HAZELNUT QUALITY SORTING USING HIGH DYNAMIC RANGE SHORT WAVE INFRARED HYPERSPECTRAL IMAGING

Roberto Moscetti[a], Wouter Saeys[b], Janos C. Keresztes[b], Mohammad Goodarzi[b], Massimo Cecchini[a], Monarca Danilo[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] KU Leuven Department of Biosystems, MeBioS, Kasteelpark Arenberg 30, 3001 Leuven, Belgium

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

* Corresponding author at: 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.).

INTRODUCTION

Hazelnut (*Corylus avellana* L.) is one of the most economically important raw materials for the confectionary, bakery and chocolate industries. Approximately 90% of the worldwide crop is absorbed by the food industry (Moscetti et al., 2012).

Hazelnuts are a rich source of energy (62.8 kcal g^{-1}), with a 61% lipid content (w/w) primarily composed of oleic acid [18:1 (n-3)] triacylglycerols (Shahidi and Miraliakbari 2005), and they contain other minor constituents, including polyphenols (Alasalvar and Shahidi 2008) and fat-soluble vitamins (Chun *et al.* 2006). Other components include carbohydrates (16.7% w/w), proteins (14.9% w/w), water (5.3% w/w), sugars (4.3% w/w) and dietary fiber (9.7% w/w) (USDA 2014). Thanks to its interesting composition, hazelnut has been identified as a contributing factor to the beneficial effects of the Mediterranean diet (Shahidi and Miraliakbari 2005) and is used to add bioactive compounds to new health-related foods, serving as a substitute for synthetic ingredients (Kibar and Öztürk 2009; Contini *et al.* 2011).

Hazelnuts, being fatty foods, are very susceptible to rancidity, leading to the development of undesirable odors and flavors and decreased nutritional value (Ghirardello *et al.* 2013). Extrinsic and intrinsic factors such as the temperature, relative humidity (RH), soil, storage duration, insect damage, fungal growth, moisture content, water activity (a_w) and presence or absence of a shell, can affect the quality of tree nuts like hazelnuts (Cecchini *et al.* 2011; Moscetti *et al.* 2013). It is important to consider that the interaction of all these factors may also be responsible for various kernel degradation processes, even if the kernel composition is initially identical. Moreover, many factors (RH, moisture content, a_w, insect damage and fungal growth) are mainly related to the occurrence of oxidative processes, changes in the activity of hydrolytic enzymes (Miles 1972; Sipahioglu and Heperkan 2000; Capinera 2001; Vaccinio *et al.* 2008; Amaike and Keller 2011) and promotion of localized Maillard chemical reactions (non-enzymatic browning) (Pearson 1999; Özdemir *et al.* 2001). The hydrolysis of oils following the lipolytic activity produces FFAs and catalysis auto-oxidation reactions (Parcerisa *et al.* 1997; Köksal *et al.* 2006), developing off-flavor, off-odor and bitterness, thereby reducing consumer acceptability (Sipahioglu and Heperkan 2000). Therefore, the rate and degree of hazelnut flaws depends primarily on the harvest, drying and storage conditions.

As hazelnuts naturally fall to the ground when mature, the first nuts dropped lie on the ground for extended periods before the remaining nuts are collected. This protracted contact between the nuts and soil increases the chance of loss in quality and processing properties as the nuts remain under humid or rainy conditions, which can trigger the aforementioned degradation factors and processes (Massantini et al. 2009).

Currently, loss in quality of hazelnuts remains an economic burden for the food industry. Thus, the ability to measure and predict quality changes that are occurring in the kernel using a fast and easy-to-use method could be very valuable for the industry.

Sensors for real-time, nondestructive systems using non-contact optical techniques (*e.g.* imaging techniques or spectroscopy) for food quality and safety inspection are becoming an integral part of the food industry's move towards automation. Imaging technologies can easily provide excellent spatial information on the external attributes of foods, such as the size, shape, color, surface texture and external defects, by ordinary means (*e.g.* RGB color camera) (ElMasry and Sun, 2010). As such, computer vision is often used for nondestructive, automated quality inspection (Nguyen Do Trong et al., 2011). Unfortunately, this imaging technique has some drawbacks that make it inefficient for certain industrial applications (*e.g.* complex classifications of objects of similar colors and detection of invisible defects) and, above all, provides very little spectral information for the measurement of internal attributes of food products (*e.g.* chemical composition) (ElMasry et al. 2012). Conversely, spectroscopic measurement of food products is one of the most successful nondestructive techniques for quality assessment and can simultaneously provide information on several quality properties related to the chemical composition and microstructure. Visible (Vis) and near infrared (NIR) spectroscopy is fast, noninvasive, flexible, easy to use, environmentally benign and well-suited to rapid on-line and off-line applications compared with conventional laboratory methods used to determine food quality (Nicolaï et al. 2007; Moscetti et al. 2014a, 2014b). Although Vis/NIR spectroscopy merges simplicity and accuracy for the prediction of the major components in foods, it provides no information on the compositional distribution within one product. Thus, Vis/NIR spectroscopy is limited to homogeneous products or products where the distribution of the components is not of interest (ElMasry et al. 2012).

Hyper- and multispectral imaging combines the advantages of Vis/NIR spectroscopy and imaging technologies by acquiring spectral information for each pixel on the product. This information then forms a 3-D dataset (hypercube) that can be used to characterize the food more reliably compared to traditional imaging or spectroscopy techniques.

The feasibility of Vis/NIR spectroscopy to detect flawed hazelnuts was already demonstrated by Moscetti et al. (2013) and even showed the potential to separate nuts into quality classes as defined by the European Commission Regulation No. 1284/2002 (European Commission 2002). However, as a spectrum has to be acquired from every individual kernel, this method is not suitable for hazelnut sorting in industry. Furthermore, mechanical detectors are commercially available for sorting hazelnuts in quality classes. However, due to high error rates in the automated devices, subsequent manual inspection by trained workers is still standard practice. Therefore, the objectives of the present study was to investigate the potential of high dynamic range–Short Wave Infrared hyperspectral imaging in the spectral range 850–2500 nm for the classification of hazelnut kernels into the quality classes currently used in industry.

MATERIALS AND METHODS

Sample preparation

Whole, raw (unroasted) and shelled hazelnut kernels (*Corylus avellana* L., cv. Tonda Gentile Romana) were provided by the Assofrutti facility (Viterbo, Central Italy). At the outset, kernels were classified into the following four quality classes, using the guidelines of the Commission Regulation No. 1284/2002 (European Commission 2002): ‘Extra Class’ (superior quality - C_X), ‘Class I’ (good quality - C_I), ‘Class II’ (fruits that do not qualify for inclusion in the higher classes, but satisfy the minimum requirements - C_{II}) and ‘Waste’ (fruits which do not satisfy the minimum requirements - W_S). The kernels were classified using a proprietary mechanical detector at the facility along with manual selection performed by expert workers. Classification was based on the presence/absence of discolorations, infestations, infections and/or detrimental disorders, including kernel size and shape. Class batches were composed of 600 fruits (~750 g) each. Hazelnut kernels had an average moisture content of 4.00±0.13%.

Hyperspectral imaging system

Image acquisition was performed using a Hyperspectral Imaging (HSI) system (Fig. 1) consisting of a Short Wave Infrared (SWIR) camera, an illumination source, a linear translation stage and a control system. The HSI setup was built to be suitable for scanning linearly-moving hazelnut samples and was well shielded in a dark room to prevent interference from external light sources.

The SWIR camera was a XEVA-FPA-2.5-320 model (Xenics Vision, Leuven, Belgium), equipped with a mercury-cadmium-tellurium (HgCdTe) detector array (850-2500 nm sensitivity, ~7.5 nm resolution) with a 320×256 pixel Focal Plane Array (FPA) with 30 μm pitches, a reflective concentric grating (Headwall Photonics Inc., Fitchburg, MA, USA) and a four-stage Peltier cooler operating at -85°C.

The illumination source consisted of four 20 W halogen spots model DC 2800 K (OSRAM, Munich, Germany) powered by a DC model 380 W (Antec, Fremont, CA, USA). Halogen lamps are the most common illumination sources used in NIR spectroscopy (ElMasry and Sun 2010). However, the two main disadvantages of halogen lamps are variation in light intensity due to voltage instabilities and spectral peak shift due to large heat generation. Therefore, a DC power supply was used to avoid operating voltage fluctuation and thus output variations, while samples were exposed to the light for the shortest time possible with the aim of circumventing spectral peak shift due to temperature change. Furthermore, light spots were arranged in an arc to improve the intensity and uniformity of the light along the field of view of the camera and to improve the responsivity of the camera.

A linear-translation stage model TLA15-400 (Franke GmbH, Aalen, Germany) was used to move the samples through the field of view of the camera. It was equipped with a twenty-hole sample holder (5×4, five samples per class) (Fig. 1) to avoid micro-vibrations of the kernels during the scanning and to consequently acquire sharp images. Positioning of the hazelnut classes on the sample holder was randomly permuted at each scan (120 iterations) to avoid misleading statistical inferences due to camera artifacts and pixel misregistration problems.

