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@upv.es

Abstract

Cocoa powder is a global product of great value that can be adulterated with low-cost raw materials such as carob flour without changing the characteristics of color, aroma and flavor of the product. The use of rapid methods, as a NIR technology combined with multivariate analysis, is of interest for this detection. In this work, 216 adulterated samples prepared by blending commercial cocoa powders with different alkalization levels (n = 12) with commercial carob flour (n = 6) in different proportions (0-60% of adulteration) were analyzed. The diffuse reflectance spectra of the samples were acquired from 1100 to 2500 nm using a Foss NIR spectrophotometer. A qualitative and quantitative analysis was done. For the qualitative analysis, a principal component analysis (PCA) and a partial least squares discriminant analysis (PLS-DA) was performed. The coefficient of determination (R^2) of the model PLS-DA was 0.969 and the coefficient of determination of the validation (R^2_{CV}), based on a full cross-validation was 0.901 indicating good calibration with good predictability. These results indicate that it is possible to distinguish between pure cocoa powders from the adulterated samples. For

the quantitative analysis a partial least squares (PLS) regression analysis was performed. The most robust model of PLS prediction was obtained with 1 factors (LV) at coefficient of determination (R^2) of 0.980 and a root mean square error of prediction (RMSEp) of 3.237 % for the external validation set. These data lead to the conclusion that NIR technology combined with multivariate analysis allows 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, due 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 stands out the formulation of beverages, confectionery, bakery and pastry products (Shankar, Levitan, Prescott, & Spence, 2009). Apart from flavor and aroma, cocoa is really appreciated as a natural coloring agent, in part because of the tendency to restrict the use of artificial colors.

During cocoa processing, cocoa colour and aroma can be modified 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 colour, mild aroma and texture of commercial natural beans (Bonvehí, 2005, Krysiak, 2006; Afoakwa, Budu, Mensah-Brown, Felix & Ofosu-Snsah, 2014). By its part, alcalizatization is an optional operation to reduce acity, bitterness and astringency and to darken cocoa color. This procedure involves the use of an alkali (generally potassium carbonate) in combination with oxigen, water and hight temperatures. This extreme conditions provoke, among others, Maiard reactions and polyphenol oxidations and polymerizations, ending up with flavour and colour modifications (from light brown (natural) to red, dark brown or extremly black) (Miller et al., 2008; Li, Feng, Zhu, Luo, Ma, & Zhong, 2012).

During recent years, cocoa powders have experienced both, an increase in demand and a tightening of supplies, which has steadily raised the price (Fadel, Mageed, Samad, & Lotfy, 2006). In consequence, there has been a demand for the development of cocoa substitutes. Some studies suggest that cocoa-like aromas can be found in roasted carobs (Arrighi, Hartman & Ho, 1997). Carob pods are characterized for a high sugar content (around 50%), composed essentially of sucrose. This high sugar content favors the same

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

Having in mind this great aromatic and visual similarity between carob flour (natural or toasted) and cocoa (natural or alkalized), some traders have seen in the sale of carob (average price of 940 US$/tonne) as cocoa (1945 US$/tonne), omitting this substitution, a profitable option to increase their benefits (Arias and Zapata, 2017; ICCO, 2017). However, this deliberate, intentional and not declared substitution of one product for another with a lower price not only is a food fraud that affects producers and consumers, but it also affects the physico-chemical properties of the manufactured product. Some studied examples comprise milk chocolates and chocolate cakes in wich certain percentages of cocoa powder were subtituted by carob flour (Salem & Ohaad Fahad, 2012 Rosa, Tessele, Prestes, Silveira, & Franco, 2015).

To detect food adulteration the three most common technologies are liquid chromatography, infrared spectroscopy and gas chromatography (Moore et al., 2012). Of those, liquied and gas chromatography analysis need large times of sample preparation, the optimization of the method as well as high cost of materials and reactives. In contrast, infrared spectroscopy is fast, realiable, less expensive and a chemical-free alternative (Ellis et al., 2012). Near infrared spectroscopy (NIR) is a type of infrared spectroscopy characterized by registering reflectance or transmittance spectrums in the region from 13000 cm^{-1} to 3300 cm^{-1}. These spectrums act as a 'fingerprint' characteristic of a particular molecule of sample and allows its identification. Some examples of the use of NIR and multivariant analysis in the cocoa sector comprise the prediction of basic food componets such as moisture, carbohydrate, fat, protein, teobromin and catechin as well as total polyphenol content (Veselá et al., 2007; Álvarez et al., 2012; X. Y. Huang et al., 2014). In other sectors, NIR in

88 combination with multivariant analysis has been employed to detect starch in onion
89 powders; acid whey, starch, maltodextrin in skim powder milk; sudan dyes in chilly
90 powders and talcum powder in teas (Lohumi et al., 2014; Capuano, Boerrigter-Eenling,
91 Koot, & van Ruth, 2015; Haughey, Galvin-King, Ho, Bell, & Elliott, 2015; Li, Zhang,
92 & He, 2016).
93 In this context, the aim of this work is to detect qualitatively and cuantitativelye the
94 adulteration of cocoa powders (with independence of their alkalization level) with carob
95 flours throug the aplication of NIR and multivariant analysis.
96
97 **2. Materials and methods**
98
99 *2.1 Raw materials*
100
101 In order to analyze a good set of samples representative of the existing variability in
102 commercial cocoa and carob flour, cocoa powders with different alkalization levels
103 (natural cocoa -NC-, lightly alkalized cocoa -LAC-, medium alkalized cocoa –MAC-
104 and strong alkalized cocoa -SAC-) (n=12) and carob flour powders with three different
105 roasting degrees (light carob flour -LCF-, medium carob flour -MCF- and dark carob
106 flour -DCF-) (n=6) were used in this study. The cocoa powders were gently donated by
107 OLAM Food Ingredients Spain (Cheste, Spain) and the carob flour podwers were
108 bought in a local specialised supermarket. The raw samples were placed in a glass
109 container and stored in a dry and dark atmosphere until use.
110
111 *2.2 Physical and Chemical characterization of raw materials*
112
113 Each of the raw sample was characterized according to their extractable pH value
114 and extrinsic colour.

For extractable pH determination, 10g of cocoa powder were suspended in 90 mL of boiling distilled water and stirred. Then, temperature was reduced to 20-25 ºC in a coldwater bath (OLAM, 2017). pH was mesured with a digital pH-meter micropH 2001 (Crison Instruments, S.A., Barcelona, Spain). Samples were classified according to their pH value in four different categories: natural cocoa powders (pH 5-6), light alkalized (pH 6-7.2), medium alkalized (pH 7.2-7.6) and strong alkalized powders (pH > 7.6) (Miller et al., 2008).

For the determination of the extrinsic color, the sample of cocoa powder was placed in a methacrylate cuvette, unifying the degree of compaction through small successive shocks. The color was measured using a spectrocolorimeter Minolta CM 3600D (Tokyo, Japón). The reflectance spectra, between 400-700 nm was used to obtained the color coordinates L*, a* and b* for D65 illuminant and 10° observer. Hue (h^*) and chroma (C^*) were estimated by the equations 1 and 2, respectively. All the measurements were performed in triplicate.

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

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

2.3 Preparation of adulterated samples

A total of 216 adulterated samples were prepared by blending the 12 cocoa powders with the 6 different carob flours at different proportions. For each one of 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 fixed on the consideration that above this

concentration the adulteration is evident due to the characteristic aroma of carob (Cantalejo, 1997). The concrete adulteration percentage within a level was determined randomly from an uniform distribution (every percentage of adulteration had the same probability to be selected) following the latin hypercube strategy (LHS) (Helton & Davis, 2003). The adulterated samples, in the same way as the raw samples, were placed in a glass container and stored in a dry and dark atmosphere until use.

