Food Chemistry 153 (2014) 15–19

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Fast and robust discrimination of almonds (*Prunus amygdalus*) with respect to their bitterness by using near infrared and partial least squares-discriminant analysis

Eva Borràs [a,*], José Manuel Amigo [b], Frans van den Berg [b], Ricard Boqué [a], Olga Busto [a]

[a] iSens Research Group, Department of Analytical Chemistry and Organic Chemistry, Universitat Rovira i Virgili, Campus Sescelades, 43007 Tarragona, Spain
[b] Department of Food Sciences, Faculty of Sciences, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark

ARTICLE INFO

Article history:
Received 2 September 2013
Received in revised form 5 December 2013
Accepted 7 December 2013
Available online 12 December 2013

Keywords:
Sweet almonds
Bitter almonds
Prunus amygdalus
NIR
Raman
PLS-DA
Classification

ABSTRACT

In this study, near-infrared spectroscopy (NIR) coupled to chemometrics is used to develop a fast, simple, non-destructive and robust method for discriminating sweet and bitter almonds (*Prunus amygdalus*) by the in situ measurement of the kernel surface without any sample pre-treatment.

Principal component analysis (PCA) and partial least-squares discriminant analysis (PLS-DA) models were built to discriminate both types of almonds, obtaining high levels of sensitivity and specificity for both classes, with more than 95% of the samples correctly classified and discriminated. Moreover, the almonds were also analysed by Raman spectroscopy, the reference technique for this type of analysis, to validate and confirm the results obtained by NIR.

1. Introduction

Almonds (*Prunus amygdalus*) are members of the family Rosaceae and the fruit is classified as a drupe in which the edible seed or kernel is the commercial product (Gradziel, 2009). Two different species can be distinguished depending on the kernel bitterness: the bitter and the non-bitter/sweet almonds. The bitter flavour of the almond is a consequence of the presence of cyanogenic glucosides, such as amygdalin and prunasin, concentrated in the kernel. When the seed tissue is damaged an enzymatic hydrolysis (beta-glucosidases) occurs that produces benzaldehyde (that confers the bitter flavour), sugars and hydrogen cyanide (HCN), which is highly toxic, though providing an effective chemical defence against herbivores, insects and pathogens (Arrázola, Sánchez-Pérez, Dicenta, & Grané, 2012; Franks et al., 2008; Gradziel, 2009; Krafft, Cervellati, Paetz, Schneider, & Popp, 2012; Micklander, Brimer, & Engelsen, 2002; Sánchez-Pérez, Jørgensen, Olsen, Dicenta, & Møller, 2008; Sánchez-Pérez, Lindberg Møller, Olsen, & Dicenta, 2009; Sánchez-Pérez et al., 2010; Zagrobelny, Bak, & Møller, 2008)

The sweet or bitter flavour characteristic of almond kernels is an inherited monogenic trait, bitter being recessive. The sweet almond is the predominant type cultivated globally, and a constant selection in the breeding is carried out to obtain sweet kernelled almond trees, with a progressive reduction of the alleles responsible for the bitter flavours. Nevertheless, some varieties, which have been grown for years and are commercially viable, carry these alleles and can produce seedlings with bitter kernels when combined with each other (Sánchez-Pérez, Sáez Belmonte, Borch, Møller, et al., 2012).

Since cyanogenic glucosides are not found in sweet almonds its detection might be a specific indicator of bitterness in the almonds (Mirrahimi et al., 2011). Prunasin (Fig. 1a), a monoglucoside, is found in unripe almonds, and is converted to amygdalin (Fig. 1b), a diglucoside, during the ripening process.

The most commonly consumed is the sweet almond, which has been recognised as a source of nutrients (Yada, Lapsley, & Huang, 2011). Nevertheless, Bitter almonds are also used, primarily in the production of flavour extracts, being processed before consumption to remove the poisonous substances. Sometimes, though, bitter almonds are mixed up with sweet almonds, causing unpleasant taste of the final processed products and, what is more important, poisoning. Some cases of poisoning through the ingestion have been described in the literature (Shragg, Albertson, & Fisher, 1982; Sánchez-Pérez et al., 2012; Toomey, Nickum, & Flurer, 2012). One bitter almond produces from 4 to 9 mg of hydrogen cyanide and a high consumption can lead to death. Because of the potential health hazard associated with the ingestion of

* Corresponding author. Tel.: +34 977 558 496.
E-mail address: eva.borras@urv.cat (E. Borràs).

