

**NEAR-INFRARED SPECTROSCOPY IS SUITABLE FOR THE  
CLASSIFICATION OF HAZELNUTS ACCORDING TO  
PROTECTED DESIGNATION ORIGIN**

|                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|-------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Journal:                      | <i>Journal of the Science of Food and Agriculture</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| Manuscript ID:                | JSFA-14-1991.R1                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| Wiley - Manuscript type:      | Research Article                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| Date Submitted by the Author: | 23-Oct-2014                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| Complete List of Authors:     | Moscetti, Roberto; Tuscia University, Department of science and technology for Agriculture, Forest, Nature and Energy<br>Radicetti, Emanuele; Tuscia University, Department of science and technology for Agriculture, Forest, Nature and Energy<br>Cecchini, Massimo; Tuscia University, Department of science and technology for Agriculture, Forest, Nature and Energy<br>Monarca, Danilo; Tuscia University, Department of science and technology for Agriculture, Forest, Nature and Energy<br>Massantini, Riccardo; Tuscia University, Department for Innovation in Biological, Agro-food and Forest system |
| Key Words:                    | Corylus avellana L., cultivars, discriminant analysis, chemometrics, nondestructive techniques                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |

SCHOLARONE™  
Manuscripts

**TITLE:**

NEAR-INFRARED SPECTROSCOPY IS SUITABLE FOR THE CLASSIFICATION  
OF HAZELNUTS ACCORDING TO PROTECTED DESIGNATION ORIGIN

**RUNNING TITLE:**

IDENTIFICATING ‘NOCCIOLA ROMANA’ PDO HAZELNUT USING NIR  
SPECTROSCOPY

**AUTHOR’S NAMES**

Moscetti, R<sup>a</sup>, Radicetti, E.<sup>a</sup>, Monarca, D<sup>a</sup>, Cecchini, M<sup>a</sup>, Massantini, R<sup>b\*</sup>

<sup>a</sup>Department of Science and Technology for Agriculture, Forest, Nature and Energy,  
Tuscia University, Via S. Camillo de Lellis snc, 01100, Viterbo, Italy

<sup>b</sup>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**

**Background:** this study investigates the possibility of using Near-Infrared spectroscopy  
for the authentication of the ‘Nocciola Romana’ hazelnut (*Corylus avellana* L. cvs.  
‘Tonda Gentile Romana’ and ‘Nocchione’) as a protected designation of origin (PDO)  
hazelnut from Central Italy. Algorithms for the selection of the optimal pretreatments  
were tested in combination with the following discriminant routines: k-Nearest

Neighbor, Soft Independent Modeling of Class Analogy, Partial Least Squares Discriminant Analysis and Support Vector Machine Discriminant Analysis.

**Results:** the best results were obtained using a Support Vector Machine Discriminant Analysis routine. Thus, classification performance rates with specificities, sensitivities, and accuracies as high as 96.0%, 95.0% and 95.5%, respectively, were achieved. Various pretreatments, such as Standard Normal Variate, mean centering and a Savitzky-Golay filter with seven smoothing points, were used. The optimal wavelengths for classification were mainly correlated with lipids, although some contribution from minor constituents, such as proteins and carbohydrates, was also observed.

**Conclusion:** from the results obtained, it seems that Near-Infrared spectroscopy could classify hazelnut according to the PDO 'Nocciola Romana' designation. Thus, the experimentation lays the foundations for a rapid, online, authentication system for hazelnut. However, model robustness should be improved taking into account agro-pedo-climatic growing conditions.

**Keywords:** *Corylus avellana* L., cultivars, discriminant analysis, chemometrics, nondestructive techniques

## INTRODUCTION

Hazelnut (*Corylus avellana* L.) is one of the most economically important raw materials for the confectionary, bakery and chocolate industries. Turkey, Italy, Spain and the USA are the most important world producers. Approximately 90% of the worldwide crop is absorbed by the food industry. Hazelnut promotes good health and contributes to the beneficial effects of the Mediterranean diet.<sup>1</sup> The kernel is also used to add bioactive compounds to new health-related foods, serving as a substitute for

synthetic ingredients.<sup>2,3</sup> Moreover, different uses of hazelnuts are being explored, including the consumption of fresh unripe fruit.<sup>4</sup>

According to qualitative standards of the confectionary industry, the main commodity-related characteristics of the fruits are expressed in relation to their process and use and can be summarized as follows: fruit morphology, moisture content, shape uniformity, percent kernel, flaw incidences and peelability due to the roasting process. Moreover, confectionary industries and markets require whole hazelnuts from specific cultivars to enhance the characteristics of the traditional product with the aim of releasing a high-quality product.

The hazelnut ‘Nocciola Romana’ is a ‘Protected Designation of Origin’ (PDO) produced and harvested in a well-defined geographical area located in the North-Central Latium region in Italy. The area lies within the territory of several municipalities of the provinces of Viterbo and Rome. The PDO designation guarantees that the quality of the product is closely linked to the main cultivars of its territorial origin. In fact, according to the Official Journal of European Union<sup>5</sup>, the PDO ‘Nocciola Romana’ hazelnut consists of both ‘Tonda Gentile Romana’ and ‘Nocchione’ cultivars, which must represent at least 90% of the selection at the farm. Other permitted cultivars are ‘Tonda di Giffoni’ and ‘Barrettona’, which must represent a maximum of 10%.

