HIGHLIGHTS

- NIR spectroscopy is a feasible method for detection of olive fruit fly damage
- QDA allows reliable detection of infested olives
- Features were selected within the 1100-1250 nm and 1400-1500 spectral bands
- The suggested approach provides the basis for a rapid online detection system
- The NIR based method yielded an accuracy of around 94%

FEASIBILITY OF NIR SPECTROSCOPY TO DETECT OLIVE FRUIT INFESTED BY *Bactrocera oleae*

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

[a]Department of Science and Technology for Agriculture, Forest, Nature and Energy, 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.

[c]Department for Innovation in Biological, Agro-food and Forest System, 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

Olive fruit fly infestation is a significant problem for the milling process. In most cases, damage from insects is 'hidden', i.e. not visually detectable on the fruit surface. Consequently, traditional visual sorting techniques are generally inadequate for the detection and removal of olives with insect damage. In this study, the feasibility of using NIR spectroscopy to detect hidden insect damage is demonstrated. Using a genetic algorithm for feature selection (from 2 to 6 wavelengths) in combination with Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA) or *k*-Nearest-Neighbors (*k*NN) routines, classification error rates as low as 0.00% false negative, 12.50% false positive, and 6.25% total error were achieved, with an AUC value of 0.9766 and a Wilk's λ of 0.3686 ($P < 0.001$). Multiplicative Scatter Correction, Savitzky-Golay spectral pre-treatment with 13 smoothing points and Mean Centering spectral pre-treatments were used. The

optimal features corresponded to *Abs*[1108 nm], *Abs*[1232 nm], *Abs*[1416 nm], *Abs*[1486 nm] and *Abs*[2148 nm].

INTRODUCTION

Olea europaea is one of the most important and widespread fruit trees cultivated in the Mediterranean basin, where it has important environmental, economical and social significance. A portion of the olive production is processed for direct human consumption but most is used for the production of oil, for which worldwide consumption multiplied 6-fold over the last 30 years.

Among the various vegetable oils, virgin olive oil is unique for many reasons, enclosing its chemical composition which differs from that of other vegetable oils used by humans, containing unique compounds that positively impact sensory, nutritional and health properties. The highest quality olive oils, denominated extra-virgin, contain a high concentration of these compounds. Extra-virgin olive oil is a key component of the Mediterranean diet and is considered, at least in part, to contribute to reduced incidence of cardiovascular diseases observed in this region (Katan et al., 1995).

The chemical composition and subsequent quality of extra-virgin olive oil depends entirely on the quality of the fruit from which it is derived. Thus, methods for improving fruit quality or removing damaged or defective fruit have a direct impact on the quality of oil produced.

Bactrocera oleae (olive fruit fly) is the most significant pest of olives worldwide (Daane and Johnson, 2010) and one of the most frequent causes of reduced olive oil quality. The detrimental effects are related to both the severity of infestation and on the stage of the fly development. Infestation occurs when the adult female pierces the fruit and lays eggs just under the surface. The developing larvae causes extensive damage by feeding, excavating deep tunnels which can reach the stone (Rice, 2000). This facilitates the penetration and development of microorganisms, with accompanying loss of fruit integrity and oil quality.

It has been demonstrated that infestation by *Bactrocera oleae* reduces oil yield and alters the chemical composition of the olive fruit, negatively affecting many olive oil qualitative parameters such as free acidity, peroxide value and ultraviolet absorption (Gómez-Caravaca et al., 2008). Moreover, the infested olives produce oil with a reduced content of antioxidant phenolics (Gómez-Caravaca et al., 2008; Gucci et al., 2012) and an altered volatile compound profile (Angerosa et al., 2004) leading to severe and unacceptable off-flavour. Consequently, the nutritional value and the sensory properties of oil extracted from infested olives are compromised, and the product often does not conform to legal specification for extra-virgin or virgin oils.

Currently, the loss of oil quality due to infested olives is unavoidable, since processing procedures authorized by the European Union for virgin olive oil do not account for infestation. However, since not all olives in an infested lot are infested, an excellent product could be produced if good product could be separate from defective product. This is the underlying motivation to utilize Near-Infrared (NIR) spectroscopy for detection and removal of olives infested by *Bactrocera olea*.

NIR spectroscopy has been proven effective for the detection of insects or insect damage in food commodities and seeds such as chestnut (Moscetti et al., 2014a, 2014b), blueberry (Peshlov et al., 2009), cherry (Xing and Guyer, 2008; Xing et al., 2008), fig (Burks et al., 2000), flour (Wilkin et al., 1986), green soybean (Sirisomboon et al., 2009), jujube (Wang et al., 2010, 2011), mangoes (Haff et al., 2013), seeds of the *Larix* species (Tigabu and Odén, 2004), seeds of *Picea abies* (Tigabu et al., 2004), seeds of *Cordia africana* (Tigabu and Odén, 2002) and others. Insects and larvae can be detected directly, due to their hemolymph, lipids and/or chitin content (Rajendran, 2005), or indirectly due to subsequent damage such as internal browning or darkening, dehydration or fungi contamination (Wang et al., 2011). However, on-line inspection systems for infestation of fresh produce, including olives, are still not in common use. NIR techniques have the potential to benefit the food market by reducing the risk of buying poor-quality products and consequently allowing compliance with consumer demands for uniform high-quality products (Butz et al., 2005).

