Fast detection of cocoa shell in cocoa powders by Near Infrared Spectroscopy and multivariate analysis

Maribel Alexandra Quelal-Vásconez[a], María Jesús Lerma-García[a], Édgar Pérez-Esteve[a,*], Alberto Arnau-Bonachera[b,c], José Manuel Barat[a], Pau Talens[a]

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

[b]*Institute for Animal Science and Technology. Universitat Politècnica de València. Camino de Vera, s/n 46022, Valencia, Spain*

[c]*Biomedical Research Institute (PASAPTA-Pathology group), Veterinary School, Universidad Cardenal Herrera-CEU, CEU Universities, Av. Seminario s/n, 46113 Moncada, Valencia, Spain*

* Corresponding author:

E-mail address: edpees@upv.es (Édgar Pérez-Esteve)

ABSTRACT

Cocoa shell must be removed from the cocoa bean before or after the roasting process. In the case of a low efficient peeling process or the intentional addition of cocoa shell to cocoa products (i.e. cocoa powders) to increase the economic benefit, quality of the final product could be unpleasantly affected. In this scenario, the Codex Alimentarius on cocoa and chocolate has established that cocoa cake must not contain more than 5% of cocoa shell and germ (based on fat-free dry matter). Traditional analysis of cocoa shell is very laborious. Thus, the aim of this work is to develop a methodology based on near infrared (NIR) spectroscopy and multivariate analysis for the fast detection of cocoa shell in cocoa powders. For this aim, binary mixtures of cocoa powder and cocoa shell containing increasing proportions of cocoa shell (up to ca. 40% w/w based on fat-free dried matter) have been prepared. After acquiring NIR spectra (1100-2500 nm) of pure samples (cocoa powder and cocoa shell) and mixtures, qualitative and quantitative analysis were done. The qualitative analysis was performed by using principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA), finding that the model was able to correctly classify all samples containing less than 5% of cocoa shell. The quantitative analysis was performed by using a partial least squares (PLS) regression. The best PLS model was the one constructed using extended multiple signal correction plus orthogonal signal correction pre-treatment using the 6 main wavelengths selected according to the Variable Importance in Projection (VIP) scores. Determination coefficient of prediction and root mean square error of prediction values of 0.967 and 2.43, respectively, confirmed the goodness of the model. According to these results it is possible to conclude that NIR technology in combination with multivariate analysis is a good and fast tool to determine if a cocoa powder contains a cocoa shell content out of Codex Alimentarius specifications.

Keywords:

Cocoa powder

Cocoa shell

NIR

PLS

PLS-DA

1. Introduction

Cocoa powder is a cocoa bean (Theobroma cacao) derivative largely consumed around the world due to its capacity to give color, flavor and eating pleasure to a myriad of food preparations (Dico et al., 2018).

The obtaining of cocoa powder from cocoa beans follows different steps. First of all, beans must be peeled, starting with the peeling of the bean before or after a roasting process. During the same peeling, cocoa cotyledon must be separated from cocoa shell (12-20% of the cocoa seed), yielding fragments of cotyledon, called nibs (Okiyama, Navarro, & Rodrigues, 2017). During the shelling step, shell should be perfectly separated, removing large parts of shells and leaving nib particles practically unbroken (Beckett, 2009). The performance of this procedure is very relevant since the presence of cocoa shell in cocoa beans derivatives (cocoa liquor, cocoa powder or chocolate) adversely affects the final product quality (Mendes & Lima, 2007). Concretely, it can have an influence in some characteristics of the final product such as the flavor or taste; it can also be responsible of off-flavors. Additionally, fiber content in cocoa shell is really high. Thus, it can be a problem for the grinding process, causing equipment abrasion in some cases. Bearing this in mind it is not surprising that shell content in cocoa powders is a quality parameter to be controlled. Concretely, the Codex Alimentarius establishes a maximum amount of 5% of cocoa shells in cocoa cake (based on fat-free dry matter) (Codex Alimentarius, 2016).

Analysis of cocoa shell in cocoa products might be done following the AOAC 968.10 or the 970.23 methods (Codex Alimentarius, 2016). The first method, called spiral vessel count consists of counting the spiral vessels in a defatted, grinded and digested sample with the help of a microscope adjusted to mold counting (field of view 1.382 mm at 100x)

(AOAC, 1984). The second method, called stone cell count, consists of microscope assisted counting the stone cells present in the samples after a really laborious preparation (AOAC, 1984).

