*Manuscript
Click here to view linked References

1 **Changes in methylxanthines and flavanols during cocoa powder processing and
2 their quantification by near-infrared spectroscopy**

3

4 Maribel Alexandra Quelal-Vásconez[a], María Jesús Lerma-García[a,]*, Édgar Pérez-
5 Esteve[a], Alberto Arnau-Bonachera[b,c], José Manuel Barat[a], Pau Talens[a]
6 [a] Departamento de Tecnología de Alimentos. Universitat Politècnica de València.
7 Camino de Vera, s/n 46022, Valencia, Spain
8 [b] Institute for Animal Science and Technology. Universitat Politècnica de València.
9 Camino de Vera, s/n 46022, Valencia, Spain
10 [c] Biomedical Research Institute (PASAPTA-Pathology group), Veterinary School,
11 Universidad Cardenal Herrera-CEU, CEU Universities, Av. Seminario s/n, 46113
12 Moncada, Valencia, Spain

13

14 ABSTRACT

15 Variation in methylxanthines (theobromine and caffeine) and flavanols (catechin and
16 epicatechin) was studied in a large set of cocoa powders (covering different origins,
17 processing parameters and alkalization levels). The content of these compounds was
18 established by high-performance liquid chromatography (HPLC), whose results showed
19 that the alkalisation process lowered the content of all analytes, whose loss was more
20 evident in flavanols. Therefore, the determination of these analytes in a huge set of
21 samples allowed not only better knowledge of the concentration variability in natural
22 commercial cocoas from different origins, but also the understanding of the effect that
23 industrial alkalisation has on these contents. The feasibility of reflectance near-infrared
24 spectroscopy (NIRS) combined with partial least square (PLS) to non-destructively
25 predict these contents, was also evaluated. All the analytes were generally well

26 predicted, with better predictions for methylxanthines (R^2_P 0.882 for both analytes;
27 RMSEP 0.020-0.061%, bias -0.027-0.006) than for flavanols (R^2_P 0.818-0.863; RMSEP
28 6.63-15.87%, bias 1.942-3.056). Thus NIRS could be an alternative fast reliable method
29 for the routine assessment of these analytes in the cocoa industry.

30

31 *Keywords:*

32 Cocoa powder

33 Methylxanthines

34 Flavanols

35 HPLC

36 Near-infrared spectroscopy

37

38 *Abbreviations:* HPLC, high performance liquid chromatography; GC, gas
39 chromatography; IR, infrared spectroscopy; NIRS, near infrared spectroscopy; PLS,
40 partial least square; PCA, principal component analysis; LV, latent variable; RMSEC,
41 root mean square error of calibration; RMSECV, root mean square error of cross-
42 validation; RMSEP, root mean square error of prediction; R^2_C, coefficient of
43 determination for calibration; R^2_{CV}, coefficient of determination for cross-validation;
44 R^2_P, coefficient of determination for prediction; RPD, ratio of prediction deviation;
45 LOD, limit of detection; LOQ, limit of quantification; S-G, Savitzky-Golay; OSC,
46 orthogonal signal correction.

47

48

49 **1. Introduction**

50

51 Cocoa powder is the most important raw material of confectionery products,
52 chocolate-flavoured bakeries, ice-creams and drinks (Miller et al., 2008). Apart from
53 technologic properties, cocoa (*Theobroma cacao* L.) and its derived products are well
54 considered for being a rich source of methylxanthines and polyphenols (Langer,
55 Marshall, Day, & Morgan, 2011).

56 Methylxanthines (i.e. theobromine and caffeine) are pharmacologically active
57 alkaloids responsible for a bitter cocoa taste and desirable physiological effects; e.g.
58 stimulation of the central nervous system and gastric secretion, diuresis,
59 bronchodilation, and stimulation of skeletal muscles in high doses (Franco, Oñatibia-
60 Astibia, & Martínez-Pinilla, 2013). They also display health benefits in diseases
61 involving cell death in the nervous system (Oñatibia-Astibia, Franco, &
62 Martínez-Pinilla, 2017).

63 The main type of polyphenols (known for their demonstrated antioxidant and anti-
64 inflammatory properties) in cocoas are flavanols. This family of compounds includes
65 catechin and epicatechin (monomeric species) and procyanidins (oligomeric and
66 polymeric fractions). Among them, epicatechin is the most abundant flavanol in cocoa
67 and accounts for 35% of the total polyphenolic fraction (Lacueva et al., 2008).

