# Comparative analysis of spectroscopic methods for rapid authentication of hazelnut cultivar and origin

B. Torres-Cobos [a,b], A. Tres [a,b], S. Vichi [a,b,*], F. Guardiola [a,b], M. Rovira [c], A. Romero [c], V. Baeten [d], J.A. Fernández-Pierna [d]

[a] Departament de Nutrició, Ciències de L'Alimentació i Gastronomia, Facultat de Farmàcia i Ciències de L'Alimentació, Universitat de Barcelona, Av Prat de La Riba, 171, 08921 Santa Coloma de Gramenet, Spain
[b] Institut de Recerca en Nutrició i Seguretat Alimentària (INSA-UB), Universitat de Barcelona, Av Prat de La Riba, 171, 08921 Santa Coloma de Gramenet, Spain
[c] Institute of Agrifood Research and Technology (IRTA), Ctra. de Reus – El Morell Km 3.8, Constantí 43120, Spain
[d] Quality and Authentication of Products Unit, Knowledge and Valorization of Agricultural Products Department, Walloon Agricultural Research Centre, Chaussée de Namur 24, 5030 Gembloux, Belgium

## HIGHLIGHTS

- Three spectroscopic methods were compared to verify hazelnut cultivars and origin.
- MIR and NIR spectroscopic methods achieved ≥93 % accuracy in classifying hazelnuts.
- Benchtop NIR spectroscopy showed superior performance for hazelnut authentication.
- Ground hazelnuts provide better results than whole kernels due to greater homogeneity.
- Models rely on protein and lipid composition for hazelnut discrimination.

## GRAPHICAL ABSTRACT
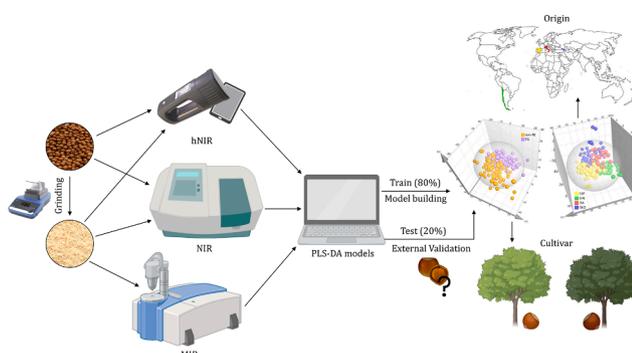


## ARTICLE INFO

## ABSTRACT

Hazelnut market prices fluctuate significantly based on cultivar and provenance, making them susceptible to counterfeiting. To develop an accurate authentication method, we compared the performances of three spectroscopic methods: near infrared (NIR), handheld near infrared (hNIR), and medium infrared (MIR), on over 300 samples from various origins, cultivars, and harvest years. Spectroscopic fingerprints were used to develop and externally validate PLS-DA classification models. Both cultivar and origin models showed high accuracy in external validation. The hNIR model effectively distinguished cultivars but struggled with geographic distinctions due to lower sensitivity. NIR and MIR models showed over 93 % accuracy, with NIR slightly outperforming MIR for geographic origin. NIR proved to be a fast and suitable tool for hazelnut authentication. This study is the first to systematically compare spectroscopic tools for authenticating hazelnut cultivar and origin using the same dataset, offering valuable insights for future food authentication applications.

---

\* Corresponding author at: Departament de Nutrició, Ciències de L'Alimentació i Gastronomia, Facultat de Farmàcia i Ciències de L'Alimentació, Universitat de Barcelona, Av Prat de La Riba, 171, 08921 Santa Coloma de Gramenet, Spain.
*E-mail address:* stefaniavichi@ub.edu (S. Vichi).

## 1. Introduction

Hazelnuts are one of the most consumed nuts in Europe, as raw and roasted fruits or included in many preparations and traditional dishes. In addition, they are widely used in the chocolate and confectionary industries. The main hazelnut-producing countries are Turkey, Italy, the USA, Azerbaijan, Chile and Georgia [1]. Hazelnut nutritional values and sensory attributes are greatly influenced by geographical and varietal origin [2–4]. Consequently, their market price also fluctuates significantly depending on cultivar and provenance. Italian and Spanish hazelnuts are among the highest priced, with values of 3,416 USD/T and 2,434 USD/T, respectively, in contrast to Georgian kernels, which are priced at 1,287 USD/T [5]. Besides, hazelnuts with special geographical indications, such as Protected Designation of Origin or Protected Geographical Indication, are also highly appreciated by consumers, which is reflected in their price.

The high economic value of hazelnuts makes them susceptible to fraudulent practices. This susceptibility is accentuated by the absence of effective fraud detection methods. Appropriate tools to verify the cultivar and origin of hazelnuts are therefore necessary to ensure their authenticity and to safeguard consumers' rights.

Different analytical approaches have addressed this issue; morphological analysis [6,7], genetic methods [8,9], metabolomic analysis of proteins, phenolic and lipidic compounds by chromatographic techniques [6,10–12] and proton nuclear magnetic resonance ($^1$H NMR) [13] have been applied for hazelnut authentication. Nevertheless, these methods are usually laborious, time consuming, not suitable to be applied on-site (i.e. in the field, in the industry or in storage/retail facilities), require highly trained personnel and have a substantial cost.

Spectroscopic methods, such as mid-infrared (MIR) and near-infrared (NIR) spectroscopy, are fast, simple, environmentally friendly and non-destructive techniques, which can be applied directly to the solid sample without complex sample pretreatment steps. They allow simultaneous analysis of different chemical compounds and can be easily adapted to on-site or on-line applications, making them suitable for routine analysis.

These techniques, coupled with chemometric data analysis tools, have been successfully applied to the varietal and geographical authentication of different nuts [14–18]. In particular, previous works used NIR spectroscopy to authenticate an Italian PDO "Nocciola Romana" [19,20] to distinguish between Turkish hazelnut cultivars [21], or to discriminate hazelnuts from different origins [22,23]. However, none of them has been applied to simultaneously authenticate the origin and cultivar of hazelnuts. To the authors' knowledge, only Manfredi et al. [24] applied MIR spectroscopy to hazelnut halves, successfully differentiating between three cultivars. No further studies using MIR spectroscopy to authenticate hazelnuts have been found, underscoring the need for further evaluation of the technique's potential. Overall, studies using large sample sets, including higher variability in terms of harvest years, origins, cultivars and producers, are still needed to further corroborate the suitability of NIR and MIR spectroscopies as tools for geographical and varietal authentication of hazelnuts.

On the other hand, the performance of different spectroscopic methods for hazelnut authentication has not been systematically compared. Such a comparison would be particularly interesting in the case of MIR and NIR, as each provides different but complementary information. NIR spectroscopy includes bands issued from overtones and combination vibrations. It allows direct analysis of highly absorbing samples. NIR measurements often result in overlapping bands and complex spectra, so the bands are less specific in the NIR range than in MIR, which may hinder their interpretability. In contrast, MIR spectra include information of fundamental vibrations of specific functional groups, providing spectra with better resolved bands that can be used for structural identification [25]. In addition, NIR provides deeper light penetration into the sample matter, performing better with bulk or intact heterogeneous samples, minimizing the need for sample preparation and having a wider scope of applications. This makes it ideal for in situ analysis, as it requires less specificity requirements and less sample preparation [26,27]. On this basis, a large number of handheld NIR (hNIR) devices have been developed to authenticate a wide range of products [28–32]. These devices are small, compact, robust, more practical and affordable; they do not require a high level of expertise and can be applied for routine in-field analysis. However, handheld systems can reduce the accuracy of measurement, which can affect spectral quality compared to benchtop alternatives [33].

