Elsevier Editorial System(tm) for
Postharvest Biology and Technology
Manuscript Draft

Manuscript Number:

Title: NEAR-INFRARED SPECTROSCOPY FOR DETECTION OF HAILSTORM DAMAGE ON OLIVE FRUIT

Article Type: Research Paper

Keywords: Olea europaea L., impact damage, bruise, Acousto-Optic Tunable Filter-Near Infrared spectroscopy, Discriminant Analysis, wavelengths selection

Corresponding Author: Prof. Riccardo Massantini,

Corresponding Author's Institution: University of Tuscia

First Author: Roberto Moscetti, PhD

Order of Authors: Roberto Moscetti, PhD; Ron P Haff; Danilo Monarca; Massimo Cecchini; Riccardo Massantini

Abstract: A rapid, robust, and economical method to detect hailstorm-damaged olive fruit (Olea europaea L.) would benefit both consumers and producers of olives and olive oil. Here, the feasibility of using Near-Infrared (NIR) spectroscopy for olive fruit sorting (cv. Canino) into hailstorm-damaged and undamaged classes is demonstrated. Features selected from the entire spectra by the genetic algorithm (two to six features per model) were input to Linear Discriminant Analysis, Quadratic Discriminant Analysis and k-Nearest Neighbor routines to develop models to classify olive fruit. Spectral pretreatment and feature selection were optimized through an iterative routine developed in R statistical software. Each model was evaluated based on false positive (α-error), false negative (β-error) and total error rates. The most accurate models yielded total error rates of less than five percent. The optimal features corresponded to R[1320 nm], R[~1460 nm], R[~1650 nm], R[~1920 nm], R[~2080 nm], R[~2200 nm] and R[~2220 nm], where R[x] represents the reflectance of light from the sample at a wavelength of x nm. The results indicate that single-point NIR spectroscopy is a feasible basis for hailstorm damage detection in olive fruit with the potential to allow on-line implementation on milling production lines.

Suggested Reviewers: Tom Pearson PhD
Agricultural Engineer/Lead Scientist , ARS, USDA
thomas.pearson@ars.usda.gov

Carlos Esquerre Fernandez PhD
School of Biosystems Engineering
carlos.esquerre@ucd.ie

Kawano Sumio PhD
Faculty of Agriculture
kawano@agri.kagoshima-u.ac.jp

Opposed Reviewers:

UNIVERSITÀ DEGLI STUDI DELLA Tuscia

DIBAF
DEPARTMENT FOR INNOVATION IN BIOLOGICAL, AGRO-FOOD AND FOREST SYSTEM

Via S. Camillo De Lellis - 01100 Viterbo

January 26, 2016

Dear Editor,

I am pleased to submit the attached manuscript, 'Near-infrared spectroscopy for detection of hailstorm damage on olive fruit', to be considered for the publication in the Journal of Postharvest Biology and Technology. The aim of the present study was to demonstrate the use of NIR spectroscopy for hailstorm damage detection on olive fruit. The expected impact is a new tool to help the olive industry maintain quality today and in the future with the expected increase in the frequency of hailstorm and other extreme weather events.

Looking forward your answer,

Prof. Riccardo Massantini

DIBAF – Department for Innovation in Biological, Agro-food and Forest system
Tuscia University, Viterbo, Italy.
Tel.: +39 0761 357496
Fax: +39 0761 357498
E-mail: massanti@unitus.it

HIGHLIGHTS

- NIR spectroscopy is feasible for hailstorm damage detection on olive fruit
- LDA and QDA allow reliable detection of hailstorm bruises
- 1400-1550-nm and 1900-2300-nm spectral bands are the most performing wavelengths
- This approach provides the basis for a rapid on/in/at-line detection system
- NIR based method yielded an accuracy of around 96%

***Potential Reviewers List**

Tom Pearson Ph.D.
Agricultural Engineer/Lead Scientist
USDA-ARS
1515 College Ave
Manhattan, KS 66503
785-776-2729
thomas.pearson@ars.usda.gov

Dr Carlos Esquerre Fernandez
School of Biosystems Engineering
Agriculture and Food Science
Belfield
Dublin 4
carlos.esquerre@ucd.ie

Dr Sumio Kawano
Faculty of Agriculture
Kagoshima University, 1-21-24 Korimoto, Kagoshima city 890-0065, Japan.
kawano@agri.kagoshima-u.ac.jp

NEAR-INFRARED SPECTROSCOPY FOR DETECTION OF HAILSTORM DAMAGE ON OLIVE FRUIT

Roberto Moscetti[a], Ron P. Haff[b], Danilo Monarca[c], Massimo Cecchini[c], Riccardo Massantini[a*]

[a]Department for Innovation in Biological, Agro-food and Forest system, Tuscia University, Via S. Camillo de Lellis snc, 01100, Viterbo, Italy

[b]United States Department of Agriculture, Agricultural Research Service, Western Regional Research Center, 800 Buchanan St, Albany, CA, 94710, United States

[a]Department of science and technology for Agriculture, Forest, Nature and Energy, Tuscia University, Via S. Camillo de Lellis snc, 01100, Viterbo, Italy

* Corresponding author: Tuscia University, Department for Innovation in Biological, Agro-food and Forest system, S. Camillo De Lellis snc, 01100 Viterbo, Italy. Tel.: +39 0761 357496; fax: +39 0761 357498. E-mail address: massanti@unitus.it (Massantini, R.).

ABSTRACT

A rapid, robust, and economical method to detect hailstorm-damaged olive fruit (*Olea europaea* L.) would benefit both consumers and producers of olives and olive oil. Here, the feasibility of using Near-Infrared (NIR) spectroscopy for olive fruit sorting (cv. Canino) into hailstorm-damaged and undamaged classes is demonstrated. Features selected from the entire spectra by the genetic algorithm (two to six features per model) were input to Linear Discriminant Analysis, Quadratic Discriminant Analysis and k-Nearest Neighbor routines to develop models to classify olive fruit. Spectral pretreatment and feature selection were optimized through an iterative routine developed in R statistical software. Each model was evaluated based on false positive (α-error), false negative (β-error) and total error rates. The most accurate models yielded total error rates of less than five percent. The optimal features corresponded to *R*[1320 nm], *R*[~1460 nm], *R*[~1650 nm], *R*[~1920 nm], *R*[~2080 nm], *R*[~2200 nm] and *R*[~2220 nm], where *R*[x] represents

the reflectance of light from the sample at a wavelength of x nm. The results indicate that single-point NIR spectroscopy is a feasible basis for hailstorm damage detection in olive fruit with the potential to allow on-line implementation on milling production lines.

Keywords: *Olea europaea* L., impact damage, bruise, Acousto-Optic Tunable Filter-Near Infrared spectroscopy, Discriminant Analysis, wavelengths selection

INTRODUCTION

Olea europaea (L.) is widely cultivated in the Mediterranean basin and its oil is a key component of the Mediterranean diet. The processed olive is either sold as table olives or further processed into oil. Worldwide production of olive oil increased from 1.5 million metric tons in 1990 to roughly three million metric tons in 2015, with 90% coming from Spain, Italy, or Greece (IOC, 2016). The unique chemical composition of virgin olive oil positively impacts sensory, nutritional and health properties as compared to the oil of other vegetable crops. These beneficial properties have been attributed to a fatty acid composition rich in monounsaturated acids such as oleic acid (Visioli and Galli, 2002), and to certain minor constituents such as phenolic antioxidants, including some derivative secoiridoids. Consumption of extra-virgin olive oil has been identified as a contributing factor to reduced rates of atherosclerosis, oxidative-stress associated diseases, autoimmune diseases, as well as some forms of cancer and diabetes (Katan et al., 1995; Tripoli et al., 2005; Bendini et al., 2007). Since olive fruit quality is directly correlated with the chemical composition and overall quality of produced oil, methods for removing defective fruit would benefit producers and consumers alike.

