoli*. Available at: http://michem.disat.unimib.it/chm/download/materiale/introduzione_alla_chemiometria.pdf.

Trapani, S. *et al.* (2017) 'Feasibility of filter-based NIR spectroscopy for the routine measurement of olive oil fruit ripening indices', *European Journal of Lipid Science and Technology*, 119(6), pp. 1–4. doi: 10.1002/ejlt.201600239.

Tugnolo, A. *et al.* (2019) 'Characterization of green, roasted beans, and ground coffee using near infrared spectroscopy: A comparison of two devices', *Journal of Near Infrared Spectroscopy*, 27(1), pp. 93–104. doi: 10.1177/0967033519825665.

Tugnolo, A. *et al.* (2020) 'A diagnostic visible/near infrared tool for a fully automated olive ripeness evaluation in a view of a simplified optical system', *Computers and Electronics in Agriculture*. Elsevier B.V., 180, p. 105887. doi: 10.1016/j.compag.2020.105887.

Zhang, X. *et al.* (2013) 'Improvement of near infrared spectroscopic (NIRS) analysis of caffeine in roasted arabica coffee by variable selection method of stability competitive adaptive reweighted sampling (SCARS)', *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*. Elsevier B.V., 114, pp. 350–356. doi: 10.1016/j.saa.2013.05.053.

Zhang, Y. and Guo, W. (2020) 'Moisture content detection of maize seed based on visible/near-infrared and near-infrared hyperspectral imaging technology', *International Journal of Food Science & Technology*. Blackwell Publishing Ltd, 55(2), pp. 631–640. doi: 10.1111/ijfs.14317.

Zhong, Y. *et al.* (2018) 'Computational intelligence in optical remote sensing image processing', *Applied Soft Computing Journal*. Elsevier Ltd, pp. 75–93. doi: 10.1016/j.asoc.2017.11.045.

Zhu, X. *et al.* (2010) 'Detection of adulterants such as sweeteners materials in honey using near-infrared spectroscopy and chemometrics', *Journal of Food Engineering*, 101(1), pp. 92–97. doi: 10.1016/j.jfoodeng.2010.06.014.

In this paper, the candidate was responsible of the formal analysis and the curation of the data along the whole sampling campaign.

PAPER 4: Prediction of olive ripening degree combining image analysis and FT-NIR spectroscopy for virgin olive oil optimization

Cristina Alamprese[a], Silvia Grassi[a], Alessio Tugnolo[b] and Ernestina Casiraghi[a]

[a]Department of Food, Environmental, and Nutritional Sciences (DeFENS), Universit`a Degli Studi di Milano, Via G. Celoria 2, 20133, Milan, Italy

[b]Department of Agricultural and Environmental Sciences (DiSAA), Università degli Studi di Milano, via G. Celoria 2, 20133 Milan, Italy.

* Corresponding author: silvia.grassi@unimi.it

Abstract

This work proposes classification models for the prediction of olive maturity index based on Fourier Transform-Near Infrared (FT-NIR) spectra of intact drupes. An image analysis (IA) method was purposely developed for the objective evaluation of the maturity index. Thirteen cultivars at different ripening stages were harvested along three years. The reliability of the IA method was confirmed by the highly significant correlation with the common visual evaluation of maturity index. Classification models were developed with Partial Least Square-Discriminant Analysis (PLS-DA), using IA results and FT-NIR spectra of olives collected in diffuse reflectance. Most PLS-DA models calculated separately for olive origin gave sensitivity and specificity values in prediction higher than 81%. The global model performed slightly worse (sensitivity, 79%; specificity, 75%), but it is definitely more robust and can provide the olive sector with a fast, green and non-destructive olive sorting method for the production of high quality virgin oil.

Keywords: Olive quality, Maturity index, Intact fruit Image analysis, Near infrared spectroscopy

Abbreviations

AR, Abruzzo Region; CA, Calabria Region; CV, cross-validation; EVOO, extra virgin olive oil; FN, false negative; FP, false positive; FT-NIR, Fourier Transform-Near Infrared; IA, image analysis; LV, Latent Variable; NIR, Near Infrared; MI, Maturity Index; PC, Principal Component; PCA, Principal Component Analysis; PLS-DA, Partial Least Square-Discriminant Analysis; PR, Apulia Region; ROI, Region of Interest; SCI, Surface Colorimetric Index; SENS, sensitivity; SPEC, specificity; SR, Sardinia Region; SNV, Standard Normal Variate; TP, true positive; TN, true negative.

INTRODUCTION

The quality of virgin and extra virgin olive oils is strongly related to the physiological conditions and the ripening stage of the fruits from which they are extracted. In general, a progressive deterioration of oil quality is observed as fruit ripening progresses, but different behaviours

have been registered in distinct olive cultivars (Camposeo et al., 2013; Garcia et al., 1996). To date, the optimal harvest time has been selected mainly using traditional approaches and personnel experience rather than scientific criteria (Garcia et al., 1996), thus making difficult the objective optimisation of the oil quality.

The most common method used to define the optimal harvest time is based on olive visual inspection (Uceda & Frías, 1975), but other evaluations have been proposed over the years, such as the measurement of fruit detachment force and fresh weight (Camposeo et al., 2013), the determination of flesh firmness (Garcia et al., 1996) or respiration rate (Ranalli et al., 1998), as well as the drupe oil (Allalout et al., 2011; Correa et al., 2019) and polyphenol (Morello ´ et al., 2004) content. However, all these methods are laborious and time-consuming, thus limiting the efficiency of controls and preventing the possibility to develop in-line applications for the olive oil industries. In order to overcome these issues and develop rapid, reliable and automatic methods for the assessment of olive ripening stage, computer vision and near infrared (NIR) spectroscopy approaches have been recently studied. For instance, Guzman ´ et al. (2015) proposed a machine vision system based on the use of infrared and visible images for the prediction of maturity index of Picual olives determined by a panel of experienced evaluators. Soto et al. (2018) presented a proposal to automatically determine olive oil quality parameters

by processing images acquired from olive fruits. They tested 84 batches of olives and the image processing was guided by experts' assessment of fruit conditions. Ram et al. (2010) applied image processing to the prediction of oil quantity in Picual and Souri olives during the ripening season. As for vibrational spectroscopy, Gracia and Leon ´ (2011) studied the evolution of oil and moisture content in intact olives during the maturity process by using a portable NIR device and testing eight Spanish cultivars. Cayuela and Camino (2010) carried out a similar study, testing the usefulness of a portable Vis/NIR spectrometer for the prediction of fruit moisture, oil content, fruit maturity index and oil free acidity of two Spanish olive varieties. Bellincontro et al. (2012) applied a portable NIR-AOTF (Acousto Optically Tunable Filter) device for the prediction of the main phenolic compounds (oleuropein, verbascoside, and 3, 4-DHPEA-EDA) and total phenols of three Italian olive cultivars. All these studies investigated the quality parameters by collecting spectra on the single fruits and considering only one harvest season.

Salguero-Chaparro et al. (2013) analysed 250 samples of intact olives belonging to fifty varieties and harvested in two crop seasons in order to test a NIR device set on a conveyor belt for the on-line determination of the oil content, moisture, and free acidity. Other similar works were reviewed by Stella et al. (2015), who concluded that the advantages of NIR spectroscopy over traditional analytical methods are clear, but the robustness of the developed models is affected by small datasets. Similarly, Nenadis and Tsimidou (2017) reviewed papers focused on vibrational techniques addressing quality and authenticity issues of olives and virgin olive oils and they discussed the need for validation guidelines and open-access spectral databanks in order to standardize these techniques and use them in the official control. They also pointed out the importance of the reference analytical methods for the successful calibration of chemometric models and the reduction of errors. In this respect, the use of visual inspection evaluations as references for the calibration of olive classification models based on maturity index can represent a source of errors preventing the development of reliable methods. Thus, the aim of this work was the development of a classification model based on Fourier Transform-Near Infrared (FT-NIR) spectra of aliquots of intact olives using as a priori information the olive maturity index objectively evaluated through an image analysis method purposely developed. A big dataset was used to set up the proposed methodology, considering thirteen Italian olive cultivars at different ripening stages, harvested during three crop seasons (from 2016 to 2018), for a total of 303 samples corresponding to as many olive aliquots.

MATERIALS AND METHODS

Experimental plan

Olives at different maturity stages were collected from September to December over three harvesting years (2016, 2017, and 2018). Thirteen cultivars typical of four different Italian regions were sampled as detailed in Table 1. For each sampling time (from T1 up to T5) and cultivar, three sample units (500 g each) were collected from different labelled trees of the same grove. The Maturity Index (MI) was evaluated picking olives from the three sample units; furthermore, two aliquots (100 g each) were withdrawn from each sample unit, to be used for surface colorimetric index assessment, image analysis, and FT-NIR spectroscopy. A total of 303 aliquots of olives in adequate hygienic and qualitative conditions were analysed after wiping olive surface with paper towel.

Table 1. Overview of the olive sampling plan. T1-T5 represent the different sampling times over the ripening period.

		Sampling times per harvesting year			
Region	**Cultivar**	**2016**	**2017**	**2018**	**Number of aliquots**
Abruzzo (AR)	*Dritta*		*T1-T2-T3-T4*		23
	Gentile		*T1-T2-T3-T4*	*T1-T2-T3*	25
	Tortiglione		*T1-T2-T3-T4*	*T1-T2*	29
Calabria (CR)	*Calipa*		*T1-T2-T3-T4-T5*	*T1-T2-T3-T4-T5*	26
	Cannavà		*T1-T2-T3-T4-T5*	*T1-T2-T3-T4-T5*	19
	Ciciariello		*T1-T2-T3-T4-T5*	*T1-T2-T3*	23
	Filogaso		*T1-T2-T3-T4-T5*		21
Apulia (PR)	*Bambina*		*T1-T2-T3*	*T1-T2-T3*	28
	Cima di melfi	*T1-T2-T3*	*T1-T2-T3*	*T1-T2-T3*	32
	Oliva Rossa	T1-T2	T1-T2-T3	T1-T2	15
Sardinia (SR)	Corsicana	T1-T2-T3-T4	T1-T2-T3-T4	T1	20
	Semidana	T1-T2-T3-T4	T1-T2-T3-T4	T1	22
	Sivigliana	T1-T2-T3-T4	T1-T2-T3-T4	T1	20

Visual assessment of ripening stage

Maturity index

From each olive sample unit, 100 drupes were randomly collected to assess the MI according to the visual approach proposed by Uceda and Frías (1975). The method classifies olives in eight classes according to both skin and pulp colour and then MI is calculated following equation (1):

$$MI = \sum_{i=0}^{7} \left(\frac{i \, x \, ni}{N} \right) \qquad (1)$$

where *i* is the class number, *ni* is the number of olives belonging to the *i* class, and *N* is the total number of considered olives (100).

Surface colorimetric index

Each olive aliquot was evaluated by a non-destructive visual assessment of the ripening stage, considering only the skin colour as reported in Table 2. Results were expressed by an index referred as Surface Colorimetric Index (SCI) calculated according to equation (2):

$$SCI = \sum_{i=1}^{4} \left(\frac{i \, x \, ni}{N} \right) \qquad (2)$$

where *i* is the class number, *ni* is the number of olives belonging to the *i* class, and *N* is the total number of considered olives (60).

Table 2. Olive maturity classes considered in the Surface Colorimetric Index (SCI).

Class number	Olive skin colour
1	100% green
2	<50% turning red, purple or black
3	>50% turning red, purple or black
4	100% purple or black

Image analysis

Images of each olive aliquot were collected using a flatbed scanner (Scanjet 8300, HP Inc., Palo Alto, CA, USA) controlled by VueScan software (v. 9.4, 2016, Hamrick Software, Sunny Isles Beach, FL, USA). The entire aliquot was placed on the scanner glass and covered with a black box to prevent light losses and avoid noise from the environment. Two images were acquired for each aliquot, at 600 dpi resolution, with a 24 bit colour depth and saved in TIFF format; the olives were blended inbetween the two acquisitions. Images were then elaborated using a Matlab (v. 2016a, Mathworks, Inc., Natick, MA, USA) routine purposely developed to objectively assess the olive ripening stage. At first, in order to distinguish olives from background, a Region of Interest (ROI) was segmented applying the Superpixel algorithm on each RGB channel. The algorithm performs a simple linear iterative clustering (Achanta et al., 2012) in order to group pixels into regions with similar intensity values. From a practical point of view, this algorithm groups together neighbour pixels based on their RGB intensity values and allows the background removal. An example of the original image and the segmented one is reported in Fig. 1. The average red channel intensity of the segmented mask, containing only the olive pixels, was then calculated and used to assign each olive aliquot to a maturity class (named IA).

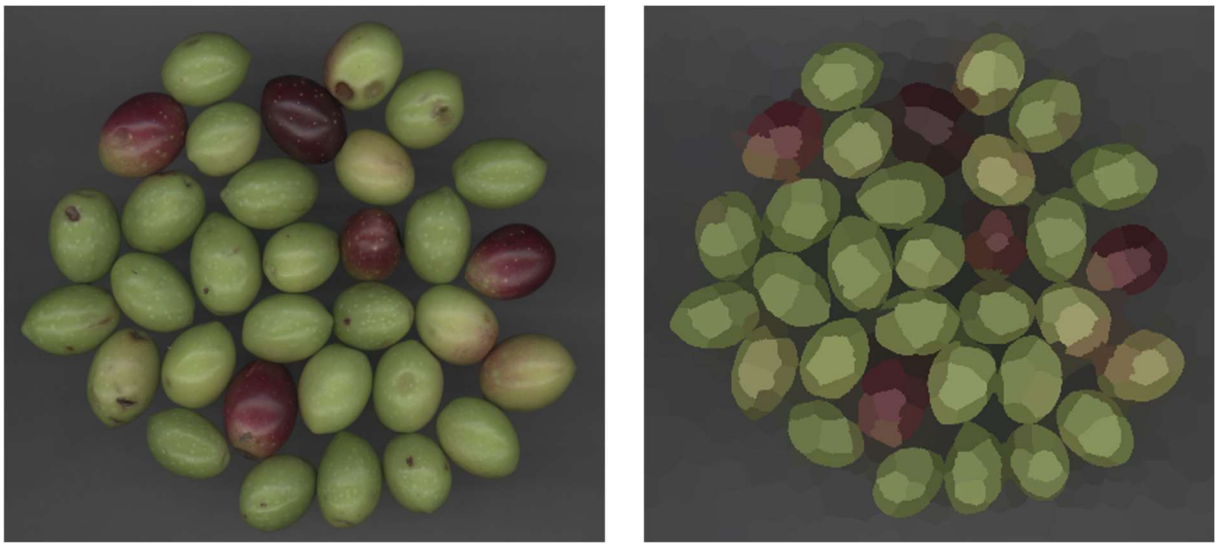


Figure 1. Example of an olive image before (left) and after (right) the segmentation by using the Superpixel algorithm.

FT-NIR spectroscopy

Each olive aliquot was placed in a glass Petri dish and analysed in diffuse reflectance by a FT-NIR spectrometer (MPA, Bruker Optics, Milan, Italy) equipped with an integrating sphere (Fig. 2). Spectra were collected in the whole NIR region (12,500–4000 cm− 1), with a resolution of 8 cm− 1 and 32 scans for both sample and background. Two spectra were collected for each aliquot, blending the olives in-between measurements.

Figure 2. Image of the FT-NIR spectrometer (MPA, Bruker Optics, Milan, Italy) equipped with an integrating sphere used for the spectral data collection. The detail shows the glass Petri dish containing an olive aliquot under analysis.

Data analysis

Linear correlations between the maturity class indices (i.e. MI vs SCI and SCI vs IA) were evaluated calculating Pearson correlation coefficients and the corresponding significance (p-value). The two FT-NIR spectra collected from each olive aliquot were averaged and organised in a global dataset. Furthermore, four independent datasets were created according to olive origin (AR, CR, PR, and SR). Spectra were reduced in the 10,500–4000 cm− 1 range and pretreated with smoothing, Standard Normal Variate (SNV), and first derivative, alone or in combination. Before further analyses, spectral data were explored by Principal Component Analysis (PCA). Then, the maturity classes identified by image analysis were used as a priori information for the development of classification models based on Partial Least Square-Discriminant Analysis (PLS-DA). A dataset (206 samples) constructed with aliquots of olives

collected in 2017 was used for model calibration and cross-validation (venetian blinds, ten splits). A test set for external validation was created merging data collected during 2016 and 2018 (97 aliquots). The PLS-DA model performance was evaluated by comparing the reference class of each sample (IA) with the class predicted by the model. The most probable approach was applied to assign samples to classes. Samples were assigned to the class reaching the highest probability regardless the probability magnitude, thus no probability threshold was considered. This approach is advantageous if there is the need of unambiguous class assignment when “no class” o “multiple class” has no meaning. Sensitivity and specificity were calculated for each maturity class in calibration, cross-validation and prediction. Sensitivity (SENS, equation (3)) expresses the model capability to correctly recognize samples belonging to the considered class. Specificity (SPEC, equation (4)) describes the model capability to correctly reject samples that do not belong to the considered class (Ballabio & Consonni, 2013).

$$SENS = \frac{TP}{(TP+\quad)} * 100 \qquad (3)$$

$$SPEC = \frac{TN}{(TN+FP)} * 100 \qquad (4)$$

where TP are the true positive samples, FN are the false negative samples, TN are the true negative samples, and FP are the false positive samples. Both SENS and SPEC can assume values between 0% and 100%, being 100% the totally correct classification of TP and the totally correct rejection of TN for sensitivity and specificity, respectively. All data analyses were performed in Matlab environment (v. 2016a, Mathworks, Inc., Natick, MA, USA), using the PLS toolbox (v. 8.5, Eigenvector Research, Inc., Seattle, WA, USA).

RESULTS AND DISCUSSION

Objective assessment of olive maturity class by image analysis

MI of olive samples ranged from 0 to 5.9, with a total mean of 2.4 and a standard deviation value of 1.4. The highest mean values (3.0–3.3) were reached by the cultivars Filogaso (CR), Corsicana (SR), Sivigliana (SR), and Dritta (AR), while the cultivar Semidana (SR) had an average MI lower than 1.0. Thus, despite the long harvesting period considered (about 4 months), none of the olive cultivars reached the highest possible value of MI (7.0). This is in

agreement with results of previous published works, in which MI ranged from 0 to 6 (Baccouri et al., 2007) or from 0 to 3 (Vidal et al., 2019), depending on the considered olive cultivars. In general, it has been reported that the best results in terms of both oil quantity and quality (mill extraction yield, induction time, polyphenol content, aromatic compounds, sensory score) are obtained with olives harvested when their degreening is limited to the skin, i.e. with a maturity index less than 4 in a 0–7 scale (Camposeo et al., 2013). The MI assessment is very time-consuming and the obtained results strongly depend on the experience of the evaluator, the olive health and physical state, and the environmental conditions, such as lighting (Guzman ´ et al., 2015). Due to these issues, a simpler and non-destructive visual evaluation based only on skin colour was also performed on all the olive samples. The resulting SCI ranged from 1 to 4, with a total mean of 2.3 and a standard deviation value of 1.0. MI and SCI values resulted highly correlated ($r = 0.95$; $p < 0.001$), thus indicating the effectiveness of the procedure developed for SCI assessment, and the higher importance of the skin colour with respect to the pulp one. However, also the SCI determination is highly dependent on analysis conditions and evaluator experience, whereas a vision system approach could overcome these drawbacks. In fact, the scanning procedure allows lighting standardisation, and image analysis gives more objective and reproducible results. Thus, a suitable image analysis procedure was developed and the average value of the red channel was used to assign the olive aliquots to a maturity class. A simpler method, based on three-class maturity assignment, as reported in Table 3, was considered instead of the four-class division of SCI. A global red channel thresholding was defined by considering the images of all the olive cultivars and sampling times of the crop seasons 2016 and 2018. The reliability of the thresholding was assessed on the images acquired in 2017. The IA index had a global mean of 2, with a standard deviation value of 1. Also in this case, a highly significant correlation was found between IA and SCI values ($r = 0.85$, $p < 0.001$), thus demonstrating that the subjective visual inspection can be substituted by the objective vision system approach, enabling the standardisation and automation of the maturity index evaluation. Furthermore, samples of the Interquartile Range (IQR) of the three IA classes had MI ranges of 0.80–1.55, 1.65–3.00, and 3.40–4.55, for class 1, 2, and 3, respectively, thus demonstrating the reliability of the IA method in discriminating olives with different ripening degree.

Table 2. Olive maturity classes defined according to the red channel average intensity calculated with the developed image analysis procedure.

Class number	Olive skin colour	Red channel average intensity
1	100% green	250.0 – 90.0
2	<50% turning red, purple or black	89.9 – 71.0
3	>50% turning red, purple or black	70.9 – 0.0

Interpretation of olive FT-NIR spectra

The raw spectra acquired for each aliquot of olives are reported in Fig. 3a. In the region between 9000 and 8000 cm− 1 it is possible to notice the characteristic bands mainly linked to the absorption of fats and oils (Correa et al., 2019). Indeed, the absorbance at 8733 and 8620 cm− 1 is related to the stretching of the C–O bonds of aliphatic esters, while the second overtone of C–H stretching vibrations of alkyl groups and alkenes occurs at 8245 and 8030 cm− 1 (Fernandez-Espinosa, ´ 2016). Other characteristic bands are present around 6900 cm− 1 (6842 cm− 1 , –CH2– 2C–H stretch + C–H deformation; 6900 cm− 1 , first overtone of –OH bond) and in the region between 6700 and 5300 cm− 1 , characterised by CH3–CH stretch (5950 cm− 1) and OH, –CO stretch (5500 cm− 1). The absorbance around 5200 cm− 1 is related to the harmonic and combination bands of O–H bonds in hydroxyl groups (Fernandez-Espinosa, 2016; Trapani et al., 2017). Besides, the relevant differences in absorbance intensity within 4500 and 4000 cm− 1 are linked to the combination of CH stretching with other vibrational modes (Casale & Simonetti, 2014). The peak at 4085 cm− 1 was characteristic of the olives assigned to the IA classes 1 (100% green skin colour) and 2 (<50% of the skin colour turning red). Actually, this peak disappeared with the ripening progress, being replaced in the spectra of the class 3 olives by two peaks at 4335 cm− 1 (CH second overtone) and 4262 cm− 1 (CH2 second overtone) (Trapani et al., 2017).

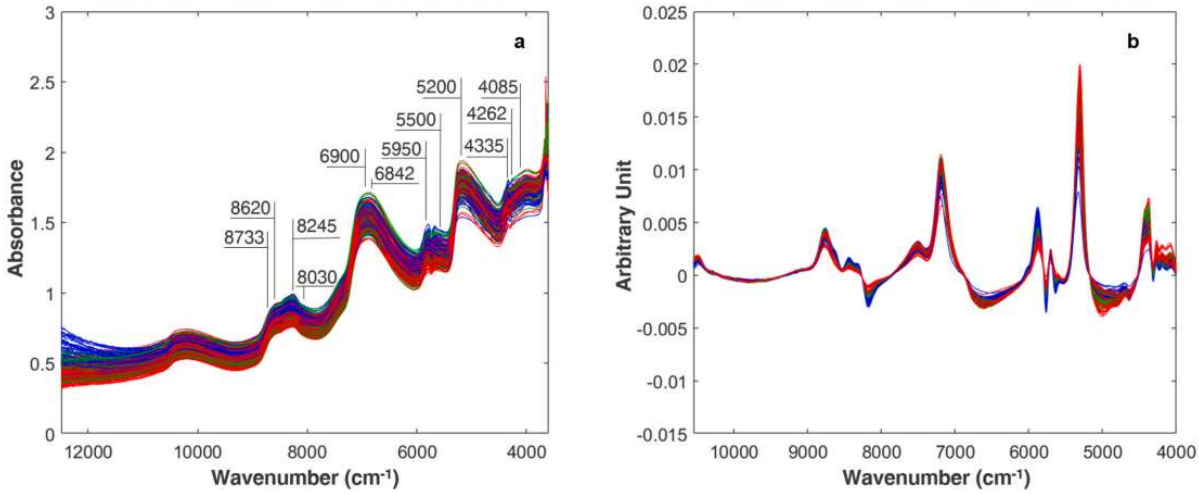


Figure 3. FT-NIR spectra of olives at different ripening stages, identified according to the image analysis maturity class (class 1, green line; class 2, red line; class 3, blue): a) raw spectra; b) reduced spectra after smoothing and first derivative transformation. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article).

Due to its scarce relevance, the region between 12,500 and 10,500 cm− 1 was removed before further analyses and the spectra were pretreated with smoothing (Savitzky–Golay, zero polynomial order, 15 points) and first derivative (Savitzky–Golay, second polynomial order, 15 points) to enhance differences (Fig. 3b). The pre-treated spectra collected from olive samples of the IA class 3 better highlighted the changes occurring with the olive ripening. In particular, the higher absorbance in the region 8600–8330 cm− 1 and at 5800 cm− 1 and the lower intensity in the 8350–8000 and 5850–5580 cm− 1 ranges of the completely ripe drupes (class 3) are linked to the higher oil content with respect to the less ripe fruits (Hernandez-Sanchez & Gomez-del-Campo, 2018).

Qualitative inspection of spectral data

The observed differences in spectra of olive samples were investigated by performing a PCA. The best differentiation among IA classes was observed for the spectra transformed by smoothing and first derivative. Sample distribution along the first two PCs (around 80% of the explained variance) was marked according to the maturity stage (Fig. 4a) and origin (Fig. 4b). The green skin drupes (IA class 1) generally had positive PC1 and PC2 values, whereas IA class 3 samples were characterised by negative PC1 scores (Fig. 4a). Accordingly, the samples with an intermediate ripening stage (IA class 2) were distributed in-between the other two classes. By the loading inspection it was possible to associate samples with positive PC1

scores, i.e. drupes with green skin colour, to higher absorbance at 6900 cm− 1 (first overtone of –OH bond), 5850–5580 cm− 1 (OH stretch, –CO stretch), and 4500–4000 cm− 1 (Fig. 4c). Sample distinction based on olive origin (Fig. 4b) was not clear, remarking that the main differences among the olives were related to the maturation progress, rather than the growing area or the cultivar. Similarly, Fernandez-Espinosa ´ (2016) reported difficulties in discriminating olive varieties using AOTF-NIR spectra. The data exploration did the groundwork for the development of a classification model able to discriminate olives according to the maturity class identified by image analysis.

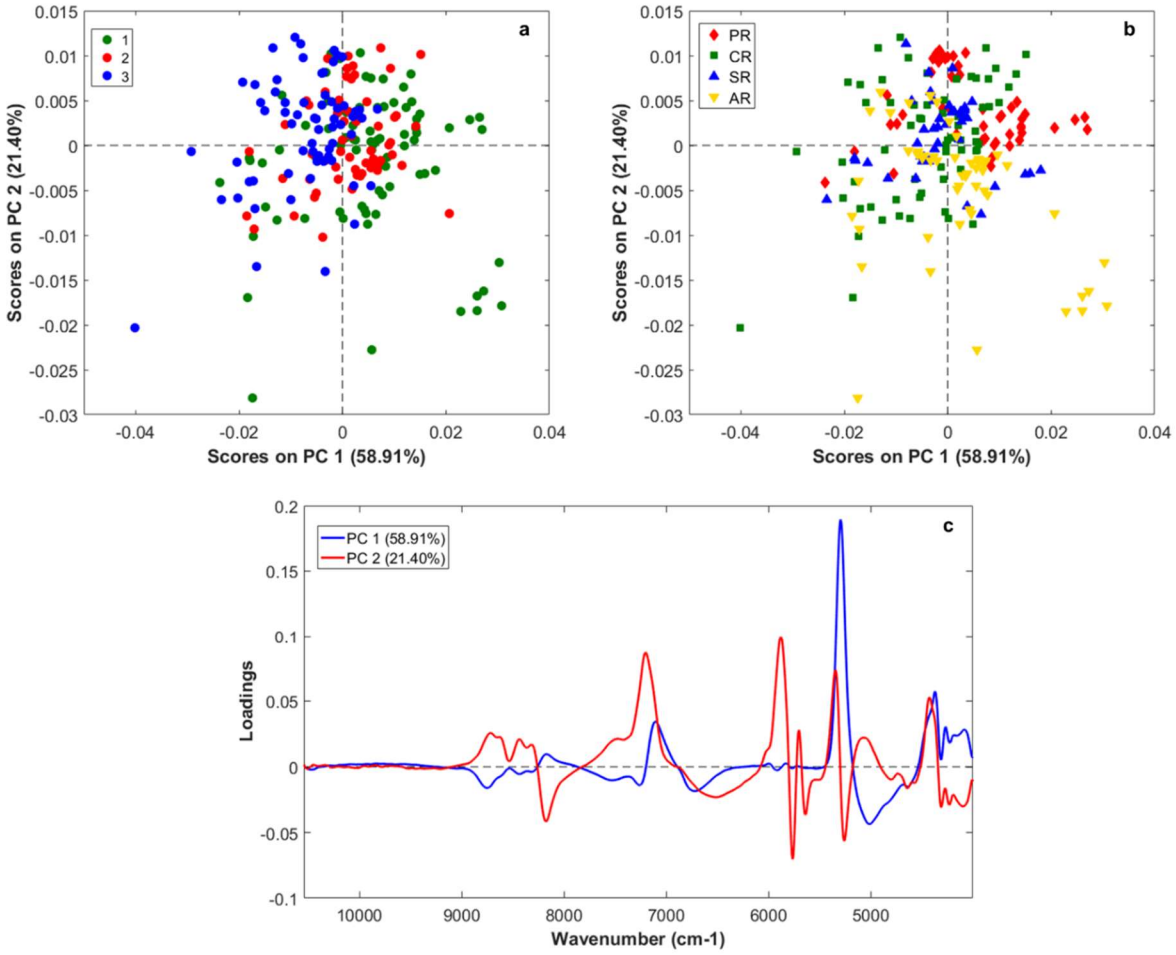


Figure 4. Principal Component Analysis results of the FT-NIR spectra of olives at different ripening stages after smoothing and first derivative: a) score plot with sample identification according to the image analysis maturity classes (dark gray, class 1; light gray, class 2; black, class 3); b) score plot with sample identification according to the origin (reversed triangle, Abruzzo region; square, Calabria region; diamond, Apulia region; triangle, Sardinia region); c) loading plot (black line, PC1; gray line, PC2).

FT-NIR spectroscopy for the assessment of olive maturity class

Olive classification models based on IA maturity classes were developed applying the PLS-DA algorithm, based on the optimisation of covariance between X and Y, substituting the Y with a dummy variable having three levels, i.e. the number of IA classes. After the regression procedure, the y predicted values for each sample were translated into class membership (classification procedure). The class assignation was then performed based on probability, in a way that each sample was assigned to the class that had the highest probability regardless of the magnitude of the probability (Ballabio & Consonni, 2013). The performance of the global model (based on data of all the olive samples) and the origin models (one model for each region of provenance of the olives) was evaluated in terms of SENS and SPEC in calibration, cross-validation, and prediction (Table 4). Different spectral pre-treatments were evaluated for the classification model development; the best performance was achieved when spectra were transformed by smoothing and first derivative. SENS and SPEC average values in prediction for the global model were 79% and 75%, respectively, being mainly affected by the high SENS value of classes 1 and 3 (>85%) and the low SENS and SPEC values of class 2 (64% and 72%, respectively). This means that samples with MI values from 0.80 to 1.55 (the IQR for class 1) were correctly classified with a SENS of 85%, similarly to samples with MI values from 3.40 to 4.55 (the IQR for class 3). The higher misclassification of the intermediate class was expected as the maturation steps are not circumscribable, but they are the result of continuous modifications along the biological process. The misclassification is also justified by the high variability of samples present in class 2. Indeed, the IQR of class 2 includes samples with a larger range of MI (from 1.65 to 3.00), and 10.7% of the samples included in this class (9 samples out of 84) has MI values outside the IQR range. Due to the limited number of papers focused on the prediction of olive maturity stages by NIR spectroscopy, a comparison with previously published data is difficult, also because of the different chemometric models and performance figures used. For instance, the work by Cayuela and Camino (2010) assessed the olive MI using Vis-NIR data, but a PLS regression model was applied. They obtained a root mean square error of prediction (RMSEP) of 0.51 with a 1–6 MI range, i.e. a RMSEP% of 15.7, that was considered a good performance being the reference method susceptible of a large error. Even though the PCA demonstrated that there was not a clear distribution of samples based on olive origin, PLS-DA models for each region of provenance were in any case developed to assess if a better performance could be achieved. This was the case of AR, CR,

and PR models, being SENS and SPEC in prediction higher than 82% and 81%, respectively. On the contrary, the SR model reached values of SENS and SPEC in prediction of 75% and 69%, respectively, lower than the figures of merit of the global model. The lower performance of the SR model could be ascribed to the lowest number of samples used to calibrate (43) and validate the model (20), which could lead to a model less stable, also considering the lower range of MI variability of the Semidana olives belonging to this dataset. Notwithstanding the slightly worse performance of the global model with respect to most of the origin-based ones, it should be considered that it is definitely more robust being constructed on a dataset of 303 samples, i.e. aliquots of olives belonging to thirteen different cultivars harvested over three years. Thus, the global model could be used on a larger scale, no matter the olive cultivar or origin.

Table 3. Figures of merit of PLS-DA models developed for olive maturity class prediction based on FT-NIR spectral data after smoothing (Savitzky–Golay, zero polynomial order, 15 points) and first derivative (Savitzky–Golay, second polynomial order, 15 points) of all the olives (Global) or of the olives coming from Abruzzo Region (AR), Calabria Region (CA), Apulia Region (PR) and Sardinia Region (SR): number of latent variables (LV), number of samples used to calibrate the model (N-Cal) and to validate the model (N-Pred), calibration, cross-validation (CV) and prediction results in terms of sensitivity (SENS) and specificity (SPEC) percentages.

Dataset		Class 1		Class 2		Class 3		Average prediction	
		SENS	SPEC	SENS	SPEC	SENS	SPEC	SENS	SPEC
Global (7 LV)	**N-Cal**	*78*		*61*		*67*		*206*	
	Calibration	86	84	66	61	92	87	82	78
	CV	83	83	66	57	90	85	80	76
	N-Pred	*41*		*28*		*28*		*97*	
	Prediction	85	67	64	72	86	84	**79**	**75**
AR (5 LV)	**N-Cal**	*12*		*22*		*19*		*53*	
	Calibration	100	100	86	77	95	97	92	90
	CV	100	100	77	77	95	85	89	85
	N-Pred	*8*		*9*		*7*		*24*	
	Prediction	100	94	67	80	100	94	**87**	**89**
CR (5 LV)	**N-Cal**	*27*		*17*		*20*		*64*	
	Calibration	100	100	100	96	100	100	100	99
	CV	89	89	71	85	95	100	86	91
	N-Pred	*10*		*6*		*9*		*25*	
	Prediction	100	60	67	95	78	100	**84**	**83**
PR (4 LV)	**N-Cal**	*23*		*14*		*10*		*47*	
	Calibration	74	8	79	64	80	76	77	76
	CV	74	75	64	64	80	76	72	72
	N-Pred	*16*		*7*		*5*		*28*	
	Prediction	81	92	86	62	80	74	**82**	**81**
SR (4 LV)	**N-Cal**	*16*		*11*		*18*		*45*	
	Calibration	94	100	75	59	100	79	93	83
	CV	94	100	62	56	100	79	90	83
	N-Pred	*7*		*6*		*7*		*20*	
	Prediction	100	61	33	50	86	92	**75**	**69**

CONCLUSIONS

In conclusion, the work proposes FT-NIR classification models for the objective evaluation of olive ripening stage based on maturity classes objectively assessed by an image analysis method. The developed methodology is easy, green, and non-destructive and it overcomes the issues of the visual evaluations commonly applied in the olive sector. Moreover, the use of

IA results for the calibration of classification models make them highly reliable, reducing the magnitude of the possible errors. Such a tool can be used for sorting olives directly at the entrance in the mill or even in field – if transferred in a portable device – thus providing olive growers and oil producers with an important decision-making support for the optimisation of virgin and extra virgin olive oil, with a major economic potential for all the olive oil chain. Moreover, it has to be pointed out that the implementation of NIR systems do the groundwork for the prediction of olive and oil quality parameters, thus justifying the use of spectroscopy over the IA used for the calibration step.