A desktop computer with Labview software v8.5 (National Instrument Corporation, Austin, TX, USA) was used as the control system for synchronizing the linear-translation stage movement with the camera acquisition. The linear translation stage speed was set to 3 mm s^{-1}, and images were consequently captured by the camera at intervals of 0.5 mm.

The hypercube was built using Matlab software R2012a (Mathworks, Natick, MA, USA).

The procedure for hazelnut class detection and comparisons of the obtained results are summarized in a flowchart (Fig. 2).

Spectra acquisition

Variations in responsiveness of the camera, also known as 'pattern noise', were corrected by performing the reflectance calibration to account for the background spectral response and the 'dark' current of the camera. A 75% Spectralon diffuse reflectance standard model RSS-08-010 (Labsphere, Inc., North Sutton, NH, USA) was used to collect the background spectral response by recording the spectral and pixel variations of the system's response. The Spectralon standard was chosen for its affinity with the hazelnuts' reflectance over the entire SWIR range, to avoid spectral saturation due to free-skin spots (i.e. yellow-white areas) on the kernel surface of some hazelnuts and to obtain higher signal-to-noise ratio. Moreover, the internal camera noises caused by the 'dark' current were stored in 'dark' images, which were acquired by covering the camera lens with a non-reflective opaque black cap. Reference and 'dark' images were acquired after every twenty scans, which corresponded to an interval of approximately 1 h and 30 min. Finally, the raw reflectance spectral images that had been acquired were converted to pixel-based, relative-reflectance spectral images using the following equation:

(1) $$I = \frac{I_0 - D}{W - D}$$

where I corresponds to the 'relative-reflectance' image, I_0 represents the 'raw-spectral' image of the sample, W stands for the 'white-reference' image, and D corresponds to the 'dark-current' image.

The 'dark' images were also used to account for the variations in the 'dark' current of the camera during the measurements.

Ideally, the signal to noise ratio of a spectral imaging system should be constant over the full wavelength range (Wang et al. 2012). However, the camera used in this experimentation had a lower sensitivity in the 1800-2500 nm range which could be especially useful for quality evaluations of the oil content in hazelnuts (Armenta et al. 2007). Therefore, preliminary tests were conducted to increase the spectral sensitivity of the system by controlling the camera exposure time. Two different exposure times (5 and 8 ms) were experimentally selected to obtain a High Dynamic Range (HDR) and thus a high signal-to-noise ratio over the full wavelength range. Therefore, one hypercube per exposure time was acquired from the upper side of each hazelnut, ignoring the position of the damaged/infected areas, if any. The 5-ms exposure time was implemented in the spectral band with high system sensitivity (850-1870 nm), where the system was even susceptible to spectral saturation. The 8-ms exposure time was used in the spectral range with relatively lower sensitivity (1870-2500 nm). The spectral regions below 1003 nm and above 2484 nm were removed because the signal-to-noise ratio was too low. The resulting HDR hypercube was obtained through a pixel-by-pixel merging of both the 1003-1870-nm range acquired at 5-ms integration time and the 1870-2484-nm range acquired at 8-ms integration time by compensating the offset between both spectral ranges. Spectral merging was performed by dividing each value for a spectrum by the maximum value observed in that spectrum (i.e. maximum normalization) (Fig. 3).

Two binary masks were used to remove the background and the edges in each HIS raw image (Fig. 4a). The edges were removed because of pixels with low reflectance signals (Fig. 4b). Finally, the mean reflectance spectrum of each hazelnut was computed from an 80% Region Of Interest (ROI) (Fig. 4c) by shrinking the binary mask around the center of mass (centroid) of the kernel (Fig. 4d).

Spectra preprocessing

Data preprocessing (transformation and data reduction) can substantially improve the classification results for spectral data which contain highly correlated variables as well as noise and other irrelevant information (Wu et al. 1996; Tallada et al. 2011). Spectral preprocessing helps to remove the spectral baseline shift and scattering effects caused by particle size differences and to enhance the spectral differences between classes. However, spectral information that can be useful for the classification models can also be lost in the data pretreatment process. Therefore, the spectral preprocessing has to be optimized for each application.

During experimentation, spectra were preprocessed following a variety of spectral pretreatments including the Standard Normal Variate (SNV), Multiplicative Scatter Correction (MSC), Savitzky-Golay first, second and third derivatives (D1*f*, D2*f* and D3*f*, respectively) with a second or third order polynomial fitted over a window of five, seven or nine variables (Savitzky and Golay 1964; Boysworth and Booksh 2008) and Mean Centering (MC). Every possible combination of preprocessings was also tested and only the best results, in terms of model performance, were retained.

Kernel classification

Partial Least Squares Discriminant Analysis (PLS-DA) is a supervised classification technique which applies Partial Least Squares (PLS) regression to calculate the probability that a sample belongs to a certain class (Workman and Weyer 2008). PLS-DA was used to classify the mean kernel spectra. As PLS regression performs a dimensionality reduction, it is essential to test the model and select the correct number of latent variables to find the optimal trade-off between under-fitting and over-fitting. Root Mean Squared Error (RMSEC, RMSECV and RMSEP) calculations were employed to evaluate each discriminant model with the purpose to select the optimal number of latent variables and thus circumvent under-fitting and over-fitting problems and avoid unrealistic results (Goodarzi et al. 2013).

PLS-DA models were computed for the binary (B_1 and B_2), ternary (T_1 and T_2) and quaternary (Q) classifications. The B_1 model regarded the classification of samples belonging to the 'Extra Class' and a cumulative batch of 'Class I' and 'Class II'. The B_2 model classified the samples between the 'Extra Class' and a cumulative batch of 'Class I', 'Class II' and 'Waste'. The T_1 model was computed to distinguish between the 'Extra Class', 'Class I' and 'Class II' samples, while the T_2 model classified the samples between the 'Extra Class', 'Class I' and a cumulative batch of 'Class II' and 'Waste'. Finally, the Q model tested the performance of PLS-DA to discriminate the samples between 'Extra Class', 'Class I', 'Class II' and 'Waste'.

The samples (2400 fruits) were randomly split as follows: 40% of the samples were assigned to the calibration set (C) (i.e. 960 fruits) and 60% were split into three prediction sets (P) (i.e. 480 fruits in each test set). Moreover, every PLS-DA model was optimized by computing a random sub-set cross validation (CV) with 30 data splits and three iterations. This CV method was preferred because it is very versatile and capable of providing a better representation when the dataset contains many samples, as in our experiments (Naes et al. 2004; Wise 2009).

The classification performance of each PLS-DA model was determined in terms of the sensitivity, selectivity, α-error, β-error and accuracy rates (Dejaegher et al. 2011). Precisely, accuracy was used to mainly select models in terms of predictivity, while sensitivity and specificity were jointly employed to perform a model sub-selection in terms of robustness (i.e. the capability of the model to resist to small changes in test conditions).

The sensitivity rate is the number of correctly classified hazelnuts in the considered class divided by the total number of hazelnuts in that class (Eq. 2). The selectivity rate is the number of correctly classified hazelnuts in the other classes divided by the total number of hazelnuts in the other classes (Eq. 3). The α-error rate is related to the presence of false positive results and is equal to '1 – Selectivity Rate' (Eq. 4), while the β-error is related to the presence of false negatives results and corresponds to '1 – Sensitivity Rate' (Eq. 5). The accuracy rate is the proportion of the true results in the population (Eq. 6). Therefore, the statistical measures of the performance of each classification test were calculated as follows:

(2) $$Sensitivity\,rate = \frac{True\,Positives}{True\,Positives + False\,Negatives}$$

(3) $$Selectivity\,rate = \frac{True\,Negatives}{False\,Positives + True\,Negatives}$$

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

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

(6) $$Accuracy\,rate = \frac{True\,Positives + True\,Negatives}{Total\,Positives + Total\,Negatives}$$

Measurements were performed for calibration, cross-validation and prediction sets.

Size and shape measurement

Hazelnut perimeter (mm), major and minor axis lengths (mm), surface area (mm^2) and kernel eccentricity were extracted from each unprocessed HIS image using Matlab software R2012a in combination with ‘Image Processing’ toolbox. Measurements were performed using the ‘regionprops’ toolbox function. Perimeter was acquired by computing the distance between each adjacent pair of pixels around the edge of the kernel. Axis lengths were measured as major and minor axis of the ellipse with the same normalized second central moments (rotational inertia) as the kernel region. Area was acquired by calculating the number of pixels in the kernel-image surface. Eccentricity was computed as the ratio of the distance between the foci of the ellipse and its major axis length. Size and shape parameters were converted from pixel-to-metric unit (mm) by scanning a reference with known dimensions.

Data handling and statistical analyses

One-way analysis of variance (ANOVA) was performed to evaluate statistical differences between the quality classes. The Tukey’s pairwise comparison method was performed, and the

Honestly Significant Difference (HSD) was calculated for an appropriate level of interaction ($P \leq 0.01$) (Montgomery 2001). Results were reported as the mean and standard error of the mean.