The quality of olive oil depends mainly on the quality of the fruit from which it is derived. Thus, reducing the incidence of external damage is a priority for the agri-food industry (Jiménez et al., 2016). Susceptibility to external damage such as bruising, scarring and impact damage depends on both internal and external factors of the fruit, such as cultivar, texture and firmness, maturity, dry matter, temperature, size and shape, as well as cell wall status, cell shape, etc. (Van linden et al., 2006; Jimenez-Jimenez et al., 2012). When cellular tissue is ruptured, the release of intracellular fluid causes enzymatic reactions resulting in oxidative and hydrolytic degradation of oil and phenolic fractions (Segovia-Bravo et al., 2009; Jiménez-Jiménez et al., 2013; Moscetti e al., 2015; Ramirez et al., 2015; Rojnić et al., 2015). The subsequent browning is visible within a few hours and increases over time (Ben-Shalom et al., 1978). The chief enzyme that leads to browning, Polyphenol oxidase (PPO), catalyzes the oxidation of both hydroxytyrosol and verbascoside to o-quinones, which condense to form brown polymers. Chemical oxidation of hydroxytyrosol may occur simultaneously (Martinez and Whitaker, 1995; Segovia-Bravo et al., 2011). Unless the damaged fruit is detected and removed the result is reduced quality in the produced oil. Thus, human inspection and classification of olive fruit according to quality regulations as defined by the European Commission Regulation No. 299/2013 (European Commission, 2013) is a fundamental tool to ensure the quality and economic value of olive oil. However, human inspection is labor intensive and highly subjective, and thus subject to high classification error rates. Thus, classification based on machine vision has great potential to improve quality (Guzmán et al., 2013), reduce labor expense, and minimize the impact of extreme weather events such as hailstorm. Botzen et al. (2010) estimate hailstorm frequency could increase by 25 to 50% by 2050 under standard climate change scenarios, with economic implications due to increased fruit damage. Most European Countries already experience severe hailstorms, especially in northern Italy where they are most frequent, although central and southern Italian regions are also affected (Baldi et al., 2014). Thus, as climate change progresses there will be a corresponding need for research to offset the damaging economic impacts on agriculture (Schiermeier, 2015).

Technological advances in visible/near-infrared (vis/NIR) spectroscopy and hyperspectral imaging have provided the means for real-time on-line quality assessment of both olive fruit and oil. A substantial amount of research has been reported addressing the assessment of olive fruit defects, such as bruising, impact damage, insect infestation, etc. (Stella et al., 2015a, 2015b). Although ongoing advances in computing speed and data transfer rates have enhanced the speed and accuracy of spectroscopic devices, data handling continues to be one of the more serious technological challenges to practical real-time applications (Wu and Sun, 2013). This is particularly true for hyperspectral imaging, where the "curse of dimensionality" results in a high computational load (Burger and Gowen, 2011). Consequently, for the foreseeable future, on-line applications for traditional hyperspectral imaging are limited and conventional Vis/NIR spectroscopic systems, which are well-suited to rapid on-line applications, are generally used. Overviews of chemometric methods used for data handling of Vis/NIR spectroscopic datasets are available (Nicolaï et al., 2007; Burger & Gowen, 2011). Finally, on-line systems customized for particular applications based on a limited number of wavelengths (multi-spectral imaging) in combination with data reduction schemes (ElMasry and Sun, 2010) have the potential to meet the speed requirements for an olive-oil production line.

The objective of this research was to demonstrate the use of NIR spectroscopy for hailstorm damage detection on olive fruit. The expected impact is a new tool to help the olive industry maintain quality today and in the future with the expected increase in the frequency of hailstorm and other extreme weather events.

MATERIALS AND METHODS

Sample preparation

Approximately 1.2 kg of olives (*Olea europaea* L., cv. Canino) was manually harvested on a local farm in Central Italy at the beginning of November 2014. The harvest was conducted 13 days after a severe storm with the following characteristics: Torro code H0-H3, size code 2 and hail diameter between 11 and 15 mm (Baldi et al., 2014). The storm caused significant damage to fruit and vegetation as well as production loss. In order to establish a homogeneous sample set, only mature fruits of the appropriate green color were collected. The use of fruit that were naturally damaged by hailstorm avoided the need for lab-based simulation of fruit damage, which would lead to less robust and transferrable classification models. Immediately after harvesting, the samples were taken to the laboratory in appropriate thermal boxes. From the original 1.2 kg, 752 olives were selected that were free from decay. The fruit were kept at 25 ± 0.5°C for 24 h to allow for temperature and moisture equilibrium prior to NIR spectra acquisitions.

NIR spectral acquisition

Olive spectra were acquired using a Luminar 5030 Acousto-Optic Tunable Filter-Near Infrared (AOTF-NIR) Miniature ‘Hand-held’ Analyzer (Brimrose Corp., Baltimore, USA). The instrument was equipped with a reflectance post-dispersive optical configuration, a pre-aligned dual beam lamp assembly and an indium gallium arsenide (InGaAs) array (range 1100-2300 nm, 2-nm resolution) with an integrating time of 60 msec. Each sample was measured in triplicate along the fruit equatorial line, and the average spectrum was used for further analysis. The reference spectrum was automatically measured by the instrument as described by Cayuela and Weiland (2010). Diffuse reflectance spectra were acquired by the SNAP! 2.04 software (Brimrose Corp.). Before the spectral acquisition, olives were visually evaluated for presence or absence of hailstorm damage, thus assigning each spectrum into damaged (Unsound) or not-damaged (Sound) classes.

Each olive was modeled as a ‘data vector’, where the spectral reflectance values (otherwise called features) were vector components. Features for use in classification were extracted from the whole spectra. Features were extracted following spectral pretreatments including Standard Normal Variate (SNV), Multiplicative Scatter Correction (MSC), and Savitzky-Golay first and second derivatives (*D1f* and *D2f*, respectively) with a second or third order polynomial fitted over a window of five (S5), nine (S9) or thirteen (S13) variables (Savitzky and Golay, 1964; Boysworth and Booksh, 2008). For each dataset, Mean Centering (MC) was also tested (Figure 1). Every possible combination of preprocessing was also tested and only the best results, in terms of model performance, were retained.

Chemometrics

Because of high correlation among spectral data, a Genetic Algorithm (GA) was used to select features for input to classification algorithms, with the goal of selecting a series of wavebands which could describe the correlation between the predictor variables and the response variables (Xing et al., 2008). The GA selects a small subset of spectral bands with biological or biochemical importance, which are representative of the entire spectral dataset. In this study, GA was used to seek *n*-feature subsets (where *n* ranged from 2 to 6) which are optimal surrogates for the whole dataset (Cerdeira et al., 2013). A maximum of 6 features was chosen to minimize overfitting.