2.4.Near-infrared spectra collection

The 234 samples (12 unaltered cocoa powders, 6 carob flour and 216 adulterated samples) were scanned using a FOSS NIR 5000 System spectrophotometer (Silver Spring, MD, USA) equipped with a transport module. A round sample cup with 3.8 cm diameter x 1cm thick quartz windows were filled with each sample (about 5g) to maintain a uniform surface and thickness during spectral collection. The instrument measures the diffuse reflectance and automatically converts it to relative absorbance (log 1/R) to obtain a linear correlation with the concentration of the chemical constituents of the product according to Beer's Law (Martens, Nielsen, & Engelsen, 2003). A total of 32 successive scans with 700 points (wavelengths) from each sample were collected, between a wavelength range of 1100 and 2500 nm at 2 nm intervals. Samples were measured twice in order to mesasure the influence of environmental conditions and to ensure reproducibility. Then a total of 468 spectra were collected.

2.5 Chemometric analysis

Analysis of variance (ANOVA) was used to determine differences in pH and extrinsic color among samples. The data were statistically processed using Statgraphics

167 Centurion XVI (Manugistics Inc., Rockville, MD, USA). Simultaneously the color
168 parameters (C*, h*, L*) and the pH were used in a principal component analysis (PCA)
169 to show the samples and their relationship.
170 Multivariate analysis was conducted by a qualitative and a quatitative analysis using
171 The Unscrambler v10.4 (CAMO Software AS, OSLO, Norway). For the qualitative
172 analysis, a PCA and a partial least squares discriminant analysis (PLS-DA) was
173 performed. The PCA was performed with the raw data while the PLS-DA (Berrueta,
174 Alonso, & Héberger, 2007) was constructed after apply to the spectra 2nd derivative
175 (Savitzky-Golay smoothing) (Savitzky & Golay, 1951) and orthogonal signal correction
176 (OSC). Both pre-treatments are applied in order to extract useful information, improve
177 the signal-to-noise ratio and remove systematic variation from the predictor matrix X
178 unrelated, or orthogonal, to the matrix Y (Wold, Antti, Lindgren, & Öhman, 1998;
179 Pizarro et al., 2004). For the quantitative analysis a partial least squares (PLS)
180 regression analysis was performed. In order to evaluate and correct multiplicative and
181 additive effects caused by different light scattering in the spectroscopic measurement
182 (Cozzolino et al., 2011; Stohner et al., 2012), five PLS models were tested. The PLS
183 were constructed using the raw spectrum and applying three pre-treatments to the
184 spectrum: 2nd derivative Savitzky-Golay smoothing, orthogonal signal correction (OSC)
185 and the combination of 2nd derivative Savitzky-Golay smoothing and orthogonal signal
186 correction (OSC)
187
188 *2.5.1 Calibration models development*
189
190 Two databases were used for the analysis. The first database consisted of 468 spectra
191 and 700 variables (wavelengths, nm) and was used for the PCA and PLS models. For
192 PLS-DA classification in three categories (0=Cocoa; 1=Adulterated samples and

193 2=Carob flour), a second database containing 135 spectra and 700 variables was created
194 in order to balance the number of samples belonging to each category. Moreover, the
195 spectra of each database were randomly separated into two different data sets. A set
196 containing the 70% of the spectra was used for the creation and evaluation of the model
197 by leave-one-out cross-validation. The other set, with 30% of the remaining samples
198 was used for the external validation. The relative performance of the constructed models
199 was assessed by the required number of latent variables (LVs), the coefficient of
200 determination for calibration (R^2_C), the root mean square error of calibration ($RMSE_C$),
201 the coefficient of determination for cross validation (R^2_{CV}) and the root mean square
202 error of leave-one-out cross validation ($RMSE_{CV}$). A model can be considered good
203 when a low number of LVs are required and it has a low $RMSE_C$ and $RMSE_{CV}$ and high
204 R^2_C and R^2_{CV}.
205
206 *2.5.2 External validation*
207
208 To assess the predictive capability of the models the coefficient of determination for
209 prediction (R^2_P), the root mean square error of prediction (RMSEp), the ratio of
210 prediction deviation (RPD = SD/SEP), where the SD was the standard deviation of the
211 Y-variable in the prediction set) and the bias were used. RPD is more meaningful than
212 only looking at the error of prediction. RPD value less than 2 is considered insufficient
213 for application, between 2 and 2.5 is considered for approximate quantification, values
214 between 2.5 and 3 as a good model, while models with RPD values more than 3 can be
215 considered an excellent and most reliable for analytical tasks (Sunoj, Igathinathane, &
216 Visvanathan, 2016). The bias estimates the difference between experimental value and
217 NIR predictions and can be positive or negative. Positive values indicate that the model
218 was overestimating, while negative values indicate otherwise. Larger bias values

indicate that the NIR predictions vary significantly from the experimental values (Cantor et al., 2011), so it is better that tends to zero. The LOD which is the minimum value of adulteration that could be detected by the model, which is the result of adding the average plus 3 times the standard deviation (Haughey et al., 2015; Lerma, Ramis, Herrero, & Simó, 2010) was calculated with the optimal model. For this the adulteration level of four pure cocoa samples (blanks) of different alkalization degree was calculated.

3. Results and Discussion

3.1 Color and extractable pH analysis

In order to classify cocoa samples in relation to their alkalization level and carob samples in line with their roasting intensity, the 12 cocoa samples and the 6 carob flour samples were characterized according to their pH value and color parameters (L*, C* and h*).

Table 1 contains the color parameters and pH values of different raw matters. As observed, the pH values obtained ranged from 5.3 (NC1) to 7.9 (SAC3). pH can be used as an indicator of the degree of alkalization occurring in production because the pH value of the cocoa powder is related to the amount and type of alkali used in the process (OLAM, 2017; Pérez et al., 2016). According to previous statement, samples were classified in four categories (natural, light alkalized, medium alkalized and strong alkalized) following the classification of Miller et al., (2008) (See section 2.2). Following these values, the samples of cacao were classified in light alkalized cocoa (PH 6-7.2), medium alkalized cocoa (pH 7.2-7.6), and strong alkalized cocoa PH > 7.6).

244 3 of the 12 samples were considered natural cocoa (NC), 3 light alkalized cocoa (LAC),
245 3 medium alkalized cocoa (MAC) and 3 strong alkalized cocoa (SAC).
246 The luminosity values (L*) measured in cocoa samples ranged from 31 (SAC1) to 50
247 (NC3). The maximum value of luminosity appears in a sample of natural cocoa (NC3).
248 The value of L* decreases progressively as a function of the degree of alkalization to
249 the minimum value in strongly alkalized (SAC) samples with a very dark color. The
250 observed differences in luminosity in natural cocoa samples (NC1, NC2 and NC3)
251 could be due to a different geographical origin or to a different processing in the
252 fermentation or roasting stages (Afoakwa, Budu, Mensah-brown, Felix, & Ofosu-ansah,
253 2014).
254 The chroma values, C*, oscillated between 11 (SAC1) and 22 (NC2). As can be seen
255 in Table 1, the higher the alkalization degree the lower the purity.
256 The hue (h*), unlike the other parameters, does not follow a linear relationship with
257 the increase in pH value. Cocoa samples evolve from a more yellow-orange hue (h * =
258 60) to more orange-red hue (h* = 43) in the alkaline cocoa samples.
259 Whit respect of carob flours the pH value ranged from 4.5 to 5.1, with no trend
260 between pH value and degree of toasting samples. Thus, carob samples could be added
261 to natural cocoa beans in high proportions without significantly changing the pH value
262 of the mixture.
263 The values of the L* ranged from 34 (DCF) to 49 (LCA) in carob flours, showing
264 that the luminosity decreases progressively as the degree of roasting increases.
265 Comparing these values with those of cocoa, it can be confirmed that attending to the
266 luminosity does not exist statistical differences ($p<0.05$) between samples of natural
267 cocoa and natural carob meal. On the other hand, there are also no differences among
268 luminosity of medium and strong alkali cocoa and roasted carob. These little differences
269 in luminosity favor the adulteration of cocoa with carob meal.

