Cortes-Lopez, V.; Talens Oliag, P.; Barat Baviera, JM.; Lerma-García, MJ. (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. https://doi.org/10.1016/j.foodcont.2018.03.040

Potential of NIR spectroscopy to predict amygdalin content established by HPLC in intact almonds and classification based on almond bitterness

Victoria Cortés, Pau Talens, José Manuel Barat, María Jesús Lerma-García*

Departamento de Tecnología de los Alimentos, Universitat Politècnica de València, Camino de Vera s/n, 46022, Spain

*Corresponding author:

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

Abbreviated running title: Amygdalin content and sweet and bitter almonds classification by NIR

ABSTRACT

In this study, 360 intact almonds, half sweet and half bitter, were assessed by near-infrared (NIR) spectroscopy to predict amygdalin content (established by high performance liquid chromatography (HPLC)) and by applying partial least squares (PLS) to the spectral data. After optimising amygdalin extraction and chromatographic conditions, the amygdalin contents found by HPLC were not detected or below to 350 $mg \cdot kg^{-1}$ for sweet almonds, and between 14,700 and 50,400 $mg \cdot kg^{-1}$ for bitter almonds. The intact almond spectra resulted in good predictions of amygdalin content with R^2_p of 0.939 and RMSEP of 0.373. Almonds were correctly classified into sweet and bitter by linear discriminant analysis (LDA), quadratic discriminant analysis (QDA) and PLS-DA, with sensitivity and specificity values higher than 0.94 for evaluation set samples. Based on these results, it

can be concluded that NIR spectroscopy is a good non-destructive alternative to be used as an automatic in-line classification system by food industry.

Keywords

HPLC

NIR spectroscopy

Amygdalin content prediction

Classification

Bitter and sweet almonds

Quality control

1. Introduction

Almonds (*Prunus amygdalus*) are an edible kernel in its natural state and a fruit of high commercial value for the food and cosmetic industries. Two different types can be distinguished depending on kernel bitterness: bitter and sweet almonds (Borrás, Amigo, van den Berg, Boqué, & Busto, 2014). Sweet almonds are widely used as a main ingredient in manufactured food products, while bitter almonds provide the main source of bitter almond oil, used as both flavouring and an ingredient in cosmetics (Salas-Salvadó, Casas-Agustench, & Salas-Huetos, 2011). The bitter almond flavour is a consequence of the presence of cyanogenic glucosides, such as amygdalin and prunasin (Sánchez-Pérez, Jørgensen, Olsen, Dicenta, & Møller, 2008). Amygdalin concentrates in almond kernels, while prunasin is a monoglycoside of roots, leaves and kernel of immature almonds that converts into amygdalin during maturation. The bitter taste occurs due to enzymatic hydrolysis by β-glucosidase that produces benzaldehydes, sugars and hydrogen cyanide to provide a chemical defence barrier against herbivores, insects and pathogens.

One complicated aspect for the almond sector is lack of homogeneity of almond batches. For example, many different shape and size varieties are marketed with the same commercial name which cause disorders for the processing industry. Additionally, the presence of bitter almonds in batches and, as a result, in the final food product, can become a public health issue that endangers almond marketing. Hence, it is necessary to develop analytical methodologies capable of discriminating between sweet and bitter almonds. One of the traditionally used methods to determine cyanogenic compounds in almonds is high performance liquid chromatography (HPLC). Several authors have studied the optimisation of the extraction process and the effect of sample preparation to evaluate amygdalin and prunasin levels in almonds (Arrazola, Grané, Martin, & Dicenta,