Currently, the origin identification of products with the PDO label remains an economic burden for the food industry because commodities acquire an added value and are more likely subject to fraud. Thus, the ability to ensure the declared origin, cultivar and quality of the hazelnut kernel using a fast and easy-to-use method may supply the industry with considerable advantages. Menesatti *et al.*<sup>6</sup> successfully used shape-based imaging techniques in order to discriminate Italian hazelnut cultivars. However, their results were strongly affected by the viewpoint of the camera. Sensors for real-time, nondestructive systems using non-contact optical techniques (e.g., Near-Infrared

spectroscopy) for food quality and safety inspection are becoming an integral part of the food industry's move toward automation.<sup>7</sup> The spectroscopic measurement of food products is one of the most successful nondestructive techniques for quality assessment and can provide several quality details simultaneously,<sup>8</sup> usually related to the chemical constituents. Vis/NIR spectroscopy is fast, noninvasive, flexible, easy to use, environmentally benign and well-suited to rapid on-line and off-line applications compared with the conventional laboratory methods used to determine food quality.<sup>9,10</sup> Moreover, NIR spectroscopy coupled with chemometrics has been proven to be a versatile and valuable tool for assessing cultivar of food and for authenticating the geographical origin of various commodities.<sup>11–13</sup>

The objectives of the present study were to calibrate a NIR spectrometer in the spectral range of 1100 to 2300 nm for the cultivar detection of hazelnut kernels and to demonstrate the feasibility of NIR spectroscopy for the characterization and the classification of hazelnut samples according to the PDO 'Nocciola Romana' designation.

## MATERIALS AND METHODS

### *Sample preparation*

Whole, raw (unroasted) and shelled hazelnut kernels (*Corylus avellana* L.) were provided directly from a local industry (Viterbo, Central Italy) early in November. At the outset, the kernels were classified as follows: 600 were hazelnuts from PDO 'Nocciola Romana' (hereafter called NRM) (523 and 77 kernels from 'Tonda Gentile Romana' and 'Nocchione' cultivars, respectively) and 600 were hazelnuts from other cultivars (300 and 300 kernels from 'Tonda di Giffoni' and 'Barrettona' cultivars, respectively, hereafter called TGB). The hazelnut cultivars were chosen because the most widely cultivated variety in the 'Nocciola Romana' PDO region, accounting for

more than 90% of the total hazelnut production. The hazelnut kernels had an average moisture content equal to 4.0%. The fruits were stored at  $24 \pm 0.3^{\circ}\text{C}$  for 24 h to allow them to reach a temperature and moisture equilibrium prior to the NIR spectra acquisition. Because of the feasibility of NIR spectroscopy to detect flawed hazelnut,<sup>14</sup> following the NIR measurement, the fruits were dissected to identify hazelnuts with hidden flaws and to discard the spectra belonging to unsound fruits.

*NIR spectra measurements*

The NIR measurements were performed 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 and an indium gallium arsenide (InGaAs) array (range from 1100 to 2300 nm, 2-nm resolution) with a scanning time of 60 ms. Each acquired spectrum averaged 10 scans. The reference spectrum was automatically measured by the instrument as described by Cayuela and Weiland.<sup>15</sup> Two spectra were acquired from opposite sides of the same hazelnut, and the spectra were averaged. Measurements was checked in real-time and wrong sampling (e.g. low signal spectrum) was immediately repeated. Diffuse reflectance spectra were acquired and transformed into absorbance ( $A = \log R^{-1}$ ) using the Matlab R2013b statistics software (MathWorks, Natick, MA, USA).

*Chemometric procedure*

The NIR data matrix had 1200 rows (samples) and 601 columns (absorbance values at different wavelengths). Each hazelnut was modeled as a ‘data vector’, in which the features were the vector components.

Principal Component Analysis (PCA)<sup>16</sup> as an unsupervised technique was performed to evaluate the between-class similarity and to uncover any existing clustering of hazelnut samples based on either cultivar or PDO designation. PCA was applied to the whole mean-centered data matrix. The scree-plot criterion<sup>16</sup> (a plot of the

1  
2  
3 129 PCs as the X-axis and the corresponding eigenvalues as the Y-axis) was used to select  
4  
5 130 the required number of PCs for describing the dimensionality of the data.  
6

7  
8 131 Several mathematical transformations of the NIR spectra were performed to  
9  
10 132 improve the spectral information. In fact, data preprocessing (transformation and data  
11  
12 133 reduction) can markedly influence the classification results of spectral data, which  
13  
14 134 contain highly correlated variables as well as noise and irrelevant information.<sup>17</sup>  
15