Regarding Chroma (C*) of the samples also decreases as the degree of roasting increases, reaching values of 21 for natural carob and reducing to 12 for strong roasted carob. Comparing the C* values between cocoa and carob can be seen that there is a similarity between the two. Thus, C* values would be equivalent between natural cocoa and natural carob meal and between medium / strong cocoa beans and roasted carob.

The values of hue (h*) for the carob flours do not seem to show significant differences with the increase of the degree of roasting, only a slight decrease. The values obtained for samples of natural carob flour have an average of 61, a value that is reduced to 58 in the samples of carob with a high degree of roasting. These values coincide with those observed in samples of natural cocoa and soft alkaline cocoa.

In general, the cocoa color parameters are affected by several factors including the degree of roasting and the alkalization. The strong alkalized has a dark color while the natural has lighter color. The roasting result in a darkening of the cocoa or carob because of the formation of brown pigments (Zyzelewicz, Krysiak, Nebesny, & Budryn, 2014) with changes in the values of individual color parameters.

Insert here Table 1

In order to know the different characteristics between the cocoa and carob flour samples, a PCA was performed with the pH and color parameters, which is presented in Fig. 1. It can be seen several groups. The scores of natural cocoa (NC) and light carob flour (LCF) are very nearly which indicates that these samples are related and have similar characteristics of pH and color parameters. These scores are negative in the component 1 and positive in the component 2. The other groups correspond to the different levels of alkalization and roasting. Positive scores in the component 1 and component 2 correspond to samples with different degree of alkalization. This position

296 and the variables values lead to the conclusion that samples with low luminosity and
297 high pH are samples of alkaline cacao; Samples with low luminosity and low pH are
298 samples of roasted carob flour (Dark (DCF) and medium (MCF)), the scores of these
299 samples are positive in component 1 and negative in the component 2. This agrees with
300 the results presented by other authors (Bulca, 2016; Yousif & Alghzawi, 2000) that
301 indicate that the carob flour could not be visually separated of the cocoa powder, even if
302 the other groups of alkalized and roasting samples, were blended.
303 Insert here Figure 1
304
305 *3.2 Spectral differences analysis of carob and cocoa powder*
306
307 Spectrums of relative absorbance of cocoa powder and carob flour are represented in
308 Figure 2 (a, b). As shown in the figure, all spectra have a similar pattern of absorbance,
309 although this pattern is different between cocoa and carob flour.
310
311 Insert here Figure 2
312
313 Raw data were preprocessed by applying the 2nd derivative and OSC. Examples of
314 the pretreated spectrums of cocoa (brown) and carob (gray) are shown in Figure 3. As
315 observed, after this pretreatment differences among both types of spectrums are more
316 evident than in non-treated spectrums. Moreover, it can be stated how divergence points
317 between both types of spectrums are located especially in the magnitude of reflectance
318 at 1438, 1728, 2312, 2324, 2350 nm. As it could be expected from composition
319 differences among cocoa powder and carob flours, these wavelengths are associated to
320 the vibration of functional groups typical from fatty acids (1800 and 1734 nm), from
321 methyl that could be related with the theobromine and caffeine contain of the cocoa

322 powder (1728 nm) (Cozzolino et al., 2011) and polyphenols like epicathechin (2312,
323 2324 nm) (Esteban, González, & Pizarro, 2004; Teye & Huang, 2015).
324
325 Insert here Figure 3
326
327 *3.2 Classification model*
328
329 A principal component analysis which is a non-supervised method of classification
330 was performed with the raw spectrums data, with the aim of evaluating the relation
331 among samples. The fig. 4 shows the score plot of the first two principal components.
332 The first PC explains the 71% of the total variance of the NIR data. It is related to the
333 sample processing. The most negative scores correspond to the different natural cocoa
334 or lightly roasted carob. In contrast the most positive scores correspond to the strong
335 alkalized cocoa powder and highly roasted carob flours. The second PC explains the
336 20% of the variability. It is related to the percentage of cocoa powder in the sample. The
337 most positive values correspond to pure cocoa powders, while the most negative
338 correspond to the pure carob flours. In the middle are allocated samples containing
339 different levels of adulteration: low (0-20%), medium (20-40%) and high (40-60%).
340 Wavelengths corresponding to the highest loading values for PC_1 were 1100, 1464,
341 1936, 2108, 2276, 2330 and 2486 nm and for the PC_2 1116, 1324 1460, 1576, 1728,
342 1914, 1976, 2106, 2262, 2310 and 2494 nm. Wavelengths from 971 and 1400 nm are
343 related to the ascending part of the water first overtone absorption peak O–H stretching
344 bonds at 1722 nm the C-H stretching is present too. Which are associated with water
345 and sugar content (Álvarez et al., 2012; Cozzolino, Smyth, & Gishen, 2003; X. Y.
346 Huang et al., 2014; Talens et al., 2013). Meanwhile, wavelengths at 1736 and 2319-
347 2328 nm are related to the absorption of the C–H bonds, CH_3 combination and C-C

348 stretching. These are features of fatty acids, proteins and polysaccharides in cocoa
349 powder. It could be associated with fat content of approximately a 10-12% (Veselá et
350 al., 2007; Westad, Schmidt, & Kermit, 2008). The absorption bands of 1728, 2108 and
351 2494 nm approximately coincide with those that have been used to predict the total
352 content of fat in cocoa beans by (Ribeiro, Ferreira, & Salva, 2011; Teye & Huang,
353 2015). The variations are related to the compositional characteristics of the cocoa
354 categories and the adulterant carob powder. The wavelengths founded are similar to the
355 study performed in cocoa beans (Teye et al., 2015b). Therefore, the absorption in the
356 wavelengths (product of the vibrational reactions) has chemical information which is
357 contributing to explain the differences observed between the carob and cocoa powder
358 pure samples and their several proportions of adulteration. Due to the generated spectra
359 correspond to a level of adulteration on a continuous scale. It is not possible to have
360 well separated categories (high, medium and low) in this PCA, especially for the
361 percentages that are in the limits. For this reason, a discriminant partial least squares is
362 necessary to generate a model with categorized spectrums. Which allow detecting gross
363 adulterations levels.
364
365 Insert here Figure 4
366
367 As the PCA was unable to classify samples in different groups according to their
368 adulteration percentage, a qualitative model using the supervised discriminant partial
369 least squares PLS-DA was performed. Moreover, to improve the accuracy of the model,
370 the original spectrums were pre-processed using second derivative with Savitzky-Golay
371 smoothing (9-point window, 2nd order polynomial) and an Orthogonal Signal
372 Correction (OSC). For PLS-DA (Figure 5), 3 latent variables (LVs) were generated with
373 most of the variation (67%) explained by the first LV and (12%) by the second. In this
374 way separation is mainly achieved using the first latent variable with most negative

scores related to the cocoa pure samples and most positive scores related to the adulterates samples and the carob powder (pure adulterant). Visually, the scores plot differences between 100% cocoa powder, adulterated cocoa powders and 100% carob powder indicating that it may be possible to use this approach to quickly screen for adulteration. Moreover, the determination coefficient (R^2) of this PLS-DA model was 0.969. The cross validation determination coefficient (R^2_{CV}) based on a full cross validation was 0.901. Those values indicate the goodness of the classification model.

Insert here Figure 5

In order to measure the robustness of the PLS-DA model, a validation with an external set of data was performed. Table 2 shows the capability of the model to classify 100% of the samples in its corresponding group (cocoa, carob or adulterated samples).

Insert here Table 2

3.3 Adulterant Prediction

For the prediction of the adulteration, a PLS was performed with a calibration set and after that the prediction was done with the validation set. The models were constructed applying different pre-treatments to the spectra. The statistical indicators of goodness of fit of each of these models are presented in the Table 3. Good models were obtained with high values for the correlation coefficients (R^2) between 0.951 and 0.980. Low values for the root mean square error of calibration (RMSEC) and root mean square error of prediction (RMSEp) between 4.397 and 3.237 depending on the processing of the spectral data. The ratio prediction deviation RPD of the models obtained were

401 between 4.66 and 6.41. All of these values are greater than 3 which means that all these
402 models, even the model whit out the preprocessing data, can be considered as excellent
403 and most reliable for analytical tasks. This indicates that the multiplicative and additive
404 effects in this type of samples and with the equipment used for the measurement in this
405 study is minimal. Although have an optimal model with a lower error of prediction is
406 always better.

