Rapid fraud detection of cocoa powder with carob flour using near infrared spectroscopy

Maribel Alexandra Quelal-Vásconez[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: pautalens@tal.upv.es

Abstract

Cocoa powder is a highly valuable global product that can be adulterated with low-cost raw materials like carob flour as small amounts of this flour would not change the color, aroma and taste characteristics of the final product. Rapid methods, like NIR technology combined with multivariate analysis, are interesting for such detection. In this work, unaltered cocoa powders with different alkalization levels, carob flours with three different roasting degrees, and adulterated samples, prepared by blending cocoa powders with carob flour at several proportions, were analyzed. The diffuse reflectance spectra of the samples of 1100 - 2500 nm were acquired in a Foss NIR spectrophotometer. A qualitative and a quantitative analysis were done. For the qualitative analysis, a principal component analysis (PCA) and a partial least squares discriminant analysis (PLS-DA) were performed. Good results (100% classification

accuracy) were obtained, which indicates the possibility of distinguishing pure cocoa powders from adulterated samples. For the quantitative analysis, a partial least squares (PLS) regression analysis was performed. The most robust PLS prediction model was obtained with one factor (LV), a coefficient of determination for prediction (R_P^2) of 0.974 and a root mean square error of prediction (RMSEP) of 3.2% for the external set. These data allowed us to conclude that NIR technology combined with multivariate analysis enables the identification and determination of the amount of natural cocoa powder present in a mixture adulterated with carob flour.

Keywords: Cocoa powder, adulteration, carob flour, NIR, PCA, PLS.

1. Introduction

Cocoa powder, thanks to its characteristic and pleasant flavor and aroma, is one of the most valued commodities around the world (Bonvehí, 2005). Among its applications in the food industry, the formulation of beverages, confectionery, bakery and pastry products stands out (Shankar, Levitan, Prescott, & Spence, 2009). Apart from flavor and aroma, cocoa is highly appreciated as a natural coloring agent, partly because of the current tendency to restrict the use of artificial colors.

During cocoa processing, it is possible to modify cocoa color and aroma through roasting and/or alkalization processes. Roasting consists of exposing cocoa beans to temperatures of 130–150°C for 15–45 min. It is used to inactivate microorganisms and to develop the characteristic brown color, mild aroma and texture of commercial natural beans (Bonvehí, 2005; Krysiak, 2006; Afoakwa, Budu, Mensah-Brown, Felix & Ofosu-Snsah, 2014). Alkalization is an optional operation to reduce acidity, bitterness and astringency, and to darken cocoa's color. This procedure involves using an alkali (generally potassium carbonate) in combination with oxygen, water and high temperatures. These extreme conditions provoke, among others, Maillard reactions and polyphenol oxidations and polymerizations, which end up with flavor and color modifications from light brown (natural) to red, dark brown or extremely black (Miller et al., 2008; Li et al., 2012).

In recent years, the demand for cocoa powder has increased and its supplies have tightened, thus its price has steadily grown (Fadel, Mageed, Samad, & Lotfy, 2006). Consequently, there is a demand to develop cocoa substitutes. Some studies suggest that cocoa-like aromas can be found in roasted carobs (Arrighi, Hartman & Ho, 1997). Carob pods are characterized by a high sugar content (around 50%), composed

61 essentially of sucrose. This high sugar content favors the same chemical reactions that
62 occur during the roasting and alkalization of cocoa: caramelization of high sugar content
63 and Maillard reactions between amino acids and sugars (Fadel et al., 2006). In this way,
64 toasted carob can provide similar aromas to cocoa.

65 Bearing in mind this striking aromatic and visual similarity between carob flour
66 (natural or toasted) and cocoa (natural or alkalized), some traders have seen that selling
67 carob (average price of 940 US$/tonne) as cocoa (1945 US$/tonne), by omitting this
68 substitution, is a profitable option to increase their benefits (ICCO, 2017). However,
69 this deliberate, intentional and undeclared substitution of one product for another with a
70 lower price is food fraud that not only affects producers and consumers, but also the
71 physico-chemical properties of the manufactured product. Some studied examples
72 comprise milk chocolates and chocolate cakes, in which some percentages of cocoa
73 powder have been substituted for carob flour (Salem & Ohaad Fahad, 2012; Rosa,
74 Tessele, Prestes, Silveira, & Franco, 2015).

75 To detect food adulteration, the three most widespread technologies are liquid
76 chromatography, infrared spectroscopy and gas chromatography (Moore, Spink, &
77 Lipp, 2012). Liquid and gas chromatography analyses need long sample preparation
78 times, method optimization, and high-cost materials and reagents, while infrared
79 spectroscopy is fast, reliable, less expensive and a chemical-free alternative (Ellis et al.,
80 2012). Near infrared spectroscopy (NIR) is an infrared spectroscopy type characterized
81 by recording reflectance or transmittance spectra within the region from 750 nm to 2500
82 nm. These spectra act as a 'fingerprint' that is characteristic of a particular sample
83 molecule and allows its identification. Some examples of using NIR and multivariate
84 analyses in the cocoa sector include the prediction of basic food components, such as
85 moisture, carbohydrate, fat, protein, theobromine and catechin and total polyphenol

content (Veselá et al., 2007; Álvarez et al., 2012; Huang et al., 2014). In other sectors, NIR in combination with a multivariate analysis has been employed to detect starch in onion powders, acid whey, starch, maltodextrin in skim powder milk, sudan dyes in chili powders, and talcum powder in teas (Lohumi, Lee, Lee, & Cho, 2014; Capuano, Boerrigter-Eenling, Koot, & van Ruth, 2015; Haughey, Galvin-King, Ho, Bell, & Elliott, 2015; Li, Zhang, & He, 2016).

In this context, the aim of this work was the rapid detection of the adulteration of cocoa powders, regardless of their alkalization level, with carob flours by applying NIR and a multivariate analysis.

2. Materials and Methods

2.1 Raw materials

In order to analyze a good representative set of samples of the variability in commercial cocoa and carob flour, cocoa powders with different alkalization levels (n=12), as well as carob flour powders with three different roasting degrees (n=6) were used in this study. The samples used were natural cocoa (NC), lightly alkalized cocoa (LAC), medium alkalized cocoa (MAC), strong alkalized cocoa (SAC), light carob flour (LCF), medium carob flour (MCF) and dark carob flour (DCF).

OLAM Food Ingredients, Spain (Cheste, Valencia), kindly donated cocoa powders. Carob flour powders were bought from a local specialized supermarket. Raw samples were placed inside a glass container and stored in a dry dark atmosphere until were used.

2.2 The physico-chemical characterization of raw materials

Each of the raw samples was characterized according to their extractable pH value and extrinsic colour. All measurements were taken in triplicate. For extractable pH determination, the process described in The Zaan Manual (Olam, 2017) was followed. For that purpose, 10g of cocoa powder were suspended in 90 mL of boiling distilled water and stirred. After decreasing temperature to 20-25ºC in a cold bath, sample pH was measured with a digital pH-meter (Crison Instruments, S.A., Barcelona, Spain) previously calibrated with 3 buffer solutions: pH 4.01, pH 7.0 and pH 9.21 (T=25 ºC). According to pH value, samples were classified in four different categories: natural cocoa powders (5< pH<6), light alkalized (6<pH<7.2), medium alkalized (7.2< pH<7.6) and strong alkalized powders (pH >7.6) (Miller et al., 2008).