Sets of features selected by the GA were input into three different classification algorithms: Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA) and *k*-Nearest Neighbors (kNN). LDA is a classification procedure based on Bayes' formula. This method renders a number of orthogonal linear discriminant functions equal to the number of categories minus 1. Thus for the case of two classes a single discriminant function is generated, allowing easier interpretation of the results. QDA is closely related to LDA and is often used when class-covariance matrices are not assumed to be identical. In case of large sample size and large differences between class-covariance matrices, QDA might outperforms LDA. kNN is an alternative method much simpler that LDA and QDA in which classification is performed by computing the sample distance from each of the samples in the training set, finding the *k* nearest ones and classifying the unknown to the class that has most members among these neighbors (Naes et al., 2004). A cross validation procedure was performed for the selection of the optimal *k* nearest neighbors (where *k* ranged from 3 to 15). The smallest *k* among those having the lowest average error was selected (Massart et al., 1988).

Fifty percent of the samples were randomly assigned to the calibration set (372 fruits) and 50% were split into three prediction sets (124 fruits each). Moreover, every model was optimized by computing a random sub-set cross validation with 6 data splits and 3 iterations. This Cross Validation method was preferred because it is very versatile and capable of providing a better representation when the dataset contains many samples, as is the case here (Naes et al., 2004; Wise, 2009).

Classification results for each discriminant model were defined as the percent of false positives (α-error rate, Sound fruit classified as Unsound) (Eq. 1), false negatives (β-error rate, Unsound fruit classified as Sound) (Eq. 2), and total error (incorrectly classified samples) (Eq. 3). Receiver Operating Characteristics (ROC) analysis was applied to each model and the respective ROC curve was plotted. The ROC curve contains a plot of the α-error rate and the true positive rate (Eq. 4) (also called sensitivity) as a function of the threshold value (Moscetti et al., 2013).

(1) $\alpha - error\ rate = \frac{False\ Positives}{False\ Positives + True\ Negatives}$

(2) $\beta - error\ rate = \frac{False\ Negatives}{True\ Positives + False\ Negatives}$

(3) $Total\ error\ rate = \frac{False\ Positives + False\ Negatives}{Total\ Positives + Total\ Negatives}$

(4) $Sensitivity = \frac{True\ Positives}{True\ Positives + False\ Negatives}$

Data analyses were performed using R 3.1.3 statistical software in combination with MASS 7.3-35, mda 0.4-7, prospectr 0.1.3 and class 7.3-11 R-packages (CRAN, 2015).

RESULTS AND DISCUSSION

Data overview

Figure 2a depicts the mean reflectance spectra for sound and unsound fruit, while Figures 2b, 2c and 2d highlight the 95% confidence interval of each calibration sample (n = 372), allowing the spectral variability of each class to be observed. Visual inspection of the spectra suggests that fruit quality is mainly correlated with spectral characteristics in the 1400-1550 nm and 1900-2300 nm wavelength regions. Reflectance in those particular spectral ranges are commonly associated with lipids, carbohydrates, proteins and water (Workman and Weyer, 2008), which represent the main constituents of the olive fruit (USDA 2015). Nevertheless, since less obvious spectral differences between classes in the other spectral regions might also be relevant the entire spectra were included in the computation.

A high percentage of features selected using the Genetic Algorithm were in wavebands around 1320 nm, 1460 nm, 1650 nm, 1920 nm, and 2000-2220 nm. Features ranging from 1150-1750 nm likely represent first overtones of O–H stretching vibrations, and overtones and combinations of –CH (CH_3–, –CH_2– and *cis* $R_1CH{=}CHR_2CH_3$–) stretches and deformations (Yildiz et al., 2001; Christy et al., 2004; Dupuy et al., 2010). Moreover, hydroxyl groups exhibit higher stretching overtones within these bands. Changes in reflectance in the aforementioned spectral band are commonly associated with phenols (~1320 nm) and hydroperoxides (~1460 nm), which are likely related to physical damage in olive fruit. In fact, phenolic content has been reported to decrease as a consequence of bruising, impact damage, pest damage, etc. (Segovia-Bravo et al., 2009, 2011), while peroxide value generally increases (Vossen, 2013). However, reflectance in that region is also correlated with aliphatic esters, alkyl groups and alkenes (~1100-1375 nm), triglycerides (~1650 nm) and water (~1460 nm) (Fernández-Espinosa, 2016). Finally, the 1920 nm and 2000-2220 nm bands can be associated with the O-H bond and the combination of C-C/C-H, C-H/C=O, and N-H/C=O groups, which are associated with water, lipids and proteins (Workman and Weyer, 2008; Vitale et al., 2013).

Differences between classes in reflectance in the water, phenols, protein/amino acid regions could be related to external damage as consequences of water loss, oxidative processes, microbial growth and/or fruit processes to repair damaged tissue. Impact damage disrupts the internal tissues, affecting moisture, fruit density and firmness (Jiménez-Jiménez et al., 2012, 2013; Jiménez et al., 2016), and these effects can be observed in the NIR spectra because of changes in light scattering. Damaged olive fruit have higher free acidity due to hydrolytic and lipolytic enzymes activity and/or concurrent freeze injury (Pereira et al., 2004; Mraicha et al., 2010; Vossen, 2013; Bustan et al., 2014).

Discriminant analysis

kNN models yielded total classification error rates on average higher than 10% and are thus not considered good candidates for implementation in on-line systems (data not shown), while both LDA and QDA models achieved more accurate discrimination. Tables 1 and 2 summarize discriminant performances for LDA and QDA, respectively, based on *n*-features subsets of pretreated spectra. In both cases, the AUC and Wilk's λ values were well correlated with the total classification error. Results of MANOVA indicate that Wilk's λ test was significant at the $Pr < 0.001$ level, suggesting that the classes differ on the basis of the extracted features for LDA and QDA classifications. Wilk's λ compares group means, with a lower value indicating a higher difference between the mean reflectance values of the classes. Thus, the higher statistical distances were observed in the best two results of both LDA and QDA algorithms.

Note that the best five results of both LDA and QDA algorithms were selected with the same combination of pretreatments and features. This suggests that: selected variables corresponded to optimal wavelengths for effectively discriminating between Sound and Unsound samples, and; classification results only differed in term of algorithms performance, which are strictly related to the characteristics of the spectral dataset. The accuracy of both LDA and QDA generally depends on the dimensionality of the training set, the multivariate normal distribution of the data and whether the class covariances are equal. Here, QDA outperformed LDA using the same pretreatments and number of features. This is likely due to large differences between class-covariance matrices, but further investigation would be necessary to confirm this.

LDA and QDA models showed very-good (< 5%) and/or good (< 10%) total classification error rates and thus potential in recognition. The best results for both LDA and QDA models were obtained using the same combination of two spectral pretreatments and five features as listed under Trial #01 in Tables 1 and 2. The QDA algorithm (1.72% α-error, 5.97% β-error, 4% total error, 0.9959 AUC) had the overall lowest error, outperforming LDA in each error class except β-error which was equal. The application of the best QDA model is shown in Figure 3a, in which only one

Sound (1.72%) and four Unsound (5.97%) olive fruits were incorrectly classified, yielding an overall accuracy of 96%. Figure 3b shows the relationship between α-error and sensitivity rates at different threshold values (i.e. posterior probability of Unsound olive fruits) and the corresponding AUC value. An AUC value close to 1.0 indicates strong discriminative ability, while an AUC value close to 0.5 indicates little discriminative power (Luo et al., 2012).

QDA Trial #2 (Figure 4) also outperformed the best LDA model and matched QDA Trial #1 in terms of total error but with a different distribution between α and β error rates (5.17% and 2.99%, respectively), a slightly lower AUC value (0.9897), and six features as opposed to five for Trial #1. The best LDA model (Table 2, Trial #01 and Figure 5a) yielded 5.17% α-error, 5.97% β-error, 5.6% total error, and a 0.9894 AUC value. Despite the higher error rates, most LDA trials yielded high AUC values indicating good ability to discriminate classes. Figure 5b shows the trend in α-error and sensitivity rates with change of the LD1-vector value (i.e. threshold value) for the best LDA model.