16 135 Spectral preprocessing helps remove the spectral baseline shift and the scattering effect  
17  
18 136 caused by particle size differences. However, spectral information that can be useful for  
19  
20 137 the classification models could be lost in the data pretreatment process. Thus, during the  
21  
22 138 experimentation, the spectra were preprocessed following a variety of spectral  
23  
24 139 pretreatments, including scattering correction (Standard Normal Variate, SNV or  
25  
26 140 Multiplicative Scatter Correction, MSC), Savitzky-Golay first, second or third  
27  
28 141 derivatives (D1f, D2f or D3f, respectively) with a second or third-order polynomial and  
29  
30 142 with a smoothing with a window size of 5, 7 or 9 variables,<sup>18</sup> and Mean Centering  
31  
32 143 (MC). Each possible combination of preprocesses was also tested (Fig. 1).  
33  
34  
35

36 144 To discriminate hazelnuts on the basis of the PDO ‘Nocciola Romana’  
37  
38 145 regulation, k-Nearest Neighbor (kNN), Soft Independent Modeling of Class Analogy  
39  
40 146 (SIMCA), Partial Least Squares Discriminant Analysis (PLSDA) and Support Vector  
41  
42 147 Machine Discriminant Analysis (SVMDA) were tested as classification techniques.<sup>19</sup>  
43  
44 148 Specifically, a classification technique is a method that assigns objects to the class with  
45  
46 149 the maximum posterior probability. It differs from a class-modeling approach, which is  
47  
48 150 based on a bilateral significance test.<sup>20</sup> SVMDA was computed using the radial basis  
49  
50 151 function (RBF) for the kernel transformation. Therefore, it is essential to test the models  
51  
52 152 to evaluate the classification performance and to reduce the risk for overly optimistic  
53  
54 153 results. A reliable way of testing the model significance is through cross validation  
55  
56 154 (CV) and by predicting the results for independent sets of samples (test sets). Thus, with  
57  
58  
59  
60

the objective of circumventing overfitting and underfitting problems,<sup>21</sup> the samples were randomly split as follows: 50% of the samples were assigned to the calibration set (600 fruits) and the other 50% were split into three prediction sets (200 fruits each). Moreover, each model was optimized by computing a random subset cross-validation with  $\sqrt{600}$  data splits and 3 iterations.

The reliability of each classification model was determined on the basis of the lower number of required preprocessings as well as the classification sensitivity, classification specificity and classification accuracy rates.<sup>22</sup> The classification sensitivity rate is the number of correctly classified hazelnuts in the considered class (i.e., NRM) divided by the total number of hazelnuts in that class (Eq. 1). The classification specificity rate is the number of correctly classified hazelnuts in the other class (i.e., TGB) divided by the total number of hazelnuts in the other class (Eq. 2). The classification accuracy rate is the proportion of the true results in the population (Eq. 3).

(1) 
$$\text{Classification sensitivity rate} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

(2) 
$$\text{Classification specificity rate} = \frac{\text{True Negatives}}{\text{False Positives} + \text{True Negatives}}$$

(3) 
$$\text{Classification accuracy rate} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Positives} + \text{Total Negatives}}$$

The *t*-test at 95% confidence was performed to compare results from calibration, cross-validation and prediction of each metric (i.e. classification sensitivity, classification specificity and classification accuracy). The pairwise *t*-test was computed with the Bonferroni's correction in which the *p*-values are multiplied by the number of comparisons.

The data handling, chemometric and statistical analyses were performed using Matlab software R2013b coupled with both Statistic Toolbox and PLS\_Toolbox software v7.3.1 (Eigenvector Research Inc., Washington, USA).

## RESULTS AND DISCUSSIONS

The PCA analysis of the data variations provides insight into the class distribution. The first three principal components (PCs) accounted for 92.32% of the total variance (Fig. 2). When the scores of the samples are represented on two-dimensional spaces defined by PC1 and PC2 (Fig. 2a) and PC2 and PC3 (Fig. 2b), a natural separation between the samples was obtained, even though the classes partially overlapped.

From the loadings of the features (Fig. 2c), the most influential features on the first principal component (PC1, accounting for 64.25% of the total variance) were 1120 nm, 1720 nm, 1920 nm, 2120 nm and 2250 nm. The main contribution to the second principal component (PC2, 21.24% of the total variance) was due to the bands at 1120 nm, 1920 nm, 1950 nm and 2250 nm. The third principal component (PC3, 6.83% of the total variance) was mainly associated with the 1200-nm, 1620-nm, 1950-nm and 2198-nm bands.