78 The objective of the present study was to investigate the feasibility of using the NIR
79 spectroscopy for detection of olives infested by the olive fruit fly and identifying combinations of
80 features (based on absorbance of light in the NIR band from 1100 to 2300 nm) having optimal
81 discriminatory ability and testing different classification methods.
82
83 **MATERIALS AND METHODS**
84 **Sample preparation**
85 Approximately 1.2 kg of olives (*Olea europaea* L., cv. canino) were manually harvested on
86 a local farm in Central Italy at the end of October, and were immediately taken to the laboratory in
87 appropriate thermal boxes. From these, 896 olives that were free from visual external impact
88 damage and/or decay were selected. The fruit were kept at 25 ± 0.5°C for 24 h to allow for
89 temperature and moisture equilibrium prior to NIR spectra acquisitions.
90 **NIR spectral acquisition**
91 Olive spectra were acquired using a Luminar 5030 Acousto-Optic Tunable Filter-Near
92 Infrared (AOTF-NIR) Miniature 'Hand-held' Analyzer (Brimrose Corp., Baltimore, USA). The
93 instrument was equipped with a reflectance post-dispersive optical configuration, a pre-aligned dual
94 beam lamp assembly and an indium gallium arsenide (InGaAs) array (range 1100-2300 nm, 2-nm
95 resolution) with an integrating time of 60 ms. Two spectra were acquired on each of two opposite
96 side of the fruit along the equatorial line and averaged. Each acquired spectrum was the average of
97 10 scans. The reference spectrum was automatically measured by the instrument as described by
98 Cayuela and Weiland (2010). Diffuse reflectance spectra were acquired and transformed into
99 absorbance ($A = \log T^{-1}$) using R 3.1.0 statistic software (CRAN, 2014). Immediately after the
100 spectral acquisition, olives were dissected to visually determine presence or absence of olive fruit
101 fly larvae, thus assigning each spectrum into infested (Unsound) or not-infested (Sound) classes.
102 **Statistical analysis of NIR spectra**

Each olive was modeled as a 'data vector', where the spectral absorbance values (otherwise called features) were vector components. Principle Component Analysis (PCA) was applied to evaluate between-class similarity. The original spectral data was converted to score and loading vectors by PCA analysis. The scree-plot criterion (Jolliffe, 2002) was used to select the required number of PCs for describing the dimensionality of the data.

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, second and third derivatives (*df*, *d2f* and *d3f*) with second order polynomial (from 5 nm to 13 nm smoothing points with a step of 4 nm) (Savitzky and Golay, 1964; Boysworth and Booksh, 2008). For each dataset, Mean Centering (MC) was also tested. Data preprocessing (transformation and data reduction) can dramatically influence the final results of recognition for spectral data, which may contain highly correlated variables, noise and irrelevant information caused directly by scattering or adsorption of NIR light due to variable interaction of the various types of compounds (Wu et al., 1996; Tallada et al., 2011). Spectra pre-treatments can help remove the influence of pericarp thickness and skin condition, and enhance spectral differences between classes. However, spectral information that can be useful for the classification models could be lost in pretreatment process. The aforementioned preprocessing techniques were tested and the worst-fit pretreatments for classification purposes were discarded by evaluating the robustness and the accuracy of each model.

A Genetic Algorithm (GA) was used to select features for input to classification algorithms, with the obvious 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 spectra 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 (*k*NN). 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). 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 outperform LDA. *k*NN 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).

Following the random subset selection method, the sample was split as follows: 50% of the samples were assigned to a training set (224 sound and 224 unsound fruits), 25% to a validation set (112 sound and 112 unsound fruits) and 25% to a test set (112 sound and 112 unsound fruits). The random subset selection circumvents overfitting problems and avoids overly optimistic results. No outlier selection was computed.

Classification results for each discriminant model were determined as the percent of false positives (fp, Sound fruit classified as Unsound), false negatives (fn, Unsound fruit classified as Sound), and total error (incorrectly classified samples). The Receiver Operating Characteristics (ROC) analysis was performed for evaluating the performance of each discriminant model, and the respective ROC curve was plotted. The ROC curve contains a plot of the 'false positive rate' and the 'true positive rate' (fraction of positives correctly classified) (also respectively called '1 – specificity' and 'sensitivity') as a function of the threshold value. For each ROC plot the relative

Area Under the ROC Curve (AUC) was additionally used to evaluate the performance of each computed discrimination function (Moscetti et al., 2013).