2013; Bolarinwa, Orfila, & Morgan, 2014; Dicenta et al., 2002; Ferrara, Maggio, & Pizzigallo, 2010; Lee, Zhang, Wood, Rogel Castillo, & Mitchell, 2013; Yıldırım, San, Koyuncu, & Yıldırım, 2010). However, such methodologies are complex, expensive, require highly trained personnel, long sample preparation times and reagent use, and are also destructive, which means that these technologies are not always available to all food industries (Liang, Slaughter, Ortega-Beltran, & Michailides, 2015). Thus simpler, faster and non-destructive techniques are required and near-infrared (NIR) spectroscopy is a good alternative. The potential of this technology has been previously demonstrated to authenticate the geographical origin of pistachio and to recognise samples with a Protected Designation of Origin (Vitale, Bevilacqua, Bucci, Magrì, Magrì, & Marini, 2013), to inspect internal damages in almonds (Nakariyakul, 2014), and to detect both fungal infection (*Aspergillus flavus and Aspergillus parasiticus*) in almond kernels (Liang et al., 2015) and hidden damage in raw almonds (Rogel-Castillo, Boulton, Opastpongkarn, Huang, & Mitchell, 2016), among others. Only one published study about the discrimination of sweet and bitter almonds using both NIR and Raman spectroscopy was found in the literature (Borrás et al., 2014), and no work is available about predicting the amygdalin content of both sweet and bitter almonds by rapid and non-destructive techniques.

The objective of this study was to investigate the feasibility of NIR spectroscopy, in combination with chemometric methods, to non-destructively predict amygdalin content in intact almonds (established by HPLC). It should be pointed that amygdalin was the only cyanogenic glucoside quantified in this work since prunasin was not present in mature almonds. Moreover, the potential of this technique was also evaluated in the discrimination of sweet from bitter almonds.

2. Materials and methods

2.1. Chemicals and samples

The following analytical grade reagents were used: amygdalin (BioXtra, ≥ 97.0% HPLC, Sigma-Aldrich, St. Louis, Missouri, USA), acetonitrile (ACN, HPLC Far UV/Gradient Grade, J.T. Baker, The Netherlands), methanol (MeOH, AGR ACS, ISO, Ph.Eur. Assay ≥ 99.8%, Labkem, Barcelona, Spain) and acetone (VWR Prolabo, Fontenay – sous – Bois, France). Deionised water was obtained using an Aquinity deionizer (Membrapure GmbH, Berlin, Germany).

The number of almonds employed in this study were 360 (180 sweet and 180 bitter almonds), which were kindly provided by Agricoop (Alicante, Spain). Sweet almonds belonged to six commercial varieties: Planeta (P), Comuna (C), Largueta (L), Rumbeta (R), Marcona (M) and Guara (G). Bitter almonds (A) were a mix of non-specific varieties. The analysed almonds were free of visual damage and were of uniform size and colour.

2.2. Instrumentation and experimental conditions

2.2.1. NIR spectroscopy

In situ recording of NIR spectra was directly carried out on the intact almond kernel (with skin) at room temperature (22±1°C) in a NIR spectrometer from Avantes BV (The Netherlands), model AVS-DESKTOP-USB2. The NIR spectrometer collected spectra by covering the 888–1,795 nm range, and data were measured every 3.535 nm in the diffuse reflectance mode using a detector model AvaSpec-NIR256-1.7 NIRLine. Two

points were acquired per sample on each almond side, and the mean of both spectra was employed for the statistical analysis. All the spectra were acquired using a bi-directional fibre optic reflectance probe, model FCR-7IR200-2-45-ME, whose tip is created at 45º to avoid back-reflection from almond face. The legs of the probe are formed by six fibre cables of 200 μm by connecting one leg with the light source (model AvaLight-HAL-S, formed by a 10-W tungsten halogen) and the other with the spectrometer. The software to perform the spectroscopic measurements was Avasoft version 7.2.

The external white reference was a 99% diffuse reflectance standard (WS-2, Avantes BV) which allowed to adjust the integration time to 500 ms for a maximum reflectance value of around 90% of saturation (Lorente, Escandell-Montero, Cubero, Gómez-Sanchis, & Blasco, 2015). The dark spectrum was recorded by completely covering the probe tip and by switching off the light source.