The interpretation of NIR spectra is complicated by the fact that specific bands in this region are overtones containing information on various molecular vibration and generic functional groups. Although no references or previous studies are available regarding the classification of PDO hazelnuts using NIR spectroscopy and thus no corresponding correlation bands are known, knowledge of the sample's chemical composition allows some insight into the particular features that were found to provide good class discrimination. Thus, the 1120-nm and 1200-nm bands can be related to the

second overtone of C-H (C-H, C-H<sub>2</sub> and C-H<sub>3</sub>) stretching vibrations,<sup>23,24</sup> which are normally associated with the fatty acid contents of the samples. The bands at 1620 nm and 1720 nm can be assigned to the first overtone of C-H (C-H<sub>2</sub>, =C-H<sub>2</sub> and *cis* R<sub>1</sub>CH=CHR<sub>2</sub>CH<sub>3</sub>) stretches and deformations, which are generally related to unsaturated fatty acids with *cis* double bonds.<sup>25-27</sup> The 1920-nm and 1950-nm bands are mainly associated with the second overtone of C=O (C=ONH and C=OOR) due to protein, acid and ester stretching vibrations. Furthermore, the 2120-nm and 2198-nm bands involve polysaccharides (O-H bending and C-O stretching combination), carbohydrates (C-H stretching and C=O combination) and lipids (C-H stretching, C=O stretching combination and C-H deformation combination).<sup>28,29</sup> Thus, the features computed in the aforementioned wavelength regions were mainly correlated with lipids, which represent the main constituent of the hazelnut with an average content of 60.75%,<sup>30</sup> with some contribution from minor constituents, such as proteins and carbohydrates. The PCA results were confirmed by Cristofori *et al.*,<sup>31</sup> who demonstrated that information related to **major** constituents, such as fatty acids, sugars and phenols, are useful for discriminating hazelnut samples according to cultivar and year of production.

In some cases, water can interfere with the NIR analysis due to the strength and the broadband characteristics of water absorption. Thus, water bands may be excluded from the computations if the goal is to develop classification methods based on other minor constituents.<sup>9,32</sup>

Table 1 summarizes the best discriminant performances based on the ability of kNN, SIMCA, PLSDA and SVMDA to authenticate the PDO ‘Nocciola Romana’ hazelnut. The selected results indicate that the SNV baseline correction was necessary to obtain the highest **classification specificity** and **classification sensitivity** and consequently improve the **classification accuracy**, with the exception of the SIMCA

models. Moreover, MC resulted in improved classification accuracy for most of the discriminant methods used, with the exception of kNN models, probably by enhancing the differences between spectra.<sup>18</sup> The Savitzky-Golay smoothing filter also showed improvement in the performance of the SIMCA, PLSDA and SVMDA models. This finding suggests that the spectral data were slightly affected by issues due to noise and other irrelevant information. Thus, removing and/or reducing the non-relevant variability from the raw dataset were crucial for the development of accurate classification models.

Considering the large variability in hazelnut characteristics due to growth and harvest conditions, the proposed methodology suggested very high classification performances. In fact, several models exhibited very good (>95%) classification specificity and classification sensitivity, which means that the models presented a higher ability to discriminate between one class with respect to that class. Consequently, the results also indicated a lower chance of incorrectly classifying one class as belonging to another.

In general, the PLSDA and SVMDA models yielded better results. However, the highest classification accuracy was obtained with the SVMDA model (Table 1, ID #4). The results indicate very good recognition of the NRM and TGB samples using the SVMDA algorithm. In fact, the most accurate classification with classification specificity, classification sensitivity and classification accuracy rates of 96.0%, 95.0% and 95.5%, respectively, was obtained using SVMDA with the SNV, a 7-point Savitzky-Golay filter and mean centering. It is fundamental to consider that SVMDA assumes that all variables are equally important. Thus, the preprocessing techniques used may have had a significant impact on the model performance. The second best result, which corresponded to classification specificity, classification sensitivity and classification accuracy rates of 92.4%, 95.0% and 93.7%, was obtained using PLSDA

(Table 1, ID #3) with the following pretreatments: SNV, 7 point Savitzky-Golay filter with first derivative and mean centering.

Regarding the SIMCA model (Table 1, ID #2), although NRM misclassification was very low (1.3%), TGB misclassification was extremely high (32.0%). Thus, the SIMCA model was not appropriate for the development of a sorting system capable of ensuring an amount of ‘Tonda di Giffoni’ and ‘Barrettona’ cultivars lower than the 10% threshold admitted by the PDO ‘Nocciola Romana’ regulation.

The kNN was capable of performing good discrimination with a classification accuracy higher than 90% (Table 1, ID #1). The kNN method has the further advantage of simple implementation; it is a robust algorithm that requires few parameters to tune the classification process. However, similarly to the best SIMCA model, kNN also showed insufficient classification specificity for discriminating ‘Barrettona’ and ‘Tonda di Giffoni’ cultivars from PDO ‘Nocciola Romana’. Moreover, it should be taken into account that the disadvantages of kNN include classification sensitivity to unbalanced training sets, substantial computational work for large datasets, and computational inefficiency compared with standard normal theory methods.<sup>19</sup>

Finally, spatial distribution plots (Fig. 3) illustrate posterior-probability differences between kNN, SIMCA, PLSDA and SVMDA. In our study, PLSDA and SVMDA outperformed both kNN and SIMCA in ensuring the declared origin and quality of the PDO ‘Nocciola Romana’ hazelnut.