Funding sources

This work was supported by AGER 2 Project, grant no. 2016–0105.

CRediT authorship contribution statement

Cristina Alamprese: Conceptualization, Methodology, Supervision, Writing - original draft. Silvia Grassi: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing - original draft, preparation. Alessio Tugnolo: Formal analysis, Data curation. Ernestina Casiraghi: Conceptualization, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(11), 2274–2282.

Allalout, A., Krichene, D., Methenni, K., Taamalli, A., Daoud, D., & Zarrouk, M. (2011). Behaviour of super-intensive Spanish and Greek olive cultivars grown in Northern Tunisia. Journal of Food Biochemistry, 35, 27–43.

Baccouri, B., Zarrouk, W., Krichene, D., Nouairi, I., Youssef, N. B., Daoud, D., & Zarrouk, M. (2007). Influence of fruit ripening and crop yield on chemical properties of virgin olive oils from seven selected oleasters (Olea europea L.). Journal of Agronomy, 6, 388–396.

Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: Linear models. PLS-DA. Analytical Methods, 5, 3790–3798.

Bellincontro, A., Taticchi, A., Servili, M., Esposto, S., Farinelli, D., & Mencarelli, F. (2012). Feasible application of a portable NIR-AOTF tool for on-field prediction of phenolic compounds during the ripening of olives for oil production. Journal of Agricultural and Food Chemistry, 60, 2665–2673.

Camposeo, S., Vivaldi, G. A., & Gattullo, C. E. (2013). Ripening indices and harvesting times of different olive cultivars for continuous harvest. Scientia Horticulturae, 151, 1–10.

Casale, M., & Simonetti, R. (2014). Near infrared spectroscopy for analysing olive oils. Journal of Near Infrared Spectroscopy, 22, 59–80.

Cayuela, J. A., & Camino, M. D. C. P. (2010). Prediction of quality of intact olives by near infrared spectroscopy. European Journal of Lipid Science and Technology, 112(11), 1209–1217.

Correa, E. C., Roger, J. M., Lleo, ´ L., Hernandez-S ´ anchez, ´ N., Barreiro, P., & Diezma, B. (2019). Optimal management of oil content variability in olive mill batches by NIR spectroscopy. Scientific Reports, 9, 1–11.

Fernandez-Espinosa, A. J. (2016). Combining PLS regression with portable NIR spectroscopy to on-line monitor quality parameters in intact olives for determining optimal harvesting time. Talanta, 148, 216–228.

Garcia, J. M., Seller, S., & P´erez-Camino, M. C. (1996). Influence of fruit ripening on olive oil quality. Journal of Agricultural and Food Chemistry, 44, 3516–3520.

Gracia, A., & Leon, ´ L. (2011). Non-destructive assessment of olive fruit ripening by portable near infrared spectroscopy. Grasas Y Aceites, 62(3), 268–274.

Guzman, E., Baeten, V., Fernandez ´ Pierna, J. A., & García-Mes, J. A. (2015). Determination of the olive maturity index of intact fruits using image analysis. Journal of Food Science & Technology, 52, 1462–1470.

Hernandez-Sanchez, N., & Gomez-del-Campo, ´ M. (2018). From NIR spectra to singular wavelengths for the estimation of the oil and water contents in olive fruits. Grasas Y Aceites, 69(4), e278.

Morellò, J.-R., Romero, M.-P., & Motilva, M.-J. (2004). Effect of the maturation process of the olive fruit on the phenolic fraction of drupes and oils from Arbequina, Farga, and Morrut cultivars. Journal of Agricultural and Food Chemistry, 52, 6002–6009.

Nenadis, N., & Tsimidou, M. Z. (2017). Perspective of vibrational spectroscopy analytical methods in on-field/official control of olives and virgin olive oil. European Journal of Lipid Science and Technology, 119, Article 1600148.

Ram, T., Wiesman, Z., Parmet, I., & Edan, Y. (2010). Olive oil content prediction models based on image processing. Biosystems Engineering, 105, 221–232.

Ranalli, A., Tombesi, A., Ferrante, M. L., & De Mattia, G. (1998). Respiratory rate of olive drupes during their ripening cycle and quality of oil extracted. Journal of the Science of Food and Agriculture, 77, 359–367.

Salguero-Chaparro, L., Baeten, V., Fernandez-Pierna, ´ J. A., & Pena-Rodríguez, ˜ F. (2013). Near infrared spectroscopy (NIRS) for on-line determination of quality parameters in intact olives. Food Chemistry, 139, 1121–1126.

Soto, J. N., Martínez, S. S., Gila, D. M., Ortega, J. G., & García, J. G. (2018). Fast and reliable determination of virgin olive oil quality by fruit inspection using computer vision. Sensors, 18, 3826.

Stella, E., Moscetti, R., Haff, R. P., Monarca, D., Cecchini, M., Contini, M., & Massantini, R. (2015). Recent advances in the use of non-destructive near infrared spectroscopy for intact olive fruits. Journal of Near Infrared Spectroscopy, 23(4), 197–208.

Trapani, S., Migliorini, M., Cherubini, C., Cecchi, L., Canuti, V., Fia, G., & Zanoni, B. (2017). Direct quantitative indices for ripening of olive oil fruits to predict harvest time. European Journal of Lipid Science and Technology, 118, 1202–1212.

Uceda, M., & Frías, L. (1975). Epocas de recoleccion. Evolucion del contenido graso del fruto y de la composicion y calidad del aceite. (Seasons of harvest. Changes on fruit oil content, oil composition and oil quality). In Proceeding of II seminario oleícola internacional. Cordoba, Spain: International Olive-oil Council.

Vidal, A. M., Alcala, S., de Torres, A., Moya, M., & Espínola, F. (2019). Characterization of olive oils from superintensive crops with different ripening degree, irrigation management, and

cultivar: Arbequina, Koroneiki, and Arbosana. European Journal of Lipid Science and Technology, 121(4), Article 1800360.

4.2 Portable analytical devices

The availability of handheld spectrophotometers has opened up the possibility of use them in many agricultural food chains for monitoring the fruit ripeness and quality. Handheld instrumentations have been made due to the rapid progress on sensing technologies such as Linear Variable Filters (LVF) or micro-electro-mechanical systems (MEMS) and its integration with micro-optics (MOEMS). The precise dimensions and alignment of MEMS devices, combined with the mechanical stability that comes with miniaturisation, make optical MEMS sensors well suited to a variety of challenging measurements. Moreover, these portable devices have been recently developed with attention to their simplification, by integrating user friendly software for statistical processing and partial automation of analysis, with the aim to fit less skilful users (Pu et al., 2021).

PAPER 5: A diagnostic visible/near infrared tool for a fully automated olive ripeness evaluation in a view of a simplified optical system

A. Tugnolo[a], V. Giovenzana[a]*, R. Beghi[a], S. Grassi[b], C. Alamprese[b], A. Casson[a], E. Casiraghi[b], R. Guidetti[a].

[a]Department of Agricultural and Environmental Sciences (DiSAA), Università degli Studi di Milano, via G. Celoria 2, 20133 Milan, Italy.

[b]Department of Food, Environmental, and Nutritional Sciences (DeFENS), Università degli Studi di Milano, via G. Celoria 2, 20133 Milan, Italy.

* Corresponding author: valentina.giovenzana@unimi.it

Abstract

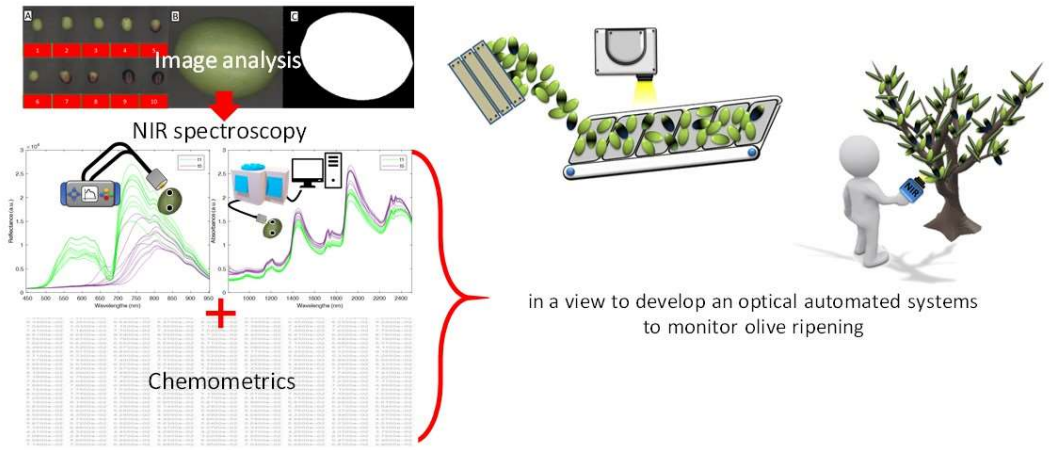
A diagnostic visible/near infrared tool calibrated by means of image analysis,was proposed to evaluate the maturation degree of oil olives in order to replace traditional subjective methods in a view of future fully automated applications.

Thirteen varieties of Olea europaea deriving from four regions of the south of Italy were analyzed. In order to objectify the ripening stage assessment, the RGB image was acquired. Spectroscopic analyses were performed using a benchtop FT-NIR and a portable vis/NIR instrument. The benchtop device was equipped with an optical fiber probe and the spectra were collected over the 800-2500 nm range, nominal resolution of 1.6 nm; the portable spectrophotometer cover the range of 400-1000 nm, nominal resolution 0.3 nm. The olive spectral data were modelled using Partial Least Squares - Discriminant Analysis (PLS-DA). The prediction capability reached by the global model (13 varieties were used) obtained from data acquired with both the devices results promising. The PLS-DA models calculated on the olives from Calabria, Sardinia and Abruzzo revealed high prediction capabilities, i.e. sensitivity, specificity and accuracy higher than 83%. The prediction capability of Apulia samples could be improved increasing the variability of the samples since for this region only 3 sampling times were considered. To compare the modelling performance between the benchtop FT-NIR and the portable vis/NIR device, a McNemar's test was performed resulting no significant difference between the PLS-DA global models. Finally, considering the good performance of the vis/NIR

model, a variable selection using the interval PLS (iPLS) algorithm was applied. To reduce the complexity keeping the performance of the model built using the whole vis/NIR spectra (1647 variables), 12 bands (1.5 nm wide) were selected. The new model showed an improvement in terms of model stability and complexity (Sensitivity 86%; Specificity 87 %; Accuracy 87%) than the two global models built with the whole vis/NIR and NIR range.

The classification performance provided the groundwork for the development of (i) simplified systems for a direct olives ripening determination on the olive tree, and (ii) automated systems to be applied both in field and at the mill for olives sorting according to the ripening degree.

Graphical abstract



Keywords: chemometrics, optical analysis, instrument comparison, wavelenght selection, field

INTRODUCTION

Olea europaea L. is a small evergreen tree which grows between 8 and 15 m tall. It is a slow-growing and extremely long-lived species, with a life expectancy up to 1000 years (Rhizopoulou, 2007). This olive tree originated from the eastern Mediterranean area and it has been cultivated for millennia for its oily fruits. The sensory and nutritional characteristics of these fruits have led to a sharp boost of the demand for the main derivative products (table olives and olive oil) in traditional producing areas and elsewhere in the world.

Olive ripening begins after a period of 25 weeks of cell growth. During this phase, the fruit reaches its final size maintaining the original green color. Therefore, this stage is known as the "green stage" corresponding to green mature fruits. Subsequently, the chlorophyll content decreases, being replaced by anthocyanins. This allows dividing the ripening process into stages according to the exhibited concentration of anthocyanins. Thus, it is possible to identify a "spotted stage", a "purple stage", and a "black stage" according to the skin color of the fruits (Uceda and Frias, 1975).

Olive main constituents are water and oil. During the ripening period, the oil begins to form by synthesis of triglycerides from fatty acids, mainly oleic acid. At a certain stage, the lipids lose their capacity to be synthesized. At this point the olive oil is of the highest quality and consequently, the harvest should begin (Fernández-Espinosa, 2016). These quality features are of great commercial importance as they influence the quality and shelf-life of olive oil. Therefore, monitoring olives ripeness is essential in order to define the most appropriate degree of ripeness to obtain a finished product of high-quality (Trapani et al., 2017).

Several destructive techniques have been used to assess the degree of olives ripeness. Besides being expensive and time-consuming, they are often not sustainable for the environment due to the use of chemical reagents (Casson et al., 2020). For these reasons, these types of analyses are hardly applied by the growers. Therefore, Uceda and Frias (1975) proposed the Maturity Index (MI), a cheap and easy destructive technique for the visual determination of the best harvesting moment. This method, based on the color changes of skin and pulp, classifies 100 olives into eight groups or categories from intense green (category 0) to black with 100% purple flesh (category 7). Although it is the most used method, MI depends

on the operator experience, thus being subjective and strongly affected by human error. Besides, color changes are dramatically different among cultivars (Guzmán et al., 2015).

Computer vision holds considerable potential for the agriculture industry thanks to its simplicity and capability to provide accurate and consistent information. In fact, the image processing has been successfully applied in several applications on different fruits and vegetables, like tomato, mandarin, table olives, and potatoes, for the estimation of color, size, shape, and texture (Ram et al., 2010). However, automatic detection of quality and grading is still difficult due to some existing challenges such as the influence of physical and biological variability (e.g. fruit size and composition), canopy effects, whole surface exploration, environmental conditions (e.g. sun and moon light exposition, relative humidity and temperature), discrimination between defects and stems/calyxes, unobvious defect detection, the robustness of the features and algorithms, as well as rapid optical detection system development. These issues can compromize the fruit or vegetable quality inspection accuracy, reducing the possibility to automize quality inspection and product grading (Zhang et al., 2018).

Among non-destructive techniques, near infrared spectroscopy (NIR) is particularly interesting, being capable to account for the chemical changes always associated to olive ripening, irrespective the color modification. In recent years, the development of new technologies used in the construction of NIR spectrometers has resulted in a significant reduction in size and cost of these devices, also in vision of fully mechanized sorting systems (Salguero-Chaparro et al., 2013; Giovenzana et al., 2018). Several studies have shown that handheld vis/NIR and NIR spectrometers can be suitably used for the determination of quality parameters, providing performance similar to that of the benchtop spectrometers (Jha et al., 2014). Moreover, due to the problems related to the image analysis, the handheld spectrophotometric systems have shown clear advantages, and despite the need of the operator, they remain a good compromise in terms of performance of the analysis (Beghi et al, 2020).

Several works in literature checked the feasibility of individual technologies application, such as image analysis, visible/near infrared (vis/NIR) spectroscopy, hyperspectral imaging (González-Cabrera, et al., 2018), and low resolution nuclear magnetic resonance (Ram et al., 2010) for the analysis of ripening of fruits in general, and olives in particular. It is necessary to undertake a further step that involves the development of complete automated diagnostic tools that are not based on subjective data or subjective classifications (i.e. MI index). This is crucial for the introduction of these tools in fully mechanized or robotic sorting systems, whose study

and application is booming (Zhao et al., 2016). A combined use of image analysis and vis/NIR spectroscopy can be an interesting solution (Gatica et al., 2013). In addition, for a real diffusion in an operative framework, the calculation of predictive models useful for a wide range of cultivars is desirable.

Therefore, the aim of this study was to evaluate a diagnostic visible/near infrared tool, coupled with chemometrics and calibrated by means of image analysis , capable to evaluate the maturation degree of oil olives in order to replace traditional subjective methods in a view of future fully automated applications. In detail, an objective method to classify ripeness degree by using image analysis was estabilished. Afterward, the ripening classes were combined with Fourier-Transformed NIR (FT-NIR) and vis/NIR spectra to develop classification models capable to predict olive ripening degree. A comparison of modelling performance between benchtop FT-NIR spectrophotometer and portable vis/NIR device was also performed. Moreover, a wavelength selection was applied to identify the most informative optical bands in view of a future instrumental simplification.

MATERIALS AND METHODS

Sampling

The experimental activity took place in the south of Italy during the ripening period of one crop season, from September to December. Thirteen non-climacteric varieties of *Olea europaea* coming from Apulia (Bambina, Cima di Melfi, and Oliva Rossa), Calabria (Ciciariello, Ottobratica Calipa, Ottobratica Cannavà, and Tonda di Filogaso), Sardinia (Corsicana, Semidana, and Sivigliana) and Abruzzo (Dritta, Gentile D'Aquila, and Tortiglione) were analyzed. The several olive varieties studied allowed to account for a wide range of agronomic traits, including oil content, fruit size, and veraison process. Samples of each variety were picked in 3 quotas (A, B, C), every two weeks, from the same trees, and delivered (in refrigerated condition) to the Università degli Studi di Milano labs for the analyses.

Two portions of about 100 g of olives were weighed from each quota in order to obtain representative portion of olives (A1, A2, B1, B2, C1, C2) to perform the analyses (Figure 1).

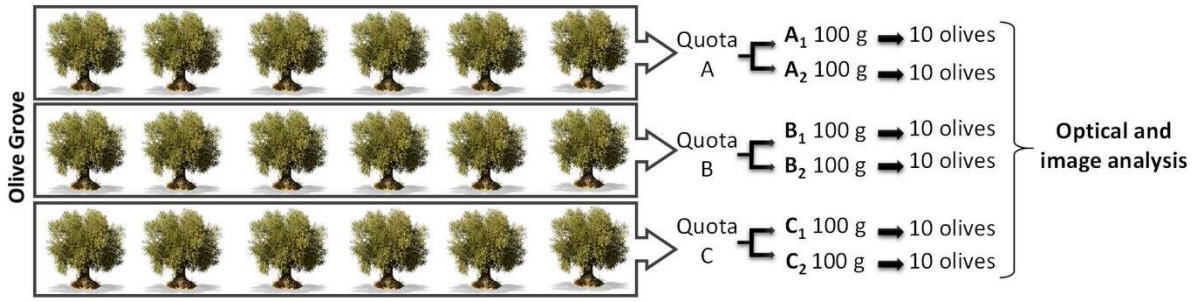


Figure 1. Sampling plan for each variety.

Each portion of olives (A1, A2, B1, B2, C1, C2) was analyzed according to the Surface Colorimetric Index (SCI), which sorts the olives into 4 different ripeness classes according only to the skin color (Figure 2):

- Class 1: totally green;
- Class 2: olives with less than 50% of purple/black surface;
- Class 3: olives with more than 50% of purple/black surface;
- Class 4: totally purple/black.

Figure 2. Olives used as a standard for classification in each of the four classes (Class 1: totally green; Class 2: olives with less than 50% of purple/black surface; Class 3: olives with more than 50% of purple/black surface; Class 4: totally purple/black).

This method is similar to MI (Uceda and Frias, 1975), but it involves a visual evaluation of the skin color only, without considering the flesh and it was applied in a perspective of simplifying the MI procedure. The olives were evaluated always by the same trained observer, under constant laboratory illumination conditions, on a white background. The SCI calculation is based on the same equation of MI (Guzmán et al., 2013):

$$SCI = \frac{1}{N}\sum_{i=1}^{4}(i \times n_i)$$

where i is the group number, ni is the number of olives in the group and N is the total number of olives considered.

From each portion (A1, A2, B1, B2, C1, C2), ten olives were selected according to the SCI classes' frequency. Therefore, for each variety of each sampling time, 60 olives were used to perform the optical and image analyses, for a total of 3180 olives.

Image acquisition and analysis

In order to objectify the ripening stage assessment, the RGB image of the ten olives of each portion was acquired using a flatbed scanner HP Scanjet 8300 (HP Inc., Palo Alto, CA, USA), covered with a black box to prevent light losses, and managed by VueScan software (v. 9.4, 2016, Hamrick Software, Sunny Isles Beach, FL, USA). A resolution of 600 DPI and a color depth of 24 bits were applied for the image acquisition.

After digitalization, images were analyzed using MATLAB® (v. R2017b, The MathWorks, Inc., Natick, MA, USA). A script was set up in order to frame a Region of Interest (ROI) which includes the olive keeping out the background. The red channel has been selected as the most powerful channel in distinguishing the olives from the background (using a cut-off of 0.1). Then the ROI was converted into a vectorial image (HSV). The intensity values (0-255) of the pixels belonging to each olive were averaged in order to obtain a mean value representative of the whole olive. The results obtained were used to build the objective method for olive ripeness assessment, the Image Analysis Classification (IAC).

Spectral acquisitions

Optical analyses were performed using a benchtop FT-NIR (MPA, Bruker Optics, Milan, Italy) and a portable vis/NIR (Jaz Modular Optical Sensing Suite, OceanOptics, Inc., Dunedin, FL, USA) spectrophotometer. The benchtop device was equipped with an optical fiber probe and the spectra were collected over the 12500-4000 cm^{-1} (800-2500 nm) range, with a resolution of 8 cm^{-1} and 32 scans for both sample and Spectralon background. The collection of each spectrum lasted 30 s. Instrument control was managed by using the OPUS software (v. 6.0 Bruker Optics, Milan, Italy). The portable vis/NIR spectrophotometer (400 -1000 nm) works using a bifurcated fiber that conveys the light from the halogen lamp to the sample and back

to the detector. The tip of the fiber was equipped with a cap that standardized the analysis distance (about 2 mm) and reduced the environmental light interference. The background was realized using a white (100% of light reflected) and a black (0% of light reflected) standard. Besides, an integration time of 50 ms was set in order to collect the best spectral dynamics using a light intensity of 1095 lumen. For both instruments, two spectra were acquired in reflectance at opposite sides of each olive. These spectra were subsequently averaged to obtain a mean spectrum that is representative of the whole olive.

Data processing

Regarding the image analysis, two thresholds were identified to separate the olives according to the degree of ripeness. To the aim, the whole dataset composed by the average values of the red channel of each olive was randomly divided into a calibration (70% of the data) and a validation set (30% of the data) The calibration set was used to assess the two thresholds, then, the training set was tested for the capability of the thresholds in discriminating the three maturation classes: 100% of the skin color green, skin color partially purple/black, 100% of the skin color purple/black olives.

Spectral data were explored by Principal Component Analysis (data not shown). This method was applied to extract the useful information from the spectra acquired from each olive (using FT-NIR and the portable vis/NIR). This method is recommended as the initial step of any multivariate analysis to give a first look at the data structure and for other aims, including the outlier detection (Wold, Esbensen and Geladi, 1987). The spectra collected with each instrument were organized in different datasets, one for each of the four Italian regions and one including all the olive samples. Before data analysis, spectra were pre-treated as follows:

- Benchtop FT-NIR: smoothing (Savitzky-Golay method, filter width 15 points; polynomial order 1), Standard Normal Variate (SNV);
- Portable vis/NIR: smoothing (Savitzky-Golay method, filter width 19 points; polynomial order 3), SNV.

The olive spectral data were modeled using Partial Least Squares - Discriminant Analysis (PLS-DA). Its objective is to allow the maximum separation among classes of objects. PLS-DA is based on the PLS regression algorithm. In detail, it is a PLS 2 (a variant of PLS algorithm that can handle multi responses) applied, in this case study,on a dummy matrix representing the membership at one of the four ripening classes of the IAC index. PLS-DA transforms the

class vector (containing the membership of samples at the *G* classes) into a dummy matrix *Y*, with n rows (samples) and *G* columns (the class information). Each entry y_{ig} of *Y* represents the membership of the *i*-th sample at the *g*-th class expressed with binary code (1 or 0). Therefore, the *n*-dimensional class vector is transformed into a binary *Y* matrix constituted by *n* rows and *G* columns. Afterward, PLS 2 algorithm is applied on *Y* matrix. The estimated class values are not either 1 or 0 perfectly, thus the most probable prediction was used to identify the membership at a specific class. This technique chooses the class that has the highest probability regardless of the magnitude of that probability (Ballabio and Consonni, 2013). The model calibration was performed using 66.6% of the samples of each dataset. The external validation used the remaining 33.3% of the samples, systematically extracted from every dataset by a Venetian blinds approach.

The model prediction performance was assessed using the following indicators:

1) *Sensitivity (%)* = $\frac{TP}{TP+FN} * 100$
2) *Specificity (%)* = $\frac{TN}{FP+TN} * 100$
3) *Classification accuracy (%)* = $\frac{TP+T}{TP+FP+FN+TN}$*100

The values of true positive (TP), true negative (TN), false negative (FN), and false positive (FP) were calculated from the confusion matrix. TP and TN values correspond, respectively, to the correctly classified olives that belong or not to a specific class, whereas FN and FP are the olives erroneously classified in a specific class ().

The ‘testcholdout’ function implemented in MATLAB® was used to statistically compare the predictive accuracies of the PLS-DA models obtained from both benchtop and portable device data. This function performs a one-tailed, mid-P-value McNemar test, which is a particular case of Fisher’s sign test. It compares the predicted labels against the true labels, and then it detects whether the difference between the misclassification rates is statistically significant. Thus, the test reveals if the models obtained from the two instruments have the same performance based on P-value. P-values higher than 0.05 confirm the acceptance of the null hypothesis, i.e. that the two models are not significantly different (Fagerland, Lydersen, and Laake, 2013; Giraudo et.al, 2019).

Therefore, the performances of the models were compared by the calculation of McNemar's value (χ2) on the prediction results, according to Eq. (1).

$$(1)\ x^2 = \frac{(n_{12}-n_{21})^2}{n_{12}+n_{21}}$$

where n_{12} represents the number of samples misclassified (both FP and FN) by the FT-NIR model and n_{21} stands for the number of samples misclassified (both FP and FN) by the vis/NIR model.

Finally, a variable selection was proposed based on vis/NIR data, considering the future objective of building an easy to use and more cost-effective device for in-field application. In this context, the interval PLS (iPLS) algorithm was applied. iPLS builds local PLS-DA models after the division of the spectra into equal non-overlapping intervals "m" with a specific number of variables "n". The main intervals (m = 12 and n = 5 in this study) capable to build a PLS-DA model with the lowest Root Mean Square Error in Cross-Validation (RMSECV) were selected(Nørgaard et al., 2000; Miaw et al.,2018). The prediction performance of the new model with few wavelengths was compared to the model including all the wavelengths in the Vis/NIR region.

All data processing was carried out in MATLAB® environment (v. 2019b, Mathworks, Inc., Natick, MA, USA), applying algorithms from the PLS Toolbox, version 8.8 (Eigenvector Research, Inc).

RESULTS AND DISCUSSIONS

Image analysis

Figure 3 shows the three main steps of the developed image analysis procedure. As stated, 10 olives were selected from each portion (Figure 3A) based on the skin color classification (SCI value). Each olive was processed individually defining the ROI (Figure 3B, 3C). The red channel information of all the pixel constituting the ROI were averaged and a three-class segmentation was proposed. The reduction of the maturity classes to three, instead of the four classes defined by SCI, was proposed as the image information originated from one olive side and not from the global drupe surface. Thus, the SCI classes 2 (olives with less than 50% of purple/black surface) and 3 (olives with more than 50% of purple/black surface) were merged

in a single class containing olives characterised by red channel values referred to skin color partially purple/black (Table 1).

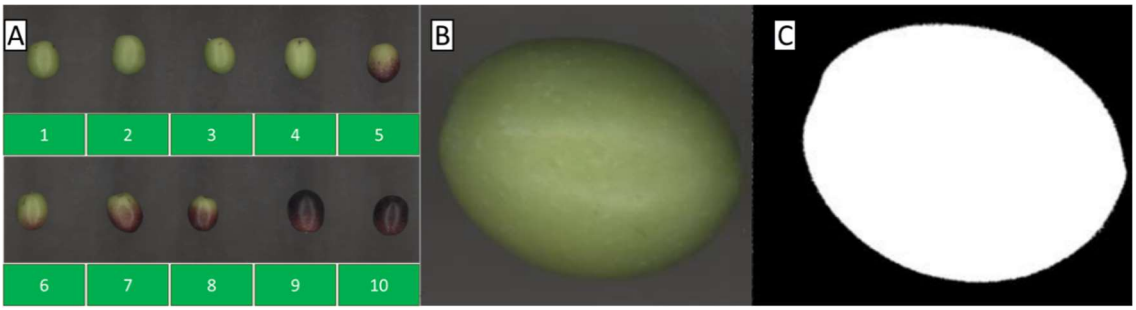


Figure 3. Image analysis process. (A) Image acquisition; (B) olive individualization; (C) background and olive isolation.

Table 1. Olive maturity classes according to red channel intensity calculated from image analysis.

IAC Class Number	Description	Red channel intensity
1	100% of the skin color green	255-120
2	skin color partially purple/black	119.9-60
3	100% of the skin color purple/black	59.9-0

Spectral analysis

Figure 4 shows the vis/NIR diffuse reflectance spectra (A, B) and FT-NIR absorbance spectra (C, D) obtained from Filogaso (a variety subject to veraison) and Semidana olives (a variety that is not subjected to veraison). In green were represented the spectra at the beginning of the crop season (t1) and in purple the spectra at the end of the ripening process (t5/t4). The main absorption peaks in the visible region are around 540 and 680 nm due to external color differences in the samples which are related to changes in the amount of pigment (anthocyanins) causing a decrease in reflectance in the visible band associated with the absorption peak at 540 nm. A high reflectance absorption could be noticed around the chlorophyll absorption peak at 680 nm (Abu-Khalaf and Hmidat, 2020). Instead, in the NIR region, the relevant peaks, located at 978, 1454 and 1930 nm, can be related to water absorption bands, since the moisture content in the olive fruits is higher than 60%. In particular, the bands at 978, 1454 and 1930 nm are related to the OH second overtone, the OH stretch first overtone and the harmonic and combination bands of OH bonds in hydroxyl groups, respectively. The absorbance at 1145 and 1160 nm is related to the stretching of the CO bonds of aliphatic esters, while the second overtone of CH stretching vibrations of alkyl groups and

alkenes occurs at 1212 and 1245 nm (Fernández-Espinosa, 2016). In the region between 1500 and 1900 nm, the spectra are characterized by CH_3-CH stretch (1680 nm) and OH, -CO stretch (1820 nm) (Fernández-Espinosa, 2016; Trapani *et al.*, 2017). Besides, the relevant differences in absorbance intensity within 2200 and 2500 nm are linked to the combination of CH stretching with other vibrational modes (Casale and Simonetti, 2014).

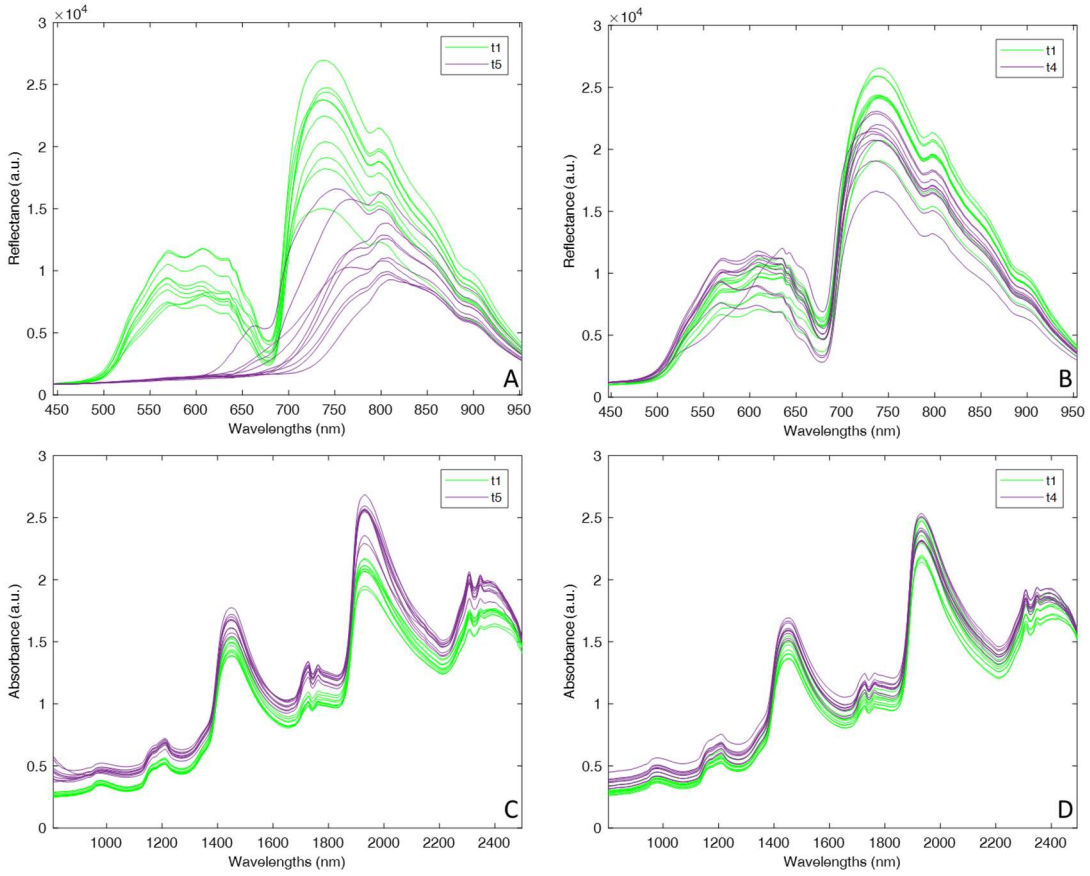


Figure 5. Vis/NIR diffuse reflectance spectra (A, B) and FT-NIR absorbance spectra (C, D) obtained from Filogaso olives, a variety subject to veraison (A, C), and Semidana olives, a variety that is not subjected to veraison (B, D). In green (t1) the spectra at the beginning of the maturation process are reported and in purple (t5/t4) the spectra obtained at the end of the ripening process.

Overall, strong evidence of the maturation process is highlighted in Figures 4A and 4C (Filogaso variety). This variety is subjected to veraison and shows a higher absorbance for both instruments at the end of the crop season. Figures 4B and 4D show the spectra of the maturation process of the variety Semidana. These olives are not subjected to veraison, thus a lower difference of absorbance is showed between 500 and 700 nm. Anyway, going towards

the pure NIR region, a higher absorbance at the final stage of the maturation process (t4) is highlighted for both devices.

Then, the PCA was applied to find two indices of distance (Hotelling's T2 and squared prediction error) useful for the outliers detection (Mahmoud, Lotfi and Langensiepen, 2016). Data, from each region, were explored and, excluding outliers, the final amount of samples correctly acquired was 3033. In table 2 were reported the number of samples for each IAC class.

Table 2. Overview of the number of olive samples for each IAC class which were analyzed and used to build (66.6%) and validate (33.3%) the PLS-DA models.

Region	N° of varieties	N° of samples in calibration			N° of samples in prediction		
		Class 1	*Class 2*	*Class 3*	*Class 1*	*Class 2*	*Class 3*
Apulia	*3*	86	206	62	41	103	32
Calabria	*4*	355	131	234	172	72	116
Sardinia	*3*	157	146	171	72	79	85
Abruzzo	*3*	172	189	115	90	94	53
Total		3033					

Models results

The spectral data were used for the elaboration of PLS-DA classification models capable to classify olives according to the IAC. Table 3 shows the results obtained from the global model (considering all the olive varieties) and from the dataset of each considered region (Apulia, Calabria, Sardinia and Abruzzo). For each PLS-DA model, the performance was expressed in terms of weighted average in prediction, in terms of sensitivity, specificity and classification accuracy.

Table 3. Sensitivity (Sen.), specificity (Spec.), classification accuracy (Acc.), and the number of latent variables (LVs) for PLS-DA models derived from data of the benchtop FT- NIR and portable vis/NIR spectrophotometers. The average results in prediction are reported.