68 While producing cocoa powder from cocoa beans, seeds are primarily fermented,
69 dried and roasted. Then broken beans (nibs) are ground, heated and liquefied. The
70 product of these operations, cocoa liquour, is pressed to obtain two different fractions:
71 cocoa powder and butter. Optionally, nibs or cake can be treated with an alkali
72 dissolved in water. This alkalisation reduces acidity, bitterness and astringency by
73 improving and enhancing the aromatic features of cocoa powder (Kongor et al., 2016).

74 Alkalisation also allows colour development by transforming the typical light brown
75 hues of natural cocoa powder into reddish or very dark ones. These colour changes
76 derive from a sequence of chemical reactions between alkalizing agents and pigments in
77 the presence of water and oxygen at high temperatures. Finally, the dispersibility of
78 cocoa powder increases with alkalisation. This property facilitates the use of cocoa
79 powders in different industries like dairy products (Afoakwa, Paterson, Fowler, & Ryan,
80 2008).

81 Cocoa powder processing not only strongly affects the sensory properties of cocoa
82 and derived products, but also alters flavonoid and methylxanthine fractions (Payne,
83 Hurst, Miller, Rank, & Stuart, 2010). However, studies to date have used small sample
84 numbers and have, thus provided conclusions that can be read only as behaviour
85 tendencies.

86 The commonest techniques employed to analyse methylxanthines and flavanols
87 from cocoa extracts or isolated fractions are high-performance liquid chromatography
88 (HPLC) or gas chromatography (GC) (Cádiz-Gurrea et al., 2014; Elwers, Zambrano,
89 Rohsius, & Lieberei, 2009; Fayeulle et al., 2018, Humston, Knowles, McShea, &
90 Synovec, 2010; Machonis, Jones, Schaneberg, Kwik-Uribe, & Dowell, 2014; Van
91 Durme, Ingels, & De Winne, 2016), which provide reliable and accurate descriptions of
92 individual cocoa qualities. Recently, the use of novel methods based on on-line
93 comprehensive two-dimensional liquid chromatography coupled to tandem mass
94 spectrometry (LC × LC-MS/MS) have allowed the characterization of new secondary
95 metabolites of cocoa beans (Toro-Uribe, Montero, López-Giraldo, Ibáñez, & Herrero,
96 2018). However, these methods are not recommended for routine raw material quality
97 controls as they are destructive, require specialised personnel, sample preparation and
98 expensive equipments. Thus simpler, faster and non-destructive techniques are required.

99 An alternative is infrared spectroscopy (IR is a fast non-destructive analytical tool
100 that needs little samples preparation), which is useful for both qualitative and
101 quantitative analyses of molecules. Finally, the application of chemometric techniques
102 to IR data provides a powerful tool to develop methods capable of classifying or
103 characterising samples (Vergara-Barberán, Lerma-García, Herrero-Martínez, & Simó-
104 Alfonso, 2015).

105 Recent studies into commercially cocoa derivates have shown the suitability of near
106 IR (NIR) for the quantification of main constituents like fat, protein, moisture and
107 carbohydrates (Veselá et al., 2007). NIRS has also been shown appropriate to determine
108 parameters like geographical origin (Teye, Huang, Dai, & Chen, 2013), fermentation
109 quality (Hue et al., 2014), fermentation index or pH (Sunoj, Igathinathane, &
110 Visvanathan, 2016), and minor valuable components like organic acids (Krähmer et al.,
111 2015), caffeine, theobromine and epicatechin in unfermented and sun-dried beans
112 (Álvarez, Pérez, Cros, Lares, & Assemat, 2012), procyanidins (Whitacre et al., 2003)
113 and adulterations with carob flour (Quelal-Vásconez, Pérez-Esteve, Arnau-Bonachera,
114 Barat, & Talens, 2018) and cocoa shell (Quelal-Vásconez, Lerma-García, Pérez-Esteve,
115 Arnau-Bonachera, Barat, & Talens, 2019). Despite these advances in cocoa
116 characterisation by NIRS, as far as we know, the effect of cocoa powder processing on
117 methylxanthine and flavanol content by this technique has not yet been studied.

118 In this context, the objective of this work is twofold. Firstly, to study the effect of
119 cocoa powder processing on methylxanthines (theobromine and caffeine) and flavanols
120 (catechin and epicatechin) contents in a large batch of samples (with different origins,
121 processing parameters and alkalisation levels). Secondly, to evaluate the feasibility of
122 reflectance NIRS combined with partial least square (PLS) to non-destructively predict
123 the content of these compounds in cocoa powders.

