Exploring the potential of NIRS technology for the *in situ* prediction of amygdalin content and classification by bitterness of in-shell and shelled intact almonds

Miguel Vega-Castellote[a], Dolores Pérez-Marín[b,*], Irina Torres[a], José-Manuel Moreno-Rojas[c], María-Teresa Sánchez[a,*]

[a] *Department of Bromatology and Food Technology, University of Cordoba, Rabanales Campus, 14071 Córdoba, Spain.*

[b] *Department of Animal Production, University of Cordoba, Rabanales Campus, 14071 Córdoba, Spain.*

[c] *Department of Food Science and Health, IFAPA-Alameda del Obispo, Avda. Menéndez Pidal, s/n. 14071, Córdoba, Spain*

Corresponding authors. Tel.: +34 957 212576; Fax: 34 957 212000

E-mail addresses: teresa.sanchez@uco.es (M.T. Sánchez) or dcperez@uco.es (D. Pérez-Marín).

Abstract

Amygdalin is a cyanogenic compound found in almonds which gives them their bitter taste. For the almond industry, it is important to prevent the presence of bitter almonds in batches of sweet almond that can affect their commercialization and even consumer safety. This study sought to ascertain the viability of near infrared spectroscopy (NIRS), as a fast and reliable candidate for non-destructive and *in situ* quantification of amygdalin levels and for classification of almonds by bitterness, when analysed in bulk. With that purpose, in-shell and shelled sweet and bitter almonds were analysed in dynamic mode using two new handheld NIRS instruments. As a first step, the amygdalin levels in in-shell and shelled almonds were determined using modified partial least squares (MPLS) and local regression algorithms. Next, classification models for bitterness were made using partial least square discriminant analysis (PLS-DA). For the discrimination between sweet and bitter almonds, two strategies to set up the optimum threshold were studied: the mean value of the discriminant variables and the value calculated using the Receiver Operating Characteristic (ROC) curves. The results for measuring amygdalin in shelled almonds showed that NIRS technology, using both regression algorithms, is a robust technology for inspection purpose at an industrial level. Additionally, excellent performances were obtained for the classification models of the two in-shell and shelled almond groups analysed in bulk with both instruments, with better results when the threshold values obtained from the ROC curves were applied.

Keywords: In-shell and shelled almonds; *In situ* bulk NIRS analysis; Amygdalin content; Discriminant analysis; ROC curve optimum threshold

1. Introduction

Almonds can be divided into two distinct groups: sweet almonds and bitter almonds. The bitter taste of almonds is due to the presence of cyanogenic compounds, such as amygdalin (D-mandelonitrile-2-D-gentiobioside), which is present in almond kernels, and its precursor, prunasin. The prunasin is a monoglycoside present in the roots, leaves and kernel of unripe almonds, which turns into amygdalin during the ripening process (Frehner et al., 1990; Barceloux, 2009). Both compounds are highly toxic and directly influence the product's sensory qualities and acceptability (Arrázola et al., 2012).

Chewing brings amygdalin into contact with the emulsine present in saliva, a β-glucosidase, which breaks this compound down into β-D-glucose, benzaldehyde, and hydrogen cyanide. The benzaldehyde is responsible for the bitter taste and the hydrogen cyanide can cause toxicity. In addition, the consumption of these compounds can lead to poisoning, depending on the amount of bitter almonds ingested (Morant et al., 2008; Mouaffak et al., 2013).

Therefore, the possible existence of bitter almonds in batches of sweet almond can lead to problems in their commercialization and can even affect consumer safety. This accounts for the need to prevent bitter almonds from being processed at an industrial level together with the sweet almonds. However, there is a great variability in the sweet almond batches due to the heterogeneity of shapes, weights and sizes, as well as their variable nutrient composition, which is mainly derived from the variety to which they belong. This variability makes it extremely difficult for the handling and processing industry to classify them correctly (Yada et al., 2013; Arrázola-Paternina et al., 2015).

Currently, the analytical technique used officially to measure the cyanogenic compounds found in these nuts is high performance liquid chromatography (HPLC), which requires a previous extraction process by which the almond has to be shelled before

68 the levels of amygdalin and prunasin can be measured (Lee et al., 2013; Xu et al., 2017;
69 Cortés et al., 2018a). This analytical technique is complex, destructive, highly expensive
70 and time consuming to obtain results. It therefore does not allow real-time responses to
71 be obtained, nor is it affordable for all almond processing industries.

72 For these reasons, the industry currently needs the development, fine-tuning and
73 implementation of faster, cheaper, non-destructive, reliable analytical methodologies to
74 detect bitter almonds which are simpler to use routinely and non-polluting, both in the in-
75 shell and the shelled product. Near infrared spectroscopy (NIRS) is one of the most
76 suitable analytical technologies for this purpose, since it combines speed in its
77 measurements with great versatility, rapid data collection and low cost per sample
78 (Sánchez and Pérez-Marín, 2011). This technology is also highly versatile and allows to
79 analyse the different parameters simultaneously and instantaneously with a single
80 spectrum and give information at different points in the value chain about the quality and
81 authentication of the product analysed *in situ*. In addition, thanks to advances in the
82 instrumentation over recent years, portable NIRS instruments are now available (Teixeira
83 Dos Santos et al., 2013; Pasquini, 2018; Yan and Siesler, 2018; Cortés et al., 2019; Beć
84 et al., 2020).

85 A number of published works have evaluated the use of NIRS technology in
86 different areas of almond production: for diagnosing fungal diseases in seeds using the
87 Foss NIRSystem 6500 spectrophotometer, which is suitable for analysing laboratory
88 samples (Liang et al., 2015); for assessing damage to raw almonds using the MicroNIR
89 2200, a portable instrument suitable for *in situ* product analysis (Rogel-Castillo et al.,
90 2016); for classifying sweet and bitter almonds using the FT-NIR MB160PH Aridzone
91 instrument (Borrás et al., 2014); and for predicting the amygdalin content and its varietal

differentiation in almonds analysed with the AvaSpec-NIR256-1.7 NIRLine instrument (Cortés et al., 2018a, b).

However, these previous works have all been carried out in individual, previously shelled kernels, and not in batches of the product, as required by the industry. In addition, no published studies have been found which deal with in-shell almonds. However, the study of both in-shell and shelled almonds in bulk is of maximum interest since it involves a practical application of NIRS technology in the almond industry. This would allow to measure the amygdalin content in product batches and to discriminate between batches of sweet and bitter almonds when they are received by the industry and throughout the production process.

The aim of this research work was to assess the viability of NIRS technology for measuring the amygdalin content in almonds and for differentiating between sweet and bitter in-shell and shelled almonds, analysed in bulk. At the same time, the performance of two portable latest generation NIRS instruments with different optical designs, which are suitable for *in situ* analysis of the product, was compared to identify the most suitable spectrophotometer for these purposes.

2. Materials and methods

2.1. Samples

A total of 145 in-shell almond samples, harvested during the season 2018-2019, were used in this study. This set in turn was comprised of a group of 84 in-shell sweet almonds (*Prunus dulcis* Mill., cv. 'Antoñeta', 'Belona', 'Guara', 'Lauranne', 'Soleta' and 'Vairon') and a group of 61 in-shell bitter almonds of different varieties. Additionally, 84

117 samples of shelled sweet almonds, of the same varieties and batches as above were
118 analysed while the 61 in-shell bitter almond samples were later manually shelled and
119 analysed. Each sample was around 750 g.
120
121 *2.2. NIR spectrum acquisition*
122
123 The near infrared (NIR) spectra of the in-shell and shelled sweet and bitter
124 almonds were taken using two portable handheld NIRS instruments of different optical
125 designs and technical specifications; the Aurora spectrophotometer (GraiNit S.r.l.,
126 Padova, Italia) and the MicroNIR™ Pro 1700 (VIAVI Solutions, Inc., San Jose,
127 California, USA), both suitable for the *in situ* analysis of the product.
128 The Aurora spectrophotometer is a robust, compact, handheld instrument based
129 on diode array technology. This instrument works in reflectance mode in the spectral
130 range 950–1650 nm, taking data every 2 nm, with an optical window of 1256 mm^2, and
131 has an internal reference, which facilitates calibration. In this work, the sensor integration
132 time was 6.57 ms and each spectrum was the mean of 50 scans. Acquisition of the spectra
133 was carried out by means of the UCal 4^{TM} software (Unity Scientific LLC, Milford, MA,
134 USA). Each sample of in-shell and shelled almonds was uniformly distributed on a white
135 plastic tray covering the whole surface. Four spectra were taken per sample by moving
136 the sensor along the tray containing the almonds (dynamic analysis mode), covering the
137 entire area of the tray.
138 The MicroNIR™ Pro 1700 instrument was also used in this study. It is a light
139 portable miniature spectrophotometer which works in reflectance mode in the spectral
140 range 908 to 1676 nm with a constant interval of 6.2 nm. This instrument incorporates
141 Linear Variable Filters (LVF) technology as the dispersion element and has an optical

window of around 227 mm^2. In this work, the sensor integration time was set at 11 ms and each spectrum was the mean of 200 scans. Spectra acquisition was carried out using the VIAVI MicroNIR software Pro version 2.2 (VIAVI Solutions, Inc., San Jose, California, USA). The instrument's performance was checked every 10 minutes. A white reference measurement was obtained using a NIR reflectance standard (Spectralon^{TM}) with 99 % diffuse reflectance, while a dark reference was obtained from a fixed point on the floor of the room. For in-shell and shelled almonds, the analysis was carried out in dynamic mode following the same procedure described above for the product uniformly distributed on white plastic trays. Four spectra were also taken per sample.