Models		Average prediction						
		Benchtop FT- NIR			Portable vis/NIR			
	N° of samples	*LVs*	*Sen. (%)*	*Spec. (%)*	*Acc. (%)*	*LVs*	*Sen. (%)*	*Spec. (%)*	*Acc. (%)*
Global	1011	14	81	82	81	4	85	87	86
Apulia	176	14	74	81	78	5	74	76	75
Calabria	360	14	86	88	87	8	91	93	92
Sardinia	236	14	83	84	83	4	86	90	88
Abruzzo	237	14	88	83	86	4	87	88	88

Overall, the results from both devices are satisfactory, considering also a large number of varieties analyzed (13). Each PLS-DA model shows good performance predicting each sample in its class. The olives of the Apulia region show the lower classification accuracy (78% and 75% for the benchtop FT- NIR and the portable Vis/NIR device, respectively). This result could probably be improved increasing the variability of the samples since for this region only 3 sampling times were considered. Instead, the models built with the olives from Calabria, Sardinia and Abruzzo revealed high prediction capabilities, i.e. sensitivity, specificity and accuracy higher than 83%. The prediction capability reached by the global model (13 varieties were used) obtained from data acquired with both the devices results promising. Indeed, the global models are the more prompt to robust prediction of the olive ripening stage, as they do consider a high number of olive varieties with different physiological and maturation characteristics.

To objectively explore if there is a significant difference between the PLS-DA models obtained with the benchtop FT-NIR and the portable vis/NIR data, a McNemar's test was performed on the prediction results of the global models. The analysis was conducted to test if the model built with the vis/NIR region data through a cost-effective device is significantly different from the model built with data in the NIR region collected with a more expensive benchtop spectrophotometer. The McNemar's test gave a p-value equal to 1, thus the two global models resulted comparable. Being the vis/NIR instrument a portable and cheap instrument its use is advisable to substitute the benchtop FT-NIR analysis.

Finally, considering the good performance of the vis/NIR model, a variable selection using the iPLS algorithm was proposed. To reduce the complexity, keeping the performance of the model built using the whole vis/NIR spectra (1647 variables), 12 bands were selected (figure 5). The number of the selected optical bands was envisaged on the basis of Authors'previous reserch (Beghi et al. 2020). Each band is composed by 5 variables, therefore, considering the portable vis/NIR resolution equal to 0.3 nm, the interval of interest is 1.5 nm wide. In a view of building a simplified optical device, the proposed selection represents a good compromise among commercial availability of light sources, constructive simplicity and cost-effective. Overall, the larger part of the usefull wavelengths comes from the pure visible region (from 500 to 660 nm) where the contribution of olive color, due to molecules like

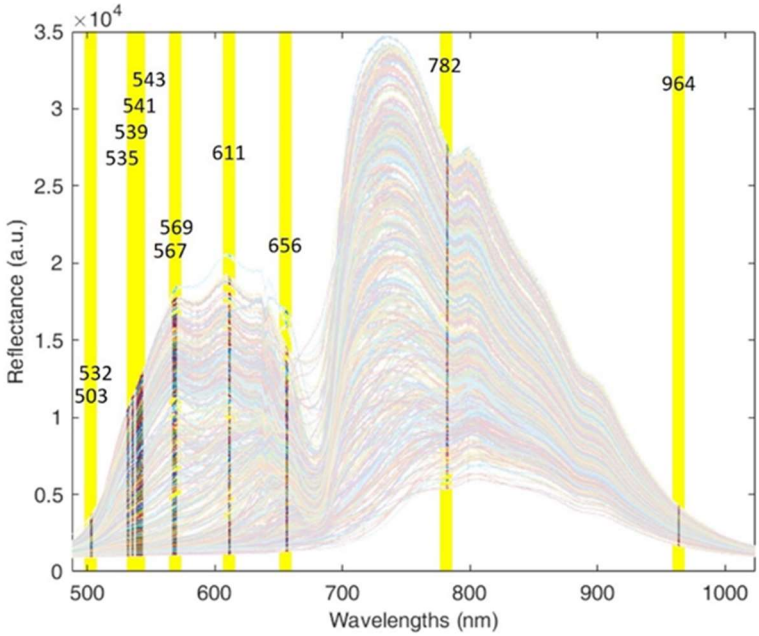


Figure 6. Variable selection performed on the vis/NIR portable device dataset using the iPLS algorithm.

chlorophylls, carotenoids, anthocyanins and polyphenols, has a great weight being the expression of the relevant changes occurring during ripening. Moreover, two interesting bands were selected in the NIR region (about 780 and 965 nm) where the -OH and -CH functional groups absorb, thus, describing the chemical modification related to water reduction and lipid synthesis during ripening.

Using the variables selected through iPLS, a new PLS-DA global model was calculated. Table 4 shows the obtained results, comparing them with the PLS-DA models obtained using the

complete FT-NIR and vis/NIR spectra. iPLS only slightly improved the performance in prediction (sensitivity 86%; specificity 87%; accuracy 87%) compared to those obtained using the full spectra. However, the sharp reduction of the wavelengths boosts the stability and reduce the complexity of the model even if 7 LVs (few variables model) were used instead of 4 LVs (full-range model). The results of this approach agree with the results obtained by Miaw et al. (2018) where three variable selection methods (iPLS, Variable Importance in Projection "VIP" scores and a Genetic Algorithm "GA") were tested to improve the effectiveness of the classification models (SIMCA and PLS-DA).

Table 4. Sensitivity (Sen.), specificity (Spec.), classification accuracy (Acc.), and the number of latent variables (LVs) for PLS-DA models calculated with the data of the benchtop FT- NIR and portable vis/NIR devices. The average results in prediction are reported.

				Average prediction		
Model	*N°of variables*	*range*	*LVs*	*Sen. (%)*	*Spec. (%)*	*Acc. (%)*
	1816	NIR	14	81	82	81
Global	1647	Vis/NIR	4	85	87	86
	60 (12 bands * 5 points wide)	Vis/NIR	7	86	87	87

CONCLUSIONS

In this work, a fully automated diagnostic visible/near infrared tool was proposed to evaluate the maturation degree of oil olives in order to replace traditional subjective methods. Through the analysis of the sample images, it was possible to objectify the olive classification on ripeness basis. Besides, it has been shown that starting from ripeness classes created by this analysis, it is possible to use FT-NIR and Vis/NIR techniques to automate olive ripening classification, in a view of future fully automated application. In fact, the global classification models developed for both instruments reached high classification accuracy (above 81%) and comparable (p=1.00) prediction capabilities. Thus, the portable vis/NIR device showed the potential to be used for in-field applications, overcoming the costs and management problems of a benchtop instrument. To further simplify the hardware (i.e., the instrument) and the software (i.e., the computation of the classification model), a variable selection strategy was proposed by the application of the iPLS algorithm on the vis/NIR range. The new model has shown an improvement in terms of model stability and complexity (Sensitivity 86%; Specificity 87 %; Accuracy 87%). This very good performance provides the groundwork for the

development of (i) simplified systems available to small olive growers for the rapid, non-destructive and direct determination of the degree of maturity of the olives from the olive tree, and (ii) automated systems to be applied both in field and at the mill for olives sorting according to the ripening degree. To achieve these objectives and reach the step of a real scale application of the proposed vis/NIR technology for the automatic detection of olive quality, some challenges have to be faced. Some issues such as environmental conditions due to canopy effects, sunlight interferences, effects related to temperature and humidity have to be overcome. Therefore further prototyping work is needed to propose an upgraded version of the device, including a compensatory system for stray light and temperature.

Acknowledgements

This work was supported by AGER 2 Project, grant n° 2016-0105. The authors are also grateful to Marco Menegon for the technical support and the data collection.

References

Abu-Khalaf, N., & Hmidat, M. (2020). Visible/Near Infrared (VIS/NIR) spectroscopy as an optical sensor for evaluating olive oil quality. Computers and Electronics in Agriculture, 173, 105445.

Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: linear models. PLS-DA. Analytical Methods, 5(16), 3790-3798.

Beghi, R., Giovenzana, V., Tugnolo, A. Casson, A., & Guidetti, R. (2020). Design of prototypes of LED based devices for the evaluation of grape (Vitis Vinifera L.) ripeness. Acta Horticlulturae (Accepted 06/2020).

Casale, M., & Simonetti, R. (2014). Near infrared spectroscopy for analysing olive oils. Journal of Near Infrared Spectroscopy, 22(2), 59-80.

Casson, A., Beghi, R., Giovenzana, V., Fiorindo, I., Tugnolo, A., & Guidetti, R. (2020). Environmental advantages of visible and near infrared spectroscopy for the prediction of intact olive ripeness. Biosystems Engineering, 189, 1-10.

Fagerland, M. W., Lydersen, S., & Laake, P. (2013). The McNemar test for binary matched-pairs data: mid-p and asymptotic are better than exact conditional. BMC

medical research methodology, 13(1), 91.

Giovenzana, V., Beghi, R., Romaniello, R., Tamborrino, A., Guidetti, R., & Leone, A. (2018). Use of visible and near infrared spectroscopy with a view to on-line evaluation of oil content during olive processing. Biosystems Engineering, 172, 102-109.

Guzmán, E., Baeten, V., Pierna, J. A. F., & García-Mesa, J. A. (2015). Determination of the olive maturity index of intact fruits using image analysis. Journal of Food Science and Technology, 52(3), 1462-1470.

Fernández-Espinosa, A. J. (2016). Combining PLS regression with portable NIR spectroscopy to on-line monitor quality parameters in intact olives for determining optimal harvesting time. Talanta, 148, 216-228.

Gatica, G., Best, S., Ceroni, J., & Lefranc, G. (2013). Olive fruits recognition using neural networks. Procedia Computer Science, 17, 412-419.

Giraudo, A., Grassi, S., Savorani, F., Gavoci, G., Casiraghi, E., & Geobaldo, F. (2019). Determination of the geographical origin of green coffee beans using NIR spectroscopy and multivariate data analysis. Food control, 99, 137-145.

González-Cabrera, M., Domínguez-Vidal, A., & Ayora-Cañada, M. J. (2018). Hyperspectral FTIR imaging of olive fruit for understanding ripening processes. Postharvest Biology and Technology, 145, 74-82.

Jha, S. N., Narsaiah, K., Jaiswal, P., Bhardwaj, R., Gupta, M., Kumar, R., & Sharma, R. (2014). Nondestructive prediction of maturity of mango using near infrared spectroscopy. Journal of Food Engineering, 124, 152-157.

Mahmoud, S., Lotfi, A., & Langensiepen, C. (2016). User activities outliers detection; integration of statistical and computational intelligence techniques. Computational Intelligence, 32(1), 49-71.

Miaw, C. S. W., Sena, M. M., de Souza, S. V. C., Ruisanchez, I., & Callao, M. P. (2018). Variable selection for multivariate classification aiming to detect individual adulterants and their blends in grape nectars. Talanta, 190, 55-61.

Nørgaard, L., Saudland, A., Wagner, J., Nielsen, J. P., Munck, L., & Engelsen, S. B. (2000). Interval partial least-squares regression (i PLS): a comparative chemometric study with an example from near-infrared spectroscopy. Applied Spectroscopy, 54(3), 413-419.

Ram, T., Wiesman, Z., Parmet, I., & Edan, Y. (2010). Olive oil content prediction models based on image processing. Biosystems engineering, 105(2), 221-232.

Rhizopoulou, S. (2007). Olea europaea L. A botanical contribution to culture. American-Eurasian Journal of Agricultural & Environmental Sciences, 2(4), 382-387.

Salguero-Chaparro, L., Palagos, B., Peña-Rodríguez, F., & Roger, J. M. (2013). Calibration transfer of intact olive NIR spectra between a pre-dispersive instrument and a portable spectrometer. Computers and electronics in agriculture, 96, 202-208.

Trapani, S., Migliorini, M., Cecchi, L., Giovenzana, V., Beghi, R., Canuti, V., ... & Zanoni, B. (2017). Feasibility of filter-based NIR spectroscopy for the routine measurement of olive oil fruit ripening indices. European Journal of Lipid Science and Technology, 119(6), 1600239.

Uceda, M., & Frías, L. (1975). Épocas de recolección. Evolución del contenido graso del fruto y de la composición y calidad del aceite. Proceedings of II Seminario Oleícola International. Córdoba, Spain, 25-46.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. Chemometrics and intelligent laboratory systems, 2(1-3), 37-52.

Zhang, B., Gu, B., Tian, G., Zhou, J., Huang, J., & Xiong, Y. (2018). Challenges and solutions of optical-based nondestructive quality inspection for robotic fruit and vegetable grading systems: A technical review. Trends in food science & technology, 81, 213-231.

Zhao, Y., Gong, L., Huang, Y., & Liu, C. (2016). A review of key techniques of vision-based control for harvesting robot. Computers and Electronics in Agriculture, 127, 311-323.

In this paper, the candidate was responsible of the methodological approach relative to the modes building, he conducted the data collection along the whole sampling campaign, wrote and reviewed the final manuscript.

PAPER 6: Near Infrared Spectroscopy as a Green Technology for the Quality Prediction of Intact Olives

Silvia Grassi [1], Olusola Samuel Jolayemi [1], Valentina Giovenzana [2], Alessio Tugnolo [2], Giacomo Squeo [3], Paola Conte [4], Alessandra De Bruno [5], Federica Flamminii [6], Ernestina Casiraghi [1] and Cristina Alamprese [1,*]

[1] Department of Food, Environmental, and Nutritional Sciences (DeFENS), Università degli Studi di Milano, via G. Celoria 2, 20133 Milan, Italy; silvia.grassi@unimi.it (S.G.); olusola.jolayemi@unimi.it (O.S.J.);
ernestina.casiraghi@unimi.it (E.C.)

[2] Department of Agricultural and Environmental Sciences (DiSAA), Università degli Studi di Milano, via G. Celoria 2, 20133 Milan, Italy; valentina.giovenzana@unimi.it (V.G.); alessio.tugnolo@unimi.it (A.T.)

[3] Department of Soil Plant and Food Sciences (DiSSPA), Università degli Studi di Bari "Aldo Moro", via Amendola 165/A, 70126 Bari, Italy; giacomo.squeo@uniba.it

[4] Department of Agricultural Sciences, Università degli Studi di Sassari, Viale Italia 39/A, 07100 Sassari, Italy; pconte@uniss.it

[5] Department of Agraria, University Mediterranea of Reggio Calabria, Vito, 89124 Reggio Calabria, Italy; alessandra.debruno@unirc.it

[6] Faculty of Bioscience and Technology for Agriculture, Food and Environment, University of Teramo, Via Balzarini 1, 64100 Teramo, Italy; fflamminii@unite.it

* Correspondence: cristina.alamprese@unimi.it; Tel.: +39-0250319187

Abstract

Poorly emphasized aspects for a sustainable olive oil system are chemical analysis replacement and quality design of the final product. In this context, near infrared spectroscopy (NIRS) can play a pivotal role. Thus, this study aims at comparing performances of different NIRS systems for the prediction of moisture, oil content, soluble solids, total phenolic content, and antioxidant activity of intact olive drupes. The results obtained by a Fourier transform (FT)-NIR spectrometer, equipped with both an integrating sphere and a fiber optic probe, and a Vis/NIR handheld device are discussed. Almost all the partial least squares regression models were encouraging in predicting the quality parameters ($0.64 < R^2_{pred} < 0.84$), with small and comparable biases ($p > 0.05$). The pair-wise comparison between the standard deviations demonstrated that the FT-NIR models were always similar except for moisture ($p < 0.05$),

whereas a slightly lower performance of the Vis/NIR models was assessed. Summarizing, while on-line or in-line applications of the FT-NIR optical probe should be promoted in oil mills in order to quickly classify the drupes for a better quality design of the olive oil, the portable and cheaper Vis/NIR device could be useful for preliminary quality evaluation of olive drupes directly in the field.

Keywords: antioxidant activity; harvesting time; olive composition; olive cultivars; olive ripening; phenolic compounds; PLS regression model; portable device; quality parameters; sustainability

INTRODUCTION

The economic significance of olive industries to the European Union is unquestionable. Europe contributed almost 70% of the world olive oil production in the 2018–2019 harvest year campaign and the resultant revenue was to the tune of five billion euro [1]. This large and continuously expanding industry is also associated with many negative environmental problems stemmed from waste production and inappropriate disposal, soil depletion, and atmospheric emissions [2]. Every phase in the olive chain is characterized by different environmental concerns. In the agronomic phase, the use of pesticides, herbicides, and fertilizers has been identified as the principal contributor to ecological challenges [3]. In the cultivation phase, activities such as irrigation, pruning, soil management, and fertilizer applications can negatively affect the environment. The impacts of these primary phases are minor when compared to olive oil production and its unit operations. Oil extraction generates the most potentially hazardous organic compounds that accompany olive wastewater and pomace, depending on the techniques [4]. Laudable efforts have been made to adopt sustainable agricultural and industrial practices in the olive value chain to mitigate these problems. For instance, adoption of organic integrated agricultural systems in the farming and cultivation of olives is an example of sustainable agricultural practice. Industrially, practices such as the two-phase olive extraction method, which reduces water consumption, extraction of bioactive phytonutrients from by-products, and overall valorization of the olive production chain have significantly reduced the negative impacts of the industry on the environment [5,6]. However, a rather less emphasized aspect of the sustainable olive system is solvent reduction and replacement strategies during laboratory chemical analyses of olives and olive oils.

These chemical analyses are fundamental to monitor olive ripeness, estimate oil extraction efficiency, and control oil quality. Free acidity, moisture, and oil contents are examples of chemical parameters serving as quick tests on olive drupes before extraction [7].

On-field information of these chemical parameters can suggest suitable harvest time and overall orchard management [8,9]. Immediate first-hand knowledge of moisture and oil content of olive drupes prior to processing can reliably predict the economic viability of the entire production process, therefore informing producers about the raw material composition is of crucial relevance [10,11]. Similarly, prediction of minor constituents such as phenols, pigments, and antioxidants contents of olives can facilitate instant classification of the resultant oils even before production, making official standard compliance and product consistency easier. Commonly used wet methods, such as Soxhlet extraction technique, gravimetry, and chromatography have many unsustainable limitations such as excess solvent consumption, limited sample size, destructive sample preparation, slow response, and technical demand [7]. Thus, for effective processing and quality control of the olive system, application of green, sustainable eco-friendly, energy-efficient, non-destructive, non-invasive, easy-to-use, and inexpensive spectroscopic methods become inevitable.

From the technological point of view, the importance of these rapid determinations before oil extraction may lie in the possibility of modulating the extraction systems based on the drupe characteristics and type of desired product. For instance, operative conditions safeguarding the phenolic content can be adopted if phenolic substances are not so high in the drupes or, vice versa, the outstanding phenolic content of some drupes can be lowered if the final product is intended for consumers who do not like bitter/pungent oil [12,13]. Knowing how to set the equipment before starting the process instead of correcting the settings once the oil has been extracted and analyzed might be of interest.

Near infrared spectroscopy (NIRS) has gained prominence in the last decade and has contributed economically to food and feed industries by ensuring on-time processing and quality control [14,15]. The technology is a formidable green chemistry tool and environmentally sustainable analytical technique capable of handling a large sample size in solid and liquid forms and it provides quick answers to quality questions. NIRS, in conjunction with appropriate chemometrics, has become a routine analytical tool for the determination of intact olive drupes moisture and fat contents [16,17]. Using a portable Vis/NIR spectral acquisition device equipped with multiple detectors, it was possible to predict several economically important olive mill parameters such as maturity index, moisture, oil content, acidity, and dry matter [18]. Another type of NIRS system with a wavelength selection tool (acousto-optically tunable filter—AOTF) was satisfactorily applied to predict phenolic compounds and to monitor ripening of olives [19,20]. In addition to intact or crushed olive quality assessment, NIRS has been found to be handy in evaluation of olive oils and olive by-products [21,22]. However, comparative performance evaluations of NIRS using different

signal acquisition devices are relatively uncommon especially for olive drupes. In this study the results obtained by a Fourier transform (FT)-NIR spectrometer (equipped with both integrating sphere and fiber optic probe) and a Vis/NIR handheld device for the prediction of quality parameters of intact olives of 13 different cultivars collected in three harvest years are discussed. In particular, the objective was to evaluate the different performance of the acquisition systems in the prediction of moisture, oil content, soluble solids, total phenol content, and antioxidant activity, in vision of suitable tools to be applied both in the field and at the mill for quick answers to quality questions in a sustainable way.

MATERIALS AND METHODS

Olive Samples

Samples of olives belonging to 13 different cultivars from Abruzzo, Apulia, Calabria, and Sardinia regions (Italy) were used; sampling was carried out at different ripening degrees. For each sampling and cultivar, three sample units (500 g each) were picked from different identified trees of the same grove, for a total of 267 sample units. Each unit was independently analyzed for the chemical parameters (moisture, oil content, soluble solid content, total phenolic content, and antioxidant activity). Two aliquots (100 g each) were taken from each sample unit for FT-NIR analysis with the integrating sphere. From each aliquot, 10 olives were selected as representative of the ripening stage [23] and used for analyses with both the FT-NIR and Vis/NIR fiber optical probes.

Chemical Analyses

Determination of moisture content (%) was carried out according to the AOAC 934.06 official method [24]. Oil content (% on fresh weight) was determined gravimetrically after the extraction of the oil from 10 g of dehydrated olive paste in a Soxhlet apparatus using petroleum ether as solvent [25]. Total soluble solids content (°Bx) was measured according to a previously published procedure [26]. Briefly, the sugar aqueous solution was prepared by homogenizing olive paste (20 g) in distilled water (40 mL) and stirring for 2 min. After centrifugation (11,000× *g* for 10 min), the supernatant solution was analyzed through a digital refractometer. Total phenol content (TPC) was determined as follows: olive pulp (1 g) was extracted using hexane (3 mL) and methanol:water (70:30 *v*/*v*; 15 mL), by stirring for 10 min at room temperature. After centrifugation (6000× *g* at 4 °C for 10 min), the supernatant phase was collected and further centrifuged (13,600× *g*, 5 min, room temperature). The obtained extracts were filtered through nylon syringe filters (pore size 0.45 µm; LLG Syringe Filter CA, Carlo Erba, Milano, Italy), properly diluted, and spectrophotometrically analyzed at 750 nm

using the Folin-Ciocalteau reagent [27]. Calibration curves were made using gallic acid and the results were expressed as grams of gallic acid equivalent per kilogram olive pulp (g_{GA}/kg). Antioxidant activity (% inhibition/mg olive pulp) was determined on the same extracts used for TPC, applying the radical 2,2 diphenyl-1-picrylhydrazyl (DPPH•) method [28]. Briefly, 200 µL extract (previously diluted 1:20 in methanol) was made to react with 2.8 mL DPPH• methanol solution (6×10^{-5} M) for 1 h at 22 °C, measuring the discoloration at 515 nm. All reagents were from Sigma-Aldrich (St. Louis, MO, USA).

Spectra Collection

Spectra were collected by using a benchtop FT-NIR spectrometer (MPA, Bruker Optics, Milan, Italy), equipped with both an integrating sphere and a fiber-optic probe, and a handheld portable Vis/NIR device (Jaz, OceanOptics Inc., Dunedin, FL, USA). The FT-NIR spectra of the two aliquots (100 g each) of each olive sample unit were collected in duplicate in diffuse reflectance by means of the integrating sphere system. The optical fiber was used to acquire, in duplicate, the FT-NIR spectra of the 10 single olives selected from each aliquot based on ripening degree [23]. For both FT-NIR sampling systems, spectra were collected within a 12,500–3600 cm^{-1} spectral range, at 8 cm^{-1} resolution and with 32 scans. The background for the integrating sphere was performed by closing the internal reference wheel of the module, while for the fiber-optic probe a Spectralon standard was used. A dedicated software (OPUS v. 6.5, Bruker Optics, Ettlingen, Germany) was used to manage the instrument. The same single olives were analyzed in duplicate also by using the Vis/NIR portable device (500–1000 nm, i.e., 20,000–10,000 cm^{-1}; 0.3 nm resolution; 5 scans) equipped with a bifurcated optical fiber provided with a cap that standardizes the distance between the head of the probe and the sample (about 2 mm) and reduces the environmental light interference. A white reference (99% reflection) was used to set the maximum reflection. Spectrum acquisition lasted 18 s for both the integrating sphere and the probe of the benchtop FT-NIR spectrometer, and 1 s for the portable Vis/NIR device. Measurements were conducted with both instruments on the same day, thus making sample storage between analyses unnecessary.

Data Analysis

Data elaborations were performed using the Unscrambler X software (v. 10.4, CAMO ASA, Oslo, Norway). The replicated spectra were averaged in order to have one spectrum for each sample unit. For FT-NIR probe and sphere, spectral ranges were reduced to eliminate non-informative and noisy regions (i.e., 3600–4000 and 10,500–12,500 cm^{-1}), whereas in the case of the portable Vis/NIR device, the whole spectral range was used. The spectral data were

independently pre-processed by standard normal variate (SNV), which removes possible interferences due to light scattering [29]. Chemical variables and all spectral data were merged in a single matrix (267 sample units × 5024 variables) and used to perform principal component analysis (PCA), autoscaling all the variables to overcome the heteroscedasticity nature of the data. The coordinate transformation of the merged spectral–chemical data matrix allowed for the selection of a calibration and a prediction data set, using the Kennard–Stone (KS) algorithm [30]. The algorithm partitioned the data in order to have 70% of samples (187 sample units) in the calibration set and 30% (80 sample units) in the prediction set.

Prediction of olive chemical characteristics based on spectral data was performed applying the partial least squares (PLS) regression to the calibration set of each spectral matrix (187 sample units x 1686 variables for the FT-NIR systems; 187 sample units × 1647 variables for the Vis/NIR equipment) using nonlinear iterative partial least squares (NIPALS) algorithm. Different pre-treatments of spectral data were tested: SNV, first derivative (d1; Savitzky–Golay algorithm, second order polynomial, 11-window size), which allows removal of baseline offset [31], and their combination. After calibration, the models were validated internally, through cross-validation (Venetian blind, 10 cancellation segments). The number of components to be considered for each model was determined based on the plot of calibration and cross-validation errors as a function of the number of latent variables (LVs). The optimal number of LVs was chosen as the number of LV allowing to minimize the cross-validation error. Afterwards, the models were externally validated by independently using the prediction set previously created with KS. Model performance was evaluated in terms of determination coefficients for calibration (R^2_{cal}), cross-validation (R^2_{cv}), and prediction (R^2_{pred}), as well as by root mean square error of calibration (RMSEC), cross-validation (RMSECV), and prediction (RMSEP), and standard error of prediction (SEP).

Prediction performances of the models obtained by the three spectral acquisition systems were compared by different approaches: (i) comparison of intermediate precisions expressed as standard error of laboratory (SEL); (ii) comparison of SEP with SEL of reference analyses; (iii) statistical tests proposed in the scientific literature [32,33]. SEL of the reference analyses and NIRS acquisition systems was calculated as follows [34]:

$$SEL = \sqrt{\frac{\sum_1^m (x_1 - x_2)^2}{m}}$$

where m is the number of olive samples and $x_1 - x_2$ is the absolute value of the difference between replicate results. In the third approach (i.e., statistical tests), first, the model biases, i.e., differences between the reference method results and those of the models predicting the

chemical parameters, were compared by a *t* confidence interval for paired samples with a 95% confidence interval. The null hypothesis (H_0) states that model biases are not different. If the calculated Fisher value is higher than the F critical value, the H_0 is rejected and the hypothesis H_1 is true (i.e., differences between models are significant) [32]. Furthermore, a pairwise comparison of the model standard deviations was performed by the calculation of the correlation coefficient between each two sets of prediction errors (r). Then, K index is calculated by the following equation:

$$K = 1 + \{[2(1 - r^2)t^2_{n-2,0.025}]/(n - 2)\} \quad (1)$$

where $t_{n-2,0.025}$ is the upper 2.5% point of the *t* distribution on n − 2 degrees of freedom. Subsequentially, L index is calculated as follows [33]:

$$L = \sqrt{[K + \sqrt{((K^2 - 1))}]} \quad (2)$$

Then, the 95% confidence interval for the ratio of the standard deviations (L-lower and L-upper limits) was calculated. If the L interval includes 1, the standard deviations are not significantly different ($p > 0.05$). The model comparison was performed in MATLAB environment (v. R2017b, The MathWorks, Inc., Natick, MA, USA).

RESULTS AND DISCUSSION

Chemical Parameters

Descriptive statistics of the chemical variables are presented in Figure 1 as box and whisker plots. The box lines represent the first and third quartiles and the median. The mean value is indicated by a cross sign. Whiskers correspond to the minimum and maximum measured values. Genetic, environmental, and cultivation factors affect olive composition, which changes during growth together with the drupe weight [5]. Actually, the tested cultivars and the different ripening stages and crop seasons accounted for a high range of variability of all the chemical parameters. This is an important point for the development of prediction models useful for different production sites. Variation ranges of the chemical parameters for the different olive cultivars are reported in Table S1.

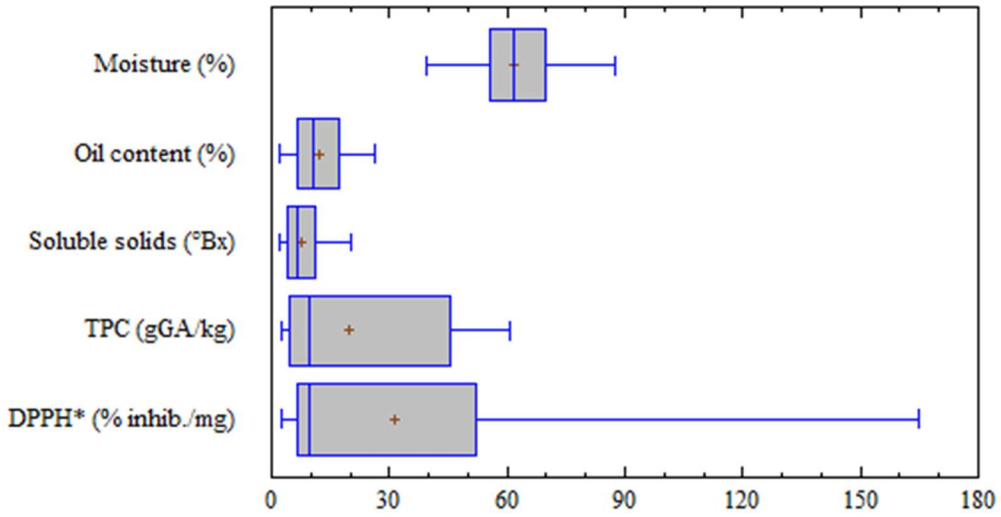


Figure 1. Box and whisker plots showing the descriptive statistics for the chemical variables tested on olive drupes. TPC: total phenol content; GA: gallic acid equivalent; DPPH•: radical 2,2 diphenyl-1-picrylhydrazyl; inhib.: inhibition.

Moisture represents the main constituent alongside oil. In the considered drupes, moisture content ranged from 39.3 to 87.2%. The obtained results agree with previously published data [18,35], considering that the moisture mean value was 63.3%, while the highest values (>80%) were obtained only in three out of thirteen cultivars, all from Calabria region. Excluding those three cultivars, the maximum value for moisture was 73.7%.

Commonly, olives intended for oil production have approximately 20% oil [36]. The samples here considered had a wide range of oil content (1.9–26.0%), suggesting the high influence of cultivar and ripening degree on this parameter. A general increase in oil content ranging from 2 to 12% was observed over ripening, depending on the considered cultivar.

TPC is an approximate estimation of total phenolic acids, phenolic alcohols, flavonoids, and secoiridoids in olive drupes. These compounds confer the bitter taste and pungent sensation on olive oils and are responsible for the well-known antioxidant properties. TPC values of the samples had a wide range of variation (2.5–60.6 g_{GA}/kg), with the highest levels (>35 g_{GA}/kg) found in three cultivars from Sardinia region. The antioxidant activity too was very different in the various samples, ranging from 2.4 to 165.0% inhibition/mg. Unexpectedly, the highest values (>70% inhibition/mg) were not found in the olives with the highest TPC, but in two cultivars from the Apulia region.

Spectral Features

Figure 2 shows the spectra of the olives obtained from the three acquisition systems. Visual features and patterns of the spectra conform with those previously reported for intact olive drupes [37,38].

Aside from the visual differences in band intensities among samples, FT-NIR spectra from the integrating sphere and the fiber-optic probe (Figure 2a,b) are quite similar, with the latter exhibiting higher absorbances in most of the observable peaks.

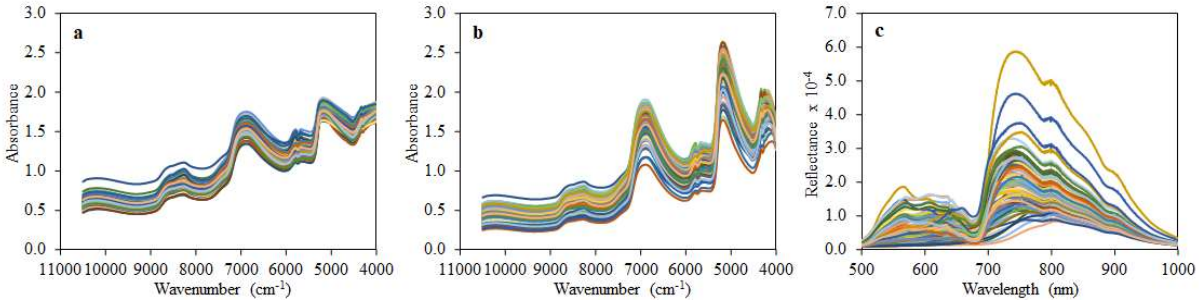


Figure 2. Spectra of olive drupes acquired with: (a) FT-NIR integrating sphere; (b) FT-NIR fiber-optic probe; (c) portable Vis/NIR device.

The low absorbance band around 8600 cm^{-1} represents a combined symmetric and asymmetric OH stretching and bending vibrations. This is followed by the second overtone of CH stretching vibrations at 8300 cm^{-1} that corresponds to methyl ($-CH_3$), methylene ($-CH_2$), and olefin (-CH=CH-) bonds [37]. The high water content of the olive drupes (39–87%) explains the two absorption bands at 7500–6100 and 5400–4500 cm^{-1}. These bands are designated as the combination of first overtone of symmetric and asymmetric OH-bending and OH-stretching bands (6900 cm^{-1}) and combined OH-bending and OH-stretching bands (5200 cm^{-1}), respectively [39]. Similarly to the second overtone of CH stretching vibrations at 8300 cm^{-1}, the two bands at 5800 and 5650 cm^{-1} represent the first overtone of CH-stretching vibrations present in the same CH_3, CH_2, and CH=CH functional groups. At the far end of the FT-NIR spectral range, two peaks at 4335 and 4262 cm^{-1} represent CH and CH_2 s overtones, respectively [35]. However, the intermediate bands between the overtones (i.e., 8600, 5800–5650, and 4350–4250 cm^{-1}) have been attributed to the oil content of the drupes [40]. Regarding olive fruit phenols, there are no reported NIR correlated bands in the literature. However, a previous study suggested that some regions (i.e., 8700–8300 and 5800–5650 cm^{-1}) are correlated with TPC of olives [19].

In the case of Vis/NIR spectra (Figure 2c), more peak variations among samples were observed, especially within the visible (550–680 nm) and near-infrared (700–790 nm) regions.

The changes around 550–680 nm correspond to some varying pigment indices. Specifically, the peak around 540 nm has been associated with anthocyanin, while that at 680 nm has been linked to chlorophyll [41]. Thus, changes in reflectance along these peaks may be due to maturation differences among the drupes. Other parameters, such as soluble solids, pH, and firmness, have been implicated within these regions in pears, especially around 340–740 nm [42]. Changes in the two absorption peaks around 750 and 850 nm could be assigned to the third overtone of H_2O and C-H functional group, respectively [43].

Principal Component Analysis

Figure 3 shows the score and loading plots of the PCA model built on the merged chemical and spectral database. The first two principal components (PCs) represent 59% of total data variance. The application of KS algorithm after PCA allowed to select evenly distributed samples for the calibration and prediction sets, highlighted with different colors in the score plot of Figure 3a. Few samples were seemingly outliers, but they were not removed in order to avoid presumptive assumption that they might adversely affect the model. Anyway, KS data splitting algorithm retained to a large extent as much variability as possible within the calibration and validation sets and this is a prerequisite for model robustness and validity in prediction. The loading plot (Figure 3b) shows a balanced contribution of both the chemical parameters and the three spectral ranges to sample distribution and consequently to the dataset partitioning.