Spectral data generally includes correlated variables, noise from physical and/or chemical processes, and irrelevant information. Removing irrelevant variability and features from the raw dataset improves accuracy and reproducibility of classification models (Naes et al., 2004; Ozaki et al., 2006; Boysworth and Booksh, 2008). The results in this study indicate that baseline correction and derivative transformation were not necessary to obtain the best classification. On the other hand, both spectral smoothing and mean centering improved the classification accuracy for all discriminant methods, presumably by enhancing signal-to-noise ratio and spectra resolution, respectively (Boysworth and Booksh, 2008). In order to ensure model robustness and transferability, and reproducibility of results, further validation on a larger data set including olives from different cultivars, regions/orchards, crop years and agro-pedo-climatic conditions is necessary. Furthermore, improvement of model robustness and transferability should be tested by selecting wavelengths which are invariant to spectral shift (i.e. isomeric features) (Naes et al., 2004) and by extracting features as ratios or differences of absorbance/reflectance values of paired wavelengths (Moscetti et al., 2013).

CONCLUSIONS

The use of NIR spectroscopy (1100-2300 nm) to detect hailstorm-damaged olive fruits has been demonstrated. The methods developed here represent a promising step forward for sorting technologies employed by mill facilities and demonstrate that single-point NIR spectroscopy has the potential to address some of the negative impacts on food quality of the expected increased frequency of extreme weather events that are a consequence of global warming.

Discriminant models obtained total error rates of less than 5% in separating olives into sound and unsound classes based on a limited number of features derived from the spectra. Specifically, the QDA algorithm yielded 96% classification accuracy and generally outperformed the LDA models when the same combination of spectral pretreatments and features was used. Nevertheless, the major advantages and disadvantages of the LDA should be taken into account prior to implementation of the classification model for on-line sorting systems.

Optimal features selected for discriminant analysis comprised wavelengths around 1320 nm, 1400-1550 nm, 1650 nm and 2000-2220 nm, which are generally associated with phenols, hydroperoxides, alkyl groups, alkenes, triglycerides, proteins and water. However, the number of wavelengths required in an on-line sorting device would be affected by the use of spectral pretreatments, since these operations generally involve greater numbers of wavelengths than are ultimately selected for the classification model.

The proposed method lays the foundations for accurate, automated, non-destructive NIR-based detection systems. However, future model development should: employ larger data sets; test wavelengths not measured by the spectrophotometer used for this study; improve the class model simplicity; increase the light-beam intensity, and; assimilate other chemometric methods. These steps should improve the accuracy, robustness and transferability of the classification models. Since hailstorms have various intensity scales (i.e. Torro code and thus size code and hail diameter), models should be further validated, updated and upgraded on real external datasets. Finally, research is needed to allow accurate simulation of hailstorm damage in the laboratory.

REFERENCES

Baldi, M., Ciardini, V., Dalu, J.D., De Filippis, T., Maracchi, G., Dalu, G., 2014. Hail occurrence in Italy: Towards a national database and climatology. Atmos. Res. 138, 268–277. doi:10.1016/j.atmosres.2013.11.012

Bendini, A., Cerretani, L., Carrasco-Pancorbo, A., Gómez-Caravaca, A.M., Segura-Carretero, A., Fernández-Gutiérrez, A., Lercker, G., 2007. Phenolic Molecules in Virgin Olive Oils: a Survey of Their Sensory Properties, Health Effects, Antioxidant Activity and Analytical Methods. An Overview of the Last Decade Alessandra. Molecules 12, 1679–1719. doi:10.3390/12081679

Ben-Shalom, N., Harel, E., Mayer, A.M., 1978. Enzymic browning in green olives and its prevention. J. Sci. Food Agric. 29, 398–402. doi:10.1002/jsfa.2740290415

Botzen, W.J.W., Bouwer, L.M., van den Bergh, J.C.J.M., 2010. Climate change and hailstorm damage: Empirical evidence and implications for agriculture and insurance. Resour. Energy Econ. 32, 341–362. doi:10.1016/j.reseneeco.2009.10.004

Boysworth, M.K., Booksh, K.S., 2008. Aspects of multivariate calibration applied to near-infrared spectroscopy, in: Burns, D.A., Ciurczak, E.W. (Eds.), Handbook of Near-Infrared Analysis. CRC Press, New York, USA, pp. 207–229.

Burger, J., Gowen, A., 2011. Data handling in hyperspectral image analysis. Chemom. Intell. Lab. Syst. 108, 13–22. doi:10.1016/j.chemolab.2011.04.001

Bustan, A., Kerem, Z., Yermiyahu, U., Ben-Gal, A., Lichter, A., Droby, S., Zchori-Fein, E., Orbach, D., Zipori, I., Dag, A., 2014. Preharvest circumstances leading to elevated oil acidity in “Barnea” olives. Sci. Hortic. (Amsterdam). 176, 11–21. doi:10.1016/j.scienta.2014.06.028

Cayuela, J. a., Weiland, C., 2010. Intact orange quality prediction with two portable NIR spectrometers. Postharvest Biol. Technol. 58, 113–120.

Cerdeira, J.O., Silva, P.D., Cadima, J., Minhoto, M., 2013. Package “subselect” 0.12-3 for use with R. Selecting variable subsets.

Christy, A. a, Kasemsumran, S., Du, Y., Ozaki, Y., 2004. The detection and quantification of adulteration in olive oil by near-infrared spectroscopy and chemometrics. Anal. Sci. 20, 935–40.

CRAN, 2015. Comprehensive R Archive Network. URL http://cran.r-project.org/. Accessed 11.09.15.

Dupuy, N., Galtier, O., Le Dréau, Y., Pinatel, C., Kister, J., Artaud, J., 2010. Chemometric analysis of combined NIR and MIR spectra to characterize French olives. Eur. J. Lipid Sci. Technol. 112, 463–475. doi:10.1002/ejlt.200900198

ElMasry, G., Sun, D.-W., 2010. Hyperspectral imaging for food quality analysis and control, in: Sun, D.-W. (Ed.), Hyperspectral Imaging for Food Quality Analysis and Control. Academic Press, London, UK, pp. 3–43. doi:10.1017/CBO9781107415324.004

European Commission, 2013. COMMISSION IMPLEMENTING REGULATION (EU) No 299/2013 of 26 March 2013 amending Regulation (EEC) No 2568/91 on the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis. Off. J. Eur. Union 90, 52–70.

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. doi:10.1016/j.talanta.2015.10.084

Guzmán, E., Baeten, V., Pierna, J.A.F., García-Mesa, J.A., 2013. Infrared machine vision system for the automatic detection of olive fruit quality. Talanta 116, 894–898. doi:http://dx.doi.org/10.1016/j.talanta.2013.07.081

International Olive Council (IOC) 2016. World Olive Oil Figures. http://www.internationaloliveoil.org/estaticos/view/131-world-olive-oil-figures. Accessed january 21, 2016.