Since those methods are really arduous, recent attempts to develop alternative methods have been done. Researchers from the Nestlé Research Center proposed a gas-liquid chromatography procedure based on the detection of fatty acid tryptamides (FATs) in the sample, since FATs are compounds more abundant in cocoa shells than in other parts of cocoa seed. This work, carried out with only cocoa originating from the Ivory Coast, demonstrated that it might be an appropriate tool for the determination and prediction of the shell content in cocoa liquor (Hug, Golay, Giuffrida, Dionisi, & Destaillats, 2006). In another work, Yang et al. (2015) proposed the employment of polysaccharide fingerprint established by high performance liquid chromatography followed by principal component analysis to identified cocoa powders adulterated with cocoa or other plant shells such as chestnut, longan, peanut, etc. However, only cocoa powders containing cocoa or other plant shell percentages higher than 15 and 10%, respectively, were detected using this methodology. Therefore, even when these methodologies (determination of FATs, HPLC polysaccharide fingerprint, etc.) are more sensible, accurate and faster than the methods proposed by the Codex Alimentarius, their use as routine techniques for shell content determination still have certain limitations such as the limit of detection or the fact that they need sample preparation, require specialized personnel and they are destructive. To avoid these drawbacks common in traditional chemical analysis techniques, recent attempts on developing accurate and sensitive analytic techniques based on near infrared (NIR) spectroscopy have been done. Due to the ability of NIR spectroscopy to provide a spectrum that acts as a ‘fingerprint’ distinctive of a particular sample, this technology is now widely used as a successful quality control tool (Lerma-García, Cortés, Talens &

Barat, 2018). Concretely, in the cocoa sector NIR spectroscopy has been employed for the prediction of majority (moisture, carbohydrate, fat, protein) or minority functional compounds (theobromine, catechin, organic acids, etc.) (Veselá, Barros, Synytsya, Delgadillo, Čopíková, & Coimbra, 2007; Álvarez et al., 2012; Krähmer et al., 2015) as well as for quality control (discrimination of cocoa beans according to geographical origin, prediction of cocoa powder adulterations, etc) (Teye, Huang, Dai & Chen, 2013; Quelal et al., 2018).

In this scenario, the goal of this work is the fast determination of cocoa shell content in cocoa powders in concentrations higher than the limit established by the Codex Alimentarius (5%) by means of NIR spectroscopy and a multivariate analysis.

2. Materials and methods

2.1. Cocoa powder and shell Samples

A total of 20 natural cocoa powders and 2 cocoa shells, gently provided by Olam Food Ingredients (Cheste, Spain) or purchased in the market from different origins (Ghana, Ivory Coast, Cameroon, Peru and Indonesia) were employed in this study. In order to predict the presence of cocoa shell in cocoa powders using partial least squares (PLS), binary mixtures containing cocoa powder and cocoa shell were prepared. The mixtures contained percentages of cocoa shells in cocoa powder (based on fat-free dry matter) from ca. 2.5 to 40%. Percentages higher than 40% were not considered since over this percentage the presence of cocoa shell is sensory evident. To improve the robustness of the PLS model, all 20 cocoa powder samples (coming from different origins and obtained after different processings) were randomly selected to perform a total of 12 binary

mixtures for each percentage (2.5, 5, 7.5, 10, 20 and 40%), in which both cocoa shell samples were also considered. Thus, a total of 72 mixtures were obtained. Once all mixtures were prepared, they were poured in hermetic plastic containers and stored at 20±2 ºC under dark conditions until use.

2.2. NIR spectra acquisition

The 94 samples (20 cocoa powders, 2 cocoa shells and 72 binary mixtures) were measured with a spectrophotometer FOSS NIR 5000 (Silver Spring, MD, USA). A uniform thickness and surface were secured during spectra scanning using a device with 380 mm of diameter and 1cm of thick with a quartz windows which was filled with 5 g of sample. The spectrophotometer gives the measurements in relative absorbance units (log 1/R), which could be correlated with chemical constituents (Liu, Sun, & Ouyang, 2010; Martens, Nielsen, & Engelsen, 2003). Each sample was scanned 32 times in a range comprised between 1100 and 2500 nm at 2 nm intervals (700 points). The samples were measured twice and no differences between them were found.

2.3. Statistical analysis

Spectral data were pre-treated and analysed using qualitative and quantitative models by means of the chemometric software Unscrambler v10.5 (CAMO Software AS, Oslo, Norway).

The PCA model was performed using raw data to identify different sample groups and to find and remove defective outliers (Adnan, Hörsten, Pawelzik, & Mörlein, 2017; Bro & Smilde, 2014).

The PLS was performed in order to predict the presence of cocoa shell in the cocoa

149 powders and the PLS-DA (Berrueta, Alonso, & Héberger, 2007; Prats-Montalbán, Jerez-
150 Rozo, Romañach, & Ferrer, 2012), was constructed to evaluate its capability in
151 classifying samples according to the following categories: cocoa powders containing less
152 than 5% cocoa shell (w/w), and cocoa powders containing from 5 to 40% cocoa shell
153 (w/w).