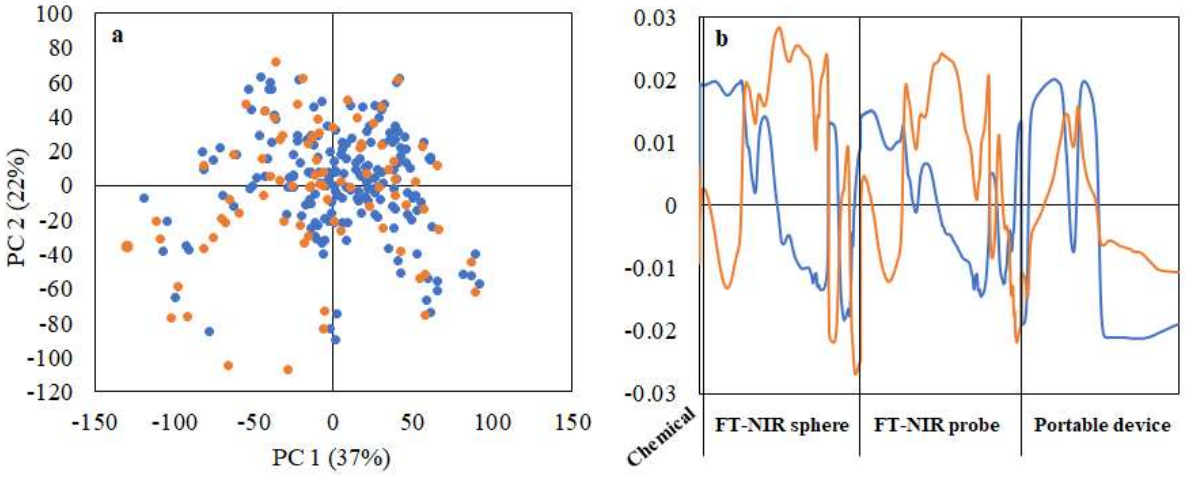


Figure 3. PCA results: (a) score plot showing the distribution of calibration (blue) and prediction (orange) set samples selected by Kennard-Stone algorithm applied on the merged chemical and spectral dataset of olive drupes; (b) loading plot of PC1 (blue) and PC2 (orange).

Regression Models

PLS regression models were built with FT-NIR and Vis/NIR spectra to quantify moisture, oil content, soluble solids, TPC, and antioxidant activity of olive drupes. In order to make data more evenly distributed, TPC and DPPH• results were transformed in the inverse and the logarithmic values, respectively. The best models based on determination coefficients and errors are reported in Table 1 for each spectral acquisition system. Predicted vs. measured plots of the models are reported in the supplementary Figure S1. In general, performances of the three acquisition systems were similar in calibration and cross-validation, while in prediction FT-NIR spectra gave better results, maybe due to the wider NIR range and the low complexity of the models resulting in a higher stability.

With respect to moisture content prediction, the three acquisition systems exhibited promising and similar prediction outcomes. The determination coefficients ranged from 0.77 to 0.92, with reasonably low values of errors (from 2.67 to 4.75%). However, the model calculated with the Vis/NIR spectra transformed in d1 showed a higher number of LVs (16 vs. 8 and 7 for the sphere and the probe, respectively), maybe due to the higher resolution of the spectra and the limited NIR range considered.

Oil content was better predicted by FT-NIR spectra, pre-treated with SNV and d1. Prediction coefficients of determination were higher than those of the portable acquisition system (0.77 and 0.78 vs. 0.64), with lower RMSEP values (2.92 and 2.86% vs. 3.74%) and LVs (9 and 5 vs. 16). The outcomes of calibration and cross-validation coefficients of determination for the FT-NIR sphere and probe (0.77–0.93) were comparable to those reported in the literature (0.78–0.84) for a smaller number of samples (183) [18].

Considering soluble solids, the regression model reliability appeared even more promising for FT-NIR spectrometer than for the portable device. Both the FT-NIR sphere and fiber-optic spectra pre-treated with a combination of SNV and d1 resulted in satisfactory determination coefficients in prediction (0.70 and 0.74, respectively) and low RMSEP (2.39 and 2.23 °Bx, respectively). The precision of the models was comparable to those observed for other fruits as, to the best of our knowledge, there is no study on NIR prediction of soluble solids in intact olives. For instance, quantitative determination of soluble solid content for quality prediction of intact strawberries using a handheld micro-electro-mechanical NIR system, resulted in R^2_{pred} of 0.37–0.47 and RMSEP of 1.02–0.87% [44]. With the spectra in the Vis/NIR range, the coefficient of determination in prediction decreased to 0.58, with a RMSEP of 3.02 °Bx.

Similar model performances in calibration and cross-validation were obtained for 1/TPC for all the spectral acquisition systems (R^2 range, 0.76–0.89), whereas in prediction FT-NIR spectra, gave better results (R^2_{pred} = 0.77–0.76) than Vis/NIR spectra (R^2_{pred} = 0.69). FT-NIR models

are better than those reported in the literature for a filter-based NIR spectrometer [35]. The authors attributed the unsatisfactory output of their model (R^2_{cal}, 0.72; SEP, 13.35 g oleuropein/kg_{dm}) to the exclusion of 8600–6900 cm^{-1} range from the spectral bands, which was instead here considered. Our models were more promising also when compared to grape TPC prediction models developed using a portable NIR-AOTF [45]; the authors observed determination coefficient values of 0.77 and 0.62 in calibration and cross-validation, respectively.

Table 1. Figures of merit of the best PLS regression models for olive chemical parameter prediction based on spectroscopic data.

				Calibration		Cross-Validation		Prediction	
Parameter	**NIR System**	**Pre-treatment**	**LVs**	**R^2_{cal}**	**RMSEC**	**R^2_{cv}**	**RMSECV**	**R^2_{pred}**	**RMSEP**
Moisture content (%)	Sphere	SNV+d1	8	0.92	2.67	0.85	3.66	0.77	4.59
	Probe	SNV+d1	7	0.88	3.56	0.85	3.87	0.84	3.97
	Portable	d1	16	0.87	3.68	0.77	4.77	0.77	4.75
Oil content (%)	Sphere	SNV+d1	9	0.93	1.62	0.82	2.62	0.77	2.92
	Probe	SNV+d1	5	0.79	2.87	0.77	2.99	0.78	2.86
	Portable	SNV+d1	16	0.81	2.72	0.67	3.58	0.64	3.74
Soluble solids (°Bx)	Sphere	SNV+d1	9	0.90	1.45	0.75	2.36	0.70	2.39
	Probe	SNV+d1	11	0.87	1.66	0.80	2.06	0.74	2.23
	Portable	SNV	13	0.79	2.11	0.75	2.34	0.58	3.02
1/TPC (kg/g_{GA})	Sphere	SNV	13	0.89	0.04	0.81	0.04	0.77	0.04
	Probe	SNV+d1	13	0.87	0.04	0.76	0.05	0.76	0.04
	Portable	SNV+d1	9	0.83	0.05	0.79	0.05	0.69	0.05
logDPPH• (log % inhib./mg)	Sphere	SNV	15	0.84	0.20	0.68	0.29	0.68	0.29
	Probe	SNV+d1	16	0.93	0.14	0.79	0.24	0.73	0.27
	Portable	d1	13	0.79	0.23	0.72	0.27	0.41	0.39

TPC: total phenolic content; GA: gallic acid equivalent; inhib.: inhibition; DPPH•: radical 2,2 diphenyl-1-picrylhydrazyl; LVs: latent variables; R^2_{cal}: calibration coefficient of determination; R^2_{cv}: cross-validation coefficient of determination; R^2_{pred}: prediction coefficient of determination; RMSEC, RMSECV, and RMSEP: root mean square errors of calibration, cross-validation, and prediction, respectively; SNV: standard normal variate; d1: first derivative.

To the best of our knowledge, there is no other published paper in which the antioxidant activity of olive drupes is tentatively determined using rapid spectroscopic techniques. Therefore, our models seem fair especially when the FT-NIR probe was used, which generated comparatively highest R^2_{pred} and lowest RMSEP among the three spectral acquisition systems. The dynamic nature of this in vitro antioxidant activity makes its adaptation to spectroscopic techniques somewhat difficult. A more accurate NIR prediction of DPPH• radical scavenging activity was recorded in bean flours (R^2_{cal}, 0.94–0.99; R^2_{val}, 0.85–0.97) [46]. On the contrary, for a more bioactive horticultural product like *Hibiscus sabdariffa*, calibration and prediction determination coefficients are reported in the literature in the ranges 0.82–0.87 and 0.75–0.86, respectively, depending on spectra pretreatments [47].

From the inspection of the weighted regression coefficients of PLS models, for both the FT-NIR sphere and the probe the relevance of 7500–6100 and 5400–4500 cm^{-1} regions for moisture and soluble solid prediction was confirmed. Moreover, the PLS model developed for oil content prediction were highly influenced by the 5800–5650 and 4350–4250 cm^{-1} regions, attributed to the oil content of the drupes [40]. The same regions showed high weighted regression coefficients for TPC and DPPH• models, which were also characterized by high relevance of the 8700–8300 cm^{-1} region, previously related to TPC of olives [19].

As for the model developed with the Vis/NIR spectra, the inspection of the weighted regression coefficients revealed that both visible and NIR range influenced the prediction of moisture, oil content, and soluble solids. In particular, the range 880–970 nm showed the highest influence in the models for moisture and oil content prediction, whereas the maximum recorded weight for soluble solids corresponded to 970 nm. Moving to TPC and DPPH• prediction, it has been noticed that the highest values of the weighted regression coefficients were related to the visible range (550–700 nm), maybe linked to the olive color modification occurring during ripening, due to compounds like chlorophylls, carotenoids, anthocyanins, and polyphenols. Actually, other authors demonstrated that during olive ripening a rise in some bands of the visible range occurs (i.e., 600–650 and 550–625 nm), due to the presence of anthocyanin and other pigments related to reddish as well as green and yellow color [18].

Regression Model Comparison

The effectiveness of the prediction ability was at first established comparing the intermediate precisions (SEL) of the regression models with those of the reference methods (Table 2). The SEL values for the different NIR systems were generally higher than those obtained for the reference analyses, except for the SEL of the oil content predicted by the FT-NIR probe measurements. Indeed, the SEL values of NIR systems are more affected by the drupe

heterogeneity, since spectra are collected on entire olives without the sample preparation phase of the chemical analyses, which is carried out by grinding and homogenizing the olive pulp.

The SEL_{ref} values were also compared with the prediction performances of the models in terms of SEP. As expected, SEP values were always higher than those of SEL_{ref}, because they include not only the sampling and analysis errors, but also the spectroscopy and model errors. The SEP obtained for the FT-NIR probe models were the lowest and the closest to the corresponding SEL_{ref} values. If the SEP is $<2SEL_{ref}$, the prediction performance of the model should be considered as good [48]. This was the case of models developed from FT-NIR probe spectra for moisture, oil content, and 1/TPC prediction.

Furthermore, the *t*-test for paired samples demonstrated that the biases for the models developed with the three spectral acquisition systems were comparable, i.e., the null hypothesis was confirmed (*p* values between 0.1 and 0.8; data not shown). On the other hand, the comparison between the standard deviations of the models [33] returned some differences as reported in the last three columns of Table 2. For moisture, the FT-NIR probe model resulted significantly different from those based on sphere and portable device spectra, due to a better performance resulting in a lower RMSEP (Table 1). All the other comparisons resulted in similar performance of the FT-NIR sphere and probe models, whereas the portable device models resulted significantly different because of the worse performance in terms of R^2_{pred}, RMSEP, and SEP.

Table 2. Comparison of regression models calculated for olive chemical parameter prediction based on three different FT-NIR and Vis-NIR acquisition systems.

Parameter	SEL_{ref}	NIR System	SEL_{NIR}	SEP	NIR System		
					Sphere	Probe	Portable Device
Moisture content (%)	2.00	Sphere	4.41	4.56	-	*	n.s.
		Probe	3.21	3.99	*	-	*
		Portable	4.49	4.72	n.s.	*	-
Oil content (%)	2.29	Sphere	3.13	2.94	-	n.s.	*
		Probe	2.18	2.88	n.s.	-	*
		Portable	2.95	3.77	*	*	-
Soluble solids (°Bx)	1.02	Sphere	2.21	2.41	-	n.s.	*
		Probe	2.31	2.24	n.s.	-	*
		Portable	1.88	3.03	*	*	-
1/TPC (kg/g_{GAE})	0.023	Sphere	0.045	0.044	-	n.s.	*
		Probe	0.044	0.043	n.s.	-	*
		Portable	0.036	0.052	*	*	-
logDPPH• (log % inhib./mg)	0.106	Sphere	0.257	0.287	-	n.s.	*
		Probe	0.282	0.267	n.s.	-	*
		Portable	0.223	0.390	*	*	-

TPC: total phenolic content; GA: gallic acid equivalent; DPPH•: radical 2,2 diphenyl-1-picrylhydrazyl; inhib.: inhibition; SEL_{ref}: standard error of laboratory for reference analyses; SEL_{NIR}: standard error of laboratory for NIR systems; SEP: standard error of prediction; n.s.: not significantly different standard deviation values ($p > 0.05$); *: statistically different standard deviation values ($p \leq 0.05$).

CONCLUSIONS

The benefits of different NIRS acquisition systems as green technology for quality characterization of intact olive drupes were explored. Generally, the calculated PLS models were remarkably encouraging in terms of determination coefficients and errors, both in internal validation and prediction. The model comparison highlighted a general better performance of both the FT-NIR sphere and probe acquisition systems with respect to the handheld device. However, the Vis/NIR device, being portable and relatively cheaper, is worthy of further investigations, because its use could be in any case very useful for preliminary quick quality assessment of olive drupes directly in the field. On the contrary, an on-line or in-line application of the FT-NIR optical probe in the olive mill should be promoted in order to quickly classify the drupes for a better quality design of the olive oil and a more sustainable management of the production chain.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Figure S1: Regression lines obtained for the prediction of entire olive chemical parameters with models developed by FT-NIR integrating sphere, FT-NIR fiber-optic probe, and portable Vis/NIR device. Table S1: Variation ranges of the chemical parameters for the different olive cultivars.

Author Contributions: Conceptualization, S.G., V.G., E.C. and C.A.; methodology, S.G., V.G., A.T., G.S., P.C., A.D.B., C.A.; formal analysis, S.G., O.S.J.; investigation, S.G., A.T., G.S., P.C., A.D.B.; resources, G.S., P.C., A.D.B., F.F.; data curation, S.G., O.S.J., C.A.; writing—original draft preparation, S.G., O.S.J., C.A.; writing—review & editing: V.G., A.T., G.S., P.C., A.D.B., F.F.; supervision, E.C., C.A.; funding acquisition, E.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by AGER 2 Project, grant no. 2016-0105.

Acknowledgments: The authors wish to thank prof. Francesco Caponio for the supervision of the whole research project.

Conflicts of Interest: The authors declare no conflict of interest. The funder had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Casson, A.; Beghi, R.; Giovenzana, V.; Fiorindo, I.; Tugnolo, A.; Guidetti, R. Visible near infrared spectroscopy as a green technology: An environmental impact comparative study on olive oil analyses. *Sustainability* **2019**, *11*, 2611, doi:10.3390/su11092611.
2. Arvanitoyannis, I.S.; Kassaveti, A. Current and potential uses of composted olive oil waste. *Int. J. Food Sci. Technol.* **2017**, *42*, 281–295. doi:10.1111/j.1365-2621.2006.01211.x.
3. Banias, G.; Achillas, C.; Vlachokostas, C. Environmental impacts in the life cycle of olive oil: A literature review. *J. Sci. Food Agric.* **2017**, *97*, 1686–1697, doi:10.1002/jsfa.8143.
4. Souilem, S.; El-Abbassi, A.; Kiai, H.; Hafidi, A.; Sayadi, S.; Galanakis, C.M. Olive oil production sector: Environmental effects and sustainability challenges. In *Olive Mill Waste. Recent Advances for Sustainable Management*; Galanakis, C.M., Ed.; Academic Press: Oxford, UK, 2017; pp. 1–28, doi:10.1016/B978-0-12-805314-0.00001-7.
5. Pantziaros, A.G.; Trachili, X.A.; Zentelis, A.D.; Sygouni, V.; Paraskeva, C.A. A new olive oil production scheme with almost zero wastes. *Biomass Conv. Bioref.* **2020**, doi:10.1007/s13399-020-00625-0.
6. Rosello-Soto, E.; Koubaa, M.; Moubarik, A.; Lopes, R.P.; Saraiva, J.A.; Boussetta, N.; Grimi, N.; Barba, F.J. Emerging opportunities for the effective valorization of wastes and

by-products generated during olive oil production process: Non- conventional methods for the recovery of high-added value compounds. *Trends Food Sci. Technol.* **2015**, *45*, 296–310, doi:10.1016/j.tifs.2015.07.003.

7. Salguero-Chaparro, L.; Palagos, B.; Peña-Rodríguez, F.; Roger, J.M. Calibration transfer of intact olive NIR spectra between a pre-dispersive instrument and a portable spectrometer. *Comp. Electron. Agric.* **2013**, *96*, 202–208, doi:10.1016/j.compag.2013.05.007.
8. Hernández-Sánchez, N.; Gómez-Del-Campo, M. From NIR spectra to singular wavelengths for the estimation of the oil and water contents in olive fruits. *Grasas Aceites* **2018**, *69*, 1–13, doi:10.3989/gya.0457181.
9. Morrone, L.; Neri, L.; Cantini, C.; Alfei, B.; Rotondi, A. Study of the combined effects of ripeness and production area on Bosana oil's quality. *Food Chem.* **2018**, *245*, 1098–1104, doi:10.1016/j.foodchem.2017.11.061.
10. Correa, E.C.; Roger, J.M.; Lleó, L.; Hernández-Sánchez, N.; Barreiro, P.; Diezma, B. Optimal management of oil content variability in olive mill batches by NIR spectroscopy. *Sci. Rep.* **2019**, *9*, 1–11, doi:10.1038/s41598-019-50342-6.
11. Giovenzana, V.; Beghi, R.; Romaniello, R.; Tamborrino, A.; Guidetti, R.; Leone, A. Use of visible and near infrared spectroscopy with a view to on-line evaluation of oil content during olive processing. *Biosyst. Eng.* **2018**, *172*, 102–109, doi:10.1016/j.biosystemseng.2018.06.001.
12. Bianchi, B.; Tamborrino, A.; Giametta, F.; Squeo, G.; Difonzo, G.; Catalano, P. Modified rotating reel for malaxer machines: Assessment of rheological characteristics, energy consumption, temperature profile, and virgin olive oil quality. *Foods* **2020**, *9*, 813, doi:10.3390/foods9060813.
13. Caponio, F.; Squeo, G.; Brunetti, L.; Pasqualone, A.; Summo, C.; Paradiso, V.M.; Catalano, P.; Bianchi, B. Influence of the feed pipe position of an industrial scale two-phase decanter on extraction efficiency and chemical-sensory characteristics of virgin olive oil. *J. Sci. Food Agric.* **2018**, *98*, 4279–4286, doi:10.1002/jsfa.8950.
14. Porep, J.U.; Kammerer, D.R.; Carle, R. On-line application of near infrared (NIR) spectroscopy in food production. *Trends Food Sci. Technol.* **2015**, *46*, 211–230, doi:10.1016/j.tifs.2015.10.002.
15. Qu, J.H.; Liu, D.; Cheng, J.H.; Sun, D.W.; Ma, J.; Pu, H.; Zeng, X.A. Applications of near-infrared spectroscopy in food safety evaluation and control: A review of recent research advances. *Crit. Rev. Food Sci. Nutr.* **2015**, *55*, 1939–1954, doi:10.1080/10408398.2013.871693.

16. Leon, L.; Garrido, A.; Downey, G. Parent and harvest year effects on near-infrared reflectance spectroscopic analysis of olive (*Olea europaea* L.) fruit traits. *J. Agric. Food Chem.* **2004**, *52*, 4957−4962, doi:10.1021/jf0496853.
17. Saha, U.; Jackson, D. Analysis of moisture, oil, and fatty acid composition of olives by near-infrared spectroscopy: Development and validation calibration models. *J. Sci. Food Agric.* **2018**, *98*, 1821–1831, doi:10.1002/jsfa.8658.
18. Cayuela, A.J.; Camino, M.D.P. Prediction of quality of intact olives by near infrared spectroscopy. *Eur. J. Lipid Sci. Technol.* **2010**, *112*, 1209–1217, doi:10.1002/ejlt.201000372.
19. Bellincontro, A.; Taticchi, A.; Servili, M.; Esposto, S.; Farinelli, D.; Mencarelli, F. Feasible application of a portable NIR-AOTF tool for on-field prediction of phenolic compounds during the ripening of olives for oil production. *J. Agric. Food Chem.* **2012**, *60*, 2665–2673, doi:10.1021/jf203925a.
20. Cirilli, M.; Bellincontro, A.; Urbani, S.; Servili, M.; Esposto, S.; Mencarelli, F.; Muleo, R. On-field monitoring of fruit ripening evolution and quality parameters in olive mutants using a portable NIR-AOTF device. *Food Chem.* **2016**, *199*, 96–104, doi:10.1016/j.foodchem.2015.11.129.
21. Azizian, H.; Mossoba, M.M.; Fardin-Kia, A.R.; Karunathilaka, S.R.; Kramer, J.K.G. Developing FT-NIR and PLS1 methodology for predicting adulteration in representative varieties/blends of extra virgin olive oils. *Lipids* **2016**, *51*, 1309–1321, doi:10.1007/s11745-016-4195-0.
22. Comino, F.; Ayora-Cañada, M.J.; Aranda, V.; Díaz, A.; Domínguez-Vidal, A. Near-infrared spectroscopy and X-ray fluorescence data fusion for olive leaf analysis and crop nutritional status determination. *Talanta* **2018**, *188*, 676–684, doi:10.1016/j.talanta.2018.06.058.
23. Tugnolo, A.; Giovenzana, V.; Beghi, R.; Grassi, S.; Alamprese, C.; Casson, A.; Casiraghi, E.; Guidetti, R. A diagnostic visible/near infrared tool for a fully automated olive ripeness evaluation in a view of a simplified optical system. *Comput. Electron. Agr.* **2020,** 105887.
24. AOAC. *Official Methods of Analysis of the Association of Official Analytical Chemists International*, 17th ed.; Official Method 934.06. Moisture in Dried Fruits; Journal of AOAC International: Gaithersburg, MD, USA, 2000.
25. Thiex, N.J.; Anderson, S.; Gildemeister, B. Crude fat, diethyl ether extraction, in feed, cereal grain, and forage (Randall/Soxtec/submersion method): Collaborative study. *J. AOAC Int.* **2003**, *86*, 888–898. doi:10.1093/jaoac/86.5.888.

26. Migliorini, M.; Cherubini, C.; Mugelli, M.; Gianni, G.; Trapani, S.; Zanoni, B. Relationship between the oil and sugar content in olive oil fruits from Moraiolo and Leccino cultivars during ripening. *Sci. Hort.* **2011**, *129*, 919–921, doi:10.1016/j.scienta.2011.05.023.
27. Singleton, V.L.; Rossi, J. Colorimetry of total phenolics with phosphomolybdic-phosphotungstic acid reagents. *Am. J. Enol. Vitic.* **1965**, *16*, 144–158.
28. Conte, P.; Squeo, G.; Difonzo, G.; Caponio, F.; Fadda, C.; Caro, A. Del; Urgeghe, P.P.; Montanari, L.; Montinaro, A.; Piga, A. Change in quality during ripening of olive fruits and related oils extracted from three minor autochthonous Sardinian cultivars. *Emir. J. Food Agric.* **2019**, *31*, 196–205, doi:10.9755/ejfa.2019.v31.i3.1923.
29. Rabatel, G.; Marini, F.; Walczak, B.; Roger, J.M. VSN: Variable sorting for normalization. *J. Chemom.* **2020**, *34*, 1–16, doi:10.1002/cem.3164.
30. Kennard, R.W.; Stone, L.A. Computer aided design of experiments. *Technometrics* **1969**, *11*, 137–148.
31. Rinnan, Å.; Van Den Berg, F.; Engelsen, S.B. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC* **2009**, *28*, 1201–1222, doi:10.1016/j.trac.2009.07.007.
32. Roggo, Y.; Duponchel, L.; Ruckebus ch, C.; Huvenne, J.P. Statistical tests for comparison of quantitative and qualitative models developed with near infrared spectral data. *J. Mol. Struct.* **2003**, *654*, 253–262, doi:10.1016/S0022-286000248-5.
33. Fearn, T. Comparing standard deviations. *NIR News* **1996**, *7*, 5–6, doi:10.1255/nirn.378.
34. The European Agency for the Evaluation of Medicinal Products. Note for guidance on the use of near infrared spectroscopy by the pharmaceutical industry and the data requirements for new submissions and variations. *EMEA/CVMP/961/01* **2003**. Available online: https://www.ema.europa.eu/en/documents/scientific-guideline/note-guidance-use-near-infrared-spectroscopy-pharmaceutical-industry-data-requirements-new_en.pdf (accessed on 15th April 2021).
35. Trapani, S.; Migliorini, M.; Cecchi, L.; Giovenzana, V.; Beghi, R.; Canuti, V.; Fia, G.; Zanoni, B. Feasibility of filter-based NIR spectroscopy for the routine measurement of olive oil fruit ripening indices. *Eur. J. Lipid Sci. Technol.* **2017**, *119*, 1600239, doi:10.1002/ejlt.201600239.
36. de la Casa, J.A.; Castro, E. Fuel savings and carbon dioxide emission reduction in a fired clay bricks production plant using olive oil wastes: A simulation study. *J. Clean Prod.* **2018**, *185*, 230–238, doi:10.1016/j.jclepro.2018.03.010.

37. Fernández-Espinosa, A.J. Combining PLS regression with portable NIR spectroscopy to on-line monitor quality parameters in intact olives for determining optimal harvesting time. *Talanta* **2016**, *148*, 216–228, doi:10.1016/j.talanta.2015.10.084.
38. Giovenzana, V.; Beghi, R.; Civelli, R.; Trapani, S.; Migliorini, M.; Cini, E.; Zanoni, A.; Guidetti, R. Rapid determination of crucial parameters for the optimization of milling process by using visible/near infrared spectroscopy on intact olives and olive paste. *Ital. J. Food Sci.* **2017**, *29*, 357–369, doi:10.14674/1120-1770/ijfs.v560.
39. Dupuy, N.; Galtier, O.; Le Dréau, Y.; Pinatel, C.; Kister, J.; Artaud, J. Chemometric analysis of combined NIR and MIR spectra to characterize French olives. *Eur. J. Lipid Sci. Technol.* **2010**, *112*, 463–475, doi:10.1002/ejlt.200900198.
40. Salguero-Chaparro, L.; Peña-Rodríguez, F. On-line versus off-line NIRS analysis of intact olives. *LWT Food Sci. Technol.* **2014**, *56*, 363–369, doi:10.1016/j.lwt.2013.11.032.
41. Beghi, R.; Giovenzana, V.; Marai, S.; Guidetti, R. Rapid monitoring of grape withering using visible near-infrared spectroscopy. *J. Sci. Food Agric.* **2015**, *95*, 3144–3149, doi:10.1002/jsfa.7053.
42. Li, J.; Huang, W.; Zhao, C.; Zhang, B. A comparative study for the quantitative determination of soluble solids content, pH and firmness of pears by Vis/NIR spectroscopy. *J. Food Eng.* **2013**, *116*, 324–332, doi:10.1016/j.jfoodeng.2012.11.007.
43. Xia, Y.; Huang, W.; Fan, S.; Li, J.; Chen, L. Effect of spectral measurement orientation on online prediction of soluble solids content of apple using Vis/NIR diffuse reflectance. *Infrared Phys. Technol.* **2019**, *97*, 467–477, doi:10.1016/j.infrared.2019.01.012.
44. Sánchez, M.T.; De La Haba, M.J.; Benítez-López, M.; Fernández-Novales, J.; Garrido-Varo, A.; Pérez-Marín, D. Non-destructive characterization and quality control of intact strawberries based on NIR spectral data. *J. Food Eng.* **2012**, *110*, 102–108, doi:10.1016/j.jfoodeng.2011.12.003.
45. Barnaba, F.E.; Bellincontro, A.; Mencarelli, F. Portable NIR-AOTF spectroscopy combined with winery FTIR spectroscopy for an easy, rapid, in-field monitoring of Sangiovese grape quality. *J. Sci. Food Agric.* **2014**, *94*, 1071–1077, doi:10.1002/jsfa.6391.
46. Carbas, B.; Machado, N.; Oppolzer, D.; Queiroz, M.; Brites, C.; Rosa, E.A.S.; Barros, A.I.R.N.A. Prediction of phytochemical composition, in vitro antioxidant activity and individual phenolic compounds of common beans using MIR and NIR spectroscopy. *Food Bioproc. Technol.* **2020**, *13*, 962–977, doi:10.1007/s11947-020-02457-2.
47. Tahir, H.E.; Xiaobo, Z.; Jiyong, S. Rapid determination of antioxidant compounds and antioxidant activity of Sudanese karkade (*Hibiscus sabdariffa* L.) using near infrared spectroscopy. *Food Anal. Methods* **2016**, *9*, 1228–1236, doi:10.1007/s12161-015-0299-z.

48. Shenk, J.S.; Westerhaus. M.O. Calibration the ISI Way. In *Near Infrared Spectroscopy: The Future Waves*; Davies, P.C., Williams, A.M.C., Eds.; NIR Publications: Chichester, UK, 1996; pp. 198–202.

4.3 Design and development of optical prototypes

In this chapter, optical IoT devices for real-time grape monitoring were designed, built, and tested. In particular, the viticulture sector has been taken into account because is one of the supply chains where the presence of optical instruments is increasing day by day.

In the winemaking industry, grape maturation control is a complex process that is critical to produce high-quality wines, but currently, maturation control is cumbersome and inefficient. This inefficient control of the maturation is related to a reduced value of the wine.

The current state of the art for grape maturation control is based on multiple wet-chemistry assays that are: (i) destructive, (ii) time consuming, (iii) labour intensive, (iv) performed on a sparse basis and (v) based in on complex sampling process.

Therefore, to shift the current paradigm of grape maturation monitoring (based on lab analysis) it is needed a new technology that: (i) allows a real-time and stand-alone monitoring with a low-cost, (ii) is non-destructive and chemical free, (iii) is capable to drastically reduce the need of manpower and (iv) provides information with temporal and spatial resolution.

Optical instruments have been present into the winery laboratories for several years but, nowadays, the trends is to move from the laboratory (especially for the classical analysis of technological and phenolic maturation) into the field where researchers have demonstrated alternative optical methods and instruments of proximal and remote sensing that allowed a more cost-effective evaluation of the condition of the grapevines directly in the field (Power et al., 2019).

The market proposes portable instruments which use a set of discrete commercial modules (LED arrays, white light sources, photodiodes, filters, micro-spectrometers) on a package. These instruments provide wide bulk of information, but, currently, they require a human operator for data.

4.3.1 Hand-held optical device development

PAPER 7: Design of cost-effective LED based prototypes for the evaluation of grape (Vitis Vinifera L.) ripeness

Alessia Pampuri, Alessio Tugnolo*, Valentina Giovenzana, Andrea Casson, Riccardo Guidetti and Roberto Beghi

Department of Agricultural and Environmental Sciences (DiSAA), Università degli Studi di Milano, via G. Celoria 2, 20133 Milan, Italy.

* Corresponding author: Alessio Tugnolo

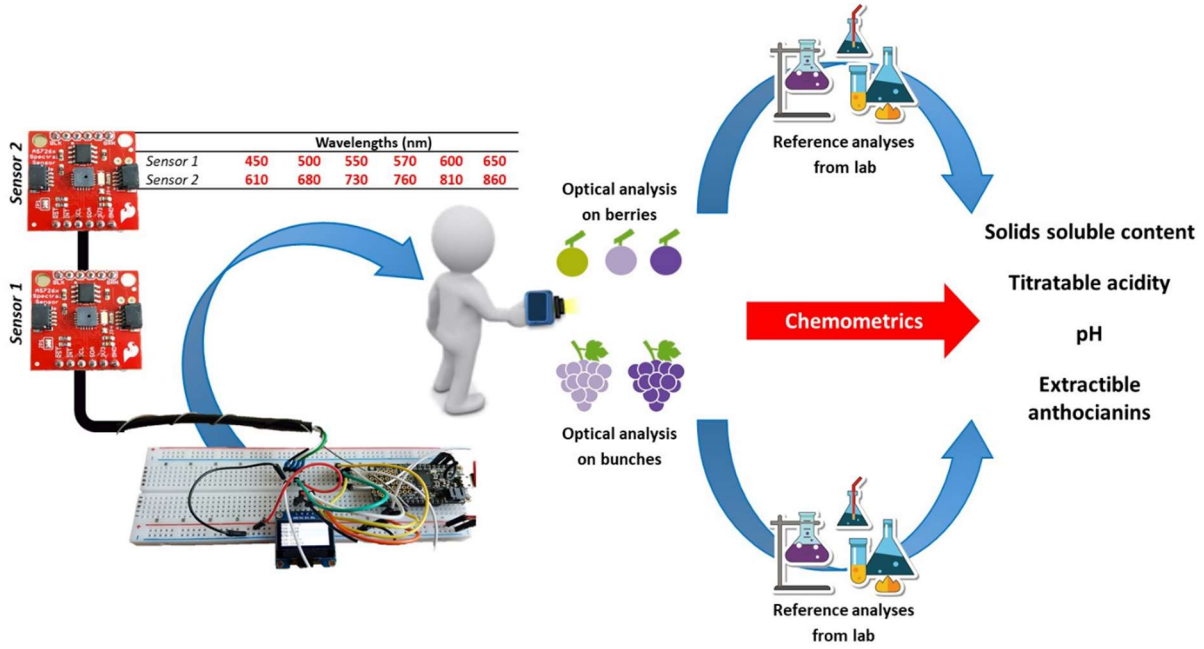
E-mail address: alessio.tugnolo@unimi.it

Abstract

Monitoring the grape ripening until the harvest is a crucial issue since berry quality is closely related to it. Therefore, the research for non-destructive methods which could explore many samples and give a rapid and comprehensive overview of ripening would be helpful. The aim of the research was to design, build and test prototypes of cost-effective and user-friendly device to support small-scale growers in planning the optimal harvest date according to grape ripening degree. A pre-prototype version of a fully integrated optical device which incorporates sensors (tuned photodiode arrays, interference filters, LEDS, optics) was presented. The system is equipped with two digital 6-channel sensors for spectral identification in the visible and SW-NIR. The sensors cover 12 independent on-device optical filters from 450 nm to 860 nm. It was tested on Nebbiolo grape red variety. The optical data were collected on grape bunches and single grape berries using the pre-prototype. Besides, reference parameters through traditional laboratory analyses of Soluble Solids Content (SSC), Titratable Acidity (TA), Extractability of Anthocyanins (EA) and pH were carried out on each sample. The MLR correlation models between the optical data from the prototypes and the reference parameters were calculated. A promising determination coefficient in cross validation (r2cv) was obtained for the prediction of SSC (r2cv=0.86) while, the models for TA, EA and pH (r2cv from 0.4 to 0.5) can be considered enough accurate to allow an initial field screening of the trend of these parameters.

Keywords: effective wavelengths, vis/NIR spectroscopy, simplified system, ripening, chemometrics

Graphical abstract



INTRODUCTION

Europe is one of the world leaders for the production of high-quality wine. In 2019 the sold European (EU) production of wine was around 16 billion litres of which 7.1 billion were exported (Eurostat, 2019).