Jiménez, M.R., Rallo, P., Rapoport, H.F., Suárez, M.P., 2016. Distribution and timing of cell damage associated with olive fruit bruising and its use in analyzing susceptibility. Postharvest Biol. Technol. 111, 117–125. doi:10.1016/j.postharvbio.2015.07.029

Jiménez-Jiménez, F., Castro-García, S., Blanco-Roldán, G.L., Agüera-Vega, J., Gil-Ribes, J.A., 2012. Non-destructive determination of impact bruising on table olives using Vis–NIR spectroscopy. Biosyst. Eng. 113, 371–378. doi:10.1016/j.biosystemseng.2012.09.007

Jiménez-Jiménez, F., Castro-García, S., Blanco-Roldán, G.L., Ferguson, L., Rosa, U.A., Gil-Ribes, J.A., 2013. Table olive cultivar susceptibility to impact bruising. Postharvest Biol. Technol. 86, 100–106. doi:10.1016/j.postharvbio.2013.06.024

Katan, M.B., Zock, P.L., Mensink, R.P., 1995. Dietary oils, serum lipoproteins, and coronary heart disease. Am. J. Clin. Nutr.

Luo, X., Takahashi, T., Kyo, K., Zhang, S., 2012. Wavelength selection in vis/NIR spectra for detection of bruises on apples by ROC analysis. J. Food Eng. 109, 457–466. doi:10.1016/j.jfoodeng.2011.10.035

Martinez, M.V., Whitaker, J.R., 1995. The biochemistry and control of enzymatic browning. Trends Food Sci. Technol. 6, 195–200. doi:10.1016/S0924-2244(00)89054-8

Massart, D.L., Vandeginste, B.G.M., Deming, S.N., Michotte, Y., Kaufman, L., 1988. Chemometrics: a textbook, in: Data Handling in Science and Technology. Elsevier Science Publishers B.V., Netherlands, pp. 339–355, 373–378.

Moscetti, R., Haff, R.P., Aernouts, B., Saeys, W., Monarca, D., Cecchini, M., Massantini, R., 2013. Feasibility of Vis / NIR spectroscopy for detection of flaws in hazelnut kernels. J. Food Eng. 1–7.

Moscetti, R., Haff, R.P., Stella, E., Contini, M., Monarca, D., Cecchini, M., Massantini, R., 2015. Feasibility of NIR spectroscopy to detect olive fruit infested by Bactrocera oleae. Postharvest Biol. Technol. 99, 58–62.

Mraicha, F., Ksantini, M., Zouch, O., Ayadi, M., Sayadi, S., Bouaziz, M., 2010. Effect of olive fruit fly infestation on the quality of olive oil from Chemlali cultivar during ripening. Food Chem. Toxicol. 48, 3235–41.

Naes, T., Isaksson, T., Fearn, T., Davies, T., 2004. A user-friendly guide to Multivariate Calibration and Classification. NIR publications, Chichester, UK.

Ozaki, Y., McClure, W.F., Christy, A.A. (Eds.), 2006. Near-Infrared Spectroscopy in Food Science and Technology. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Pereira, J.A., Alves, M.R., Casal, S., Oliveira, M.B.P.P., 2004. Effect of olive fruit fly infestation on the quality of olive oil from cultivars cobrançosa, madural and verdeal transmontana.

Ramirez, E., Sanchez, A.H., Romero, C., Brenes, M., 2015. Combined use of nitrogen and coatings to improve the quality of mechanically harvested Manzanilla olives. Food Chem 171, 50–55. doi:10.1016/j.foodchem.2014.08.103

Rojnić, I.D., Bažok, R., Barčić, J.I., 2015. Reduction of olive fruit fly damage by early harvesting and impact on oil quality parameters. Eur. J. Lipid Sci. Technol. 117, 103–111. doi:10.1002/ejlt.201400150

Savitzky, A., Golay, M.J.E., 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Anal. Chem. 36, 1627–1639.

Schiermeier, Q., 2015. Quest for climate - proof farms. Nat. 523 396–397.

Segovia-Bravo, K.A., García-garcía, P., Antonio, L., Garrido-fern, A., 2011. Effect of Bruising on Respiration , Superficial Color , and Phenolic Changes in Fresh Manzanilla Olives (Olea europaea pomiformis): Development of Treatments To Mitigate Browning. J. Agric. Food Chem. 59, 5456–5464.

Segovia-Bravo, K.A., Jarén-Galán, M., García-García, P., Garrido-Fernández, A., 2009. Browning reactions in olives: Mechanism and polyphenols involved. Food Chem. 114, 1380–1385. doi:10.1016/j.foodchem.2008.11.017

Stella, E., Moscetti, R., Haff, R.P., Monarca, D., Cecchini, M., Contini, M., Massantini, R., 2015a. Review : Recent advances in the use of non-destructive near infrared spectroscopy for intact olive fruits. J. Near Infrared Spectrosc. 23, 197. doi:10.1255/jnirs.1169

Stella, E., Moscetti, R., Massantini, R., Monarca, D., Cecchini, M., Haff, R.P., Contini, M., 2015b. Recent advances in the use of NIR spectroscopy for qualitative control and protection of extra virgin olive oil. La Riv. Ital. della sostanze grasse XCII, 155–173.

Tripoli, E., Giammanco, M., Tabacchi, G., Di Maio, D., Giammanco, S., La Guardia, M., 2005. The phenolic compounds of olive oil: structure, biological activity and beneficial effects on human health. Nutr. Res. Rev. 1, 98–112.

USDA, 2015. National Nutrient Database for Standard Reference – Release 26. URL http://ndb.nal.usda.gov/ndb/search/list. Accessed 11.09.15.

Van linden, V., Scheerlinck, N., Desmet, M., De Baerdemaeker, J., 2006. Factors that affect tomato bruise development as a result of mechanical impact. Postharvest Biol. Technol. 42, 260–270. doi:10.1016/j.postharvbio.2006.07.001

Visioli, F., Galli, C., 2002. Biological properties of olive oil phytochemicals. Crit. Rev. Food Sci. Nutr. 42, 209–21. doi:10.1080/10408690290825529

Vitale, R., Bevilacqua, M., Bucci, R., Magrì, A.D., Magrì, A.L., Marini, F., 2013. A rapid and non-invasive method for authenticating the origin of pistachio samples by NIR spectroscopy and chemometrics. Chemom. Intell. Lab. Syst. 121, 90–99.

Vossen, P., 2013. Handbook of Olive Oil - Analysis and Properties, in: Aparicio, R., Harwood, J. (Eds.), Handbook of Olive Oil Analysis and Properties. Springer, New York, USA, pp. 19–56.

Wise, B.M., 2009. Eigen vector research wiki - Using Cross-Validation. URL http://wiki.eigenvector.com. Accessed 11.09.2015.

Workman, J., Weyer, L., 2008. Practical Guide to Interpretive Near-Infrared Spectroscopy. CRC Press, London, UK.

Wu, D., Sun, D.-W., 2013. Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review — Part II: Applications. Innov. Food Sci. Emerg. Technol. 19, 15–28. doi:10.1016/j.ifset.2013.04.016

Xing, J., Guyer, D., Ariana, D., Lu, R., 2008. Determining optimal wavebands using genetic algorithm for detection of internal insect infestation in tart cherry. Sens. Instrum. Food Qual. Saf. 2, 161–167.

Yildiz, G., Wehling, R.L., Cuppett, S.L., 2001. Method for determining oxidation of vegetable oils by near-infrared spectroscopy. J. Am. Oil Chem. Soc. 78, 495–502.