0308-8146/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.foodchem.2013.12.032

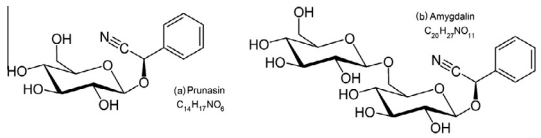

Fig. 1. Structures of the molecules of prunasin (a) and amygdalin (b).

cyanide or for possible economic adulterations it is important to differentiate between sweet and bitter almonds.

There are indirect ways to estimate the presence or content of cyanogenic compounds, based on enzymatic or chemical hydrolysis of the cyanogenic compounds to cyanide and measured with methods such as gravimetry, spectrophotometry or biosensors. These methods do not allow the identification and quantification of the specific compounds responsible for cyanide release. Therefore, direct ways should be used to detect the genuine cyanogenic substances. In consequence, separation techniques such as thin layer chromatography (TLC), high performance liquid chromatography (HPLC), gas chromatography (GC) and micellar capillary electrophoresis (MCE), coupled with detection systems like ultraviolet–visible (UV–Vis) spectrophotometry and mass spectrometry (MS) have been applied for the analysis of the cyanogenic glucosides (Dicenta et al., 2002; Gradziel, 2009; Yıldırım, San, Koyuncu, & Yıldırım, 2009). However, all these methodologies are complex, time and reagent consuming and sample destructive. Thus simpler, faster and non-destructive techniques are required.

Vibrational spectroscopy techniques like Near Infrared (NIR) (Galvez-Sola et al., 2010; Guidetti, Beghi, & Giovenzana, 2012; Roggo et al., 2007) and Raman (Micklander et al., 2002; Thygesen, Lokke, Micklander, & Engelsen, 2003) are complementary techniques that provide information on the molecular structure and are able to measure the sample in situ without any sample pre-treatment. Most of the studies related to spectroscopy are based on Raman analysis because it provides strongly intense bands containing selective chemical information. Nevertheless, Raman is usually affected by the interference from fluorescent compounds, it is more expensive than NIR and the heat generated by the laser may alter or destroy the sample during the measurement, becoming a potentially sample destructive technique (Thygesen et al., 2003). On the contrary, NIR has the advantage of being extremely fast in the measurement, robust and no alteration of the sample whatsoever is needed. The main drawback of NIR is the wide non selective bands of the spectral profile. This has promoted very few studies of almond with NIR (Galvez-Sola et al., 2010; Pearson, 1999). With the evolution of chemometrics (multivariate data analysis and modelling) NIR is gaining strong acceptance in food science and technology. Multivariate exploratory, classification and prediction methods are required to extract information from the spectral data that is related to the desired predictor (bitterness in our case).

The aim of this study is to develop a simple, fast, non-destructive and robust methodology to discriminate bitter and sweet almonds by the in situ measurement of the kernel surface without any sample pre-treatment and employing NIR spectroscopy. Principal component analysis (PCA) and partial least square discriminant analysis (PLS-DA) were applied to develop a method to classify almonds with respect to their bitterness. Raman spectroscopy was used as a reference technique to validate and confirm the results obtained by NIR.

2. Materials and methods

2.1. Almond samples

One kilogram of both bitter and sweet almonds was provided by *La Morella Nuts* (Group Barry Callebaut, A/S, Zurich). The almonds were stored in vacuum sealed in bags and in dark and cold place (fridge at 4 °C). Therefore, no significant oxidation was observed. All analyses were completed within 2 months after opening the bags in order to protect them against oxidation reactions.

