A comparison between NIR and ATR-FTIR spectroscopy for varietal discrimination of Spanish intact almonds

Victoria Cortés[1], José Manuel Barat[1], Pau Talens[1], José Blasco[2], María Jesús Lerma-García[1,*]

[1]*Departamento de Tecnología de los Alimentos, Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain*

[2]*Centro de Agroingeniería. Instituto Valenciano de Investigaciones Agrarias (IVIA), Ctra. CV-315, km. 10,7, 46113, Moncada, Valencia, Spain*

*Corresponding author:

María Jesús Lerma-García; e-mail: malerga1@tal.upv.es

Abbreviated running title: NIR and ATR-FTIR spectroscopy for almond varietal discrimination

ABSTRACT

The rapid and easy discrimination of almond varieties with similar morphology, different quality properties and in most cases different prices is interesting to protect both almond industry and consumers from fraud. Therefore, in this work, intact almond kernels coming from four Spanish varieties (‘Guara’, ‘Rumbeta’, ‘Marcona’ and ‘Planeta’) were analysed using both near infrared (NIR) and attenuated total reflectance Fourier-transform infrared (ATR-FTIR) spectroscopy. After spectra measurement, an attempt to classify almonds according to their variety was tried using two classification methods (partial least square-discriminant analysis (PLS-DA) and quadratic discriminant analysis (QDA)) applied to both NIR and ATR-FTIR pre-treated spectral data. An overall accuracy of 94.45% was obtained with both PLS-DA of ATR-FTIR and QDA of NIR data. These results confirm that both spectroscopic techniques, if the optimal statistical model is selected, are powerful tools to reliable discriminate almonds according to their varieties.

Keywords

NIR

ATR-FTIR

Intact almond

Varietal discrimination

PLS-DA

QDA

1. Introduction

Almond (Prunus dulcis (Mill.) D.A.Webb) is one of the main nut tree crops in terms of commercial production around the world (FAOSTAT, 2012). Spain is the second largest almond world producer after the US (López-Ortiz, Prats-Moya, Sanahuja, Maestre-Pérez, Grané-Teruel, & Martín-Carratalá, 2008), being almond trees very extended due to Spanish mild weather conditions that favour its cultivation (Vázquez-Araújo, Enguix, Verdú, García-García, & Carbonell-Barrachina, 2008). In addition, Spain is also an important consumer country, in which almonds are consumed raw, roasted, fried or as an important ingredient in different foodstuffs like ice creams or sweets as “turrón”, among others. Almond quality covers different features, such as kernel and shell physical aspect and kernel organoleptic characteristics and composition (with its different protein, lipid and sugar contents, among others). All these characteristics are influenced by almond variety, which could define the industrial use of each one of them (Cordeiro, Oliveira, Ventura, & Monteiro, 2001). There are several almond varieties grown in Spain. Among them, the Marcona variety is the principal one (Varela, Chen, Fiszman, & Povey, 2006), which is mainly consumed as roasted or fried snack or as the main ingredient of the Protected Designations of Origin of the “turrones” Jijona and Alicante. However, Marcona variety is very expensive due to their excellent organoleptic properties and low production rate (Vázquez-Araújo et al., 2008). Another variety to be highlighted is Guara, which has experimented and important commercial triumph due to its late-flowering, self-compatibility and high quality (Kodad, Estopañán, Juan, Alonso, & Espiau, 2014). Other important Spanish varieties due to their large production volume are ‘Largueta’, ‘Planeta’, ‘Rumbeta’ or ‘Desmayo’, among others. Therefore, it is important to find analytical methodologies able to discriminate almond varieties with similar morphology or with lower prices in order to protect both almond industry and consumers from fraud.

There are some studies published in literature that cover almond variety discrimination (Gil Solsona, Boix, Ibáñez, Sancho, 2017; Piscopo, Romeo, Petrovicova, & Poiana, 2010; Prats-Moya, Grané-Teruel, Berenguer-Navarro, & Martín-Carratalá, 1997; García-López, Grané-Teruel, Berenguer-Navarro, García-García, & Martín-Carratalá, 1996), or in which almond components or physical characteristics from different varieties have been established and compared (Oliveira, Meyer, Afonso, Ribeiro, & Gonçalves, 2018; Zamany, Samadi, Kim, Keum, & Saini, 2017; Yada, Lapsley, & Huang, 2011; Valdés, Vidal, Beltrán, Canals, & Garrigós, 2015, Kodad et al., 2014; Özcan, Ünver, Erkan, & Arslan, 2011; López-Ortiz et al., 2008; Cherif, Sebei, Boukhchina, Kallel, Belkacemi, & Arul, 2004; Cordeiro et al., 2001); however, the analytical techniques employed are in most cases expensive, destructive and time-consuming, and sample pre-treatment is normally required. Therefore, there is a need of non-destructive and fast alternative methodologies able to cover this issue. In this regard, the employment of spectroscopic techniques, such as infrared (IR) spectroscopy, could be an excellent alternative. The potential of this technique in both, near and medium IR regions, has been demonstrated in several previous works in the almond field. For example, Fourier-transform infrared spectroscopy (FTIR) has been applied to quality control of medicinal almonds (Chun-Song et al., 2017), while near infrared spectroscopy (NIR) has been used to detect hidden damage in raw almonds (Rogel-Castillo, Boulton, Opastpongkarn, Huang, & Mitchell, 2016), to inspect internal damages in almonds (Nakariyakul, 2014), to discriminate sweet and bitter almonds (Borrás, Amigo, van den Berg, Boqué, Busto, 2014; Cortés, Talens, Barat, & Lerma-García, 2018), and to detect fungal infection in almond kernels (Liang, Slaughter, Ortega-Beltran, & Michailides, 2015), among others. We have only found three articles regarding almond discrimination according to their variety using IR data. Two of these articles were from a research group of the University of Alicante (Beltrán Sanahuja, Prats Moya, Maestre Pérez, Grané Teruel, Martín Carratalá, 2009; Beltrán, Ramos, Grané, Martín, & Garrigós, 2011), in which almond varieties were discriminated after almond oil extraction according to its thermal stability after application of a forced oxidative treatment. For this purpose, oil degradation was studied by registering the changes produced in the most abundant fatty acids (established by gas chromatography (GC)) (Beltrán Sanahuja et al., 2009) or volatile compounds (established by headspace solid-phase microextraction/GC–mass spectrometry (HS-SPME/GC–MS) (Beltrán et al., 2011) and to changes produced in the FTIR spectra (Beltrán Sanahuja et al., 2009; Beltrán et al., 2011). Using stepwise linear discriminant analysis (LDA), authors were able to classify almond varieties using fatty acid contents and FTIR data in the first work (Beltrán Sanahuja et al., 2009), and using HS-SPME/GC–MS data in the second one (Beltrán et al., 2011).In the third article, Valdés et al. (Valdés, Beltrán, & Garrigós, 2013) employed FTIR and two thermal analysis techniques (differential scanning calorimetry and thermogravimetric analysis) to classify almonds according to their cultivar, after almond grounding and sieving. Next, LDA models were constructed using FTIR and thermal data all together and separately. With these models, good almond classifications according to their variety were obtained. However, an as far as were are concern, any article has been published regarding the employment and comparison of both NIR and FTIR data to classify almonds according to their variety by directly measuring spectra on intact almonds surface.