CONCLUSIONS

Currently, among the spectroscopic techniques, no NIR methods are used to perform hazelnut sorting based on the PDO label. In this study, calibration and characterization techniques consisting of a series of tests utilizing different classification techniques were successfully used for the development of a nondestructive system with

high classification sensitivity over the whole spectral range of the spectrophotometer (from 1100 nm to 2300 nm). Specifically, the feasibility of NIR spectroscopy for the robust authentication of the PDO 'Nocciola Romana' hazelnut was demonstrated. In fact, the information extracted in the spectral profile in combination with chemometrics enabled an unbiased differentiation of PDO 'Nocciola Romana' from other cultivars, fulfilling the hazelnut industry's need for traditional products with enhanced characteristics of quality and added value.

The large sample size used in this study guarantees negligible overfitting effects. Therefore, the models allowed the precise estimation of parameters and thus the predictions. In general, the SVMDA and PLSDA models were found to be promising in terms of both their classification accuracy and robustness based on the NIR spectroscopy application for the prediction of the PDO 'Nocciola Romana' label. Specifically, both SVMDA and PLSDA showed very good recognition performances (classification accuracy > 95%).

Finally, this method lays the foundations for a superior NIR detection system in terms of classification sensitivity, classification specificity, nondestructive analysis and automation. However, although the results obtained are promising, a larger validation sample must be used to address the additional possible variations expected from growing hazelnuts in different regions, crop years, agro-pedo-climatic conditions and degree of ripeness, in addition to other cultivars. Moreover, the potential use of Vis/NIR spectroscopy to separate hazelnuts according to their cultivar should become the subject of future research. Based on the results of this study, we expect that this classification system should be feasible.

## ACKNOWLEDGMENTS

305 The research has been financially supported by Mipaaf through the project  
306 ‘Miglioramento della filiera corilicola laziale – Mi.F.CO.L.’ represented by the ‘AOP  
307 Nocciola Italia Soc. Cons. s.r.l’ and ‘CeFAS - Azienda speciale CCIAA Viterbo’.

309 **REFERENCES**

310 1 Shahidi F and Miraliakbari H, in *Bailey’s Ind. Oil Fat Prod.*, ed. Fereidoon  
311 Shahidi, John Wiley & Sons Inc., Hoboken, New Jersey, United States, pp. 175–  
312 193 (2005).

313 2 Contini M, Baccelloni S, Massantini R and Anelli G, Extraction of natural  
314 antioxidants from hazelnut (*Corylus avellana* L.) shell and skin wastes by long  
315 maceration at room temperature. *Food Chem.*, **110**:659–669 (2008).

316 3 Kibar H and Öztürk T, The effect of moisture content on the physico-mechanical  
317 properties of some hazelnut varieties. *J. Stored Prod. Res.*, **45**:14–18 (2009).

318 4 Moscetti R, Frangipane MT, Monarca D, Cecchini M and Massantini R,  
319 Maintaining the quality of unripe, fresh hazelnuts through storage under modified  
320 atmospheres. *Postharvest Biol. Technol.*, **65**:33–38 (2012).

321 5 European Commission, Regulation (EC) No 510/2006 “Nocciola Romana” PDO.  
322 *Off. J. Eur. Union*, **C308**:19–22 (2008).

323 6 Menesatti P, Costa C, Paglia G, Pallottino F, D’Andrea S, Rimatori V and  
324 Aguzzi J, Shape-based methodology for multivariate discrimination among  
325 Italian hazelnut cultivars. *Biosyst. Eng.*, **101**:417–424 (2008).

326 7 Ballabio D and Todeschini R, in *Infrared Spectrosc. food Qual. Anal. Control*,  
327 ed. Sun D, Academic Press, London, UK, pp. 83–104 (2009).

328 8 Near-Infrared Spectroscopy in Food Science and Technology. John Wiley &  
329 Sons, Inc., Hoboken, NJ, USA, pp. 424 (2006).

330 9 Moscetti R, Haff RP, Saranwong S, Monarca D, Cecchini M and Massantini R,  
331 Nondestructive detection of insect infested chestnuts based on NIR spectroscopy.  
332 *Postharvest Biol. Technol.*, **87**:88–94 (2014).

333 10 Nicolai BM, Beullens K, Bobelyn E, Peirs A, Saeys W, Theron KI and  
334 Lammertyn J, Nondestructive measurement of fruit and vegetable quality by  
335 means of NIR spectroscopy: A review. *Postharvest Biol. Technol.*, **46**:99–118  
336 (2007).

337 11 Camina JM, Pellerano RG and Marchevsky EJ, Geographical and Botanical  
338 Classification of Honeys and Apicultural Products by Chemometric Methods. A  
339 Review. *Curr. Anal. Chem.*, **8**:408–425 (2012).