154 Both analyses were performed using the pre-treated spectra. The spectral pre-
155 treatments tried included extended multiple signal correction (EMSC) (Martens et al.,
156 2003), standard normal variation (SNV), 2^{nd} derivative with the Savitzky-Golay (S-G),
157 orthogonal signal correction (OSC) and combinations of all of them with OSC.

158 To construct both PLS and PLS-DA models, two data matrices were used. The first
159 one employed for the PCA and PLS model construction, contained the spectra of all
160 samples (N = 94) and the same 700 X-variables. In this case, all individual cocoa shell
161 percentages were considered as Y-variable. The second matrix, employed for the PLS-
162 DA model construction, included the spectra of 92 samples (in which the spectra of cocoa
163 shells were not considered since the considered categories were cocoa shell contents
164 below 5% and between 5-40%) and 700 predictors or X-variables (wavelengths), and also
165 a dependent Y-variable containing the 2 categories previously described (<5% and 5-40%
166 cocoa shell based on fat-free dry matter, w/w).

167 For both, PLS-DA and PLS models construction, the use of all spectra wavelengths
168 was considered, jointly with the use of the most important wavelengths. The PLS and the
169 score of Variable Importance in Projection (VIP) were combined together for these
170 selection (Botelho, Reis, Oliveira, & Sena, 2015).

171 To select the optimal factor number and to avoid the over-fitting of both PLS and PLS-
172 DA models, leave-one-out cross-validation was used using 70% of the data, which were
173 randomly selected. The remaining 30% of the data were used as an external validation

set.

PLS models accuracy was evaluated by the required number of latent variables (LVs), the coefficient of determination of calibration (R^2_C), RMSEC, the coefficient of determination of cross-validation (R^2_{CV}), RMSECV, the coefficient of determination for prediction (R^2_P), the root mean square error of prediction (RMSEP), the ratio of prediction deviation (RPD, which is calculated as ratio between the standard deviation of reference values in training set and RMSEP) and the bias value (which establishes the difference between experimental values and NIR predictions). Bias value can be positive (overestimating) or negative (underestimating), indicating values near to zero a minimum deviation from experimental and predicted values (Cantor, Hoag, Ellison, Khan, & Lyon, 2011).

On the other hand, the number of latent variables (LVs) for the PLS-DA model was determined by the low value of the root mean square error of calibration (RMSEC), and the root mean square error of leave-one-out cross validation (RMSECV) (Botelho et al., 2015). The PLS-DA classification performance was evaluated by sensitivity, specificity and by the non-error rate (NER). Sensitivity is the model ability related to a correct classification of the samples with different levels of cocoa shell content. The model capacity to correctly determine the samples which not correspond to the class and correctly refuse them is the specificity (Almeida, Fidelis, Barata, & Poppi, 2013). The non-error rate (NER) is the average of the sensitivities of the different categories (Manfredi, Robotti, Quasso, Mazzucco, Calabrese, & Marengo, 2018).

3. Results and discussion

3.1. Cocoa powder and shell spectra, pre-treated spectra and PCA analysis

The mean raw spectra of cocoa powders, cocoa shells and binary mixtures of them at different percentages are shown in Fig. 1a. As shown in this figure, the main bands observed appeared at 1470, 1930 and 2130 nm, although other bands at 1730, 2310 and 2350 nm were also evidenced. Although all spectra have a similar pattern of absorbance, the relative absorbance of these bands is different for the different types of samples: cocoa shell is characterized by the highest relative absorbance, which decreased when the content of cocoa shell in the samples decreased. The signal at 1470 nm correspond to the first overtone of O-H and N-H stretching which is associated with a $CONH_2$ structure (peptide) and related to a protein (Osborne, Fearn, & Hindle, 1993). The signal at 1930 nm is related with asymmetric stretching and rocking of water, weakly bounded water, proteins, and aromatics (Veselá et al., 2007), while the wavelength at 2130 nm can be assigned to N–H combination bands ($CONH_2$) (Ribeiro, Ferreira, & Salva, 2011). On the other hand, the band at 1730 nm could be assigned to the first overtone of C-H (Ribeiro et al., 2011), while 2310 and 2350 nm are mostly related to stretching and rocking vibrations of CH_2 of polysaccharides (Veselá et al., 2007).