FIGURE CAPTIONS

Figure 1. *Flowchart of the iterative algorithm used to optimize the classification model for hailstorm damage detection on olive fruit based on the AOTF-NIR spectroscopy system.*

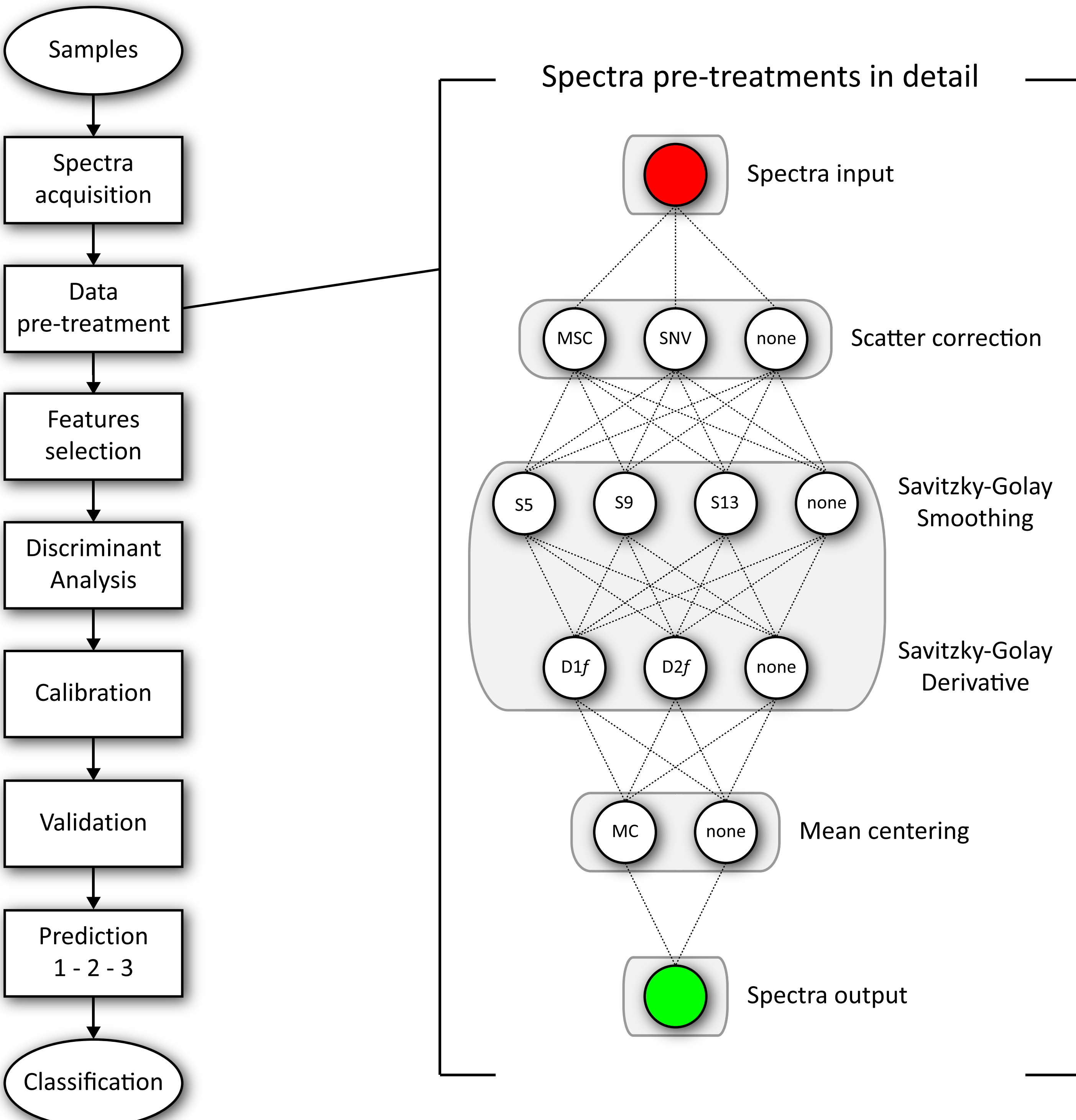
Figure 2. *Mean reflectance spectra (a) and 95% confidence intervals for different spectral bands: 1100-1498-nm (b), 1500-1898-nm (c) and 1900-2300-nm (d).*

Figure 3. *Class spatial distribution plot (a) and ROC curve plot and relative AUC value (b) referred to the 5-features QDA model corresponding to the trial #01 of Table 1.*

Figure 4. *Class spatial distribution plot (a) and ROC curve plot and relative AUC value (b) referred to the 6-features QDA model corresponding to the trial #02 of Table 1.*

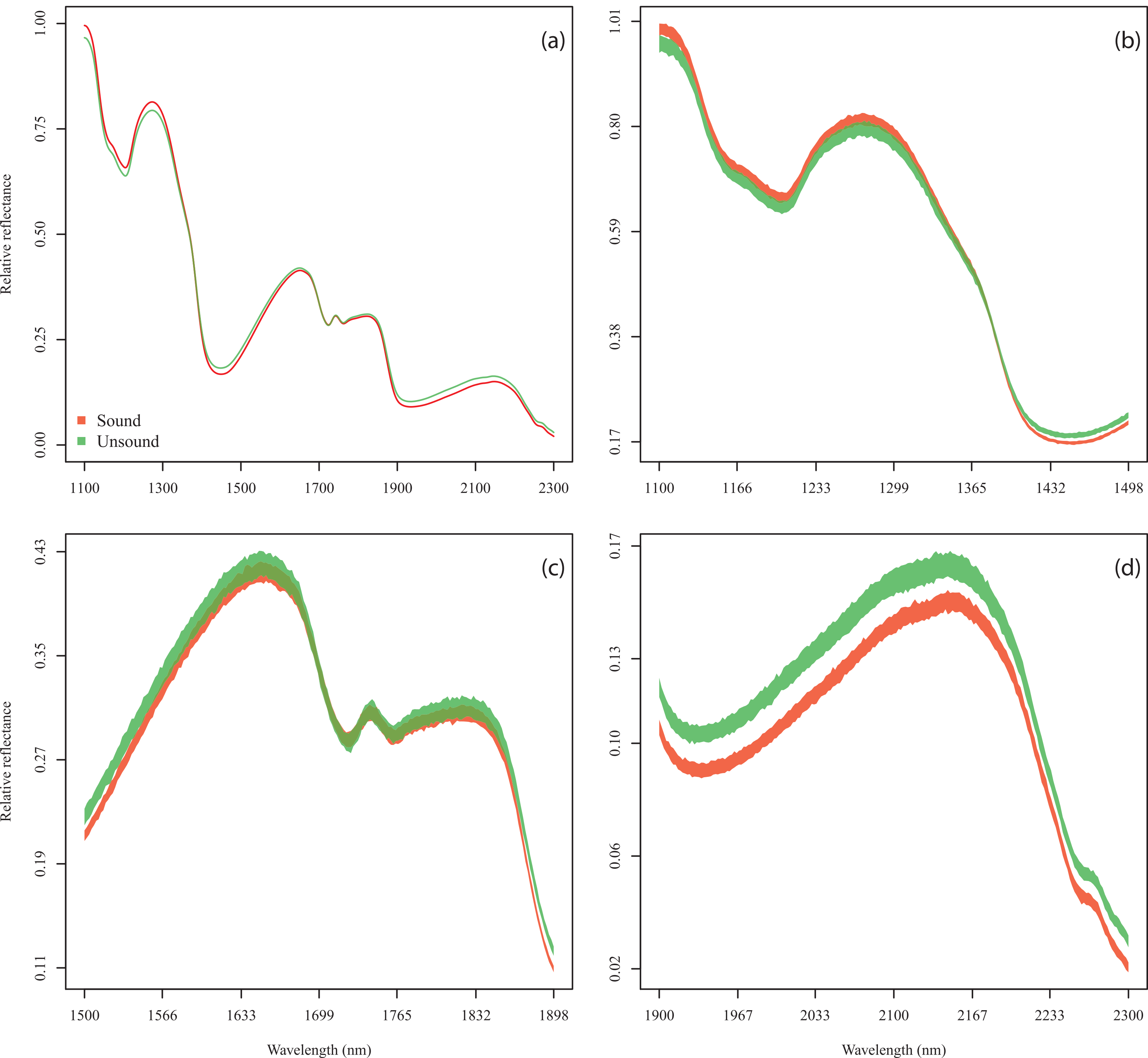
Figure 5. *Class spatial distribution plot (a) and ROC curve plot and relative AUC value (b) referred to the 5-features LDA model corresponding to the trial #01 of Table 2.*