2.2.2. Amygdalin extraction and HPLC

In order to proceed with amygdalin extraction, almond skins were removed by immersion in hot water for 5 min before drying almonds at room temperature. Then, almonds were crushed in a porcelain mortar. In order to obtain ca. 100% of amygdalin recovery from the samples, different solvents and extraction times were attempted. The best results were obtained when the grounded almond was suspended in 20 mL MeOH, and kept at constant agitation for 24 h using a magnetic stirrer. Finally, the obtained supernatant was filtered through a 0.22 μm PTFE syringe filter. For sweet almonds, the filtered solution was injected directly into the chromatograph, while for bitter almonds this solution was 1:10 (v/v) diluted with MeOH to obtain an amygdalin concentration that fell within the linear range of the calibration curve.

Amygdalin determination was performed in a liquid chromatograph from Hitachi Ltd. (Tokyo, Japan) model LaChrom Elite. The chromatograph was composed of an auto-sampler and a UV detector (models L-220 and L-2400, respectively). Amygdalin determination was performed using a 5 µm analytical column, model Liquid Purple C18 (250 x 4.6 mm i.d.) from Análisis Vínicos (Tomelloso, Spain). After testing different chromatographic parameters and mobile phase compositions, the best results were obtained with an isocratic elution using a mobile phase that contained water and ACN (80:20 v/v). The other chromatographic conditions were: UV detection, 218 nm; injection volume, 20 µL; flow rate, 1.0 mL min^{-1}.

2.3. Spectral pre-treatment

Prior to spectral pre-treatment, all the spectra were analysed by principal component analysis (PCA) to identify and eliminate defective spectral outliers, and to explore the data structure between objects based on Hotelling's T^2 and squared residual statistics (Beghi, Giovenzana, Tugnolo, & Guidetti, 2017). Then, the diffusive reflectance data were transformed into absorption spectra by *log (1/R)* transformation in order to linearise the correlation with the analyte concentration (Hernández, Lobo, & González, 2006). Moreover, the spectral range was trimmed to a region of 1,000-1,750 nm to reduce spectral noise. In this work, several spectral pre-treatments were simultaneously applied: Savitzky-Golay smoothing using a 3-point gap (Carr, Chubar, & Dumas, 2005), extended multiplicative scatter correction (EMSC) and the second derivate with a 2.3-gap-segment. When used together, the signal-to-noise ratio improved (Gorry, 1990; Savitzky & Golay, 1964), the parallel translation of spectra was eliminated (He, Li, & Shao, 2006; Martens, Nielsen, & Engelsen, 2003; Bruun, Søndergaard, & Jacobsen, 2007), and useful

information was retrieved (Cortés, Ortiz, Aleixos, Blasco, Cubero, & Talens, 2016; Rodriguez-Saona, Fry, McLaughlin, & Calvey, 2001). The criterion to select among the different pre-treatments was to obtain the best predictive ability, which is the equivalent to the highest robustness of the analytical method in the experimental domain (Xiaobo, Jiewen, Povey, Holmes, & Hanpin, 2010).

2.4. Chemometric data processing

Spectral data were organised in a matrix, where rows represented the number of samples (N = 360, 180 sweet and 180 bitter almonds) and columns denoted variables (X-variables and Y-variables). The X-variables, or predictors, were the spectral signals. The Y-variables, or responses, were the amygdalin percentages determined per sample by HPLC or the dummy variable for the classification models.

To develop the prediction and discriminant models, a training set was used that consisted in randomly selecting 80% of samples. Each model was internally validated by the leave-one-out cross-validation technique (Huang, Yu, Xu, & Ying, 2008). An independent evaluation set composed of the remaining 20% of samples was used to evaluate the constructed models (Soares, Gomes, Galvão Filho, Araújo, & Galvão, 2013).

Both spectral pre-treatment and multivariate analysis were performed with the statistical software program 'The Unscrambler X' (version 10.3, Camo Process SA, Trondheim, Norway).