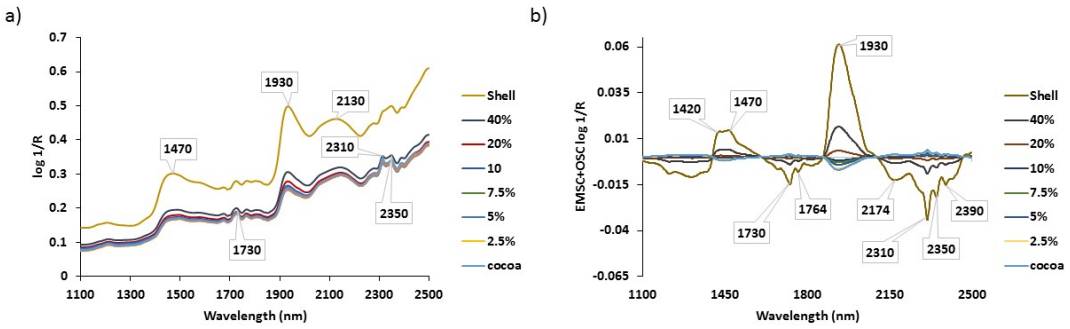

Fig 1. M.A. Quelal-Vásconez et al.

Fig 1. Mean spectra of cocoa powders and shells and mixtures of them at different percentages from (a) raw and (b) pre-treated with EMSC-OSC spectra.

The mean spectra obtained after the application of the EMSC-OSC pre-treatment is shown in Fig. 1b. In this case, the principal wavelengths were 1420, 1470, 1730, 1764, 1930, 2174, 2310, 2350 and 2390 nm. Most of the bands have been previously described, while the other ones could be attributed to the first overtones of symmetric and anti-symmetric C-H stretch vibration (CH_2-groups) (1764 nm) (Krähmer et al., 2015), to a combination of C-H (2174 nm) (Ma et al., 2017) and to the combination of C-H stretch and C-H deformation modes (2390 nm) (Wang et al., 2018).

In order to have a more precise idea about the relation between samples and variables a PCA model, a non-supervised method was performed with the raw spectra data to identify possible sample groupings. The score plot of the two first principal components (PCs) is shown in Fig. 2. A total of 98% of the variance is explained by these two first PCs (87 and 11% for PC_1 and PC_2, respectively). Along PC_1, cocoa shell samples were

clearly separated from the remaining ones, in which any clear tendency was observed, although samples containing high cocoa shell percentages (40% w/w) seemed to be located closer to the PC_1 values of cocoa shell. According to the X-loading values (data not shown), the wavelengths with higher discrimination power were 1930, 1420 and 1470 nm for the PC_1 and for the PC_2 were 1644, 1326, 2146, 2310 and 2350 nm. Some of these peaks (1930, 1470, 2310 and 2350 nm) matched with the main peaks observed in raw spectra, which have been previously mentioned. The other bands corresponded to the fist overtone of the hydroxyl and amino groups (1420 nm) and first overtone of C-H (1644 nm) (Ribeiro et al., 2011), the second overtone of C-H (1326 nm) (Ma, Wang, Chen, Cheng, & Lai, 2017) and the combination of C-C and C-H stretching (2146 nm) (Workman, & Weyer, 2008).

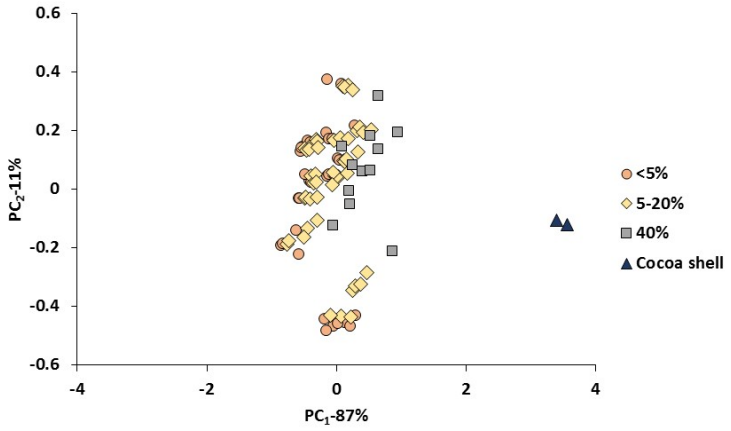

Fig 2. M.A. Quelal-Vásconez et al.

Fig 2. PCA score plot of the two first PCs showing the distribution of all the samples considered in this study. Samples were labelled as follows: cocoa shell content < 5%, comprised between 5 and 20%, 40% and pure cocoa shells.