124

125 **2. Materials and methods**

126

127 *2.1. Reagents and samples*

128

129 The employed reagents were: caffeine, theobromine, catechin and epicatequin
130 (Sigma-Aldrich, St. Louis, Missouri, USA), acetonitrile (J.T. Baker, The Netherlands),
131 methanol (Labkem, Barcelona, Spain) and acetic acid glacial (Sharlau, Barcelona,
132 Spain). Deionised water was obtained using an Aquinity deioniser (Membrapure
133 GmbH, Berlin, Germany).

134 Samples comprised 86 cocoa powders provided by Olam Food Ingredients
135 Company (Cheste, Spain) or purchased in different nationals and international markets,
136 to assure variability in cocoa origin (South America, Africa And Asia), year of
137 production (2017 or 2018) and processing practices (natural or alkalised cocoa
138 powders). The origins of samples were Ivory Coast, Ghana, Indonesia, Ecuador, Peru,
139 or undeclared.

140 Before analysing, samples were characterised according to their extractable pH and
141 colour. Extractable pH determination was performed by the procedure described in the
142 ADM Cocoa Manual (2017). CIE L*a*b colour coordinates were obtained by
143 measuring the reflection spectrum using a 10° observer and D65 illuminant (Minolta,
144 CM 3600D, Tokyo, Japan). Extractable and colour data were used to classify cocoa
145 powders into five categories: natural (NC) (pH 5.0-6.0), light-alkalised (LAC) (pH 6.0-
146 7.2), medium-alkalised (MAC) (pH 7.2-7.6), strong-alkalised (SAC) (pH > 7.6) and
147 black powders (BC) (pH > 7.6 and very low L values) (Miller et al., 2008).

148

149 *2.2. Instrumentation and experimental conditions*

150

151 *2.2.1. HPLC determination of methylxanthines and flavanols*

152 In order to extract methylxanthines and flavanols from cocoa powders, the
153 protocol of Lacueva et al. (2008) was adapted: 0.5 g of cocoa powder was weighed,
154 suspended in 5 mL H_2O at 100ºC+20 mL methanol, mixed for 20 min at 36ºC by
155 constant agitation, and centrifuged for 10 min at 10,000 rpm at room temperature.
156 Finally, the obtained supernatants were filtered using 0.22 μm pore size PTFE filters
157 (Scharlab, Barcelona, Spain). The obtained samples were immediately injected into the
158 HPLC system or stored at -20ºC.

159 Analyte determination was performed in a liquid chromatograph model
160 LaChrom Elite (Hitachi Ltd., Tokyo, Japan), equipped with an auto-sampler and a UV
161 detector (models L-220 and L-2400, respectively). A 5-μm analytical column was used,
162 model Liquid Purple C18 (250 x 4.6 mm i.d.) from Análisis Vínicos (Tomelloso,
163 Spain). The mobile phase was prepared by mixing 2% aqueous acetic acid (phase A)
164 and a ternary mixture composed of acetonitrile–H_2O–acetic acid in the 40:9:1 v/v/v
165 proportion (phase B). The following gradient (adapted from Elwers et al., 2009) was
166 used: 0-8 min, 10% B; 8-20 min, 10-15% B; 20-35 min, 15-90% B; 35-50 min, 90% B,
167 then back to the initial conditions. Other chromatographic conditions were: UV
168 detection, 280 nm; column temperature, 40ºC; injection volume, 20 µL; flow rate, 1.2
169 mL min^{-1}.

170 The differences in analyte content with the various different alkalisation levels
171 were established by a multifactorial ANOVA (95% confidence level of LSD; $p<0.05$),
172 constructed using Statgraphics Centurion XV from Manugistics Inc. (Rockville, MD,
173 USA).

174

175 *2.2.2. NIR spectra collection*

176 The 86 cocoa powders were scanned in a FOSS NIR 5000 System
177 spectrophotometer, equipped with a transport module (Silver Spring, MD, USA).
178 Samples (ca. 5.0 g) were measured by filling a round sample cup (3.8 cm diameter x 1
179 cm-thick quartz windows) to preserve unvarying surface and thickness during spectral
180 collection. For each sample, 32 sequential scans with 700 wavelengths were measured
181 from 1100 and 2500 nm at 2-nm intervals. Samples were measured twice, and the mean
182 spectra were employed for the statistical analysis.