To determine the extrinsic color, a cocoa powder sample was placed in a methacrylate cuvette by unifying the degree of compaction through small successive shocks. Color was measured in a spectrocolorimeter Minolta CM 3600D (Tokyo, Japan). Reflectance spectra (between 400 - 700 nm) were used to obtain color coordinates L*, a* and b* for D65 illuminant and 10° observer. Hue (h^*) and chroma (C^*) were estimated by Equation 1 and Equation2, respectively.

$$h^* = arctg\frac{b^*}{a^*} \quad (1)$$

$$C^* = \sqrt{a^{*2} + b^{*2}} \quad (2)$$

2.3 Preparing adulterated samples

In this study, two batches of 234 samples composed of 12 unaltered cocoa powders, 6 carob flours, and 216 adulterated samples, were used. The adulterated samples were prepared by blending the 12 cocoa powders with the 6 different carob flours at different proportions. For all the 72 possible cocoa-carob combinations, three different levels of adulteration were prepared: low adulteration LA (0-20%), medium adulteration MA (20-40%) and high adulteration HA (40-60%). The upper limit (60%) was set by considering that above this concentration, adulteration would become evident due to the characteristic carob aroma (Cantalejo, 1997). The specific adulteration percentage at a given level was determined randomly from a uniform distribution (each adulteration percentage had the same probability of being selected), following the Latin Hypercube Strategy (LHS) (Helton & Davis, 2003). The adulterated samples were placed in a glass container and stored in a dry dark atmosphere until used.

2.4 Collecting near-infrared spectra

All samples were scanned, by triplicate, in a FOSS NIR 5000 System spectrophotometer (Silver Spring, MD, USA) equipped with a transport module. Round sample cups (3.8 cm diameter x 1cm thick quartz windows) were filled with each sample (about 5 g) so that the surface and thickness remained uniform during spectral collection. The instrument measures diffuse reflectance and automatically converts it into relative absorbance (log 1/R) to obtain a linear correlation with the concentration of the product's chemical constituents (Martens, Nielsen, & Engelsen, 2003). Thirty-two successive scans with 700 points (wavelengths) from each sample were collected within a wavelength range from 1100 nm to 2500 nm at 2-nm intervals.

160 *2.5 Chemometric analysis*

161

162 An analysis of variance (ANOVA) was used to determine the differences in pH and
163 extrinsic color among samples. Data were statistically processed using Statgraphics
164 Centurion XVI (Manugistics Inc., Rockville, MD, USA). Simultaneously, color
165 parameters C*, h* and L* and pH were used in a principal component analysis (PCA) to
166 show the samples and their relationship. Before the analysis, an autoscaling was
167 performed in order to improve the weights of the variable with small values.
168 A multivariate analysis was conducted by a qualitative analysis and a quantitative
169 analysis by The Unscrambler v10.4 (CAMO Software AS, OSLO, Norway). For the
170 qualitative analysis, a PCA and a partial least squares discriminant analysis (PLS-DA)
171 was performed. The PCA was run with raw data, while the PLS-DA (Berrueta, Alonso,
172 & Héberger, 2007) was constructed after applying spectra 2^{nd} derivative Savitzky-Golay
173 smoothing (2^{nd} derivative S-G) (Savitzky & Golay, 1951) and orthogonal signal
174 correction (OSC). Both pre-treatments were applied to acquire useful information,
175 improve the signal-to-noise ratio and remove systematic variation from the predictor
176 matrix X unrelated, or orthogonal, to matrix Y (Wold, Antti, Lindgren, & Öhman, 1998;
177 Pizarro et al., 2004). For the quantitative analysis, a partial least squares (PLS)
178 regression analysis was performed. In order to evaluate and correct the multiplicative
179 and additive effects caused by different light scattering in the spectroscopic
180 measurement (Cozzolino et al., 2011; Stohner et al., 2012), four PLS models were
181 tested. The PLS were constructed using the raw spectrum and by applying three pre-
182 treatments to the spectrum: 2^{nd} derivative S-G, OSC and the combination of them.

183

184 *2.5.1 Developing calibration models*

Two databases were used for the analysis. The first database consisted of 468 spectra and 700 variables (wavelengths, nm) was used for the PCA and PLS models. For the PLS-DA classification samples were divided into three categories (0=Cocoa; 1=Adulterated samples and 2=Carob flour) and a second database with 142 spectra and 700 variables was created to balance the number of samples that belonged to each category. Moreover, the spectra of each database were randomly separated into two different data sets. A set with 70% of the spectra was used to create and evaluate the model by leave-one-out cross-validation. The other set, with 30% of the remaining samples, was used for external validation. The relative performance of the constructed models was assessed by the required number of latent variables (LVs), the coefficient of determination for calibration (R^2_C), the root mean square error of calibration (RMSEC), the coefficient of determination for cross validation (R^2_{CV}) and the root mean square error of leave-one-out cross validation (RMSECV). A model can be considered good when a few LVs are required, and when it has low RMSEC and RMSECV and high R^2_C and R^2_{CV}. A cut-off value of ±0.5 was used for the classification of the samples (Dong, Zhao, Hu, Dong, & Tan, 2017).

2.5.2 External validation

To assess the models' predictive capability, the coefficient of determination for prediction (R^2_P), the root mean square error of prediction (RMSEP), the ratio of prediction deviation (RPD = SD/RMSEP), where SD was the standard deviation of the Y-variable in the prediction set, and bias were used. The RPD is more meaningful than only looking at the error of prediction. An RPD value lower than 2 is considered insufficient for application, one between 2 and 2.5 is considered for approximate

quantification, and values between 2.5 and 3 are taken as a good model, while models with RPD values above 3 can be considered excellent and most reliable for analytical tasks (Sunoj, Igathinathane & Visvanathan, 2016). The bias estimates the difference between the experimental value and NIR predictions, and can be positive or negative. Positive values indicate that the model overestimates, while negative values suggest otherwise. Higher bias values indicate that NIR predictions vary significantly from the experimental values (Cantor, Hoag, Ellison, Khan, & Lyon, 2011), so it is better if it comes close to zero.

3. Results and Discussion

3.1 Raw materials characterization

Table 1 contains the color parameters and pH values of the different raw materials. As observed, the obtained pH values ranged from 5.3 (NC1) to 7.9 (SAC3). According to these values and following the Miller Classification, twelve samples were considered natural cocoas (NC; 5<pH<6), three samples light alkalized cocoas (LAC; 5<pH<6.2), tree samples medium alkalized cocoas (MAC; 7.2<pH<7.6) and three samples strong alkalized cocoas (SAC; PH > 7.6). pH can be used as an indicator of the degree of alkalization that occurs during production because the pH value of cocoa powder is related to the amount and type of alkali used in the process (OLAM Cocoa Manual, 2017; Pérez, Lerma, Fuentes, Palomares, & Barat, 2016). The inclusion of cocoas with different degrees of alkalization during the model-building phase assures that it might be used with independence of the cocoa powder processing.

234 The lightness (L*) values measured in the cocoa samples ranged from 31 (SAC1) to
235 50 (NC3). The maximum lightness value appeared in a NC sample (NC3). The L* value
236 progressively lowered according to the degree of alkalization to the minimum value in
237 the SAC samples with a very dark color. The differences in the lightness in the NC
238 samples (NC1, NC2 and NC3) could be due to a different geographical origin or to
239 distinct processing in the fermentation or roasting stages (Afoakwa, et al., 2014).
240 The chroma (C*) values oscillated between 11 (SAC1) and 22 (NC2). As seen in
241 Table 1, the higher the alkalization degree, the lower purity becomes.
242 Hue (h*), unlike the other parameters, does not follow a linear relationship with an
243 increased pH value. Cocoa samples evolve from a more yellow-orange hue (h * = 60) to
244 a more orange-red one (h* = 43) in the alkaline cocoa samples.
245 The pH of carob flours ranged from 4.5 to 5.1, with no trend observed between the
246 pH value and the degree of toasting samples. Thus, carob samples could be added to the
247 NC beans in high proportions without significantly changing the mixture's pH value.
248 The L* values in the carob flours ranged from 34 (DCF) to 49 (LCF), which meant
249 that lightness progressively lowered as the degree of roasting increased. When these
250 values were compared with those of cocoa, were found no statistical differences
251 ($p<0.05$) between the NC samples and natural carob meal (LCF samples), nor between
252 the lightness of MAC and SAC samples and roasted carob (MCF and DCF). These
253 minor differences in lightness would favor the adulteration of cocoa with carob meal.
254 The chroma (C*) of the samples also decreased as the degree of roasting rose, with
255 values of 23.7 for LCF samples that lowered to 13 for strong roasted carob (DCF
256 samples). When comparing the C* values between cocoa and carob, we found a
257 similarity between both. Thus the C* values would be the equivalent between natural

cocoa and natural carob meal, and between medium/strong cocoa beans and roasted carob.