2.4.1. Predicting amygdalin content using PLS

PLS was the selected chemometric technique to predict the amygdalin content of both sweet and bitter almonds. For PLS, covariance was maximised between the linear functions of the spectral variations (X-variables) and the corresponding defined value of amygdalin content (Y-variable). PLS model accuracy was judged according to the values of: the root mean square error of calibration, cross-validation and prediction (RMSEC, RMSECV and RMSEP, respectively) and the coefficient of determination for calibration, cross-validation and prediction (R^2_C, R^2_{CV}, R^2_P, respectively), and also by the required number of latent variables (LV).

2.4.2. Classifying almonds according to their bitterness

Classification of almonds in terms of bitterness (sweet and bitter) was performed by constructing three different discriminant models: LDA, QDA and PLS-DA. These models are supervised algorithms based on the relationship between spectral intensity and sample characteristics; in this case using the spectral variations as X-variables and the two established categories (sweet and bitter) as Y-variables. For PLS-DA, the Y-variable was a discrete numerical value (zero for sweet and one for bitter almonds), while LDA and QDA assumed a categorical value created by assigning different letters to sweet and bitter almonds. Hence, these discriminant analyses sought to correlate the spectral variations with the defined classes in attempt to maximise the covariance between both types of variables.

For LDA and QDA, the number of samples in the training set had to be larger than the number of variables included in the model (Kozak & Scaman, 2008; Sádecká, Jakubíková, Májek, & Kleinová, 2016). Thus, variable reduction was necessary. This reduction was achieved using the PCA scores as input data since the linear combinations

of the original variables called principal components (PCs) were not correlated (Rodriguez-Campos, Escalona-Buendía, Orozco-Avila, Lugo-Cervantes, & Jaramillo-Flores, 2011). In this study, the first nine PCs were used to supersede the original data (He et al., 2006).

The three classification models were then evaluated for sensitivity and specificity, where sensitivity relates to the probability that the sample possessing the desired characteristic gives a positive test result, while the latter is the probability that the sample without the desired characteristic gives a negative test result (Amodio, Ceglie, Chaudhry, Piazzolla, & Colelli, 2017). This also leads to the development of valuable indices, such as the non-error rate (NER) or classification rate, which represents the percentage of the correctly classified samples, and is the average of the sensitivity calculated over the various classes, as indicated in Eq. (1):

$$NER = \frac{\sum_{a=1}^{A} Sn_a}{n} \quad (1)$$

where *Sn* is the sensitivity for each *a* class and *n* is the total number of classes.

3. Results and discussion

3.1. Optimising amygdalin extraction

To achieve the best extraction conditions to recover amygdalin from almonds, two parameters, extraction solvent and time, were optimised. The best conditions were selected by establishing the recovery percentages, which were estimated by considering the quantity of amygdalin in the almond and the quantity recovered after applying the extraction conditions. All the experiments were repeated 3 times.

To perform the experiment, sweet almonds in which amygdalin content was not detected were employed. For each test, 2 mg amygdalin standard was added to 1 g of triturated almond. Initially, three solvents (MeOH, H_2O and MeOH:H_2O 80:20 (v/v)) were tested and their extraction efficiency was compared. For this purpose, 20 mL of each solvent were added to the spiked sample to be continuously stirred for 24 h. Recoveries ca. 100% were obtained using MeOH, while the other solvents (H_2O and MeOH:H_2O 80:20 (v/v)) only provided recovery values ca. 15 and 2%, respectively. For this reason, MeOH will be next used.

After optimizing the extraction solvent, the extraction time was next optimised. For this purpose, times between 15 min and 32 h were assayed. Recovery values were ca. 100% only after 24 h of extraction. For this reason, this time was selected for further studies.

3.2. Optimizing chromatographic conditions

The optimisation of the chromatographic conditions was performed using the 2mg g^{-1} spiked almond sample to achieve a satisfactory resolution between the amygdalin peak and other matrix peaks that absorb at 218 nm (maximum amygdalin wavelength). Most manuscripts previously reported in literature (Ferrara et al., 2010; Arrazola et al., 2013; Yıldırım et al., 2010; Dicenta et al., 2002) employed mixtures of ACN and H_2O as the mobile phase; then mixtures of ACN and H_2O at different percentages were tested by using two types of elution: isocratic and gradient. In all cases, 20 µL of sample were injected at a 1 mL min^{-1} flow rate. The best results considering both resolution and analysis time were provided by isocratic elution with 80% H_2O and 20% ACN, which was selected.