253 *3.2. Prediction of the added cocoa shell percentage in cocoa powders by PLS*

254

255 A total of 8 PLS models using all the available wavelengths (700) as variables, one for
256 each pre-treatment considered in the study, were performed. The results obtained are
257 summarized in Table 1. At the sight of the results, the best PLS model was the one
258 constructed using the EMSC+OSC pre-treatment. In order to reduce the high
259 dimensionality of the spectral data, the most important wavelengths were selected
260 according to the VIP scores (figure 3). These VIP scores determine the significance of
261 each variable in the projection used by a given PLS model by means of their coefficients
262 in every component, jointly with the significance of each component in regression
263 (Botelho et al., 2015). As it could be observed in Fig. 3, the most important variables are
264 wavelengths at 1930, 1420 and 1470 nm at positive values of LV_1, and 2310, 2350 and
265 1730 nm at negative values of LV_1. These wavelengths are mostly the same previously
266 mentioned in both raw and pre-treated spectra, which demonstrated their importance in
267 cocoa shell content prediction. Most of these wavelengths have been previously described
268 in literature in the prediction of several compounds (such as fat, carbohydrates,
269 polysaccharides, moisture, polyphenols, etc.) of cocoa beans and derived products
270 (Huang et al., 2014; Krähmer et al., 2015; Quelal-Vásconez et al., 2018; Veselá et al.,
271 2007). Using the EMSC+OSC pre-treatment and the six wavelengths obtained in the VIP
272 scores as variables, another PLS model was constructed. The results obtained for this
273 model are also shown in Table 1. Compared to the best model obtained with the same
274 pre-treatment but using all the available wavelengths, this model is less complex although
275 all the other parameter values are very similar.

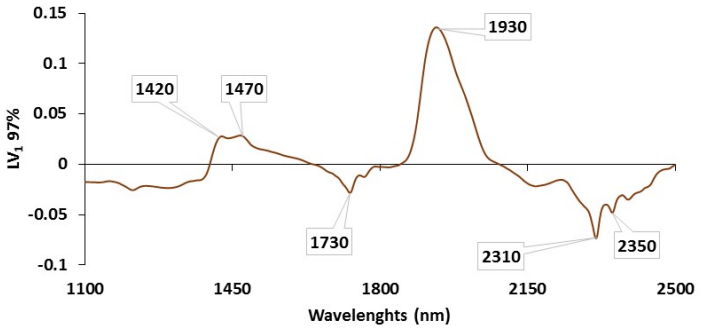

Fig. 3. M.A. Quelal-Vásconez et al.

Fig 3. Variable importance in projection (VIP) scores of the PLS model constructed to predict cocoa shell percentages.

Table 1

Results of the PLS models constructed for predicting cocoa shell percentage using different pre-treatments and different number of wavelengths with a calibration and validation sets.

Pretreatment	#W	#LV	Calibration		Cross-validation		Validation			
			R^2_C	RMSEC	R^2_{CV}	RMSECV	R^2_P	RMSEP	Bias	RPD
Raw data	700	7	0.908	3.68	0.694	6.83	0.930	3.52	0.351	3.46
EMSC	700	7	0.936	3.06	0.857	4.64	0.941	3.24	0.095	3.77
SNV	700	7	0.931	3.18	0,862	4.55	0.940	3.27	0.057	3.72
2nd Der. (S-G)	700	7	0.967	2.20	0.936	3.09	0.955	2.96	-0.021	4.11
OSC	700	1	0.990	1.20	0.989	1.25	0.851	5.16	-0.059	2.36
EMSC-OSC	*700*	*1*	*0.974*	*1.92*	*0.973*	*2.01*	*0.967*	*2.41*	*0.204*	*5.06*
SNV+OSC	700	1	0.978	1.79	0.976	1.89	0.967	2.55	-0.278	4.77
2nd Der. (S-G)+OSC	700	3	0.944	2.85	0.942	2.96	0.939	3.33	-0.104	3.66
EMSC-OSC	***6***	***1***	***0.975***	***1.91***	***0.973***	***2.01***	***0.967***	***2.43***	***0.195***	***5.03***

#W = number of wavelengths used to construct de model; #LV = latent variables; R^2 = determination coefficient; RMSEC = Root mean square error of calibration; RMSECV = Root mean square error of cross-validation; RMSEP = Root mean square error of

prediction; RPD = Ratio prediction deviation; EMSC = Extended multiple scatter correction; 2nd Der. (S-G) = Second derivative and Savitzky Golay smoothing, SNV = Standard Normal Variate, OSC = Orthogonal signal correction.

The plot representing the predicted versus the measured cocoa shell percentages of the prediction set samples constructed with PLS data of the model constructed using the 6 wavelengths as variables is shown in Fig. 4. A good linear fit due to the closer relationship between the reference values and the NIR spectra is observed, displaying the reliability and accuracy of the NIR in determining the percentage of cocoa shell present in the cocoa powders.

