Postharvest Biology and Technology 99 (2015) 58–62

Contents lists available at ScienceDirect

Postharvest Biology and Technology

journal homepage: www.elsevier.com/locate/postharvbio

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

ARTICLE INFO

Article history:
Received 29 April 2014
Accepted 24 July 2014
Available online 29 August 2014

Keywords:
Olea europaea L.
Insect damage
Acousto-optic tunable filter-near infrared spectroscopy
Discriminant analysis
Wavelengths selection

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].

1. 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 *B. 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,

* Corresponding author. Tel.: +39 0761 357496; fax: +39 0761 357498.
E-mail address: massanti@unitus.it (R. Massantini).

http://dx.doi.org/10.1016/j.postharvbio.2014.07.015
0925-5214/

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).

The objective of the present study was to investigate the feasibility of using the NIR spectroscopy for detection of olives infested by the olive fruit fly and identifying combinations of features (based on absorbance of light in the NIR band from 1100 to 2300 nm) having optimal discriminatory ability and testing different classification methods.

2. Materials and methods

2.1. Sample preparation

Approximately 1.2 kg of olives (*Olea europaea* L., cv. canino) were manually harvested on a local farm in Central Italy at the end of October, and were immediately taken to the laboratory in appropriate thermal boxes. From these, 896 olives that were free from visual external impact damage and/or decay were selected. The fruit were kept at $25 \pm 0.5\,^{\circ}C$ for 24 h to allow for temperature and moisture equilibrium prior to NIR spectra acquisitions.

2.2. 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 ms. Two spectra were acquired on each of two opposite side of the fruit along the equatorial line and averaged. Each acquired spectrum was the average of 10 scans. The reference spectrum was automatically measured by the instrument as described by Cayuela and Weiland (2010). Diffuse reflectance spectra were acquired and transformed into absorbance ($A = \log R^{-1}$) using R 3.1.0 statistic software (CRAN, 2014). Immediately after the spectral acquisition, olives were dissected to visually determine presence or absence of olive fruit fly larvae, thus assigning each spectrum into infested (unsound) or not-infested (sound) classes.

2.3. 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 six 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.