After selecting the previous conditions, the influence of flow rate on separation performance was evaluated. Flow rates between 0.5 and 1.5 mL min^{-1} were tested. Using a 0.5 mL min^{-1}, peak width and retention time increased when compared with a flow rate of 1.0 mL min^{-1}. When 1.5 mL min^{-1} was tested, amygdalin and a matrix peak, which were resolved using lower flow rates, partially overlapped. Therefore, a 1.0 mL min^{-1} flow rate was established as the optimal one.

Lastly, different injection volumes were also tested (from 5 to 30 µL). A 20 µL volume was adopted as the best compromise between peak resolution and sensitivity.

3.3. HPLC analytical figures of merit and amygdalin determination in almonds

The different parameters to evaluate the analytical performance of the HPLC method are shown in Table 1. The method's precision was evaluated by the repeatability values obtained within one day and for three days. As observed, the relative standard deviation (RSD) values for the retention times and peak areas were lower than 0.31 and 1.41, respectively.

Amygdalin quantification was performed by the external calibration curves of the peak areas. To construct them, six amygdalin standard solutions at different concentrations between the ranges shown in Table 1 were prepared and injected. The concentrations employed for the first calibration curve were 0.1, 1, 5, 10, 25 and 50 mg L^{-1}, while they were 50, 100, 250, 500, 750 and 1000 mg L^{-1} for the second one. The calibration curve constructed within the 0.1-50 mg L^{-1} amygdalin range was employed to quantify amygdalin in sweet almonds, whereas the other calibration curve (50-1000 mg L^{-1}) was used to quantify the analyte in bitter almonds after a 1:10 (v/v) sample dilution with MeOH to obtain an amygdalin concentration that fell within the linear range of the

second calibration curve. The obtained determination coefficients were higher than 0.9996 (see Table 1). The other parameters included in Table 1 were the limit of detection (LOD) and limit of quantification (LOQ) of amygdalin. LODs and LOQs were estimated following the ICH guidelines (1996). As shown in Table 1, the LOD and LOQ values were 0.02 and 0.07 mg L^{-1}, respectively. These values were lower than others found in bibliography (Ferrara et al., 2010; Arrazola et al., 2013; Bolarinwa et al., 2014). Moreover, in order to assure that no matrix effect was observed while quantifying amygdalin in almonds, standard addition calibration curves (taking into account the linearity ranges of Table 1) were also constructed. Both curves provided $R^2 > 0.9995$ and similar slopes as the external calibration curves. Therefore, it was concluded that the external calibration curves were correctly used to quantify amygdalin in almonds.

The efficiency of amygdalin extraction from almonds was estimated by a recovery study. To carry out this study, sweet almonds in which amygdalin was not detected were spiked with different amygdalin contents that ranged from 0.1 to 60 mg per 1g of almond (see Table 2). All the obtained recovery values were comprised between 98.4% and 102.9%, which demonstrated excellent amygdalin extraction efficiency.

Next, all the almond samples considered in this study (360 almonds) were injected into the HPLC system. According to the amygdalin content found in these samples, they were classified into two groups: sweet almonds, whose amygdalin content was under the LOD or below 350 mg·kg^{-1}, and bitter almonds, whose content ranged from 14,700 and 50,400 mg·kg^{-1}. In sweet almonds, the amygdalin content varied among the different varieties considered in this study: the lowest content was obtained for the Planeta and Comuna varieties, in which the amygdalin content of several almonds was below the LOD. The highest content was found for the Guara variety (350 mg·kg^{-1}). These contents

were consistent with the amygdalin contents previously reported in the literature (Lee et al., 2013).