The aim of this research was to systematically apply, compare and evaluate the effectiveness of NIR and MIR spectroscopies for hazelnut authentication. The main goal was to develop an accurate method to simultaneously discriminate hazelnuts according to their geographical and varietal origin. For this purpose, in order to ensure a direct comparison between the different spectroscopic techniques and to properly evaluate their performances as hazelnut authentication tools, we analysed the same set of hazelnuts (both whole kernels and ground samples) from different origins, cultivars, and harvest years. This analysis was carried out by a NIR spectrometer, a hNIR spectrometer, and a MIR spectrometer (the latter only for ground samples). Subsequently, individual PLS-DA classification models for cultivar and origin discrimination were built for each technique. The geographical origin was discriminated between samples of the same cultivar ('Tonda di Giffoni', TG) produced in different countries, and the cultivar was discriminated between samples of the same origin (Spain). The cultivar models focused on discriminating TG hazelnuts, one of the most widespread cultivars worldwide, from other (non-TG) cultivars, while the geographical origin models were designed to classify samples according to their country of origin. The classification models were externally validated and evaluated for both fit and predictive ability.

## 2. Material and methods

### 2.1. Samples

A set of 302 traceable hazelnut samples was obtained directly from producers within the framework of the TRACENUTS project (PID2020-117701RB-I00). Samples were collected over four consecutive harvest years, from 2019 to 2022. From these samples, 200 were of the 'Tonda di Giffoni' cultivar (TG) from Chile (CHL, n = 40), Spain (ESP, n = 91), Georgia (GEO, n = 34) and Italy (ITA, n = 35), while 102 were from different cultivars (non-TG) produced in Spain (Table S1, Supplementary material). Samples were stored vacuum-packed at 4 °C until analysis.

### 2.2. Sample preparation

Raw hazelnut kernels with skin were directly analysed by NIR spectroscopy (benchtop and handheld device). Then, 15 g of the sample was ground with an IKA TUBE MILL control, (IKA, Staufen, Germany), and was analysed by NIR (benchtop and handheld device) and MIR spectroscopy.

### 2.3. NIR spectroscopy

#### 2.3.1. Benchtop NIR spectrometer

The NIR measurements of the whole and ground hazelnut kernels were performed on a benchtop DS3 FOSS spectrometer (FOSS Analytics, Hilleroed, Denmark) acquiring spectra every 2 nm within the wavelength range of 400–2500 nm, (spectral resolution of 0.5 nm). Whole kernels were presented in a cell quarter cup, while ground samples were placed in ring cups. A background reference spectrum was collected before each measure. Two consecutive measures, being each the average of seven spectra at 4 different points of the cup (28 spectra), were collected for every sample. Both measures were averaged previously to chemometric analysis. ISIscan Nova ™ (2021, FOSS) was used for NIR

spectra recording.

### 2.3.2. Handheld NIR spectrometer (hNIR)

NIR spectra of the whole and ground samples, presented in the same way as before, were also acquired using a handheld device, NeoSpectra Scanner (Si-Ware Systems Inc., California, U.S.). Data was collected with the NeoSpectra Scan software v7.1. Spectra was acquired every 2 nm within the wavelength range of 1350–2500 nm, (spectral resolution of 16 nm) and a scanning time of 3 s. Two consecutive replicates were collected per sample and calibration was run every 10 measures (5 samples). Prior to data analysis the corresponding replicated spectra were averaged.

### 2.4. MIR spectroscopy

MIR measurements of hazelnut ground samples were performed on a Vertex 70 spectrometer (Bruker Optics, Ettlingen, Germany) equipped with Attenuated Total Reflectance (ATR) and an integrated press. Samples were directly deposited on the crystal and pressed against it to ensure optimum contact with the diamond. Each measure was the average spectrum of 64 scans. Spectra was acquired every 1.93 cm$^{-1}$ within the wavelength range of 600–4000 cm$^{-1}$, (spectral resolution of 4 cm$^{-1}$). Background readings were collected prior to each sample spectrum collection. Three consecutive replicates were collected per sample and averaged prior to chemometric analysis. OPUS software (version 8.2.28) was used for MIR spectra recording.

### 2.5. Chemometrics

Spectral data matrices were extracted as csv files and processed using SIMCA v13.0© (Sartorius Lab Instruments GmbH & Co. KG, Gottingen, Germany).

### 2.5.1. Partial least squares discriminant analysis (PLS-DA)

For each method and sample preparation (NIR, whole or ground; hNIR, whole or ground; and MIR, ground), individual PLS-DA classification models were built to authenticate cultivar or geographical origin. Cultivar models were binary models to discriminate between two classes of Spanish samples (n = 193): TG cultivar (n = 91) and non-TG cultivar (n = 102). Origin models were multi-class models aimed at classifying TG samples (n = 200) according to their country of origin: CHL (n = 40), ESP (n = 91), GEO (n = 34) or ITA (n = 35).

Prior to model building, for each type of authentication model (origin or cultivar), each sample set was randomly split into training (80 % of the samples of each category: cultivar model, n = 154; origin model, n = 160) and validation set (20 % of the samples of each category: cultivar model, n = 39; origin model n = 40). The splitting was run seven times (7 iterations) for each authentication model to increase the robustness of the external validation, resulting in seven different training sets and their corresponding validation sets. Although randomly split, a stratified sampling strategy was followed by maintaining the variability and proportions of the sample set in the validation and training sets (Table S1, Supplementary material). The exact same splitting and resulting training and validation sets were used for all the methods to ensure the direct comparability between them.

In cultivar models (by binary PLS-DA), classes were expressed as PLS dummy variables, where 'non-TG' was represented as 1 and 'TG' as 0. The PLS predicted value (PV) of each sample was used for its classification into one of the two classes based on a classification threshold (PV = 0.5). Origin models (by multi-class PLS-DA) operated as multiple binary models, each comparing one class against the others. A dummy Y matrix held a set of classification vectors equal to the number of classes, where each vector had a value of 1 for one class (CHL, ESP, GEO or ITA) and 0 for all the other classes (non-CHL, non-ESP, non-GEO or non-ITA). Each sample was classified into the class corresponding to the vector leading to the highest PV, provided it was above the classification

threshold (here, PV = 0.5). Samples with PV below the classification threshold (PV < 0.5) for all vectors were not assigned to any class.

Models were developed on the training sets and first internally validated through leave 10 %-out cross-validation using the samples of the training sets. The optimal number of latent variables (LV) and pre-processing were selected according to the lowest Root Mean Squared Error of Cross Validation (RMSEcv) criteria. Hotelling's T$^2$ and Q residuals were used to detect outliers [34]. Permutation test (n = 20 permutations) and ANOVA on the cross-validated predictive residuals (p-value) [35] were carried out to assess the risk of model overfitting. Finally, models were externally validated by predicting the class of the samples in the respective validation set, which had not been used to build the models. The suitability of each PLS-DA model was evaluated by the Q$^2$ values and efficiency, which was expressed as the rate of correct classification of each class. Additionally, for the binary cultivar models the sensitivity, Eq. (1) and specificity, Eq. (2) were also assessed. True positives were the non-TG samples correctly classified, and true positives + false negatives corresponded to the total non-TG samples. True negatives were the TG samples correctly classified, and true negatives + false positives corresponded to the total TG samples [36].

$$Sensitivity = \frac{true\ positives}{[true\ positives + false\ negatives]} \qquad (1)$$

$$Specificity = \frac{true\ negatives}{[true\ negatives + false\ positives]} \qquad (2)$$

The performance of models from each method and sample preparation was compared to determine the most suitable one for authentication.

### 2.5.2. Evaluation of PLS-DA regression coefficients

For the methods that gave the most promising results, models were rebuilt using all samples (cultivar n = 193, or origin n = 200), and their regression coefficients were explored to identify the most relevant variables for the classification (cultivar or origin) and tentatively link them to chemical species. The jack-knife standard error of cross-validation (SEcv) was used to evaluate the significance of the regression coefficients, with values exceeding their corresponding SEcv considered significant. Out of the significant variables, only those with the highest absolute values (25 % higher than the average of the coefficients) were considered.

## 3. Results and discussion

### 3.1. PLS-DA classification models

#### 3.1.1. Whole kernels

For the PLS-DA models developed on the whole kernels analysed by the benchtop NIR and hNIR spectrometers data, the optimal pre-processing, according to the lowest RMSEcv criterion, was a first derivative, Savitzky–Golay smoothing (second order with a polynomial filter of 15-point window) and mean centring and scaling to unit of variance. For models developed with the benchtop NIR spectrometer, data was also pre-processed with the SNV. No outliers were detected according to the Hotelling's T$^2$ range and Q residuals.