Moreover, each feature subset was analyzed using multivariate analysis of variance (MANOVA), and the Wilks' λ Test was used to determine if the differences between classes were significant.

Data analyses were performed using R 3.1.0 statistical software in combination with MASS 7.3-32, mda 0.4-4, pcaMethods 1.54.0, prospectr 0.1.3 and class 7.3-10 R-packages (CRAN, 2014).

RESULTS AND DISCUSSION

The PCA analysis of data variations provides insight into the class distribution of the non-pretreated spectra (Fig. 1). The first two principal components (PCs) accounted for 98.34% of the variance. Score values for Unsound olives were tightly clustered over the two-dimensional space, while scores for Sound fruit were widely distributed. Mean-raw spectra showed higher absorbance in Unsound fruit in the range of 1900 nm, and showed low absorbance between ~1100 nm and ~1400 nm, and between ~1550 and ~1860 nm (Fig. 1b), suggesting possible differences between the two classes.

Tables 1 and 2 summarize discriminant performances for LDA and QDA based on *n*-features subsets of pre-treated or untreated spectra. In both cases, the AUC and Wilk's λ values were well correlated with the total classification error. For QDA the correlation was weaker, although a high AUC still coincides with a low total classification error (Table 2, Trial #01). The results indicate that baseline correction was necessary to obtain the lowest classification errors and the best model accuracy in most cases. Mean centering improved the classification accuracy for most of the discriminant methods used, probably by enhancing the differences between spectra (Boysworth and Booksh, 2008).

The interpretation of NIR spectra is complicated by the fact that specific bands are composites of many bands containing information on various molecular vibration and generic

functional groups. However, knowledge of sample chemical composition allows some speculation regarding the wavelengths selected for class discrimination: the major constituents in raw olives are water (~80.0%), lipids (~10.7%), carbohydrates (~6.3%), and fibers (~3.2%) (USDA, 2014).

In some cases, water could interfere with NIR analysis due to the strength and the broadband characteristics of water absorption (Rasco et al., 1991; Peshlov et al., 2009). Consequently, water bands may be excluded from the computations if the goal is to develop infestation-classification methods based on other minor constituents (Ghaedian and Wehling, 1997; Peshlov et al., 2009). As stated earlier, however, for olives the change in water content due to infestation is expected to improve discriminant performances and the water bands were not excluded. In fact, in the case of qualitative analysis, the moisture content can be one of the most important factors for classifying samples (Tsuchikawa, 2007).

Features at ~1100-1250 nm and ~1400-1500 nm were frequently paired as features by the GA. Changes in absorbance in each of these wavelength regions are commonly associated with hydroperoxides and phenols related with the fruit fly damage. In fact, phenolic content has been reported to decrease as a consequence of the olive fruit fly attack (Gómez-Caravaca et al., 2008; Tamendjari et al., 2009). Highly infested olives generally exhibit a strong increase in peroxide value (Pereira et al., 2004; Gucci et al., 2012).

Differences in absorption in the water and protein/amino acid regions (~1170 nm, ~1400-1500 nm and ~2150-2250 nm) could be related to infestation as consequences of microbial growth and/or fruit processes to repair damaged tissue (Workman and Weyer, 2007). In fact, fungal growth also disrupts the internal tissues, affecting fruit density and firmness, and thus the overall information of the NIR spectra because of changes in light scattering. The microbe-damaged olive fruit have higher free acidity due to hydrolytic enzymes and lipolytic activity of bacteria and fungi (Pereira et al., 2004; Mraicha et al., 2010).

LDA and QDA results indicate good recognition. Data preprocessing, including data transformation and feature selection, improved recognition accuracy. Spectral data are usually

affected by issues due to correlated variables, noise and irrelevant information. Removing or reducing non-relevant variability as well as selecting relevant features from the raw dataset were crucial for the development of accurate classification models.

The *k*NN model yielded the worst classification results (data not shown). Both LDA and QDA models were capable of performing accurate discrimination of Unsound fruit. The accuracy of 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. In this case, QDA outperformed LDA due to large differences between class-covariance matrices. The score plot (Fig. 1a) suggests a much larger variance for the Sound vs. the Unsound class.

The most accurate classification of 12.50% fp, 0.00% fn, and 6.25% total error was achieved using QDA (Table 2, Trial #01) using MSC, 13 point Savitsky-Golay filter and MC. Second best corresponded to 18.75% fp, 0.00% fn, and 9.38% total error, using LDA with MSC, 13 point Savitsky-Golay filter and MC. In terms of total error, QDA performed better than LDA, with a lower number of features and the same pretreatments. AUC values higher for QDA indicating superior discrimination. Spatial distribution plots (Figure 2) illustrate posterior-probability differences between LDA and QDA. The application of the best QDA model is shown in Figure 2b, in which only 14 Sound olives (12.50%) were incorrectly classified, yielding an average accuracy of 93.75%.