407 In Fig. 6 are presented observed (x-axis) *versus* predicted (y-axis) values. Predicted
408 values were obtained with a model using 2nd Derivative algorithm with Sawitzky-Golay
409 smoothing (9-point window and 2nd order polynomial) and orthogonal signal correction
410 is presented. It can be observed that PLS algorithm gave a very good prediction with a
411 correlation coefficient (R^2) of 0.980 and RMSEC of 2.856. The root mean square error
412 of cross validation ($RMSE_{CV}$) was 2.897 %. The prediction of the external group of
413 validation gave a low root mean square error of prediction ($RMSE_P$) 3.237 %. The
414 similarity among $RMSE_C$, $RMSE_{CV}$ and $RMSE_P$, shows that the possibility of over-
415 fitting the model is very low and it confirms its good capacity of prediction.

416

417 Insert here Table 3

418

419 The results indicated that PLS model with 2nd Derivative algorithm with Sawitzky-
420 Golay and orthogonal signal correction showed low values of $RMSE_C$ and $RMSE_{CV}$ and
421 high values of coefficients of determination (R^2). Which indicate good performance of
422 the predict model with an improve in the ratio prediction deviation (RPD), which is
423 37.55 % higher respect the PLS model with the raw data, and with only 1 latent
424 variables, other studies have found good models with 1 LV with the use of orthogonal
425 signal correction (Esteban, González, & Pizarro, 2004).

426

427 Insert here Figure 6

The relative notorious improve of the RPD of the pretreated model could be because of NIR signal can be affected by the moisture, particle size distribution of the product. Those phisical properterties can produce significant differences because of the light scatter effects. These factors varying the effective sample pathlength and result in aditive, multiplicative and wavelenght dependent effects. The baseline shifts, tilt or a curvature scaling variation in some instances are related with the wavelength-dependent scattering . The spectra variations could mask any subtle chemical variations it can lead to inaccurate results, so the pretreatment is effective decreasing the mentioned effects (Huang et al., 2010).

The LOD (mean + 3 standard deviations) was calculated from four pure cocoa powder samples between alkalized and natural and it was 6.073 %. As with the NIR based calibration models, future work should include more variation by using cocoa powders and carob powder from more different sources.

4. Conclusions

Near infrared spectroscopy (NIR) in combination with the discriminant partial least squares (PLS-DA) and partial least squares (PLS) statistical models has been shown to be a rapid and 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.

Through the PLS-DA analysis, 100% of the samples were correctly classified into three groups: cocoa, carob flour and mixtures. On the other hand, by means of a PLS analysis it was possible to quantify the percentage of adulteration of the samples. The PLS model was obtained with 1 factor at R^2 of 0.980 and 0.974 and a mean squared error of 2.856 and 3.237 for the calibration and validation sets, respectively.

This technology is therefore an important tool for cocoa merchants, who will be able to obtain a better control of the quality of the product, avoiding the use of destructive techniques that require a complex preparation of the sample or techniques that imply an important expense for the company. Due to the excellent results achieved, we can expect that this method will become increasingly important in the cocoa industry, contributing to the reduction of food fraud.

Acknowledgments

The authors wish to acknowledge the financial assistance provided 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. Olam Food Ingredients Company is aknowledged for proving part of the cocoa samples used in the study.

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.

Arias Mesía, L. N. & Zapata Yarlequé, F. N. (2017). Estudio de prefactibilidad para la instalación de una planta para la elaboración de galletas enriquecidas con harina de algarroba (Prosopis pallida). Universidad de Lima.

481 Arrighi, W. J., Hartman, T. G., & Ho, C. T. (1997). Carob bean aroma dependence on
482 roasting conditions. *Perfumer & flavorist*, 22(1), 31-41.
483 Berrueta, L. A., Alonso, R. M., & Héberger, K. (2007). Supervised pattern recognition
484 in food analysis. *Journal of Chromatography* A, 1158(1–2), 196–214.
485 Bonvehí, J. S. (2005). Investigation of aromatic compounds in roasted cocoa powder.
486 *European Food Research and Technology*, 221(1–2), 19–29.
487 Bulca, S. (2016). Some properties of carob pod and its use, XX, 142–147.
488 Cantalejo, M. J. (1997). Effects of Roasting Temperature on the Aroma Components of
489 Carob (Ceratonia siliqua L.). *Journal of Agricultural and Food Chemistry*, 45(4),
490 1345–1350.
491 Cantor, S.L., Hoag, S.W., Ellison, C.D., Khan, M.A., Lyon, R.C., 2011. NIR
492 spectroscopy applications in the development of a compacted multiparticulate
493 system for modified release. *AAPS PharmSciTech* 12, 262–278.
494 Capuano, E., Boerrigter-Eenling, R., Koot, A., & van Ruth, S. M. (2015). Targeted and
495 Untargeted Detection of Skim Milk Powder Adulteration by Near-Infrared
496 Spectroscopy. *Food Analytical Methods*, 8(8), 2125–2134.
497 Cozzolino, D., Smyth, H. E., & Gishen, M. (2003). Feasibility Study on the Use of
498 Visible and Near-Infrared Spectroscopy Together with Chemometrics to
499 Discriminate between Commercial White Wines of Different Varietal Origins.
500 *Journal of Agricultural and Food Chemistry*, 51(26), 7703–7708.
501 Cozzolino, Cynkar, W. U., Shah, N., & Smith, P. (2011). Multivariate data analysis
502 applied to spectroscopy: Potential application to juice and fruit quality. *Food*
503 *Research International*, 44(7), 1888–1896.
504 Ellis, D. I., Brewster, V. L., Dunn, W. B., Allwood, J. W., Golovanov, A. P., &
505 Goodacre, R. (2012). Fingerprinting food: current technologies for the detection of
506 food adulteration and contamination. *Chemical Society Reviews*, 41(17), 5706.

507 Esteban, I., González, J. M., & Pizarro, C. (2004). Prediction of sensory properties of
508 espresso from roasted coffee samples by near-infrared spectroscopy. *Analytica*
509 *Chimica Acta*, 525(2), 171–182.
510 Fadel, H. H., Mageed, M. A. A., Samad, A. K. M. A., & Lotfy, S. N. (2006). Cocoa
511 substitute: Evaluation of sensory qualities and flavour stability. *European Food*
512 *Research and Technology*, 223(1), 125-131.
513 Haughey, S. A., Galvin-King, P., Ho, Y.-C., Bell, S. E. J., & Elliott, C. T. (2015). The
514 feasibility of using near infrared and Raman spectroscopic techniques to detect
515 fraudulent adulteration of chili powders with Sudan dye. Food Control, 48, 75–83.
516 Helton, J. C., & Davis, F. J. (2003). Latin hypercube sampling and the propagation of
517 uncertainty in analyses of complex systems. Reliability *Engineering and System*
518 *Safety*, 81(1), 23–69.
519 ICCO (2017). International Cocoa Organization Dailyy Prices of cocoa Beans
520 https://www.icco.org/statistics/cocoa-prices/daily-prices.html, Consulted: 13th
521 January 2018.
522 Krysiak, W. (2006). Influence of roasting conditions on coloration of roasted cocoa
523 beans. *Journal of food engineering*, *77*(3), 449-453.
524 Huang, J., Romero, S., & Moshgbar, M. (2010). Practical Considerations in Data Pre-
525 treatment for NIR and Raman Spectroscopy. *American Pharmaceutical Review*,
526 13(6). http://www.americanpharmaceuticalreview.com . Consulted: 2 December
527 2017
528 Huang, X. Y., Teye, E., Sam-Amoah, L. K., Han, F. K., Yao, L. Y., & Tchabo, W.
529 (2014). Rapid measurement of total polyphenols content in cocoa beans by data
530 fusion of NIR spectroscopy and electronic tongue. *Analytical Methods*, 6(14),
531 5008–5015.
532 Krähmer, A., Engel, A., Kadow, D., Ali, N., Umaharan, P., Kroh, L. W., & Schulz, H.