The cross-validation results of the models built from training sets (7 iterations) were promising for TG cultivar discrimination for both NIR and hNIR models, with correct classification rates of 96.0 % and 87.2 %, respectively (Table S2, Supplementary material). However, the performance of the model for discrimination according to origin appears to be lower, in particular when using hNIR data, achieving only a 66.3 % of correct classification (Table S3, Supplementary material).

These findings were corroborated through external validation by predicting the class of the samples of the corresponding validation sets. Tables 1 and 2 present the mean values obtained from seven iterations of the external validation of each type of authentication model (Cultivar:

TG/non-TG; Geographical origin: CHL/ESP/GEO/ITA) developed with NIR and hNIR data from whole kernels. In all cases involving whole kernel analysis, the NIR models outperformed the hNIR models with higher sensitivity, specificity, and total correct classification rate (91.2 % vs 82.4 % in cultivar model; 81.1 % vs 53.2 % in origin model). The NIR and hNIR cultivar models showed high sensitivity (0.90 and 0.83) and specificity (0.93 and 0.82), proving their efficiency in discriminating TG samples from other cultivars (Table 2). Nonetheless, the origin models showed lower rates of correct classification for both NIR and hNIR, especially for GEO and ITA (Table 3). As observed, neither the NIR nor the hNIR were able to accurately classify the ITA samples, which were mostly not assigned to any class or were misclassified as ESP samples. This could be attributed to the proximity of these two regions and the similarity of their pedoclimatic conditions, which may result in similar effects on hazelnut composition. The poor performance of the hNIR origin models indicates that it is not suitable for accurate classification of whole hazelnuts according to their geographical origin.

### 3.1.2. Ground samples

For NIR models based on ground samples, a first derivative and smoothing by Savitzky–Golay (second-order polynomial filter with a 15-point window) was applied along with SNV and mean centring and scaling to unit of variance. For hNIR models, the optimal pre-processing was SNV followed by mean centring and scaling to unit of variance, whereas for MIR models a first derivative and smoothing (Savitzky–Golay second-order polynomial filter with a 5-point window) was applied. No outliers were detected according to the Hotelling's $T^2$ range and Q residuals. Both types of authentication models (origin and cultivar) developed on ground samples by NIR and MIR presented successful cross-validation results with a high overall rate of correct classification (98.7–99.8 % cultivar model; 95.4–98.0 % origin model), while the accuracy was lower for hNIR models (89 % overall correct classification for both models) (Tables S4 and S5 of Supplementary information).

According to the external validation results, hNIR model of ground

**Table 1**

External validation of PLS-DA cultivar models developed on whole kernels analysed by benchtop NIR and handheld NIR (hNIR) spectrometers. Results are mean values (±standard deviation) obtained from seven iterations.

| Whole kernel – Cultivar model: TG/non-TG | | | | | |
|---|---|---|---|---|---|
| NIR spectrometer (LVs = 9, $Q^2$ = 0.47, RMSEcv = 0.38)[a] | | | | | |
| | n | Non-TG (n) | TG (n) | Correct classification (%) | Sensitivity | Specificity |
| Non-TG | 21 | 18.9 ± 1.3 | 2.1 ± 1.3 | 89.8 ± 6.4 | 0.90 ± 0.06 | |
| TG | 18 | 1.3 ± 1.1 | 16.7 ± 1.1 | 92.9 ± 6.2 | | 0.93 ± 0.06 |
| Total | 39 | | | 91.2 ± 4.4 | | |
| **hNIR spectrometer** (LVs = 6, $Q^2$ = 0.35, RMSEcv = 0.41)[a] | | | | | |
| | n | Non-TG (n) | TG (n) | Correct classification (%) | Sensitivity | Specificity |
| Non-TG | 21 | 17.4 ± 1.7 | 3.6 ± 1.7 | 83.0 ± 8.2 | 0.83 ± 0.08 | |
| TG | 18 | 3.3 ± 1.3 | 14.7 ± 1.3 | 81.7 ± 7.0 | | 0.82 ± 0.07 |
| Total | 39 | | | 82.4 ± 6.4 | | |

For all models, ANOVA p-value < 0.05.
[a] Model parameters: mean values obtained with the training sets from 7 iterations. LVs: latent variables of the training model; $Q^2$: Cumulative fraction of Y variation predicted by the X training model up to the specified latent variable, according to cross-validation; RMSEcv: root mean square error of the cross validation of the training model. TG: 'Tonda di Giffoni'; non-TG: other cultivars.

**Table 2**

External validation of PLS-DA origin models developed on whole kernels analysed by benchtop NIR (NIR) and handheld NIR (hNIR) spectrometers. Results are mean values (±standard deviation) obtained from seven iterations.

| Whole kernel – Geographical origin model: CHL/ESP/GEO/ITA | | | | | | |
|---|---|---|---|---|---|---|
| NIR spectrometer (LVs = 14, $Q^2$ = 0.27, RMSEcv = 0.36)[a] | | | | | | |
| | n | CHL (n) | ESP (n) | GEO (n) | ITA (n) | Not assigned (n) | Correct classification (%) |
| CHL | 8 | 6.6 ± 1.1 | 0.6 ± 0.8 | 0.0 ± 0.0 | 0.0 ± 0.0 | 0.9 ± 1.2 | 82.1 ± 14.2 |
| ESP | 18 | 0.0 ± 0.0 | 16.7 ± 1.0 | 0.1 ± 0.4 | 0.3 ± 0.5 | 0.9 ± 0.9 | 92.9 ± 5.3 |
| GEO | 7 | 0.3 ± 0.5 | 0.4 ± 0.8 | 5.0 ± 1.5 | 0.3 ± 0.5 | 1.0 ± 1.0 | 71.4 ± 21.8 |
| ITA | 7 | 0.0 ± 0.0 | 1.6 ± 1.0 | 0.0 ± 0.0 | 4.1 ± 1.3 | 1.3 ± 1.7 | 59.2 ± 19.2 |
| Total | 40 | | | | | 4.0 ± 2.2 | 81.1 ± 7.2 |
| **hNIR spectrometer** (LVs = 9, $Q^2$ = 0.23, RMSEcv = 0.36)[a] | | | | | | |
| | n | CHL (n) | ESP (n) | GEO (n) | ITA (n) | Not assigned (n) | Correct classification (%) |
| CHL | 8 | 3.7 ± 1.6 | 0.9 ± 0.7 | 0.0 ± 0.0 | 0.4 ± 0.5 | 3.0 ± 1.2 | 46.4 ± 20.0 |
| ESP | 18 | 0.9 ± 0.9 | 12.9 ± 2.0 | 0.6 ± 0.5 | 0.0 ± 0.0 | 3.7 ± 2.3 | 71.4 ± 10.8 |
| GEO | 7 | 0.0 ± 0.0 | 1.9 ± 0.9 | 2.1 ± 1.1 | 0.1 ± 0.4 | 2.9 ± 1.6 | 30.6 ± 15.3 |
| ITA | 7 | 0.7 ± 0.5 | 2.1 ± 1.5 | 0.0 ± 0.0 | 2.6 ± 0.8 | 1.6 ± 1.3 | 36.7 ± 11.2 |
| Total | 40 | | | | | 11.1 ± 3.2 | 53.2 ± 5.7 |

For all models, ANOVA p-value < 0.05.
[a] Model parameters: mean values obtained with the training sets from 7 iterations. LVs: latent variables of the training model; $Q^2$: Cumulative fraction of Y variation predicted by the X training model up to the specified latent variable, according to cross-validation; RMSEcv: root mean square error of the cross validation of the training model. CHL: Chile; ESP: Spain; GEO: Georgia; ITA: Italy.

samples could be useful to discriminate TG cultivar from other cultivars (0.80 sensitivity, 0.89 specificity and 83.9 % correct classification) (Table 3) but it was still unable to accurately distinguish Georgian and Italian samples from those produced in other regions (Table 4). This could be attributed to the lower sensitivity of the hNIR spectrometer compared to a benchtop spectrometer.