3.4. Spectral analysis

Fig. 1 shows the raw spectra (Fig. 1a) and the pre-treated mean spectra (Fig. 1b) of sweet and bitter almonds, where the presence of signal peaks at wavelengths of 1125, 1195, 1250, 1380, 1440, 1625 and 1730 nm were evidenced. The region at 1,370-1,400 nm corresponded to the first vibrational overtones which is associated with the O-H stretching modes of water absorption (Clément, Dorais, & Vernon, 2008; Lestander & Geladi, 2005; Magwaza, Opara, Nieuwoudt, Cronje, Saeys, & Nicolaï, 2012). It is known that sugars display bands in the wavelength regions of 1,100-1,600 nm and 1,700-2,300 nm (Tewari, & Irudayaraj, 2004); hence, the signal peaks observed within these regions could correspond to the second and first overtones of the C-H stretching associated with sugars (Osborne, Fearn, & Hindle, 1993; Golic, & Walsh, 2006; Walsh, Golic, & Greensill, 2004).

3.5. Predicting amygdalin content using PLS

In order to predict the amygdalin content of the almond samples, a PLS model was constructed. For the calibration samples, the lowest RMSEC value was 0.28 when seven LV were included in the calibration model, with a R^2_C of 0.967. When the model was validated using the leave-one-out cross-validation technique, the obtained RMSECV was 0.337, with a R^2_{CV} of 0.954. Finally, when PLS model performance was evaluated by evaluation set samples, the RMSEP was 0.373 with a R^2_P of 0.939. The good prediction

performance obtained for the evaluation set samples is shown in Fig. 2. Thus, the obtained results demonstrated that the calibration model optimised by the leave-one-out cross-validation was representative, and that the model could accurately predict amygdalin content in different almond lots with unknown content.

3.6. Classifying almonds according to their bitterness using LDA, QDA and PLS-DA

The possibility of classifying sweet and bitter almonds was evaluated by constructing and comparing three discrimination methods: LDA, QDA and PLS-DA. As previously mentioned, a PCA model was first constructed for LDA and QDA to reduce the variables. The first nine PCs explained 95% of the spectral data. Thus, these PCs were used for LDA and QDA model construction purposes. When the three models were constructed, all the training set samples were correctly classified. The results obtained for the evaluation set samples for all the models are shown in Table 3. As observed in the confusion table, only two samples of each category were not correctly assigned for LDA and QDA, while all the samples were correctly classified for PLS-DA. The same result is observed in Fig. 3. Moreover, the corresponding sensitivity and specificity values for each model are also included in Table 3. The best results were obtained with the PLS-DA model (sensitivity, specificity and NER of 100%), although the results of the LDA and QDA models were also satisfactory (sensitivity, specificity and NER of 94.4%). Thus, it can be concluded that all the discriminant models led to a satisfactory almond classification according to their bitterness.

4. Conclusions

In this work, the potential of NIR spectroscopy to predict amygdalin content in intact almonds and to classify almonds according to their bitterness was demonstrated. After optimising different experimental parameters and chromatographic conditions, the amygdalin content of 360 almonds was established by HPLC. The amygdalin content of the sweet almonds of different commercial varieties was not detected or below 350 $mg \cdot kg^{-1}$, whereas amygdalin content ranged between 14,700 and 50,400 mg kg^{-1} for bitter almonds. Using PLS, the amygdalin content of these samples was satisfactorily predicted with R^2_P and with RMSEP of 0.939 and 0.373, respectively. Moreover, both sweet and bitter almonds were correctly classified into these categories by the construction of the LDA, QDA and PLS-DA discriminant models, where the best results were obtained for the PLS-DA model. Thus, it can be concluded that the NIR spectroscopy technique is a very promising non-destructive alternative to discriminate between sweet and bitter almonds, which could be implemented into industry as an automatic in-line classification system to ensure satisfactory almond quality control.

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 donating the almonds.

References

Amodio, M.L., Ceglie, F., Chaudhry, M.M.A., Piazzolla, F., & Colelli, G. (2017). Potential of NIR spectroscopy for predicting internal quality and discriminating among strawberry fruits from different production systems. *Postharvest Biology and Technology*, *125*, 112-121.