2.2. Spectroscopic analysis

2.2.1. Near infrared (NIR) analysis

Spectral reflectance NIR measurements were collected by using a Bomen FT-NIR MB160PH Aridzone spectrometer equipped with an InGa detector, a reflection accessory, and the Grams/LT 7.00 data collection software programme. A representative surface of the almond with skin was directly deposited on the reflection measurement area with a complete and reproducible contact between the sample and the window. Spectra were recorded in the reflectance mode from 4100 to 10,000 cm^{-1} (1000–2500 nm), at 4 cm^{-1} resolution and taking 16 scans per sample. The absorbance was computed against a background spectrum of Spectralon. The reflection window plate was carefully cleaned with a soft tissue to eliminate the presence of residues between measurements. Two replicates were recorded for each sample.

2.2.2. Raman analysis

Raman spectra were collected with a Perkin Elmer System 2000 FT-Raman spectrometer. The radiation source was a Nd:YAG laser centred at 1064 nm with a power adjusted to 100 mW at the sample. The spectra were collected at a resolution of 8 cm^{-1} in the range of 100–3500 cm^{-1}, and represented the average of 32 scans. The reference method for Raman measurements states that the almond must be measured on the skin peeled surface (Micklander et al., 2002; Thygesen et al., 2003). Two replicates were recorded for each sample.

2.3. Data analysis

A final matrix of 320 samples and 3060 variables was built from the NIR spectra and a matrix of 78 samples and 3401 variables was obtained in the case of Raman measurements. Both NIR and Raman spectra were pre-processed with Standard Normal Variate (SNV) scaling and mean centering previous any multivariate analysis.

2.3.1. Principal component analysis (PCA)

PCA is a bilinear decomposition method that decomposes the **X** matrix of spectra into a score matrix, **T**, and a loading matrix, **P**, which describe the original data in a more condensed way, being the residuals collected in matrix **E** (Eq. (1)):

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathbf{T}} + \mathbf{E}. \quad (1)$$

The goal of PCA is to express the main information contained in the original variables in a lower number of variables, called principal components (PCs), which describe the main sources of variation in the data. These PCs are linear combinations of the original variables (Beebe, Pell, & Seasholiz, 1998). Some properties of PCs are that they are orthogonal (i.e. uncorrelated to each other), hierarchical (i.e. the first PC retains the main information of the data, the second PC retains the main information that is not included in the first, and successively), and they are calculated sequentially (Bro & Andersson, 1998).

2.3.2. Partial least squares-discriminant analysis (PLS-DA)

In PLS-DA, a PLS regression model is calculated that relates the independent variables (**X** matrix of spectra) to a vector **y** containing the codified classes as integer numbers, for instance ones (1) if the training sample belongs to a given class of interest, and zeros (0) if the sample belongs to a different class. Classification of an