The hue (h*) values for the carob flours gave no significant differences with an increasing degree of roasting, but only a slight decrease. The values obtained for the carob flour samples were 61 on average. These values coincided with those observed in the NC samples.

Cocoa color parameters are generally affected by several factors, including the degree of roasting and alkalization. The strong alkalized ones were dark, while the natural ones were lighter. The roasting result was darkened cocoa or carob due to the formation of brown pigments (Zyzelewicz, Krysiak, Nebesny, & Budryn, 2014), with changes noted in the values of the individual color parameters.

Insert Table 1 here

In order to know how the physico-chemical properties explained the different characteristics between the cocoa and carob flour samples, a PCA was performed with the pH and color parameters. Figure 1 shows the two-dimensional scatter plot of scores for two principal components (PCs) from projection results and allows the visualization of the distribution of the scores of the samples of cocoa and carob powder. The two PCs explain over 94% of the variation. The first PC explains 78 % of the variance and is related with the roasted degree of the carob, or alkalization of the cocoa powder, whereas the second PC explains 16 % of the total variance and is related with the difference between level of alkalized cocoa and the carob flour.

The natural cocoa NC and light carob flour LCF scores were close, which indicated that these samples were related and had similar pH characteristics and color parameters..

The positive scores on component 1 and component 2 corresponded to the samples with different degrees of alkalization. This position and the loading values of the variables led to the conclusion that the samples with low lightness and high pH were the alkaline cacao samples, while the samples with low lightness and low pH were those of roasted carob flour (Dark (DCF) and medium (MCF)). This agrees with the results presented by other authors (Bulca, 2016; Yousif & Alghzawi, 2000), which indicated that carob flour could not be visually separated from cocoa powder, not even when the other groups of the alkalized and roasting samples were blended.

Insert Figure 1 here

3.2 Spectral differences analysis of carob and cocoa powder

The spectra of the relative absorbance of cocoa powder and carob flour are represented in Figure 2 (a, b). All the cocoa spectra display a similar absorbance pattern, this pattern differs between cocoa and carob flour in relation to the absorbance intensity.

Insert Figure 2 here

Raw data were preprocessed by applying the 2nd derivative S-G and OSC. Examples of the pretreated spectra of cocoa (brown) and carob (gray) are shown in Figure 3. As observed, after this pretreatment the differences between both spectra types became more evident than in the untreated spectra. It can be stated how divergence pointed between both spectra types being located especially in the magnitude of reflectance at 1438, 1728, 2312, 2324, 2350 nm. As expected from the compositional differences,

between cocoa powder and carob flours, these wavelengths were associated with the vibration of the functional groups that cocoa powder contains like theobromine and caffeine (1728 nm) (Cozzolino et al., 2011), and epicathechin (2312, 2324 nm) (Esteban, González, & Pizarro, 2004; Teye & Huang, 2015).

Insert Figure 3 here

3.2 Classification model

A PCA was performed as a non-supervised learning algorithm with the raw spectra data to evaluate the relationship among samples. Figure 4 shows the score plot of the first two principal components (PC). The first two PC explain 91% of the total variance among the samples. The first PC explains 71% of total variance and might be related to sample processing. The different natural cocoas are found in the negative region, whereas the alkalinized samples are distributed across the negative and positive regions. These differences could be due to alkali, the stage in which it has been alkalized (bean or cake), and the degree of alkalinization that can produce different color changes (red or dark brown) (Miller et al., 2008). The second PC explains 20% of variability and might be related to the percentage of cocoa powder in the sample. Pure cocoa powders are located in the positive region, while pure carob flours are found in the negative region. The samples with different levels of adulteration lie in the middle: low (0-20%), medium (20-40%) and high (40-60%).

The wavelengths that corresponded to the highest loading values were 1100, 1464, 1936, 2108, 2276, 2330 and 2486 nm for the first PC, and 1116, 1324 1460, 1576, 1728, 1914, 1976, 2106, 2262, 2310 and 2494 nm for the second PC. The wavelengths from 971 and 1400 nm were related to the ascending part of the water first overtone

absorption peak O–H stretching bonds at 1722 nm C-H stretching was also present, which are associated with water and sugar content (Álvarez et al., 2012; Cozzolino, Smyth, & Gishen, 2003; X. Y. Huang et al., 2014; Talens et al., 2013). The wavelengths at 1736 and 2319-2328 nm were related to the absorption of the C–H bonds, CH_3 combination and C-C stretching. These are features of fatty acids, proteins and polysaccharides in cocoa powder and could be associated with a fat content of approximately 10-12% (Veselá et al., 2007; Westad, Schmidt, & Kermit, 2008). The absorption bands of 1728, 2108 and 2494 nm coincided approximately with those that were used to predict the total fat content in cocoa beans by (Ribeiro, Ferreira, & Salva, 2011; Teye & Huang, 2015). Variations were related to the compositional characteristics of the cocoa categories and the adulterant carob powder. The found wavelengths were similar to a study performed in cocoa beans (Teye et al., 2015b). Therefore, absorption in wavelengths (as a result of vibrational reactions) contains chemical information that helps explain the observed differences between the carob and cocoa powder pure samples and their several adulteration proportions.

Since the generated spectra correspond to an adulteration level on a continuous scale, it was not possible to see well-separated groups (high, medium and low adulterated) in this PCA, especially for the percentages that fell within the limits. With this information, a PLS-DA analysis was created to generate a model with categorized spectra, which allowed the detection of gross adulterations levels.

Insert Figure 4 here

As the PCA was unable to see samples in the different groups according to their adulteration percentages, a qualitative model that used the supervised PLS-DA was

358 employed. In order to improve the model's accuracy, the original spectra were pre-
359 processed using 2nd derivative S-G (9-point window, second-order polynomial) and an
360 OSC. For the PLS-DA (Figure 5), three latent variables (LVs) were generated with most
361 of the variation (67%) explained by the first LV and 12% by the second. In this way,
362 separation was achieved mainly by using the first latent variable with the most negative
363 scores for the pure cocoa samples, and the most positive scores related to the adulterate
364 samples and carob powder (pure adulterant). In visually terms, the scores plot
365 differences among the 100% cocoa powder, adulterated cocoa powders and 100% carob
366 powder indicated the possibility of using this approach to quickly screen for
367 adulteration. The determination coefficient (R^2) of this PLS-DA model was 0.969. The
368 cross-validation determination coefficient (R^2_{CV}), based on full cross-validation, was
369 0.901. Those values indicate the goodness of the classification model.
370
371 Insert Figure 5 here
372
373 In order to measure the robustness of the PLS-DA model, validation with an external
374 data set was performed. Table 2 shows the model's capability to classify 100% of the
375 samples in their corresponding groups (cocoa, carob or adulterated samples).
376
377 Insert Table 2 here
378
379 *3.3 Adulterant Prediction*
380
381 A PLS was performed with the calibration set. The prediction was done with the
382 validation set. The models were constructed by applying different pre-treatments to the

383 spectra. The statistical indicators of the goodness of fit of each model are presented in
384 Table 3.