Arrazola, P., Grané, N., Martin, M. L., & Dicenta, F. (2013). Determination of cyanogenic compound amygdalin and prunasin in almond kernels (prunus dulcis l) by using liquid chromatography. *Revista Colombiana de Química, 42(3),* 365–386.

Beghi, R., Giovenzana, V., Tugnolo, A., & Guidetti, R. (2017). Application of visible/near infrared spectroscopy to quality control of fresh fruits and vegetables in large-scale mass distribution channels: a preliminary test on carrots and tomatoes. *Journal of the Science of Food and Agriculture,* DOI: 10.1002/jsfa.8768.

Bolarinwa, I.F., Orfila, C., & Morgan, M.R. (2014). Amygdalin content of seeds, kernels and food products commercially-available in the UK. *Food chemistry, 152,* 133–139.

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.

Bruun, S.W., Søndergaard, I., & Jacobsen, S. (2007). Analysis of protein structures and interactions in complex food by near-infrared spectroscopy. 2. Hydrated Gluten. *Journal of Agricultural and Food Chemistry, 55,* 7244–7251.

Carr, G.L, Chubar, O., & Dumas, P. (2005). Multichannel detection with a synchrotron light source: Design and potential. Spectrochemical Analysis Using Multichannel

405 Detectors Analytical Chemistry Series, edited by *Bhargava P., Levin I. Chapter 3, pp.*
406 *56–84. Oxford: Wiley-Blackwell.*

407 Clément, A., Dorais, M., & Vernon, M. (2008). Multivariate approach to the
408 measurement of tomato maturity and gustatory attributes and their rapid assessment by
409 Vis-NIR spectroscopy. *Journal of Agricultural and Food Chemistry, 56*, 1538–1544.

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

413 Dicenta, F., Martinez-Gomez, P., Grane, N., Martin, M. L., Leon, A., Canovas, J. A.,
414 & Berenguer, V. (2002). Relationship between cyanogenic compounds in kernels, leaves,
415 and roots of non-bitter and bitter kernelled almonds. *Journal of Agricultural and Food*
416 *Chemistry, 50(7)*, 2149−2152.

417 Ferrara, G., Maggio, P., & Pizzigallo, M.D.R. (2010). Cyanogenic d-amygdalin
418 contents of the kernels of cultivated almonds and wild Amygdalus *webbii Spach*. *Journal*
419 *of Horticultural Science and Biotechnology, 85*, 410–414.

420 Golic, M., & Walsh, K.B. (2006). Robustness of calibration models based on near
421 infrared spectroscopy for the in-line grading of stone fruit for total soluble solids content.
422 *Analytica Chimica Acta, 555(2),* 286–291.

423 Gorry, P.A. (1990). General least-squares smoothing and differentiation by the
424 convolution (Savitzky-Golay) method. *Analytical Chemistry, 62*, 570–573.

425 He, Y., Li, X.L., & Shao, Y.N. (2006). Discrimination of varieties of apple using near
426 infrared spectra based on principal component analysis and artificial neural network
427 model. *Spectroscopy and Spectral Analysis, 26*, 850–853.

428 Hernández, Y., Lobo, M.G., & González, M., (2006). Determination of vitamin C in
429 tropical fruits: a comparative evaluation of methods. *Food Chemistry, 96(4)*, 654– 664.

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.

International Conference on Harmonization (ICH guidelines). (1996). Validation of analytical procedures: Text and methodology. *Geneva: ICH-Q2.*

Kozak, M., & Scaman, C.H. (2008). Unsupervised classification methods in food sciences: discussion and outlook. *Journal of the Science of Food and Agriculture, 88,* 1115–1127.

Lee, J., Zhang, G., Wood, E., Rogel Castillo, C., & Mitchell, A.E. (2013). Quantification of amygdalin in nonbitter, semibitter, and bitter almonds (Prunus dulcis) by UHPLC-(ESI) QqQ MS/MS. *Journal of agricultural and food chemistry, 61(32),* 7754–7759.