(2015). Fast and neat – Determination of biochemical quality parameters in cocoa using near infrared spectroscopy. *Food Chemistry*, 181, 152–159.

Lerma, M. J., Ramis, G., Herrero, J. M., & Simó, E. F. (2010). Authentication of extra virgin olive oils by Fourier-transform infrared spectroscopy. *Food Chemistry*, 118(1), 78–83.

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.

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.

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

559 coffee by near infrared spectroscopy. *Analytica Chimica Acta*, 509, 217–227.
560 Ribeiro, J. S., Ferreira, M. M. C., & Salva, T. J. G. (2011). Chemometric models for the
561 quantitative descriptive sensory analysis of Arabica coffee beverages using near
562 infrared spectroscopy. *Talanta*, 83(5), 1352–1358.
563 Rosa, C. S., Tessele, K., Prestes, R. C., Silveira, M., & Franco, F. (2015). Effect of
564 substituting of cocoa powder for carob flour in cakes made with soy and banana
565 flours. *International Food Research Journal*, 22(5), 2111–2118.
566 Salem, E. M., & Ohaad Fahad, A. A. (2012). Substituting of Cacao by Carob Pod
567 Powder In Milk Chocolate Manufacturing. *Australian Journal of Basic and Applied*
568 *Sciences*, 6(3), 572–578.
569 Savitzky, A., & Golay, M. (1951). Smoothing and Differentiation of Data by Simplified
570 Least Squares Procedures. Z. Physiol. *Chem. Chem. & Ind*, 40(42).
571 Shankar, M. U., Levitan, C. A., Prescott, J., & Spence, C. (2009). The influence of color
572 and label information on flavor perception. *Chemosensory Perception*, 2(2), 53–58.
573 Stohner, J., Lukas, B., Suter, M., Zucchetti, B., Deuber, F., & Hobi, F. (2012). NIRS of
574 chocolate and its chemometric analysis. *Newfood*, 15(6), 21–28.
575 Sunoj, S., Igathinathane, C., & Visvanathan, R. (2016). Nondestructive determination of
576 cocoa bean quality using FT-NIR spectroscopy. *Computers and Electronics in*
577 *Agriculture*, 124, 234–242.
578 Talens, P., Mora, L., Morsy, N., Barbin, D. F., ElMasry, G., & Sun, D.-W. (2013).
579 Prediction of water and protein contents and quality classification of Spanish
580 cooked ham using NIR hyperspectral imaging. *Journal of Food Engineering*,
581 117(3), 272–280.
582 Teye, E., & Huang, X. (2015). Novel Prediction of Total Fat Content in Cocoa Beans by
583 FT-NIR Spectroscopy Based on Effective Spectral Selection Multivariate
584 Regression. *Food Analytical Methods*, 8(4), 945–953.

585 Teye, E., Huang, X., Sam-Amoah, L. K., Takrama, J., Boison, D., Botchway, F., &
586 Kumi, F. (2015). Estimating cocoa bean parameters by FT-NIRS and chemometrics
587 analysis. *Food Chemistry*, 176, 403–410.
588 Veselá, A., Barros, A. S., Synytsya, A., Delgadillo, I., Čopíková, J., & Coimbra, M. A.
589 (2007). Infrared spectroscopy and outer product analysis for quantification of fat,
590 nitrogen, and moisture of cocoa powder. *Analytica Chimica Acta*, 601(1), 77–86.
591 Westad, F., Schmidt, A., & Kermit, M. (2008). Incorporating chemical band-assignment
592 in near infrared spectroscopy regression models. *Journal of Near Infrared*
593 *Spectroscopy*, 16(1), 265.
594 Williams, P.C. (1987). Variables affecting near-infrared reflectance spectroscopic
595 analysis. In: Williams, P., Norris, A. (Eds.), Near-infrared Technology in the
596 Agricultural and Food Industries. *American Association of Cereal Chemists*, St.
597 Paul, MN, pp. 143–167.
598 Wold, S., Antti, H., Lindgren, F., & Öhman, J. (1998). Orthogonal signal correction of
599 near-infrared spectra. *Chemometrics and Intelligent*, 44, 175–185.
600 Yousif, A. K., & Alghzawi, H. M. (2000). Processing and characterization of carob
601 powder. *Food Chemistry*, 69(3), 283–287.
602 Zyzelewicz, D., Krysiak, W., Nebesny, E., & Budryn, G. (2014). Application of various
603 methods for determination of the color of cocoa beans roasted under variable
604 process parameters. *European Food Research and Technology*, 238(4), 549–563.

605 **Figure captions**

606

607 **Fig 1.** Score plot of the first and the second principal components of PCA model using
608 the color parameters L*, C*, h* and pH of pure carob and cocoa powder samples (n =
609 18 by triplicate).

610

611 **Fig 2.** Spectra with raw data in the range of 1100 to 2500 nm (a) Cocoa. (b) Carob flour.

612

613 **Fig 3.** Second derivative, Savitzky Golay smoothing and orthogonal signal correction
614 pretreated cocoa (brown) and carob (grey) spectra in the range of 1100 to 2500 nm.

615

616 **Fig 4.** (a) NIR PCA score plot for the separation of pure cocoa powder and different
617 levels of adulteration with carob flour (high adulteration HA (40-60%), low adulteration
618 LA (0-20%) and medium adulteration MA (20-40%)).

619

620 **Fig 5.** NIR PLS-DA score plot from latent variable 1 and 2, pure cocoa blue, carob
621 powder grey and adulterations brown.

622

623 **Fig 6.** Predicted versus observed values of adulterant percentage (n = 140) since the
624 pure cocoa, carob powder to different levels of adulterated samples.