Finally, the four spectra were averaged to obtain a mean spectrum per sample for each presentation form and instrument.

2.3. Reference data

Prior to the extraction of amygdalin, 200 g of shelled almonds were ground in a SK-3 Cutter-Blender (Sammic, Guipúzcoa, Spain) for 60 seconds. Then, 0.6 g of ground almonds were inserted in a 50 mL screw cap tube and mixed with 15 mL of methanol. After that, the mixture was homogenized using a T25 Ultra-Turrax (IKA-Werke Staufen, Germany) for 1 minute. The extraction was performed under constant stirring for 24 hours at 30 ºC. Then, the tubes were centrifuged (Selecta Medifriger-BL, Barcelona, Spain) at 4000 rpm for 15 minutes at 6 ºC. Next, the supernatant was filtered using a 0.45 µm polytetrafluoroethylene (PTFE) syringe filter. The samples were extracted in duplicate and stored at -80 °C for high performance liquid chromatography diode array detector (HPLC-DAD) analysis. Additionally, the amygdalin standard was prepared by dissolving the pure compound in methanol (1 g L^{-1}).

Amygdalin determination was performed using in a HPLC Perkin Elmer series 200 (Waltham, MA, USA), consisting of an HPLC pump, a diode array detector (DAD), and an autosampler operating at 4 °C, and following the method described by Cortes et al. (2018a), with some modifications. Briefly, the amygdalin was separated on a 150 × 4.6 mm i.d. Luna 3 µm C18 (2) column and a 4.0 × 3.0 mm guard column from Analytical Phenomenex (Torrance, CA, USA) and maintained at 40 °C. In the mobile phases, A: deionized water and B: acetonitrile, were pumped at a flow rate of 0.3 mL min^{-1} using an isocratic method (80 % A–20 % B) for 12 minutes. The injection volume was 10 µL and detection was carried out at 218 nm. The linearity of the method was determined by a regression analysis of the area *versus* the amygdalin concentrations. Thus, standard solutions of amygdalin in concentrations ranging from 0.001 to 10 g L^{-1} were prepared and analysed in triplicate. The determination coefficients (R^2) obtained for the standard curves were higher than 0.9945.

2.4. Spectral data pre-processing and definition of calibration and validation sets

Data pre-processing and chemometric treatments were performed using the WinISI II software package version 1.50 (Infrasoft International LLC, Port Matilda, PA, USA) (ISI, 2000) and Matlab R2019a (The Mathworks, Inc., Natick, MA, USA).

The sample set used to carry out the quantitative models for amygdalin determination consisted of 145 in-shell and 145 shelled samples. The structure and variability of the population available was studied using the CENTER algorithm (Shenk and Westerhaus, 1995a), which was applied to the four sets of in-shell and shelled almonds analysed with both instruments used, previous to calibration development.

The algorithm performs a principal component analysis (PCA) and calculates the global Mahalanobis distance (GH) of each sample to the centre of the population in the new n-dimensional space, which enables to sort the samples by their GH distance. An in-depth study of those samples considered as potential outliers or anomalous spectra (GH > 3.5) was carried out. The CENTER algorithm was applied using a combination of mathematical pre-treatments — Standard Normal Variate (SNV) and De-trending (DT) for scatter correction (Barnes et al., 1989), together with the 1,5,5,1 Norris derivative treatment, where the first digit is the order of the derivative, the second is the gap over which the derivative is calculated, the third is the number of data points in a running average or smoothing and the fourth is the second smoothing (Shenk and Westerhaus, 1995b).

Having ordered the sample sets by spectral distances from smallest to largest from the centre, a structured selection of the validation set, i.e. one out of every four samples in the overall set of shelled almonds analysed with Aurora instrument, was performed ($N_{validation}$ = 35). The remaining samples were used to build the calibration set ($N_{calibration}$ = 110) (Shenk and Westerhaus, 1991). Similarly, these samples were then selected to form the calibration and validation groups for the other three groups - i.e. the shelled almonds analysed with Aurora and the in-shelled and shelled almonds analysed with MicroNIR™ Pro 1700.

2.5. *NIRS quantitative models for the prediction of amygdalin content using linear and non-linear regression procedures*

2.5.1. MPLS regression

To develop the NIRS calibration models to predict amygdalin content in intact almonds, the modified partial least squares (MPLS) regression with five cross-validation groups was used (Fig. 1), using the combined pre-treatment of SNV + DT and first or second derivative, 1,5,5,1 and 2,5,5,1 treatment (Shenk and Westerhaus, 1995a).

The best models were selected using the statistics, coefficient of determination for cross validation (R^2_{cv}), standard error of cross validation (SECV) and the residual predictive deviation for cross validation (RPD_{cv}), and were then subjected to an external validation process. For this, the validation samples were used, and the external validation protocol proposed by Windham et al. (1989) was applied to assess their predictive capacity.

2.5.2. LOCAL algorithm

In addition, in this study, we applied a non-linear regression method based on local calibrations. Thus, the LOCAL algorithm (ISI, 2000) was used to predict amygdalin content in shelled almonds analysed with the two handheld instruments tested (Fig. 1).

LOCAL algorithm works by selecting, for each sample to be predicted, those samples which belong to the spectral library available and most resemble the unknown sample. The selected samples are then used to compute a specific calibration equation for each sample to be predicted, based on PLS regression (Shenk et al., 1997; Pérez-Marín et al., 2007).

The calibration samples were selected taking into account the coefficient of correlation value between the spectrum of the unknown sample and those comprising the spectral data base (Shenk et al., 1997). The parameters defined to run and optimize the algorithm for this study of viability were: the number of calibration samples (*k*) from 30 to 50 in steps of 10, the minimum number of calibration samples, fixed at 15, the

maximum number of PLS factors (*l*), which was set at eight, and the number of the first PLS factors to be removed, fixed at three. Furthermore, the same mathematical signal pre-treatments indicated for MPLS regression were also evaluated.

The coefficient of regression for prediction (R_p^2), the standard error of prediction (SEP), the bias, the standard error of prediction corrected for bias (SEP(c)) and the slope value were all used to assess the performance of the LOCAL algorithm using the different settings defined above. After that, the accuracy of prediction of the LOCAL algorithm was compared to the SEP and R_p^2 of MPLS regression.

2.6. Study of the sweet and bitter almond population and construction of NIRS classification models

The discriminant study of the sweet and bitter almonds was carried out using a set of 139 samples (84 sweet and 55 bitter almond samples). In a study conducted by the California Almond Board it was established that semi-bitter and bitter almonds had an amygdalin content of 520-1800 mg kg^{-1} and superior to 33,000 mg kg^{-1}, respectively (Lee et al., 2013). Considering that, 6 samples with amygdalin levels between 62-374 mg kg^{-1} initially considered as bitter were not used for classification purposes.

First, a PCA was performed using the full set of 139 shelled samples analysed with the Aurora instrument and the scores and loadings of this PCA were studied to explore the potential differences between the sweet and bitter almond groups.

Next, the CENTER algorithm was applied to the eight sets of in-shelled and shelled sweet and bitter almonds analysed with both instruments, prior to the development of the qualitative models. The samples that showed a GH > 3.5 were considered as

potential outlier samples and consequently, the spectral and chemical characteristics of those samples were studied in detail.