- 340 12 Versari A, Laurie VF, Ricci A, Laghi L and Parpinello GP, Progress in  
341 authentication, typification and traceability of grapes and wines by chemometric  
342 approaches. *Food Res. Int.*, **60**:2–18 (2014).
- 343 13 Gad H, El-Ahmady SH, Abou-Shoer MI and Al-Azizi MM, Application of  
344 chemometrics in authentication of herbal medicines: a review. *Phytochem. Anal.*,  
345 **24**:1–24 (2013).
- 346 14 Moschetti R, Haff RP, Aernouts B, Saeys W, Monarca D, Cecchini M and  
347 Massantini R, Feasibility of Vis/NIR spectroscopy for detection of flaws in  
348 hazelnut kernels. *J. Food Eng.*, **118**:1–7 (2013).
- 349 15 Cayuela J a. and Weiland C, Intact orange quality prediction with two portable  
350 NIR spectrometers. *Postharvest Biol. Technol.*, **58**:113–120 (2010).
- 351 16 Jolliffe IT, Principal Component Analysis. Springer, New York, USA, pp. 488  
352 (2002).
- 353 17 Wu W, Mallet Y and Walczak B, Comparison of regularized discriminant  
354 analysis linear discriminant analysis and quadratic discriminant analysis applied  
355 to NIR data. *Anal. Chim. Acta*, **2670**:257–265 (1996).
- 356 18 Boysworth MK and Booksh KS, in *Handb. Near-Infrared Anal.*, eds. Burns DA  
357 and Ciurczak EW, CRC Press, New York, USA, pp. 207–229 (2008).
- 358 19 Naes T, Isaksson T, Fearn T and Davies T, A user-friendly guide to Multivariate  
359 Calibration and Classification. NIR publications, Chichester, UK, pp. 354 (2004).
- 360 20 Boggia R, Zunin P, Lanteri S, Rossi N and Evangelisti F, Classification and  
361 Class-Modeling of “ Riviera Ligure ” Extra-Virgin Olive Oil Using Chemical –  
362 Physical Parameters. *J. Agric. Food Chem.*, **50**:2444–2449 (2002).
- 363 21 Gries D and Schneider FB, in *Guid. to Intell. Data Anal.*, eds. Gries D and  
364 Schneider FB, Springer-Verlag, London, UK, pp. 81–114 (2010).
- 365 22 Dejaegher B, Dhooghe L, Goodarzi M, Apers S, Pieters L and Vander Heyden Y,  
366 Classification models for neocryptolepine derivatives as inhibitors of the  $\beta$ -  
367 haematin formation. *Anal. Chim. Acta*, **705**:98–110 (2011).
- 368 23 Goddu RF, Determination of Unsaturation by Near-Infrared Spectrophotometry.  
369 *Anal. Chem.*, **29**:1790–1794 (1957).
- 370 24 Holman RT and Edmondson PR, Near-Infrared Spectra of Fatty Acids and Some  
371 Related Substances. *Anal. Chem.*, **28**:1533–1538 (1956).
- 372 25 Christy A a, Kasemsumran S, Du Y and Ozaki Y, The detection and  
373 quantification of adulteration in olive oil by near-infrared spectroscopy and  
374 chemometrics. *Anal. Sci.*, **20**:935–40 (2004).

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

375 26 Kavdir I, Buyukcan MB, Lu R, Kocabiyik H and Seker M, Prediction of olive  
376 quality using FT-NIR spectroscopy in reflectance and transmittance modes.  
377 *Biosyst. Eng.*, **103**:304–312 (2009).

378 27 Yildiz G, Wehling RL and Cuppett SL, Method for determining oxidation of  
379 vegetable oils by near-infrared spectroscopy. *J. Am. Oil Chem. Soc.*, **78**:495–502  
380 (2001).

381 28 Downey G, McIntyre P and Davies AN, Detecting and Quantifying Sunflower  
382 Oil Adulteration in Extra Virgin Olive Oils from the Eastern Mediterranean by  
383 Visible and Near-Infrared Spectroscopy. *J. Agric. Food Chem.*, **50**:5520–5525  
384 (2002).

385 29 Workman J and Weyer L, Practical Guide to Interpretive Near-Infrared  
386 Spectroscopy. CRC Press, London, UK, pp. 344 (2007).

387 30 USDA, National Nutrient Database for Standard Reference – Release 26. URL  
388 <http://ndb.nal.usda.gov/ndb/search/list>. Accessed 03.12.14. (2014).

389 31 Cristofori V, Ferramondo S, Bertazza G and Bignami C, Nut and kernel traits and  
390 chemical composition of hazelnut (*Corylus avellana* L.) cultivars. *Jorunal Sci.*  
391 *Food Agric.*, **88**:1091–1098 (2008).

392 32 Moscetti R, Monarca D, Cecchini M, Haff RP, Contini M and Massantini R,  
393 Detection of Mold-Damaged Chestnuts by Near-Infrared Spectroscopy.  
394 *Postharvest Biol. Technol.*, **93**:83–90 (2014).