Results of MANOVA indicate that Wilk's λ test (Tables 1 and 2) was significant at the $P < 0.001$ 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 higher difference between the mean absorbance values of the classes. Thus, the higher statistical distance was observed in Trial #01, Table 1.

CONCLUSIONS

In general, the use of NIR spectroscopy in the long wavelength region (1100-2300 nm) as an automated, non-destructive and rapid means to detect larvae attack olives appeared feasible, with classification accuracy as high as almost 94%. Misclassified fruit were primarily assigned to Sound olives. As such, NIR spectroscopy has the potential for detection of olive fruit fly infestation in olives. Optimal features selected for discriminant analysis comprised wavelengths that are generally associated with water, peroxides and phenols (~1100-1250 nm and ~1400-1500 nm).

Classification accuracy was highest for QDA with a 13 point SG filter, MSC and mean centering, yielding 0.00% fn, 12.50% fp and 6.25% total error. Second lowest error rates of 18.75% fp, 0.00% fn and 9.38% total error were obtained using LDA with a 13 point SG filter, MSC and mean centering. Nevertheless, the major advantages and disadvantages of the LDA should be taken into account prior to implementation of the classification model for an automated sorting system.

The methods used here provide the basis for a superior detection system in terms of sensitivity, selectivity, non-destructive analysis and automation. However, the number of wavelengths required in an online sorting device would be conditioned by the spectral pre-treatment used.

The good results obtained represent a promising step forward for sorting technologies employed by mill facilities. Moreover, the proposed NIR method could be useful as a screening tool for quality control and for studying olive fruit fly incidence directly in the orchard by using a portable spectrophotometer.

Finally, future research should include testing wavelengths not measured by the spectrophotometer used for this study, improving the class model simplicity, increasing the light-beam intensity, and/or combining other chemometric methods, as these could improve the accuracy and the robustness of the sorting system.

REFERENCES

257 Angerosa, F., Servili, M., Selvaggini, R., Taticchi, A., Esposto, S., Montedoro, G., 2004. Volatile
258 compounds in virgin olive oil: occurrence and their relationship with the quality. J.
259 Chromatogr. A 1054, 17–31.
260 Boysworth, M.K., Booksh, K.S., 2008. Aspects of multivariate calibration applied to near-infrared
261 spectroscopy, in: Burns, D.A., Ciurczak, E.W. (Eds.), Handbook of Near-Infrared Analysis.
262 CRC Press, New York, USA, pp. 207–229.
263 Burks, C., Dowell, F., Xie, F., 2000. Measuring fig quality using near-infrared spectroscopy. J.
264 Stored Prod. Res. 36, 289–296.
265 Butz, P., Hofmann, C., Tauscher, B., 2005. Recent developments in noninvasive techniques for
266 fresh fruit and vegetable internal quality analysis. J. Food Sci. 70, 131–141.
267 Cayuela, J. a., Weiland, C., 2010. Intact orange quality prediction with two portable NIR
268 spectrometers. Postharvest Biol. Technol. 58, 113–120.
269 Cerdeira, J.O., Silva, P.D., Cadima, J., Minhoto, M., 2013. Package "subselect" 0.12-3 for use with
270 R. Selecting variable subsets.
271 CRAN, 2014. Comprehensive R Archive Network. URL http://cran.r-project.org/. Accessed
272 04.28.14.
273 Daane, K.M., Johnson, M.W., 2010. Olive fruit fly: managing an ancient pest in modern times.
274 Annu. Rev. Entomol. 55, 151–69.
275 Ghaedian, A.., Wehling, R.L., 1997. Discrimination of sound and granary-weevil-larva-infested
276 wheat kernels by near-infrared diffuse reflectance spectroscopy. J. AOAC Int. 80, 997–
277 1005.
278 Gómez-Caravaca, A.M., Cerretani, L., Bendini, A., Segura-Carretero, A., Fernández-Gutiérrez, A.,
279 Del Carlo, M., Compagnone, D., Cichelli, A., 2008. Effects of fly attack (Bactrocera oleae)
280 on the phenolic profile and selected chemical parameters of olive oil. J. Agric. Food Chem.
281 56, 4577–83.

Gucci, R., Caruso, G., Canale, A., Loni, A., Raspi, A., Urbani, S., Taticchi, A., Esposto, S., Servili, M., 2012. Qualitative changes of olive oils obtained from fruits damaged by Bactrocera oleae (Rossi). HortScience 47, 301–306.

Haff, R.P., Saranwong, S., Thanapase, W., Janhiran, A., Kasemsumran, S., Kawano, S., 2013. Automatic image analysis and spot classification for detection of fruit fly infestation in hyperspectral images of mangoes. Postharvest Biol. Technol. 86, 23–28.