385 Insert Table 3 here

386

387 Good models were obtained with high R^2 values and low RMSE values for the
388 calibration, cross-validation and prediction, depending on spectral data processing. The
389 RPD values were higher than 3, which meant that all these models, even the model
390 without the preprocessing data, could be considered excellent and most reliable for the
391 analytical tasks. This indicated that the multiplicative and additive effects in the spectra
392 of cocoa powder, and with the equipment used for the measurements in this study, were
393 minimal due to even the model without pretreatment to correct that effects was
394 excellent. However, it is important to point out that the models with pre-treatments
395 obtained a smaller number of LV, which made the model more parsimonious. Figure 6
396 presents the observed (x axis) *versus* predicted (y axis) values. The predicted values
397 were obtained with a model that used 2nd Derivative S-G and OSC. We can observe that
398 the PLS algorithm predicted very well with an R_{CV}^2 of 0.979 and an RMSECV of
399 2.897%. The prediction of the external validation group gave a low RMSEP of 3.237%
400 and an R_P^2 of 0.974. The similarity among RMSEC, RMSECV and RMSEP indicated
401 that the possibility of over-fitting the model was very low and confirmed its good
402 prediction capacity. The 2nd derivative S-G and the OSC pretreatment improved the
403 RPD, which was 35.48% higher compared to the PLS model with the raw data, and
404 used only one latent variable (LV). Other studies have found good models with one LV
405 when orthogonal signal correction was used (Esteban et al., 2004). The relative
406 notorious improvement of the RPD in the pretreated model could be due to the NIR
407 signal being affected by environmental (moisture) and physical factors (product's

particle size distribution). According to Huang et al., those factors generated light scattering and, consequently, significant differences arose. Additionally, these factors affect the effective sample pathlength and result in additive, multiplicative and wavelength-dependent effects. In some cases, wavelength-dependent scattering is related with baseline shifts, tilt or curvature scaling variation. In certain instances, spectra variations mask any subtle chemical variation, which can produce inaccurate results. Thus pretreatment is effective for cushioning the aforementioned effects (Huang et al., 2010).

Insert Figure 6 here

4. Conclusions

Near infrared spectroscopy (NIR) combined with PLS-DA and PLS statistical models has been shown to be a rapid effective method to identify adulterations of cocoa powder with Carob flour, regardless of the alkalization or roasting level. In contrast, these adulterations would not be readily detectable by routine techniques such as determination of pH analysis and color measurement.

With the PLS-DA analysis, all (100%) the samples were correctly classified into three groups: cocoa, carob flour and mixtures. The PLS analysis enabled the percentage of adulteration to be calculated with the samples. The PLS model was obtained with one factor with an R^2 of 0.979 and 0.974, and a mean squared error of 2.9 and 3.2 for the calibration and external validation sets, respectively.

This technique is, therefore, an important tool for cocoa merchants, who will be able to better control the product's quality by avoiding the use of destructive techniques that require complex sample preparations or techniques that imply much expense for

companies. Given the excellent results obtained, we expect this method to become increasingly important in the cocoa industry and to reduce food fraud.

Acknowledgments

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

References

Afoakwa, E. O., Budu, A. S., Mensah-brown, H., Felix, J., & Ofosu-ansah, E. (2014). Effect of Roasting Conditions on the Browning Index and Appearance Properties of Pulp Pre-Conditioned and Fermented Cocoa (Theobroma Cacao) Beans. *J Nutrition Health Food Sci*, 2(1), 1–5.

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.

Arrighi, W. J., Hartman, T. G., & Ho, C. T. (1997). Carob bean aroma dependence on roasting conditions. *Perfumer & flavorist*, 22(1), 31-41.

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

Bonvehí, J. S. (2005). Investigation of aromatic compounds in roasted cocoa powder.

458 *European Food Research and Technology*, 221(1–2), 19–29.
459 Bulca, S. (2016). Some properties of carob pod and its use. Scientific Bulletin. Series F.
460 Biotechnologies, Vol. XX, 142–147.
461 Cantalejo, M. J. (1997). Effects of Roasting Temperature on the Aroma Components of
462 Carob (Ceratonia siliqua L.). *Journal of Agricultural and Food Chemistry*, 45(4),
463 1345–1350.
464 Cantor, S.L., Hoag, S.W., Ellison, C.D., Khan, M.A., Lyon, R.C., 2011. NIR
465 spectroscopy applications in the development of a compacted multiparticulate
466 system for modified release. *AAPS PharmSciTech* 12, 262–278.
467 Capuano, E., Boerrigter-Eenling, R., Koot, A., & van Ruth, S. M. (2015). Targeted and
468 Untargeted Detection of Skim Milk Powder Adulteration by Near-Infrared
469 Spectroscopy. *Food Analytical Methods*, 8(8), 2125–2134.
470 Cozzolino, D., Smyth, H. E., & Gishen, M. (2003). Feasibility Study on the Use of
471 Visible and Near-Infrared Spectroscopy Together with Chemometrics to
472 Discriminate between Commercial White Wines of Different Varietal Origins.
473 *Journal of Agricultural and Food Chemistry*, 51(26), 7703–7708.
474 Cozzolino, D., Cynkar, W. U., Shah, N., & Smith, P. (2011). Multivariate data analysis
475 applied to spectroscopy: Potential application to juice and fruit quality. *Food*
476 *Research International*, 44(7), 1888–1896.
477 Dong, W., Zhao, J., Hu, R., Dong, Y., & Tan, L. (2017). Differentiation of Chinese
478 robusta coffees according to species, using a combined electronic nose and tongue,
479 with the aid of chemometrics. Food Chemistry, 229, 743-751.
480 Ellis, D. I., Brewster, V. L., Dunn, W. B., Allwood, J. W., Golovanov, A. P., &
481 Goodacre, R. (2012). Fingerprinting food: current technologies for the detection of
482 food adulteration and contamination. *Chemical Society Reviews*, 41(17), 5706.

Esteban, I., González, J. M., & Pizarro, C. (2004). Prediction of sensory properties of espresso from roasted coffee samples by near-infrared spectroscopy. *Analytica Chimica Acta*, 525(2), 171–182.

Fadel, H. H., Mageed, M. A. A., Samad, A. K. M. A., & Lotfy, S. N. (2006). Cocoa substitute: Evaluation of sensory qualities and flavour stability. *European Food Research and Technology*, 223(1), 125-131.

Haughey, S. A., Galvin-King, P., Ho, Y.-C., Bell, S. E. J., & Elliott, C. T. (2015). The feasibility of using near infrared and Raman spectroscopic techniques to detect fraudulent adulteration of chili powders with Sudan dye. *Food Control*, 48, 75–83.

Helton, J. C., & Davis, F. J. (2003). Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliability *Engineering and System Safety*, 81(1), 23–69.

ICCO (2017). International Cocoa Organization Dailyy Prices of cocoa Beans https://www.icco.org/statistics/cocoa-prices/daily-prices.html, Consulted: 13th January 2018.

Krysiak, W. (2006). Influence of roasting conditions on coloration of roasted cocoa beans. *Journal of Food Engineering*, *77*(3), 449-453.

Huang, J., Romero, S., & Moshgbar, M. (2010). Practical Considerations in Data Pre-treatment for NIR and Raman Spectroscopy. *American Pharmaceutical Review*, 13(6). http://www.americanpharmaceuticalreview.com. Consulted: 2 December 2017

Huang, X. Y., Teye, E., Sam-Amoah, L. K., Han, F. K., Yao, L. Y., & Tchabo, W. (2014). Rapid measurement of total polyphenols content in cocoa beans by data fusion of NIR spectroscopy and electronic tongue. *Analytical Methods*, 6(14), 5008–5015.

Li, Y., Feng, Y., Zhu, S., Luo, C., Ma, J., & Zhong, F. (2012). The effect of alkalization on the bioactive and flavor related components in commercial cocoa powder. *Journal of Food Composition and Analysis*, *25*(1), 17-23.

Li, X., Zhang, Y., & He, Y. (2016). Rapid detection of talcum powder in tea using FT-IR spectroscopy coupled with chemometrics. Scientific Reports, 6(February), 30313.

Lohumi, S., Lee, S., Lee, H., & Cho, B. K. (2015). A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. *Trends in Food Science and Technology*, 46(1), 85–98.

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.

Miller, K. B., Hurst, W. J., Payne, M. J., Stuart, D. A., Apgar, J., Sweigart, D. S., & Ou, B. (2008). Impact of alkalization on the antioxidant and flavanol content of commercial cocoa powders. *Journal of Agricultural and Food Chemistry*, 56(18), 8527–8533.

Moore, J. C., Spink, J., & Lipp, M. (2012). Development and Application of a Database of Food Ingredient Fraud and Economically Motivated Adulteration from 1980 to 2010. Journal of Food Science, 77(4).

OLAM (2017). De Zaan® Cocoa Manual. The Netherlands: Archer Daniels Midland Company BV.