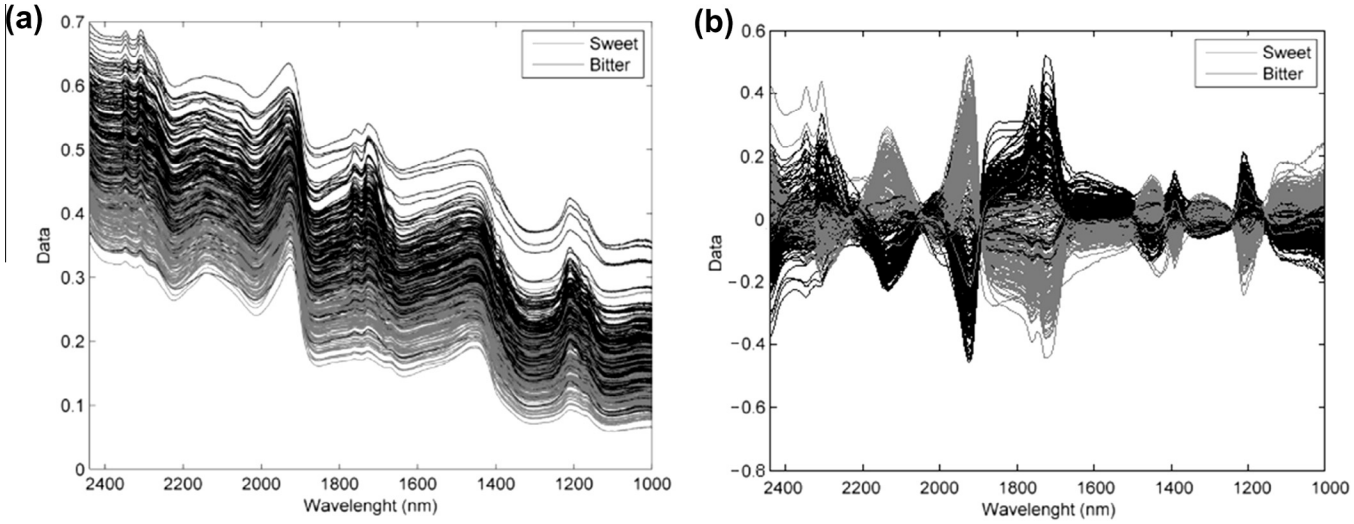

Fig. 2. (a) Raw spectra and (b) pre-processed (SNV and mean centred) spectra of all the almonds analysed by NIR (gray: sweet almonds, black: bitter almonds).

unknown sample is derived from the value predicted by the PLS model. This value is a real number, not an integer, which should ideally be close to the values used to codify the class (here either 0 or 1). A cut-off value between 0 and 1 is established so that a sample is assigned to class 1 if the prediction is larger than the cut-off value, or assigned to class 0 otherwise. The method uses the appropriate number of latent variables (or factors); which are linear combinations of the initial selected variables that maximize the discrimination among the groups (Roggo et al., 2007).

The selection of the optimal number of latent variables in PLS-DA was done using the criterion of lowest prediction error in leave-one-out cross validation (LOOCV). Additionally, the optimal PLS-DA model was further validated by predicting an external validation set. Quality assessment of the results was performed by computing the sensitivity (samples of the class of interest correctly assigned to their class), specificity (samples not belonging to the class of interest correctly not assigned to that class) and overall classification error of calibration (Cal), cross-validation (CV) and prediction (Pred).

The collected spectra were imported into MATLAB v. 7.8 (The Mathworks AS, Massachusetts, USA). Two data subsets were randomly prepared, one for calibration and another one for externally validate the calibration model. All calculations were performed using the PLS_Toolbox v. 6.2 (Eigenvector Research, Manson, WA, USA) working under MATLAB environment.

3. Results and discussion

3.1. Near infrared (NIR) analysis

A set of 160 almond samples of sweet and bitter types were measured on both sides of the outer skin and in different days, to account for the instrumental and day-to-day variability. The corresponding raw spectra are plotted in Fig. 2a. The spectra were pre-processed with Standard Normal Variate (SNV) and mean centred previous to PCA (Fig. 2b).

A preliminary data exploration with PCA was carried out in the whole dataset after spectral pre-processing. The PCA score plot shows a clear differentiation of the samples corresponding to the sweet and bitter types. The first component (PC1) is the main