Lestander, T.A., & Geladi, P. (2005). NIR spectral information used to predict water content of pine sedes from multivariate calibration. *Canadian Journal of Forest Research, 35,* 1139–1148.

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.

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.

Magwaza, L.S., Opara, L.U., Nieuwoudt, H., Cronje, P.J.R., Saeys, W., & Nicolaï, B. (2012). NIR spectroscopy applications for internal and external quality analysis of citrus fruit—a review. *Food and Bioprocess Technology, 5,* 425–444.

Martens, H., Nielsen, J.P., & Engelsen, S.B. (2003). Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures. *Analytical Chemistry, 75(3)*, 394–404.

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

Osborne, B.G., Fearn, T., & Hindle, P.H. (1993). Practical NIR spectroscopy with applications in food and beverage analysis. *Burnt Mill, Harlow, Essex, England: Longman Group,* 2nd ed., 123–132.

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.

Rodriguez-Saona, L.E., Fry, F.S., McLaughlin, A., & Calvey, E.M. (2001). Rapid analysis of sugars in fruit juices by FT-NIR spectroscopy. *Carbohydrate Research, 336*, 63–74.

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.

Salas-Salvadó, J., Casas-Agustench, P., & Salas-Huetos, A. (2011). Cultural and historical aspects of Mediterranean nuts with emphasis on their attributed healthy and nutritional properties. *Nutrition, Metabolism & Cardiovascular Diseases, 21*, S1–S6.

478 Sánchez-Pérez, R., Jørgensen, K., Olsen, C.E., Dicenta, F., & Møller, B.L. (2008).
479 Bitterness in Almonds. *Plant Physiology, 146*, 1040–1052.
480 Savitzky, A., & Golay, M.J.E. (1964). Smoothing and differentiation of data by
481 simplified squares procedures. *Analytical Chemistry, 36*, 1627–1639.
482 Soares, S.F.C., Gomes, A.A., Galvão Filho, A.R., Araújo, M.C.U., & Galvão, R.K.H.
483 (2013). The successive projections algorithm. *Trends in Analytical Chemistry, 42*, 84–98.
484 Tewari, J., & Irudayaraj, J. (2004). Quantification of Saccharides in Multiple Floral
485 Honeys Using Fourier Transform Infrared Microattenuated Total Reflectance
486 Spectroscopy. *Journal of Agricultural and Food Chemistry, 52*, 3237–3243.
487 Vitale, R., Bevilacqua, M., Bucci, R., Magrì, A.D., Magrì, A.L., & Marini, F. (2013).
488 A rapid and non-invasive method for authenticating the origin of pistachio samples by
489 NIR spectroscopy and chemometrics. *Chemometrics and Intelligent Laboratory Systems,*
490 *121*, 90–99.
491 Walsh, K.B., Golic, M., & Greensill, C.V. (2004). Sorting of fruit and vegetables using
492 near infrared spectroscopy: application to soluble solids and dry matter content. *Journal*
493 *of Near Infrared Spectroscopy, 12*, 141–148.
494 Xiaobo, Z., Jiewen, Z., Povey, M. J., Holmes, M., & Hanpin, M. (2010). Variables
495 selection methods in near-infrared spectroscopy. *Analytica chimica acta, 667*, 14–32.
496 Yıldırım, A.N., San, B., Koyuncu, F., & Yıldırım, F. (2010). Variability of phenolics,
497 α-tocopherol and amygdalin contents of selected almond (*Prunus amygdalus Batsch.*)
498 genotypes. *Journal of Food, Agriculture & Environment, 8(1)*, 76–79.
499

Figure captions

Fig. 1. Spectra of almonds obtained from (a) raw data and (b) pre-treated data.

Fig. 2. Measured versus predicted amygdalin content by PLS in the prediction set.

Fig. 3. Discrimination plots of the (a) LDA, (b) QDA and (c) PLS-DA models constructed to classify the evaluation set almonds according to their bitterness.