Therefore, the aim of this work was to explore the viability of both NIR and FTIR data to reliable classify Spanish almonds according to their variety. For this purpose, almonds belonging to four of the main varieties cultivated in Spain (‘Guara’, ‘Rumbeta’, ‘Marcona’ and ‘Planeta’) were directly measured on both spectrometers. Using both, NIR and FTIR data, two different classification methods (partial least square discriminant analysis (PLS-DA) and quadratic discriminant analysis (QDA)) were constructed and their overall accuracies compared.

2. Materials and methods

2.1. Raw material

A total of 120 almonds, coming from four different Spanish varieties (‘Guara’ (G), ‘Rumbeta’ (R), ‘’Marcona (M) and ‘Planeta’ (P)), were analysed in this study. All samples, gently provided by Agricoop (Alicante, Spain), were free of visual damage and of uniform colour and size.

2.2. Spectra acquisition

2.2.1. NIR

An AvaSpec-NIR256-1.7 NIRLine spectrometer (AVS-DESKTOP-USB2, Avantes BV, The Netherlands) was used for collecting NIR spectra of intact almond kernel (with skin) over the range of 1000–1700 nm at an interval of 3.535 nm. The instrument is equipped with a 10-W tungsten halogen light source (AvaLight-HAL-S, Avantes BV, The Netherlands). Almond spectra were acquired in diffuse reflectance mode using a bi-directional fibre-optic probe (FCR-7IR200-2-45-ME, Avantes BV, The

Netherlands) designed under an angle of 45º to prevent direct back-reflection from almond surface. The probe, composed by two legs, is connected to the light source and to the spectrometer. The integration time (500 ms) was adjusted using a 99% reflective white reference (WS-2, Avantes BV, The Netherlands), so that the maximum reflectance value was over 90% of saturation (Lorente, Escandell-Montero, Cubero, Gómez-Sanchis, & Blasco, 2015). The dark spectrum was obtained by turning off the light source and covering the tip of the reflectance probe.

A personal computer equipped with the commercial software AvaSoft version 7.2 (Avantes, Inc.) was used to acquire the spectra. For each sample, five replicates were collected on both almond sides and mean spectra values were used for the analysis. All measurements have been performed at room temperature (22±1 °C).

2.2.2. ATR-FTIR

ATR-FTIR spectra were obtained using a Tensor 27 spectrometer (Bruker Optics, Milan, Italy) coupled to a deuterated triglycine sulphate (DTGS) detector and to an ATR accessory (Specac Inc., Woodstock, Georgia, USA) composed of a zinc selenide (ZnSe) crystal. Absorbance spectra were obtained in the wavenumber range from 4000 to 600 cm^{-1} acquiring 32 scans per sample at a resolution of 4 cm^{-1}. After every scan, a new reference air background spectrum was taken. Each intact almond kernel was put on the ZnSe crystal for measurements, and the crystal was carefully cleaned by scrubbing with acetone and dried with a soft tissue before measuring the next sample. The system was operated using the OPUS software version 5.0 provided by Bruker Optics. Two measurements were acquired for each almond (one measurement on each almond face), being spectra mean employed for statistical analysis.

2.3. Data pre-processing and multivariate analysis

To execute the pre-treatments and multivariate procedures, 'The Unscrambler X' software version 10.3 (Camo Process SA, Trondheim, Norway) was used.

Before multivariate analysis, the dispersion of almond NIR spectra was corrected by simultaneously applying Savitzky-Golay (S-G) smoothing (3 points gap), extended multiplicative scatter correction (EMSC) and second derivative (with a 2.3 gap-segment). In the case of ATR-FTIR, standard normal variate (SNV) and S-G second derivate (with a second order polynomial) spectral pre-treatments were applied.

After spectra pre-treatments, principal component analysis (PCA) models were constructed to obtain qualitative information about the possible varietal discrimination and to identify possible outliers.

In order to construct the chemometric models for both, NIR and ATR-FTIR data, the full sample set (N= 120) was divided into training (70% of almonds) and test sets (remaining 30% of almonds). Once the models were constructed, and before external validation with the test set, model were internally validated using full cross-validation (CV; leave-one-out method) (Huang, Yu, Xu, & Ying, 2008).

Two classification models (PLS-DA and QDA) to differentiate almond varieties were constructed with both NIR and ATR-FTIR data. The PLS-DA models were constructed using the PLS algorithms (Wold, Sjöström, & Eriksson, 2011), where the variables in the *X*-matrix (which corresponded to the spectral data) were related to the classes included in the *Y*-matrix. This matrix contained dummy variables that describe the belonging of each training set sample to a given category. The *Y*- or dummy-matrix is composed by 4 columns (one column for each variety) with ones and zeros, such that the entry in the first column is unity and the entry of the rest of the columns is zero for the samples of the first variety, and so on until completing the 4 columns. Almond classification according to their variety was performed using the 0.5 cutoff value (Cortés, Ortiz, Aleixos, Blasco, Cubero, & Talens, 2016). Predicted values higher than 0.5 indicated that the sample belongs to a given class, while values lower than 0.5 indicated that the samples does not belong to this category.