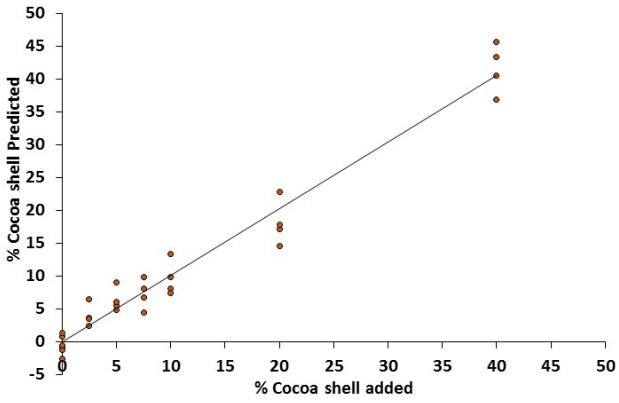

Fig. 4. M.A. Quelal-Vásconez et al.

Fig 4. Predicted versus measured cocoa shell percentages by PLS model constructed using the 6 main wavelengths in the prediction set.

304 *3.3. Classification of cocoa powder samples according to the added level of cocoa shell*

305

306 Since PCA is a non-supervised method, and it is not possible to observe a clear
307 separation between the different sample categories, a supervised discriminant model,
308 PLS-DA, was next constructed using all the available wavelengths (700) and the EMSC-
309 OSC pre-treatment. The best model was obtained with 2 LVs with RMSEC and RMSECV
310 values of 0.24 and 0.28, respectively, with most of the variability explained by the LV_1
311 (72%).

312 Next, using the 6 most relevant wavelengths as variables, another PLS-DA model was
313 constructed. The discriminant plot obtained using the two LVs for the classification of
314 samples according to the different categories is shown in Fig. 5. As it can be observed in
315 this figure, separation between the two categories is achieved along LV_1, with negative
316 scores related to the samples containing < 5% cocoa shell, and positive scores related to
317 samples containing 5-40% cocoa shell. Once constructed, the model was validated with
318 the external validation set samples. The results obtained for both calibration and external
319 validation sets for this model are included in Table 2. As it can be observed in the
320 confusion table for the calibration samples, all samples were correctly classified. On the
321 other hand, for the external validation set, all samples of the <5% category were correctly
322 classified, while 3 samples of the 5-40% category were misclassified. Even if the number
323 of misclassified samples is very low, it should be highlighted that all the "misclassified
324 samples" corresponded to samples containing a 5% cocoa shell (based on fat-free dry
325 matter), which is the limit established by the Codex Alimentarius, and thus the borderline
326 of both categories. Next, the PLS-DA classification performance was evaluated by the
327 sensitivity, specificity and NER values, which are also included in Table 2. Taking into
328 account the values reported and the comments previously mentioned, it could be

concluded that the PLS-DA model constructed is able to reliable discriminate between samples containing cocoa shell percentages below and upper 5%.

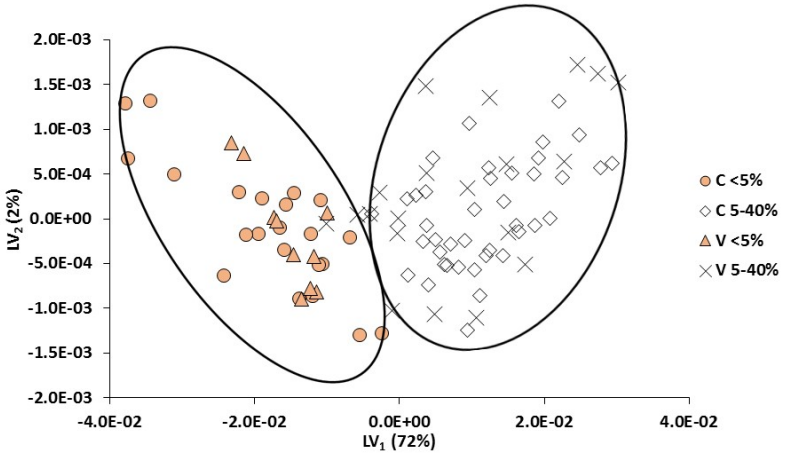

Fig. 5. M.A. Quelal-Vásconez et al.

Fig 5. PLS-DA discriminant plot constructed using the two first LVs of the model constructed using the 6 main wavelengths to classify cocoa powders according to the following categories: cocoa shell content < 5% and cocoa shell content comprised between 5 and 40%. Both calibration (C<5% and C 5-40%) and external validation (V<5% and V 5-40%) set samples have been included and represented with different symbols.

346 **Table 2**

347

348 Confusion table, sensitivity (SENS), specificity (SPEC) and non-error prediction rates
349 (NER) of the PLS-DA model constructed with variable selection to discriminate cocoa
350 powders into two categories: cocoa powders with < 5% and between 5-40% cocoa shell.