After applying the CENTER algorithm and ordering the set of samples by spectral distances, the structured selection of training and validation groups was carried out, following the procedure proposed by Shenk and Westerhaus (1991). To select the validation set, one out of every nine sweet samples and one out of every six bitter samples were selected from the group of shelled samples analysed, using the Aurora instrument. The validation set therefore consisted of a total of 20 samples, 10 sweet and 10 bitter, while the remaining samples were used to make up the training set (N_{sweet} = 74 and N_{bitter} = 45). Similarly, the same samples were selected from the other six groups (sweet and bitter in-shell almonds tested with Aurora, and sweet and bitter in-shell and shelled almonds tested with MicroNIR^{TM} Pro 1700) to make up their respective training and validation sets.

The classification models for the sweet and bitter almonds were carried out using partial least squares-discriminant analysis (PLS-DA) (Fig. 1) for supervised classification (Naes et al., 2002). Specifically, the PLS2 algorithm was used, which generates as many discriminant variables as there are classes in the learning group. To develop these models, six cross-validation groups were used and a maximum number of 10 PLS terms was considered. The same signal pre-treatments described earlier for quantitative analysis were also tested for qualitative model development.

The performance of the models was assessed in terms of the sensitivity (fraction of the true positives divided by the true positives and false negatives), specificity (fraction of true negatives divided by true negatives and false positives) and non-error rate (NER), which represents the percentage of correctly classified samples.

Initially, these models were carried out considering the mean value (1.5) of the discriminant variables as the threshold to discriminate between bitter (class 1) and sweet (class 2) almonds. However, according to Downey (2000), this may not be the optimal limit when the models are not balanced as regards the number of samples of the two types. Consequently, and due to the great importance of eradicating the presence of bitter almonds from the marketing channels has for producers of sweet almonds intended for consumption as snacks and other products, an optimum threshold value using the Receiver Operating Characteristic (ROC) curves was also calculated (Serrano-Lourido et al., 2012; Martínez -Cagigal, 2020).

The aim here was to maximize the sensitivity and specificity values obtained with the models developed with a different number of samples per type. In this study, the strategy aimed at optimizing the threshold value was considered more suitable than the one which the models are balanced on, with an equal number of samples per class to discriminate: this involves removing a large number of samples from the type with the most samples, and this information can be very useful when developing the classification models.

The ROC curve is a two-dimensional mapping of the 'false positive rate' and the 'true positive rate' (also respectively called '1 – specificity' and 'sensitivity') for all the possible threshold values between the two classes being studied (Unal, 2017). However, to obtain the optimal threshold, threshold values were sought that would maximize the sensitivity and specificity of the model. In those trials which did not have a single threshold value, but a range of values which maximized sensitivity and specificity, the optimal threshold was taken from the midpoint of the range (Tena et al., 2014).

Finally, the best classification models obtained were subjected to an external validation process, using those samples belonging to the validation group.

3. Results and discussion

3.1. Prediction of amygdalin content in almonds using MPLS regression and LOCAL algorithm

When the CENTER algorithm was applied to the 145 samples available for amygdalin determination, four samples presented a GH > 3.5 (3.70, 3.86, 3.89 and 6.43) when the analysis was carried out in in-shell almonds with the Aurora instrument, plus four (GH = 3.52, 4.03, 5.34 and 8.20) using the MicroNIR™ Pro 1700, three of which were included in the four samples with GH > 3.5 identified using the Aurora instrument. Only one sample belonging to the group of shelled almonds analysed using the Aurora instrument showed a GH value above the limit (5.52). No shelled samples analysed using the MicroNIR™ Pro 1700 instrument presented GH values higher than 3.5.

It is worth noting that most of the samples presenting GH values above 3.5 belonged to the group of samples analysed in-shell, which is, based on previous studies developed by this research group, the sample presentation form that reported the lowest repeatability compared to those NIRS analyses carried out with shelled almonds. None of the samples which had a GH > 3.5 were eliminated, since according to a detailed study, there were no reasons to justify the elimination of these samples.

Table 1 shows the cross validation results for the best prediction models for the amygdalin content in in-shell and shelled almonds analysed with the two instruments tested, using MPLS regression.

According to Shenk and Westerhaus (1996) and Williams (2001), the models developed to predict amygdalin content in in-shell almonds with both instruments would

enable to discriminate between almonds with low, medium and high amygdalin content. The *in situ* quantification of the amygdalin content in in-shell almonds allows to conduct a first screening of the product when received by the industry. Carrying out this screening at the reception points in the industry is of great importance, since the industrial destination of the product will depend on the amount of amygdalin it contains.

This screening in turn would enable to avoid not only the consumption of poisonous substances, but also the typical unpleasant taste of bitter almonds. No previous studies can be found in the literature which focus on predicting the amygdalin content in in-shell almonds.

The models of amygdalin content in shelled almonds showed an excellent predictive capacity for both instruments, with R^2_{cv} values of 0.95 or higher, and RPD_{cv} values higher than 4 (Shenk and Westerhaus, 1996; Williams, 2001). A study conducted by Cortés et al. (2018a) proved NIR spectroscopy to be a suitable tool to quantify the amygdalin content in intact shelled almonds when the product was analysed as individual kernels. However, in the present research, the suitability of NIRS technology was proven when the almond kernels were analysed in batches. This can be very useful when it comes to the quantification of the amygdalin content of the batches, which makes this tool extremely useful for managing the industrial destination of these batches.

Validation of the best calibration models developed with in-shell and shelled almonds and the two instruments tested using MPLS regression was carried out to predict the external validation sets. The negative NIRS predicted values for the amygdalin content are shown as zero (Fig. 2).

The models developed with in-shell almonds complied with the protocol from Windham et al. (1989) in terms of the standard error of prediction corrected for bias (SEP (c)) and the bias, but neither of them complied with the coefficient of determination for

prediction (R^2_p) and only the model developed with samples analysed with the Aurora instrument did so for slope. These prediction results indicate a limited predictive capacity when in-shell almonds are used to develop the model. However, the R^2_p, SEP(c), bias and slope values of the models developed for shelled almonds with both instruments were within the confidence limits established in the protocol established by these authors. According to Nicolaï et al. (2007), the RPD_p values presented by both models developed with shelled almonds indicate an excellent predictive capacity, and these equations can therefore be applied routinely.

As regards amygdalin content parameters, it is common to find groups of almonds with very different amygdalin contents, which in practice form two very different populations, sweet almonds and bitter almonds. The LOCAL algorithm was therefore used only at the feasibility study level, since the number of samples available was small. However, the methodology was considered eminently suitable for sampling groups of this type and for facilitating the prediction of amygdalin by the industry.

The best results to predict amygdalin content in shelled almonds using the LOCAL algorithm are shown in Table 2. When the Aurora instrument was used, the value for R^2_p was improved by 2 % and the SEP was reduced by 16 % as compared to the prediction results obtained for the MPLS model developed with shelled almonds analysed with this instrument. These results highlight that the application of the LOCAL algorithm constitutes an excellent strategy to obtain accurate predictions of the amygdalin content in shelled almonds. Although Shenk et al. (1997) recommend using the LOCAL algorithm with large databases, in this research we aimed to give a hint of the potential of non-linear methods such as LOCAL algorithm to address the problem of underrepresentation of samples in the 10,000–30,000 mg kg^{-1} amygdalin range, which

leads to the presence of two different groups of samples based on their content of this cyanogenic compound.

3.2. Exploratory study of the sweet and bitter almond population

The first and second principal components (PCs) scores plot enabled to evidence the separation between the sweet and the bitter shelled almonds analysed using the Aurora instrument (Fig. 3a). The sweet almonds were associated with PC2 negative values, whereas the bitter ones tended to present positive values for this PC. Seven out of the nine bitter almonds with slightly negative values for PC2 presented amygdalin reference values under 7,200 mg kg^{-1}, with the amygdalin range the bitter almonds 922.97-80,980.13 mg kg^{-1}.

The loading plot (Fig. 3b) showed the main regions for differentiating between the two classes of almond. PC1 showed a peak at 1212 nm, which could be related to the second overtone of C-H bonds and in turn to the presence of lipids, and a peak at 1390 nm characteristic of C-H combination, probably related to fatty acids and carbohydrates. Likewise, PC2 exhibited three main peaks at around 1136 nm that might be attributed to the second overtone of the C-H stretch, 1152 nm, which could correspond to the C-H links of aromatic compounds, and 1406 nm, which could be linked to the first overtone of the O-H functional groups (Shenk et al., 2008; Rogel-Castillo et al., 2016; Zhang et al., 2018; Firmani et al., 2019).