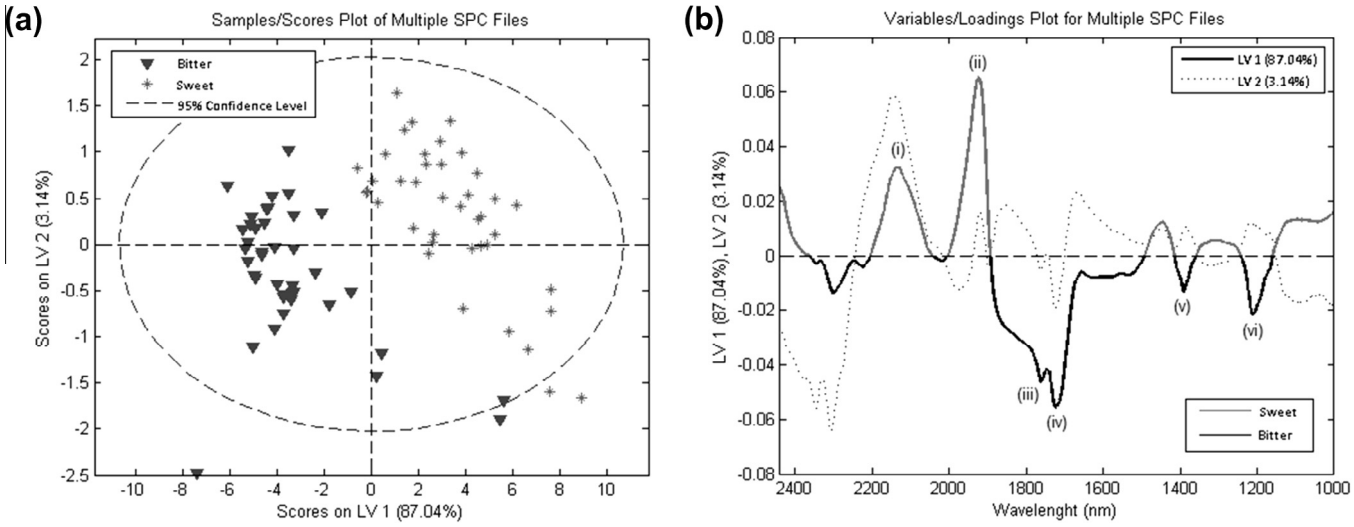

Fig. 3. (a) Scores plot and (b) loadings plot of the first two factors of the PLS-DA model built with the NIR spectra of the calibration set of almonds.

Table 1
PLS-DA model statistics for the NIR data.

Modeled class (%)	Bitter	Sweet
Sensitivity (Cal[a])	95	100
Specificity (Cal[a])	100	95
Sensitivity (CV[b])	95	100
Specificity (CV[b])	100	95
Sensitivity (Pred[c])	99.2	96.7
Specificity (Pred[c])	96.7	99.2
Class Err (Cal[a])	2.5	2.5
Class Err (CV[b])	2.5	2.5
Class Err (Pred[c])	2.1	2.1

[a] Cal: calibration.
[b] CV: cross-validation.
[c] Pred: prediction.

Table 2
PLS-DA model statistics for the Raman data.

Modeled class (%)	Bitter	Sweet
Sensitivity (Cal[a])	100	100
Specificity (Cal[a])	100	100
Sensitivity (CV[b])	100	95.7
Specificity (CV[b])	95.7	100
Sensitivity (Pred[c])	100	100
Specificity (Pred[c])	100	100
Class Err (Cal[a]):	0.0	0.0
Class Err (CV[b])	2.2	2.2
Class Err (Pred[c])	0.0	0.0

[a] Cal: calibration.
[b] CV: cross-validation.
[c] Pred: prediction.

responsible of the separation, explaining the 85.70% of the original variance. The combination of PC1 and PC4 (explaining a total of 87.45% of the original variance), shows the best separation of the groups. No significant differences between days of analysis were observed, thus indicating a robust model. PC1 loadings shows the variables related to each type of almond: negative loadings mainly describe sweet almonds and positive loadings mainly describe bitter almonds.

A PLS-DA calibration model was built by using the almonds measured on the first day (80 spectra). The model was leave-one-out cross validated and the optimal number of factors was chosen on the basis of the minimum value of the Root Mean Square Error of Cross-Validation (RMSECV). The optimal model was built with two factors (explaining a total of 90.18% of the variation in the spectra (**X**) and 81.37% of the variation in the vector of classes (y)). Using this calibration model, the other measurement set (240 almonds measured on different days) was used as external prediction set. The results from the PLS-DA model are shown in Fig. 3a (scores plot), Fig. 3b (loadings plot) and Table 1.

The PLS-DA scores plot shows an outstanding discrimination between bitter and sweet almonds. The loadings plot (Fig. 3b) shows the wavelengths responsible for the class separation. Mainly, the sweet almonds (green, positive values in factor 1) are explained by wavelengths 2135 nm (i) and 1923 nm (ii), due to O–H bend/ C–O stretch combinations and O–H stretch/HOH deformation combination (water), respectively. Bitter almonds (red, negative values for factor 1) are mainly described by wavelengths 1760 nm (iii) and 1726 nm(iv), corresponding to C–H stretch 1st overtone; 1390 nm (v) for C–H combination and 1210 nm (vi) for stretch 2nd overtone.

Table 1 shows the classification results obtained with the NIR data. High sensitivity and specificity levels were obtained for both classes, with more than 95% of the samples correctly classified as well as low classification errors (below 2.5%). The false positive/negative values are below 5% both in the cross-validation and the prediction confusion matrices.