625

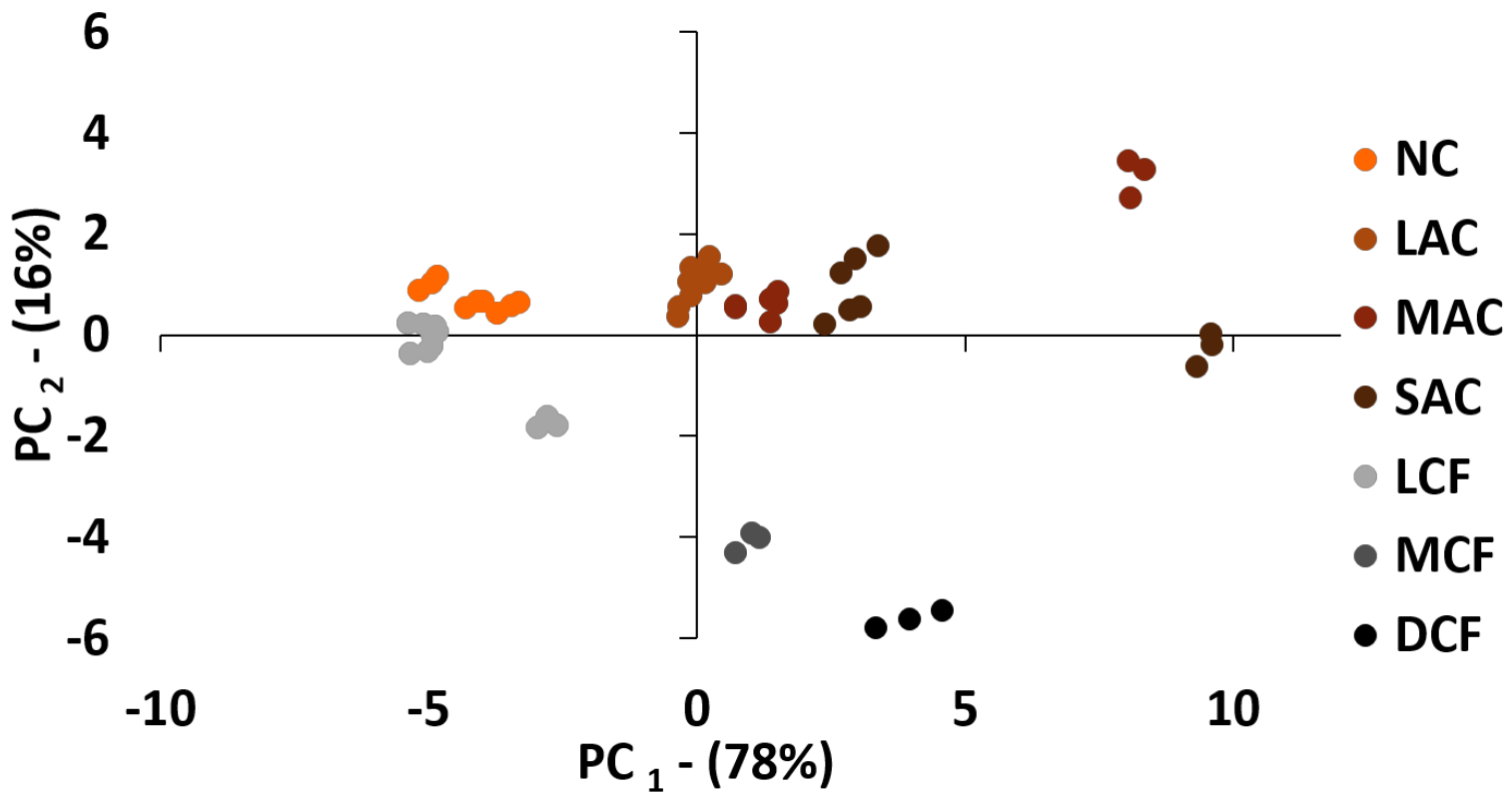

Fig 1. M.A. Quelal et al.

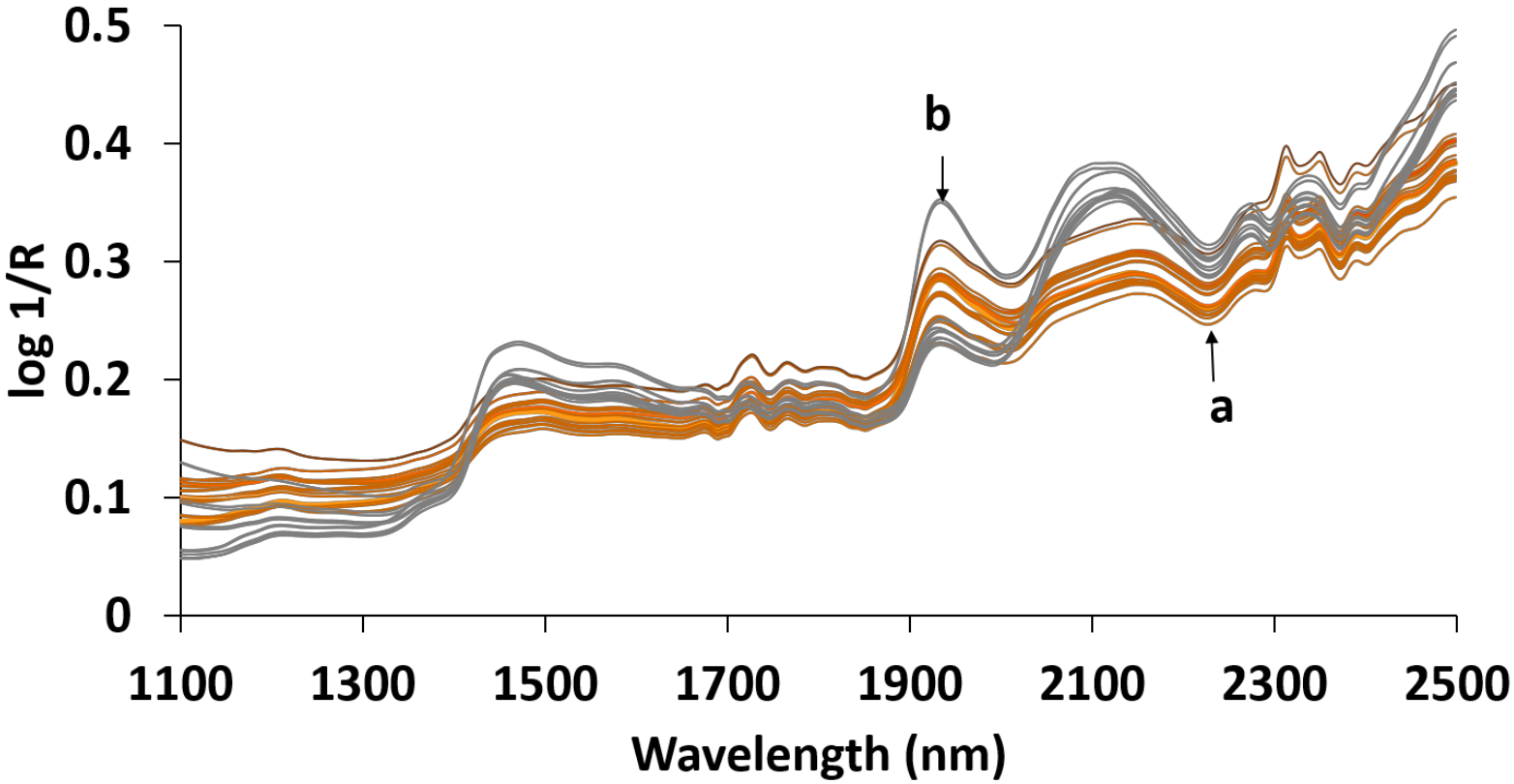

Fig 2. M.A. Quelal et al.

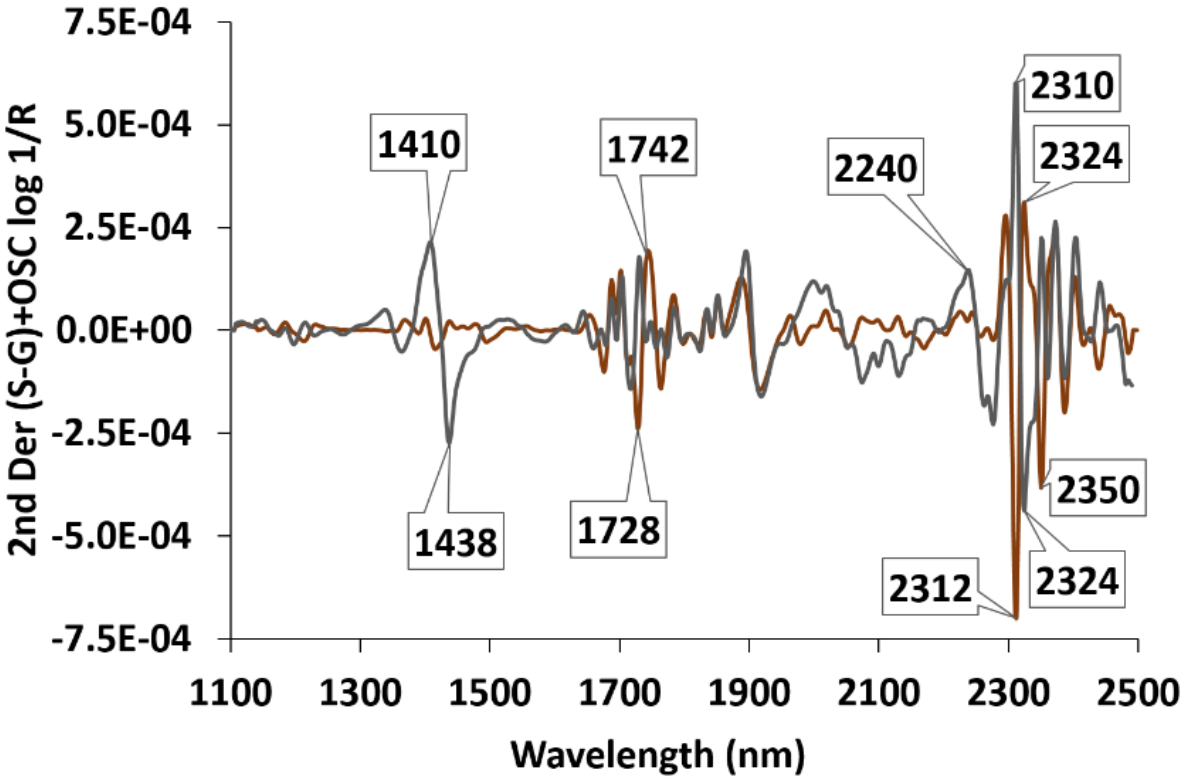

Fig 3. M.A. Quelal et al.