The positive values found in the PC2 axis of the bitter almond samples could be attributed partly to the peak observed in the 1152 nm wavelength of the loading values for this PC. As has been mentioned above, the absorption band at around 1152 nm might be related to the aromatic compounds of almonds, and could associate, therefore, with

412 bitter almonds with a higher content of aromatic compounds compared to sweet ones.
413 Kesen et al. (2018) showed that the amount of aromatic compounds in bitter almond oil
414 (315,283 μg kg^{-1}) was much higher than the amount of these compounds in sweet almonds
415 (3,002 μg kg^{-1}), which supports the statement formulated above.
416
417 *3.3. Classification of almonds by bitterness*
418
419 When the CENTER algorithm was applied to the sweet and bitter almond sets
420 separately, two samples (GH = 3.61 and 4.76) belonging to the group of bitter almonds
421 analysed in-shell with the Aurora instrument, plus one sample (G = 3.79) belonging to
422 the group of bitter samples analysed in-shell using the MicroNIR™ Pro 1700, presented
423 GH values higher than 3.5. No justifiable reasons were found to eliminate these samples
424 from the set and these samples were therefore not discarded.
425 Table 3 shows the results of the classification models obtained, considering a pre-
426 defined threshold value of 1.5 in terms of sensitivity, specificity and NER. The models
427 correctly classified by cross-validation 74/74 samples of sweet almonds and 44/45
428 samples of bitter almonds, while in external validation they correctly classified 10/10
429 sweet samples and 9/10 bitter samples, for the group of in-shell almonds analysed using
430 the Aurora instrument. When the in-shell almond group was analysed using the
431 MicroNIR™ Pro 1700, 73/74 sweet and 42/45 bitter samples were well-classified,
432 respectively, in cross-validation, while in external validation, all were correctly classified.
433 Although the models developed with both instruments classified the majority of the
434 samples correctly, the difference in terms of sensitivity and specificity in cross-validation
435 between the two instruments could be due to the larger window size of the Aurora
436 spectrophotometer, which allows to obtain a more representative measurement of the

437 sample and consequently, greater precision in discriminating between the two types being
438 studied.

439 The models developed with shelled almonds showed 100 % of correctly classified
440 samples in all cases.

441 However, Naes et al. (2002) and Brereton (2009) have shown that when the types
442 are unbalanced in terms of number of samples, the PLS-DA prediction boundary will be
443 biased towards the smaller type, and therefore, a greater number of poorly classified bitter
444 samples will occur (Fig. 4).

445 Fig. 5 and Fig. 6 show the sensitivity and specificity values against the threshold
446 values and the ROC curves, respectively. In all cases, there is a range of threshold values
447 which maximizes sensitivity and specificity (Fig. 5). The midpoint of this interval was
448 chosen as the optimal cut-off point and can be seen in the ROC curves (Fig. 6), as it
449 corresponds to the point of the curve closest to point x = 0 (specificity = 1) and y = 1 (
450 sensitivity = 1).

451 The threshold values calculated were slightly different to the average value of 1.5
452 previously established as the discriminatory limit. Threshold values of 1.53 and 1.64 were
453 obtained for the tests in in-shell almonds carried out using the Aurora and MicroNIR^{TM}
454 Pro 1700 instruments, respectively. For shelled almonds, threshold values of 1.58 and
455 1.60 were obtained for those tests carried out with the Aurora and MicroNIR^{TM} Pro 1700
456 instruments, respectively.

457 Table 3 also shows the results obtained for the best classification models
458 considering the new threshold values obtained from the ROC curves to classify almonds
459 by bitterness using the two sample presentations and instruments tested. In this case, the
460 models correctly classified 74/74 sweet samples and 44/45 bitter samples in cross-
461 validation and 10/10 sweet samples and 9/10 bitter samples in external validation for the

462 group of in-shell almonds analysed with the Aurora instrument. When the in-shell
463 almonds were analysed using the MicroNIR™ Pro 1700, 72/74 sweet samples and 44/45
464 bitter samples were well-classified in cross validation, while 9/10 sweet samples and
465 10/10 bitter samples were well-classified in external validation. The models developed
466 with shelled almonds produced 100 % of correctly classified samples in all cases.

467 The displacement of threshold values towards the sweet class when using the
468 optimum threshold value that maximizes the sensitivity and specificity allowed to obtain
469 a larger number of correctly classified samples in the cross-validation of the model
470 developed using in-shell almonds analysed with the MicroNIR™ Pro 1700. This, in turn,
471 enabled to obtain a higher NER value for cross-validation. It is also important to note that
472 the displacement of the threshold value enabled to minimise the number of poorly-
473 classified bitter samples in this model in cross-validation, where the discrimination
474 capacity for bitter almonds was worst affected, and therefore, the specificity of the model
475 improved. In turn, the sensitivity for the external validation collective of in-shell almonds
476 analysed with MicroNIR™ Pro 1700 was lower when using the threshold value obtained
477 from the ROC curves: in this case, one sweet sample was classified as bitter, although it
478 presented a predicted value very close to the established limit.

479 For both threshold strategies, the results showed that the shape of the in-shell
480 almonds made the surface of the samples on which the NIRS analysis was carried out less
481 homogeneous than in the case of the shelled almonds, making it more difficult to analyse
482 the in-shell almonds, so the discrimination capacity of models developed was inferior for
483 the in-shell product.

484 The results obtained are of great importance for the sweet almond processing
485 industry, which produces almonds for consumption as snacks as well as for making
486 cakes/desserts, since they allow to eliminate the presence of bitter almonds from the

marketing channels quickly, and at reduced cost. Further studies could be focused on the detection of bitter almonds that could be mixed with sweet almonds in response to the high demand from the almond industry to receive batches of this product which are totally free of bitter almonds.

The predictive capacity of the models developed in this research work was superior to that of those carried out by Borrás et al. (2014), who reported 99.2 % and 96.7 % of correctly-classified bitter and sweet shelled intact almonds, respectively, using PLS-DA for the external validation set. These results were the same as in Cortés et al. (2018a), who also reported 100 % classification accuracy for the external validation sets of sweet and bitter almonds using PLS-DA. The former worked with a FT-NIR MB160PH Aridzone instrument in the 1000-2500 nm spectral range, while the latter used a AvaSpec-NIR256-1.7 NIRLine instrument, both of which are adequate instruments for the at-line analysis of the product. However, it should be noted that both the studies cited above were conducted using spectral information obtained from representative areas of the almond kernel when analysed individually, which is not the optimal mode of analysis for the large-scale control required at the industrial level.

Conclusions

The results obtained showed that NIRS technology can be used in routine analysis in the industry to quantify the amygdalin content of shelled almonds *in situ* with great accuracy and precision, which represents a huge advantage for the almond industry in comparison with the official methods normally used to measure this cyanogenic compound. However, the presence of the shell in the product makes it difficult to predict

511 the amygdalin content, and here, the results reflect a low predictive capacity of the
512 developed models.

513 However, the discrimination of sweet and bitter almonds based on qualitative
514 analysis strategies did allow to accurately detect bitter almonds, both in-shell and shelled.
515 The non-error rate, together with the sensitivity and specificity obtained in the
516 classification models developed, confirms the feasibility of using NIRS technology for
517 the *in situ* discrimination of these two almond classes in bulk, both in-shell and shelled,
518 thus allowing to discriminate between batches of sweet and bitter almonds when the
519 product is received in the industry and during processing.

520 In addition, it confirms the convenience of using the ROC curves to establish an
521 optimal discrimination threshold to obtain a larger number of correctly classified samples,
522 which can help improve the classification ratio for fraudulent products or products that
523 should never reach the consumer, thus increasing the reliability of safety alert systems for
524 this product.

525

526 **CRediT authorship contribution statement**

527

528 **Miguel Vega-Castellote**: Data acquisition, Methodology, Formal analysis,
529 Investigation, Software, Data curation, Validation, Writing - original draft, Writing -
530 review & editing, Visualization. **Dolores Pérez-Marín**: Conceptualization,
531 Methodology, Validation, Investigation, Resources, Writing – original draft, Writing -
532 review & editing, Visualization, Supervision, Project administration, Funding
533 acquisition. **Irina Torres:** Data acquisition, Formal analysis, Investigation, Software,
534 Data curation, Writing - original draft, Writing - review & editing, Visualization. **José**
535 **Manuel Moreno-Rojas:** Data acquisition, Methodology, Investigation, Writing -

original draft. **María-Teresa Sánchez**: Conceptualization, Methodology, Validation, Investigation, Resources, Writing – original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition.