PLS-DA models accuracy was evaluated by the number of latent variables (LVs), the coefficient of determination of calibration (R^2_C), the root mean square error of calibration (RMSEC), the coefficient of determination for cross-validation (R^2_{CV}) and the root-mean square error of cross-validation (RMSECV).

For QDA models, a categorical value (Y-variable) was assigned with a different letter (G, M, P and R) for each variety. To construct QDA models, a number of variables lower than the number of objects is required (Sádecká, Jakubíková, Májek, & Kleinová, 2016). Then, a variable reduction is needed before model construction. This variable reduction is performed using PCA scores, since principal components (PCs) are found as linear transformations that are uncorrelated (Rodriguez-Campos, Escalona-Buendía, Orozco-Avila, Lugo-Cervantes, & Jaramillo-Flores, 2011).

Finally, PLS-DA and QDA models performance was evaluated by considering the percentage of correctly classified test samples.

3. Results and discussion

3.1. Characteristics of NIR and ATR-FTIR almond spectra

Fig. 1 represents the typical raw and pre-processed (a) NIR and (b) ATR-FTIR almonds spectra. The main absorbance bands in the NIR spectra (Fig. 1a) were evidenced at 1120, 1200 and 1440 nm. These bands are representative of the chemical or functional groups of components present in the almonds. The 1120 and 1200 nm bands denote absorptions that may occur due to the second overtone vibration of C-H stretching, while the band at 1440 nm may belong to the first overtone of O-H stretching of water (Workman Jr, & Weyer, 2008).

Fig. 1b represents the almond ATR-FTIR spectra showing the major peaks at 2940, 2460, 2350, 2220, 1860, 1750, 1390, 1220 and 1040 cm^{-1}. Absorbance at 2940 cm^{-1} is due to the asymmetric bands arising from CH_2 stretching vibrations (Sinelli, Cosio, Gigliotti, & Casiraghi, 2007), whereas the peaks at 2460, 2350 and 2220 cm^{-1} could be assigned to alkane stretching (Kök, Varfolomeev, & Nurgaliev, 2017). The two absorption peaks at 1860 and 1750 cm^{-1} are the characteristic peaks of the C=O stretching vibrations (Beltrán Sanahuja et al., 2009; Vlachos, Skopelitis, Psaroudaki, Konstantinidou, Chatzilazarou, & Tegou, 2006; Zhang, Guo, & Zhang, 2002). The peak at 1390 cm^{-1} may be due to CH bending (Hernández, & Zacconi, 2009), while the peak at 1220 cm^{-1} could be associated with the C-O stretching vibration (Paradkar, Sakhamuri, & Irudayaraj, 2002). Finally, the peak at 1040 cm^{-1} may be due to combination of vibrations of C(1)H bending (that is C-H bond at C1 position) of carbohydrates (Paradkar et al., 2002).

3.2. PCA analysis

Both NIR and ATR-FTIR spectra were pre-processed before PCA model construction. A preliminary data exploration with PCA was carried out with the training set samples. As observed in the PCA score plots (Fig. 2a,b), an evident separation of almonds according to the different varieties is observed with both NIR and ATR-FTIR data. The two first PCs summarized 76% and 97% accumulative contribution of the original data for NIR and ATR-FTIR data, respectively, which means that nearly all the variation of the variables were explained by these PCs. Next, the X-loading plots (Fig. 2c,d) were analysed to evidence which variables showed the greatest separation among almond varieties. As observed in PC1 and PC2 X-loading plots for the NIR data (Fig. 2c), the most prominent peaks were observed at 1150 nm (second overtone vibration of C-H stretching) (Workman Jr et al., 2008), 1490 and 1520 nm (O–H bond stretching and first water overtone) (Blanco, Coello, Iturriaga, Maspoch, & Pages, 2000), 1570 nm (N–H first overtone) (Kaddour, Mondet, & Cuq, 2008) and 1610 nm (related to carbohydrate content) (Teena, Manickavasagan, Ravikanth, & Jayas, 2014), while for the ATR-FTIR data the most relevant peaks were those located at 2350 (alkane stretching) and 1750 cm^{-1} (C=O stretching vibrations) (Kök et al., 2017; Zhang et al., 2002).

3.3. Classification of almonds according to their variety

Two different classification techniques (PLS-DA and QDA) were applied to both NIR and ATR-FTIR pre-processed spectra in order to discriminate almonds according to their variety.

The PLS-DA models were constructed using 7 and 14 LVs for NIR and ATR-FTIR spectra, respectively. The accuracy of the PLS-DA models obtained using both NIR and ATR-FTIR pre-treated data with the training set samples is included in Table 1. As it can be observed in this table, both spectroscopic techniques provided similar and good results, with R^2_{CV} and RMSECV values comprised between 0.85-0.92 and 0.12-0.18, respectively. When these models were validated with the test set samples, satisfactory classification rates were obtained (see Table 2). The best PLS-DA model which produce the highest overall rate of correct classification was obtained using ATR-FTIR data, with a 94.45% of correctly classified almonds, being this value lower (86.13% of overall accuracy) for the model constructed with NIR data. The same results are confirmed in Fig. 3.

Next, QDA models using both spectroscopic techniques data were constructed using the first 9 PCs. An overall rate of 100% and 96% of correct classified samples of the training set samples were obtained using NIR and ATR-FTIR data, respectively. The results obtained for the test set samples are shown in Table 2. As it can be observed in this table for NIR data, the almonds coming from 'Guara' and 'Rumbeta' varietieswere both 100% correctly classified, while the samples of 'Marcona' and 'Planeta' varitieswere both 88.9% correctly classified. In the case of ATR-FTIR data, the overall accuracy classification is lower (77.8%) than those obtained using NIR (94.45%). Concretely, the samples of 'Planeta' and 'Rumbeta' varieties were both 88.9% correctly classified, while samples of 'Marcona' and 'Guara' varieties provided a 77.8% and 55.6% correctly classified samples, respectively. The QDA plots obtained with both NIR and ATR-FTIR data are shown in Fig. 4. The same results of Table 2 are also evidenced in this figure, where there is a good classification of samples into their corresponding category for the QDA model constructed with NIR data (Fig. 4a). On the other hand, the QDA model constructed with the ATR-FTIR data (Fig. 4b) evidenced several misclassified samples.