395  
396

| ID # | Model | Preprocess      |                     |                 |                 | Feat. <sup>e</sup> | Classification specificity (%) |   |       |   |       |   | Classification sensitivity (%) |   |       |   |       |   | Classification accuracy (%) |   |       |   |       |   |
|------|-------|-----------------|---------------------|-----------------|-----------------|--------------------|--------------------------------|---|-------|---|-------|---|--------------------------------|---|-------|---|-------|---|-----------------------------|---|-------|---|-------|---|
|      |       | SC <sup>a</sup> | Deriv. <sup>b</sup> | SG <sup>c</sup> | MC <sup>d</sup> |                    | C                              |   | CV    |   | P     |   | C                              |   | CV    |   | P     |   | C                           |   | CV    |   | P     |   |
| 1    | kNN   | SNV             | -                   | -               | -               | -                  | 91.00                          | a | 89.89 | a | 91.33 | a | 92.00                          | a | 91.52 | a | 92.33 | a | 91.50                       | a | 90.71 | a | 91.83 | a |
| 2    | SIMCA | -               | -                   | 5               | Yes             | 5                  | 67.67                          | a | 65.01 | a | 68.00 | a | 98.33                          | a | 97.44 | a | 97.67 | a | 83.00                       | a | 81.23 | a | 82.84 | a |
| 3    | PLSDA | SNV             | D1 <sup>f</sup>     | 7               | Yes             | 4                  | 94.02                          | a | 92.58 | a | 92.38 | a | 95.32                          | a | 94.76 | a | 94.95 | a | 94.67                       | a | 93.67 | a | 93.67 | a |
| 4    | SVMDA | SNV             | -                   | 7               | Yes             | -                  | 98.34                          | a | 95.35 | b | 96.04 | b | 100.00                         | a | 94.98 | b | 94.99 | b | 99.17                       | a | 95.17 | b | 95.52 | b |

<sup>a</sup> Scatter correction

<sup>b</sup> Derivative

<sup>c</sup> Savitzky-Golay filter - smoothing points

<sup>d</sup> Mean Centering

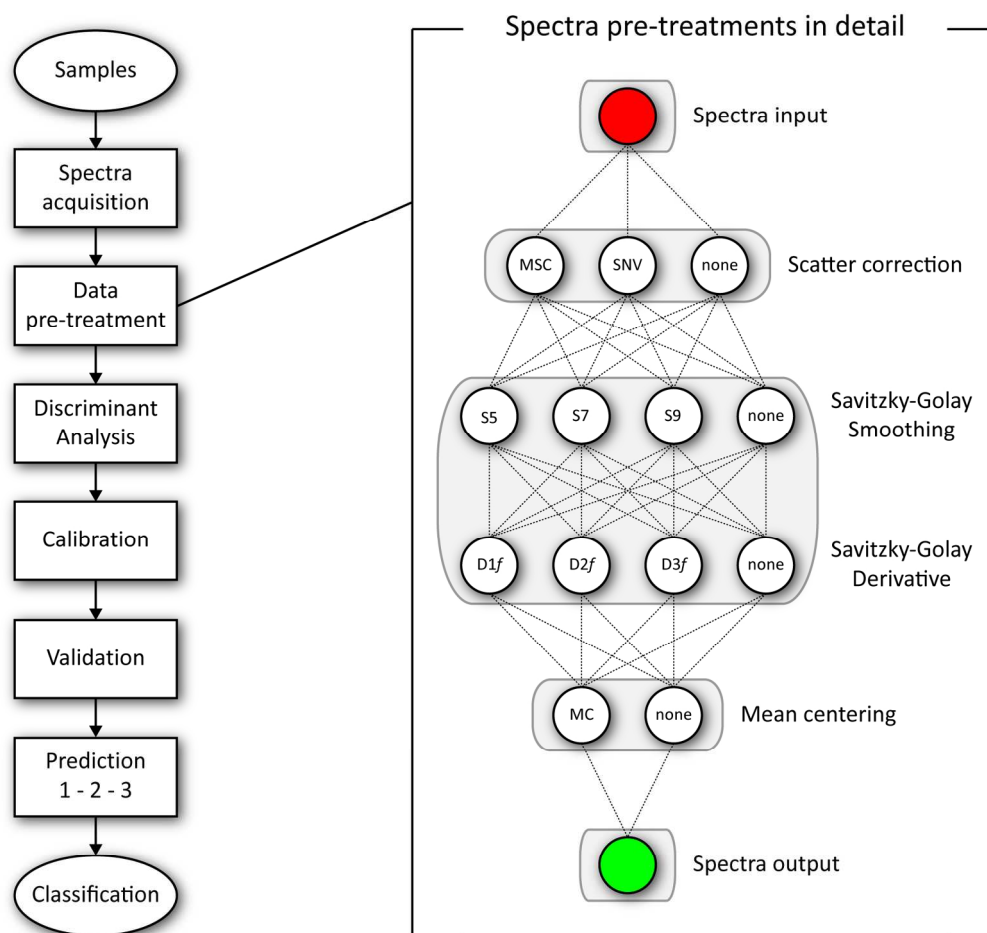
<sup>e</sup> Features: latent variables or principal components