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.

Acknowledgments

This research was carried out as part of the research project P-12018024 'Measuring the quality of almonds grown in the Guadalquivir Valley (Cordoba)', funded by Desarrollo y Aplicaciones Fitotécnicas, DAFISA. The authors are grateful to Mrs. María Carmen Fernández from the Animal Production Department for her technical assistance.

References

Arrázola, G., Sánchez-Pérez, R., Dicenta, F., Grané, N., 2012. Content of the cyanogenic glucoside amygdalin in almond seeds related to the bitterness genotype. Agron. Colomb. 30, 260–265.

559 Arrázola-Paternina, G., Dicenta-López-Higuera, F., Grané-Teruel, N., 2015. Evolution of
560 the amygdalin and prunasin content during the development of almond (*Prunus*
561 *dulcis* Miller). Rev. Fac. Agron. 32, 63–81.
562 Barceloux, D.G., 2009. Cyanogenic foods (cassava, fruit kernels, and cycad seeds). In:
563 Barceloux, D.G. (Ed.), Medical Toxicology of Natural Substances: Foods, Fungi,
564 Medicinal Herbs, Plants, and Venomous Animals. John Wiley & Sons, Inc.,
565 Hoboken, NJ, USA, pp. 44–53. http://dx.doi.org/10.1002/9780470330319.ch5.
566 Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1989. Standard Normal Variate transformation
567 and De-trending of near infrared diffuse reflectance spectra. Appl. Spectrosc. 43,
568 772–777. http://dx.doi.org/10.1366/0003702894202201.
569 Beć, K.B., Grabska, J., Siesler, H.W., Huck, C.W., 2020. Handheld near-infrared
570 spectrometer: Where are we heading? NIR news 0 (0), 1–8.
571 http://dx.doi.org/10.1177/0960336020916815.
572 Borrás, E., Amigo, J.M., Van den Berg, F., Boqué, R., Busto, O., 2014. Fast and robust
573 discrimination of almond (*Prunus amygdalus*) with respect to their bitterness by
574 using near infrared and partial least squares-discriminant analysis. Food Chem.
575 153, 15–19. http://dx.doi.org/10.1016/j.foodchem.2013.12.032.
576 Brereton, R. G., 2009. Chemometrics for Pattern Recognition. John Wiley and Sons,
577 Chichester, West Sussex, UK.
578 Cortés, V., Talens, P., Barat, J.M., Lerma-García, M.J., 2018a. Potential of NIR
579 spectroscopy to predict amygdalin content established by HPLC in intact almonds
580 and classification based on almond bitterness. Food Control 91, 68–75.
581 http://dx.doi.org/10.1016/j.foodcont.2018.03.040.
582 Cortés, V., Talens, P., Barat, J.M., Lerma-García, M.J., 2018b. A comparison between
583 NIR and ATR-FTIR spectroscopy for varietal differentiation of Spanish intact

almonds. Food Control 94, 241–248. http://dx.doi.org/10.1016/j.foodcont.2018.07.020.

Cortés, V., Blasco, J., Aleixos, N., Cubero, S., Talens, P., 2019. Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review. Trends Food Sci. Technol. 85, 138–148. http://dx.doi.org/10.1016/j.tifs.2019.01.015.

Downey, G., 2000. Discriminant PLS–questions and answers from a listserver. NIR News 11 (1), 9–12.

Firmani, P., Bucci, R., Marini, F., Biancolillo, A., 2019. Authentication of "Avola almonds" by near infrared (NIR) spectroscopy and chemometrics. J. Food Compos. Anal. 82, 103235, 1–5. http://dx.doi.org/10.1016/j.jfca.2019.103235.

Frehner, M., Scalet, M., Conn, E.E., 1990. Pattern of the cyanide potential in developing fruits. Plant Physiol. 94, 28–34. http://dx.doi.org/10.1104/pp.94.1.28.

ISI., 2000. The Complete Software Solution Using a Single Screen for Routine Analysis, Robust Calibrations, and Networking. Manual, FOSS NIRSystems/TECATOR. Infrasoft International, LLC, Sylver Spring, MD, USA.

Kesen, S., Amanpour, A., Selli, S., 2018. Comparative evaluation of the fatty acids and aroma compounds in selected Iranian nut oils. Eur. J. Lipid Sci. Technol. 120, 1800152, 1–9. http://dx.doi.org/10.1002/ejlt.201900210.

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. J. Agric. Food Chem. 61, 7754–7759. http://dx.doi.org/10.1021/jf402295u.

Liang, P.S., Slaughter, D.C., Ortega-Beltrán, A., Michailides, T.J., 2015. Detection of fungal infection in almond kernels using near-infrared reflectance spectroscopy.

609 Biosyst. Eng. 137, 64–72.
610 http://dx.doi.org/10.1016/j.biosystemseng.2015.07.010.
611 Martínez-Cagigal, V., 2020. ROC Curve. MATLAB Central File Exchange. Available at:
612 https://www.mathworks.com/matlabcentral/fileexchange/52442-roc-curve,
613 Accessed: 09/06/2020.
614 Morant, A.V., Jørgensen, K., Jørgensen, C., Paquette, S.M., Sánchez-Pérez, R., Møller,
615 B.L., Bak, S., 2008. β-Glucosidases as detonators of plant chemical defense.
616 Phytochemistry 69, 1795–1813.
617 http://dx.doi.org/10.1016/j.phytochem.2008.03.006.
618 Mouaffak, Y., Zegzouti, F., Boutbaoucht, M., Najib, M., El Adib, A.G., Sbihi, M.,
619 Younous, S., 2013. Cyanide poisoning after almond ingestion. Ann. Trop. Med.
620 Public Health 6, 679–680. http://dx.doi.org/10.4103/1755-6783.140262.
621 Naes, T., Isaksson, T., Fearn, T., Davies, A., 2002. A User-Friendly Guide to Multivariate
622 Calibration and Classification. NIR Publications, Chichester, UK.
623 Nicolaï, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I., Lammertyn,
624 J., 2007. Nondestructive measurement of fruit and vegetable quality by means of
625 NIR spectroscopy: a review. Postharvest Biol. Technol. 46, 99–118.
626 http://dx.doi.org/10.1016/j.postharvbio.2007.06.024.
627 Pasquini, C., 2018. Near infrared spectroscopy: A mature analytical technique with new
628 perspectives - A review. Anal. Chim. Acta 1026, 8–36.
629 http://dx.doi.org/10.1016/j.aca.2018.04.004.
630 Pérez-Marín, D., Garrido-Varo, A., Guerrero, J.E., 2007. Non-linear regression methods
631 in NIRS quantitative analysis. Talanta 72, 28–42.
632 http://dx.doi.org/10.1016/j.talanta.2006.10.036.

633 Rogel-Castillo, C., Boulton, R., Opastpongkarn, A., Huang, G., Mitchell, A.E., 2016. Use
634 of near-infrared spectroscopy and chemometrics for the nondestructive
635 identification of concealed damage in raw almonds (*Prunus dulcis*). J. Agric. Food
636 Chem. 64, 5958–5962. http://dx.doi.org/10.1021/acs.jafc.6b01828.
637 Sánchez, M.T., Pérez-Marín, D., 2011. Non-Destructive Measurement of Fruit Quality
638 by NIR Spectroscopy. In: Advances in Post-Harvest Treatments and Fruit Quality
639 and Safety. Vázquez, M., Ramírez, J.A. (Eds.). Nova Science Publishers, Inc.,
640 New York, USA. pp. 101–163.
641 Serrano-Lourido, D., Saurina, J., Hernández-Cassou, S., Checa, A., 2012. Classification
642 and characterization of Spanish red wines according to their appellation of origin
643 based on chromatographic profiles and chemometric data analysis. Food Chem.
644 135, 1425–1431. http://dx.doi.org/10.1016/j.foodchem.2012.06.010.
645 Shenk, J.S., Westerhaus, M.O., 1991. Population structuring of near infrared spectra and
646 modified partial least squares regression. Crop Sci. 31, 1548–1555.
647 http://dx.doi.org/10.2135/cropsci1991.0011183X003100060034x.
648 Shenk, J.S., Westerhaus, M.O., 1995a. Analysis of Agriculture and Food Products by
649 Near Infrared Reflectance Spectroscopy. Monograph. NIRSystems, Inc., Silver
650 Spring, MD, USA.
651 Shenk, J.S., Westerhaus, M.O., 1995b. Routine Operation, Calibration, Development and
652 Network System Management Manual. NIRSystems, Inc., Silver Spring, MD,
653 USA.
654 Shenk, J.S., Westerhaus, M.O., 1996. Calibration the ISI way. In: Davies, A.M.C.,
655 Williams, P.C. (Eds.), Near Infrared Spectroscopy: The Future Waves. NIR
656 Publications, Chichester, UK. pp. 198–202.