Finally, when PLS-DA and QDA models obtained using NIR and ATR-FTIR data were compared, it is possible to conclude that the best results in terms of overall performance were obtained using PLS-DA of ATR-FTIR and with QDA of NIR data. Therefore, these results confirm that both spectroscopic techniques, if the optimal statistical model is selected, are useful for almond varietal discrimination.

4. Conclusions

The results obtained by the two classification methods (PLS-DA and QDA) applied to both NIR and ATR-FTIR pre-processed data demonstrated that, when the optimal classification method was applied, it is possible to correctly discriminate Spanish almonds according to their variety. Concretely, the best overall accuracies (94.45%) were obtained with the PLS-DA model of ATR-FTIR and the QDA model of NIR data. Therefore, both spectroscopic techniques could be successfully applied for the rapid and non-destructive varietal classification of intact almonds. The developed methodology could be very useful to protect both almond industry and consumers from fraud, since the almond varieties studied are from similar appearance and cover different price ranges in the market.

Acknowledgements

Victoria Cortés López thanks the Spanish Ministry of Education, Culture and Sports for the FPU grant (FPU13/04202). The authors wish to thank the cooperative Agricoop for kindly providing the samples used in the experiments. This work was partially funded by INIA and FEDER funds through project RTA2015-00078-00-00.

References

Beltrán Sanahuja, A., Prats Moya, M.S., Maestre Pérez, S. M., Grané Teruel, N., & Martín Carratalá, M.L. (2009). Classification of four almond cultivars using oil degradation parameters based on FTIR and GC data. *Journal of the American Oil Chemists' Society, 86,* 51–58.

Beltrán, A., Ramos, M., Grané, N., Martín, M. L., & Garrigós, M. C. (2011). Monitoring the oxidation of almond oils by HS-SPME–GC–MS and ATR-FTIR: Application of volatile compounds determination to cultivar authenticity. *Food Chemistry, 126(2),* 603–609.

Blanco, M., Coello, J., Iturriaga, H., Maspoch, S., & Pages, J. (2000). NIR calibration in non-linear systems: different PLS approaches and artificial neural networks. *Chemometrics and Intelligent Laboratory Systems, 50(1),* 75–82.

Borrás, E., Amigo, J. M., van den Berg, F., Boqué, R., & Busto, O. (2014). Fast and robust discrimination of almonds (*Prunus amygdalus*) with respect to their bitterness by using near infrared and partial least squares-discriminant analysis. *Food Chemistry, 153,* 15–19.

Cherif, A., Sebei, K., Boukhchina, S., Kallel, H., Belkacemi, K., & Arul, J. (2004). Kernel fatty acid and triacylglycerol composition for three almond cultivars during maturation. *Journal of the American Oil Chemists' Society, 81(10),* 901–905.

Chun-Song, C., Can-Jian, W., Liang, J., Chi-Chou, L., Hua, Z., Zhin-Feng, & Zhang, Z. F. (2017). A new approach for identification of medicinal almonds by fourier transform infrared spectroscopy and systematic clustering of characteristic peaks. *Chinese Journal of Natural Medicines, 15,* 703–709.

Cordeiro, V., Oliveira, M., Ventura, J., & Monteiro, A. (2001). Study of some physical characters and nutritive composition of the Portuguese's (local) almond varieties. *Options Méditerranéennes, 56,* 331–337.

Cortés, V., Ortiz, C., Aleixos, N., Blasco, J., Cubero, S., & Talens, P. (2016). A new internal quality index for mango and its prediction by external visible and near-infrared reflection spectroscopy. *Postharvest Biology and Technology, 118,* 148–158.

Cortés, V., Talens, P., Barat, J. M., & Lerma-García, M. J. (2018). Potential of NIR spectroscopy to predict amygdalin content established by HPLC in intact almonds and classification based on almond bitterness. *Food Control*, *91*, 68–75.

FAOSTAT (2012). FAO Statistical Databases. Food and Agriculture Organisation of the United Nations, Statistics Division, Available at (http://faostat3.fao.org).

García-López, C., Grané-Teruel, N., Berenguer-Navarro, V., García-García, J. E., & Martín-Carratalá, M. L. (1996). Major fatty acid composition of 19 almond cultivars of different origins. A chemometric approach. *Journal of Agricultural and Food Chemistry*, *44(7),* 1751–1755.

Gil Solsona, R., Boix, C., Ibáñez, M., & Sancho, J. V. (2017). The classification of almonds (Prunus dulcis) by country and variety using UHPLC-HRMS-based untargeted metabolomics. *Food Additives & Contaminants: Part A*, https://doi.org/10.1080/19440049.2017.1416679.

Hernández, S. A., & Zacconi, F. (2009). Aceite de almendras dulces: Extracción, caracterización y aplicación. *Quimica Nova, 32(5),* 1342–1345.

Huang, H., Yu, H., Xu, H., & Ying, Y. (2008). Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: a review. *Journal of Food Engineering, 87(3),* 303–313.

Kaddour, A. A., Mondet, M., & Cuq, B. (2008). Application of two-dimensional cross-correlation spectroscopy to analyse infrared (MIR and NIR) spectra recorded during bread dough mixing. *Journal of Cereal Science, 48(3),* 678–685.

Kodad, O., Estopañán, G., Juan, T., Alonso, J. M., & Espiau, M. T. (2014). Oil content, fatty acid composition and tocopherol concentration in the Spanish almond genebank collection. *Scientia Horticulturae, 177,* 99–107.

Kök, M. V., Varfolomeev, M. A., & Nurgaliev, D. K. (2017). Thermal characterization of crude oils in the presence of limestone matrix by TGA-DTG-FTIR. *Journal of Petroleum Science and Engineering, 154,* 495–501.

Liang, P.S., Slaughter, D.C., Ortega-Beltran, A., & Michailides, T.J. (2015). Detection of fungal infection in almond kernels using near-infrared reflectance spectroscopy. *Biosystems Engineering, 137*, 64–72.