Jolliffe, I.T., 2002. Principal Component Analysis. Springer, New York, USA.

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

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., Saranwong, S., Monarca, D., Cecchini, M., Massantini, R., 2014a. Nondestructive detection of insect infested chestnuts based on NIR spectroscopy. Postharvest Biol. Technol. 87, 88–94.

Moscetti, R., Monarca, D., Cecchini, M., Haff, R.P., Contini, M., Massantini, R., 2014b. Detection of Mold-Damaged Chestnuts by Near-Infrared Spectroscopy. Postharvest Biol. Technol. 93, 83–90.

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.

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.

Peshlov, B., Dowell, F., Drummond, F., Donahue, D., 2009. Comparison of three near infrared spectrophotometers for infestation detection in wild blueberries using multivariate calibration models. J. Near Infrared Spectrosc. 17, 203.

Rajendran, S., 2005. Detection of insect infestation in stored foods. Adv. Food Nutr. Res. 49, 163–232.

Rasco, B.A., Miller, C.E., King, T.L., 1991. Utilization of NIR spectroscopy to estimate the proximate composition of trout muscle with minimal sample pretreatment. J. Agric. Food Chem. 39, 67–72.

Rice, R., 2000. Bionomics of the olive fruit fly, Bactrocera (Dacus) oleae. Univ. Calif. plant Prot. Q. 10, 1–5.

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

Sirisomboon, P., Hashimoto, Y., Tanaka, M., 2009. Study on non-destructive evaluation methods for defect pods for green soybean processing by near-infrared spectroscopy. J. Food Eng. 93, 502–512.

Tallada, J.G., Wicklow, D.T., Pearson, T.C., Armstrong, P.R., 2011. Detection of fungus infected corn kernels using near infrared reflectance spectroscopy and color imaging. Trans. ASABE 54, 1151–1158.

Tamendjari, A., Angerosa, F., Mettouchi, S., Bellal, M.M., 2009. The effect of fly attack (Bactrocera oleae) on the quality and phenolic content of Chemlal olive oil. Grasas y Aceites 60, 509–515.

Tigabu, M., Odén, P., 2002. Multivariate classification of sound and insect-infested seeds of a tropical multipurpose tree, Cordia africana, with near infrared reflectance spectroscopy. J. Near Infrared Spectrosc. 10, 45.

334 Tigabu, M., Odén, P.C., 2004. Simultaneous detection of filled, empty and insect-infested seeds of
335 three Larix species with single seed near-infrared transmittance spectroscopy. New For. 27,
336 39–53.
337 Tigabu, M., Odén, P.C., Shen, T.Y., 2004. Application of near-infrared spectroscopy for the
338 detection of internal insect infestation in Picea abies seed lots. Can. J. For. Res. 34, 76–84.
339 Tsuchikawa, S., 2007. Sampling techniques, in: Ozaki, Y., McClure, W.F., Christy, A.A. (Eds.),
340 Near-Infrared Spectroscopy in Food Science and Technology. John Wiley & Sons, Inc.,
341 Hoboken, New Jersey, United States, pp. 133–144.
342 USDA, 2014. National Nutrient Database for Standard Reference – Release 26. URL
343 http://ndb.nal.usda.gov/ndb/search/list. Accessed 03.12.14.
344 Wang, J., Nakano, K., Ohashi, S., 2011. Nondestructive detection of internal insect infestation in
345 jujubes using visible and near-infrared spectroscopy. Postharvest Biol. Technol. 59, 272–
346 279.
347 Wang, J., Nakano, K., Ohashi, S., Takizawa, K., He, J.G., 2010. Comparison of different modes of
348 visible and near-infrared spectroscopy for detecting internal insect infestation in jujubes. J.
349 Food Eng. 101, 78–84.
350 Wilkin, D.R., Cowe, I.A., Thind, B.B., McNicol, J.W., Cuthbertson, D.C., 1986. The detection and
351 measurement of mite infestation in animal feed using near infra-red reflectance. J. Agric.
352 Sci. 107, 439–448.
353 Workman, J., Weyer, L., 2007. Practical Guide to Interpretive Near-Infrared Spectroscopy. CRC
354 Press, London, UK.
355 Wu, W., Mallet, Y., Walczak, B., 1996. Comparison of regularized discriminant analysis linear
356 discriminant analysis and quadratic discriminant analysis applied to NIR data. Anal. Chim.
357 Acta 2670, 257–265.
358 Xing, J., Guyer, D., 2008. Comparison of transmittance and reflectance to detect insect infestation
359 in Montmorency tart cherry. Comput. Electron. Agric. 64, 194–201.