Shenk, J.S., Westerhaus, M.O., Berzaghi, P., 1997. Investigation of a LOCAL calibration procedure for near infrared instruments. J. Near Infrared Spectrosc. 5, 223–232. http://dx.doi.org/10.1255/jnirs.115.

Shenk, J.S., Workman, J.J., Westerhaus, M.O., 2008. Application of NIR spectroscopy to agricultural products. In: Burns, D.A., Ciurczak, E. (Eds.), Handbook of Near-Infrared Analysis. Marcel Dekker Inc., New York, NY, USA. pp. 347–386.

Teixeira dos Santos, C.A., Lopo, M., Páscoa, R.N.M.J., Lopes, J.A., 2013. A review on the applications of portable near-infrared spectrometers in the agro-food industry. Appl. Spectrosc. 67, 1215–1233. http://dx.doi.org/10.1366/13-07228.

Tena, N., Fernández-Pierna, J.A., Boix A., Baeten, V., von Holst. C., 2014. Differentiation of meat and bone meal from fishmeal by near-infrared spectroscopy: Extension of scope to defatted samples. Food Control 43, 155–162. http://dx.doi.org/10.1016/j.foodcont.2014.03.001.

Unal, I., 2017. Defining an optimal cut-point value in ROC analysis: an alternative approach. Comput. Math. Method. M. 2017, 3762651. 1–14. http://dx.doi.org/10.1155/2017/3762651.

Williams, P.C., 2001. Implementation of near-infrared technology. In: Williams, P.C., Norris, K.H. (Eds.), Near-Infrared Technology in the Agricultural and Food Industries. AACC, Inc., St. Paul, MN, USA. pp. 145–171.

Windham, W.R., Mertens, D.R., Barton II, F.E., 1989. Protocol for NIRS calibration: sample selection and equation development and validation. In: Martens, G.C., Shenk, J.S., Barton II, F.E. (Eds.), Near Infrared Spectroscopy (NIRS): Analysis of Forage Quality. Agriculture Handbook nº643. USDA-ARS, Government Printing Office, Washington, DC. pp. 96–103.

Xu, S., Xu, X., Yuan, S., Liu, H., Liu, M., Zhang, Y., Zhang, H., Gao, Y., Lin, R., Li, X., 2017. Identification and analysis of amygdalin, neoamygdalin and amygdalin amide in different processed bitter almonds by HPLC-ESI-MS/MS and HPLC-DAD. Molecules 22, 1425–1434. http://dx.doi.org/10.3390/molecules22091425.

Yada, S., Huang, G., Lapsley, K., 2013. Natural variability in the nutrient composition of California-grown almonds. J. Food Compos. Anal. 30, 80–85. http://dx.doi.org/10.1016/j.jfca.2013.01.008.

Yan, H., Siesler, H.W., 2018. Hand-held near-infrared spectrometers: state-of-the-art instrumentation and practical applications. NIR News 29, 8–12. http://dx.doi.org/10.1177/0960336018796391.

Zhang, C., Fei, L., He, Y., 2018. Identification of coffee bean varieties using hyperspectral imaging: influence of preprocessing methods and pixel-wise spectra analysis. Sci. Rep. 8, 2166, 1–11. http://dx.doi.org/10.1038/s41598-018-20270-y.

Table 1

Calibration statistics for the best equations obtained to predict the amygdalin content (mg kg^{-1}) in in-shell and shelled almonds. MPLS regression.

Sample presentation	Instrument	Mathematical treatment	[a] N	Range	[b] Mean	[c] SD	[d] R^2_{cv}	[e] SECV	[f] RPD_{cv}
In-shell	Aurora	2,5,5,1	103	2-80980	15884	28366	0.58	18226	1.56
	MicroNIR™ Pro 1700	2,5,5,1	102	2-80980	16013	28476	0.55	19060	1.49
Shelled	Aurora	2,5,5,1	102	2-80980	16013	28476	0.95	6633	4.29
	MicroNIR™ Pro 1700	2,5,5,1	101	2-80980	15857	28574	0.96	5617	5.09

[a] Number of samples.

[b] Mean of the calibration set.

[c] Standard deviation of the calibration set.

[d] Coefficient of determination of cross validation.

[d] Standard error of cross validation.

[f] Residual predictive deviation for cross validation.

Table 2

Validation statistics for the best models to predict amygdalin content in shelled almonds using the LOCAL algorithm.

Parameter	Instrument	Math treatment	Calibration samples (*k*)	Predicted samples	Factors (*l*)	[a] SEP	[b] $SEP_{(c)}$	Bias	[c] R^2_p	[d] RPD_p	Slope
Amygdalin (mg kg^{-1})	Aurora	2,5,5,1	30	35	8 (-3)	5185	4981	1668	0.98	6.12	1.09
	MicroNIR™ Pro 1700	2,5,5,1	30	35	8 (-3)	7449	7400	858	0.95	4.32	1.05

[a] Standard error of prediction.

[b] Standard error of prediction corrected for bias.

[c] Coefficient of determination of prediction.

[d] Residual predictive deviation for prediction.

713 **Table 3.**

714 Sensitivity, specificity and non-error rate values for the classification models of in-shell

715 and shelled sweet and bitter intact almonds considering mean and ROC threshold values.

Sample presentation	Instrument		Threshold value			
			Mean value		ROC value	
			Training set	Prediction set	Training set	Prediction set
In-shell	Aurora	Sensitivity	100 %	100 %	100 %	100 %
		Specificity	98 %	90 %	98 %	90 %
		Non error rate	99 %	95 %	99 %	95 %
	MicroNIR™	Sensitivity	99 %	100 %	97 %	90 %
	Pro 1700	Specificity	94 %	100 %	98 %	100 %
		Non error rate	97 %	100 %	97 %	95 %
Shelled	Aurora	Sensitivity	100 %	100 %	100 %	100 %
		Specificity	100 %	100 %	100 %	100 %
		Non error rate	100 %	100 %	100 %	100 %
	MicroNIR™	Sensitivity	100 %	100 %	100 %	100 %
	Pro 1700	Specificity	100 %	100 %	100 %	100 %
		Non error rate	100 %	100 %	100 %	100 %

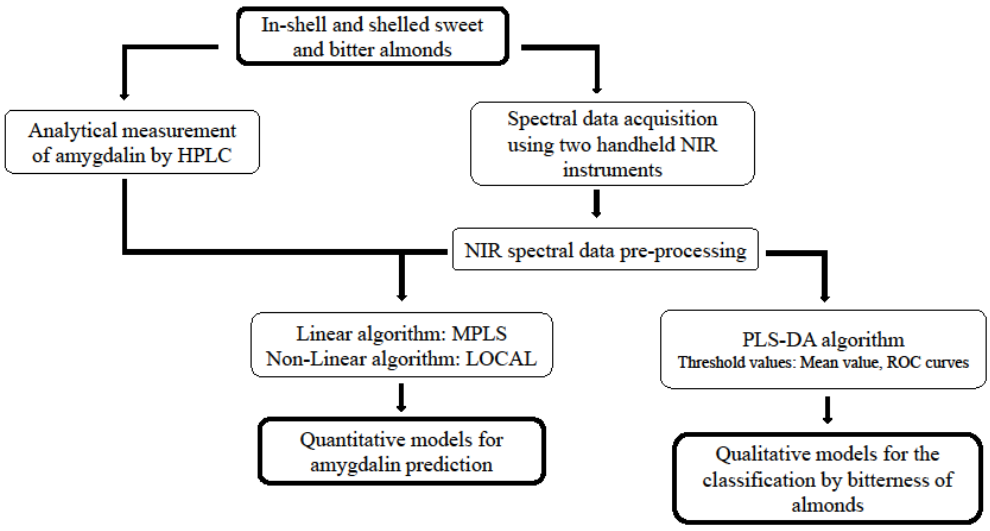
716

717

718 **Fig. 1.** Flowchart for amygdalin prediction and classification by bitterness of almonds