On the other hand, the NIR and MIR models based on ground sample data performed satisfactorily in external validation. For cultivar models, both techniques achieved a sensitivity equal or higher than 0.92 and a specificity of 0.98, resulting in an overall classification rate of 95 % for both techniques (Table 3). Concerning the geographical origin models, NIR outperformed MIR in classifying hazelnuts from GEO (91.8 % vs 85.7 %) and ITA (91.8 % vs 83.7 %), providing slightly better overall classification results (96.4 % vs 93.9 %, respectively) (Table 4). Consequently, NIR proved to provide the most successful spectroscopic model for hazelnut varietal and geographical authentication.

In all cases, the results for the ground samples outperformed those obtained with the whole kernels, in line with studies performed on other nuts [37]. This could be attributed to the higher homogeneity and representativeness of the ground samples compared to analysing only a

**Table 3**

External validation of PLS-DA cultivar models developed on ground samples analysed by benchtop NIR (NIR), handheld NIR (hNIR) and MIR spectrometers. Results are mean values (±standard deviation) obtained from seven iterations.

Ground samples – Cultivar model: TG/non-TG

NIR spectrometer (LVs = 10, $Q^2$ = 0.53, RMSEcv = 0.34)[a]

| | n | Non-TG (n) | TG (n) | Correct classification (%) | Sensitivity | Specificity |
|---|---|---|---|---|---|---|
| Non-TG | 21 | 19.3 ± 0.8 | 1.7 ± 0.8 | 91.8 ± 3.6 | 0.92 ± 0.04 | |
| TG | 18 | 0.4 ± 0.5 | 17.7 ± 0.5 | 98.4 ± 2.7 | | 0.98 ± 0.03 |
| Total | 39 | | | 94.9 ± 2.1 | | |

**hNIR spectrometer** (LVs = 9, $Q^2$ = 0.39, RMSEcv = 0.39)[a]

| | n | Non-TG (n) | TG (n) | Correct classification (%) | Sensitivity | Specificity |
|---|---|---|---|---|---|---|
| Non-TG | 21 | 16.1 ± 1.0 | 4.3 ± 1.0 | 79.6 ± 4.5 | 0.80 ± 0.05 | |
| TG | 18 | 2.0 ± 0.8 | 16.0 ± 0.8 | 88.9 ± 4.5 | | 0.89 ± 0.05 |
| Total | 39 | | | 83.9 ± 2.4 | | |

**MIR spectrometer** (LVs = 10, $Q^2$ = 0.63, RMSEcv = 0.30)[a]

| | n | Non-TG (n) | TG (n) | Correct classification (%) | Sensitivity | Specificity |
|---|---|---|---|---|---|---|
| Non-TG | 21 | 19.6 ± 1.4 | 1.4 ± 1.4 | 93.2 ± 6.7 | 0.93 ± 0.07 | |
| TG | 18 | 0.4 ± 0.5 | 17.6 ± 0.5 | 97.6 ± 3.0 | | 0.98 ± 0.03 |
| Total | 39 | | | 95.2 ± 2.3 | | |

For all models, ANOVA p-value < 0.05.

[a] Model parameters: mean values obtained with the training sets from 7 iterations. LVs: latent variables of the training model; $Q^2$: Cumulative fraction of Y variation predicted by the X training model up to the specified latent variable, according to cross-validation; RMSEcv: root mean square error of the cross validation of the training model. TG: 'Tonda di Giffoni'; non-TG: other cultivars.

single or a few kernels per sample. Additionally, grinding not only improves sample homogeneity but also causes oil release from the cells, which affects the primary spectral signals. In whole kernels with skin, the dominant signals arise primarily from the lignocellulosic structure, whereas in ground samples, the signal from the released oil becomes more prominent. Consequently, the difference in model performance can be partly attributed to this shift in dominant signals between the two sample types.

### 3.2. Regression coefficients

The regression coefficients of the two best performing methods (NIR and MIR spectroscopic analysis of ground hazelnuts) were explored to identify the most informative variables in PLS-DA models for cultivar or origin discrimination and to relate them to the chemical composition of the samples.

#### 3.2.1. NIR spectroscopy

The interpretation of NIR spectra might be hindered by the fact that some of the bands in the analysed range are overtones containing generic information corresponding to different molecular vibrations of different functional groups. Even so, prior knowledge of the hazelnut's chemical composition enables the tentative identification of variables relevant in classifying the samples within each category. These variables can then be correlated with the main compositional constituents of the samples (lipids, proteins, carbohydrates).

**Table 4**

External validation of PLS-DA origin models developed on ground samples analysed by benchtop NIR (NIR), handheld NIR (hNIR) and MIR spectrometers. Results are mean values (±standard deviation) obtained from seven iterations.

Ground samples – Geographical origin model: CHL/ESP/GEO/ITA

NIR spectrometer (LVs = 14, $Q^2$ = 0.66, RMSEcv = 0.25)[a]

| | n | CHL (n) | ESP (n) | GEO (n) | ITA (n) | Not assigned (n) | Correct classification (%) |
|---|---|---|---|---|---|---|---|
| CHL | 8 | 8.0 ± 0.0 | 0.0 ± 0.0 | 0.0 ± 0.0 | 0.0 ± 0.0 | 0.0 ± 0.0 | 100.0 ± 0.0 |
| ESP | 18 | 0.0 ± 0.0 | 17.7 ± 0.5 | 0.0 ± 0.0 | 0.0 ± 0.0 | 0.3 ± 0.5 | 98.4 ± 2.7 |
| GEO | 7 | 0.0 ± 0.0 | 0.3 ± 0.5 | 6.4 ± 0.4 | 0.0 ± 0.0 | 0.4 ± 0.5 | 91.8 ± 7.6 |
| ITA | 7 | 0.1 ± 0.4 | 0.0 ± 0.0 | 0.0 ± 0.0 | 6.4 ± 0.8 | 0.4 ± 0.5 | 91.8 ± 11.2 |
| Total | 40 | | | | | 1.1 ± 0.9 | 96.4 ± 2.4 |

**hNIR spectrometer** (LVs = 9, $Q^2$ = 0.50, RMSEcv = 0.29)[a]

| | n | CHL (n) | ESP (n) | GEO (n) | ITA (n) | Not assigned (n) | Correct classification (%) |
|---|---|---|---|---|---|---|---|
| CHL | 8 | 7.3 ± 0.5 | 0.0 ± 0.0 | 0.0 ± 0.0 | 0.0 ± 0.0 | 0.7 ± 0.5 | 91.1 ± 6.1 |
| ESP | 18 | 0.1 ± 0.4 | 17.1 ± 0.9 | 0.0 ± 0.0 | 0.0 ± 0.0 | 0.7 ± 1.0 | 95.2 ± 5.0 |
| GEO | 7 | 0.1 ± 0.4 | 0.7 ± 0.5 | 4.1 ± 1.1 | 0.3 ± 0.8 | 1.7 ± 1.3 | 59.2 ± 15.3 |
| ITA | 7 | 0.0 ± 0.0 | 0.6 ± 0.5 | 0.3 ± 0.5 | 4.6 ± 0.5 | 1.6 ± 0.8 | 65.3 ± 7.6 |
| Total | 40 | | | | | 4.7 ± 1.8 | 82.9 ± 1.7 |

**MIR spectrometer** (LVs = 12, $Q^2$ = 0.61, RMSEcv = 0.26)[a]

| | n | CHL (n) | ESP (n) | GEO (n) | ITA (n) | Not assigned (n) | Correct classification (%) |
|---|---|---|---|---|---|---|---|
| CHL | 8 | 8.0 ± 0.0 | 0.0 ± 0.0 | 0.0 ± 0.0 | 0.0 ± 0.0 | 0.0 ± 0.0 | 100.0 ± 0.0 |
| ESP | 18 | 0.0 ± 0.0 | 17.7 ± 0.5 | 0.0 ± 0.0 | 0.0 ± 0.0 | 0.3 ± 0.5 | 98.4 ± 2.7 |
| GEO | 7 | 0.0 ± 0.0 | 0.6 ± 0.5 | 6.0 ± 0.6 | 0.0 ± 0.0 | 0.4 ± 0.8 | 85.7 ± 8.2 |
| ITA | 7 | 0.0 ± 0.0 | 0.1 ± 0.4 | 0.0 ± 0.0 | 5.9 ± 0.7 | 1.0 ± 0.6 | 83.7 ± 9.9 |
| Total | 40 | | | | | 1.7 ± 1.1 | 93.9 ± 2.4 |

For all models, ANOVA p-value < 0.05.