The wine sector is evolving in an increasingly competitive international scenario characterised by new producing countries with innovative strategies in production and trade. In this highly competitive market, it is now well accepted that the quality of a wine depends mainly on the qualitative features in terms of chemical characteristics of the grapes used to produce it (Giovenzana et al., 2018).

This is well understood by the consumers which are also more exigent on everything is related to the sensory characteristics of a specific wine. Moreover, sustainability aspects are becoming one of the crucial factors which define the added value of the final product (Vallone et al., 2019). This is causing a growing interest in sustainable high-quality production especially in the agro-food sector where the use of more sustainable, automated and precise non-destructive monitoring analytical systems have become a crucial and competitive aspect (Miranda et al., 2019). In vinification, the phenolic (anthocyanin and polyphenols) and technological (soluble solids content, pH and acidity) features of the grape (defined by chemical analysis) can influence the decision making for the wine production.

The optimal harvest decision is made by winegrowers and winemakers who work closely together using an integrated approach toward maturity assessment/monitoring. Therefore, to support the growers in these fundamental decisions, many techniques starting from the simply visual evaluation of the grape (by the operator) going toward more sophisticated techniques (which belong to a new concept of viticulture defined "viticulture 4.0") are available (Aleixandre-Tudo et al., 2019).

The conventional grape maturation assessment methodologies rely on wet-chemistry analysis of the grape composition in the laboratory. These methods are reliable but they are affected by the limited number of samples tested, the distance to field and sometime a time gap between sample collection and results. Furthermore, they are (i) destructive, (ii) time-consuming, (iii) labour-intensive, and (iv) generate significant amounts of chemical waste which are critical factors in a view of a more sustainable production (Casson et al., 2019).

Over the years, many approaches have been followed towards more effective methods which could explore a large number of samples and give a rapid and comprehensive overview of the ripening process. Optical sensing techniques based on ultra-violet (UV), visible (Vis) and infrared (IR) are widely used in agriculture and for food fingerprinting. Most of these techniques offer the possibility of analysing many samples in a non-destructive way allowing a simultaneous determination of several chemical properties. Moreover, they are more sustainable than the traditional physical and chemical methods (Tugnolo et al., 2021, Casson et al., 2019).

In literature, optical techniques showed a good ability to assess the quality parameters of fruit and vegetables (Nicolaı et al., 2007). In grape and wine field, several works were performed. Boca-Bocanegra et al. (2019) studied spectra of intact grapes and skins at harvest in two crop seasons (2016 and 2017) using a portable NIR device (908-1676 nm) to evaluate the capability to determine the levels of extractable phenolic compounds of red grapes. Costa et al. (2019) used a spectroradiometer (450-1800 nm) to develop predictive models for quality and maturation stage attributes (SSC, total anthocyanins and yellow flavonoids) of wine grapes. Even though several studies were performed using portable NIR spectrometers, the development of a new generation of simplified and customised optical sensors have not been fully studied, especially in practical conditions.

Potential users, in particular personnel from the agri-food sector, are interested in scientific studies regarding the prediction performance and practical application of these equipment (Donis-González I et al., 2020).

The limited adoption of NIR technology by the viticulture sector can be attributed to the costs and technical limitations (Magwaza et al., 2012). Therefore, in order to support the winemakers, the research is moving towards simplified, easy to use, low-cost systems for real-time assessment of fruit ripeness directly in-situ (Giovenzana et al., 2015).

Giovenzana et al. (2015) and Ribera-Fonseca et al. (2016) investigated the potential use of a portable device based on few wavelengths (630, 690, 750, 850 nm and 560,640 nm respectively), to measure parameters as SSC, titratable acidity, firmness and anthocyanins getting promising results. The predictive multiple linear models for SSC and TA developed by Giovenzana et al. (2015) showed (in validation) a determination coefficient (R2) of 0.66 and 0.85 and a root mean square error (RMSE) of 1.9 and 1.8 respectively. The model developed by Ribera-Fonseca et al. (2016) instead, showed that IAD (Index of Absorbance Difference) values were significantly correlated to SSC (R2 = 0.92), TA (R2 = 0.87), firmness (R2 = 0.89), and monomeric and total anthocyanin concentration (R2 ranging from 0.68 to 0.97).

Donis-Gonzàlez et al. (2020) have conducted analyses between 740 and 1070 nm using two commercially available portable spectrometers to determine table grape and peach quality attributes, showing significative variations of the two instrument's performance potentially due to their design.

All these innovative approaches are becoming crucial in a view of a more efficient and completely inter-connected wine production process (Lee et al., 2015). For winemakers, the development of optical instrumentation equipped with very inexpensive sensors that can be placed directly in contact with the fruits for continuous monitoring without operator (thanks to cloud data storage) could be an interesting opportunity of development. In the context of precision agriculture, the development of new sensors, especially based on spectroscopy, enables high resolution data acquisition that could be used to track crop development and ripening. The capability to assess ripening in a fast, non-destructive way would substantially and positively impact the processes of harvesting (operating procedures, scheduling and classification).

Nowadays, costs (instruments plus operation) are still an issue for large-scale application in agri-food chains, discouraging, in many cases, the technology transfer of the laboratory research. For this reason, the aim of the research was to design, build and test prototypes of miniaturised low-cost and user-friendly devices to support small-scale growers in determining the optimal harvest date according to grape ripening degree. The optical bench can be made up of highly miniaturized and robust components now available on the market from (multinational) producers of optoelectronic components. This peculiarity guarantees the quality

of the components at an extremely low cost. The development of new simple and interconnected optical systems would also allow the creation of remote storage of optical databases, allowing the continuous updating and refinement of the prediction performance of the predictive models with the aim of a continuous upgrading of the control services at different and crucial levels of the winemaking process.

MATERIALS AND METHODS

Prototype layout

The new optical prototype (technology readiness level estimated equal to 3, Fig. 1) is composed by tuned photodiode arrays, interference filters, LEDs, optics. In detail, the device incorporates MEMS (Micro Electro-Mechanical Systems) equipped with 6-channel digital sensors each one (AMS, models AS7262 visible and AS7263 NIR, Premstaetten, Austria-Europe) for spectral measurement in the visible (vis) and Short Wave Near-infrared (SW-NIR) region. The vis and SW-NIR sensors are 4.5 × 4.4 mm in size and are classified as ultra-low power consumption sensors. They have a 16-bit radiometric resolution and 12 independent on-device optical filters from 450 nm to 860 nm as summarized in table 1.

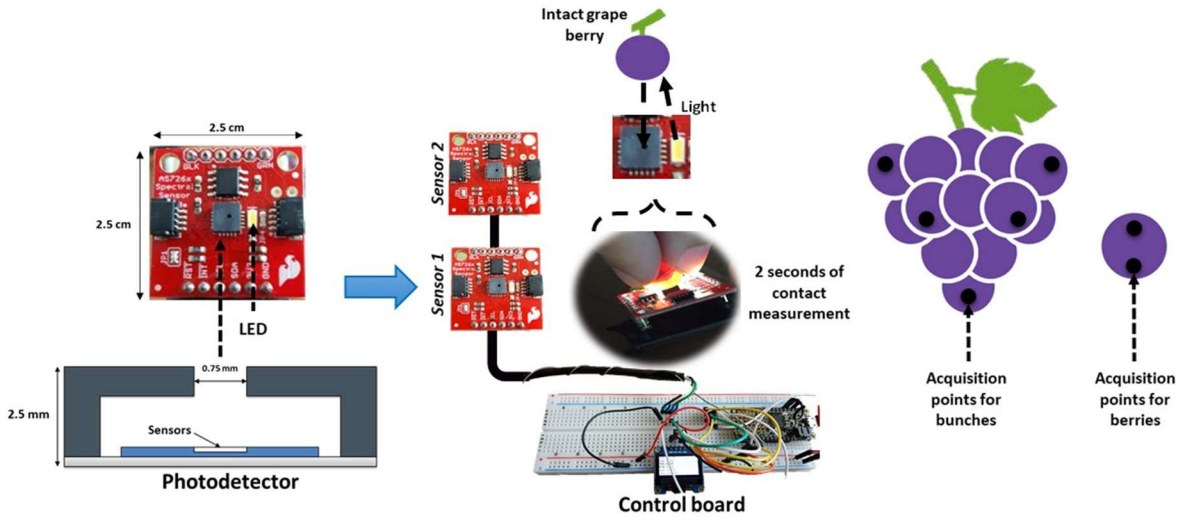


Figure 1. Prototype layout and optical acquisition setup carried out on bunches and single berries.

Table 1. Sensors wavelengths description.

	Wavelengths (nm)					
Sensor 1	450	500	550	570	600	650
Sensor 2	610	680	730	760	810	860

The prototype enables chip-scale spectral analysis by integrating Gaussian filters into standard complementary metal-oxide-semiconductor (CMOS) silicon via nano-optic deposited interference filter technology. The six-channel vis sensor is sensitive to the 400 - 700 nm spectral range with centre wavelengths of 450 nm, 500 nm, 550 nm, 570 nm, 600 nm, and 650 nm (interesting to get also colour information, a crucial feature to get indications regarding the ripening progress). The full-width half-maximum of the sensors is 40 nm. The six-channel SW-NIR sensor is sensitive to the 600 - 900 nm spectral range. Centre wavelengths of the six sensors are 610 nm, 680 nm, 730 nm, 760 nm, 810 nm, and 860 nm. The full-width half-maximum of these sensors is 20 nm.

The visible and short wave near-infrared sensitive spectrometers are available in the form of breakout boards, which include sensors and auxiliary electronic components (Fletcher & Fisher, 2018).

Sampling

The experimental activity took place in the viticulture area of Treiso at “Azienda Agricola Renato Fenocchio” (Cuneo, Piedmont, Italy, 44°41′21.2″N 8°05′13.1″) during the ripening period among August and October 2019 in order to make the last sampling date coincide with the harvest date.

Sampling was performed on Nebbiolo grape (Vitis vinifera L.) on a hillside with ~20-degree slope and a western exposure. Nebbiolo is one of the most important and autochthonous red cultivars of the north of Italy used to produce high-quality aged wines such as Barolo and Barbaresco. Nebbiolo is very sensitive to terroir and is characterized by high vigour and reduced berry skin colour.

A total of 150 bunches were analysed weekly till the harvest for five sampling dates (30 bunches each date). The samples were collected (for each sampling date) from 30 grapevines placed in different areas of the vineyard in order to be representative at best of the ripening process of the whole estate. Furthermore, a specific sampling on 180 single berries was also performed in the same dates.

Optical analysis

Optical acquisitions were performed (straight away before the wet-chem reference analyses) on bunches and single berries directly in the winery laboratory without any sample preparation and immediately after the sampling in the vineyard. In order to avoid any sample degradation, samples were conserved in refrigerated conditions and analysed within 15 minutes after picking. Data were acquired using the LED fully integrated pre-prototype (Fig. 1).

The wavelengths used in the sensors are characteristic of certain peaks of absorption: 630 and 690 nm are near chlorophyll peak, 730 nm is close to the third overtone of OH stretching, while 810 and 860 nm are close to the combination band of OH group of sugars (Giovenzana et al., 2015).

The same measurements were carried out using both optical sensors. For each grape bunch, 5 optical measurements randomly from the top to the bottom of the bunches were performed and then averaged to obtain a mean value while, for each berry the optical measurements were 2 as shown in figure 1. Therefore, two different datasets have been obtained (one from the bunch data collection and one from the berry data collection). In particular, since the berry dataset contains an optical output of each berry, the optical acquisitions have been collected following a colour criterion from green to purple in order to cover at best the variability associated with the grape ripening stage.

The pre-prototype was configurated (using a portable computer) to perform an average of 10 scans for each acquisition point in order to reduce the experimental noise related to the light source.

Chemical analyses

The wet-chemical analyses (reference parameters) were also performed on the same samples. Soluble Solids Content (SSC) was measured both on single berries (one measurement) and on bunches (three averaged replicates from each bunch) using a digital refractometer (PAL-1 ATAGO, Tokyo, Japan, accuracy refractive index ±0.2 °Brix) which measures the refractive index of the juice derived from the mashing of the sample due to the total content of soluble solids and expresses the result in °Brix. Moreover, on each bunch were defined the Titratable Acidity (TA, three averaged replicates from each bunch), the Extractability of Anthocyanins (EA, three averaged replicates from each bunch) and pH (three averaged replicates from each bunch). TA (gtartaric acid dm-3) was identified using an automatic titrator (TitroMatic KF 1S, Crison Instruments, Milan, Italy). EA (%), or the capacity of the grapes to release anthocyanins, was calculated by means of Glories method (Ribéreau-Gayon et al., 2006) based on the optical density (OD) measurement at 280 and 520 nm using an UV/vis spectrophotometer. This technique involves the rapid extraction of anthocyanins from the sample using hydrochloric acid which can facilitate the process by destroying the cell membranes of the grape. Finally, pH was defined using a portable pH meter PCE-PHD 1 (PCE Inst. GmbH, Meschede, Germany).

Data processing

The reference parameters (SSC, TA, EA and pH) and the optical reflectance data obtained from the LED pre-prototype were analysed using The Unscrambler X software package (Camo Software, Oslo, Norway). In this first phase of analysis, the data obtained from the two sensors were processed and analysed as data obtained from one single pre-prototype in order to take advantage of all wavelengths considered.

Principal Component Analysis (PCA) was carried out in order to understand the relationships among all variables and among variables and samples (Giovenzana et al., 2018; Malegori et al., 2018). Considering the inhomogeneous physical structure of the grape samples (size and shape) and the various positions and distances where the sample can interact with the sensor, a correction of the baseline vertical shifts (offsets) and of the global intensity effects (typically arising from unwanted light scattering) was performed, applying the Standard Normal Variate (SNV) transform. Afterwards, a data scaling phase was performed in order to make the different variables comparable in importance before applying scale-dependent multivariate

analysis methods (such as PCA or Multiple Linear Regression, MLR). For this purpose, the unit variance scaling (or autoscaling) where variables are divided by their respective standard deviations was applied. The method is commonly used to data sets containing variables with different units and scales in order toto impose equal weights in the analysis (Marini et al., 2017, Tugnolo et al., 2019). Moreover, an outlier detection procedure was applied on PCA scores using the 'Hotelling T2 computation' function (α value was set to 0.05).

The reference parameters on bunches (SSC, TA, EA and pH) and on single berries (SSC) were used for the calculation of MLR predictive models, since the number of these parameters is not particularly high, and since they can be considered independent from each other but all linearly related to the response variable. The MLR, indeed, allows to relate the variations in a response variable (reference parameter) to the variations of several predictors (optical data from pre-prototype) (Livingstone et al., 2005). Finally, the prediction capability of the MLR models was verified to evaluate the efficiency of the selected wavelengths.

To evaluate MLR model accuracy, the statistical parameters used were the RMSE (root mean square error), as well as bias and R2 (coefficient of determination); the lower the error and the bias and the higher the R2 (as maximum equal to 1), the better the model performances. Besides, the RPD (ratio between the standard deviation of the response variable and RMSE) was calculated. An RPD between 1.5 and 2 means that the model can discriminate low from high values of the response variable; a value between 2 and 2.5 indicates that coarse quantitative predictions are possible, and a value between 2.5 and 3 or above corresponds to good and excellent prediction accuracy, respectively (Nicolaï et al., 2007). The whole set of parameters for the evaluation of model goodness were calculated not only in calibration but also in cross-validation (CV, internal validation method in which some samples are omitted from the calibration and used for validation) leave more out (venetian blinds with five cancellation groups).

RESULTS AND DISCUSSIONS

Figure 2 shows the sensors' readouts (12 wavelengths). The data have been handled by merging both sensors. However, even though the data do not come from a single sensor, the final output provide a complete profile which is comparable with the typical vis/NIR spectrum of dark grape berries as reported by Beghi et. al 2015.

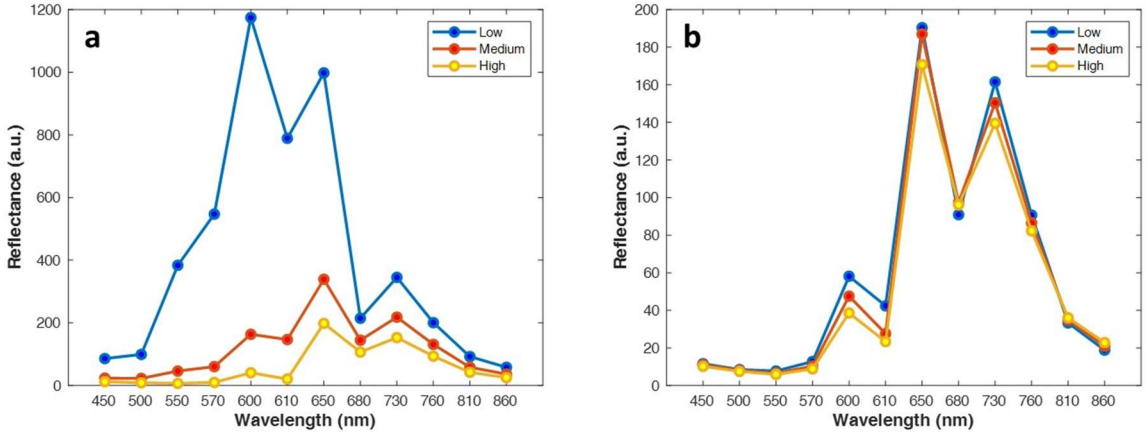


Figure 2. Sensors readouts on single berries (a) and bunches (b) grouped according to three arbitrary levels of soluble solids content (SSC) (Table 2).

The optical outputs come from two different sampling methods: on single berries (a) and bunches (b) which have generated two different datasets. The single berry method of analysis provides an optical output which comes from a single berry. Instead, the bunch method of analysis provides an optical output which comes from an average of five acquisitions performed randomly on the grape bunch.

Due to these two different sampling methods, two different reflectance intensities have been highlighted in figure 2. Indeed, higher reflectance values were obtained from the optical analyses on single berries (Figure 2a). The high optical output is related to complete green samples (very unripe berries). This feature is typical at the beginning of the ripening process where the concentration of anthocyanins is still low. Then, as the maturation progresses, the colour evolution due to the accumulation of phenolic compounds lead to lower absorbance values.

Concerning the bunch analysis, lower reflectance values have been obtained (figure 2b). The random averaging process has reduced the reflected light intensity between 550 and 650 nm due to the presence largely of already pigmented berry in the bunch (Bigard et al., 2019). In order to describe at best the data, the samples were labelled according to arbitrary levels of SSC (low, medium and high concentration) as described in Table 2.

Table 2. Arbitrary levels of Soluble Solids Content (SSC) used for sample grouping.

	Low	**Medium**	**High**
SSC of single berries (°Brix)	13.0-17.5	17.6-21.9	22.0-26.4
SSC bunches (°Brix)	16.5-22.5	22.6-24.0	24.1-25.9

Observing the readouts is evident how the reflectance values decrease as SSC interval considered increases and this is consistent with the grape ripening process

Figure 3 shows the PCA biplot used to explore the data structure deriving from the optical signals registered with the 12 wavelengths of the simplified LED based pre-prototype. The samples are divided into the same three levels of soluble solids content as reported in Table 2. In these plots, both scores for samples and loading for variables are represented in the space of the two principal components (PCs) obtained from PCA. Thus, the relationships between samples and variables can be seen in the same plot. The biplots show the behaviour of the reflectance data obtained from the single berries (a) and grapes bunches (b).

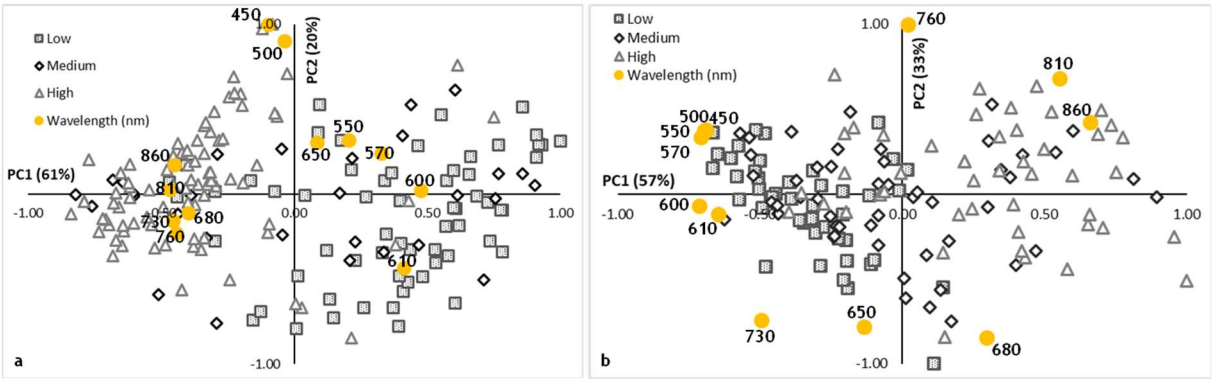


Figure 3. PCA biplot of the samples of single berries (a) and bunches (b) grouped according to three arbitrary levels of soluble solids content (SSC).

Regarding PCA (a), PC1 and PC2 explained 61% and 20% of the total variability, respectively. While, regarding the PCA (b), the first two PCs explained 90% of the total data variance together. In general, the PCs show the ripening trend which goes from high positive values to high negative values of PC1 for the berries readouts (a) and from negative to positive for the bunches readouts (b). In spite of that, it was clear how the different variables (wavelengths) describe the chemicals characteristics (SSC) of the samples. Overall, a low concentration of sugars corresponds to high reflectance values in the pure visible region and low reflectance in the SW-NIR region as can also be seen in figure 2. However, from the berry dataset, PCA shows a quite clear sample grouping (two clusters) described by the first principal component suggesting that the main information contained into this dataset is given by the colour evolution (from green to purple). The wavelengths between 570 and 610 nm are the most important to describe the yellow/green berries (SSC between 13.0-17.5 °Brix) while the wavelengths between 680 and 860 describe the second major group characterized by the totally pigmentated samples where the chlorophyll concentration which express the green colour is low. On the contrary, no sample grouping was expressed in the dataset of the grape bunches due to the absence of totally green samples. However, a clear trend is highlighted in PC1

confirming the maturation progress starting from already pigmented samples. Moreover, seeing both components together, a slightly influence related to the coupling of two different sensors suggest that a more balanced integration of the sensors to reduce this impact is needed.

At the same time, it was clear how according to the ripening process the SSC increase ad this is well described by the high values of reflectance obtained in the SW-NIR region.

Besides, in order to detect the presence of outliers, a 'Hotelling T2 computation' function (α value was set to 0.05) was included in score plots and revealed potential outliers lying outside the ellipse. In Table 3 was summarised the data of PCA results for the optical acquisition, the pre-treatments and the number of outliers detected.

Table 3. Principal Component Analysis (PCA) performed on the optical data of single berry and bunch datasets pre-treated using Standard Normal Variate (SNV) and autoscaling. The amount of explained variance were reported in percentage for the first and the second Principal Component (PC1 and PC2).

	Pre-treatments	**PC1 explained variance**	**PC2 explained variance**	**Outliers**
180 (Single berries)	SNV +	61%	20%	6
150 (Bunches)	autoscaling	57%	33%	8

In Table 4 are summarized the descriptive statistics related to the reference analyses for grape ripening parameters (SSC, TA, pH and EA). The dataset, based on the individual berries, showed a lower mean value for SSC than that referred to the bunches. This was possible thanks to the collection of totally unripe to very ripe single berries, while for the bunches the experimentation began on samples already pigmented and closest to ripeness. For the same reason, the bunches did not show a large variability for the four parameters considered.

Table 4. Descriptive statistics of the sample of single berries and grapes bunches for ripeness parameters (SSC, TA, pH and EA).[1]

Sample	**Property**	**Units**	**No.**	**Mean**	**Median**	**SD**	**Min.**	**Max.**
Berry	*SSC*	°Brix	180	20.6	21.6	4.0	13.0	26.4
Bunch	*SSC*	°Brix	150	22.9	23.2	1.9	16.5	25.9
Bunch	*TA*	g tartaric acid dm^{-3}	150	6.9	6.7	1.2	4.8	10.2
Bunch	*pH*		150	3.33	3.34	0.12	3.06	3.67
Bunch	*EA*	%	150	36.30	37.01	7.64	18.86	52.64

[1] SSC = soluble solids content, TA = titratable acidity, EA = extractible anthocyanins, SD = standard deviation, Min. = minimum and Max. = maximum.

In Table 5 are reported the MLR models obtained by the two pre-prototypes system for grape ripening parameters (SSC, TA, pH and TP). In the table is also reported the values of RPD defined as the ratio of the standard deviation of the response variable to RMSECV.

Table 5. Figure of merit of the Multiple Linear Regression (MLR) models (for grape berries and bunches) calculated using optical data pre-treated with Standard Normal Variate (SNV) and autoscaling.[2]

Sample	Property	N°	Treatment	R^2_{cal}	RMSEC	r^2_{CV}	RMSECV	RPD
Berry	*SSC*	174	SNV + autoscaling	0.88	1.37	0.86	1.51	2.65
Bunch	*SSC*	142	SNV + autoscaling	0.50	1.19	0.40	1.31	1.45
Bunch	*TA*			0.57	0.75	0.47	0.83	1.44
Bunch	*pH*			0.55	0.08	0.45	0.09	1.33
Bunch	*EA*			0.59	4.72	0.50	5.21	1.46

[2] N° = number of samples; R^2_{cal} = coefficient of determination in calibration; RMSEC = root mean square error of calibration; r^2_{cv} = coefficient of determination in cross-validation; RMSECV = root mean square error of cross validation.

For SSC, a r2cv = 0.86 and 0.40, RMSECV = 1.51 and 1.31 and RPD = 2.65 and 1.45 for single berries and bunches datasets have been obtained, respectively. SSC prediction showed better correlation and a comparable error for the berry dataset compared to bunch dataset. This behaviour could be justified by the one-to-one correlation (optical output and reference analysis) used for model calculation on single berry and by a wider ripening range available for the single berry's dataset. Instead, the optical measurement on bunches showed low r2cv due to the lower variability which the bunch dataset has considered. However, the relatively slightly lower error for SSC prediction expresses the capability of the sensors to be able to obtain better results using a more complete dataset which includes samples coming from the whole maturation process.

Instead, concerning TA, pH and EA models (from the bunch dataset) the r2cv are poor due to the low variability covered (as for the SSC model). However, the RMSECV and the RPD show the capability of the models to be enough accurate to allow an initial field screening of the trend of these parameters.

CONCLUSIONS

In this work, a pre-prototype of a simplified optical vis/NIR device (12 wavelengths from 450 nm to 860 nm) based on LED technology was tested for rapid estimation of the ripening

parameters of Nebbiolo grape. Correlations between the optical data matrix and some of the most interesting ripening parameters were carried out using the MLR method.

Overall, very interesting results were obtained for the prediction of SSC. Instead, concerning TA, pH and EA models are still poor due to the low variability covered and further experimental activities with larger datasets will certainly be necessary. However, the research has shown the capability of the models to be enough accurate to allow an initial field screening of the maturation trend of these parameters.

This new generation of optical devices could be a starting point to build a new concept of inexpensive IoT sensors which could be distributed in the vineyard at fruit set (open bunch) and that will then be enveloped by the grape bunch as the grape grows on it. The stand-alone instrument should be able to acquire and predict the most important ripening parameters directly from measurements in field inside the bunch. This approach could shift the current paradigm of grape maturation monitoring bringing the laboratory directly into the vineyard, without human intervention for data sampling.

The integration of simple multivariate models in the microcontroller software would easy calculate and visualize the real-time values of the predicted parameters in the cloud (online data bank). The cost of the device is expected to be fairly low (about few tens of euros) in order to install many devices in crucial points of the vineyard to obtain an average value related to the ripening status of the whole vineyard.

Therefore, further studies both for model improvement and for the design of the system in the vineyard that allows automated acquisitions, the remote control of the modules, and remote data sharing through cloud storage system, are needed. In a view of viticulture 4.0, a similar tool, in combination with the weather stations, will be able to lead to precise and rapid analyses that will be capable to guarantee better monitoring of the ripening process in order to predict the best moment of harvest and provide to the wineries grapes with specific features in terms of quality attributes.

Acknowledgements

The authors are also grateful to Officina delle Soluzioni s.n.c. for the technical support and Azienda Agricola Renato Fenocchio for the data collection.

References

Aleixandre-Tudo, J. L., Nieuwoudt, H., & du Toit, W. (2019). Towards on-line monitoring of phenolic content in red wine grapes: A feasibility study. Food chemistry, 270, 322-331.

Baca-Bocanegra, B., Hernández-Hierro, J. M., Nogales-Bueno, J., & Heredia, F. J. (2019). Feasibility study on the use of a portable micro near infrared spectroscopy device for the “in vineyard” screening of extractable polyphenols in red grape skins. Talanta, 192, 353-359.

Barnes, R. J., Dhanoa, M. S., & Lister, S. J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Applied spectroscopy, 43(5), 772-777.

Beghi, R., Giovenzana, V., Marai, S., & Guidetti, R. (2015). Rapid monitoring of grape withering using visible near-infrared spectroscopy. Journal of the Science of Food and Agriculture, 95(15), 3144-3149.

Bigard, A., Romieu, C., Sire, Y., Veyret, M., Ojeda, H., & Torregrosa, L. (2019). The kinetics of grape ripening revisited through berry density sorting. OENO One, 53(4).

Casson, A., Beghi, R., Giovenzana, V., Fiorindo, I., Tugnolo, A., & Guidetti, R. (2019). Visible Near Infrared Spectroscopy as a Green Technology: An Environmental Impact Comparative Study on Olive Oil Analyses. Sustainability, 11(9), 2611.

Donis-González, I. R., Valero, C., Momin, M. A., Kaur, A., & C Slaughter, D. (2020). Performance Evaluation of two commercially available portable spectrometers to non-invasively determine table grape and peach quality attributes. Agronomy, 10(1), 148.

dos Santos Costa, D., Mesa, N. F. O., Freire, M. S., Ramos, R. P., & Mederos, B. J. T. (2019). Development of predictive models for quality and maturation stage attributes of wine grapes using vis-nir reflectance spectroscopy. Postharvest biology and technology, 150, 166-178.

Eurostat. (2019). https://ec.europa.eu/eurostat/web/products-eurostat-news/-/edn-20201119-2

Fletcher, R. S., & Fisher, D. K. (2018). A Miniature Sensor for Measuring Reflectance, Relative Humidity, and Temperature: A Greenhouse Example. Agricultural Sciences, 9(11), 1516.

Giovenzana, V., Beghi, R., Tugnolo, A., Brancadoro, L., & Guidetti, R. (2018). Comparison of two immersion probes coupled with visible/near infrared spectroscopy to assess the must infection at the grape receiving area. Computers and electronics in agriculture, 146, 86-92.

Giovenzana, V., Civelli, R., Beghi, R., Oberti, R., & Guidetti, R. (2015). Testing of a simplified LED based vis/NIR system for rapid ripeness evaluation of white grape (Vitis vinifera L.) for Franciacorta wine. Talanta, 144, 584-591.

Giovenzana, V., Tugnolo, A., Casson, A., Guidetti, R., & Beghi, R. (2019). Application of visible-near infrared spectroscopy to evaluate the quality of button mushrooms. Journal of Near Infrared Spectroscopy, 27(1), 38-45.

Khanal, S., KC, K., Fulton, J. P., Shearer, S., & Ozkan, E. (2020). Remote sensing in agriculture—accomplishments, limitations, and opportunities. Remote Sensing, 12(22), 3783.

Lee, J., Bagheri, B., & Kao, H. A. (2015). A cyber-physical systems architecture for industry 4.0-based manufacturing systems. Manufacturing letters, 3, 18-23.

Livingstone, D.J., Salt, D.W., and Chemquest, (2005) J. Med. Chem 48, 43.

Magwaza, L.S., Linus Opara, U., Nieuwoudt, H., R Cronje, P.J., Saeys, W., and Nicolaï, B. NIR Spectroscopy Applications for Internal and External Quality Analysis of Citrus Fruit-A Review.

Malegori, C., Grassi, S., Ohm, J. B., Anderson, J., & Marti, A. (2018). GlutoPeak profile analysis for wheat classification: Skipping the refinement process. Journal of Cereal Science, 79, 73-79.

Marini, F., de Beer, D., Walters, N. A., de Villiers, A., Joubert, E., & Walczak, B. (2017). Multivariate analysis of variance of designed chromatographic data. A case study involving fermentation of rooibos tea. Journal of Chromatography A, 1489, 115-125.

Mena, A., Civelli, R., Guidetti, R., Best, S., Leòn Gutiérrez, L. F., Giovenzana, V., & Beghi, R. (2012). Quick Quality Evaluation of Chilean Grapes by a Portable vis/NIR Device. In I International Workshop on Vineyard Mechanization and Grape and Wine Quality 978 (pp. 93-100).

Miranda, J., Ponce, P., Molina, A., & Wright, P. (2019). Sensing, smart and sustainable technologies for Agri-Food 4.0. Computers in Industry, 108, 21-36.

Nicolai, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. Postharvest biology and technology, 46(2), 99-118.

Ribera-Fonseca, A., Noferini, M., Jorquera-Fontena, E., & Rombolà, A. D. (2016). Assessment of technological maturity parameters and anthocyanins in berries of cv. Sangiovese (Vitis vinifera L.) by a portable vis/NIR device. Scientia horticulturae, 209, 229-235.

Ribéreau-Gayon, P., Glories, Y., Maujean, A., & Dubourdieu, D. (Eds.). (2006). Handbook of enology, volume 2: the chemistry of wine-stabilization and treatments (Vol. 2). John Wiley & Sons.

Still, C., Powell, R., Aubrecht, D., Kim, Y., Helliker, B., Roberts, D., Richardson, A.D., Goulden, M. (2019). Thermal imaging in plant and ecosystem ecology: applications and challenges. Ecosphere, 10(6).

Todeschini, R. (1998). Introduzione alla chemiometria: strategie, metodi e algoritmi per l'analisi e il modellamento di dati chimici, farmacologici e ambientali. EdiSES.

Tugnolo, A., Beghi, R., Giovenzana, V., & Guidetti, R. (2019). Characterization of green, roasted beans, and ground coffee using near infrared spectroscopy: A comparison of two devices. Journal of Near Infrared Spectroscopy, 27(1), 93-104.

Tugnolo, A., Giovenzana, V., Beghi, R., Grassi, S., Alamprese, C., Casson, A., Casiraghi, E. Guidetti, R. (2021). A diagnostic visible/near infrared tool for a fully automated olive ripeness evaluation in a view of a simplified optical system. Computers and Electronics in Agriculture, 180, 105887.

Vallone, M., Alleri, M., Bono, F., & Catania, P. (2019). Quality evaluation of grapes for mechanical harvest using vis NIR spectroscopy. Agricultural Engineering International: CIGR Journal, 21(1), 140-149.

4.3.2 Stand-alone optical device development

Concerning the sensors development, a first step has been done designing optical sensors managed by the operators in the grape maturation control topic.

However, the technology is in sharp development and is reaching a considerable level of miniaturization. The consumer electronics industry is driving the convergence of digital circuitry, wireless transceivers, and microelectro-mechanical systems (MEMS), which makes it possible to integrate sensing, data processing, wireless communication, and power supply into low-cost millimeter-scale devices (Spachos and Gregori, 2019). The resulting miniaturization and cost reduction of electronic components is leaving space to a completely new method of data acquisition and management using wireless sensor networks (WSNs) based on small battery-powered nodes.

A WSN consist of small and low-cost Internet of Things (IoT) devices in a network of peripheral nodes. The nodes are equipped with sensors and a wireless module for data transmission to an online database, where the data are stored and accessible to the end-user. The nodes are energy independent and are installed in areas which are more representative of the vineyard variability (Spachos, 2020). WSN technology is widely used to monitor environmental factors (temperature, moisture, relative humidity, leaf wetness etc.) which are essential for the growers decision making (Patil and Thorat, 2016).

Therefore, the final stage of this PhD journey is the development of a fully integrated, small, cost-effective, stand-alone smart system used for grape maturation monitoring.