719 using NIRS technology

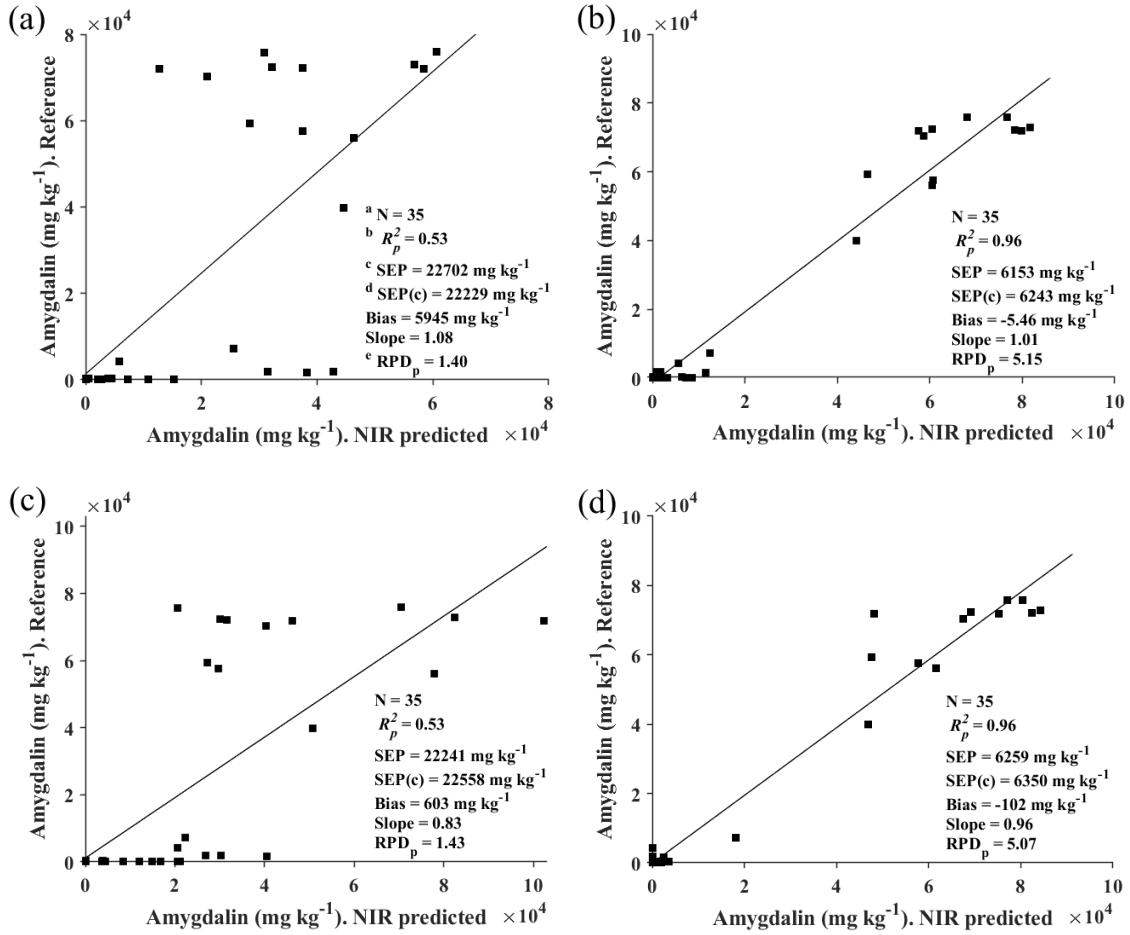
720



721

722

Fig. 2. Reference and NIR predicted values for the amygdalin content of the samples analysed in-shell (a) and shelled (b) with the Aurora instrument and of the samples analysed in-shell (c) and shelled (d) with the MicroNIR™ Pro 1700 instrument. MPLS regression.



[a] Number of samples for the validation set

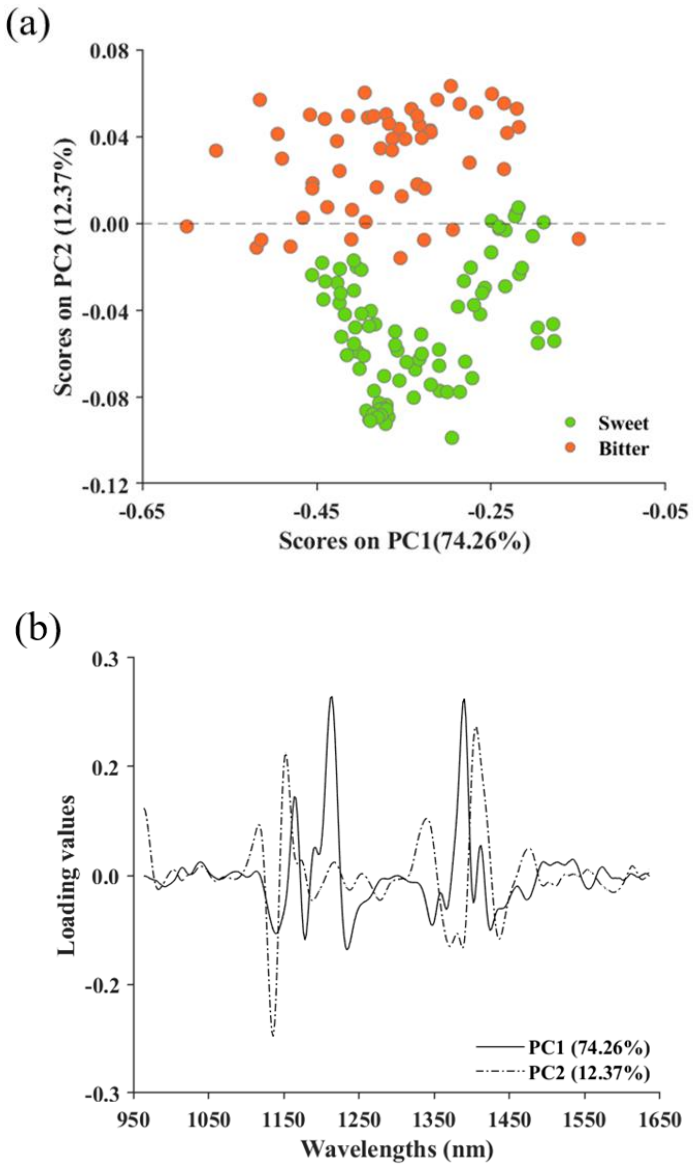
[b] Coefficient of determination of prediction.

[c] Standard error of prediction.

[d] Standard error of prediction corrected for bias.

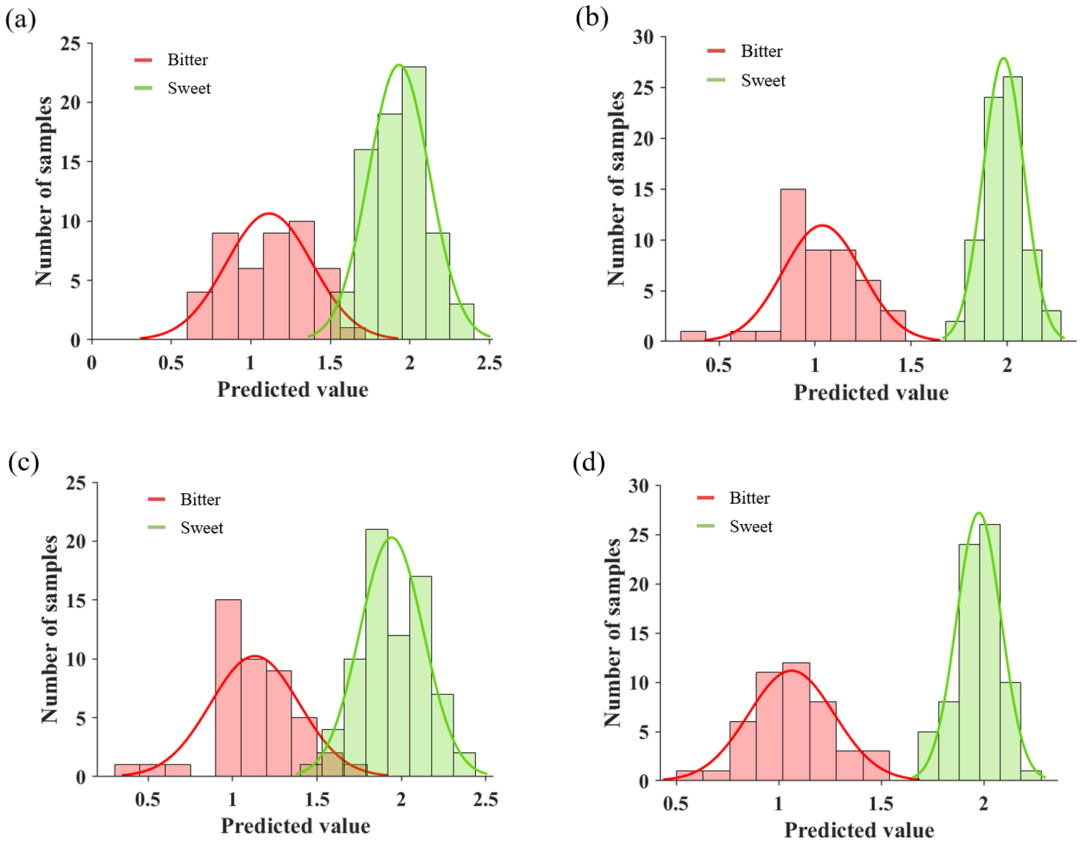
[e] Residual predictive deviation for prediction.