Pérez, É., Lerma, M. J., Fuentes, A., Palomares, C., & Barat, J. M. (2016). Control of undeclared flavoring of cocoa powders by the determination of vanillin and ethyl vanillin by HPLC. *Food Control*, 67, 171–176.

Pizarro, C., Esteban, I., Nistal, A.-J., & González, J.-M. (2004). Influence of data pre-processing on the quantitative determination of the ash content and lipids in roasted coffee by near infrared spectroscopy. *Analytica Chimica Acta*, 509, 217–227.

Ribeiro, J. S., Ferreira, M. M. C., & Salva, T. J. G. (2011). Chemometric models for the quantitative descriptive sensory analysis of Arabica coffee beverages using near infrared spectroscopy. *Talanta*, 83(5), 1352–1358.

Rosa, C. S., Tessele, K., Prestes, R. C., Silveira, M., & Franco, F. (2015). Effect of substituting of cocoa powder for carob flour in cakes made with soy and banana flours. *International Food Research Journal*, 22(5), 2111–2118.

Salem, E. M., & Ohaad Fahad, A. A. (2012). Substituting of Cacao by Carob Pod Powder In Milk Chocolate Manufacturing. *Australian Journal of Basic and Applied Sciences*, 6(3), 572–578.

Savitzky, A., & Golay, M. (1951). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Z. Physiol. *Chem. Chem. & Ind*, 40(42).

Shankar, M. U., Levitan, C. A., Prescott, J., & Spence, C. (2009). The influence of color and label information on flavor perception. *Chemosensory Perception*, 2(2), 53–58.

Stohner, J., Lukas, B., Suter, M., Zucchetti, B., Deuber, F., & Hobi, F. (2012). NIRS of chocolate and its chemometric analysis. *Newfood*, 15(6), 21–28.

Sunoj, S., Igathinathane, C., & Visvanathan, R. (2016). Nondestructive determination of cocoa bean quality using FT-NIR spectroscopy. *Computers and Electronics in Agriculture*, 124, 234–242.

Talens, P., Mora, L., Morsy, N., Barbin, D. F., ElMasry, G., & Sun, D.-W. (2013). Prediction of water and protein contents and quality classification of Spanish cooked ham using NIR hyperspectral imaging. *Journal of Food Engineering*, 117(3), 272–280.

Teye, E., & Huang, X. (2015). Novel Prediction of Total Fat Content in Cocoa Beans by FT-NIR Spectroscopy Based on Effective Spectral Selection Multivariate Regression. *Food Analytical Methods*, 8(4), 945–953.

Teye, E., Huang, X., Sam-Amoah, L. K., Takrama, J., Boison, D., Botchway, F., & Kumi, F. (2015). Estimating cocoa bean parameters by FT-NIRS and chemometrics analysis. *Food Chemistry*, 176, 403–410.

Veselá, A., Barros, A. S., Synytsya, A., Delgadillo, I., Čopíková, J., & Coimbra, M. A. (2007). Infrared spectroscopy and outer product analysis for quantification of fat, nitrogen, and moisture of cocoa powder. *Analytica Chimica Acta*, 601(1), 77–86.

Westad, F., Schmidt, A., & Kermit, M. (2008). Incorporating chemical band-assignment in near infrared spectroscopy regression models. *Journal of Near Infrared Spectroscopy*, 16(1), 265.

Wold, S., Antti, H., Lindgren, F., & Öhman, J. (1998). Orthogonal signal correction of near-infrared spectra. *Chemometrics and Intelligent*, 44, 175–185.

Yousif, A. K., & Alghzawi, H. M. (2000). Processing and characterization of carob powder. *Food Chemistry*, 69(3), 283–287.

Zyzelewicz, D., Krysiak, W., Nebesny, E., & Budryn, G. (2014). Application of various methods for determination of the color of cocoa beans roasted under variable process parameters. *European Food Research and Technology*, 238(4), 549–563.

577 **Figure captions**

578

579 **Fig 1.** Score plot of the first and the second principal components of the PCA model
580 using color parameters L*, C*, h* and pH of pure carob and cocoa powder samples
581 (n=18, in triplicate). NC: natural Cocoa; LAC: light alkalized cocoa; MAC: medium
582 alkalized cocoa; SAC: strong alkalized cocoa; LCF: light carob flour; MCF: medium
583 carob flour; DCF: dark carob flour.

584

585 **Fig 2.** Spectra with raw data within the 1100 - 2500 nm range (a) Cocoa. (b) Carob
586 flour.

587

588 **Fig 3.** Second derivative, Savitzky Golay smoothing and orthogonal signal correction
589 pretreated the cocoa (brown) and carob (grey) spectra within the 1100 - 2500 nm range.

590

591 **Fig 4.** The NIR PCA score plot to separate pure cocoa powder and different levels of
592 adulteration with carob flour (high adulteration HA (40-60%), low adulteration LA (0-
593 20%) and medium adulteration MA (20-40%)).

594

595 **Fig 5.** NIR PLS-DA score plot from latent variable 1 and 2, pure cocoa (brown
596 triangles) carob powder (gray squares) and adulterations (blue circles).

597

598 **Fig 6.** Predicted *versus* observed values of the adulterant percentages (n=140) of the
599 pure cocoa and carob powder at different levels of adulterated samples.

Table 1

Table 1. Color parameters and pH (mean and standard deviation) values for the carob and cocoa pure samples.

Product	Color Parameters			pH ± sd
	L* ± sd	C* ± sd	h* ± sd	
LCF1	48.6±0.4^{de}	23.6±0.3^{e}	61.0±0.3^{b}	5.03±0.01^{a}
LCF2	47.70±0.06^{de}	24.1±0.2^{e}	60.98±0.11^{b}	5.12±0.01^{a}
LCF3	46.1±0.2^{de}	26.3±0.7^{e}	61.5±0.3^{b}	4.67±0.01^{a}
LCF4	44.17±0.3^{de}	20.7±0.2^{e}	61.1±0.3^{b}	4.911±0.01^{a}
MCF	37.6± 0.4^{ab}	16.9±0.7^{bc}	60.2±0.5^{a}	4.851±0.01^{a}
DCF	34.5±1.5^{a}	12.9±0.9^{a}	60±2^{a}	4.817±0.01^{a}
NC1	48.7±0.2^{e}	20.1±0.5^{de}	58.8±0.4^{c}	5.391±0.01^{a}
NC2	48.33±0.13^{e}	22.3±0.4^{de}	59.5±0.3^{c}	5.46±0.01^{b}
NC3	50.3±0.6^{e}	22.19±1.02^{de}	60.0±0.4^{c}	5.70±0.01^{b}
LAC1	42.3±0.6^{c}	22.4±0.7^{cd}	54.3±0.4^{c}	6.901±0.01^{c}
LAC2	44.2±0.5^{c}	18.63±1.02^{cd}	55.0±0.9^{c}	6.96±0.02^{c}
LAC3	41.7±0.5^{b}	19.80±0.13^{bc}	54.5±0.5^{c}	6.98±0.01^{d}
MAC1	44.9±1.5^{c}	18±2^{cd}	55.7±0.6^{c}	7.24±0.01^{c}
MAC2	41.9±0.7^{b}	18.0±0.6^{bc}	54.2±0.5^{c}	7.34±0.03^{d}
MAC3	35.85±1.05^{b}	16.0±0.8^{bc}	43.0±0.6^{c}	7.43±0.01^{d}
SAC1	32.1±0.8^{a}	11.6±0.9^{b}	46.5±0.6^{c}	7.81±0.01^{e}
SAC2	39.4±0.5^{a}	19.76±0.99^{b}	51.4±0.8^{c}	7.84±0.01^{e}
SAC3	40.1±0.2^{a}	17.3±0.8^{b}	53.2±0.6^{c}	7.92±0.01^{e}

Values in the same column followed by the same letter(s) are not significantly different according to ANOVA at a 95% Confidence level. For cocoas (N): Natural cocoa (NC), light alkalized cocoa (LAC), medium alkalized cocoa (MAC) and strong alkalized cocoa (SAC). For carob flours (A): light carob flour (LCF), medium carob flour (MCF) and dark carob flour (DCF).