The main steps to fulfil the PhD project purpose were:

a) setting up of a miniaturized low-cost and stand-alone optical prototype
b) build multivariate predictive models for the prediction of the main grape ripening parameters;
c) test the prototype in field conditions.

The device consisted of an optical detection head (flexible strip or transparent canopy) connected into the grape bunch, including power, signal pre-processing, and wireless communications. The detection head is optically based on the integration of LED sources and photodiode/interference filter arrays (in the vis/NIR optical range) at wafer level or wafer package level. The sensor concept consisted of an optical detection head connected to the top of the grape bunch, including power, signal pre-processing, and communications. Reflectance measurements were used at various wavelengths to probe spectral signatures

for both technological (pH, total acidity and total soluble solids) and phenological (phenols and anthocyanins) maturation.

As already highlighted in Paper 1 and 7, the literature reports different approaches based on both fluorescence and reflectance principles for grape maturation monitoring. This systematic analysis of the literature has made possible to identify the optical models or indexes that are more suited to be used as predictors of different parameters that are controlled during the maturation (classically using wet chemistry).

At this stage, the bibliographic analysis on grape maturation control was complemented by the analysis of historical datasets from Sogrape (Portuguese winery that has collaborated in the development of these sensors) and data coming a previous experience (which includes detailed optical information and the respective chemical measurements from the cv. Nebbiolo, highlighted in Paper 7). This added a second layer of detailed information about the most promising reflectance bands that could be used in this new generation of prototypes totally designed in house.

SPECIFICATIONS

Even though a first design of optical sensors was presented (Paper 7), a new conceptualization of the sensor has been done thanks to the different features which the sensor should have in order to remain in the vineyard and pick data autonomously. Therefore, a more detailed bibliographic analysis has been done in order to have a more comprehensive idea about which wavelengths (thus, LEDs and photodetectors) to includes in the device. Tables 4.1 and 4.2 show an overview of the indices used to predict many interesting qualitative parameters of the grape. Several methods were identified in terms of optical techniques (reflectance and fluorescence) using ratio between wavelengths or a multivariate statistical approach which is required to extract the main information about quality attributes in the vis/NIR region. The multivariate approach is strongly required to manage optical data because the spectrum may be dependent by scattering effects, tissue heterogeneities, instrumental noise, ambient effects and other sources of variability. As a consequence, it is difficult to assign specific absorption bands to specific functional groups let alone chemical components (Nicolai et al., 2007; Wold et al., 2001).

Table 4.1 Reflectance bibliography indices; “R” is the percentage of reflectance of each wavelength,“I”is the intensity of each wavelength and “R^2” is the coefficient of determination.

Component	Formula (nm)	R^2	Reference
Total chlorophylls	R_{800}/R_{678}	0.59	Merzlyak, M. N.; Solovchenko, A. E.; Gitelson, A. A.; 2003b: Reflectance spectral features and non-destructive estimation of chlorophyll, carotenoid and anthocyanin content in apple fruit. Postharvest Biol. Technol. 27, 197-211.
Chlorophyll a (C_aI)	$\log[(R_{800}/R_{675})-R_{800}/R_{660})]$	0.66	Rocchi, L., Rustioni, L., & Failla, O. (2016). Chlorophyll and carotenoid quantifications in white grape (Vitis vinifera L.) skins by reflectance spectroscopy. Vitis, 55(1), 11-16.
Chlorophyll b (C_bI)	$\log[(R_{800}/R_{650})-R_{800}/R_{630})]$	0.40	
Total chlorophylls (C_{tot}I)	C_aI + C_bI	0.62	
Carotenoids (CarI)	a*C_aI + b	0.72	
Total Polyphenols (TP)	$60.75*I_{670} - 254.93*I_{730} + 382.73*I_{780} - 162.20$	0.70	Giovenzana, V., Beghi, R., Malegori, C., Civelli, R., & Guidetti, R. (2014). Wavelength selection with a view to a simplified handheld optical system to estimate grape ripeness. American Journal of Enology and Viticulture, 65(1), 117-123
Solids Soluble Content (SSC)	$13.15*I_{670} - 38.55*I_{730} + 37.81*I_{780} + 15.69$	0.71	
SSC	Ratio between 560 & 640	0.92	Ribera-Fonseca, A., Noferini, M., Jorquera-Fontena, E., & Rombolà, A. D. (2016). Assessment of technological maturity parameters and anthocyanins in berries of cv. Sangiovese (Vitis vinifera L.) by a portable vis/NIR device. Scientia Horticulturae, 209, 229-235.
Titratable Acidity (TA)		0.87	
Durofel Index (DI_firmness)		0.89	
Monomeric and total antocyanin		0.68 - 0.97	

Table 4.2 Fluorescence bibliography indices

Component	Excitation (nm)	Emission (nm)	R^2	Reference
Anthocyanins	550 (green) and 650 (red)	740	0.93	Agati, G., Traversi, M. L., & Cerovic, Z. G. (2008). Chlorophyll fluorescence imaging for the noninvasive assessment of anthocyanins in whole grape (Vitis vinifera L.) bunches. Photochemistry and Photobiology, 84(6), 1431-1434.
Anthocyanins	515 (green) and 630 (red)	red (RF) and far-red (FRF)	0.75 - 0.83	Agati, G., D'Onofrio, C., Ducci, E., Cuzzola, A., Remorini, D., Tuccio, L., ... & Mattii, G. (2013). Potential of a multiparametric optical sensor for determining in situ the maturity components of red and white Vitis vinifera wine grapes. Journal of agricultural and food chemistry, 61(50), 12211-12218.
Anthocyanins	516 (green) and 637 (red)	red (680-690) and far-red (730-780)	0.87	Tuccio, L., Remorini, D., Pinelli, P., Fierini, E., Tonutti, P., Scalabrelli, G., & Agati, G. (2011). Rapid and non-destructive method to assess in the vineyard grape berry anthocyanins under different seasonal and water conditions. Australian Journal of Grape and Wine Research, 17(2), 181-189.
Anthocyanins	516 (green) and 637 (red)	red (680-690) and far-red (730-780)	0.98	Cerovic, Z. G., Moise, N., Agati, G., Latouche, G., Ghozlen, N. B., & Meyer, S. (2008). New portable optical sensors for the assessment of winegrape phenolic maturity based on berry fluorescence. Journal of Food Composition and Analysis, 21(8), 650-654.
Flavonoids	UV (375)	red (680-690) and far-red (730-780)	0.91	

The literature analysis shows a redundancy of the use of specific wavelengths for the prediction of the several parameters which describe the maturation process. Therefore, considering the most interesting and used wavelengths in the bibliography and the physical characteristics of the optical prototype, the first generation of optical modules will be designed in order to accommodate a compromise among these requirements.

Moreover, the possibility of applying fluorescence spectroscopy is particularly interesting for the evaluation of phenolic parameters (anthocyanins and flavonoids). In the Vis-NIR spectral range (400-800 nm) it is possible to exploit spectral bands for the excitation and the signal emission. In particular, it is possible to exploit the excitation bands of green (520-550 nm) and red (630-640 nm) and the emission bands of red (680-690 nm) and the far-red (730-780 nm). A robust literature is published on the prediction of anthocyanins content based on fluorescence acquisition at these spectral bands, reporting encouraging results (Table 4.2).

Regarding flavonoids, it could be necessary to move to UV spectral range for the excitation (around 375 nm). The evolution of the literature based shows that there is a trend of using the far-red band for detecting the emission signal from chlorophyll. Although this band has lower

sensitivity that the red band, it is less exposed to the absorption peak of the chlorophyll at 680 nm, being less prone to inteferences from the auto-absorption of the chlorophyll.

Historical dataset analysis

Grape harvest decisions are critical for wine quality and relies upon the analysis of SSC (or TSS), organic acids, pH, berry weight, total polyphenol content, anthocyanin and tannins, evolution during the grape maturation period. Table 4.3 shows a list of the most important wet-chemistry parameters used to assess grape maturation quality (Table 4.3).

Table 4.3 Wet-chemistry parameters used to assess the grape quality.

Parameter	Reference values	Current periodicity	Ideal periodicity
Total Sugars	0-300 g/L	Weekly	Weekly until 170 g/L Daily after 170 g/L
Berry weight	0-3000 mg	Weekly	Weekly until sugars = 170 g/L Daily after above criteria
pH	2.5 - 4.5	Weekly	Daily
Total Acidity	0 - 20 g/L ac. Tartárico	Weekly	Weekly until sugars = 170 g/L Daily after above criteria
Total Antocyanins	0 - 2000 mg/L	Weekly	Daily
Total Polyphenols	0 - 150	Weekly	Daily
Assimilable Nitrogen	0 - 500 mg/L	Once/harvest	Weekly
Aromatic precursors (terpenes, thiols, pyrazine, etc.)	Currently not used		Daily, after de end of sugar accumulation

The literature highlighted TSS and the total acidity (TA) as the qualitative parameters most used by the vine-growers to determinate the harvesting date (Watson, 2003). Moreover, a strong effect of the phenolic compounds (polyphenols and anthocyanins) about the quality of the grape and wine (astringency, body, etc.) was recognized.

To confirm these statements and identify the qualitative parameters to be measured, a wet-chemical historical dataset provided by Sogrape was analysed. The dataset includes the above mentioned relevant wet-chemistry parameters used to assess grape maturation of samples, collected on a weekly-basis during grape ripening season, from different grape varieties (cv. Touriga Nacional, cv. Touriga Franca, cv. Tinta Roriz, cv. Tinta Barroca, cv. Tinto Cão) grown in different and well-characterized vineyard-blocks of the project's pilot-vineyard

(Sograpе's owned vineyards located in the Douro region), covering the time-period 1991-2018. No grapevine spectral reflectance is provided for this dataset.

Considering the large amount of data provided by the historical dataset of Sograpе, a chemometrics approach was followed to pick-up at best the information contained into the data. A qualitative Principal Component Analysis was performed using the Sograpе dataset with the goal to identify the qualitative parameters which show the highest variability during the ripening process (Figure 4.1).

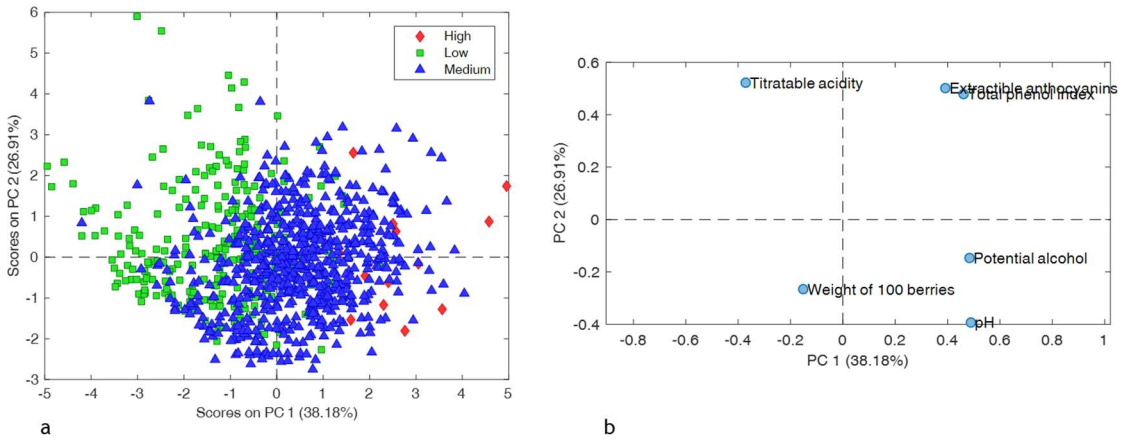


Figure 4.1 PCA (a) scores plot coloured according to increasing values of concentration of solids soluble arbitrary defined (Low 6.9-11.6 °Brix; medium 11.7-16.3 °Brix; High 16.4-21.1 °Brix); (b) loadings plot of the wet-chemical variables described by PC1 and PC2 .

Figure 4.1a shows the behaviour of the samples in the space of the Principal Components (PC). The samples were labelled according to the concentration of solids soluble content (SSC) arbitrary divided into three classes (Low 6.9-11.6 °Brix; medium 11.7-16.3 °Brix; High 16.4-21.1 °Brix). As already described, the SSC is one of the most important quality parameters which describe the maturation process (Liu, H. F. et al., 2006) and constantly increase from the veraison till the end of the ripening process and, therefore, the moment of harvest. PC1 and PC2 together explained 65.09% of the total variability with a clear description of the ripening process from negative and positive values of PC1 and PC2 respectively and to positive and negative values of PC1 and PC2 respectively. Instead, to identify the most important parameters which describe the maturation process the loadings plot (Figure 1b) was analysed. PCA confirmed the results obtained from the analysis of the bibliography. Thus, potential alcohol (quantifiable also with TSS) and titratable acidity are the two most important inversely related parameters which describe the maturation process and the extractible anthocyanins and total phenols (directly related) are interesting for the quality of the wine.

Reflectance wavelengths definition

In summary, based on the analysis of the historical datasets from Sogrape, on the information reported in the literature, and on the previous experiments using data from Nebbiolo variety (Paper 7), the most interesting parameters which have been considered for building predictive models for the ripening process were:

- Total solid soluble (TSS);
- potential alcohol (PA);
- titratable acidity (TA);
- polyphenols;
- anthocyanins.

Additionally, regarding the definition of the optical bands of interest, it was considered: (a) the wavelengths used in Paper 7, (b) the wavelengths most reported in literature, and (c) the availability of LED (Light emitting Diodes) light sources and the simplicity of the arrangement of the LEDs and PDs (Photodiodes) on the layout of the prototype. Therefore, four wavelengths were defined:

- 530 nm;
- 625 nm;
- 690 nm;
- 730 nm.

Two different configurations for the modules were envisioned: (i) reflectance - white and red grapes, modules with 4 wavelengths (4 LEDs and 4 photodetectors, PDs) for prediction of TSS, PA, TA and phenolics compounds and (ii) fluorescence for the prediction of phenolics compounds (which is still under investigation).

Sensor specs

General assembling and use

Concerning miniaturization and usability requirements in field, it was developed a “stripe” design in which the spectrometric components were mounted on a long, flexible substrate which can be placed onto or inside the grape bunch. The multiple spectrometers were placed along the stripe (currently 2, module 1, M1 and module 2, M2), enabling simultaneous measurements at different parts of the grape bunch to have more representative information of the entire bunch. Therefore, four light-emitting diodes (LEDs) were used for illumination of the grape (530 nm, green, 630 nm, orange, 690 nm, red and 730 nm, red). Placed close to these, but optically isolated using an opaque barrier, four photodiodes (with an active area of 520 × 520 μm^2) assembled with spectral filters to allow intensity measurements at the desired wavelengths (the relative spectra sensitivity is reported in figure 1b) have been used. The components were encapsulated in a hermetically sealed yet optically transparent layer, assuring weatherproof operation of the entire system and reducing stray light (Figure 4.2).

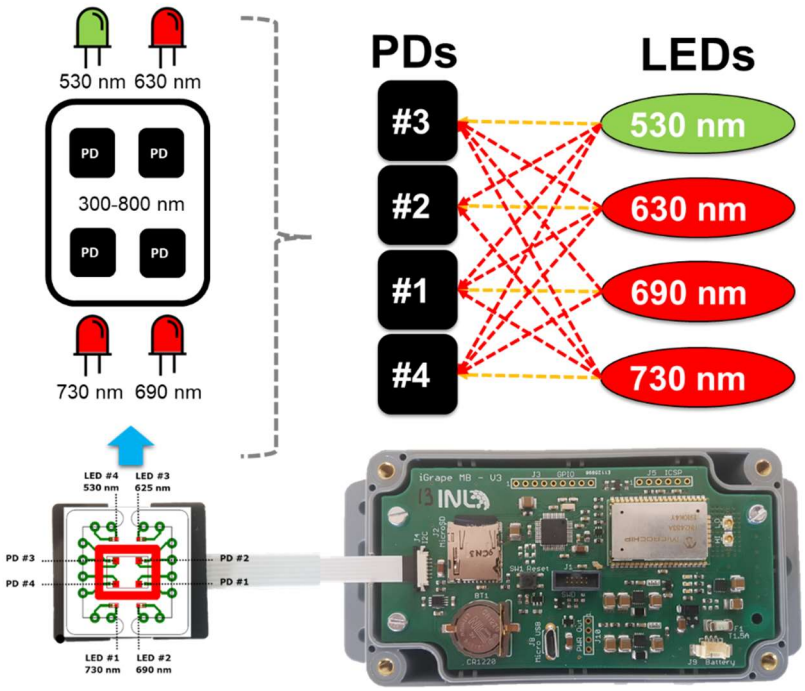


Figure 4.2 Sensor specifications.

The light emitted from each LED hits the sample and the diffuse reflectance light is collected by each PD. The electromagnetic signal is converted into electronic signal and expressed in counts (arbitrary units from 0 to 4095). From each sample, 20 readouts were obtained (one readout from each PD at each LED on and one with LEDs off for background info).

Finally, the device internal software was set in order to return the averaged optical values coming from 16 acquisitions.

Concerning the data transmission, the sensors were configurated in order to share the optical outputs coming from the sensors placed in a target parcel to a local IoT gateway LoRaWAN (able to receive data from the sensors in a communication range about 15 km). Then, the local server broadcasts the data to a web app (IoT database) in order to process the data with the chemometric models developed during the sampling campaign 2020 using samples analysed in lab and direct in field (Figure 4.3).

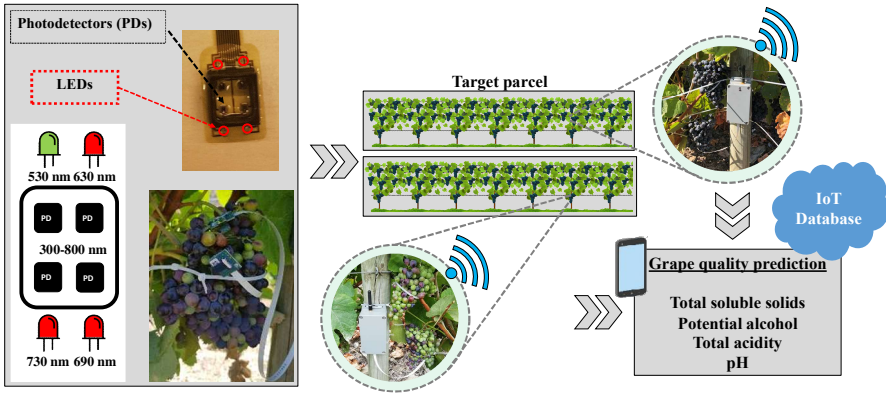


Figure 4.3 *Sensor technical features, installation and data transmission.*

LEDs

LEDs were the excitation light chosen for both reflectance and fluorescence measurements. The selected wavelengths are 530 nm, 630 nm, 690 nm and 730 nm. As already described, these bands are associated with the evolution of the maturation of the grapes.

LEDs are commonly manufactured in other processes than CMOS, like AlInGaP which makes integration only possible at packaging level. Inexpensive LED light sources can thus be found in die format (wafer). Due to the non-uniformities at wafer-level fabrication, LEDs in the same wafer have large variations of its peak wavelength and therefore LED manufacturers do the LED binning guaranteeing that LEDs in that bin are within a certain wavelength range (5nm). As a consequence it was no possible to obtain the exact peak wavelength as defined above but close enough to the requirements of reflectance and fluorescence.

LED dies are supplied in bulk, all with similar chip size dimensions at approximately 12mil (~313µm side). The following selection were the available bins provided by the manufacturers.

- CREE C527E2290 with average peak at 530.3 nm, Vf = 3V, DC forward current 50mA
- C4L C4L-S12T5 with average peak at 637.121 nm, Vf = 2.2V, DC forward current 70mA

- LA LAAI13WP3 with average peak at 685.7nm, Vf = 2.2V, DC forward current 70mA
- LA LACI13WP3 with average peak at 730.7nm, Vf = 1.55V, DC forward current 100mA

Photosensors

The photosensors had a sensitivity from 500 nm up to 800 nm in order to be combined at best with the wavelengths of interest either for diffuse reflectance or fluorescence measurements.

For the discrete module assembly, a commercial photodetector with the following specification has been used. The LA PD28AP1 Light Avenue Premium Edition detector series is designed for high performance consumer applications. This chip is a high speed and highly sensitive PIN photodiode chip with 0.27 mm^2 sensitive area detecting visible light similar to the human eye and a peak sensitivity at 560 nm. Figure 4.4 shows the spectral sensitivity which is in line with spectral band requirements mentioned above.

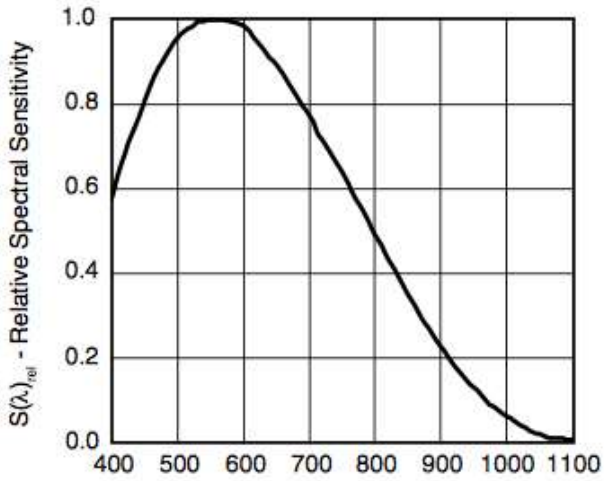


Figure 4.4: Spectral sensitivity for the LA PD28AP1

Spectrometer architecture and configuration

The architecture of the spectrometer has three main components:

1. The optical module
2. The host module that carries the optical module
3. The controller board and power management

The optical module has a combination of LEDs, Photodetectors. The assembly process was based on a conventional Chip-On-Board using a conductive resin to attach the LEDs and photodetectors to PCB (through the anode terminal). A single wire bonding completes the assembly process connecting the cathode to the PCB pad. A transparent glob top is dispensed

and cured on top of the LEDs and Photodetectors to protect dies and bond wires. Filters can be either directly fabricated onto the photodetectors or if needed incorporated onto the optical frame/mount. Figure 4.5 shows module assembled.

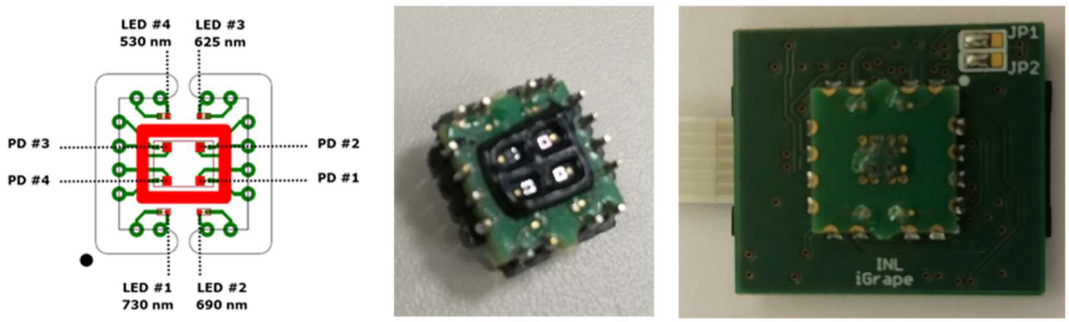


Figure 4.5 Optical module. Left: PCB layout , Center: optical module assembled with LEDs and Photodiodes/filters, Right: Optical module assembled in host module.

The host module is a generic and small PCB which was designed to be small in size as it would need to fit in the grape bunch. It includes a power down switch preventing the modules to draw power.

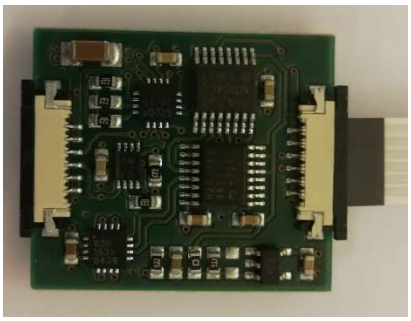


Figure 4.6 Host module (LED driver and Photodiode signal acquisition).

As mentioned before, the photodiode is biased in reverse mode and was used a low offset and drift, low-noise OPAMP from AMS (AS89000). This part provides 4 channel TIA and thus the number of channels necessary for the sensor which carries 4 photodetectors. The transimpedance (gain) can be set in 8 stages and its range is: 25kΩ, 100 kΩ, 500 kΩ, 1MΩ, 2MΩ, 5MΩ, 10MΩ and 20MΩ. As for the ADC the choice was on the Maxim MAX11614 with 12-bit resolution. This ADC had single supply operation at 5V.

The controller module (Figure 4.7) is based on an ARM Cortex-M0+ microcontroller running at 48MHz. This Microcontroller had ultra-low power performance consuming under 35 μA/MHz in active mode and 200nA in Sleep mode. The controller module also carried, battery

management (charger and step-up DC-DC), SD card data storage and a Lorawan module for wireless communications support (IoT).

The controller module connects to host module via I2C bus and can support up to 4 modules with addresses physically set by two bits in the host module PCB.

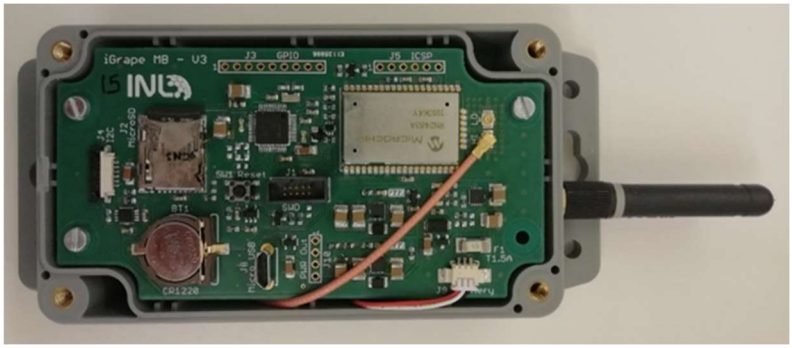


Figure 4.7 Controller module assembled.

Data collection for models 'computation

Reference analysis

The reference analyses of: (i) Total Soluble Solids (TSS, °Brix) using a digital pocket refractometer (PAL-1 ATAGO, Japan), (ii) Potential Alcohol (PA, % vol), (iii) Titratable Acidity (TA, $g_{tartaric\ acid}\ L^{-1}$) using an automatic titrator (TitroMatic KF 1S, Crison Instruments, Italy), (iv) Total Poliphenols (TP, mg dm^{-3}) calculated by means of the Folin-Ciocalteu method (Ough & Amerine, 1988), (v) Extractable Anthocyanins (EA, mg dm^{-3}) calculated by means of Glories method (Ribéreau-Gayon *et al.*, 2006) based on the Optical Density (OD) measurement at 280 and 520 nm using an UV/vis spectrophotometer and (vi) pH (pH meter, PCE Inst. GmbH, Germany) were performed.

Optical analysis

The optical acquisitions were managed considering: (i) the noise produced by environmental light, (ii) the physical features of the grape berries, (iii) the increase in size of the bunch, (iv) the variable optical gap and (v) the position of the field sensor which change during the ripening process. In order to minimize these issues and maximize the info picked from the sensors, the measurements were performed overnight (for in field analysis, from midnight to 5 a.m.) and using a dark room (for in lab analysis). The readouts coming from each PD were analysed individually as a variable in order to obtain the maximum information that can be collected.

Therefore, a total of 16 optical variables (light emitted by 4 LEDs reads by 4 PDs) were used for the model building (Figure 4.2).

Sampling methodology

The experimental activity took place in the viticulture area of the Douro Valley (Sograpе's owned Quinta do Seixo, Portugal) from the end of July to mid-September for a total of six sampling dates. Sampling was performed on *cv.* Touriga Nacional (TN) and Touriga Franca (TF) (from 18 parcels placed in different areas of the vineyard in order to be representative at best of the maturation stage of the whole vineyard) using the LED fully integrated stand-alone prototype (without any sample preparation) and wet-chemical analyses described above.
To summarize, figure 4.8 shows the experimental setup to analyse each sample by the sensor and the conventional wet-chemistry assays. Therefore, from a random sampling (200 berries for each sample) performed in the pilot-vineyard, 30 berries were analysed optically and the entire amount of 200 berries (with the 30 included) were then analysed using the classic protocol based on wet chemistry assays adopted by the winery (Sograpе).

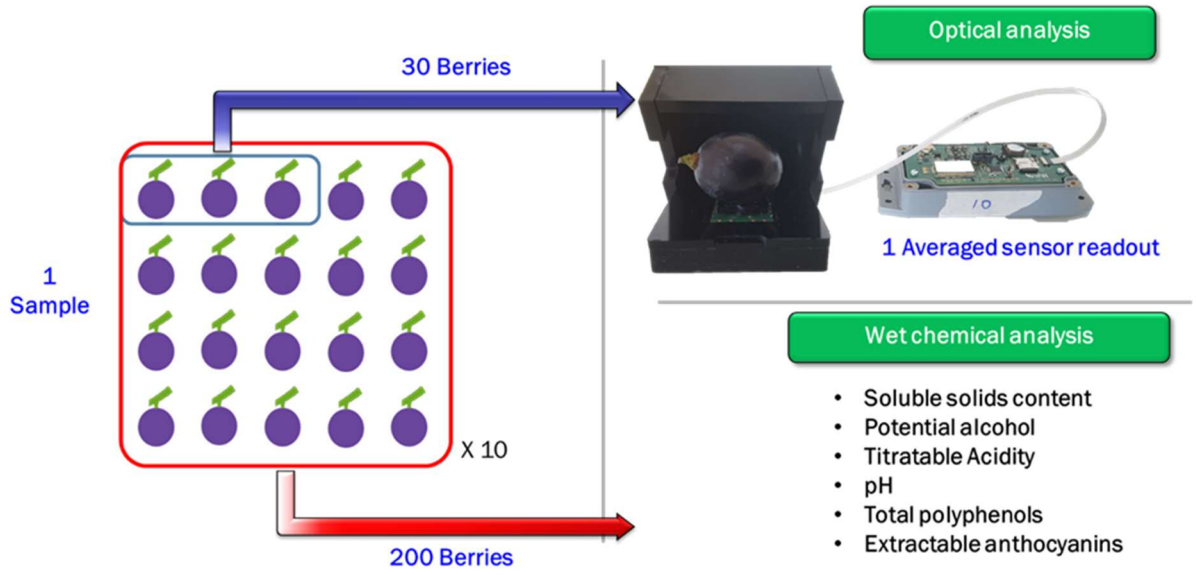


Figure 4.8 Sampling experimental setup used during the lab-scale analysis.

Modelling

A multivariate analytical approach was followed in order to explore the variability contained into the data using PCA. A latent variable modelling using the PLS method, which maximizes the covariance among the sensor readouts and the reference qualitative parameters (TSS, PA, pH and TA), was performed (Oliveri *et al.*, 2019). Model accuracy was evaluated (in calibration, cross-validation and in prediction) using the RMSE (root mean square error), as

well as bias and R^2 (coefficient of determination); the lower the error and the bias and the higher the R^2 (as maximum equal to 1), the better the model performances. The data elaboration was performed in Matlab environment (The MathWorks. Inc, USA) using both PLS-Toolbox package (Eigenvector Research, Inc. Manson, Washington) and in-house functions.

Models computation

Figure 4.9 shows the optical outcome from the lab and field sensor (figure 4.9b and 4.9c) used for modelling. While for a better understanding of the grape optical behaviour, the general mean readout coming from the reading by each PD at all LEDs (including LED off) have been reported in figure 4.9a.

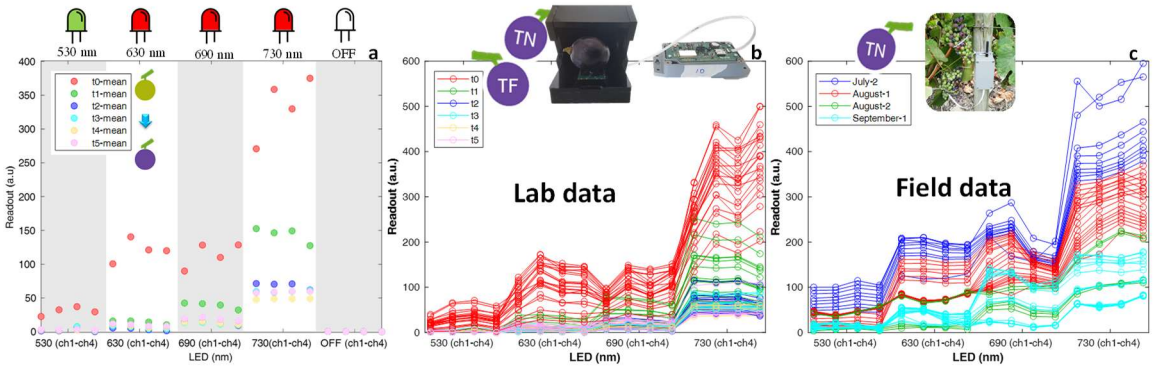


Figure 4.9 Sensor readouts of each LED read by each PD. a) mean optical outputs obtained each sampling time (including the background condition with LED off); (b) lab readouts (samples labelled according to the sampling times) and (c) field readouts (samples labelled according to rang of weeks) used for model calibration.

The descriptive statistics of the wet-chemical analysis related to the grape ripening parameters (TSS, PA, pH and TA) are summarized in figure 4.12. Overall, the technological maturation curve of TN and TF grapes is well described by the sampling campaign performed during the crop season 2020. Each sample was analysed optically (by lab sensor) and chemically (by the reference instruments) to proceed with the models calculation. In addition, a reference measurement needs to be associated also to the optical outputs from the field sensor in order to develop general predictive models computing data from both lab and field. Therefore, the wet-chem outcome was used to develop the evolution curve (for each specific parameter) to extrapolate for each day the reference values to be associated with the optical outcomes deriving from the field sensor.

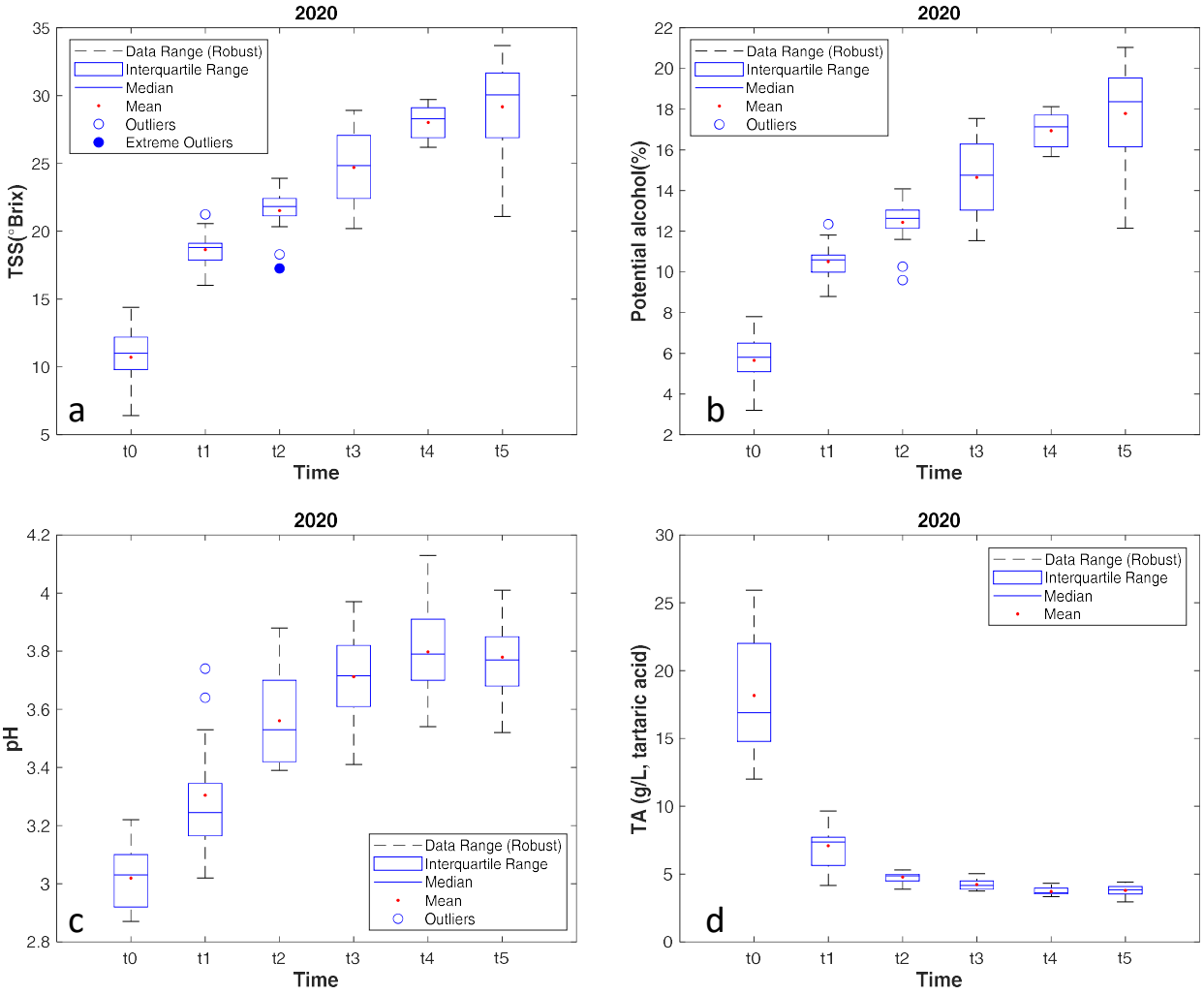


Figure 4.10 Descriptive statistics of the wet-chemical analysis for TSS (a), PA (b), pH (c) and TA (d) obtained from grape samples of cv. Touriga Nacional and Touriga Franca (campaign 2020).