López-Ortiz, C. M., Prats-Moya, S., Sanahuja, A. B., Maestre-Pérez, S. E., Grané-Teruel, N., & Martín-Carratalá, M. L. (2008). Comparative study of tocopherol homologue content in four almond oil cultivars during two consecutive years. *Journal of Food Composition and Analysis, 21(2),* 144–151.

Lorente, D., Escandell-Montero, P., Cubero, S., Gómez-Sanchis, J., & Blasco, J. (2015). Visible-NIR reflectance spectroscopy and manifold learning methods applied to the detection of fungal infections on citrus fruit. *Journal of Food Engineering, 163*, 17–21.

Nakariyakul, S. (2014). Internal damage inspection of almond nuts using optimal near-infrared waveband selection technique. *Journal of Food Engineering, 126*, 173–177.

Oliveira, I., Meyer, A., Afonso, S., Ribeiro, C., & Gonçalves, B. (2018). Morphological, mechanical and antioxidant properties of Portuguese almond cultivars. *Journal of Food Science and Technology, 55(2),* 467–478.

Özcan, M. M., Ünver, A., Erkan, E., & Arslan, D. (2011). Characteristics of some almond kernel and oils. *Scientia Horticulturae, 127(3),* 330–333.

Paradkar, M. M., Sakhamuri, S., & Irudayaraj, J. (2002). Comparison of FTIR, FT-raman, and NIR spectroscopy in a maple syrup adulteration study. *Journal of Food Science, 67(6),* 2009–2015.

Piscopo, A., Romeo, F. V., Petrovicova, B., & Poiana, M. (2010). Effect of the harvest time on kernel quality of several almond varieties (Prunus dulcis (Mill.) DA Webb). *Scientia Horticulturae, 125(1),* 41–46.

Prats-Moya, S., Grané-Teruel, N., Berenguer-Navarro, V., & Martín-Carratalá, M. L. (1997). Inductively coupled plasma application for the classification of 19 almond cultivars using inorganic element composition. *Journal of Agricultural and Food Chemistry, 45(6),* 2093–2097.

Rodriguez-Campos, J., Escalona-Buendía, H. B., Orozco-Avila, I., Lugo-Cervantes, E., & Jaramillo-Flores, M. E. (2011). Dynamics of volatile and non-volatile compounds in cocoa (Theobroma cacao L.) during fermentation and drying processes using principal components analysis. *Food Research International, 44,* 250–258.

Rogel-Castillo, C., Boulton, R., Opastpongkarn, A., Huang, G., & Mitchell, A.E. (2016). Use of Near-Infrared Spectroscopy and Chemometrics for the Nondestructive Identification of Concealed Damage in Raw Almonds (Prunus dulcis). *Journal of Agricultural and Food Chemistry, 64(29),* 5958–5962.

Sádecká, J., Jakubíková, M., Májek, P., & Kleinová, A. (2016). Classification of plum spirit drinks by synchronous fluorescence spectroscopy. *Food Chemistry, 196,* 783–790.

Sinelli, N., Cosio, M. S., Gigliotti, C., & Casiraghi, E. (2007). Preliminary study on application of mid infrared spectroscopy for the evaluation of the virgin olive oil "freshness". *Analytica Chimica Acta, 598,* 128–134.

Teena, M. A., Manickavasagan, A., Ravikanth, L., & Jayas, D. S. (2014). Near infrared (NIR) hyperspectral imaging to classify fungal infected date fruits. *Journal of Stored Products Research, 59,* 306–313.

Valdés, A., Beltrán, A., & Garrigós, M. C. (2013). Characterization and classification of almond cultivars by using spectroscopic and thermal techniques. *Journal of Food Science, 78,* 138–144.

Valdés, A., Vidal, L., Beltran, A., Canals, A., & Garrigós, M. C. (2015). Microwave-assisted extraction of phenolic compounds from almond skin byproducts (prunus amygdalus): A multivariate analysis approach. *Journal of Agricultural and Food Chemistry, 63*, 5395–5402.

Varela, P., Chen, J., Fiszman, S., & Povey, M. J. W. (2006). Crispness assessment of roasted almonds by an integrated approach to texture description: texture, acoustics, sensory and structure. *Journal of Chemometrics, 20(6-7),* 311–320.

Vázquez-Araújo, L., Enguix, L., Verdú, A., García-García, E., & Carbonell-Barrachina, A. A. (2008). Investigation of aromatic compounds in toasted almonds used for the manufacture of turrón. *European Food Research and Technology, 227(1),* 243–254.

Vlachos, N., Skopelitis, Y., Psaroudaki, M., Konstantinidou, V., Chatzilazarou, A., & Tegou, E. (2006). Applications of Fourier transform-infrared spectroscopy to edible oils. *Analytica Chimica Acta, 573,* 459–465.

Wold, S., Sjöström, M., & Eriksson, L. (2011). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems, 58,* 109–30.

Workman Jr, J., & Weyer, L. (2008). Practical guide to interpretive near-infrared spectroscopy. United States: CRC press, Taylor & Francis.

Yada, S., Lapsley, K., & Huang, G. (2011). A review of composition studies of cultivated almonds: Macronutrients and micronutrients. *Journal of Food Composition and Analysis, 24(4-5),* 469–480.

Zamany, A. J., Samadi, G. R., Kim, D. H., Keum, Y. S., & Saini, R. K. (2017). Comparative study of tocopherol contents and fatty acids composition in twenty almond cultivars of Afghanistan. *Journal of the American Oil Chemists' Society, 94(6),* 805–817.

Zhang, L. F., Guo, B. H., & Zhang, Z. M. (2002). Synthesis of multifunctional polypropylene via solid phase cografting and its grafting mechanism. *Journal of Applied Polymer Science, 84,* 929–935.

Figure captions

Fig. 1. Representative raw and pre-treated (a) NIR and (b) ATR-FTIR spectra of intact almonds.

Fig. 2. PCA score and X-loading plots of the two first PCs using (a,c) NIR and (b,d) ATR-FTIR pre-treated spectral data, respectively.

Fig. 3. Predicted values for the test set almonds of the PLS-DA models constructed with (a) NIR and (b) ATR-FTIR data.

Fig. 4. QDA plots constructed with (a) NIR and (b) ATR FT-IR data for the discrimination of the test set almonds according to their variety.