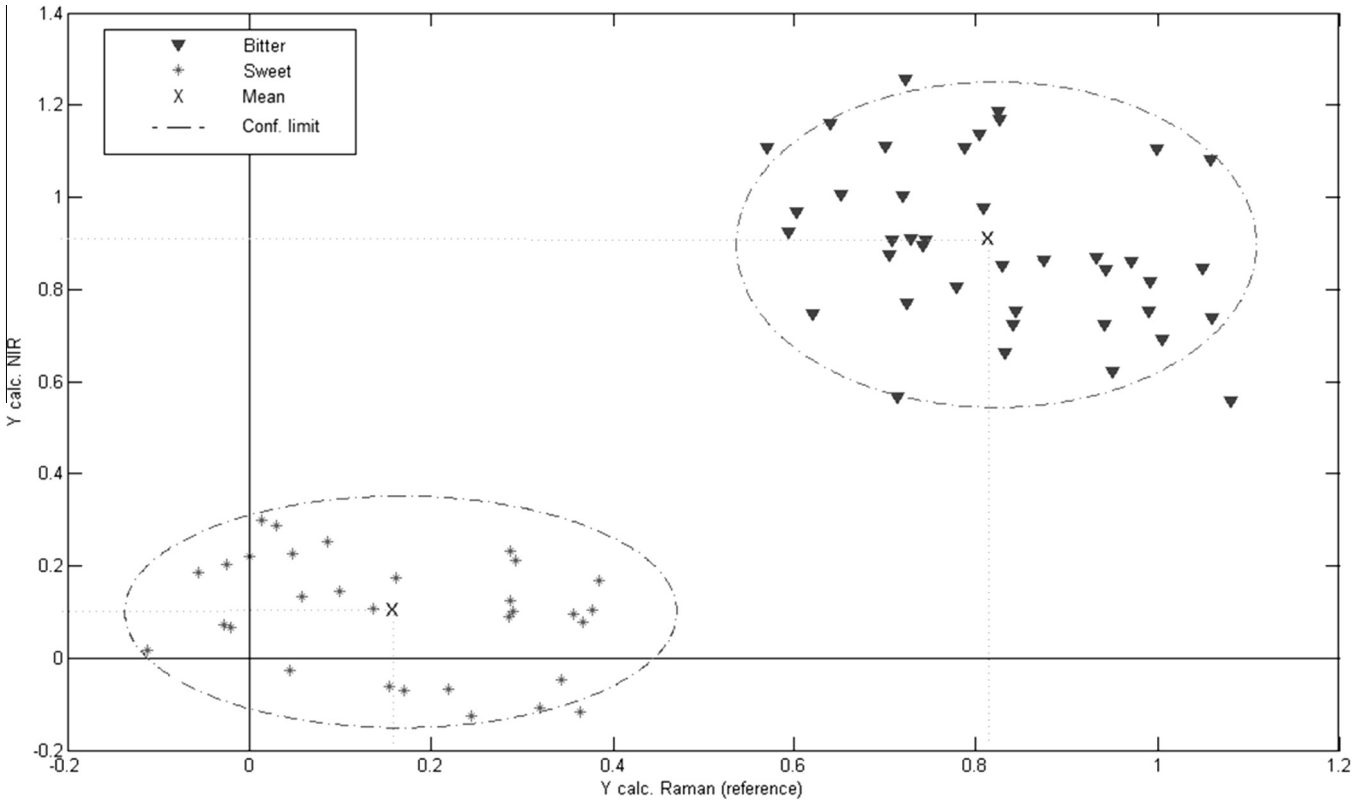

Fig. 4. Discrimination plot between sweet and bitter almonds for the two instrumental techniques used (NIR and Raman). Confidence intervals are centred on the mean of the predicted values and built with two times the standard deviation of the predictions.

3.2. Raman spectroscopy

Raman spectroscopy was applied to a set of 78 samples, both sweet (34) and bitter (44). The measurements were performed on peeled almonds, as previous studies proved to contain the higher amount of amygdalin (Micklander et al., 2002; Thygesen et al., 2003). From the collected raw spectra some of the characteristic bands related to the chemistry of amygdalin molecule were selected (previously selected by PCA). Each region was first preprocessed individually by Standard Normal Variate (SNV) and mean centred and then joined to build the final PLS-DA model.