[a] Model parameters: mean values obtained with the training sets from 7 iterations. LVs: latent variables of the training model; $Q^2$: Cumulative fraction of Y variation predicted by the X training model up to the specified latent variable, according to cross-validation; RMSEcv: root mean square error of the cross validation of the training model. Chile; ESP: Spain; GEO: Georgia; ITA: Italy.
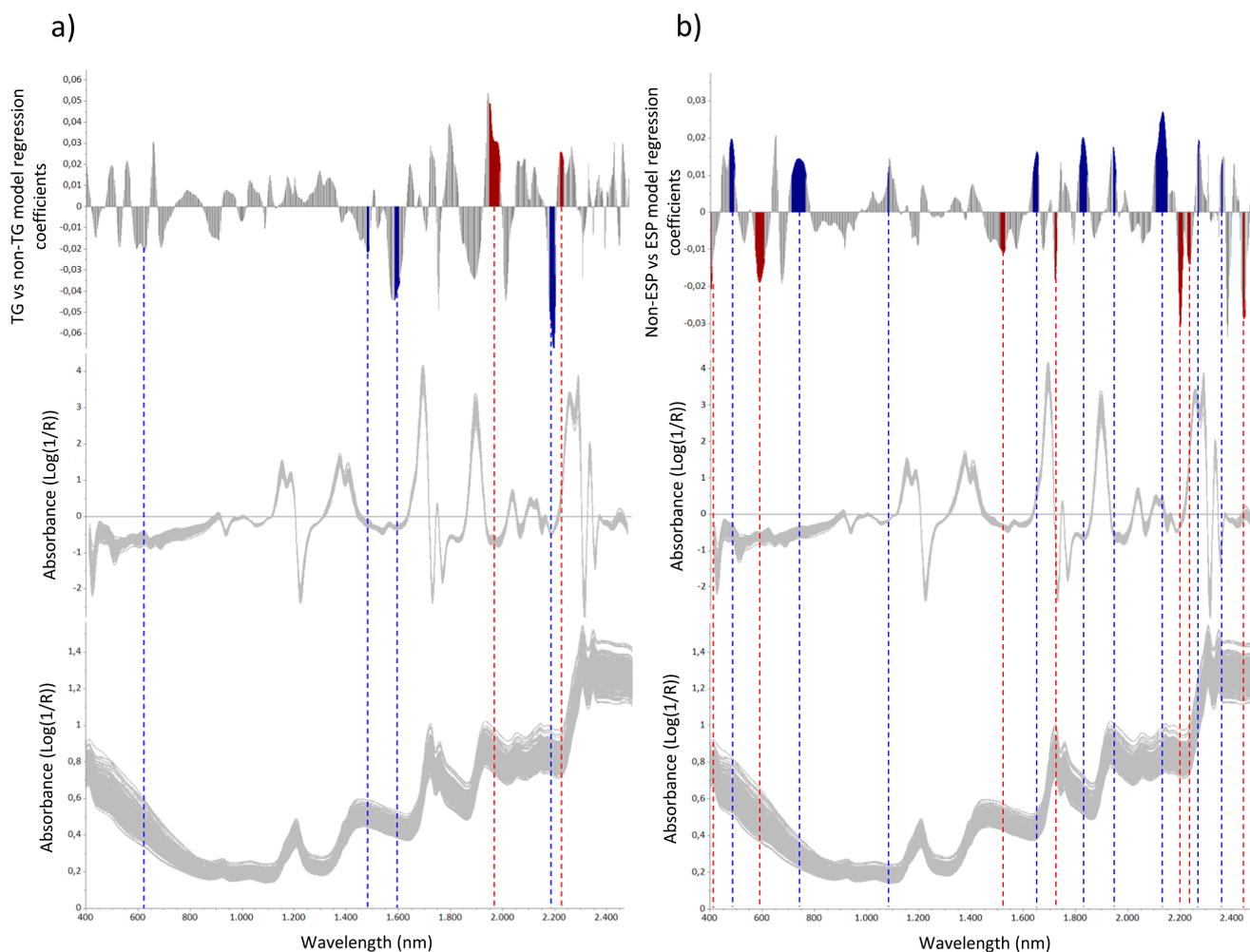
In Fig. 1, the regression coefficients of the NIR PLS-DA models for cultivar (a) and origin (b) are displayed against the pre-processed and the raw spectra. For the cultivar model, the highest regression coefficients of the TG class corresponded to bands around 1482–1490 nm, 1588–1606 nm, and 2180–2220 nm. All these bands were related to protein functional groups. The first two could be assigned to the first

overtone of the N—H stretching of the peptide bond and amino acids side chains, while the last ones belonged to the combination of amide I and amide III bonds [21,38]. The most relevant coefficients for the non-TG class were assigned to the bands around 1952–1990 nm and 2220–2232 nm, which were related to primary amide groups and to the methylene group combination bands of fatty acids, respectively [20,38]. For the origin model, the most relevant coefficients for the ESP class corresponded to the following bands: 722–774 nm, belonging to an O—H stretching overtone, which can originate from alkyl, and primary alcohols, phenols or water [38]; 1640–1658 nm, assigned to the first C—H stretching overtone associated with a secondary alkyl group, which could be related to unsaturated fatty acids [19]; 1816–1844 nm, corresponding to a combination of second O—H/C—O stretching overtones of cellulose [38]; 2110–2150 nm related to combination bands of C—H/C═O stretching associated to lipids [38]; 2266–2270 nm, corresponding to a combination of O—H/C—O stretching of cellulose or to the peptide bond CONH$_2$ in β-sheet structures; and 2476–2482 nm, assigned to the combination bands of the C—H stretching in methyl group of lipid and aliphatic compounds [20,38]. For the non-ESP class, the most significant bands were: 582–615 nm, assigned to the fourth OH stretching overtone in alkyl alcohols [38]; 1516–1532 nm, related to the first NH stretching overtone of the amide group [38]; 1724–1728 nm, corresponding to first overtone of the C—H vibration of triolein [39,40]; 2192–2208 nm, associated to the combination of amide I and amide III

bonds [21,38]; 2226–2238 nm and 2440–2444, both assigned to combination bands of the C—H stretching in methyl group of lipidic compounds [20].

Although some of the main discriminant bands coincided in both cultivar and origin models, such as the band around 2180–2220 nm (a combination band associated to amide bonds) and that around 2220–2240 nm (related to methyl groups in lipid compounds), in general, the most relevant variables for each discriminant model were different. The protein-related bands, together with those associated to lipids, were the most relevant variables to distinguish the TG cultivar. These findings are consistent with previous studies [20,21], which reported that the most influential spectral bands in hazelnut cultivar classification corresponded to protein and lipid compounds.

On the other hand, the most significant bands for discriminating the geographical origin were related to a wider variety of compounds: mainly lipids (1640–1658, 1724–1728, 2110–2150, 2226–2238 and 2440–2482 nm), proteins (1516–1532, 2192–2208 and 2266–2270 nm), complex carbohydrates (1816–1844 and 2266–2270 nm) and a few bands associated to compounds containing hydroxy groups (722–774 nm). These results agreed with previous research [22,23] reporting similar findings in Italian and Georgian hazelnuts and in samples from other origins (France, Germany and Turkey).



**Fig. 1.** Regression coefficients of the PLS-DA models developed on ground samples analysed by NIR benchtop spectrometer; a) cultivar model ('Tonda di Giffoni' TG vs non-TG) b) origin model ('Spain' ESP vs non-ESP). Regression coefficients (above) are plotted against the pre-processed (middle) and raw spectra (below). For each model, the most relevant coefficients for the prediction of the TG and ESP classes are highlighted in blue and those relevant for non-TG, non-ESP are highlighted in red.