351

Calibration set samples

	Category		# Samples			
	<5%	5-40%		SENS (%)	SPEC (%)	NER (%)
<5%	22	0	22	100	100	100
5-40%	0	40	40	100	100	
	22	40	62			

352

External validation set samples

	Category		# Samples			
	<5%	5-40%		SENS (%)	SPEC (%)	NER (%)
<5%	10	0	10	100	85	92.5
5-40%	3	17	20	85	100	
	13	17	30			

353

354 **4. Conclusions**

355

356 NIR spectroscopy in combination with PLS and PLS-DA statistical models has
357 been shown to be a rapid and effective method to determine cocoa shell content in cocoa
358 powders. Using a PLS analysis, it was possible to quantify the percentage of cocoa shell
359 present in cocoa powders. The best PLS prediction model was constructed using the 6
360 main wavelengths (1420, 1470, 1730, 1930, 2310 and 2350) selected according to the
361 VIP scores, obtaining 1 LV with R^2_C and R^2_{CV} of 0.975 and 0.973, respectively, and
362 RMSEC and RMSECV of 1.91 and 2.01, respectively. Regarding the validation samples,
363 R^2_P was 0.967 while RMSEP was 2.43, confirming the goodness of the model. On the

other hand, the PLS-DA analysis show that 92.5% of the validation set samples were correctly classified into two groups: samples with a shell content lower than 5% (considered the acceptance limit in cocoa powders by the Codex Alimentarius) and shell contents between 5 and 40%. These results indicate that this technology is therefore an important tool for cocoa producers and clients, who will be able to discriminate among samples in or out specifications, avoiding the use of destructive techniques that require a complex preparation of the sample or techniques that imply an important expense for the company.

Acknowledgments

The authors wish to acknowledge the financial assistance provided the Spanish Government and European Regional Development Fund (Project RTC-2016-5241-2). M. A. Quelal thanks the Ministry of Higher Education, Science, Technology and Innovation (SENESCYT) of the Republic of Ecuador for her PhD grant. Olam Food Ingredients Company is aknowledged for providing part of the cocoa samples used in the study.

References

Adnan, A., Hörsten, D. von, Pawelzik, E., & Mörlein, D. (2017). Rapid Prediction of Moisture Content in Intact Green Coffee Beans Using Near Infrared Spectroscopy. *Foods, 6*(6), 38.

Almeida, M. R., Fidelis, C. H. V, Barata, L. E. S., & Poppi, R. J. (2013). Classification of Amazonian rosewood essential oil by Raman spectroscopy and PLS-DA with reliability estimation. *Talanta, 117*, 305–311.

Álvarez, C., Pérez, E., Cros, E., Lares, M., Assemat, S., Boulanger, R., & Davrieux, F. (2012). The use of near infrared spectroscopy to determine the fat, caffeine, theobromine and (−)-epicatechin contents in unfermented and sun-dried beans of Criollo cocoa. *Journal of Near Infrared Spectroscopy*, *20*(2), 307–315.

AOAC, 1984. AOAC, *Official methods of analysis*. Association of Official Analytical Chemists. 14th ed., Arlington, VA.

Berrueta, L. A., Alonso, R. M., & Héberger, K. (2007). Supervised pattern recognition in food analysis. *Journal of Chromatography A*, *1158*(1–2), 196–214.

Beckett, S. T. (2009). *Industrial chocolate manufacture and use*. (4th ed.). Hoboken, NJ: Wiley-Blackwell.

Botelho, B. G., Reis, N., Oliveira, L. S., & Sena, M. M. (2015). Development and analytical validation of a screening method for simultaneous detection of five adulterants in raw milk using mid-infrared spectroscopy and PLS-DA. *Food Chemistry, 181*, 31–37.

Bro, R., & Smilde, A. K. (2014). Principal component analysis. *Analytical Methods*, *6*(9), 2812–2831.

Cantor, S. L., Hoag, S. W., Ellison, C. D., Khan, M. A., & Lyon, R. C. (2011). NIR