The whole dataset was split into a training set (52 samples) and a validation set (26 samples) using Nearest Neighbour Thinning algorithm (selecting a subset of samples by removing nearest neighbours) (Shenk & Westerhaus, 1991). The optimal number of factors of the PLS-DA model, found by leave-one-out cross-validation, was two. The results from the PLS-DA model are shown in Table 2.

The model obtained explained a 44.32% of the initial variance (**X**) and 86.44% of the variation in the vector of classes (y). In the scores plot of the samples on these two factors sweet and bitter almonds are well separated. High sensitivity and specificity levels were obtained (Table 2), with a percentage of correct classification greater than 95.7% and low classification errors (below 2.2%). Low false positive/negative values (below 4.3%) were obtained both for cross-validation and prediction samples.

The wavenumbers responsible of the Raman classification correspond to the characteristic bands of the selected amygdalin regions. These results correspond to the negative values of the first factor, associated to the bitter almonds that contain this compound. Nitrile vibrational band at 2245 cm^{-1} is a highly specific signal from the cyanide. Aromatic ring bands agree with the rest of the bitter representative loadings obtained at 3060, 1604, 1031, 1003, 850 and 620 cm^{-1}.

Fig. 4 shows the predicted class of the samples of the test set for both techniques (NIR and Raman).The confidence intervals are centred at the mean value of the predictions for each class and built with two times the standard deviation of the predictions. The plot shows a good discrimination for both techniques between sweet and bitter almonds. Good class division for both techniques is achieved with clear separation of the classes.

4. Conclusion

This work demonstrates the feasibility of building a robust, fast and non-destructive methodology to discriminate between sweet and bitter almonds by combining NIR and PLS-DA obtaining an outstanding discrimination of the classes with high sensitivity and specificity levels. Raman methodology requires pre-treatment of the samples and dictates more control over measurement parameters and, therefore, more skilled operators than with NIR. NIR methodology is simpler, economically attractive and with a direct measurement on the sample it can be easily adapted and used as an automated method in industry, suitable to be implemented for quality assurance and control of raw material or final product (i.e. final packaged almonds or almonds intended for baking or other secondary products).

Acknowledgements

This research is supported by the Spanish Ministry of Science and Technology ('iSens' AGL2011-26456). This project is proposed and developed by Department of Analytical Chemistry of the University Rovira i Virgili in cooperation with the Department of Food Chemistry from the University of Copenhaguen. We thank La Morella Nuts for the samples supply.

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. *Agronomía Colombiana, 30*(2), 260–265.

Beebe, Kenneth R., Pell, Randy J., Seasholiz, Mary Beth (1998). *Chemometrics: A practical guide* (pp. 348). NY: John Wiley & Sons.

Bro, R., & Andersson, C. A. (1998). Improving the speed of multiway algorithms: Part II: Compression. *Chemometrics and Intelligent Laboratory Systems, 42*(1–2), 105–113.

Dicenta, F., Martínez-Gómez, P., Grané, N., Martín, M. L., León, A., Cánovas, J. A., et al. (2002). Relationship between cyanogenic compounds in kernels, leaves, and roots of sweet and bitter kernelled almonds. *Journal of Agricultural and Food Chemistry, 50*(7), 2149–2152.

Franks, T. K., Yadollahi, A., Wirthensohn, M. G., Guerin, J. R., Kaiser, B. N., Sedgley, M., et al. (2008). Erratum to: A seed coat cyanohydrin glucosyl transferase is associated with bitterness in almond (*Prunusdulcis*) kernels. *Functional Plant Biology, 35*(4). 346-346.

Galvez-Sola, L., Moral, R., Perez-Murcia, M. D., Perez-Espinosa, A., Bustamante, M. A., Martinez-Sabater, E., et al. (2010). The potential of near infrared reflectance spectroscopy (NIRS) for the estimation of agroindustrial compost quality. *Science of the Total Environment, 408*(6), 1414–1421.