Table 2. Results for classification accuracy of the PLS-DA model

	Cocoa	Carob	Adulterated	Classification
Cocoa	7	0	0	100%
Carob	0	4	0	100%
Adulterated	0	0	32	100%

Table 3

Table 3. Results of the PLS models constructed for the prediction of carob flour content in cocoa powders.

Pre-treatment	#LV	Calibration			Cross-validation				Prediction				
		R^2_C	RMSEC	Slope	R^2_{CV}	RMSECV	Bias	Slope	R^2_P	RMSEP	Bias	Slope	RPD
Raw data	7	0.951	4.5	0.951	0.945	4.8	0.019	0.951	0.961	4.4	0.197	0.968	4.7
2nd Der. S-G	4	0.969	3.6	0.969	0.965	3.8	0.013	0.968	0.969	3.9	0.711	0.998	5.2
OSC	1	0.975	3.2	0.975	0.975	3.2	0.006	0.976	0.974	3.6	0.474	0.992	5.8
2nd Der. S-G + OSC	1	0.980	2.9	0.981	0.979	2.9	0.006	0.981	0.974	3.2	0.626	1.004	6.3

2nd Der. S-G = Second derivative-Savitzky Golay; OSC = Orthogonal signal correction; #LV = latent variables; R^2_C = coefficient of determination for calibration; RMSEC = root mean square error of calibration; R^2_{CV} = coefficient of determination for cross-validation; RMSECV = root mean square error of cross-validation; R^2_P = coefficient of determination for prediction; RMSEP = root mean square error of prediction; Bias = estimation of the difference between the experimental value and NIR predictions; RPD = ratio of prediction deviation.

Figure 1

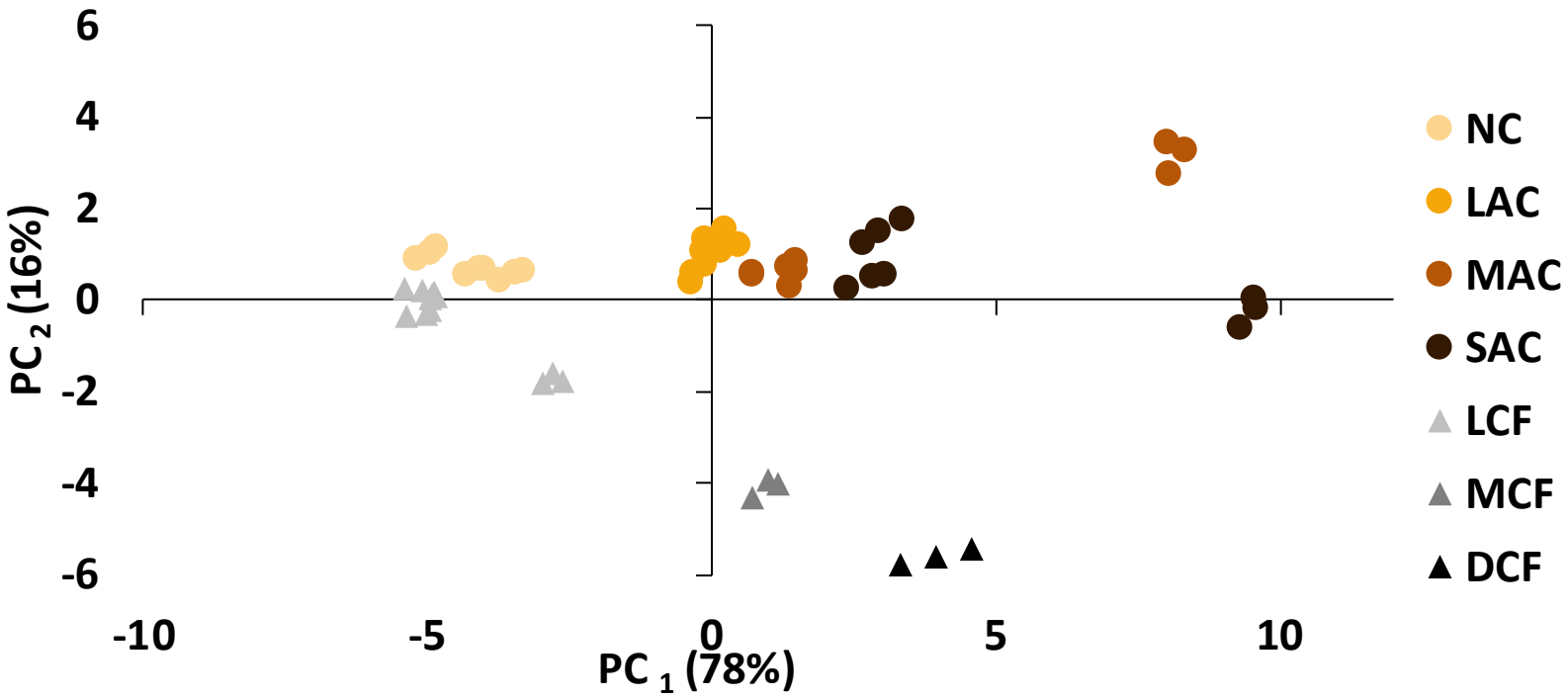

Fig 1. M.A. Quelal et al.

Figure 2

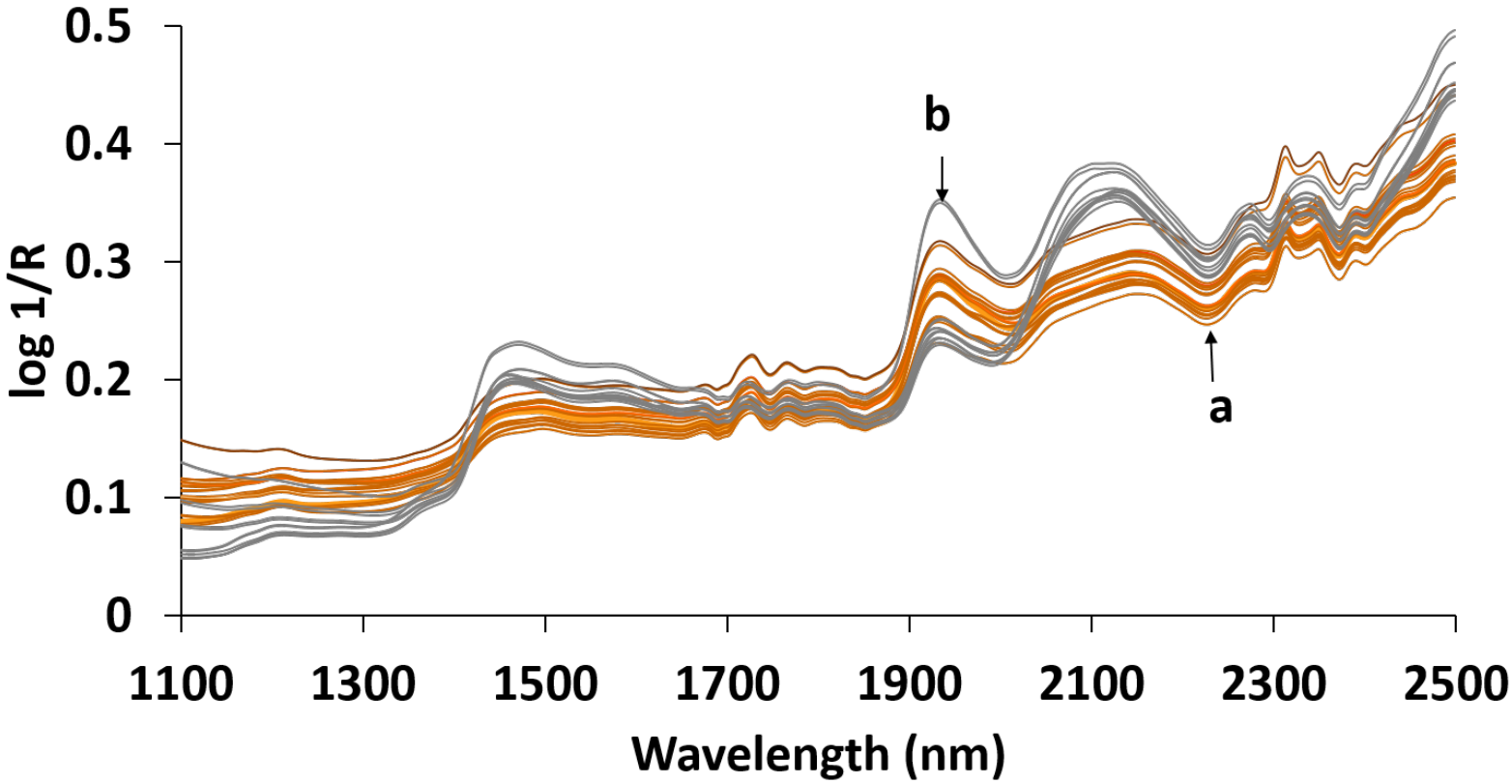

Fig 2. M.A. Quelal et al.