406 Spectroscopy Applications in the Development of a Compacted Multiparticulate
407 System for Modified Release. *AAPS PharmSciTech, 12*(1), 262–278.
408 Codex Alimentarius (2016). Standard for cocoa (cacao) mass (cocoa/chocolate liquour)
409 and cocoa cake. Codex Stan 141-1983.
410 Dico, G. M. L., Galvano, F., Dugo, G., D'ascenzi, C., Macaluso, A., Vella, A., &
411 Ferrantelli, V. (2018). Toxic metal levels in cocoa powder and chocolate by ICP-MS
412 method after microwave-assisted digestion. *Food Chemistry, 245*, 1163–1168.
413 Huang, X. Y., Teye, E., Sam-Amoah, L. K., Han, F. K., Yao, L. Y., & Tchabo, W. (2014).
414 Rapid measurement of total polyphenols content in cocoa beans by data fusion of
415 NIR spectroscopy and electronic tongue. *Analytical Methods, 6*(14), 5008–5015.
416 Hug, B., Golay, P. A., Giuffrida, F., Dionisi, F., & Destaillats, F. (2006). Development
417 of a gas−liquid chromatographic method for the analysis of fatty acid tryptamides in
418 cocoa Products. *Journal of Agricultural and Food Chemistry, 54*(9), 3199–3203.
419 Krähmer, A., Engel, A., Kadow, D., Ali, N., Umaharan, P., Kroh, L. W., & Schulz, H.
420 (2015). Fast and neat – Determination of biochemical quality parameters in cocoa
421 using near infrared spectroscopy. *Food Chemistry, 181*, 152–159.
422 Lerma-García, M. J., Cortés, V., Talens, P., & Barat, J. M. (2018). Variety discrimination
423 of fruits, edible plants, and other foodstuffs and beverages by Infrared Spectroscopy.
424 In J. Lopes, & C. Sousa (eds.), *Comprehensive Analytical Chemistry volume 80* (pp.
425 127–163). Oxford: Elsevier.
426 Liu, Y., Sun, X., & Ouyang, A. (2010). Nondestructive measurement of soluble solid
427 content of navel orange fruit by visible-NIR spectrometric technique with PLSR and
428 PCA-BPNN. *LWT - Food Science and Technology, 43*(4), 602–607.
429 Ma, H. L., Wang, J. W., Chen, Y. J., & Lai, Z. T. (2017). Rapid authentication of starch
430 adulterations in ultrafine granular powder of Shanyao by near-infrared spectroscopy

coupled with chemometric methods. *Food chemistry*, *215*, 108-115.

Manfredi, M., Robotti, E., Quasso, F., Mazzucco, E., Calabrese, G., & Marengo, E. (2018). Fast classification of hazelnut cultivars through portable infrared spectroscopy and chemometrics. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, *189*, 427–435.

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.

Mendes, F. A. T., & Lima, E. L. (2007). Perfil Agroindustrial do Processamento de Amêndoas de Cacau em Pequena Escala no Estado do Pará. SEBRAE/PA.

Osborne, B. G., Fearn, T., & Hindle, P. H. (1993). *Practical NIR spectroscopy with applications in food and beverage analysis*. Longman Scientific & Technical, Wiley, Harlow, Essex, England, New York.

Okiyama, D. C., Navarro, S. L., & Rodrigues, C. E. (2017). Cocoa shell and its compounds: Applications in the food industry. *Trends in Food Science & Technology*, *63*, 103–112.

Prats-Montalbán, J. M., Jerez-Rozo, J. I., Romañach, R. J., & Ferrer, A. (2012). MIA and NIR Chemical Imaging for pharmaceutical product characterization. *Chemometrics and Intelligent Laboratory Systems*, *117*, 240–249.

Teye, E., Huang, X., Dai, H., & Chen, Q. (2013). Rapid differentiation of Ghana cocoa beans by FT-NIR spectroscopy coupled with multivariate classification. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *114*, 183–189.

Quelal-Vásconez, M. A., Pérez-Esteve, É., Arnau-Bonachera, A., Barat, J. M., & Talens,

456 P. (2018). Rapid fraud detection of cocoa powder with carob flour using near
457 infrared spectroscopy. *Food Control, 92*, 183–189.
458 Ribeiro, J. S., Ferreira, M. M. C., & Salva, T. J. G. (2011). Chemometric models for the
459 quantitative descriptive sensory analysis of Arabica coffee beverages using near
460 infrared spectroscopy. *Talanta, 83*(5), 1352–1358.
461 Veselá, A., Barros, A. S., Synytsya, A., Delgadillo, I., Čopíková, J., & Coimbra, M. A.
462 (2007). Infrared spectroscopy and outer product analysis for quantification of fat,
463 nitrogen, and moisture of cocoa powder. *Analytica Chimica Acta, 601*(1), 77–86.
464 Wang, J., Zhang, X., Sun, S., Sun, X., Li, Q., & Zhang, Z. (2018). Online determination
465 of quality parameters of dried soybean protein–lipid films (Fuzhu) by NIR
466 spectroscopy combined with chemometrics. Journal of Food Measurement and
467 Characterization, https://doi.org/10.1007/s11694-018-9762-z.
468 Workman Jr, J., & Weyer, L. (2007). *Practical guide to interpretive near-infrared*
469 *spectroscopy*. Taylor and Francis Group, Boca Raton, FL: CRC Press.
470 Yang, W., Hu, M., Chen, S., Wang, Q., Zhu, S., Dai, J., & Li, X. (2015). Identification of
471 Adulterated Cocoa Powder Using Chromatographic Fingerprints of Polysaccharides
472 Coupled with Principal Component Analysis. *Food Analytical Methods, 8*(9), 2360–
473 2367.
474