*Highlights (for review)

Highlights

- Varietal classification of intact Spanish almonds using NIR and ATR-FTIR.
- QDA and PLS-DA were applied to both NIR and ATR-FTIR pre-treated spectral data.
- A performance of 94.45% was obtained with both PLS-DA of ATR-FTIR and QDA of NIR.

Figure 1

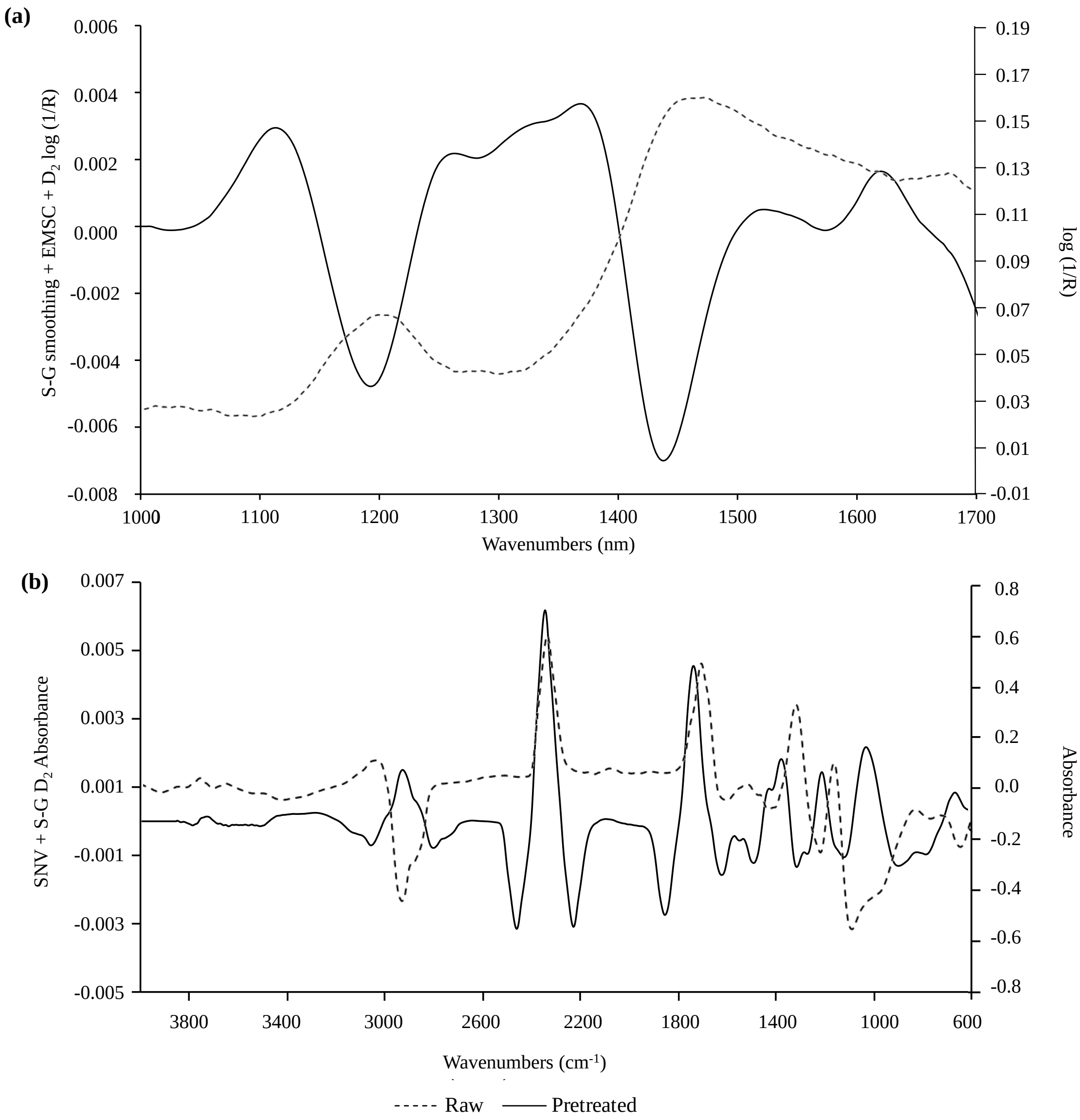

Figure 1. V. Cortés et al.