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

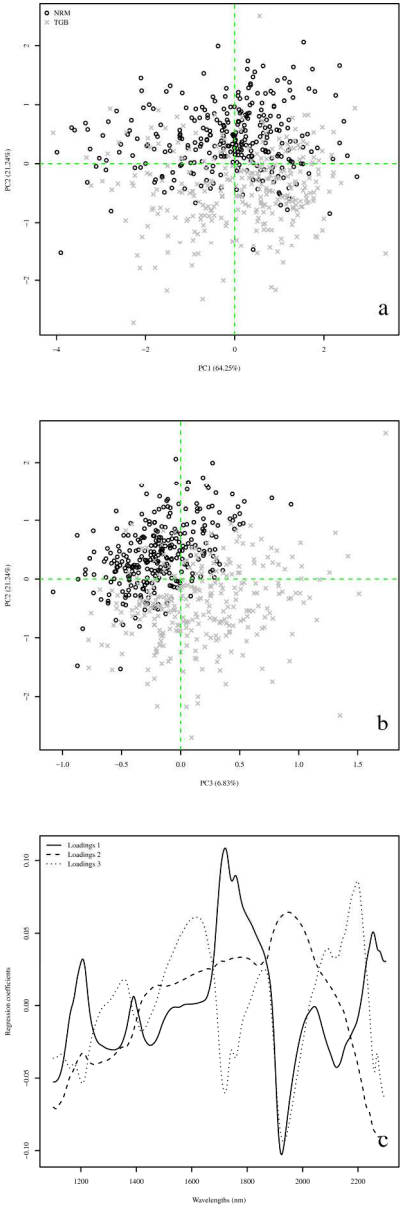
TABLE CAPTION

**Table 1.** Summarized best results for k-Nearest Neighbor (kNN), Soft Independent Modeling of Class Analogy (SIMCA), Partial Least Square Discriminant Analysis (PLSDA) and Support Vector Machine Discriminant Analysis (SVMDA) algorithms. Metric values within each model separated by different letters are significantly different ( $p < 0.05$ ).

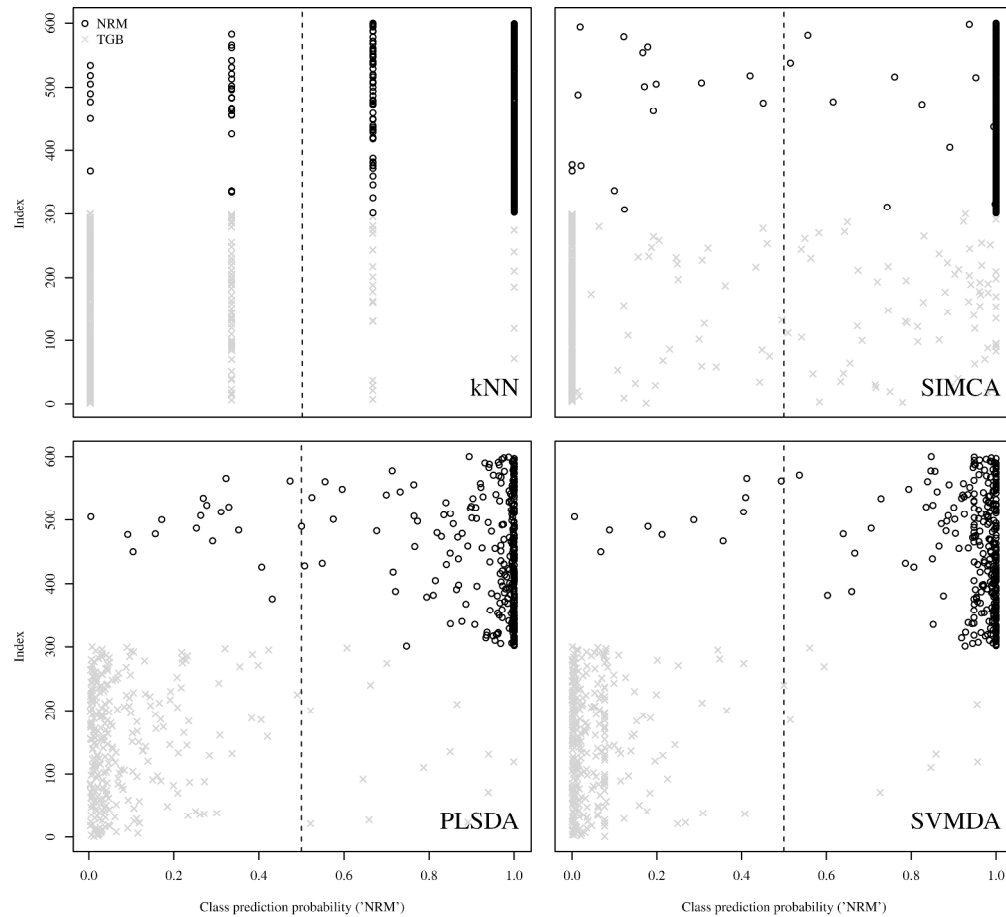
For Peer Review



160x151mm (300 x 300 DPI)



97x288mm (300 x 300 DPI)



181x164mm (600 x 600 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**FIGURE CAPTIONS**

- Figure 1.** Flowchart of PDO ‘Nocciola Romana’ hazelnut detection using an AOTF-NIR spectroscopy system.
- Figure 2.** Score plots for the PC1 and PC2 (a), and for PC2 and PC3 (b). Loading plots for the PC1, PC2 and PC3 (c).
- Figure 3.** Class prediction probability plots obtained from kNN, SIMCA, PLSDA and SVMDA models.

For Peer Review