360 Xing, J., Guyer, D., Ariana, D., Lu, R., 2008. Determining optimal wavebands using genetic
361 algorithm for detection of internal insect infestation in tart cherry. Sens. Instrum. Food Qual.
362 Saf. 2, 161–167.
363

***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

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	fp	fn	total			
01	MSC	Yes	13	-	1178	1226	1246	1566	2150	2246	18.75	0.00	9.38	0.9531	0.3525	< 0.001
02	SNV	Yes	9	*df*	1168	1420	1602	1712	-	-	12.50	6.25	9.38	0.9531	0.4590	< 0.001
03	MSC	Yes	13	-	1108	1232	1416	1486	2148	-	6.25	18.75	12.50	0.9570	0.3686	< 0.001
04	MSC	Yes	9	-	1232	1764	2248	-	-	-	12.50	18.75	15.63	0.9141	0.4658	< 0.001
05	SNV	-	5	*df*	1420	1710	-	-	-	-	31.25	12.50	21.88	0.8984	0.5328	< 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	fp	fn	total			
01	MSC	Yes	13	-	1108	1232	1416	1486	2148	-	12.50	0.00	6.25	0.9766	0.3686	< 0.001
02	-	-	13	*df*	1108	1168	1418	1466	1710	1984	6.25	12.50	9.38	0.9922	0.3758	< 0.001
03	-	Yes	13	*df*	1138	1418	1464	-	-	-	12.50	18.75	15.63	0.9297	0.4588	< 0.001
04	SNV	-	9	*df*	1168	1420	1602	1712	-	-	25.00	6.25	15.63	0.9531	0.4590	< 0.001
05	MSC	Yes	5	*df*	1422	1712	-	-	-	-	31.25	12.50	21.88	0.8984	0.5874	< 0.001

TABLE CAPTIONS

Table 1. 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 (Multiplicative Scatter Correction – MSC, Standard Normal Variate – SNV)

[b] Mean Centering

[c] Savitzky-Golay filter smoothing points

[d] Derivative (first - *df*, second - *d2f* and third - *d3f*)

Table 2. 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 (Multiplicative Scatter Correction – MSC, Standard Normal Variate – SNV)

[b] Mean Centering

[c] Savitzky-Golay filter smoothing points

[d] Derivative (first - *df*, second - *d2f* and third - *d3f*)