Figure 1



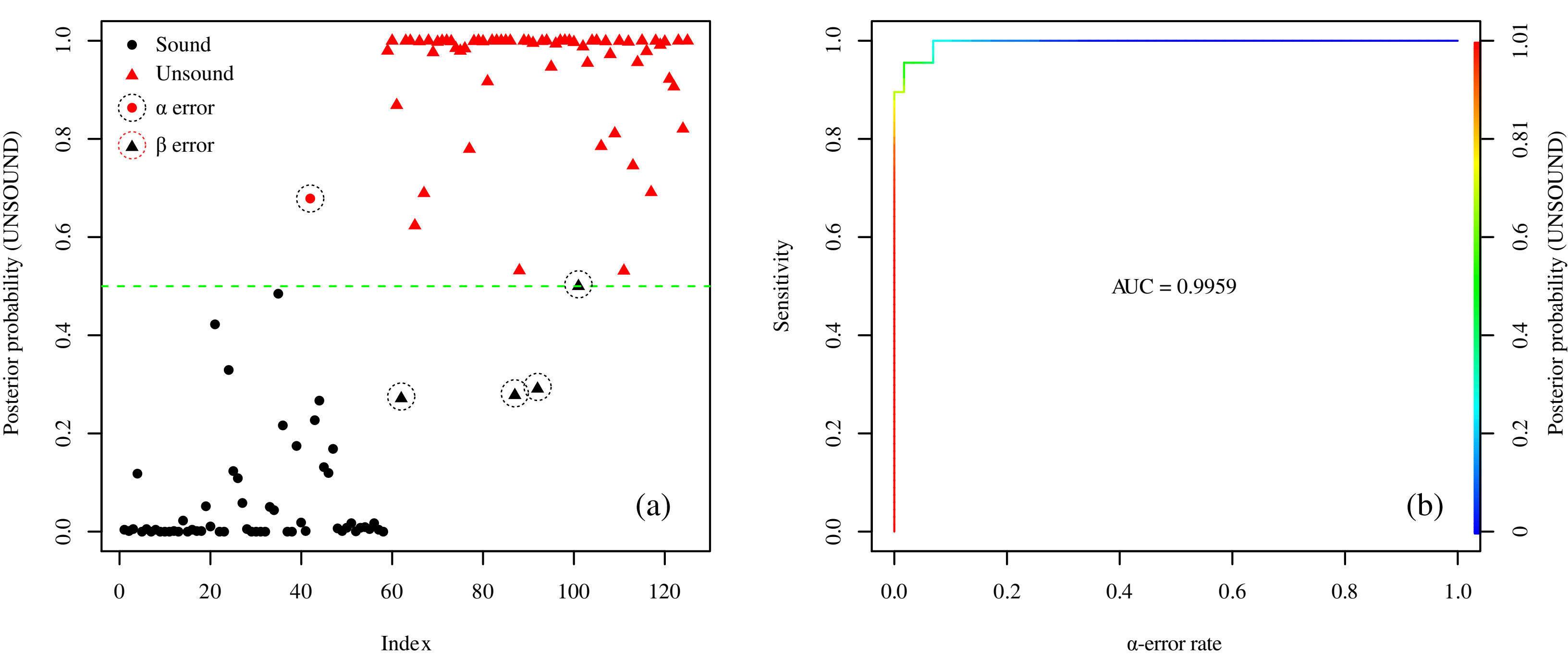
Samples
Spectra acquisition
Data pre-treatment
Features selection
Discriminant Analysis
Calibration
Validation
Prediction 1 - 2 - 3
Classification
Spectra pre-treatments in detail
Spectra input
MSC
SNV
none
Scatter correction
S5
S9
S13
none
Savitzky-Golay Smoothing
D1*f*
D2*f*
none
Savitzky-Golay Derivative
MC
none
Mean centering
Spectra output

Figure 2



(a)
(b)
(c)
(d)
Sound
Unsound
Relative reflectance
Wavelength (nm)
1.00
0.75
0.50
0.25
0.00
1100
1300
1500
1700
1900
2100
2300
1.01
0.80
0.59
0.38
0.17
1166
1233
1299
1365
1432
1498
0.43
0.35
0.27
0.19
0.11
1566
1633
1699
1765
1832
1898
0.17
0.13
0.10
0.06
0.02
1967
2033
2100
2167
2233

Figure 3



Sound
Unsound
α error
β error
Posterior probability (UNSOUND)
Index
(a)
Sensitivity
α-error rate
AUC = 0.9959
Posterior probability (UNSOUND)
(b)