Four PLS models were developed for the prediction of the qualitative parameters of interest. The models were developed using 70% of the total amount of the data for calibration and 30% as external validation (prediction).

In figure 4.11 were summarized the models figure of merit in pure prediction of the external validation set showing the relation between the measured values (using the reference instruments) and the predicted values (using the model based on the sensor readouts). The external validation samples set were labelled in diamonds (TF) and in squares (TN) for the samples coming from lab analysis and field samples labelled in triangles (B01).

In detail, it was concluded that:

- The best models were obtained for TSS, and consequently PA considering an R^2 about 0.90 and RMSEP of 2.22 and 1.54, respectively using 6 LVs;

- A very promising model was also obtained concerning the TA with an R^2 equal to 0.93 and RMSEP of 1.39 (using 4 LVs);
- The pH model (using 4 LVs) showed a lower performance than previous parameters (R^2 of 0.76 and RMSEP 0.15) but still with potential for being used with further improvements.

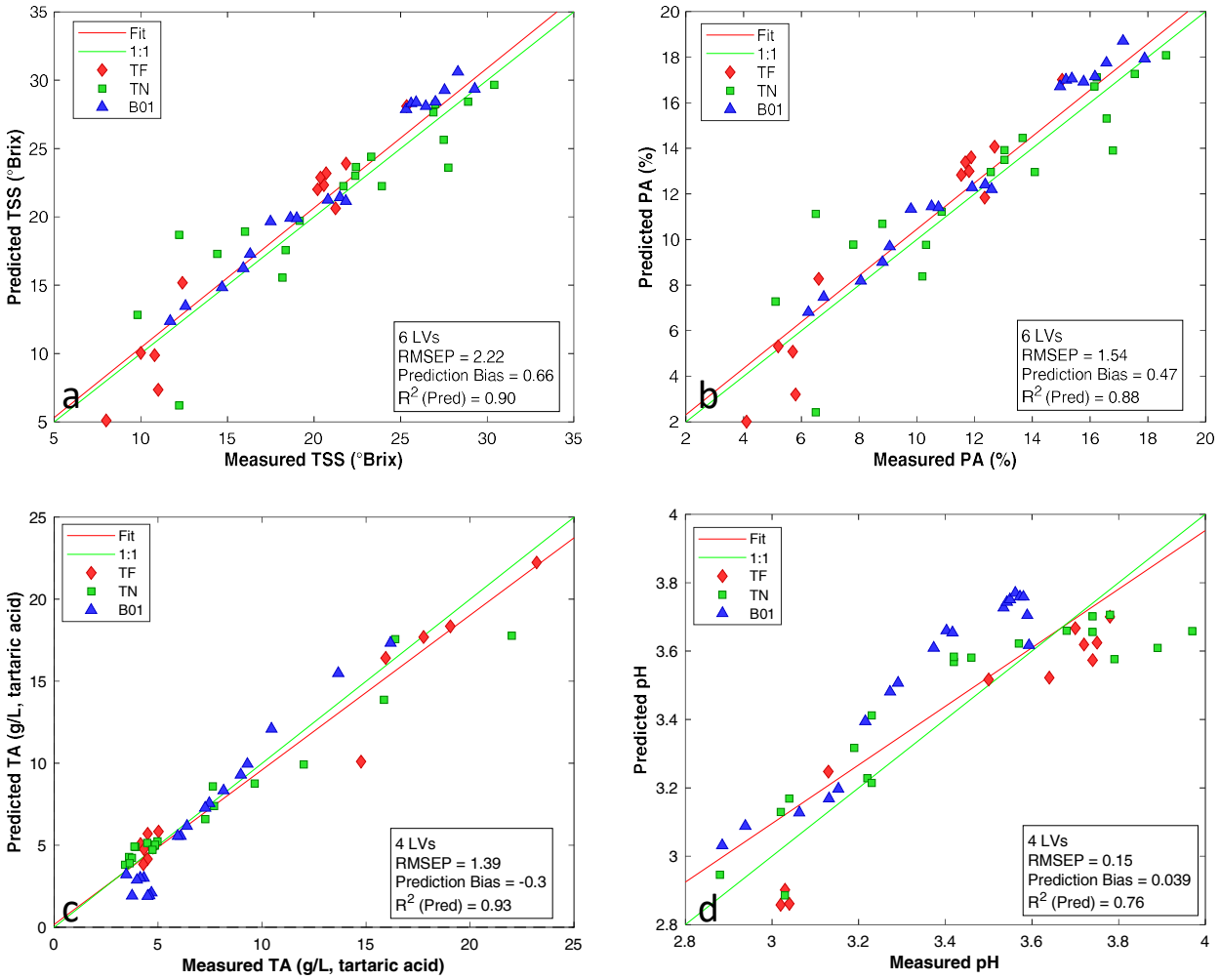


Figure 4.11 Figure of merit of the PLS models. The prediction results were expressed in terms of Root Mean Squared Error in prediction (RMSEP), coefficient of determination (R^2) and prediction bias. Lab samples labelled in diamonds (TF) and in squares (TN) and field samples labelled in triangles (B01).

Field sensors outcome

Once calibrated, the models have been integrated in the software online in order to compute the field data (Quinta do Seixo, Tabuaco, Portugal) coming from the distributed sensors. This system of management guarantees a real time evaluation of the maturity condition of the monitored grape bunch. Figure 4.12 shows the sensors in real operative conditions during the campaign 2021. Moreover, the relative readouts of the sensor's measurements were reported

labelling the data according to the days of the crop season 2021. However, due to problems of sensor deployment (at the beginning of the crop season) and grape harvesting (at the end of the crop season) not all days of the 2021 campaign were monitored (as noticed in the colorbar) with these three sensors.

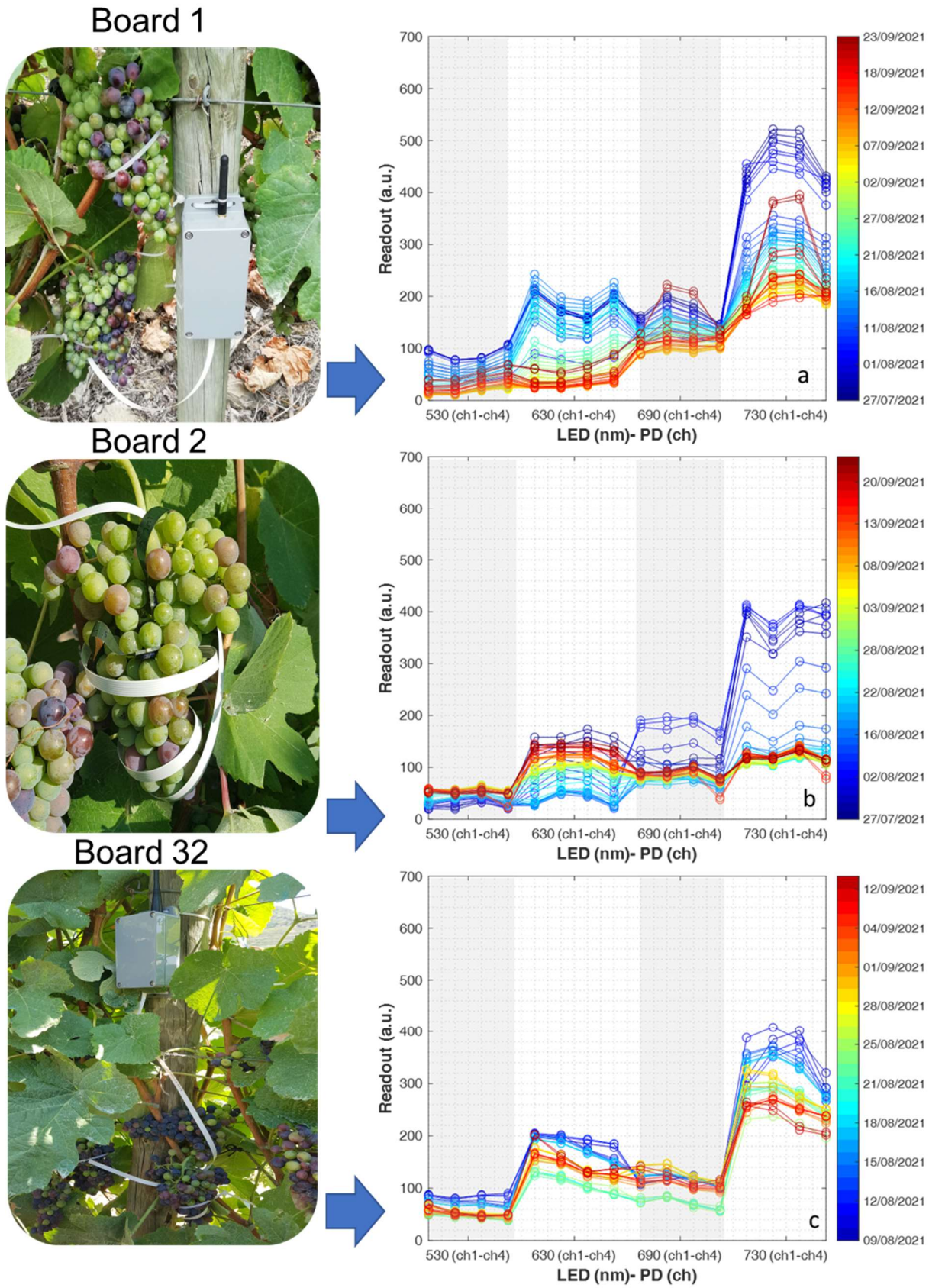


Figure 4.12 Field sensors (board 1, board 2, board 32) measurement from campaign 2021.

Figure 4.13 shows the predictions of the readouts coming from board 1, board 2, board 32. Overall, a good prediction has been obtained. Indeed, the maturation curve is consistent (especially for board 1 and 2) with the wet-chemical analysis performed by the winery during the crop season 2021 (Figure 4.14). Instead, for board 32 a lower performance (probably due to the late positioning of the sensor into the grape bunch) has been obtained.

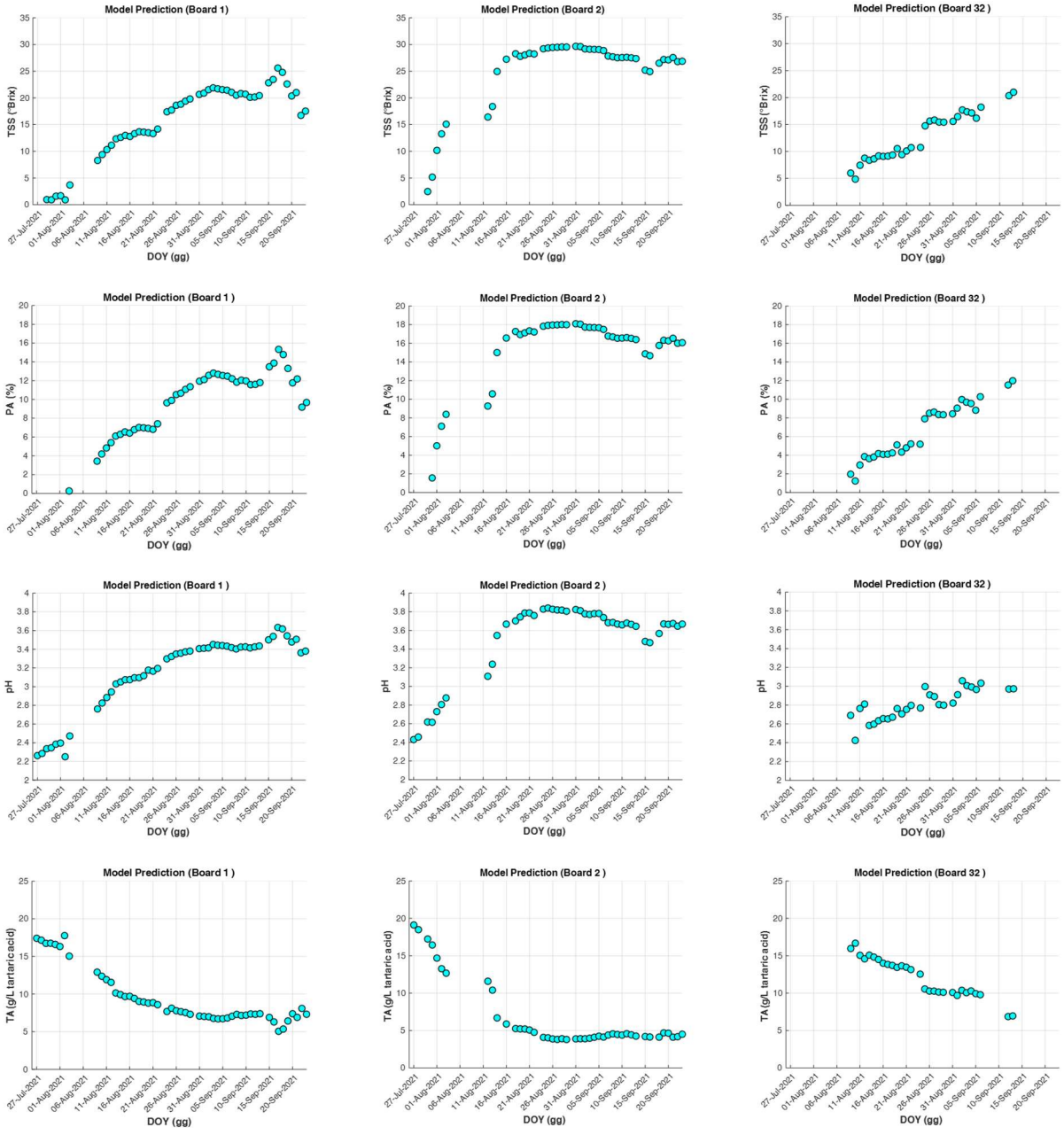


Figure 4.13 Optical prediction (TSS, PA, pH and TA) from field sensors (board 1, board 2, board 32) from the campaign 2021.

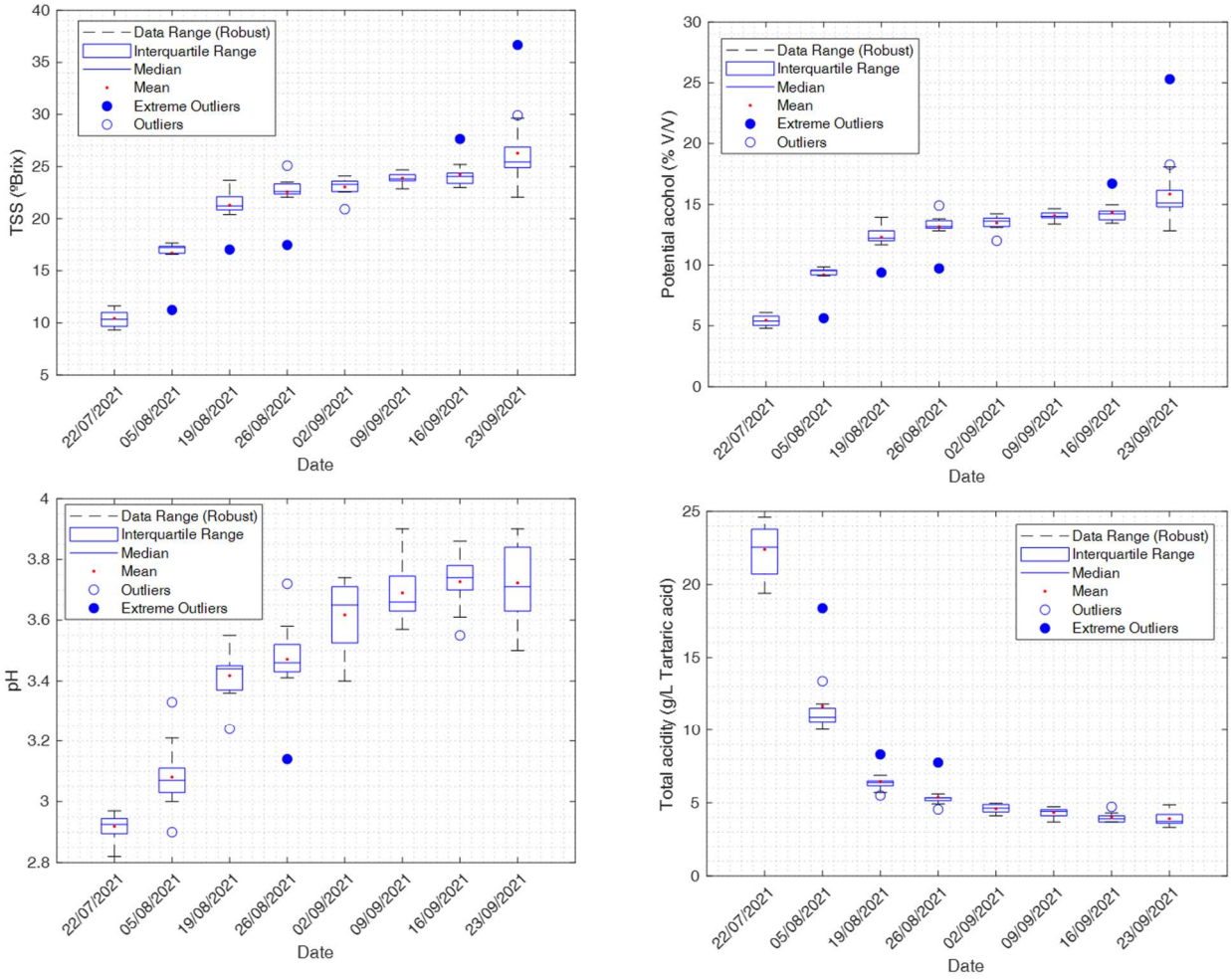


Figure 4.14 Descriptive statistics of the conventional wet-chemical analysis (TSS, PA, pH and TA) performed by the winery during campaign 2021.

Conclusions and future studies

To conclude, a good prediction capability has been reached for each qualitative parameter envisaging a real application of this device in a more sophisticated network of sensors in order to give the possibility to the wine industry to bring the laboratory to the field. However, further experiments must be carried out and different operational strategies to obtain reliable optical data need to be fine-tuned.

Firstly, a filtering procedure on optical data coming from field sensors must be developed in order to detect possible failures on data acquisition due to: (i) technical issues (sensor breaks and errors in data transmission), (ii) possible detachment of the sensor from the bunch, (i) environmental conditions (e.g., precipitations) and (iv) deterioration of the grape bunch (infections and dehydration).

Then, in order to spread the application of this sensor for the whole viticulture sector, other models for the prediction of qualitative parameters need to be calibrated to include white grape varieties into the package of potential applications of this new type of sensors.

Finally, given the low capability of these sensors (which work in diffuse reflectance in the vis/NIR range) to give back optical outputs strongly related to the concentration of polyphenols and anthocyanins (performance is consistent with previous works reported in literature, data not shown), the development of a fluorescence module is currently under study and it will be taken in consideration for a future sensor upgrade.

5. GENERAL CONCLUSIONS

This PhD project regarded different applications of non-destructive optical techniques to evaluate the quality of agri-food products as well as the development of customized optical devices.

In the coffee industry, this work laid the foundation for implementing a real-scale application of the NIR technology as a routine standard method for moisture content evaluation of coffee directly on-line proposing this method as an international standard.

Instead, in the olive supply chain, several works have been done to highlight the advantages of optical sensing techniques. FT-NIR benchtop instruments and portable commercial devices were used to develop classification models for the objective evaluation of olive ripening stage based on maturity classes objectively assessed by an image analysis method. Moreover, predictive regression models (from FT-NIR benchtop instruments and portable commercial devices) were set up and compared in order to estimate qualitative parameters for olive maturation. The model comparison highlighted a general better performance the FT-NIR acquisition systems with respect to the portable device. However, the Vis / NIR device, being portable and relatively cheaper, is worthy of further investigation, because its use could be in any case very useful for preliminary quick quality assessment of olive drupes directly in the field. On the contrary, an on-line or in-line application of the FT-NIR optical device in the olive mill has been promoted in order to quickly classify the drupes for a better quality design of the olive oil and more sustainable management of the production chain.

Moreover, to improve the production of high-quality wines, optical hand-held and stand-alone prototypes were designed, built, and tested. Overall, very interesting results were obtained for the prediction of solids soluble, potential alcohol, titratable acidity and pH. This new generation of optical devices could be the starting point to build a new concept of inexpensive IoT sensors used or distributed in the vineyard. The integration of simple multivariate models in the microcontroller software easily calculates and visualizes in real-time the values of the predicted parameters in the cloud (online data bank). The cost of these devices has to remain fairly low (about few tens of euros) in order to be able to install many devices in crucial points of the vineyard to obtain an average value related to the ripening status of the whole vineyard. Therefore, even though further studies have to be completed yet in terms of data filtering and interpretation from the field, good results were obtained envisaging a real application of these devices in a more sophisticated network of sensors in order to shift the current paradigm of grape maturation monitoring bringing the laboratory to the field.

From these experiences (carried out during the three years of Ph.D.) the candidate has developed different skills in the analytical, chemometric, engineering and computational fields. in each work reported in the thesis, the multivariate data approach has always been the basis

of every activity during which he was able to combine concepts of sustainability, image analysis, software and hardware design. Therefore, thanks to the skills obtained and the desire to acquire new ones, a further challenge was faced and described in the chapter below, laying the foundations for possible further research activities.

6. IMPLICATIONS AND FUTURE DIRECTIONS

Starting from benchtop and process instruments (developing optical predictive models) and finishing with the developments of cost- effective optical sensors, this thesis gave a complete overview about the application of vis/NIR spectroscopy and highlighted new frontier for the use of sensors base on it.

Especially in the viticulture sector, the introduction of these new optical technologies as the one proposed for supporting vineyard management allows the efficiency and quality of production to be improved with simultaneously reduction of the environmental footprint with obvious cost gains.

Recent technological developments have provided the necessary tools for the development of an integrated optical platform able to accommodate the required degree of customization necessary for each particular application. Regarding wine industry, this will make possible the implementation of an innovative sensing approach that will complement the current analytical tools (both proximal and remote) for the monitoring of the vineyard status. The results of the project have a strong economic relevance considering the control of the quality variability during the ripening process, and will significantly impact on multiple areas:

- Economic Impact on the photonics Industry
- Economic Impact on the IoT industry.
- Economic Impact for the micro-systems component Industry (in the process monitoring efficiency in the food industry and reduction of the environmental impact of the production)
- Economic impact for the agri-food industry-wine (increased competitiveness of the process and product monitoring equipment industry)

Given the evidence that this proposed proximal sensing technology has a considerable impact on many fields (especially in the winemaking industry), a further step can be moved in terms of cost reduction for a more complex optical technique which can gather even more information, the hyperspectral imaging. As mentioned in chapter 2, the advantaging of imaging techniques is characterized by the presence of the spatial resolution (Sx and Sy) and the spectral resolution (Sλ) which measure the variations in illumination within the image pixels as a function of wavelength. The advantage of also having spatial information available makes (in some cases) this technique much more efficient and useful than normal spectroscopy. For example, remaining in the viticulture sector, the measurement of the water status based on the measurement of the water potential in pre-dawn (before the sun rises) by means of the Scholander pressure chamber is very explored in literature. For some time, attempts have been made to use optical techniques to replace or complement this measure. In particular, the vis/NIR spectroscopic techniques (De Bei et al., 2011; Giovenzana et al., 2017) have produced

encouraging results but with a strong need of improvement for the development of dedicated instruments and sensors.

Thanks to the presence of the spatial information provided by an hyperspectral imaging approach could solve different issues in terms of data accuracy and application in real operative conditions.

However, even though the costs for spectroscopic sensors can be easily affordable by industry, the common hyperspectral cameras available on the market are expensive and this is an issue to be solved in order to provide the advantages of this technology to wider users. For the companies, the cost limitations are not strictly related to the device itself but for the specific applications. Indeed, even though the hyperspectral imaging technique can collect a large amount of data, the application of only one device (in some cases) is not enough to cover all the critical points which an industry can have. The production process in a firm or the monitoring in field require systems which are distribute in different specific locations in order to collect data and provide information. In these circumstances, if we consider the application of several hyperspectral devices, the costs become prohibitive for the majority of the companies and the research is going toward the development of hyperspectral sensors taking into account a considerable cost reduction

Therefore, during the experience abroad in this PhD, a first step for the development of a cost-effective hyperspectral imaging device has been done and described below.

6.1 First steps for the development of an hyperspectral imaging device

The proposed hyperspectral imaging (HSI) device can export hyperspectral images with 116*110 pixels, up to 315 wavebands from 400–1052 nm, and a size of 35 Mb. This device has a total weight of 500 g, with a dimension of 18*9*7 cm.
At this stage, the prototype spans the visible and part of the NIR spectral range and does not possess movable parts in order to decrease possible errors due to the vibrations.

6.1.1 Hardware specs

In Table 6.1 is reported the list of all the parts needed during the device building, the costs, the material type and the source where it was bought.

Table 6.1: list of the components used for the HIS device

Component	Cost (€)	Material Type
Raspberry Pi 3 B + with charger	51	Electronics
Raspberry Pi NoIR V2 Camera 8 Megapixels	37	Electronics
Mini SD memory, 32 Gb	18	Electronics
Macro + 10 lens	15	Optics
35 mm C Series Fixed Focal Length Lens. Model: 59872	320	Optics
Double-axis diffraction grating	11	Plastic
Case	≈13	Resin
Front lens	≈3	Resin
Square aperture	≈1	Resin
50mm extension	≈7	Resin
50mm extension	≈7	Resin
Diffraction grating	≈1	Resin
Camera base	≈1	Resin
Seal extension	≈2	Resin
X extension	≈2	Resin
Lid	≈3	Resin
Hanger	≈4	Resin
Hanger clip	≈1	Resin
	Tot. ≈517€	

Figure 6.1 shows all the components which compose the proposed HIS device.

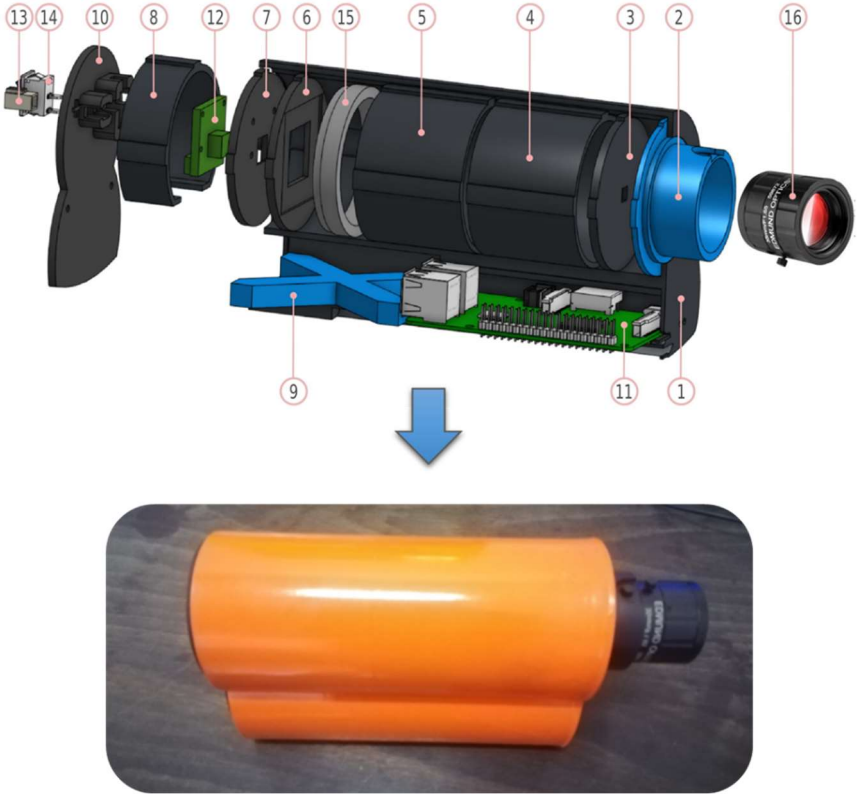


Figure 6.1 HSI composition 1) Case; 2) Front lens; 3) Square aperture; 4-5) Extension 1 and 2; 6) Diffraction grating; 7) Camera base; 8) Seal extension; 9) X extension; 10) Lid. 11) RPI01 Raspberry Pi 3 B+. 12) RPI02 Raspberry Pi NoIR V2 Camera 8 Megapixels. 13) PS02 Power Supply Connector. 14) PS03 Power On/Off Switch. 15) +10 Macro 52 mm Lens. 16) LENS02 35 mm C Series. Edmundoptics, model 59872.

6.1.2 Calibration

After the hardware assembling, a calibration phase has been done in order associate the wavelengths split by the double axis diffraction grating. Firstly, in darkroom two images have been acquired using an halogen lamp (to have a diffracted light with a continues spectrum) and one image using a fluorescent lamp to have a diffracted light with three or more well-known wavelength peaks (405, 440, 545, 613, and 710 nm) (Figure 6.2). These images were used to calibrate the device and set the region of interest (ROI), the blue, green and red peaks, and in conclusion the range limits in which the software has to find the spectral information. With this phase every photograph was red with the same pattern created and saved in a calibration function.

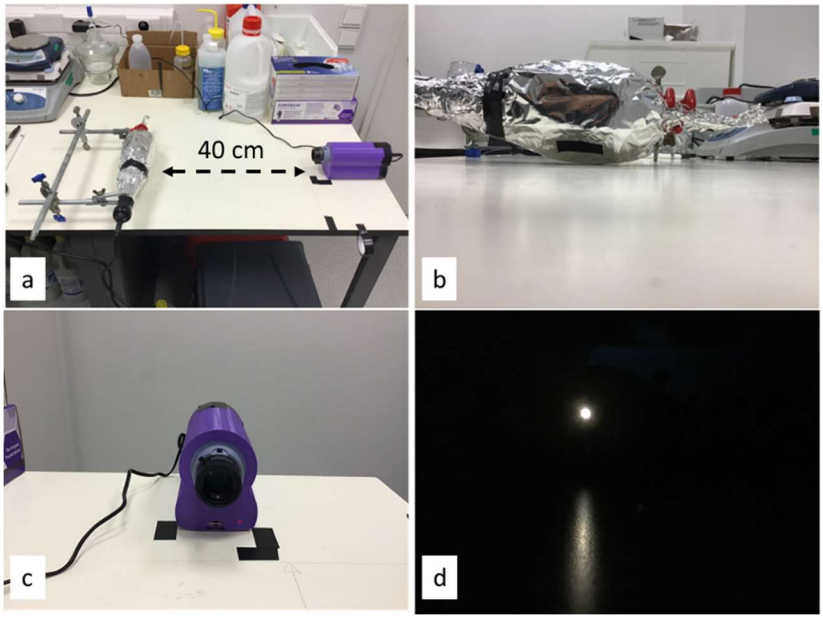


Figure 6.2 HSI calibration set-up in dark room.

Figure 6.3 shows the three images acquired with incandescent and fluorescent lamp. The images show two distinguishable areas:

1. Zero-order mode (ZM);
2. first-order mode.

ZM occurs when the light is not diffracted. It appears as a small square, located at the centre of the image. Instead, the first-order mode is the first time that the light is diffracted. It has a higher intensity than higher modes (e.g., second order diffraction mode) and its perimeter contains ZM. Hence, this area will contain the mayor amount of information and energy. Therefore, the first-order mode is the ROI (Figure 6.4).

Figure 6.3a was used to define the ROI to be used. Figure 6.3b contains the diffracted light (produced using a fluorescent lamp) with three or more well-known wavelength peaks at 405, 440, 545, 613, and 710 nm (figure 6.5). Instead, figure 6.3c contains diffracted light with a continues spectrum.

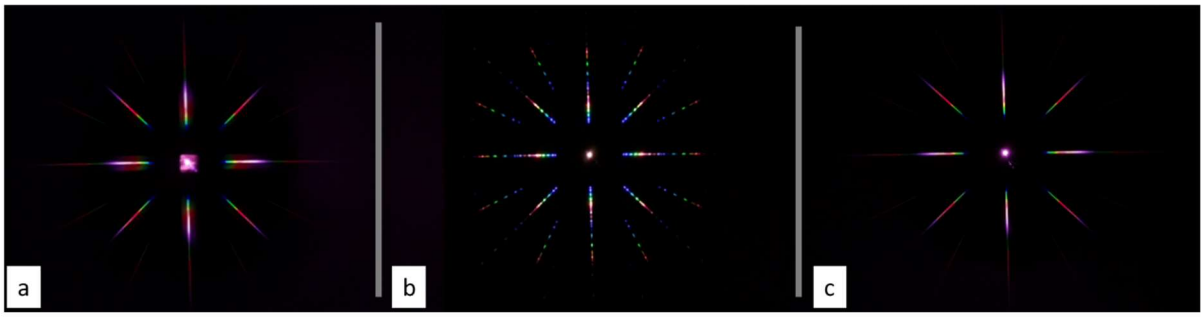


Figure 6.3 set of reference images. (a) image with halogen lamp and without darkroom. (b) image with fluorescent lamp in darkroom. (c) image with halogen lamp in darkroom.

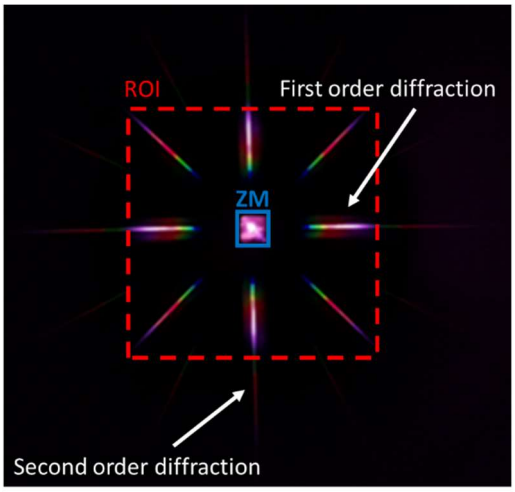


Figure 6.4 image with halogen lamp and without darkroom to define the zero-order mode (ZM) and the region of interest (ROI).

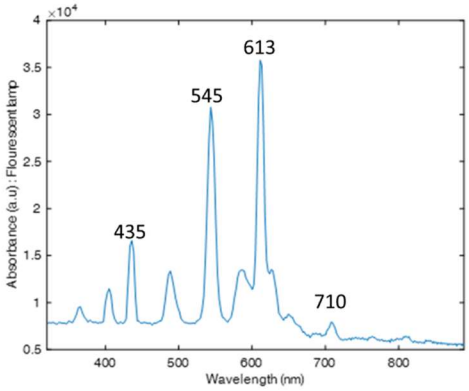


Figure 6.5 Spectra of absorbance from the fluorescent lamp (acquired using a commercia spectrophotometer from Hamamatsu, C12880MA, spectral response range from 340 to 850 nm).