All data handling and statistical analyses were performed using Matlab software R2012a coupled with PLS_Toolbox software v7.3.1 (Eigenvector Research Inc., WA, USA) for the spectral preprocessing and PLS-DA model building.

RESULTS AND DISCUSSION

Data overview

Fig. 5a depicts the resulting mean relative reflectance spectra obtained from the SWIR hyperspectral imaging system by combining scans at 5- and 8-ms exposure times. Fig. 5b, 5c and 5d were also plotted to zoom into the 95% confidence intervals of each calibration sample (n = 240), allowing to observe the spectral variability of each hazelnut class. From a first visual inspection, it seems that the hazelnut quality mainly affected the spectral characteristics in the 1934-2484-nm wavelength region. Changes in reflectance in that wavelength range are commonly associated with lipids, carbohydrates and proteins (Workman and Weyer 2008), which represent the main constituents of the hazelnut (USDA 2014). However, as less apparent changes in the other spectral regions might be also relevant for hazelnut quality prediction the entire spectra were included in the statistical analyses.

Discriminant analysis

In Table 1 the characteristics and performance metrics are summarized for the combinations of preprocessing and PLS-DA model complexity which gave the best classification performance for the binary, ternary and quaternary classifications.

Removing and/or reducing non-relevant variability from the raw spectral data seems to be crucial for the development of accurate classification models as all best models use some kind of preprocessing. As shown in Fig. 6, the preprocessing techniques significantly reduced the RMSEs for calibration, cross-validation and prediction. From the table it can be inferred that SNV and MSC scatter correction were not necessary to obtain the lowest classification errors. This may either mean that the variation in light scattering is rather limited or that it also provides information on the different quality classes. For binary (B_1 and B_2) and ternary (T_1 and T_2) models did the first derivative processing lead to better classification performance compared to other preprocessing techniques without derivation. However, for the quaternary (Q) model, derivative processing did not improve the classification results. Savitzky-Golay smoothing filter improved the classification performance for all classification challenges. This suggests that the spectral data was slightly affected by issues due to noise and other irrelevant information.

Considering the large variability in hazelnut quality due to growth and harvest conditions, the proposed methodology suggests a very high classification performance. In fact, all of the models possessed very-good (> 90%) or good (>80%) selectivity and sensitivity for C_X, C_I and C_{II} classes. However, each model should be further tested on external prediction sets (i.e. on fruits from different orchards, cultivars, harvesting years, etc). Models had a higher ability to discriminate between the aforementioned classes with respect to other classes. Consequently, the related α-error and β-error rates also indicate a lower chance of incorrectly classifying one class as belonging to another class than to incorrectly classify other classes as belonging to that class. All models had the highest selectivity and sensitivity for the excellent class C_X. This indicates that the classification of hazelnuts to belong or not belong to the excellent quality class was the most accurate. In general, we observed that the accuracy rate of a PLS-DA model for a specific class was positively related with the quality score of that class (*i.e.* the lower the quality score of a hazelnut class, the lower the model accuracy for that class). Moreover, it should be noted that the best predictive ability was obtained by excluding the W_S class from the classification models. In fact, both selectivity and sensitivity significantly declined when the W_S class was included in a model. This can probably be explained by the fact that the variation in the properties of the hazelnuts classified as waste was much higher than the variation in the properties of the hazelnuts classified as good or excellent. So, in practice it would be most efficient to classify hazelnuts as being good or excellent quality.

Binary models (B_1 and B_2) were built to classify the 'Class Extra' from the other quality classes and led to impressive classification results. In fact, the B_1 and B_2 models had accuracy rates of 0.98 and 0.93, respectively, by using four and five latent variables, respectively. In the case of discriminative models for multi-class kernel detection (high Y-complexity level), the best predictive ability was observed in the T_1 model for 'Class Extra' (Fig. 7a), the T_2 model for 'Class I' (Fig. 7b) and the Q model for 'Class II' (Fig. 7c) by using four, five and seven latent variables, respectively.

However, because of the better capability to discriminate the highest quality classes, it is recommended to only use the T_2 and Q models in a sorting device for industrial implementation. The T_2 model had the highest accuracy when identifying 'Class Extra' (0.95) and 'Class I' (0.94), even though it was incapable of sorting 'Class II' kernels from waste. Conversely, the Q model allowed for waste sorting, but showed a worse accuracy for both 'Class Extra' (0.93) and 'Class I' (0.90), although both selectivity (0.84) and sensitivity (0.81) for 'Class II' were higher than those of the T_1 model. In fact, the T_1 model was characterized by a selectivity and sensitivity for 'Class II' equal to 0.82 and 0.78, respectively.

Regression coefficient for the T_2 model (Fig. 8) shows the most informative wavelengths in the entire spectral region. Specifically, features around 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). Hydroxyl groups exhibit higher stretching overtones within these bands. Moreover, features computed in that region could be correlated with lipids, water (~1460 nm), phenols (~1400-1440 nm), hydroperoxides (~1460 nm) and starch (~1540 nm) (Workman and Weyer 2008). The 2000-2220-nm bands matched with the combination of C-C/C-H, C-H/C=O, and N-H/C=O groups (Workman and Weyer 2008; Vitale et al. 2013), which may be normally associated to lipids and proteins.

Finally, yet importantly, it is fundamental to consider that hazelnut fruits are commonly sorted in the industry on the basis of their size and shape. In fact, describing hazelnut shape and size is necessary for a range of different industrial purposes, including standardization of confectionery products, compliance with consumer preference and cultivar selection. Hyperspectral imaging not only allows the measurement of chemical constituents or internal quality attributes of the fruit, but also provides spatial distribution data (*i.e.* shape information) of the products. Thus, in addition to the development of classification models based on the spectral profiles of hazelnuts, we also explored the possibilities offered by hyperspectral imaging for quantifying external or surface characteristics of each kernel, measuring its perimeter, major and minor axis lengths, surface area and eccentricity. The results of these computer-vision-based shape and size estimations are summarized in Table 2. It can clearly be seen that there were significant differences in the size and shape parameters of the different hazelnut quality classes. More specifically, waste samples had the largest dimensions and lowest eccentricity. Consequently, although the analysis was restricted to two dimensions, it is clear that the spatial domain could have an added value for improving the classification of the hazelnuts into the different quality classes, especially for the lowest quality classes (*i.e.* 'Class II' and 'Waste'). Therefore, it is recommended to effectively exploit both the spectral and spatial information to improve the classification accuracy through data fusion as suggested by Ottavian et al. (2014). However, further research is required to confirm this hypothesis. Moreover, it is important to consider that the only spatial domain cannot be used for quality classification purposes. In fact, the aforementioned Commission Regulation (EC) No. 1284/2002 defines the hazelnut-classes characteristics on the basis of both size and incidence of defects.

CONCLUSIONS

In this study, the potential of high dynamic range (HDR) hyperspectral imaging in the SWIR range (850-2500 nm) has been evaluated for sorting of hazelnut kernels (*Corylus avellana* L. cv. Tonda Gentile Romana) into the main quality classes as stated by the UE Regulation n. 1284/2002. In general, the PLS-DA models obtained a high accuracy in classifying the hazelnut kernels into the quality classes based on the extracted mean kernel spectra. Specifically, HDR-SWIR hyperspectral imaging showed a very-good classification (accuracy >90%) for hazelnut fruits belonging to high quality classes (*i.e.* 'Class Extra' and 'Class I'). On the contrary, misclassified kernels were primarily assigned to low-quality classes (*i.e.* 'Class II' and Waste). Thus, this method lays the foundations for an accurate hyperspectral/multispectral detection system in terms of sensitivity, selectivity, nondestructive analysis and automation. However, prior to implementation in industry this approach should be validated further on a larger sample size covering the most important variations expected from growing hazelnuts in different regions, orchards, crop years and agro-pedo-climatic conditions, in addition to other cultivars.

Additionally, one-way ANOVA on kernel size and shape parameters showed strongly significant differences between quality classes ($P \leq 0.01$). Therefore, spatial data could potentially be used to improve the classification accuracy compared to only using the mean spectral profiles of the kernels. However, this possibility has to be investigated further in the future.

ACKNOWLEDGEMENTS

This research has been financially supported by Mipaaf through the project "Miglioramento della filiera corilicola laziale - Mi.F.CO.L.' represented by the 'AOP Nocciola Italia Soc. Cons. s.r.l' and 'CeFAS - Azienda speciale CCIAA Viterbo'. The authors gratefully acknowledge I.W.T.-Flanders for the financial support through the Chameleon project (IWT 100021).

REFERENCES

Alasalvar, C., & Shahidi, F. (2008). *Tree Nuts: Composition, Phytochemicals, and Health Effects*. (Alasalvar, C. & Shahidi, F., Eds.) (1st ed.). Boca Raton, FL: CRC Press.

Amaike, S., & Keller, N. P. (2011). Aspergillus flavus. *Annual review of phytopathology*, *49*, 107–133.