### 3.2.2. MIR spectroscopy

Fig. 2 shows the regression coefficients of the MIR PLS-DA cultivar (a) and origin (b) models against the corresponding pre-processed and raw spectra. In the cultivar model, some of the most significant regression coefficients, for both TG and non-TG classes, corresponded to bands in the low-frequency region (600–900 cm$^{-1}$). This region is associated to N—H wagging of primary and secondary amines and amides [41]. The highest regression coefficients of both classes corresponded to the band around 900–1200 cm$^{-1}$, which belonged to the C—O stretching of the ester groups in triacylglycerols [41,42]. For the non-TG class, several relevant regression coefficients corresponded to a broad interval with numerous bands around 1230–1470 cm$^{-1}$ that were assigned to C—H bending of CH$_2$ and CH$_3$ of lipids, and to the bands in the region 1438–1480 cm$^{-1}$, which are characteristic of the unsaponifiable fraction compounds [42–44]. Finally, the band around 1500–1570 cm$^{-1}$, associated with amide II vibrations, arises from mixed N—H bending and C—N stretching vibration in protein structures [24,45].

For the origin model, the most significant coefficients are in the 1000–1750 cm$^{-1}$ range. For the ESP class, the most discriminant bands were: the broad interval with numerous bands associated to CH$_2$ bending of lipids and unsaponifiable fraction compounds (1300–1470 cm$^{-1}$) [43]; the amide II vibration band around 1500–1570 cm$^{-1}$ related to protein structures; the amide I band associated C=O stretching of protein amides (1600–1670 cm$^{-1}$) and the narrow band at 1710–1780 cm$^{-1}$ corresponding to C=O stretching of ester groups of triglycerides [24,45]. Two of these bands, the broad CH$_2$ band around

1300–1470 cm$^{-1}$ and the amide I band (1600–1670 cm$^{-1}$), were also relevant for the discrimination of the non-ESP class. Additionally, the band around 900–1200 cm$^{-1}$ assigned to the C-O stretching of the ester groups in triacylglycerols was particularly significant for the non-ESP class.

In general, for both types of models, the most discriminating variables were associated to protein, lipid and unsaponifiable fraction compounds and were found in the 800–1800 cm$^{-1}$ region of the spectral fingerprint. The absorption pattern in this region is complex but contains valuable information, as it is highly specific for each molecular species. In the cultivar models, the low frequency region of the spectra also appeared to be relevant for discrimination. In contrast, in the origin model, the final section of the fingerprint region, between 1300 and 1700 cm$^{-1}$, was the most significant for discrimination.

Unlike the NIR method, which seemed to find proteins as the most discriminant compounds between TG and non-TG classes, the MIR method relied mainly on lipids, including both triacylglycerols and unsaponifiable fraction compounds, to achieve this discrimination. Regarding the geographical discrimination, the NIR model identified bands related to unsaturated fatty acids, other lipid compounds and complex carbohydrates such as cellulose as very distinctive for the ESP category, whereas in the MIR models, triacylglycerols and unsaponifiable fraction compounds were more relevant in distinguishing between origins.
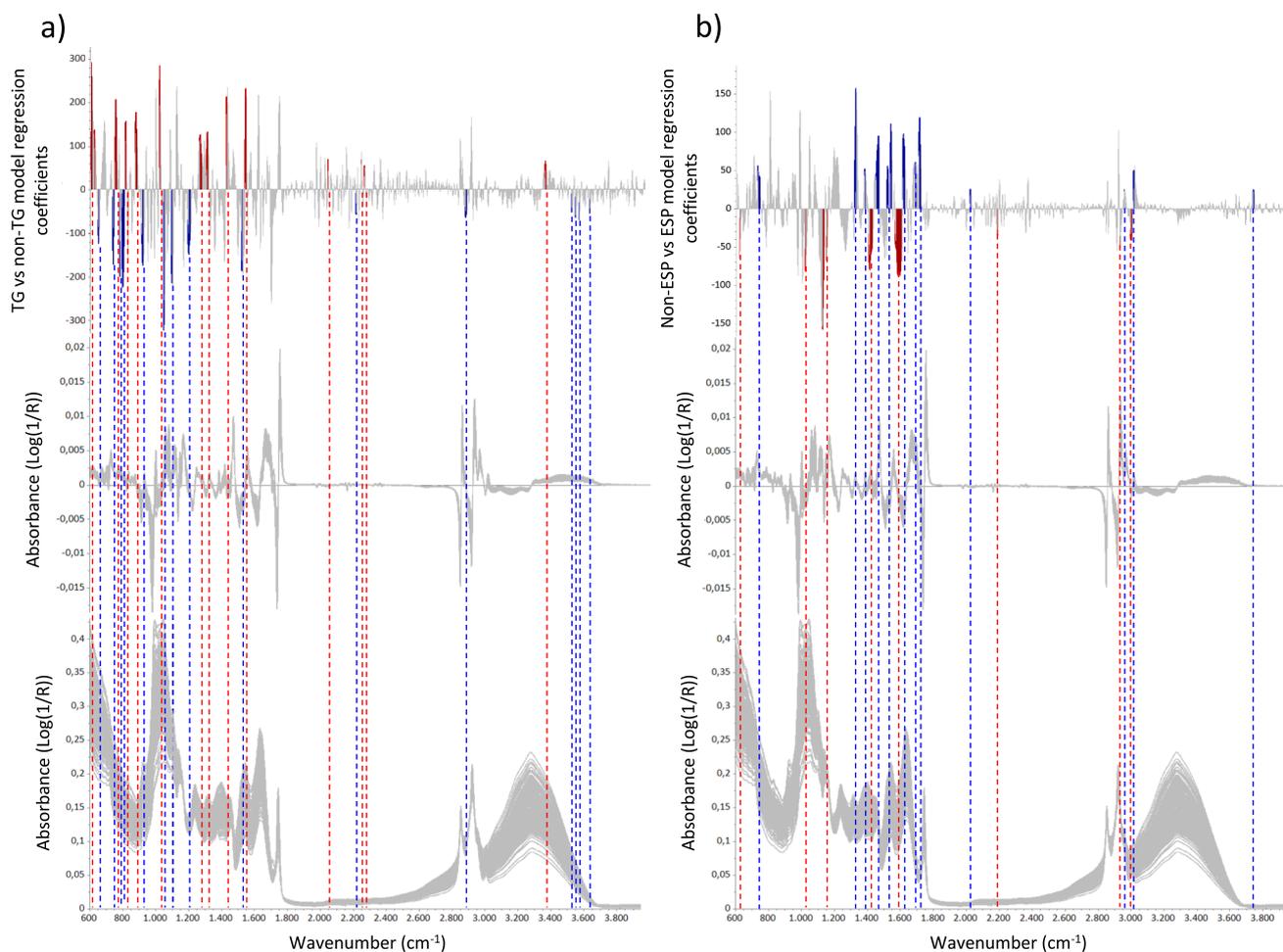


**Fig. 2.** Regression coefficients of the PLS-DA models developed on ground samples analysed by MIR spectrometer; (a) cultivar model ('Tonda di Giffoni' TG vs non-TG) (b) origin model ('Spain' ESP vs non-ESP). Regression coefficients (above) are plotted against the pre-processed (middle) and raw spectra (below). For each model, the most relevant coefficients for the prediction of the TG and ESP classes are highlighted in blue and those relevant for non-TG, non-ESP are highlighted in red.

## 4. Conclusions

Three different spectroscopic methods were tested to authenticate hazelnut cultivar and geographical origin: NIR spectroscopy using a benchtop instrument, NIR spectroscopy using a handheld device, and MIR spectroscopy. The analysis of ground hazelnuts yielded significantly better results than whole kernels, owing to the greater homogeneity, improved sample representativeness, and the prominence of oil content signals. Among the analysis conducted on the ground samples, the benchtop NIR spectrometer demonstrated superior performance, with a sensitivity of 0.92 and a specificity of 0.98 for cultivar models, as well as high correct classification rates for all origins ($\geq$91 %). This resulted in overall correct classification rates of 95 % and 96 %, for cultivar and origin models respectively, closely followed by the MIR method.