Figure 3

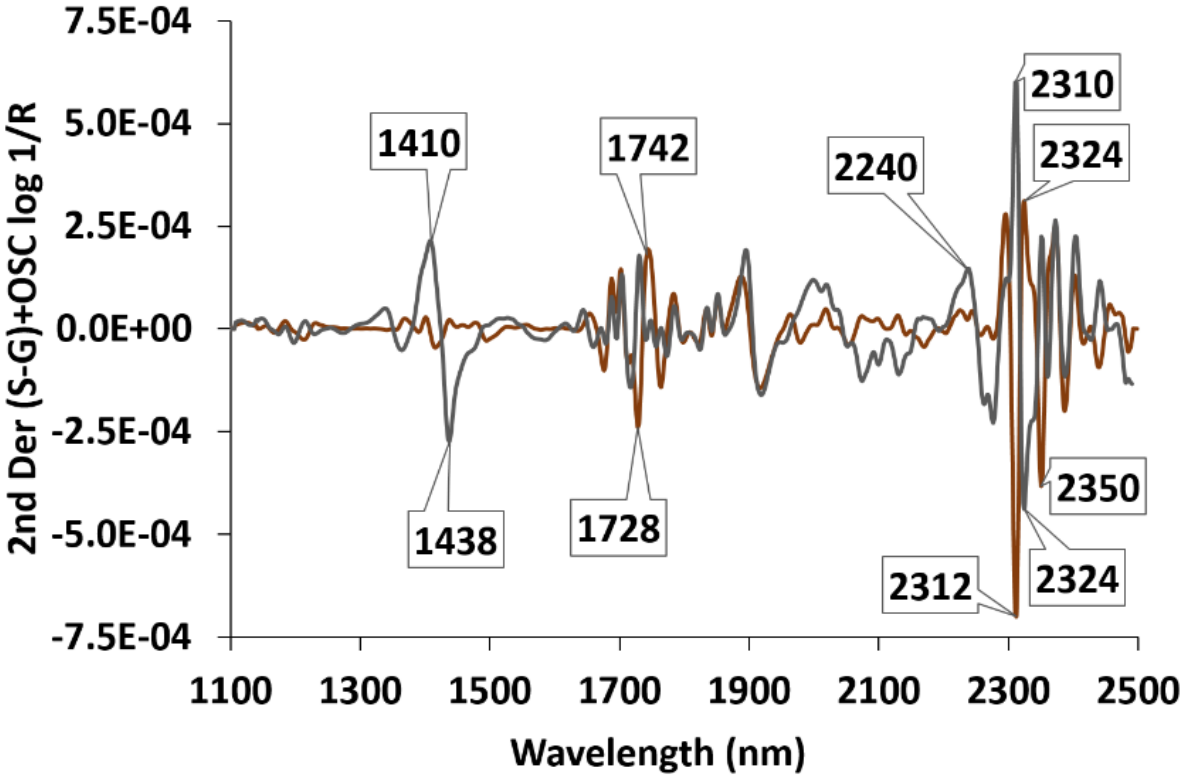

Fig 3. M.A. Quelal et al.

Figure 4

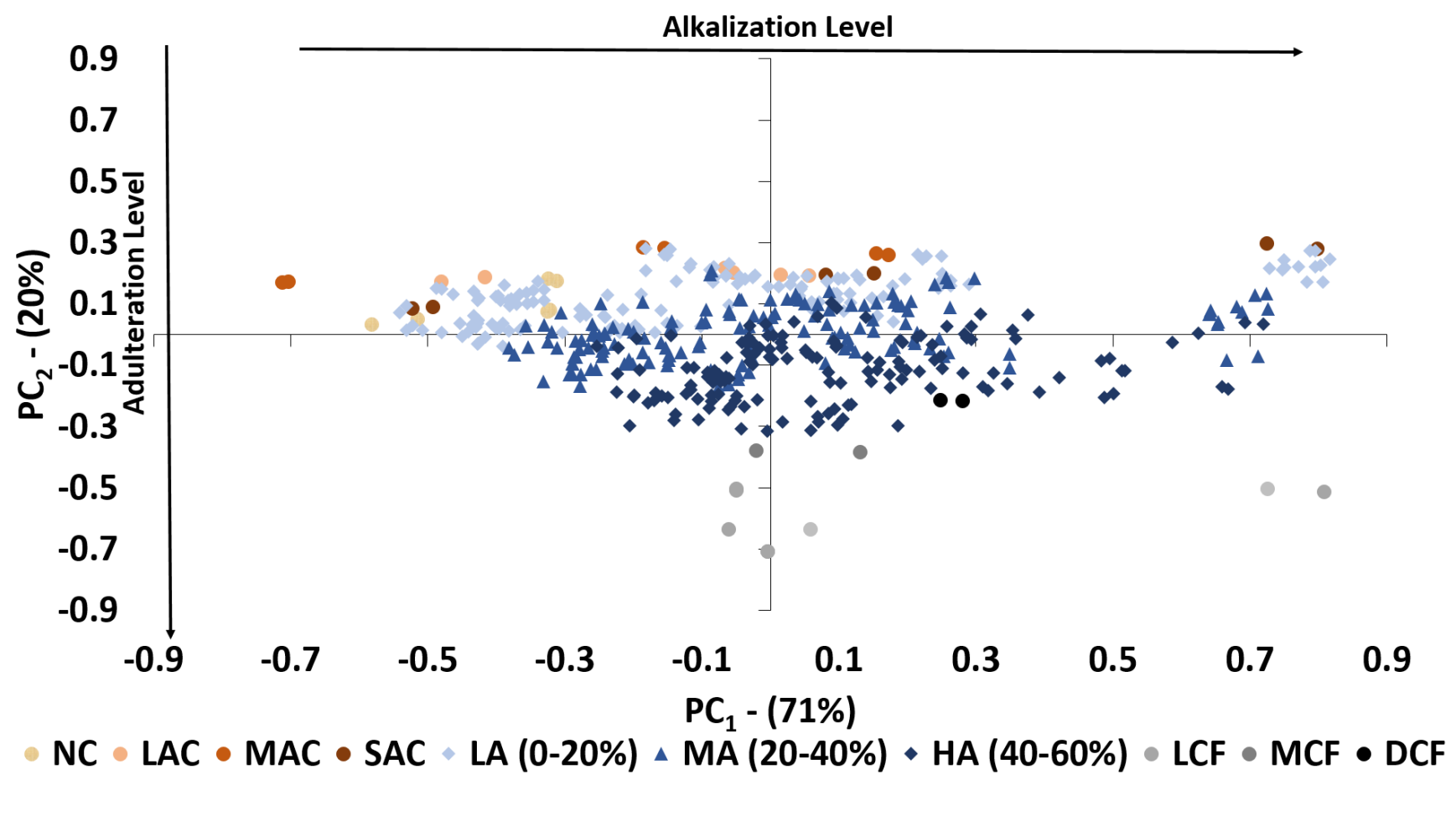

Fig 4. M.A. Quelal et al.