734 **Fig. 3.** Scores plot (a) and loading values (b) for the first (PC1) and second (PC2)
735 principal component of the shelled intact almonds analysed using the Aurora instrument.
736



737

738 **Fig. 4.** Cross validation predicted values for the sweet and bitter almonds from the four
739 sample sets tested: in-shell almonds and Aurora instrument (a), shelled almonds and
740 Aurora instrument (b), in-shell almonds and MicroNIR™ Pro 1700 instrument (c),
741 shelled almonds and MicroNIR™ Pro 1700 instrument (d).

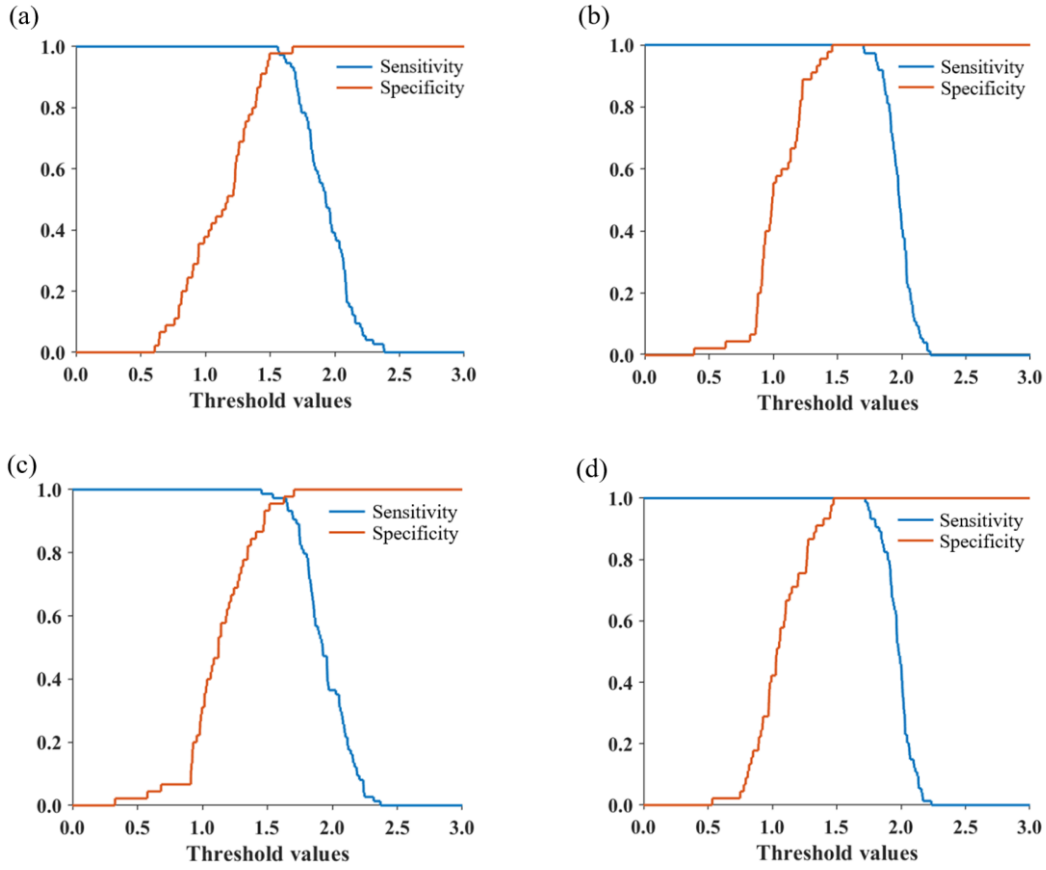


742

743

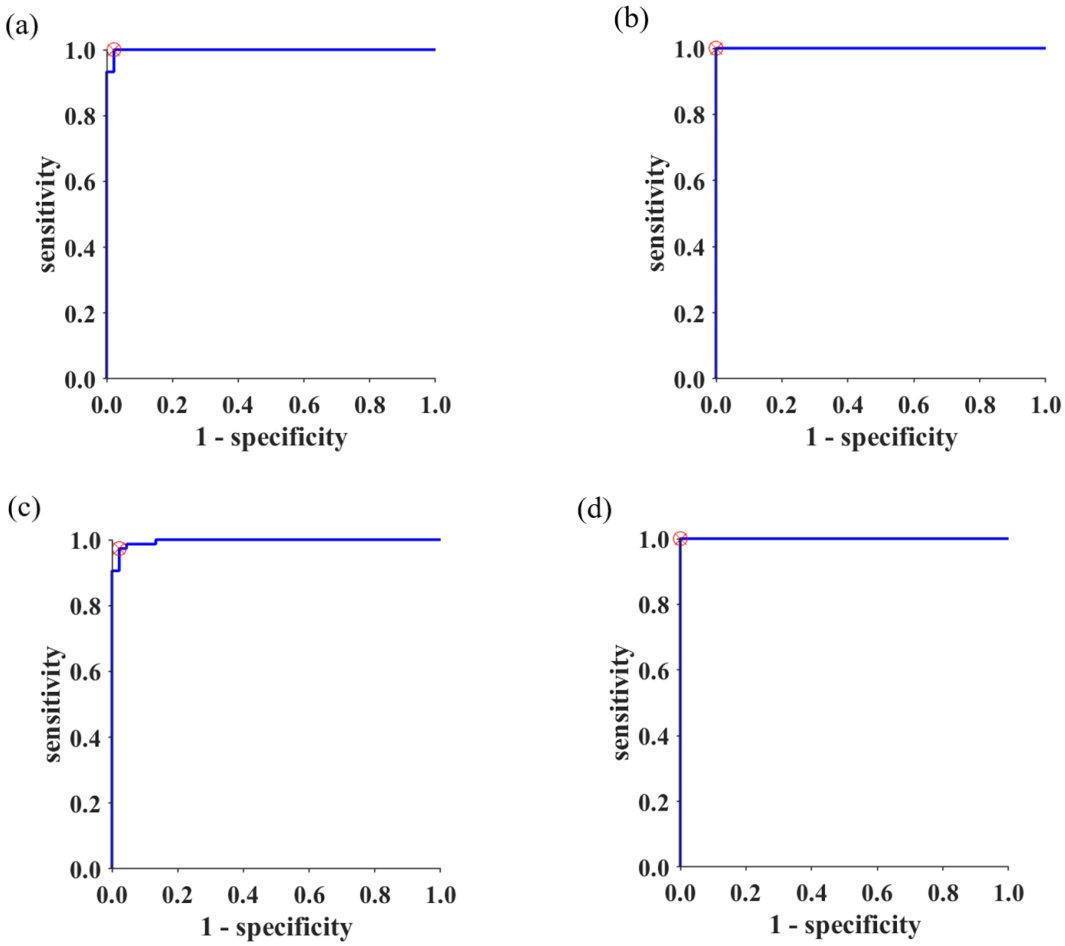
744 **Fig. 5.** Sensitivity and specificity *versus* threshold values for the samples analysed in-
745 shell (a) and shelled (b) with the Aurora instrument and the samples analysed in-shell (c)
746 and shelled (d) with the MicroNIR^{TM} Pro 1700 instrument.

747



748

749 **Fig. 6.** ROC curves and closest points to x = 0 and y = 1 for the samples analysed in-shell
750 (a) and shelled (b) with the Aurora instrument and the samples analysed in-shell (c) and
751 shelled (d) with the MicroNIR^{TM} Pro 1700 instrument.
752



753