Exploring the regression coefficients of the most promising models, based on MIR and NIR applied to ground samples, revealed their reliance on distinct sets of information for discrimination. It was observed that the discrimination of hazelnut cultivar and origin was mainly driven by proteins and lipid composition.

In conclusion, this study allowed for a straightforward comparison of three spectroscopic techniques that offer valuable insights into their performance when applied to exactly the same dataset of hazelnuts from different origins and cultivars. The present work showed that the NIR method could be a fit-for-purpose tool for hazelnut geographical and varietal authentication. However, optimal models need to be further developed and evaluated through extensive datasets, including higher natural heterogeneity of samples, producing regions, main cultivars and multiple harvest years.

## 5. Research data

Torres-Cobos, B., Tres, A., Vichi, S., Guardiola, F., Rovira, M., Romero, A., Baeten, V., & Fernández-Pierna, J.A. (2024). Near Infrared (NIR) and Mid Infrared (MIR) spectra of whole and ground hazelnuts [dataset]. CORA. Repositori de Dades de Recerca, https://doi.org/10.34810/data1725.

## CRediT authorship contribution statement

**B. Torres-Cobos:** Formal analysis, Investigation, Methodology, Validation, Data curation, Visualization, Writing – original draft. **A. Tres:** Conceptualization, Writing – review & editing. **S. Vichi:** Conceptualization, Resources, Project administration, Funding acquisition, Writing – review & editing. **F. Guardiola:** Writing – review & editing. **M. Rovira:** Resources, Writing – review & editing. **A. Romero:** Resources, Writing – review & editing. **V. Baeten:** Conceptualization, Methodology, Supervision, Writing – review & editing. **J.A. Fernández-Pierna:** Conceptualization, Methodology, Supervision, Writing – review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.saa.2024.125367.

## Data availability

Data is published in a repository: Torres-Cobos, B., Tres, A., Vichi, S., Guardiola, F., Rovira, M.,Romero, A., Baeten, V., & Fernández-Pierna, J. A. (2024). Near Infrared(NIR) and Mid Infrared (MIR) spectra of whole and ground hazelnuts[dataset]. CORA. Repositori de Dades de Recerca, https://doi.org/10.34810/data1725.

## References

[1] International Nut and Dried Fruits (INC), Nuts & Dried Fruits Statistical Yearbook 2022/2023, https://inc.nutfruit.org/publications/ (accessed Apr 10, 2024).

[2] V. Cristofori, S. Ferramondo, G. Bertazza, C. Bignami, Nut and kernel traits and chemical composition of hazelnut (*Corylus avellana* L.) cultivars, Sci. Food Agric. 88 (2008) 1091–1098, https://doi.org/10.1002/jsfa.3203.

[3] K. Król, M. Ganter, A. Piotrowska, Morphological traits, kernel composition and sensory evaluation of hazelnut (*Corylus avellana* L.) cultivars grown in Poland, Agronomy 9 (2019) 703, https://doi.org/10.3390/agronomy9110703.

[4] E. Cazzaniga, N. Cavallini, A. Giraudo, G. Gavoci, F. Geobaldo, M. Pariani, D. Ghirardello, G. Zeppa, F. Savorani, Lipids in a nutshell: quick determination of lipid content in hazelnuts with NIR spectroscopy, Foods 12 (2022) 34, https://doi.org/10.3390/foods12010034.

[5] Food and Agriculture Organization of the United Nations (FAO), agricultural data 2022, https://www.fao.org/faostat/es/#data/PP (accessed Mar 23, 2024).

[6] L.F. Ciarmiello, M.F. Mazzeo, P. Minasi, A. Peluso, A. De Luca, P. Piccirillo, R. A. Siciliano, V. Carbone, Analysis of different European hazelnut (*Corylus avellana* L.) cultivars: authentication, phenotypic features, and phenolic profiles, J. Agric. Food Chem. 62 (2014) 6236–6246, https://doi.org/10.1021/jf5018324.

[7] K. Król, M. Gantner, Morphological traits and chemical composition of hazelnut from different geographical origins: a review, Agriculture 10 (2020) 375, https://doi.org/10.3390/agriculture10090375.

[8] C. Lang, N. Weber, M. Möller, L. Schramm, S. Schelm, O. Kohlbacher, M. Fischer, Genetic authentication: differentiation of hazelnut cultivars using polymorphic sites of the chloroplast genome, Food Control 130 (2021) 108344, https://doi.org/10.1016/j.foodcont.2021.108344.

[9] T. Giulia, G. Vallauri, V. Pavese, N. Valentini, P. Rufa, R. Botta, D. Torello-Marinoni, Identification of the hazelnut cultivar in raw kernels and in semi-processed and processed products, Eur. Food Res. Technol. 248 (2022) 2431–2440, https://doi.org/10.1007/s00217-022-04058-z.

[10] S. Klockmann, E. Reiner, R. Bachmann, T. Hackl, M. Fischer, Food fingerprinting: metabolomic approaches for geographical origin discrimination of hazelnuts (*Corylus avellana*) by UPLC-QTOF-MS, J. Agric. Food Chem. 64 (2016) 9253–9262, https://doi.org/10.1021/acs.jafc.6b04433.

[11] S. Ghisoni, L. Lucini, G. Rocchetti, G. Chiodelli, D. Farinelli, S. Tombesi, M. Trevisan, Untargeted metabolomics with multivariate analysis to discriminate hazelnut (*Corylus avellana* L.) cultivars and their geographical origin, J. Sci. Food Agric. 100 (2020) 500–508, https://doi.org/10.1002/jsfa.9998.

[12] B. Torres-Cobos, B. Quintanilla-Casas, M. Rovira, A. Romero, F. Guardiola, S. Vichi, A. Tres, Prospective exploration of hazelnut's unsaponifiable fraction for geographical and varietal authentication: a comparative study of advanced fingerprinting and untargeted profiling techniques, Food Chem. 441 (2024) 138294, https://doi.org/10.1016/j.foodchem.2023.138294.

[13] R. Bachmann, S. Klockmann, J. Haerdter, M. Fischer, T. Hackl, 1H NMR spectroscopy for determination of the geographical origin of hazelnuts, J. Agric. Food Chem. 66 (2018) 11873–11879, https://doi.org/10.1021/acs.jafc.8b03724.

[14] R. Vitale, M. Bevilacqua, R. Bucci, A.D. Magrì, A.L. Magrì, F. Marini, A rapid and non-invasive method for authenticating the origin of pistachio samples by NIR spectroscopy and chemometrics, Chemom. Intell. Lab. Syst. 121 (2013) 90–99, https://doi.org/10.1016/j.chemolab.2012.11.019.

[15] P. Firmani, R. Bucci, F. Marini, A. Biancolillo, Authentication of "Avola almonds" by near infrared (NIR) spectroscopy and chemometrics, J. Food Compos. Anal. 82 (2019) 103235, https://doi.org/10.1016/j.jfca.2019.103235.

[16] L. Amendola, P. Firmani, R. Bucci, F. Marini, A. Biancolillo, Authentication of sorrento walnuts by NIR spectroscopy coupled with different chemometric classification strategies, Appl. Sci. 10 (11) (2020) 4003, https://doi.org/10.3390/app10114003.

[17] H. Zhu, J. Xu, Authentication and provenance of walnut combining Fourier transform mid-infrared spectroscopy with machine learning algorithms, Molecules 25 (21) (2020) 4987, https://doi.org/10.3390/molecules25214987.

[18] M. Arndt, M. Rurik, A. Drees, C. Ahlers, S. Feldmann, O. Kohlbacher, M. Fischer, Food authentication: determination of the geographical origin of almonds (*Prunus dulcis* Mill.) via near-infrared spectroscopy, Microchem. J. 160 (B) (2021) 105702, https://doi.org/10.1016/j.microc.2020.105702.

[19] R. Moscetti, E. Radicetti, D. Monarca, M. Cecchini, R. Massantini, Near infrared spectroscopy is suitable for the classification of hazelnuts according to Protected Designation of Origin, J. Sci. Food Agric. 95 (13) (2015) 2619–2625, https://doi.org/10.1002/jsfa.6992.