Figure 2

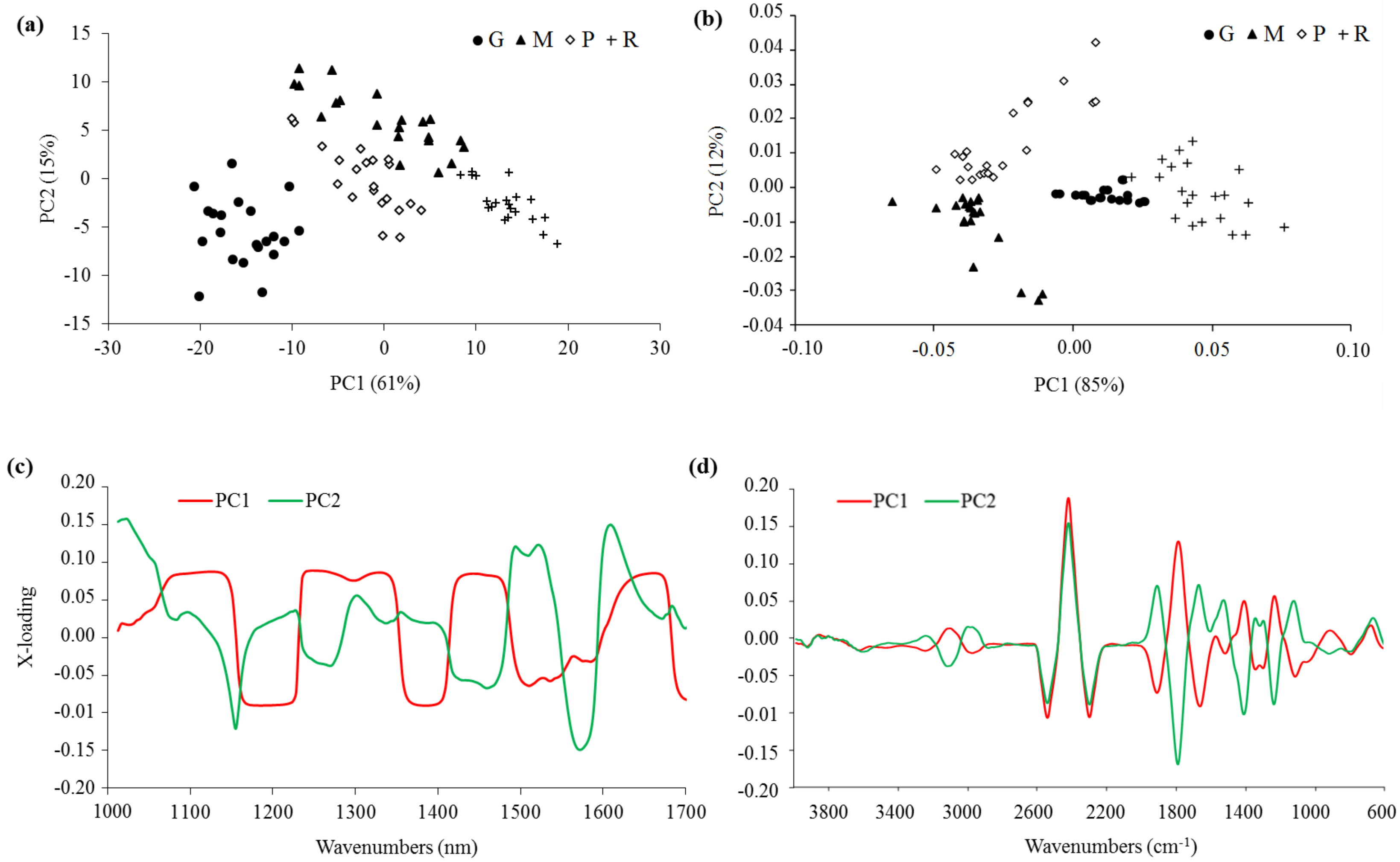

Figure 2. V. Cortés et al.

Figure 3

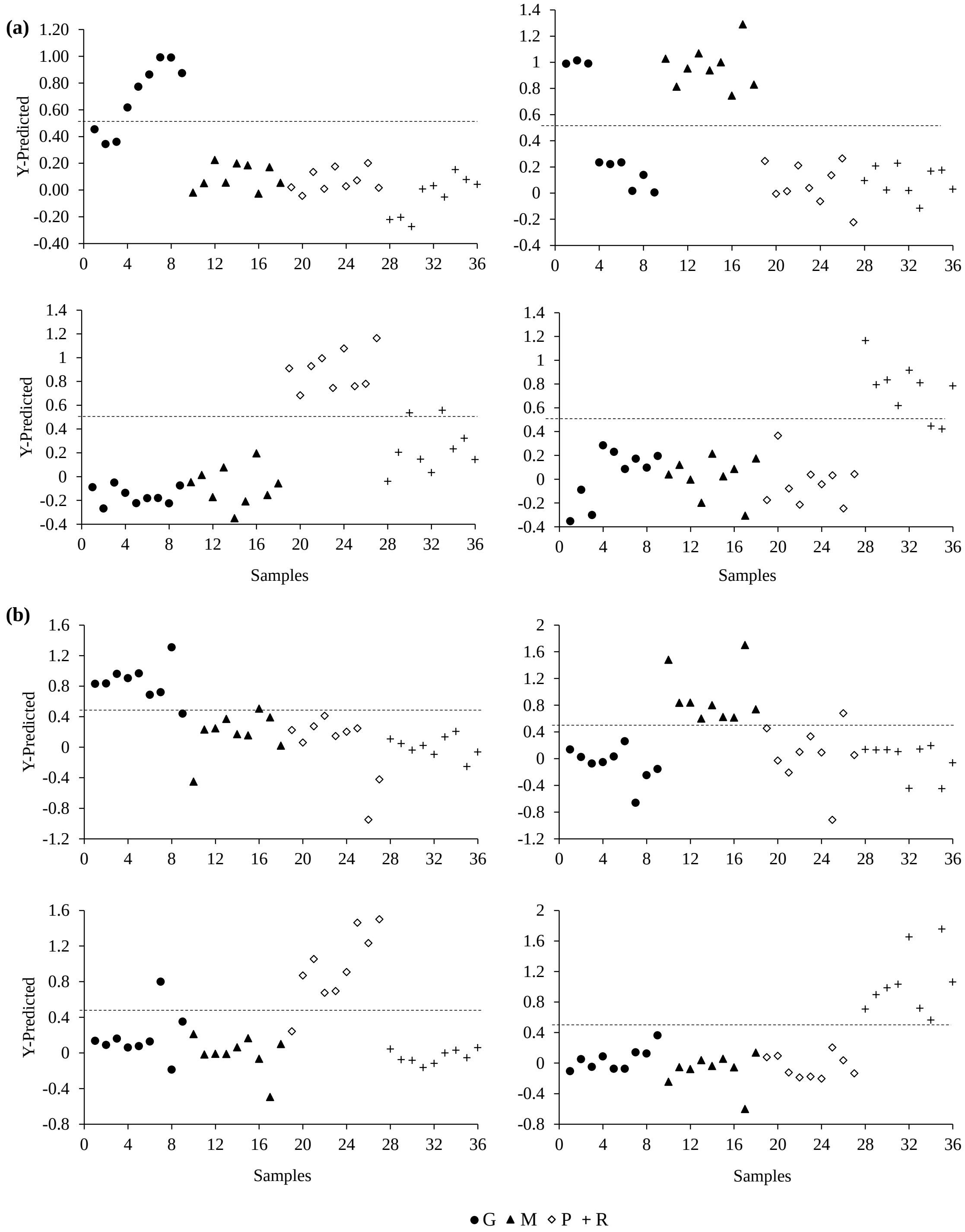

Figure 3. V. Cortés et al.