Gradziel, T. (2009). *Almond (Prunusdulcis) breeding* (pp. 1–31). New York: Springer (pp. 1–31).

Guidetti, R., Beghi, R., & Giovenzana, V. (2012). Chemometrics in food technology. In Kurt Varmuza (Ed.), *Chemometrics in practical applications*.

Krafft, C., Cervellati, C., Paetz, C., Schneider, B., & Popp, J. (2012). Distribution of amygdalin in apricot (*Prunus armeniaca*) seeds studied by Raman microscopic imaging. *Applied Spectroscopy, 66*(6), 644–649.

Micklander, E., Brimer, L., & Engelsen, S. B. (2002). Noninvasive assay for cyanogenic constituents in plants by Raman spectroscopy: Content and distribution of amygdalin in bitter almond (*Prunus amygdalus*). *Applied Spectroscopy, 56*(9), 1139–1146.

Mirrahimi, A., Srichaikul, K., Esfahani, A., Banach, M. S., Sievenpiper, J. L., Kendall, C. W. C., et al. (2011). Chapter 18 – Almond (*Prunus dulcis*) seeds and oxidative stress. In Victor R. Preedy, Ronald Ross Watson, & Vinood B. Patel (Eds.), *Nuts and seeds in health and disease prevention* (pp. 161–166). San Diego: Academic Press.

Pearson, T. C. (1999). Use of near infrared transmittance to automatically detect almonds with concealed damage. *LWT – Food Science and Technology, 32*(2), 73–78.

Roggo, Y., Chalus, P., Maurer, L., Lema-Martinez, C., Edmond, A., & Jent, N. (2007). A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *Journal of Pharmaceutical and Biomedical Analysis, 44*(3), 683–700.

Sánchez-Pérez, R., Lindberg Møller, B., Olsen, C. E. & Dicenta, F. (2009). Cyanogenic glucoside patterns in sweet and bitter almonds. ISHS Acta Horticulturae 814: XII EUCARPIA *Symposium on fruit breeding and genetics* Vol. 814, pp. 481-486.

Sánchez-Pérez, R., Howad, W., Garcia-Mas, J., Arús, P., Martínez-Gómez, P., & Dicenta, F. (2010). Molecular markers for kernel bitterness in almond. *Tree Genetics & Genomes, 6*(2), 237–245.

Sánchez-Pérez, R., Jørgensen, K., Olsen, C. E., Dicenta, F., & Møller, B. L. (2008). Bitterness in almonds. *Plant Physiology, 146*(3), 1040–1052.

Sánchez-Pérez, R., Sáez Belmonte, F., Borch, J., Dicenta, F., Møller, B. L., et al. (2012). Prunasin hydrolases during fruit development in sweet and bitter almonds. *Plant Physiology, 158*(4), 1916–1932.

Shenk, J. S., & Westerhaus, M. O. (1991). Population definition, sample selection, and calibration procedures for near infrared reflectance spectroscopy. *Crop Science, 31*(2), 469–474.

Shragg, T. A., Albertson, T. E., & Fisher, C. J. (1982). Cyanide poisoning after bitter almond ingestion. *The Western Journal of Medicine, 136*(1), 65–69.

Thygesen, L. G., Lokke, M. M., Micklander, E., & Engelsen, S. B. (2003). Vibrational microspectroscopy of food. Raman vs. FT-IR. *Trends in Food Science and Technology, 14*(1), 50–57.

Toomey, V. M., Nickum, E. A., & Flurer, C. L. (2012). Cyanide and amygdalin as indicators of the presence of bitter almonds in imported raw almonds. *Journal of Forensic Sciences, 57*(5), 1313–1317.

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.

Yıldırım, A. N., San, B., Koyuncu, F., & Yıldırım, F. (2009). Variability of phenolics, α-tocopherol and amygdalin contents of selected almond (*Prunus amygdalus Batsch.*) genotypes. *Food, Agriculture & Environment (JFAE), 8*(1), 76–79.

Zagrobelny, M., Bak, S., & Møller, B. L. (2008). Cyanogenesis in plants and arthropods. *Phytochemistry, 69*(7), 1457–1468.