183

184 *2.3. Spectral data analysis*

185 Calibration models to predict the methylxanthines (theobromine and caffeine)
186 and flavanols (catechin and epicatechin) contents (previously established by HPLC)
187 were constructed by PLS analysis of the spectral data using Unscrambler v10.4 software
188 from the CAMO Software AS (OSLO, Norway). PLS is a multivariate calibration
189 method that could be used to correlate NIR spectra data with chemical component
190 contents. Spectral data were organised in a matrix containing the number of cocoa
191 samples (N = 86) in rows and the X- and Y-variables in columns. The X-variables
192 corresponded to the 700 wavelengths, while the Y-variables were the HPLC-determined
193 analyte contents. For PLS model construction, the 86 samples were randomly separated
194 into two datasets: 80% samples were employed to create and evaluate the PLS models
195 by leave-one-out cross-validation, while the remaining 20% samples were used as an
196 external validation set. Since samples belonged to different origins, years of production
197 and processing practices both sets were considered independent.

198 Before any spectral treatment, all spectra were used to construct a principal
199 component analysis (PCA) model, which was employed to identify and remove
200 defective spectral outliers using the Q residual values and the Hotelling T^2 with a 95%
201 confidence limit (Bro & Smilde, 2014). Moreover, with the aim of detecting similarities
202 among the samples, NIR spectra data were employed to build a clustering model by
203 using the hierarchical complete-linkage method (HCL).

204 Finally, PLS models were constructed with no spectral pretreatment (raw data)
205 and with three different pre-treatments: 2nd derivative performed with the Savitzky-
206 Golay smoothing algorithm (2nd derivative S-G) (Savitzky & Golay, 1951), orthogonal
207 signal correction (OSC) (Wold, Antti, Lindgren, & Öhman, 1998), and their
208 combination. The PLS models' accuracy was evaluated by: the required number of
209 latent variables (LVs), the root mean square error of calibration (RMSEC) and cross-
210 validation (RMSECV), and the coefficient of determination for calibration (R^2_C) and
211 cross-validation (R^2_{CV}). The PLS models' predictive capability was judged by
212 considering: the coefficient of determination for prediction (R^2_P), the root mean square
213 error of prediction (RMSEP), the bias, the standard error of prediction (SEP) and the
214 ratio of prediction deviation (RPD) (calculated as the ratio between the standard
215 deviation of the reference values -training set- and RMSEP). Performance of the
216 different models was considered good when the number of LVs, and RMSE and SE
217 values were low, when R^2 value tended to unit, when RPD was higher than 3 and when
218 bias values were near to zero, for both the calibration and external prediction parameters
219 (Quelal-Vásconez et al., 2019).

220

221 **3. Results and discussion**

222

223 *3.1. Alkalisation level evaluation*

224 The 86 cocoa powders were physico-chemically characterised and divided into
225 five categories according to their extractable pH values and colour (Miller et al., 2008).
226 By contemplating these values, 23 cocoa samples were classified as natural, 19 as light-
227 alkalised, 21 as medium-alkalised, 19 as strong-alkalised and 4 as black powders.
228 Considering the linear correlation between alkalisation and pH (Pérez-Esteve, Lerma-
229 García, Fuentes, Palomares, & Barat, 2016), this classification was taken into account to
230 study the relationship between alkalisation intensity and the changes found in caffeine,
231 theobromine, catechin and epicatechin contents.

232

233 *3.2. HPLC determination of methylxanthines and flavanol contents in cocoa powders*

234 By applying the experimental conditions included in Section 2.2.1, analyte peak
235 identification was achieved by comparing the retention times of the sample peaks with
236 the retention times of the standards. Four peaks were observed at retention times of 5.1,
237 9.1, 12.0 and 15.8 min, which respectively corresponded to theobromine, catechin,
238 caffeine and epicatechin (see Fig. 1 as an example). These compounds were quantified
239 using external calibration curves. To construct them, six standard solutions at different
240 concentrations within the ranges showed in Table 1 were prepared and injected. In all
241 cases, R^2 above 0.9996 were observed. The other parameters in Table 1 were limit of
242 detection (LOD) and limit of quantification (LOQ), which were estimated following the
243 ICH guidelines (1996). The obtained values ranged between 0.005-0.1 and 0.017-0.33
244 mg L^{-1} for LODs and LOQs, respectively. These values were lower than those
245 previously reported by others (Gottumukkala, Nadimpalli, Sukala, & Subbaraju, 2014;
246 Risner, 2008; Srdjenovic, Djordjevic-Milic, Grujic, Injac, & Lepojevic, 2008). In order
247 to assure that no matrix effect was observed in the quantification of analytes, standard

248 addition calibration curves (considering the linearity ranges in Table 1) were
249 constructed. The four curves provided R^2 above 0.9995 and similar slopes to the
250 external calibration curves. Therefore, it was concluded that the external calibration
251 curves were correctly used to quantify these analytes in cocoa powders.