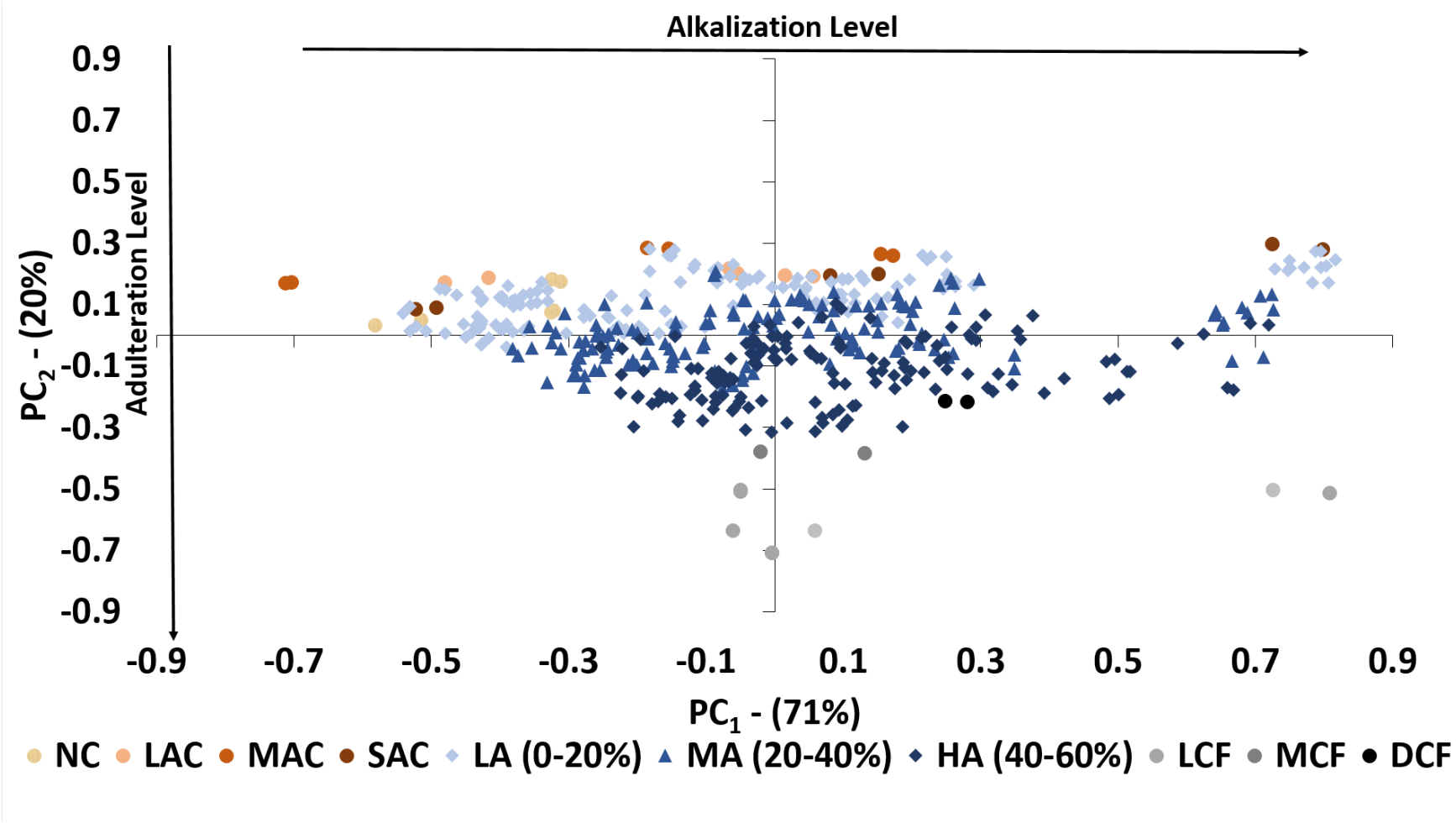

Fig 4. M.A. Quelal et al.

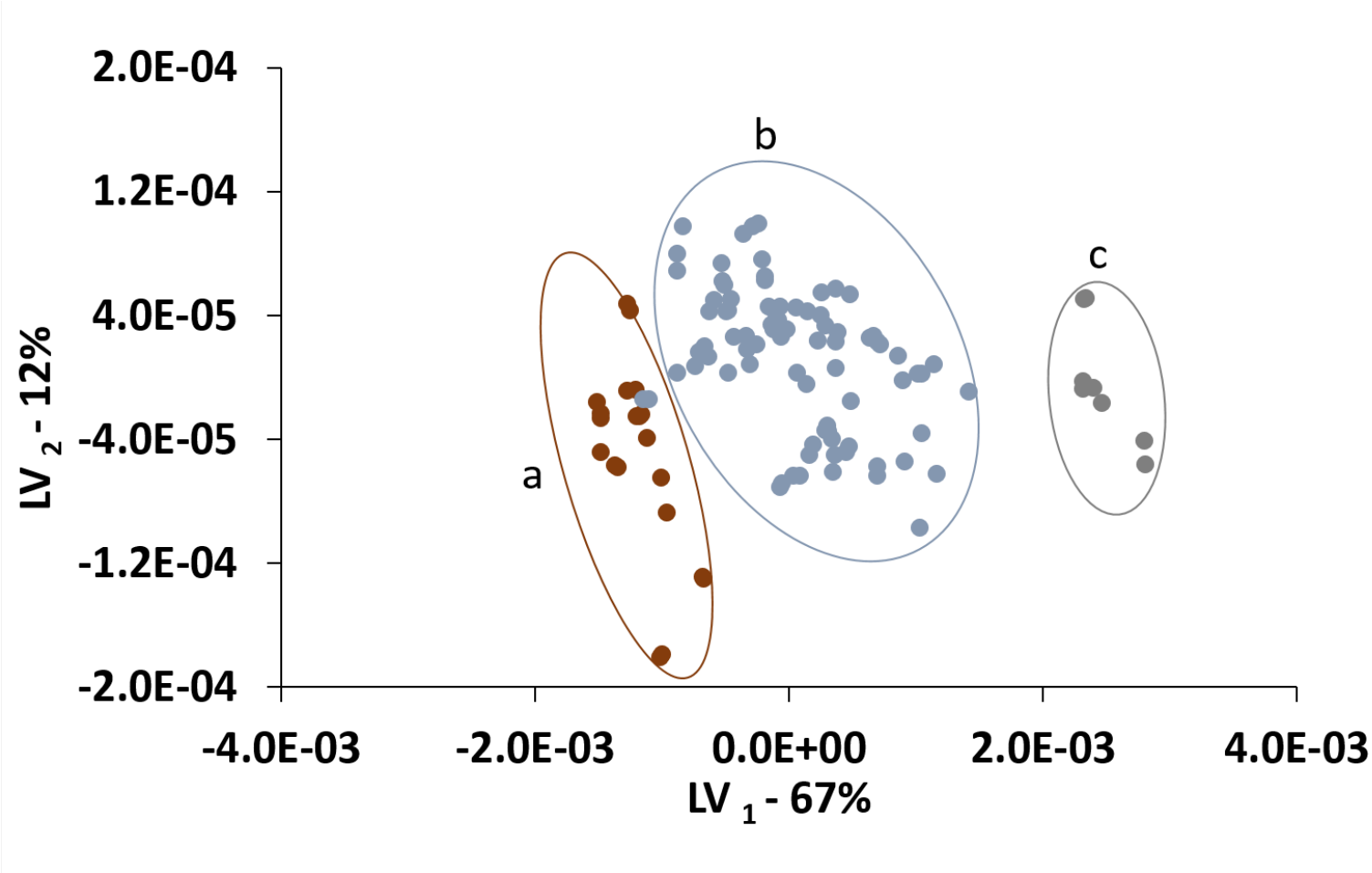

Fig 5. M.A. Quelal et al.