Armenta, S., Garrigues, S., & de la Guardia, M. (2007). Determination of edible oil parameters by near infrared spectrometry. *Analytica chimica acta*, *596*(2), 330–7. doi:10.1016/j.aca.2007.06.028

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* (3rd ed., pp. 207–229). New York, USA: CRC Press.

Capinera, J. L. (2001). Oder Hemiptera – bugs. In *Handbook of vegetable pests* (pp. 243–278). New York, NY, USA: Academic Press.

Cecchini, M., Contini, M., Massantini, R., Monarca, D., & Moscetti, R. (2011). Effects of controlled atmospheres and low temperature on storability of chestnuts manually and mechanically harvested. *Postharvest Biology and Technology*, *61*, 131–136.

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. *Analytical sciences : the international journal of the Japan Society for Analytical Chemistry*, *20*(6), 935–40.

Chun, J., Lee, J., & Eitenmiller, R. R. (2006). Vitamin E and Oxidative Stability During Storage of Raw and Dry Roasted Peanuts Packaged under Air and Vacuum. *Journal of Food Science*, *70*(4), 292–297.

Contini, M., Frangipane, M. T., & Massantini, R. (2011). Antioxidants in Hazelnuts (Corylus avellana L.). In Preedy, V. R., Watson, R. R., & Patel, V. B. (Eds.), *Nuts & Seeds in Health and Disease Prevention* (pp. 611–625). London, UK: Academic Press.

Dejaegher, B., Dhooghe, L., Goodarzi, M., Apers, S., Pieters, L., & Vander Heyden, Y. (2011). Classification models for neocryptolepine derivatives as inhibitors of the β-haematin formation. *Analytica chimica acta*, *705*(1-2), 98–110.

ElMasry, G., Kamruzzaman, M., Sun, D.-W., & Allen, P. (2012). Principles and applications of hyperspectral imaging in quality evaluation of agro-food products: a review. *Critical reviews in food science and nutrition*, *52*(11), 999–1023.

ElMasry, G., & Sun, D.-W. (2010). Principles of Hyperspectral Imaging Technology. In Sun, D.-W. (Ed.), *Hyperspectral imaging for food quality analysis and control* (pp. 3–43). London, UK: Academic Press.

European Commission. (2002). Commission Regulation (EC) No. 1284/2002 of 15 July 2002 laying down the marketing standard for hazelnuts in shell. *Official Journal of the European Union (L)*, *187*, 14–20.

Ghirardello, D., Contessa, C., Valentini, N., Zeppa, G., Rolle, L., Gerbi, V., & Botta, R. (2013). Effect of storage conditions on chemical and physical characteristics of hazelnut (Corylus avellana L.). *Postharvest Biology and Technology*, *81*, 37–43.

Goodarzi, M., Heyden, Y. Vander, & Funar-Timofei, S. (2013). Towards better understanding of feature-selection or reduction techniques for Quantitative Structure–Activity Relationship models. *TrAC Trends in Analytical Chemistry*, *42*, 49–63.

Kibar, H., & Öztürk, T. (2009). The effect of moisture content on the physico-mechanical properties of some hazelnut varieties. *Journal of Stored Products Research*, *45*(1), 14–18.

Köksal, A. İ., Artik, N., Şimşek, A., & Güneş, N. (2006). Nutrient composition of hazelnut (Corylus avellana L.) varieties cultivated in Turkey. *Food Chemistry*, *99*(3), 509–515.

Massantini, R., Moscetti, R., Monarca, D., Cecchini, M., Contini, M., & Mordacchini, L. (2009). The influence of cover crops and double harvest on storage of fresh hazelnuts (Corylus avellana L .). *Advances in Horticultural Science*, *23*(4), 231–237.

Miles, P. W. (1972). The Saliva of Hemiptera. *Advances in Insect Physiology*, *9*, 183–255.

Montgomery, D. C. (2001). *Design and analysis of experiments* (Fifth edit., p. 699). New York, USA: Wiley & Sons, Inc.

Moscetti, R., Frangipane, M. T., Monarca, D., Cecchini, M., & Massantini, R. (2012). Maintaining the quality of unripe, fresh hazelnuts through storage under modified atmospheres. *Postharvest Biology and Technology*, *65*, 33–38.

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. *Journal of Food Engineering*, *118*, 1–7.

Moscetti, R., Haff, R. P., Saranwong, S., Monarca, D., Cecchini, M., & Massantini, R. (2014). Nondestructive detection of insect infested chestnuts based on NIR spectroscopy. *Postharvest Biology and Technology*, *87*, 88–94.

Moscetti, R., Monarca, D., Cecchini, M., Haff, R. P., Contini, M., & Massantini, R. (2014). Detection of Mold-Damaged Chestnuts by Near-Infrared Spectroscopy. *Postharvest Biology and Technology*, *93*, 83–90.

Naes, T., Isaksson, T., Fearn, T., & Davies, T. (2004). *A user-friendly guide to Multivariate Calibration and Classification* (p. 344). NIR publications.

Nguyen Do Trong, N., Tsuta, M., Nicolaï, B. M., De Baerdemaeker, J., & Saeys, W. (2011). Prediction of optimal cooking time for boiled potatoes by hyperspectral imaging. *Journal of Food Engineering*, *105*(4), 617–624.

Nicolaï, 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.

Ottavian, M., Fasolato, L., Serva, L., Facco, P., & Barolo, M. (2014). Data Fusion for Food Authentication: Fresh/Frozen–Thawed Discrimination in West African Goatfish (Pseudupeneus prayensis) Fillets. *Food and Bioprocess Technology*, *7*(4), 1025–1036.

Özdemir, M., Seyhan, F. G., Bakan, A. K., I□lter, S., Özay, G., & Devres, O. (2001). Analysis of internal browning of roasted hazelnuts. *Food Chemistry*, *73*(2), 191–196.

Parcerisa, J., Richardson, D. G., Rafecas, M., Codony, R., & Boatella, J. (1997). Fatty Acid Distribution in Polar and Nonpolar Lipid Classes of Hazelnut Oil (Corylus avellana L.). *Journal of Agricultural and Food Chemistry*, *45*(10), 3887–3890.

Pearson, T. C. (1999). Spectral Properties and Effect of Drying Temperature on Almonds with Concealed Damage. *LWT - Food Science and Technology*, *32*(2), 67–72.

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, *36*(8), 1627–1639.

Shahidi, F., & Miraliakbari, H. (2005). Tree Nut Oils. In Fereidoon Shahidi (Ed.), *Bailey's Industrial Oil and Fat Products* (6th ed., pp. 175–193). Hoboken, New Jersey, United States: John Wiley & Sons Inc.

Sipahioglu, H. N., & Heperkan, D. (2000). Lipolytic activity of Trichothecium roseum on hazelnut. *Food Microbiology*, *17*(4), 401–405.

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. *Transactions of the ASABE*, *54*(3), 1151–1158.

USDA. (2014). National Nutrient Database for Standard Reference – Release 26. URL http://ndb.nal.usda.gov/ndb/search/list. Accessed 03.12.14.

Vaccinio, P., Guidone, L., Corbellini, M., & Tavella, L. (2008). Detection of damage due to bug feeding on hazelnut and wheat by biochemical techniques. *Bullettin of Insectrology*, *61*(1), 189–190.

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. *Chemometrics and Intelligent Laboratory Systems*, *121*, 90–99.

Wang, W., Li, C., Tollner, E. W., Rains, G. C., & Gitaitis, R. D. (2012). A liquid crystal tunable filter based shortwave infrared spectral imaging system: Calibration and characterization. *Computers and Electronics in Agriculture*, *80*, 135–144.

Wise, B. M. (2009). Eigen vector research wiki - Using Cross-Validation. URL http://wiki.eigenvector.com. Accessed 02.21.2014.

Workman, J., & Weyer, L. (2008). *Practical Guide to Interpretive Near-Infrared Spectroscopy* (p. 344). London, UK: CRC Press.

Wu, W., Mallet, Y., & Walczak, B. (1996). Comparison of regularized discriminant analysis linear discriminant analysis and quadratic discriminant analysis applied to NIR data. *Analytica chimica acta*, *2670*(96), 257–265.

Yildiz, G., Wehling, R. L., & Cuppett, S. L. (2001). Method for determining oxidation of vegetable oils by near-infrared spectroscopy. *Journal of the American Oil Chemists' Society*, *78*(5), 495–502.