Figure 1

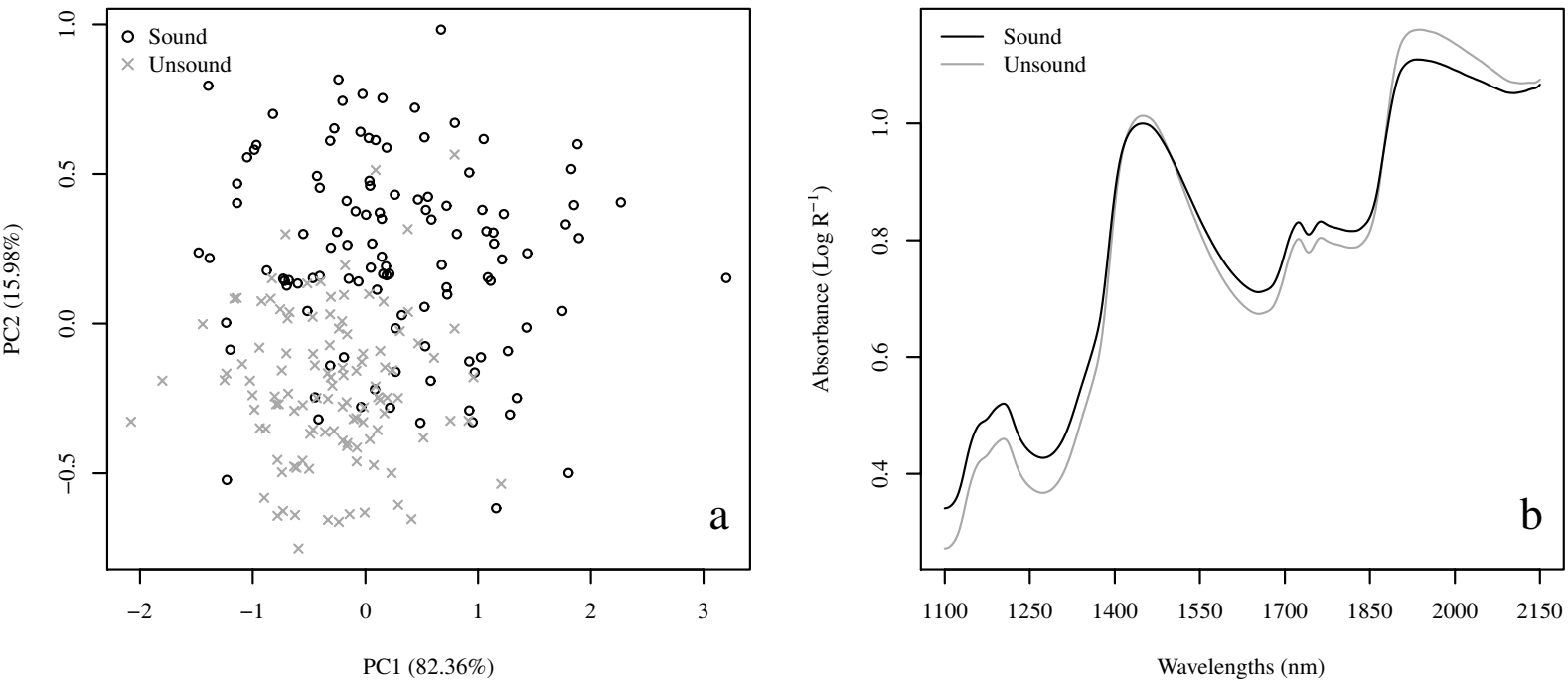
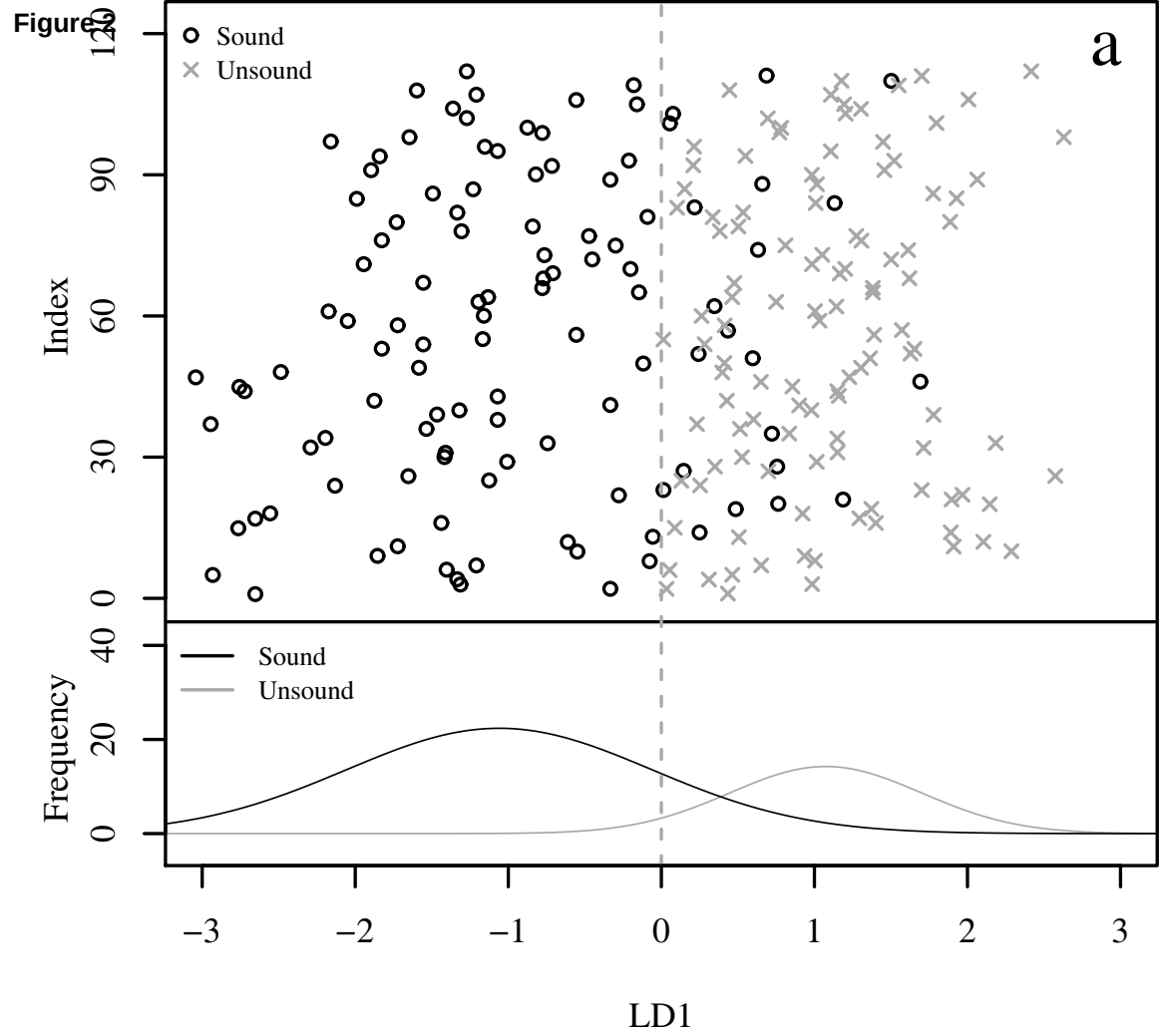

Sound
Unsound
a
PC2 (15.98%)
PC1 (82.36%)
−2
−1
0
1
2
3
−0.5
0.0
0.5
1.0
Sound
Unsound
b
Absorbance (Log R^{-1})
Wavelengths (nm)
0.4
0.6
0.8
1.0
1100
1250
1400
1550
1700
1850
2000
2150