[20] A. Biancolillo, S. De Luca, S. Bassi, L. Roudier, R. Bucci, A.D. Magrì, F. Marini, Authentication of an Italian PDO hazelnut ("Nocciola Romana") by NIR spectroscopy, Environ. Sci. Pollut. Res. 25 (2018) 28780–28786, https://doi.org/10.1007/s11356-018-1755-2.

[21] H. Ayvaz, R. Temizkan, H.E. Genis, M. Mortas, D.O. Genis, M.A. Dogan, B. A. Nazlim, Rapid discrimination of Turkish commercial hazelnut (*Corylus avellana* L.) varieties using Near-Infrared Spectroscopy and chemometrics, Vib. Spectrosc. 119 (2022) 103353, https://doi.org/10.1016/j.vibspec.2022.103353.

[22] N. Shakiba, A. Gerdes, N. Holz, S. Wenck, R. Bachmann, T. Schneider, S. Seifert, M. Fischer, T. Hackl, Determination of the geographical origin of hazelnuts (*Corylus avellana* L.) by Near-Infrared spectroscopy (NIR) and a Low-Level Fusion with nuclear magnetic resonance (NMR), Microchem. J. 174 (2022) 107066, https://doi.org/10.1016/j.microc.2021.107066.

[23] G. Sammarco, C. Dall'Asta, M. Suman, Near infrared spectroscopy and multivariate statistical analysis as rapid tools for the geographical origin assessment of Italian hazelnuts, Vib. Spectrosc. 126 (2023) 103531, https://doi.org/10.1016/j.vibspec.2023.103531.

[24] M. Manfredi, E. Robotti, F. Quasso, E. Mazzucco, G. Calabrese, E. Marengo, Fast classification of hazelnut cultivars through portable infrared spectroscopy and chemometrics, Spectrochim. Acta Part A: Mol. Biomol. Spectrosc. 189 (2018) 427–435, https://doi.org/10.1016/j.saa.2017.08.050.

[25] L.E. Rodriguez-Saona, M.E. Allendorf, Use of FTIR for rapid authentication and detection of adulteration of food, Annu. Rev. Food Sci. Technol. 2 (2011) 467–483, https://doi.org/10.1146/annurev-food-022510-133750.

[26] V. Bellon-Maurel, A. McBratney, Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils – critical review and research perspectives, Soil Biol. Biochem. 43 (7) (2011) 1398–1410, https://doi.org/10.1016/j.soilbio.2011.02.019.

[27] G.A. de Oliveira, F. de Castilhos, C.M.G.C. Renard, S. Bureau, Comparison of NIR and MIR spectroscopic methods for determination of individual sugars, organic acids and carotenoids in passion fruit, Food Res. Int. 60 (2014) 154–162, https://doi.org/10.1016/j.foodres.2013.10.051.

[28] J. Nogales-Bueno, L. Feliz, B. Baca-Bocanegra, J.M. Hernández-Hierro, F. J. Heredia, J.M. Barroso, A.E. Rato, Comparative study on the use of three different near infrared spectroscopy recording methodologies for varietal discrimination of walnuts, Talanta 206 (2020) 120189, https://doi.org/10.1016/j.talanta.2019.120189.

[29] C. McVey, T.F. McGrath, S.A. Haughey, C.T. Elliott, A rapid food chain approach for authenticity screening: the development, validation and transferability of a chemometric model using two handheld near infrared spectroscopy (NIRS) devices, Talanta 222 (2021) 121533, https://doi.org/10.1016/j.talanta.2020.121533.

[30] Y. Pu, D. Pérez-Marín, N. O'Shea, A. Garrido-Varo, Recent advances in portable and handheld NIR spectrometers and applications in milk, cheese and dairy powders, Foods 10 (2021) 2377, https://doi.org/10.3390/foods10102377.

[31] J. Müller-Maatsch, S.M. van Ruth, Handheld devices for food authentication and their applications: a review, Foods 10 (2021) 2901, https://doi.org/10.3390/foods10122901.

[32] G.S. Folli, L.P. Santos, F.D. Santos, P.H.P. Cunha, I.F. Schaffel, F.T. Borghi, I.H.A. S. Barros, A.A. Pires, A.V.F.N. Ribeiro, W. Romão, P.R. Filgueiras, Food analysis by portable NIR spectrometer, Food Chem. Adv. 1 (2022) 100074, https://doi.org/10.1016/j.focha.2022.100074.

[33] D. Sorak, L. Herberholz, S. Iwascek, S. Altinpinar, F. Pfeifer, H.W. Siesler, New developments and applications of handheld Raman, mid-Infrared, and near-Infrared spectrometers, Appl. Spectrosc. Rev. 47 (2) (2012) 83–115, https://doi.org/10.1080/05704928.2011.625748.

[34] S.N. Thennadil, M. Dewar, C. Herdsman, A. Nordon, E. Becker, Automated weighted outlier detection technique for multivariate data, Control Eng. Pract. 70 (2018) 40–49, https://doi.org/10.1016/j.conengprac.2017.09.018.

[35] L. Eriksson, J. Trygg, S. Wold, CV-ANOVA for significance testing of PLS and OPLS® models, J. Chemom. 22 (2008) 594–600, https://doi.org/10.1002/cem.1187.

[36] B. Magnusson, U. Örnemark, Eurachem guide: The Fitness for Purpose of Analytical Methods - A Laboratory Guide to Method Validation and Related Topics, second edition, Eurachem, Belgium, 2014.

[37] M. Arndt, M. Rurik, A. Drees, K. Bigdowski, O. Kohlbacher, M. Fischer, Comparison of different sample preparation techniques for NIR screening and their influence on the geographical origin determination of almonds (*Prunus dulcis* MILL.), Food Control 115 (2020) 107302, https://doi.org/10.1016/j.foodcont.2020.107302.

[38] J. Workman, L. Weyer, Practical Guide to Interpretive Near-infrared Spectroscopy, first edition, CRC Press, Boca Raton, Florida, 2007, doi: 10.1201/9781420018318.

[39] T. Sato, S. Kawano, M. Iwamoto, Near infrared spectral patterns of fatty acid analysis from fats and oils, J. Am. Oil Chem. Soc. 68 (1991) 827–833, https://doi.org/10.1007/BF02660596.

[40] V. Baeten, R. Aparicio, N. Marigheto, R. Wilson, Olive oil analysis by infrared and Raman spectroscopy: methodologies and applications, in: J. Harwood, R. Aparicio (Eds.), Handbook of Olive Oil: Analysis and Properties, Springer, Boston, MA, 2000, pp. 209–248, https://doi.org/10.1007/978-1-4757-5371-4_8.

[41] R.M. Silverstein, F.X. Webster, Spectrometric Identification of Organic Compounds, sixth edition, John Wiley & Sons, New Jersey, 1998, ISBN978-0471134572.

[42] S. Ng, O. Lasekan, K. Muhammad, R. Sulaiman, N. Hussain, Effect of roasting conditions on color development and Fourier transform infrared spectroscopy (FTIR-ATR) analysis of Malaysian-grown tropical almond nuts (*Terminalia catappa* L.), Chem. Cent. J. 8 (2014) 55, https://doi.org/10.1186/s13065-014-0055-2.

[43] V. Baeten, J.A. Fernández Pierna, P. Dardenne, M. Meurens, D.L. García-González, R. Aparicio-Ruiz, Detection of the presence of hazelnut oil in olive oil by FT-Raman and FT-MIR spectroscopy, J. Agric. Food Chem. 53 (16) (2005) 6201–6206, https://doi.org/10.1021/jf050595n.

[44] A. Valdés García, A. Beltrán-Sanahuja, M.C. Garrigós-Selva, Characterization and classification of almond cultivars by using spectroscopic and thermal techniques, J. Food Sci. 78 (2) (2013) 138–144, https://doi.org/10.1111/1750-3841.12031.

[45] A. Dogan, G. Siyakus, F. Severcan, FTIR spectroscopic characterization of irradiated hazelnut (*Corylus avellana* L.), Food Chem. 100 (2007) 1106–1114, https://doi.org/10.1016/j.foodchem.2005.11.017.