Figure 1
Click here to download Figure: Figure 1.eps

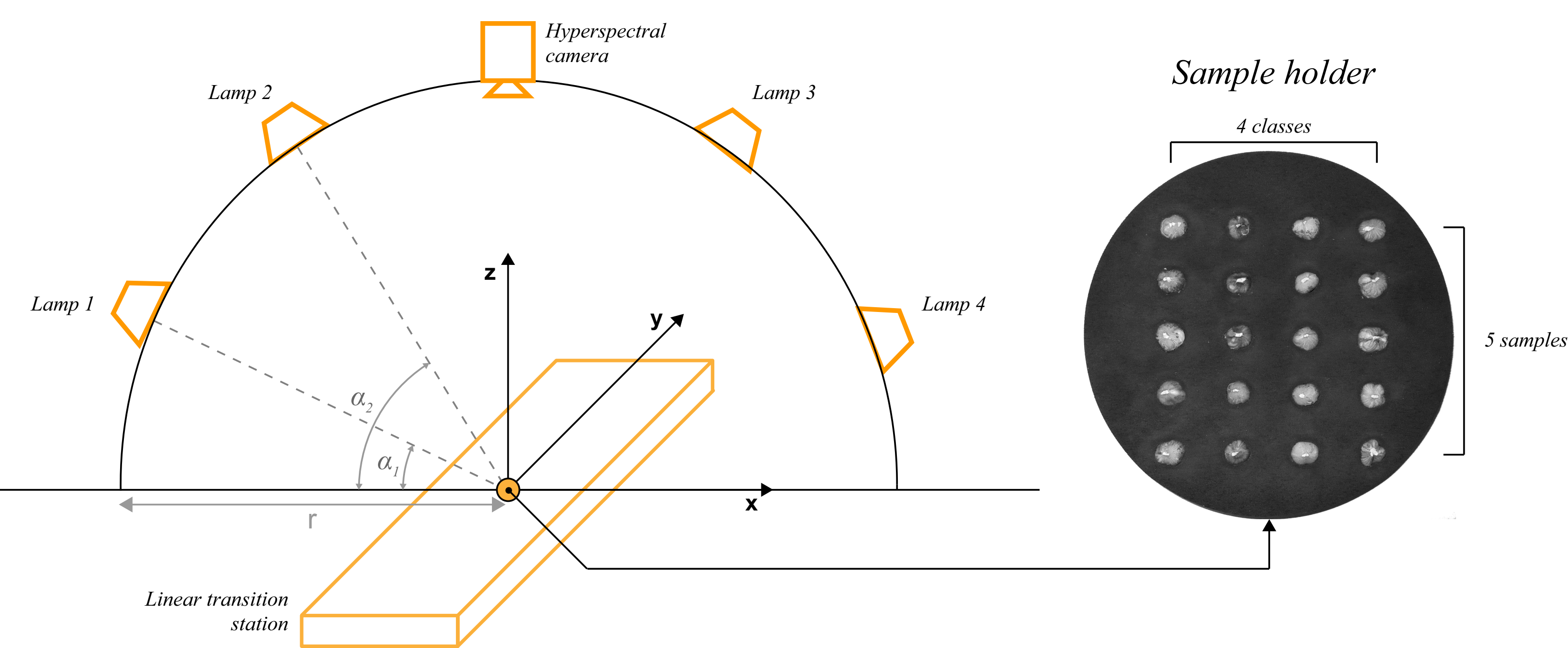

Figure 3

Click here to download Figure: Figure 3.eps

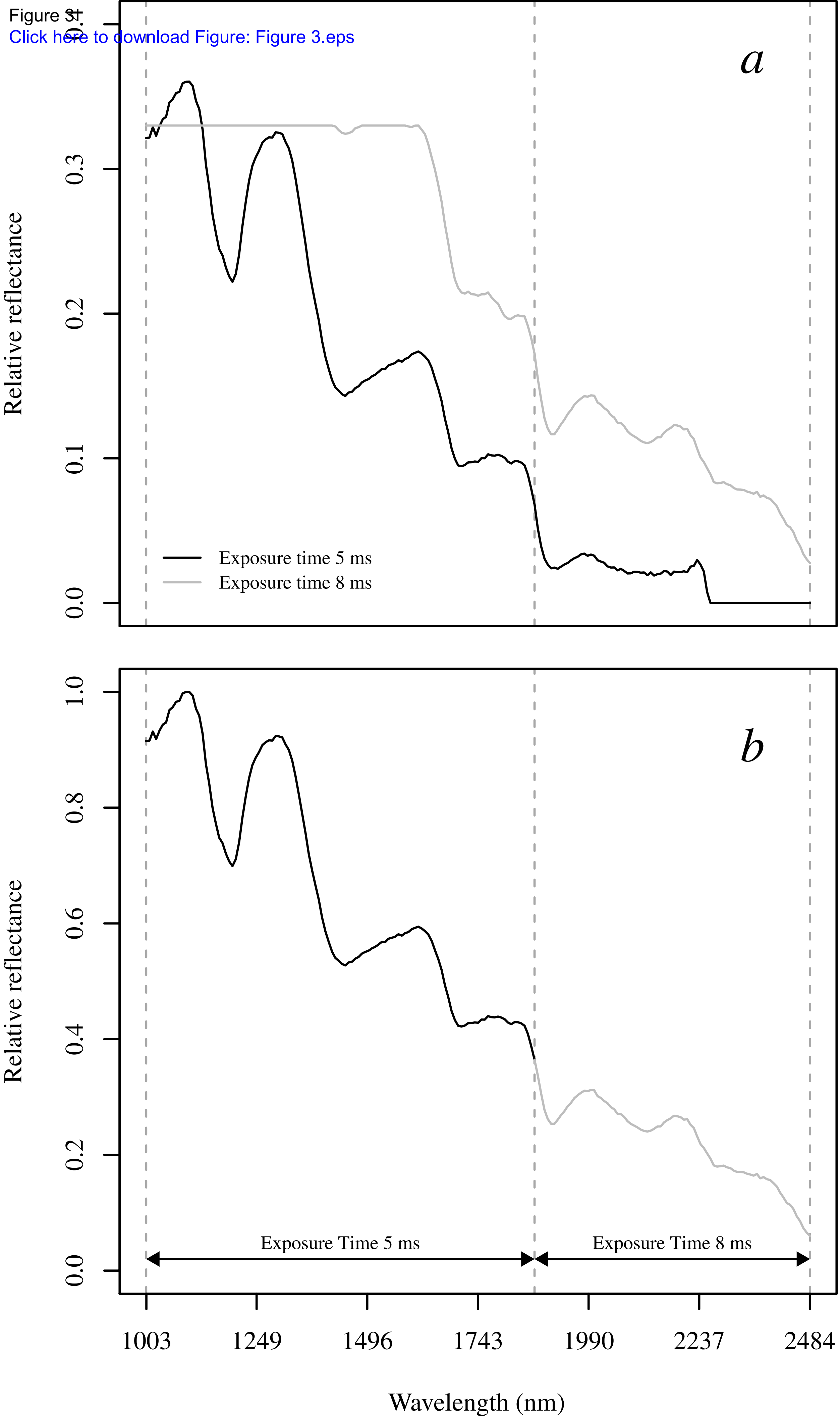

Figure 4
Click here to download Figure: Figure 4.eps

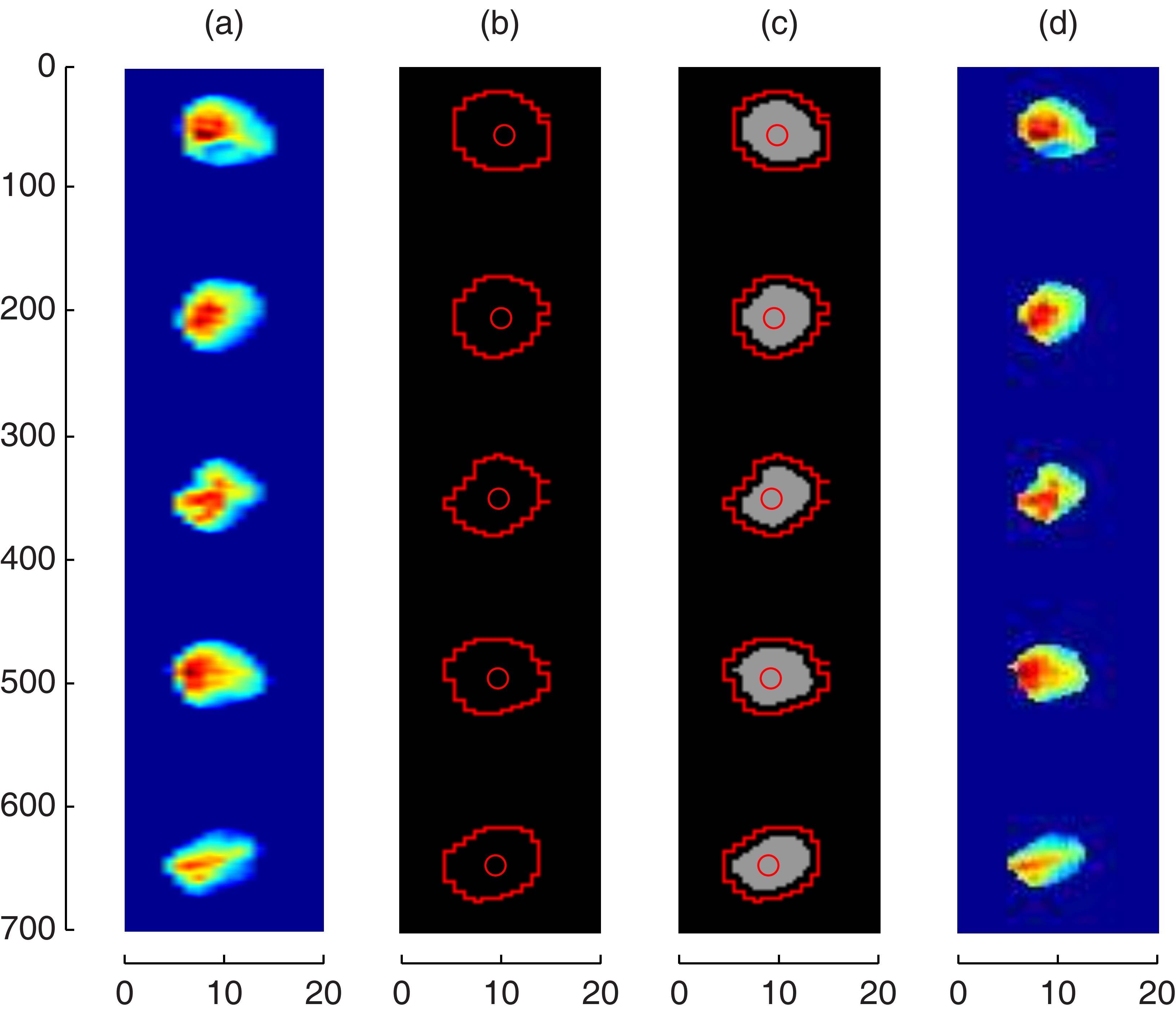

Figure 5
Click here to download Figure: Figure 5.eps

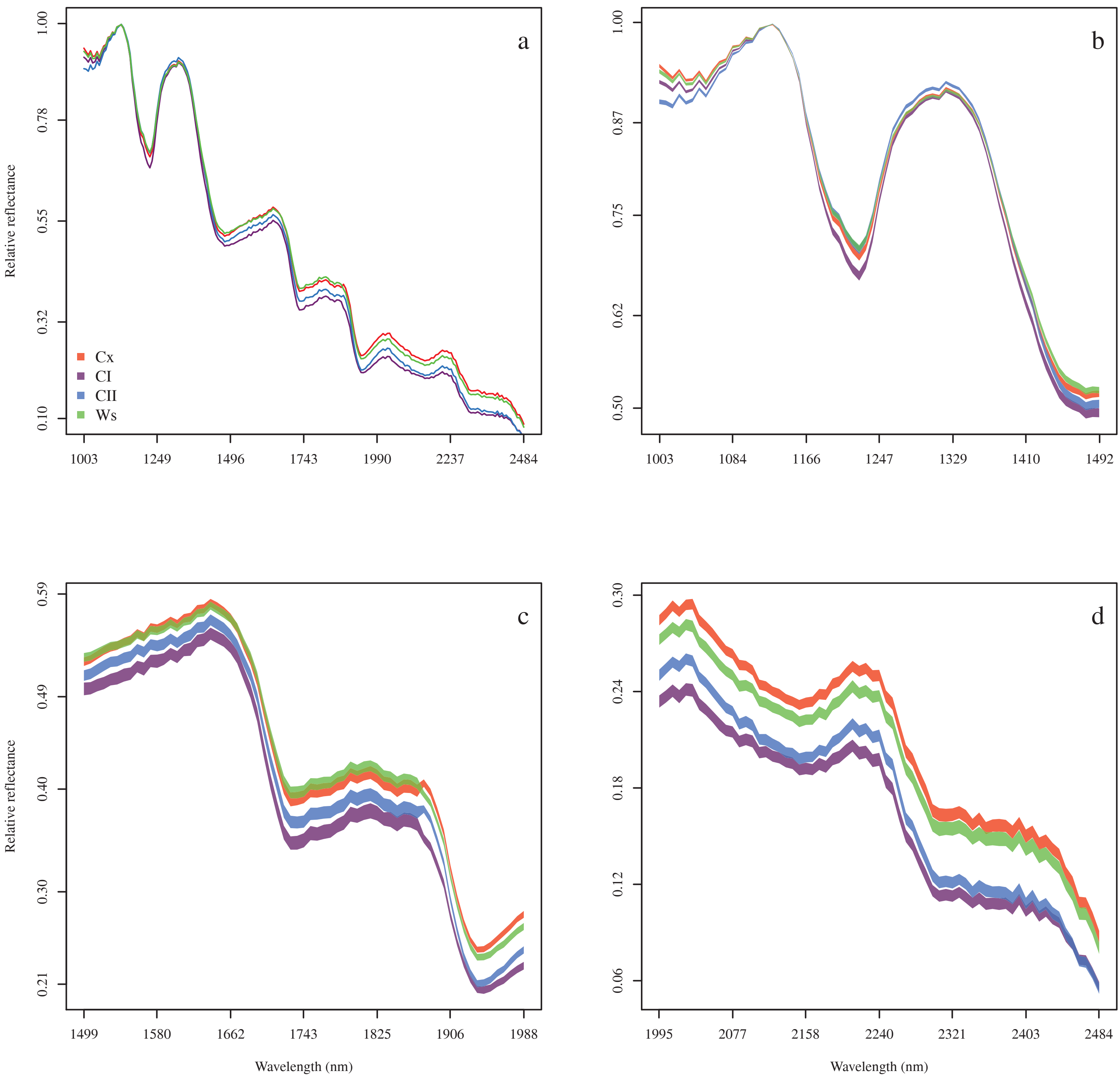

Figure 6

Click here to download Figure: Figure 6.eps

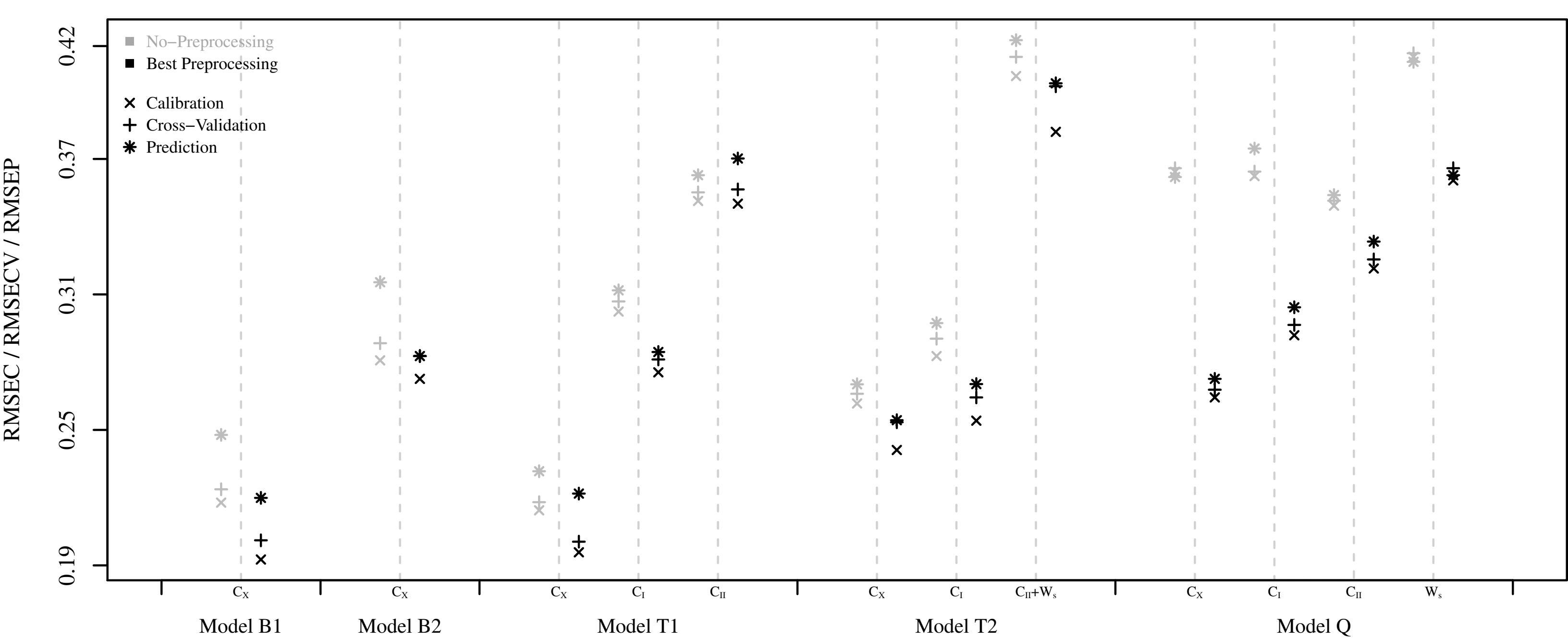

Figure 7

Click here to download Figure: Figure 7.eps

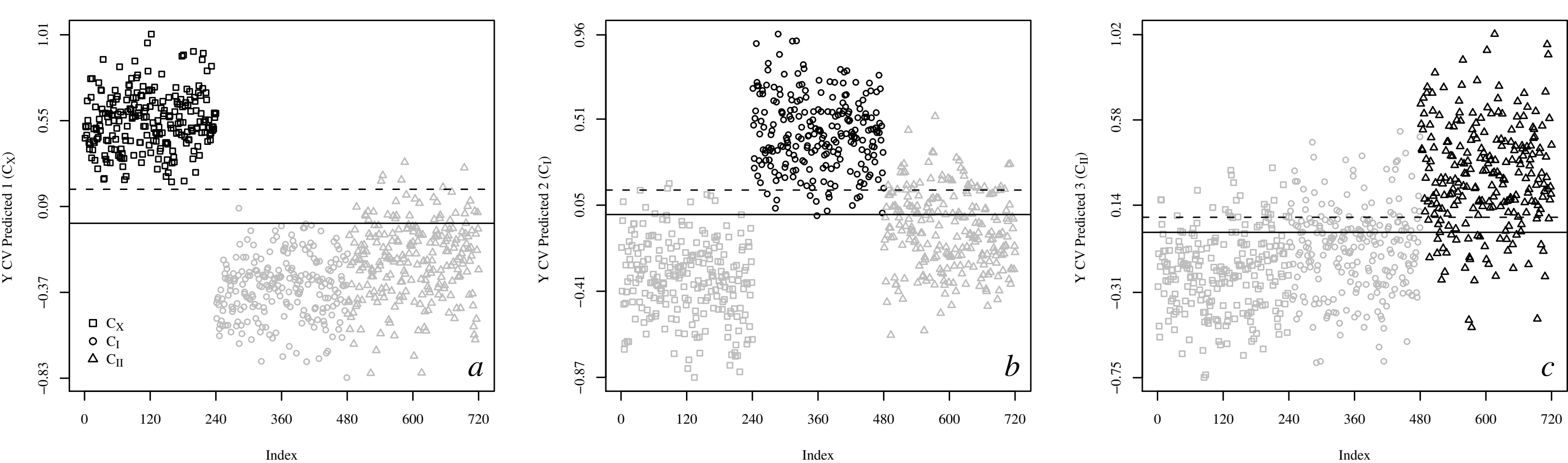

Figure 2

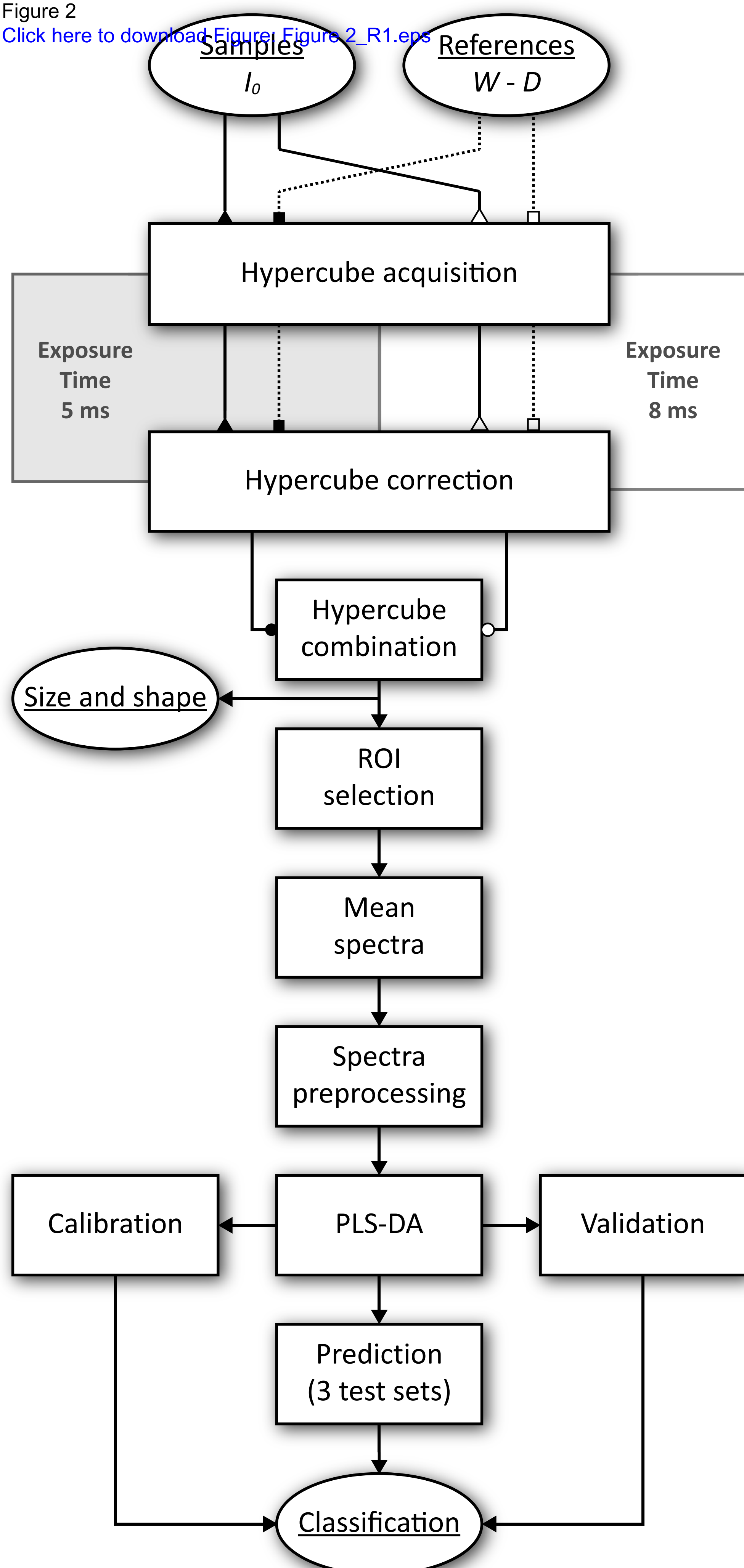
Samples
I_0
References
$W - D$
Hypercube acquisition
Exposure Time 5 ms
Exposure Time 8 ms
Hypercube correction
Hypercube combination
Size and shape