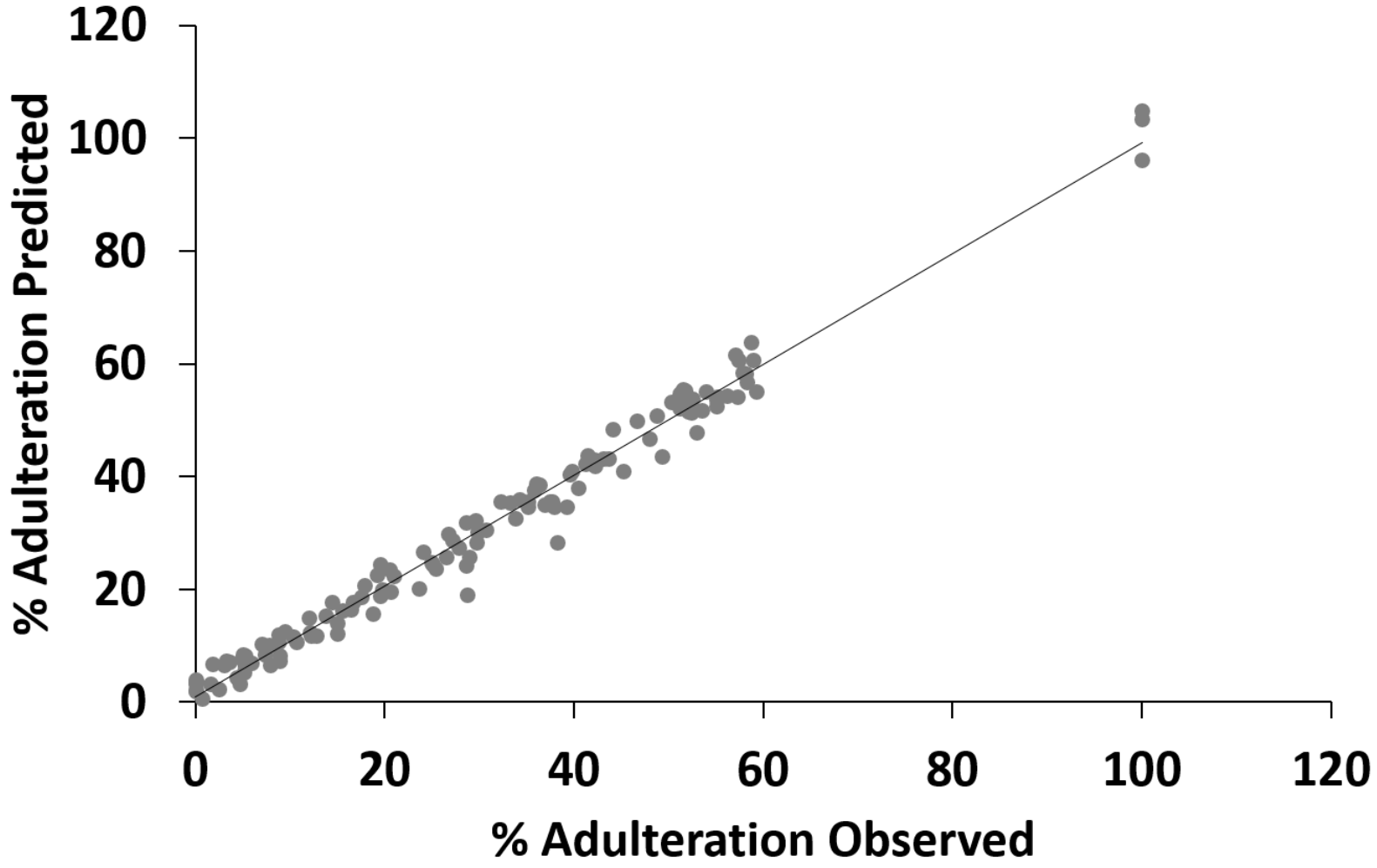

Fig 6. M.A. Quelal et al.

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}	5.033±0.012^{a}
LCF2	47.70±0.06^{de}	24.1±0.2^{e}	60.98±0.11^{b}	5.123±0.006^{a}
LCF3	46.1±0.2^{de}	26.3±0.7^{e}	61.5±0.3^{b}	4.667±0.006^{a}
LCF4	44.17±0.3^{de}	20.7±0.2^{e}	61.1±0.3^{b}	4.913±0.006^{a}
MCF	37.6± 0.4^{ab}	16.9±0.7^{bc}	60.2±0.5^{a}	4.850±0.010^{a}
DCF	34.5±1.5^{a}	12.9±0.9^{a}	60±2^{a}	4.867±0.006^{a}
NC1	48.7±0.2^{e}	20.1±0.5^{de}	58.8±0.4^{c}	5.390±0.010^{a}
NC2	48.33±0.13^{e}	22.3±0.4^{de}	59.5±0.3^{c}	5.457±0.006^{b}
NC3	50.3±0.6^{e}	22.19±1.02^{de}	60.0±0.4^{c}	5.703±0.006^{b}
LAC1	42.3±0.6^{c}	22.4±0.7^{cd}	54.3±0.4^{c}	6.903±0.015^{c}
LAC2	44.2±0.5^{c}	18.63±1.02^{cd}	55.0±0.9^{c}	6.963±0.021^{c}
LAC3	41.7±0.5^{b}	19.80±0.13^{bc}	54.5±0.5^{c}	6.987±0.006^{d}
MAC1	44.9±1.5^{c}	18±2^{cd}	55.7±0.6^{c}	7.243±0.006^{c}
MAC2	41.9±0.7^{b}	18.0±0.6^{bc}	54.2±0.5^{c}	7.340±0.026^{d}
MAC3	35.85±1.05^{b}	16.0±0.8^{bc}	43.0±0.6^{c}	7.430±0.010^{d}
SAC1	32.1±0.8^{a}	11.6±0.9^{b}	46.5±0.6^{c}	7.810±0.010^{e}
SAC2	39.4±0.5^{a}	19.76±0.99^{b}	51.4±0.8^{c}	7.837±0.006^{e}
SAC3	40.1±0.2^{a}	17.3±0.8^{b}	53.2±0.6^{c}	7.923±0.012^{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	4	0	0	100%
Carob	0	4	0	100%
Adulterated	0	0	32	100%

Table 3.

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

Pre-treatment	#LV	Calibration		Cross-validation		Validation				
		R^2_C	$RMSE_C$	R^2	$RMSE_{CV}$	R^2_P	$RMSE_P$	SEP	Bias	RPD
Raw data	7	0.951	4.530	0.945	4.785	0.961	4.397	4.400	0.197	4.66
2nd Der. S-G	5	0.978	3,082	0.974	3.28	0.979	3.271	3.195	0.749	6.39
OSC	1	0.975	3.165	0.975	3.214	0.974	3.555	3.537	0.474	5.75
2nd Der. (S-G)+OSC	1	0.980	2.856	0.979	2.897	0.974	3.237	3.187	0.626	6.41

2nd Der. S-G = Second derivative-Savitzky Golay; OSC = Orthogonal signal correction; #LV = latent variables; R^2_C = coefficient of determination for calibration; $RMSE_C$ = root mean square error of calibration; R^2_{CV} = coefficient of determination for cross-validation; $RMSE_{CV}$ = root mean square error of cross-validation; R^2_P = coefficient of determination for prediction; $RMSE_P$ = root mean square error of prediction; SEP = standard error of prediction; RPD = ratio of prediction deviation.