Figure 5

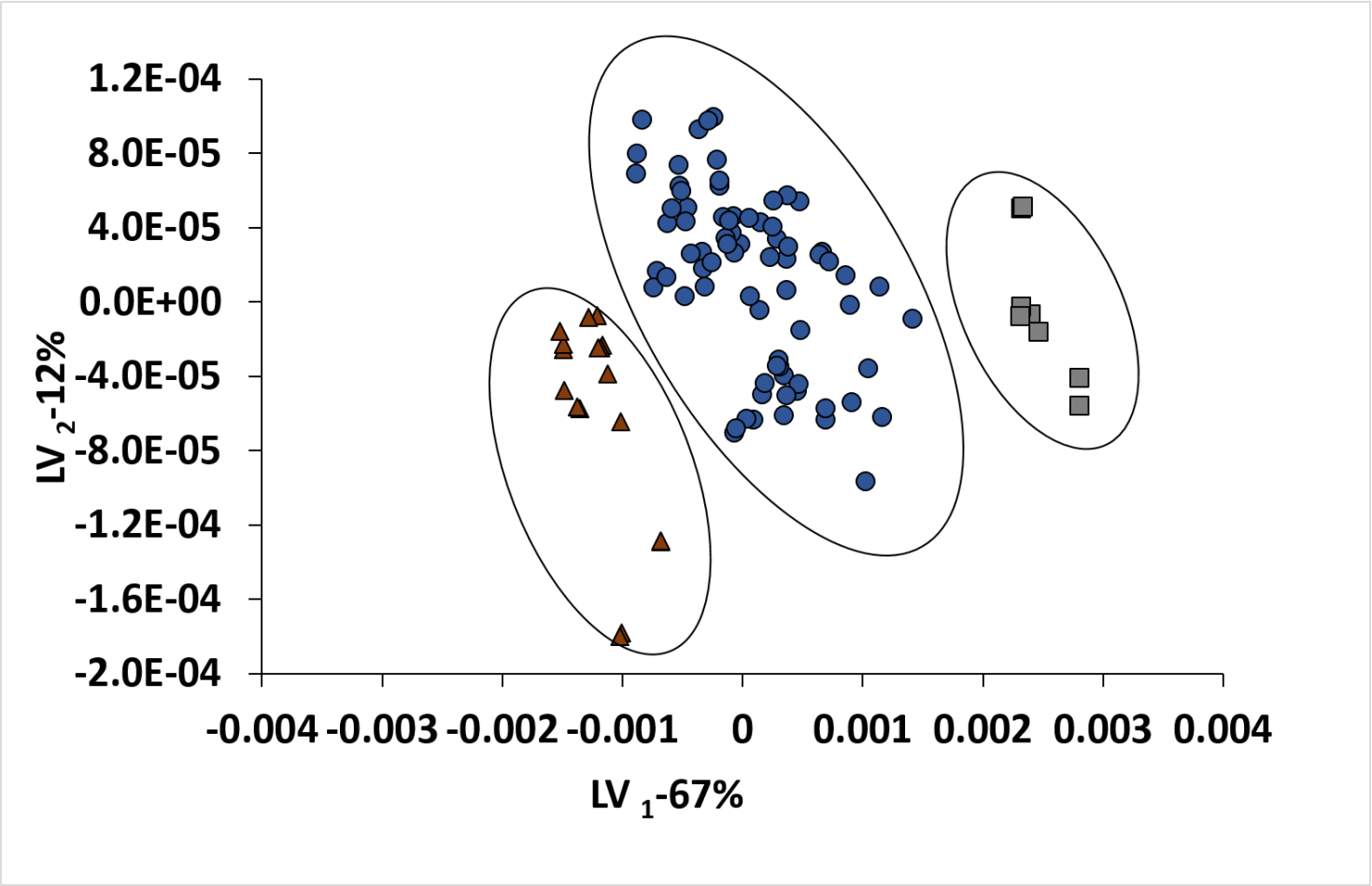

Fig 5. M.A. Quelal et al.

Figure 6

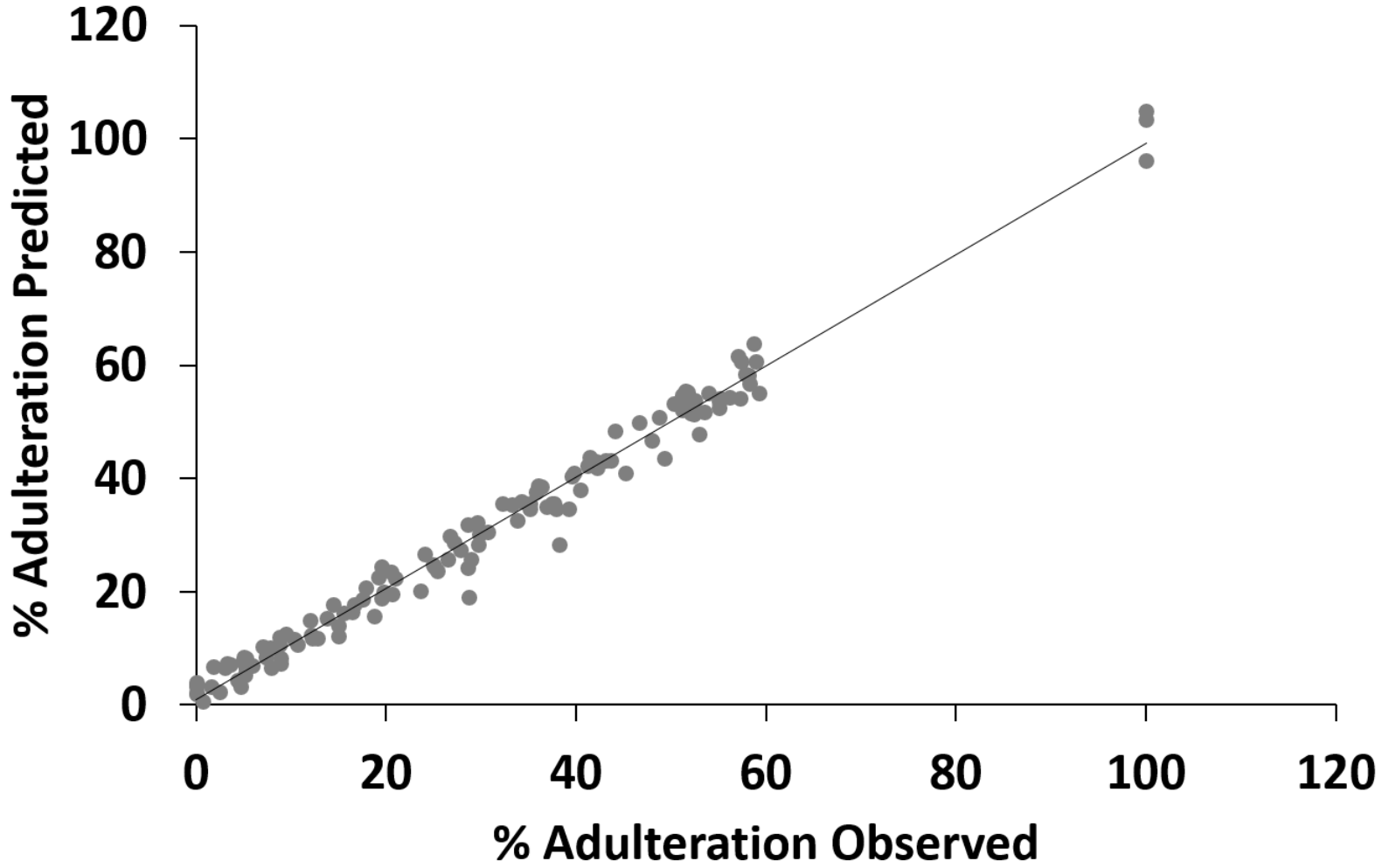

Fig 6. M.A. Quelal et al.

Highlights

- Near infrared spectroscopy can detect cacao adulterated with carob flour
- The method is rapid, useful and simple for quality cacao analyses
- PLS-DA distinguishes pure cocoa powders from those adulterated with carob flour
- A PLS analysis has the potential to quantify the percentage of adulteration