Figure 4

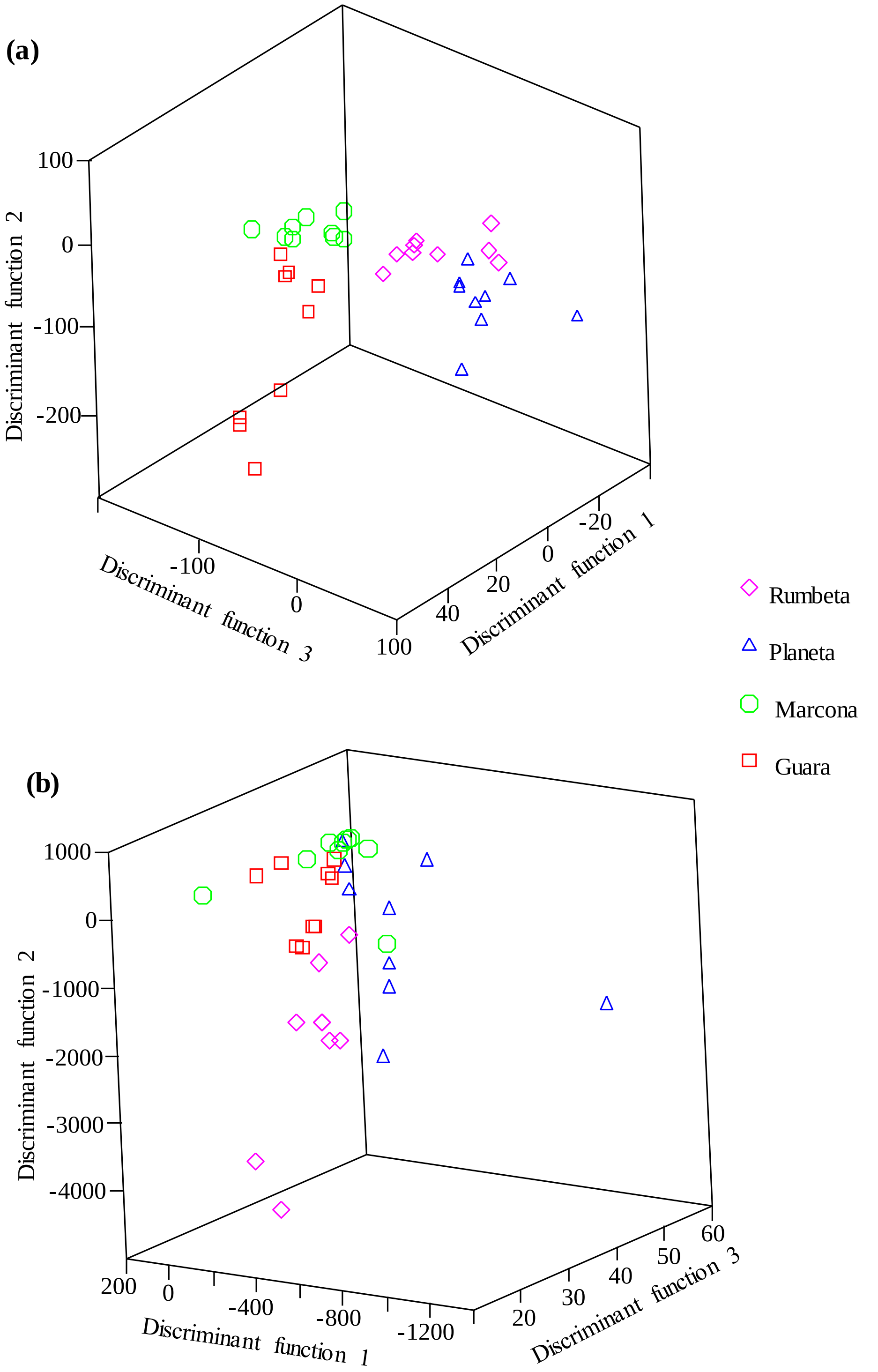

Figure 4. V. Cortés et al.

Table 1

Results of the accuracy of the PLS-DA models constructed to classify almonds according to their variety using training set samples.

	Categories	Calibration		Cross-validation	
		R^2_C	RMSEC	R^2_{CV}	RMSECV
NIR	*Guara*	0.93	0.10	0.91	0.12
	Marcona	0.94	0.11	0.91	0.13
	Planeta	0.93	0.11	0.90	0.14
	Rumbeta	0.91	0.14	0.85	0.18
ATR-FTIR	*Guara*	0.94	010	0.87	0.15
	Marcona	0.93	0.11	0.86	0.16
	Planeta	0.97	0.08	0.92	0.13
	Rumbeta	0.96	0.09	0.91	0.13

R^2_C = coefficient of determination for calibration; RMSEC = root mean square error of calibration; R^2_{CV} = coefficient of determination for cross-validation; RMSECV = root mean square error of cross-validation.

Table 2

PLS-DA and QDA classification results of test set almond samples using NIR and ATR-FTIR data.

			Correct classification				
PLS-DA	***NIR***	Categories	*Guara*	*Marcona*	*Planeta*	*Rumbeta*	Total (%)
		Guara	6/9 (66.7%)	3	0	0	86.13
		Marcona	0	9/9 (100%)	0	0	
		Planeta	0	0	9/9 (100%)	0	
		Rumbeta	0	0	2	7/9 (77.8%)	
	ATR-FTIR	Categories	*Guara*	*Marcona*	*Planeta*	*Rumbeta*	Total (%)
		Guara	8/9 (88.9%)	0	1	0	94.45
		Marcona	0	9/9 (100%)	0	0	
		Planeta	0	1	8/9 (88.9%)	0	
		Rumbeta	0	0	0	9/9 (100%)	
QDA	***NIR***	Categories	*Guara*	*Marcona*	*Planeta*	*Rumbeta*	Total (%)
		Guara	9/9 (100%)	0	0	0	94.45
		Marcona	1	8/9 (88.9%)	0	0	
		Planeta	0	0	8/9 (88.9%)	1	
		Rumbeta	0	0	0	9/9 (100%)	
	ATR-FTIR	Categories	*Guara*	*Marcona*	*Planeta*	*Rumbeta*	Total (%)
		Guara	5/9 (55.6%)	2	2	0	77.80
		Marcona	1	7/9 (77.8%)	1	0	
		Planeta	1	0	8/9 (88.9%)	0	
		Rumbeta	0	0	1	8/9 (88.9%)	