Figure 2


Sound
Unsound
a
Index
0
30
60
90
120
Frequency
0
20
40
Sound
Unsound
−3
−2
−1
0
1
2
3
LD1

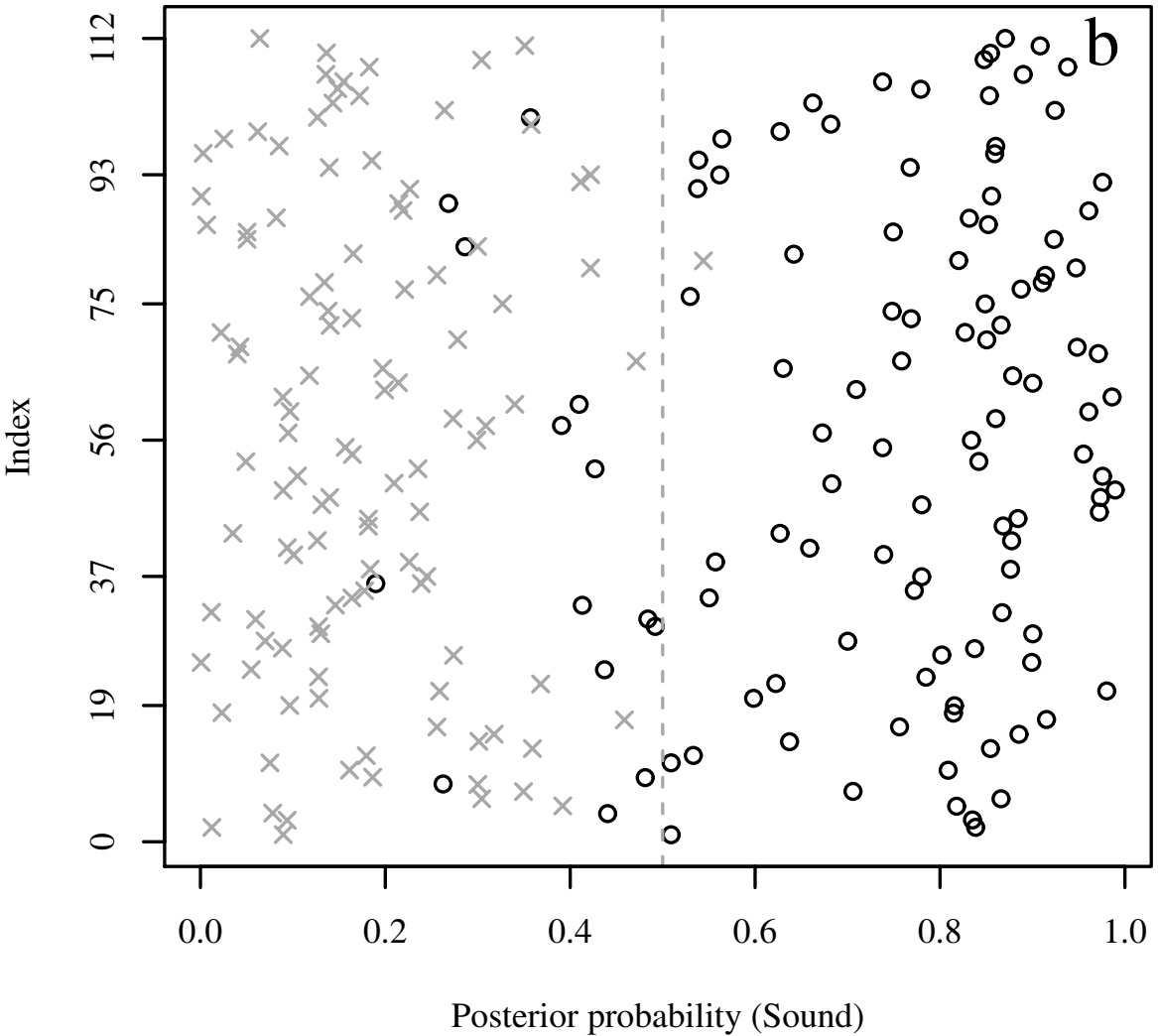

b
Index
0
19
37
56
75
93
112
0.0
0.2
0.4
0.6
0.8
1.0
Posterior probability (Sound)

FIGURE CAPTIONS

Figure 1. PCA score plot computed from the whole spectra acquired from not-infested (Sound) and infested (Unsound) olive fruit (a). Sound and Unsound spectra at the range of 1100-2150 nm (b).

Figure 2. Class spatial distribution plots obtained from LDA (a) (Trial #1, Table 1) and QDA (b) (Trial #1, Table 2) models for each group of features selected. Lower part of the (a) represents the frequency plot of the LD1 function computed from the six selected features.