ROI selection
Mean spectra
Spectra preprocessing
Calibration
PLS-DA
Validation
Prediction (3 test sets)
Classification

FIGURE CAPTIONS

Figure 1. *Scheme of the hyperspectral imaging system and a picture of the 20-hole sample holder used during the scans.*

Figure 2. *Flowchart of hazelnut class detection using hyperspectral imaging at different exposure time.*

Figure 3. *Example of reflectance spectra (a) acquired by applying both the 5-ms and 8-ms exposure times and the resulting spectrum (b) after having combined both spectra. The merged spectrum was obtained by fixing the gap between both 1003-1870-nm and 1870-2484-nm spectral ranges and then normalizing the resulting spectrum itself.*

Figure 4. *Region of interest (ROI) selection steps. (a) represents the raw image; (b) shows the mask representing the hazelnut centroid and edge; (c) shows the resulting ROI mask (area in light-gray color); (d) image after applying the ROI mask.*

Figure 5. *Mean relative reflectance spectra (a) and 95% confidence intervals for different wavelength regions: 1003-1492-nm (b), 1499-1988-nm (c) and 1995-2484-nm (d).*

Figure 6. *Comparison of the RMSE results for each PLS-DA model and for calibration (RMSEC), cross-validation (RMSECV) and prediction (RMSEP) sets of both preprocessed and non-preprocessed spectra.*

Figure 7. *Classes' spatial distribution plots obtained from T_1 (a), T_2 (b) and Q (c) models for 'Class Extra', 'Class I' and 'Class II', respectively.*

Figure 8. *Regression coefficient plot for the T_2 model for the entire spectral region.*