Therefore, using the coordinates of the light's origin (in ZM) and the coordinates of the diffracted wavelength peaks (first order diffraction) is possible to calculate the relationship between wavelength-to-pixels and vice versa and build the hypercube which includes spectral and spatial information.

6.1.3 Test

This part has the aim to show the performance of the proposed device il lab-scale conditions. Firstly, an imaging system apparatus has been built in order to acquire images under controlled conditions (Figure 6.5). The system is composed by:

1. Integrating sphere (hollow hemisphere with perfectly diffusing internal surface which allows a complete reflection of the light toward the sample);
2. Camera holder (to keep the HSI at 70 cm from the sample);
3. 4 Light bulbs.

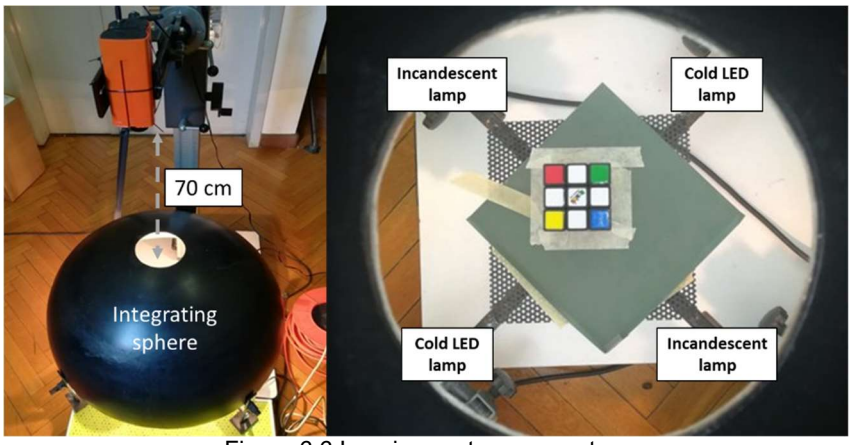


Figure 6.6 Imaging system apparatus.

In particular, the four light bulbs were: two Incandescent and two cold LED. This set-up allows to have a source of light which cover at best the vis-NIR spectral range between 400 and 1000nm. Figure 6.7 shows the light emitted from the Incandescent and cold LED lamp measured using a commercial vis/NIR spectrophotometer (Jaz, Ocean Optics, Dunedin, FL). Even though the shape of the spectrum depends also from the sensitivity of the spectrophotometer detector, it is clear that the incandescent lamp is poor at 400nm because the emission range of this type of lamps doesn't cover waveband under 500nm. Therefore, this lack at low wavelength has been balanced by the cold LED lamp that have a pick of emission at around 400nm. Whereas, at high wavelength the LED lamp has a very poor emission, and this has been balanced by the incandescent lamp.

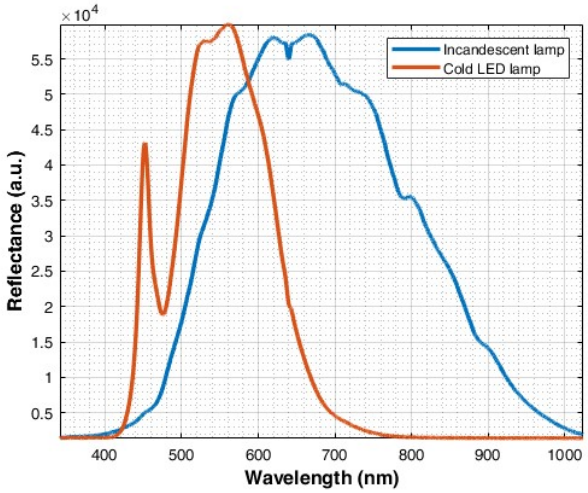


Figure 6.7 Emission of two different lamps. The blue line is referred to the incandescent lamp and the orange line to the cold LED lamp.

Then, a first test has been performed using a common 3 x 3 Rubik's cube in order to validate the calibration and the acquisition set-up. This experiment aimed to obtain an hyperspectral image (the hypercube) and use it to evaluate the accuracy of the hyperspectral device. The Rubik's cube has been arranged with a white cross and four different colours (red, green, blue and yellow) at the corners (Figure 6.8).

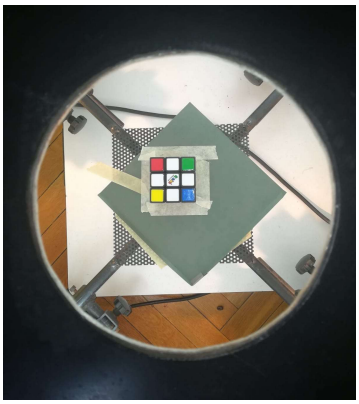


Figure 6.8 3 x 3 Rubik's cube arrangement for the first test.

Once the cube has been placed at 70 cm from the front lens of the device the system is ready for the image acquisition. Figure 6.9 shows the diffracted image taken using a Shutter Speed of 1500μs and a Square Shutter Speed of 2μs. These settings were the best compromise to improve the quality of the photograph.

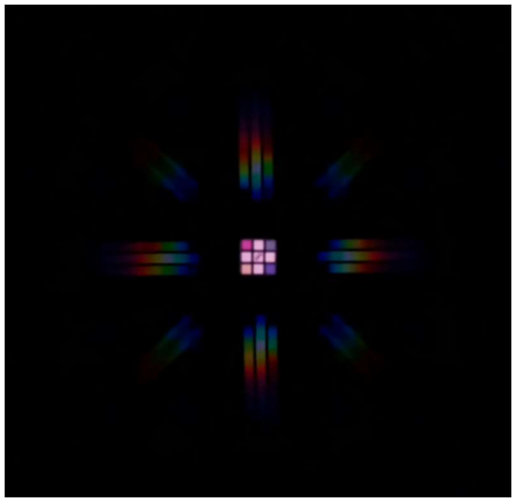


Figure 6.9 Diffracted image of the 3 x 3 Rubik's cube face.

Once obtained the diffracted image, using the hypercube function it was possible to generate the hypercube to be processed in MATLAB environment.

The hyperspectral cube has been evaluated using the software Hypertools (Graphical User Interface (GUI) for the analysis of multispectral and hyperspectral images in MATLAB). Figure 6.10 shows, from the diffracted image (Figure 6.10a), the composite image (Figure 6.10a) in false colours generated by the diffracted image of the Rubik's cube. Therefore, from each pixel which composes the composite image, it is possible to extract the relative spectral profile from 400 to 1000 nm (Figure 6.10c). One pixel for each representative square was taken in order to visualize the spectra of each color of the Rubik's cube. The final result is a spectral output which depends from the intensity of the environmental light (in the imaging system apparatus in lab-scale) and from the sensitivity of the image sensor of the Raspberry Pi NoIR V2 (CMOS image sensor, IMX219, Sony). Obviously, higher is the light reflected from the Rubik's cube (which depends from the color of the squares) higher is the magnitude of the relative spectrum. For instance, the white square shows the highest reflectance (as expected) and the three picks related to the red, green and blue peaks are clearly viewable. However, a shift about 100 nm is noticeable taking into account the real emission of the blu (around 450nm), the green (around 530nm) and the red (around 650nm) suggesting a non-perfect calibration of the camera or an inefficiency of the low-cost holographic transmissive grating.

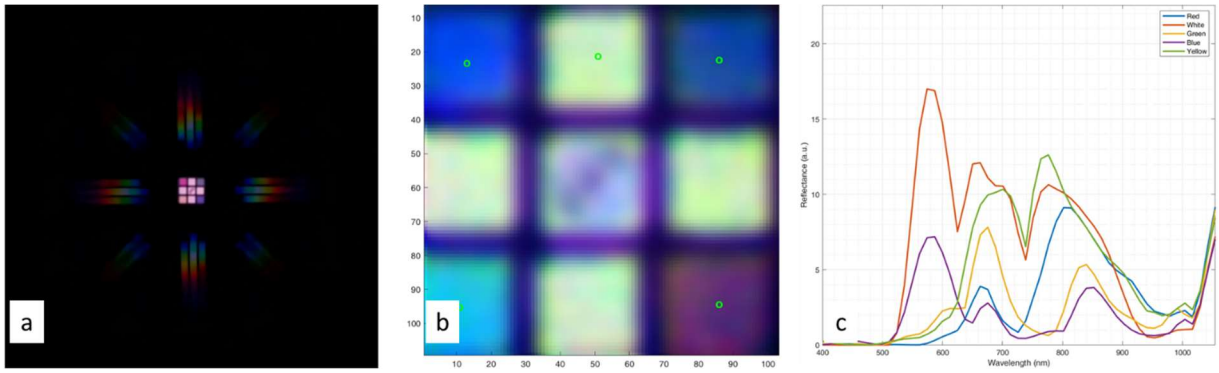


Figure 6.10 (a) diffracted image; (b) composite image with pixels; (c) spectra of the selected pixels.

Despite the obvious problem to centre the respective emission peaks at the correct wavelengths, a preliminary exploratory study was performed using PCA. PCA is usually performed on two-dimensional matrices, therefore, three-dimensional hypercube has to be unfolded in order to obtain the two-dimensional matrix with the wavelengths on the columns and pixels on the rows.

At first, the image has been masked using the first PC scores of the PCA (with mean-centred data) in order to avoid any black border of the Rubik's cube. In real samples, the masking procedure (which can be carried out with different strategies) allows to remove the background or the uninformative pixels isolating the part that contains the real information. The final output of the masking phase is an hypercube with only the squares which compose the Rubik's cube. Then, another PCA (with mean-centred data) has been performed in other to see the remaining variability within the dataset. Figure 6.11 shows the PCA scores and loadings of the masked Rubik's cube dataset.

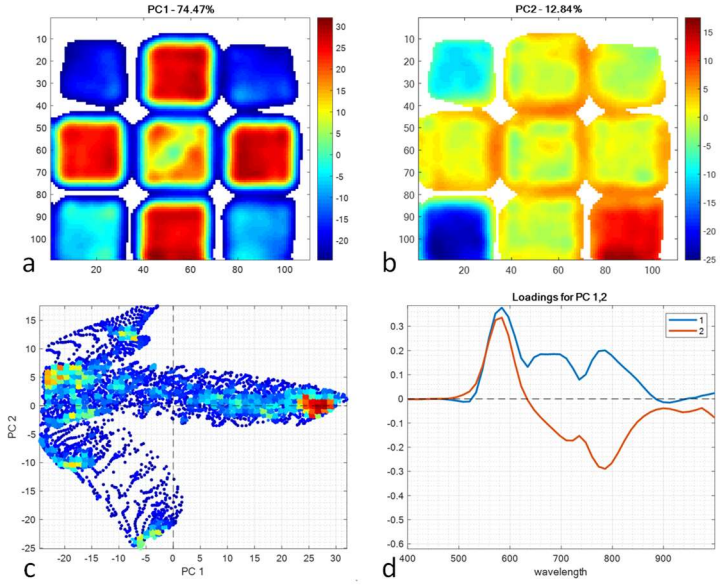


Figure 6.11 PCA scores and loadings of the Rubik's cube hyperspectral image.

The first and the second principal components, describe the 74,47% and 12,84% of the total variability present in the masked hyperspectral image, respectively. Looking at the scores of PC1, the differences are related to white squares with respect to the remaining colours while, PC2 (Figure 6.11b) highlight the differences between the coloured squares from blue (bottom right square) one to yellow one (bottom left square) with the green (up right square) in the middle of the colorbar as it comes from a mix of blue and yellow. Moreover, Figure 6.11 c and d show the behaviour of each pixel in the space of PC1 and PC2 and the weight of the original variables (wavelengths) that contribute to describe these two components. In these terms, no particular assumption could be done due to the problem of wavelengths shift already highlighted.

6.1.4 Conclusions

To concludes, this experience paved the ground for the development of cost-effective hyperspectral devices. Very interesting advantages in terms of costs (around 500€) have been reached using commercial stuff and products of additive manufacturing. However, several improvements need to be done in order to obtain a better resolution and in the final image (e.g. better calibration and better holographic diffraction grating).

The applications of devices (like the one described) are various and after these 3 years of PhD more studies will be done in the field of precision agriculture which is one of the sectors where the presence of distributed low-cost devices able to monitor the crop conditions (e.g. water status) is strongly required.

7.APPENDICES

Awards

Award 1: "WHAT FOR" AWARD 5th EDITION

Sponsored by Federalimentare in collaboration with the national network of doctoral programs of research in food science, technology and biotechnology.

Award 2: Best oral presentation at NIRItalia online 2021

NIR Italia online

24-25 February 2021

waiting for Slovenia 2022

SISNIR–National Symposium of NIR Spectroscopy
Via A. Lombardo, 11 – 26900 LODI - c.f. 92544830158

NIRItalia online 2021

awards

BEST ORAL PRESENTATION AWARD

to

Alessio Tugnolo

Stand-alone LED sensors for future field monitoring of grape (Vitis vinifera L.) ripeness

tine.sukljan@InnoRenew CoE.eu

Congratulations!

Organizing Committee

Dr. Anna Sandak
(Head of the Organizing Comittee)

Scientific Committee

Dr. Monica Casale
(President of SISNIR)

Koper, 26 February 2021

Awards

- Winner of the SISNIR grant offered by Società Italiana di Spettroscopia NIR to attend the 19th biennial meeting of the International Council for NIR Spectroscopy (NIR2019)

- Winner of the SISNIR grant offered by Società Italiana di Spettroscopia NIR to attend the 20th biennial meeting of the International Council for NIR Spectroscopy (NIR2021)

Conferences attendance

Conferences attendance

Model-It, International Symposium on Applications of Modelling as an Innovative Technology in the Horticultural Supply Chain

(2019)

ORAL
Design of prototypes of LED based devices for the evaluation of grape (Vitis Vinifera L.) ripeness

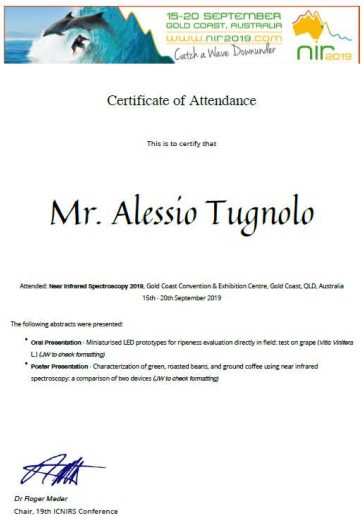
POSTER
Visible/near infrared spectroscopy for horticulture:case studies from pre-harvest to post-harvest

International Conference of NIR Spectroscopy

(2019)

ORAL
Miniaturised LED prototypes for ripeness evaluation directly in field: test on grape (Vitis Vinifera L.)

POSTER
NIR spectroscopy for coffee quality evaluation in a view of at-line and in-line applications



XXIV Workshop on the developments in the italian PhD research on food science, technology and biotechnology

(2019)

POSTER
Feasibility studies and engineering of optical simplified and stand-alone instruments for agro-food applications

Conferences attendance

2° Workshop sull'innovazione nella meccanica e nell'impiantistica applicata ai biosistemi agro-alimentarie forestali

(2019)

ORAL
Optical LED prototypes for ripeness evaluation of grape (Vitis Vinifera L.)

XXV Workshop on the developments in the italian PhD research on food science, technology and biotechnology

(2021)

ORAL
Feasibility studies and engineering of optical simplified and stand-alone instruments for agro-food applications


I virtual Workshop on the Developments in the Italian PhD
Research on Food Science, Technology and Biotechnology

Palermo, September 14th-15nd, 2021

Certificate of Workshop Attendance and Successful Learning Outcomes

Student Alessio Tugnolo
born in……
date……
resident in……
address……
Matriculated in the……year of the PhD Course in
……
has participated to the plenary lecture and
has acquired n° **4** credits (**CFU**) with
a grade **A** (score 92,5 out of 100)

Head of PhD Course/curriculum

1

Conferences attendance

NIRItalia Online

(2021)

ORAL
Stand-alone LED sensors for future field monitoring of grape (Vitis vinifera L.) ripeness


NIR Italia online 24-25 February 2021

waiting for Slovenia 2022

SISNIR–National Symposium of NIR Spectroscopy

NIRItalia online 2021

This is to certify that

Dr Alessio Tugnolo

UNIMI

participated in the NIRItalia online 2021 event that was held on the 24th and 25th February 2021.

Organizing Committee

Dr. Anna Sandak
(Head of the Organizing Committee)

Scientific Committee

Dr. Monica Casale
(President of SISNIR)

Koper, 26 February 2021

International Conference of NIR Spectroscopy

(2021)

ORAL
Stand-alone LED sensors for future field monitoring of grape (Vitis vinifera L.) ripeness

Certificate of Attendance

This is to certify that

Alessio Tugnolo

attended the NIR2021 International Conference held as an on-line meeting in Beijing, China, 18-21 October 2021

NIR2021 Chairman

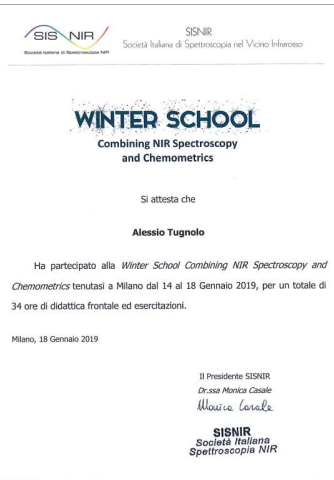
Prof. Hongfu Yuan

Beijing University of Chemical Technology

Scientific training courses

Scientific training courses

WINTER SCHOOL: Combining NIR Spectroscopy and Chemometrics (2019)



SISNIR
Società Italiana di Spettroscopia nel Vicino Infrarosso

WINTER SCHOOL
Combining NIR Spectroscopy and Chemometrics

Si attesta che

Alessio Tugnolo

Ha partecipato alla *Winter School Combining NIR Spectroscopy and Chemometrics* tenutasi a Milano dal 14 al 18 Gennaio 2019, per un totale di 34 ore di didattica frontale ed esercitazioni.

Milano, 18 Gennaio 2019

Il Presidente SISNIR
Dr.ssa Monica Casale

SISNIR
Società Italiana Spettroscopia NIR

School of Multivariate Analysis (2019)

UNIVERSITY OF GENOA
Department of Pharmacy
Research Group of Analytical Chemistry and Chemometrics
Viale Cembrano 4 – 16148 GENOVA - Italy

Tugnolo Alessio attended the SCHOOL OF MULTIVARIATE ANALYSIS, Genoa, September 30 – October 4 2019, for a total of 30 hours.

Prof. Riccardo Leardi

Genoa, October 4th 2019

International School of Chemometrics (2020)

ikerbasque
Basque Foundation for Science

October 30th, 2020

CERTIFICATE OF ATTENDANCE ISC-2020

My name is José Manuel Amigo, Research Professor of Ikerbasque and the department of analytical chemistry of the university of the Basque Country (UPV/EHU).

Serve this letter to certify that

Mr. Alessio Tugnolo

Has attended the ISC-2020 International School of Chemometrics held in the Marine Institute of Plentzia, Spain, between the 8th of October until the 30th of October, 2020.

During this period, the student has learned:

- Introduction to Matlab (3 ECTS)
- Basic chemometrics (2.5 ECTS)
- Intermediate chemometrics (2.5 ECTS)

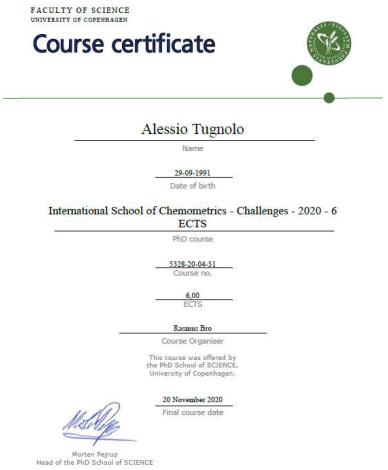
In total, the student has been granted with 8 ECTS after passing all the tests favorably.

Sincerely,

José Manuel Amigo
Ikerbasque Research Professor
Faculty of Science and Technology
University of the Basque Country, Bilbao, Spain
josemanuel.amigo@ehu.eus

Scientific training courses

International School of Chemometrics-Challenges (2020)



FACULTY OF SCIENCE
UNIVERSITY OF COPENHAGEN

Course certificate

Alessio Tugnolo
Name

29-09-1991
Date of birth

International School of Chemometrics - Challenges - 2020 - 6 ECTS
PhD course

5328-20-04-31
Course no.

6.00
ECTS

Rasmus Bro
Course Organiser

This course was offered by the PhD School of SCIENCE, University of Copenhagen.

20 November 2020
Final course date

Morten Pejrup
Head of the PhD School of SCIENCE

School of Experimental Design (2021)

Università di Genova | DIFAR DIPARTIMENTO DI FARMACIA

Sezione di Chimica e Tecnologie Farmaceutiche e Alimentari
Gruppo di Ricerca di Chimica Analitica e Chemiometria
Viale Cembrano 4 – 16148 GENOVA

This is to certify that **Tugnolo Alessio**, DeFENS, Università degli Studi di Milano, attended the SCHOOL OF EXPERIMENTAL DESIGN, Microsoft Teams, November 8-12, 2021, for a total of 32 hours.

Riccardo Leardi

Prof. Riccardo Leardi

Genoa, 12/11/2021

Other experiences

PhD period abroad
July – October 2020: Braga (PT), International Nanotechnology Laboratory (INL)

Tutor: Dott. Joao Piteira

Research topic: design of cost-effective hyperspectral camera

Credits and Soft skills

Course UNIMI catalogue	Professor	date	Food Systems course	cfu
Advanced spectroscopic methods in food systems	Bonomi	May 2019	yes	3
Introduction to statistical analysis of ecological and environmental data	Roberto Ambrosini	March 2019	no	5
The methodology of life cycle assessment (LCA) in the food	Riccardo Guidetti	December-2020	yes	4
		Total		**12**

Journal Club	oral presentation	speaker extra hours	duration (hours)	cfu
15/02/2019	yes	4	6	1.25
24/01/2020	no		3	0.375
26/02/2021	no		2.2	0.275
	Total			**1.9**

"Food Systems" seminars	speaker	date	duration (hours)	cfu
Recent research trends and funding sources in food and human nutrition in the US	Klimis-Zacas	01/07/2019	1	0.125
A South African perspective on grape cultivation during drought conditions	Blancquaert	01/07/2019	1	0.125
Global metabolomics for discovery of natural product bioactivity	Stevens	11/12/2018	1	0.125
	Total			**0.375**

Other "Food Systems" activities	date	presentation	speaker extra hours	duration (hours)	cfu
workshop 2019 "Food PhD"	11-13/9/2019			16	2
Final exam	18/02/2021			4	0.5
Final exam	19/02/2021			4	0.5
workshop 2021 "Food PhD"	14-15/09202 1	oral	32	13	9.625
Final exam	03/11/2021			2	0.25
	Total				**8.625**

Other approved activities	from-to	oral or poster in international event	speaker/ poster extra hours	duration (hours)	cfu
Bet on the research: a warranty for the future	26/11/2018			3	0.187
WINTER SCHOOL: Combining NIR Spectroscopy and Chemometrics	14-18/01/2019			37	2.31
NIR 2019 (International conference of Near Infrared Spectroscopy)	15-20/9/2019	oral-poster	72	44	7.25
Model-it (VI Symposium on applications of modelling an innovative technology in the horticultural supplay chain)	9-12/6/2019	oral	64	24	5.5
2° Workshop sull'innovazione nella meccanica e nell'impiantistica applicata ai biosistemi agro-alimentari e forestali	10/10/2019	oral	64	8	4.5
School of multivariate analysis	30/09/2019-4/10/2019			32	2
SIMEI 2019	20/11/2019	Oral	64	4	4.25
International school of chemometrics (ISC 2020)	08/10/2020-20/11/2020			192	12
NIR italia	24-25/02/2021	Oral	64	10	4.625
ICNIRS 2021 (International conference of Near Infrared Spectroscopy)	18-21/10/2021	Oral	64	15	4.937
School of experimental design	8-12/11/2021			37	2.312
Total					**45.875**

Thesis supervisor	student	bachelor/master	cfu
Valutazione del grado di maturazione di olive da olio mediante analisi dell'immagine e spettroscopia nel visibile e vicino infrarosso	Marco Menegon	master	0.5
Applicazione della spettroscopia nel visibile e nel vicino infrarosso su caffè espresso: prove preliminari.	Mattia Saulle	master	0.5
Analisi dell'immagine e sensoriale di un vino spumante prodotto con remuage ultrasonico	Alessandro Salvini	master	0.5
Sviluppo di un sistema ottico semplificato per monitorare la maturazione di olive da olio	Carlo Pozzi	master	0.5
The use of NIR spectroscopy to monitor the coffee production chain	Matteo Curatitoli	master	0.5
Ingegnerizzazione e sperimentazione di un device ottico semplificato per l'analisi della maturazione del pomodoro	Giulia Rossi	master	0.5
Sviluppo di un sensore stand alone per il monitoraggio della maturazione di uva in un'ottica di viticoltura 4.0	Claudio Vanerio	master	0.5
Analisi impiantistica e proposte progettuali per una cantina a conduzione familiare	Filippo Groppi Garlandini	bachelor	0.25

Applicazione on-line di uno spettrofotometro nir per il monitoraggio del contenuto di umidita' del caffe' e confronto con metodi analitici consolidati	Alice Bacchetta	Master	0.5
Evaluation of the ripening stage of Chardonnay grapes through non destructive optical techniques	Alessia Pampuri	Master	0.5
Design of a cost-effective hyperspectral camera	Giacomo Busnelli	Master	0.5
Total			**2**

Total amount of cfu	**74.775**

Transferable skills	date
Open access – open data e il mondo delle pubblicazioni	13/12/2018
Bet on the research: a warranty for the future (Assolombarda)	26/11/2018
Valutazione della ricerca	01/02/2019
L'importanza della comunicazione (Assolombarda)	18/02/2019
Comunicare risultati complessi (Assolombarda)	04/03/2019
Comunicare: principi comportamentali (Assolombarda)	11/03/2019
Comunicare attraverso i digital media (Assolombarda)	25/03/2019
Valutare la comunicazione (Assolombarda)	01/04/2019
GRANTMANSHIP Parte I	20/01/2020
Research Integrity	16/12/2019
tutelare e valorizzare sul mercato i risultati della ricerca	29/11/2019
FAKE NEWS, disinformazione, divulgazione e ricerca scientifica	09/04/2020
Reasearch Integrity	20/05/2020
SELF BRANDING - Dottorato e mercato del lavoro: spettive professionali e competenze da valorizzare	19/06/2020
GRANTMANSHIP Parte I	14/09/2020
Communication on new media parte 1	03/02/2021
Communication on new media parte 2	11/02/2021
Lezione avanzata sull'utilizzo dell'IP per fare innovazione	24/03/2021
Lezione avanzata sull'utilizzo dell'IP per fare innovazione	09/04/2021
Valorizzare Creando Impresa: Fare Spin off in Università degli Studi di Milano	28/04/2021 05/05/2021 19/05/2021

REFERENCES

The literature below reports all the papers used as a reference over the literature used in already published papers of this Ph.D. experience.

1. Amigo, J. M. (2020) 'Hyperspectral and multispectral imaging: setting the scene', Data Handling in Science and Technology, 32, pp. 3–16. doi: 10.1016/B978-0-444-63977-6.00001-8.
2. Ballabio, D. and Consonni, V. (2013) 'Classification tools in chemistry. Part 1: Linear models. PLS-DA', Analytical Methods. Royal Society of Chemistry, pp. 3790–3798. doi: 10.1039/c3ay40582f.
3. Beghi, R., Buratti, S., Giovenzana, V., Benedetti, S., & Guidetti, R. (2017). Electronic nose and visible-near infrared spectroscopy in fruit and vegetable monitoring. Reviews in Analytical Chemistry, 36(4).
4. Biancolillo, A. and Marini, F. (2018) Chemometrics Applied to Plant Spectral Analysis. 1st edn, Comprehensive Analytical Chemistry. 1st edn. Elsevier B.V. doi: 10.1016/bs.coac.2018.03.003.
5. Bonilla, I., De Toda, F. M. and Martínez-Casasnovas, J. A. (2015) 'Vine vigor, yield and grape quality assessment by airborne remote sensing over three years: Analysis of unexpected relationships in cv. Tempranillo', Spanish Journal of Agricultural Research, 13(2), pp. 1–8. doi: 10.5424/sjar/2015132-7809.
6. Borràs, E. et al. (2015) 'Data fusion methodologies for food and beverage authentication and quality assessment - A review', Analytica Chimica Acta, 891, pp. 1–14. doi: 10.1016/j.aca.2015.04.042.
7. Boulet, J. C. and Roger, J. M. (2012) 'Pretreatments by means of orthogonal projections', Chemometrics and Intelligent Laboratory Systems. Elsevier B.V., 117, pp. 61–69. doi: 10.1016/j.chemolab.2012.02.002.
8. Bro, R. and Smilde, A. K. (2014) 'Principal component analysis', Analytical Methods, 6(9), pp. 2812–2831. doi: 10.1039/c3ay41907j.
9. Casale, M. and Simonetti, R. (2014) 'Review: Near infrared spectroscopy for analysingolive oils', Journal of Near Infrared Spectroscopy. IM Publications LLP, pp. 59–80. doi: 10.1255/jnirs.1106.
10. Cortés, V. et al. (2019) 'Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review', Trends in Food Science and Technology. Elsevier Ltd, pp. 138–148. doi: 10.1016/j.tifs.2019.01.015.

11. Corti, M. et al. (2018) ‘Does remote and proximal optical sensing successfully estimate maize variables? A review’, European Journal of Agronomy. Elsevier, 99(June), pp. 37–50. doi: 10.1016/j.eja.2018.06.008.
12. De Bei, R., Cozzolino, D., Sullivan, W., Cynkar, W., Fuentes, S., Dambergs, R., ... & Tyerman, S. (2011). Non-destructive measurement of grapevine water potential using near infrared spectroscopy. Australian Journal of Grape and Wine Research, 17(1), 62-71.
13. Fernández-Cuevas, I. et al. (2015) ‘Classification of factors influencing the use of infrared thermography in humans: A review’, Infrared Physics and Technology, 71, pp. 28–55. doi: 10.1016/j.infrared.2015.02.007.
14. Giovenzana, V., Beghi, R., Brancadoro, L., & Guidetti, R. (2017). Classification of wine grape based on different phytosanitary status by using visible/near infrared spectroscopy. In 37th CIOSTA and CIGR section 5th Conference.
15. Gómez-Caravaca, A. M., Maggio, R. M., & Cerretani, L. (2016). Chemometric applications to assess quality and critical parameters of virgin and extra-virgin olive oil. A review. Analytica Chimica Acta, 913, 1-21.
16. Guidetti, R.; Beghi, R.; Bodria, L. Evaluation of grape quality parameters by a simple Vis/NIR system. Transactions of the ASABE. 2010, 53, 477-484.
17. Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2065), 20150202.
18. Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. Sensors, 18(8), 2674.
19. Liu, H. F., Wu, B. H., Fan, P. G., Li, S. H., & Li, L. S. (2006). Sugar and acid concentrations in 98 grape cultivars analyzed by principal component analysis. Journal of the Science of Food and Agriculture, 86(10), 1526-1536.
20. Magwaza, L. S., Opara, U. L., Nieuwoudt, H., Cronje, P. J., Saeys, W., & Nicolaï, B. (2012). NIR spectroscopy applications for internal and external quality analysis of citrus fruit—a review. Food and Bioprocess Technology, 5(2), 425-444.
21. Malegori, C. et al. (2017) ‘Comparing the analytical performances of Micro-NIR and FT-NIR spectrometers in the evaluation of acerola fruit quality, using PLS and SVM regression algorithms’, Talanta, 165, pp. 112–116. doi: 10.1016/j.talanta.2016.12.035.
22. Malegori, C. et al. (2018) ‘GlutoPeak profile analysis for wheat classification: Skipping the refinement process’, Journal of Cereal Science. doi: 10.1016/j.jcs.2017.09.005.
23. Marquez, A.J.; Díaz, A.M.; Reguera, M.P. Using optical NIR sensor for on-line virgin olive oils characterization. Sens Actuator B Chem 2005, 107(1), 64-68

24. Nicolai, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. Postharvest biology and technology, 46(2), 99-118.
25. Oliveri, P. and Downey, G. (2013) Discriminant and class-modelling chemometric techniques for food PDO verification. 1st edn, Comprehensive Analytical Chemistry. 1st edn. Copyright © 2013 Elsevier B.V. All rights reserved. doi: 10.1016/B978-0-444-59562-1.00013-X.
26. Oliveri, P., Malegori, C., Simonetti, R., & Casale, M. (2019). The impact of signal pre-processing on the final interpretation of analytical outcomes–A tutorial. Analytica chimica acta, 1058, 9-17.
27. Patil, S.S., Thorat, S.A., 2016. Early detection of grapes diseases using machine learning and IoT, in: Proceedings - 2016 2nd International Conference on Cognitive Computing and Information Processing, CCIP 2016. https://doi.org/10.1109/CCIP.2016.7802887
28. Pu, Y., Pérez-Marín, D., O'Shea, N., & Garrido-Varo, A. (2021). Recent advances in portable and handheld NIR spectrometers and applications in milk, cheese and dairy powders. Foods, 10(10), 2377.
29. Rinnan, Å. (2014) 'Pre-processing in vibrational spectroscopy-when, why and how', Analytical Methods, 6(18), pp. 7124–7129. doi: 10.1039/c3ay42270d.
30. Spachos, P., 2020. Towards a Low-Cost Precision Viticulture System Using Internet of Things Devices. IoT 1, 5–20. https://doi.org/10.3390/iot1010002
31. Still, C., Powell, R., Aubrecht, D., Kim, Y., Helliker, B., Roberts, D., Richardson & Goulden, M. (2019). Thermal imaging in plant and ecosystem ecology: applications and challenges. Ecosphere, 10(6).
32. Stuart, M. B., McGonigle, A. J. S. and Willmott, J. R. (2019) 'Hyperspectral imaging in environmental monitoring: A review of recent developments and technological advances in compact field deployable systems', Sensors (Switzerland), 19(14). doi: 10.3390/s19143071.
33. Tefas, A. and Pitas, I. (2016) 'Principal component analysis', in Intelligent Systems. doi: 10.1201/b17700-1.
34. Todeschini, R. (1998) 'Introduzione alla chemiometria', EdiSES, Napoli. Available at: http://michem.disat.unimib.it/chm/download/materiale/introduzione_alla_chemiometria.pdf.
35. Tugnolo, A. et al. (2019) 'Characterization of green, roasted beans, and ground coffee using near infrared spectroscopy: A comparison of two devices', Journal of Near Infrared Spectroscopy, 27(1), pp. 93–104. doi: 10.1177/0967033519825665.

36. Watson B. 2003. Evaluation of wine grape maturity. In E. W. Hellman (ed.) Oregon Viticulture. Oregon states University press. Corvallis, Oregon. pp 235-245.
37. Wold, S. (1995). Chemometrics; what do we mean with it, and what do we want from it?. Chemometrics and Intelligent Laboratory Systems, 30(1), 109-115.
38. Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. Chemometrics and intelligent laboratory systems, 58(2), 109-130.
39. Yousaf, A., Khurshid, K., & Abbas, A. (2018). Modern trends in hyperspectral image analysis: A review. Ieee Access, 6, 14118-14129.
40. Zhong, Y., Ma, A., soon Ong, Y., Zhu, Z., & Zhang, L. (2018). Computational intelligence in optical remote sensing image processing. Applied Soft Computing, 64, 75-93.

A great thank goes to the members of my team

Riccardo Guidetti

Roberto Beghi

Valentina Giovenzana

Andrea Casson

Alessia Pampuri

… and to the people which helped me to reach this result!!