Figure 4

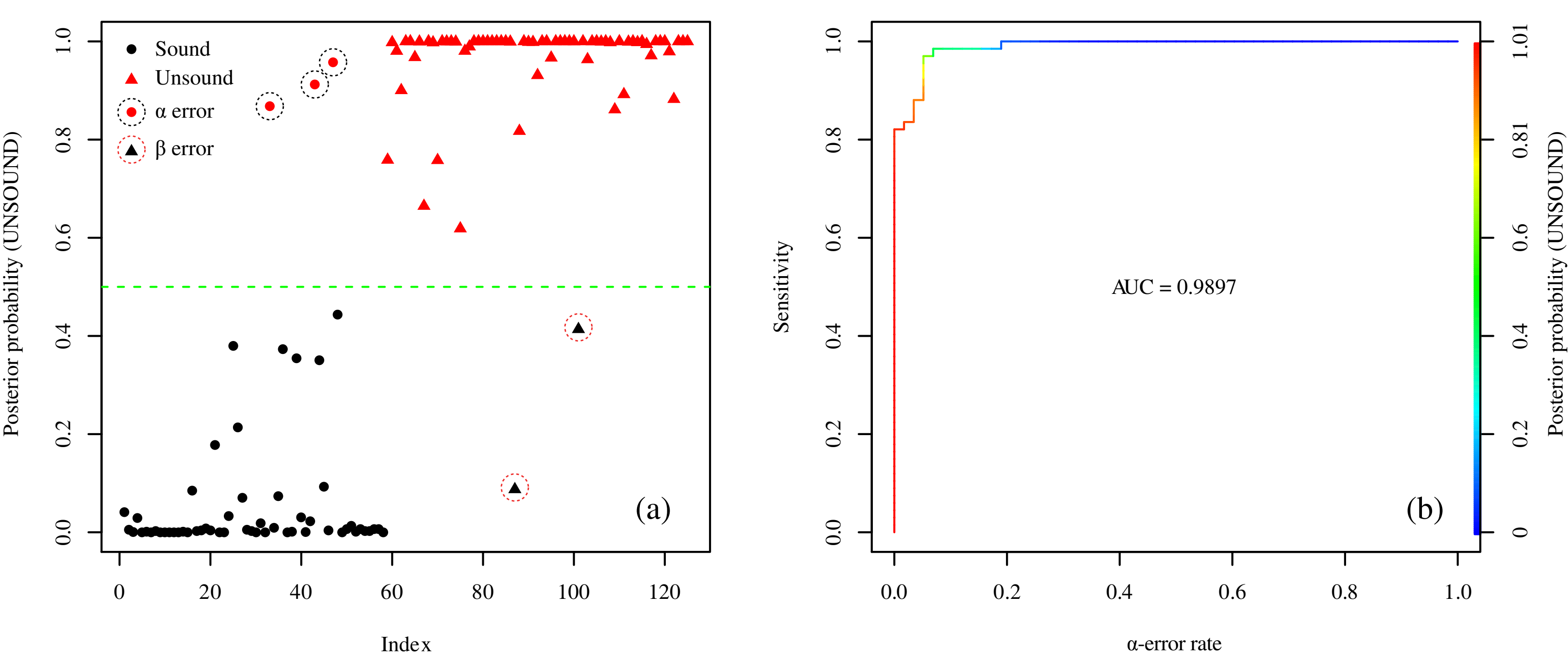
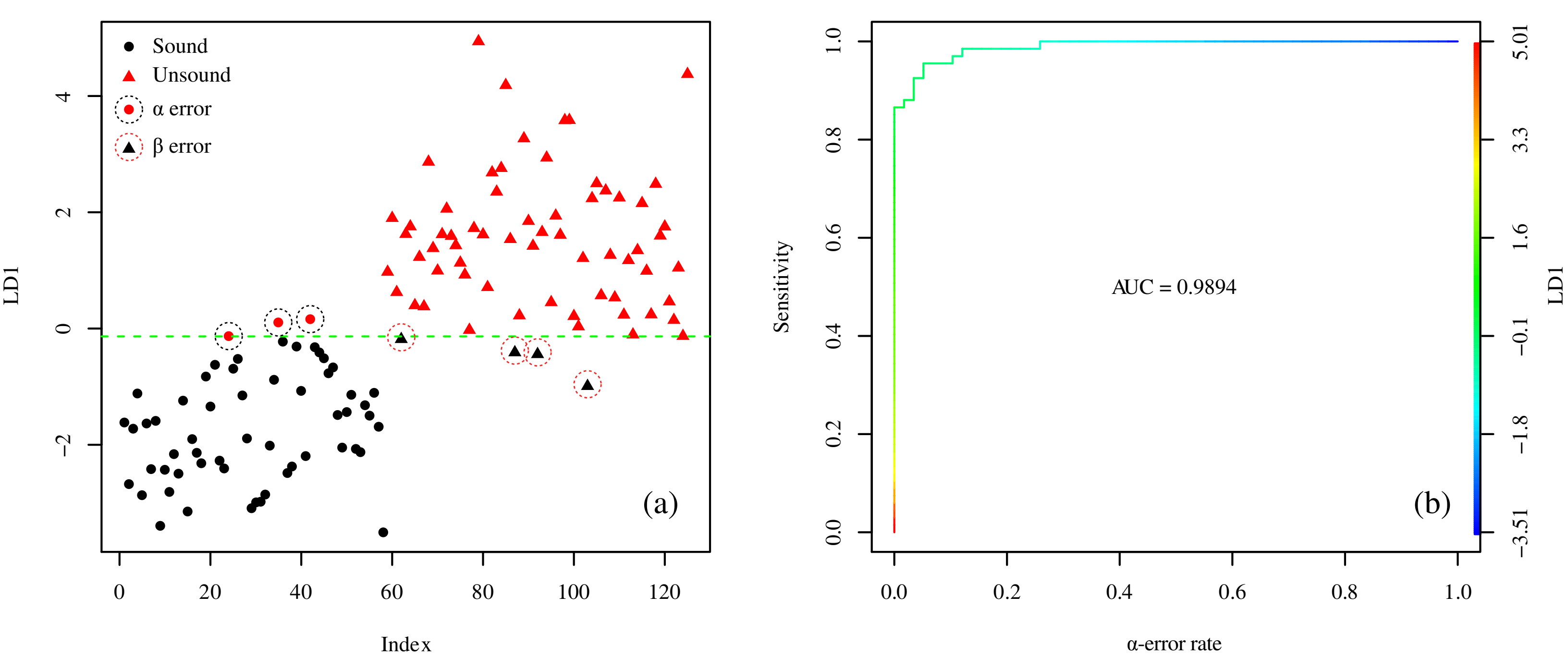

Sound
Unsound
α error
β error
Posterior probability (UNSOUND)
Index
(a)
Sensitivity
α-error rate
AUC = 0.9897
(b)
Posterior probability (UNSOUND)

Figure 5



Sound
Unsound
α error
β error
LD1
Index
(a)
Sensitivity
α-error rate
AUC = 0.9894
(b)
LD1
5.01
3.3
1.6
−0.1
−1.8
−3.51

TABLE CAPTIONS

Table 1. Results based on pretreatments, GA and QDA performed on the spectra acquired from olive fruit. The selected features corresponded to the wavelengths yielding the lowest error rates of classification per each dataset.

[a] Scatter correction

[b] Mean Centering

[c] Savitzky-Golay filter smoothing points

[d] Derivative (first – *D1f* and second - *D2f*)

Table 2. Results based on pretreatments, GA and LDA performed on the spectra acquired from olive fruit. The selected features corresponded to the wavelengths yielding the lowest error rates of classification per each dataset.

[a] Scatter correction

[b] Mean Centering

[c] Savitzky-Golay filter smoothing points

[d] Derivative (first – D1*f* and second - *D2f*)

Table 1

Trial #	Dataset				Selected Features (nm)						Error rate (%)			AUC	Wilks' λ	Pr (> *F*)
	Scatt. C.[a]	MC[b]	SG[c]	Deriv.[d]	1	2	3	4	5	6	α	β	Total			
01	-	Yes	9	-	1452	1922	2088	2198	2222	-	1.72	5.97	4.00	0.9959	0.3194	< 0.001
02	-	Yes	9	-	1320	1478	1648	2082	2198	2220	5.17	2.99	4.00	0.9897	0.2928	< 0.001
03	-	Yes	9	-	1450	1918	2086	2204	-	-	6.90	5.97	6.40	0.9781	0.3634	< 0.001
04	-	Yes	9	-	1896	2090	2196	-	-	-	10.34	11.94	11.20	0.9689	0.4069	< 0.001
05	-	Yes	9	*D1f*	1398	1714	-	-	-	-	31.03	17.91	24.00	0.7910	0.7713	< 0.001

Table 2

Trial #	Dataset				Selected Features (nm)						Error rate (%)			AUC	Wilks' λ	Pr (> *F*)
	Scatt. C.[a]	MC[b]	SG[c]	Deriv.[d]	1	2	3	4	5	6	α	β	Total			
01	-	Yes	9	-	1452	1922	2088	2198	2222	-	5.17	5.97	5.60	0.9894	0.3194	< 0.001
02	-	Yes	9	-	1320	1478	1648	2082	2198	2220	6.90	5.97	6.40	0.9866	0.2928	< 0.001
03	-	Yes	9	-	1896	2090	2196	-	-	-	12.07	10.45	11.20	0.9624	0.4069	< 0.001
04	-	Yes	9	-	1450	1918	2086	2204	-	-	8.62	13.43	11.20	0.9704	0.3634	< 0.001
05	-	Yes	9	*D1f*	1398	1714	-	-	-	-	29.31	19.40	24.00	0.7849	0.7713	< 0.001