Figure 8
Click here to download Figure: Figure 8.eps

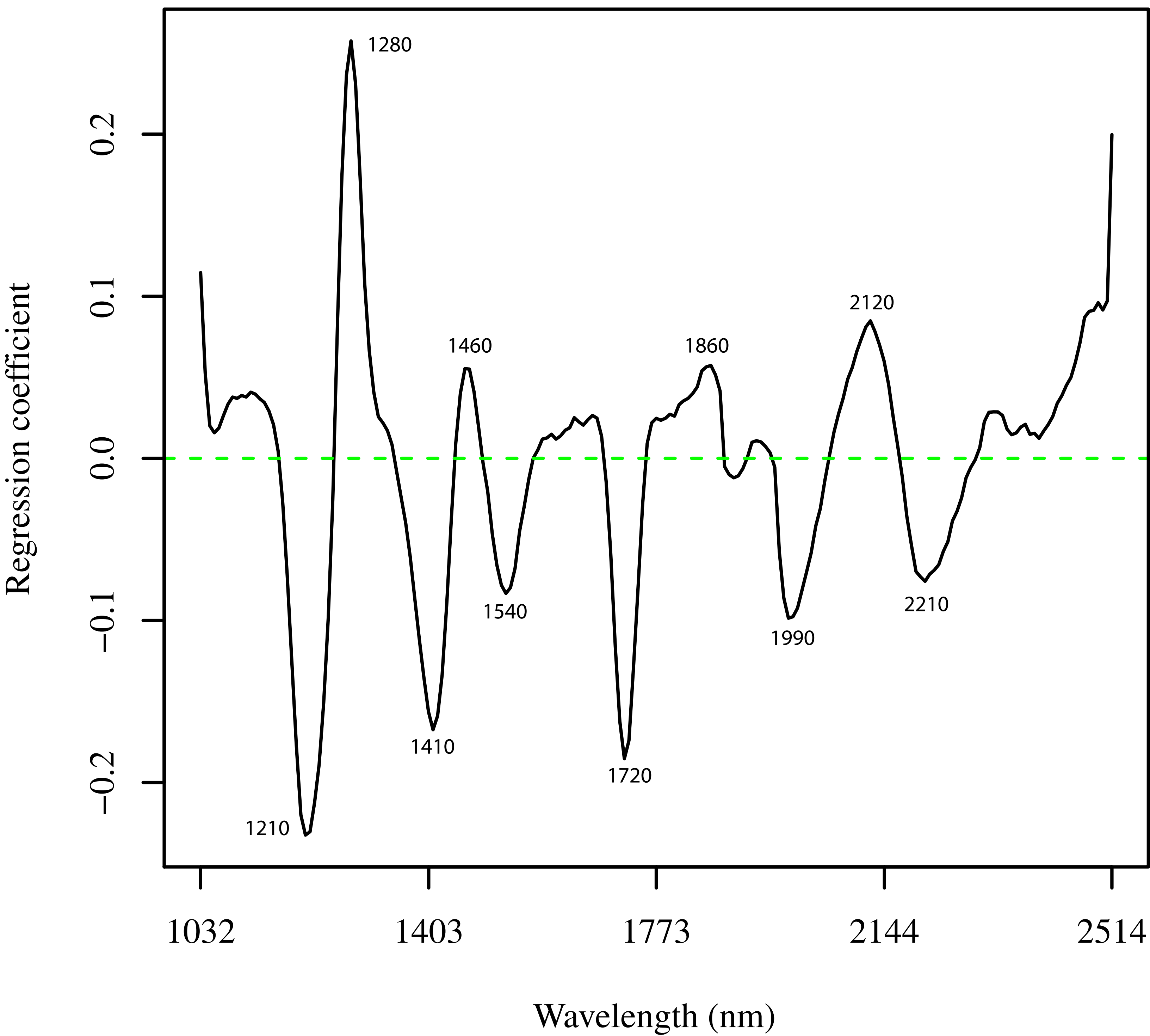

Table 2

Class ID	Perimeter (mm)	Axis Length (mm) Minor	Axis Length (mm) Major	Area (mm^2)	Eccentricity (10^2)
C_x	54.86 ± 0.17 d	11.95 ± 0.03 d	14.38 ± 0.03 d	133.06 ± 0.53 d	53.95 ± 0.40 b
C_I	58.03 ± 0.18 b	12.85 ± 0.04 b	15.33 ± 0.04 b	153.04 ± 0.71 b	52.70 ± 0.42 bc
C_{II}	56.99 ± 0.19 c	12.39 ± 0.03 c	15.11 ± 0.04 c	145.56 ± 0.69 c	55.62 ± 0.38 a
W_s	61.65 ± 0.18 a	13.53 ± 0.04 a	16.06 ± 0.03 a	168.22 ± 0.62 a	51.83 ± 0.44 c

TABLE CAPTIONS

Table 1. *Summary of the characteristics and performance metrics for the combinations of preprocessing and PLS-DA model complexity which gave the best results: B_1 – binary classification (C_X vs C_I + C_{II}); B_2 – binary classification (C_X vs C_I + C_{II} + W_s); T_1 – ternary classification (C_X vs C_I vs C_{II}); T_2 – ternary classification (C_X vs C_I vs C_{II} + W_s); Q – quaternary classification (C_X vs C_I vs C_{II} vs W_s).*

Table 2. *Results of computer-vision-based shape and size estimations for each hazelnut quality class. Results are reported as mean ± standard deviation. Different letters indicate significant differences ($P \leq 0.01$) between quality classes (Tukey's pairwise comparisons).*

Table 1

Model			Preprocess				LVs[e]	Selectivity			Sensitivity			α-error			β-error			Accuracy		
ID	Type	Class	SC[a]	Deriv.[b]	SG[c]	MC[d]		C[f]	CV[g]	P[h]	C	CV	P	C	CV	P	C	CV	P	C	CV	P
B1	Binary	C_x	-	D1*f*	5	Yes	4	0.9958	0.9931	0.9722	1.0000	0.9958	0.9806	0.0042	0.0069	0.0278	0.0000	0.0042	0.0167	0.9979	0.9945	0.9764
		$C_I + C_{II}$						-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
B2	Binary	C_x	-	D1*f*	9	Yes	5	0.9458	0.9319	0.9083	0.9375	0.9319	0.9583	0.0542	0.0681	0.0917	0.0625	0.0681	0.0417	0.9417	0.9319	0.9333
		$C_I + C_{II}$ + Ws						-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T1	Ternary	C_x	-	D1*f*	5	Yes	5	0.9917	0.9903	0.9764	1.0000	0.9972	0.9917	0.0083	0.0097	0.0236	0.0000	0.0028	0.0083	0.9944	0.9926	0.9815
		C_I						0.9188	0.9174	0.9278	0.9542	0.9361	0.9389	0.0813	0.0826	0.0722	0.0458	0.0639	0.0611	0.9306	0.9236	0.9315
		C_{II}						0.8771	0.8674	0.8167	0.8167	0.8125	0.7778	0.1229	0.1326	0.1833	0.1833	0.1875	0.2222	0.8569	0.8491	0.8037
T2	Ternary	C_x	-	D1*f*	7	Yes	9	0.9604	0.9500	0.9486	0.9750	0.9639	0.9639	0.0396	0.0500	0.0514	0.0250	0.0361	0.0361	0.9653	0.9546	0.9537
		C_I						0.9521	0.9444	0.9500	0.9542	0.9403	0.9306	0.0479	0.0556	0.0500	0.0458	0.0597	0.0583	0.9528	0.9430	0.9435
		C_{II} + Ws						0.7833	0.7590	0.7542	0.7625	0.7319	0.7250	0.2167	0.2410	0.2458	0.2375	0.2681	0.2750	0.7764	0.7500	0.7445
Q	Quaternary	C_x	-	-	5	Yes	7	0.9431	0.9403	0.9269	0.9375	0.9333	0.9444	0.0569	0.0597	0.0731	0.0625	0.0667	0.0556	0.9417	0.9385	0.9312
		C_I						0.9069	0.8991	0.8954	0.9250	0.9194	0.8972	0.0931	0.1009	0.1046	0.0750	0.0806	0.1028	0.9115	0.9042	0.8958
		C_{II}						0.8597	0.8579	0.8435	0.8250	0.8181	0.8111	0.1403	0.1421	0.1565	0.1750	0.1819	0.1889	0.8510	0.8479	0.8354
		Ws						0.7653	0.7607	0.7602	0.8417	0.8333	0.8222	0.2347	0.2394	0.2398	0.1583	0.1667	0.1778	0.7844	0.7788	0.7757

[a] Scatter correction

[b] Derivative

[c] Savitzky-Golay filter smoothing points

[d] Mean Centering

[e] Latent variables

[f] Calibration

[g] Cross validation

[p] Prediction

ABSTRACT

A rapid, robust, unbiased and inexpensive discriminant method capable of classifying hazelnut (*Corylus avellana*, L.) in compliance with the statements set out by the Commission Regulation (EC) No.1284/2002 is economically important to the fresh and processed industries. Thus, in this study, the feasibility of High Dynamic Range (HDR) hyperspectral imaging for hazelnut kernel sorting (cv. Tonda Gentile Romana) of four quality classes ('Class Extra', 'Class I', 'Class II' and 'Waste') has been investigated. Two different exposure times (5 and 8 ms) were selected for experiments, and the respective spectra were combined to obtain a HDR over the full spectral range. The illumination setup was optimized to improve the intensity and uniformity of the light along the field of view of the camera. PLS-DA was used to classify the kernels based on their spectra and the spectral pre-treatment was optimized through an iterative routine. The performance of each PLS-DA model was defined based on its accuracy, sensitivity, selectivity rates. All of the selected models provided a very-good (> 90%) or good (>80%) sensitivity and selectivity for the predefined classes. Misclassified kernels were primarily assigned to the low-quality classes (*i.e.* 'Class II' and Waste). Moreover, the spatial domain was used to evaluate the feasibility of distinguishing hazelnut classes on the basis of their size and shape. It was found that hazelnut dimensions can be used to improve the accuracy of the classification of the kernels. Thanks to this combination of both spectral and spatial information spectral imaging could be used for quality sorting of hazelnuts.