252 Finally, the 86 cocoa powders were subjected to HPLC (see Table 2 for the
253 results). The most abundant alkaloid was theobromine, whose content ranged from
254 1.53±0.03 to 2.41±0.18 g/100g when considering all the samples. Statistical differences
255 in theobromine content were found among alkalisation ($p<0.05$) as BC possessed lower
256 contents than the other categories. The caffeine content ranged from 0.1503±0.0006 to
257 0.412±0.006 g/100g. As with theobromine, significant differences were observed
258 among samples with different alkalisation degrees ($p<0.05$), and this behaviour has been
259 reported by other authors (Brunetto et al. 2007; Li et al., 2012).

260 Among flavanols, epicatechin was the most abundant analyte. The highest
261 content (536.59±0.13 mg/100g) went to a NC sample labelled as Equator origin. Other
262 natural samples exhibited an average content of ca. 160 mg/100 g. Contents statistically
263 lowered as the alkalisation process became more intense, and reached average values of
264 ca. 80, 33, 28 mg/100 g and with LODs in the LAC, MAC, SAC and BC, respectively.
265 The same behaviour was found for catechin contents. In the NC samples, catechin
266 content ranged from 15.2±0.9 to 167±2 mg/100g. The highest value went to another
267 sample from Equator. In the other natural samples, the average value was ca. 80 mg/100
268 g. According to Table 2, no significant differences in catechin content were found in the
269 LAC cocoa powders, but contents statistically decreased in the medium-alkalised
270 powders, with undetected quantities in the black powders. Similar results have been
271 found by other authors. In a study done with 11 cocoa powder samples, Lacueva et al
272 (2008) reported values of 200 and 70 mg/100 g for epicatechin and catechin in natural

273 powders, and 30 and 25 mg/100 g in alkalised powders, respectively. Similar contents
274 were observed by Payne et al. (2010), who reported 223 and 88 mg/100 g (natural), 69
275 and 70 mg/100 g (light-alkalised), 26 and 36 mg/100 g (medium-alkalised) and 4 and 9
276 mg/100 g (strong-alkalised) for epicatechin and catechin, respectively.

277

278 *3.3 Clustering analysis of the samples*

279 The clustering analysis of the NIR spectra shows four main groups (see Fig, 2).
280 First group comprised only one sample of BC, while the second one included the other
281 BCs and ca. 16% of the SAC samples. The third group included most NCs, ca. 74% of
282 LACs, ca. 20% of MACs and ca. 6% of SACs. Finally, the fourth group included the
283 81% of MACs, ca. 79% of SACs, 26% of LACs and the rest of NCs. In general, it can
284 be observed that BC samples (minimum content of catechin and epicatechin detected)
285 belonged to the first and second group and that the rest groups are mostly linked to
286 strong (group 4) or mild alkalization conditions (group 3). However, this grouping
287 cannot be completely linked neither to the content or flavanols nor methylxanines. For
288 instance, samples with the highest flavanols content (Ecuadorian samples –i.e. NC 67 or
289 NC 52-) are not clearly separated from other natural or alkalised samples. Therefore, it
290 could be concluded that in sample clustering, besides quantified analytes or alkalization
291 degrees, there are other chemical signals (i.e. proteins, sugars, volatiles, etc) that are
292 affecting sample clustering.

293

294 *3.4. Prediction of theobromine, catechin, caffeine and epicatechin content by PLS*

295 The raw spectra of the samples (between 1100-2500 nm) are shown in Fig. 3a.
296 As this spectral range could contain both useful and irrelevant information, it was pre-
297 treated to enhance the final outcome. For all spectra, the first applied pre-treatment was

298 the 2nd derivative S-G, followed by the OSC pre-treatment and their combination. Then,
299 the optimal pre-treatment method to enhance the PLS models' performance was
300 selected by taking into account the values included in Table 3. As can be observed in
301 this table, these values differed depending on the analyte to be predicted. To predict
302 theobromine and catechin contents, the best results were observed after applying the 2nd
303 derivative S-G+OSC; to predict caffeine content, the OSC pre-treatment provided the
304 best results, while no pre-treatment was selected for epicatechin. The spectra obtained
305 after applying the OSC pre-treatment and 2nd derivative S-G + OSC pre-treatments are
306 shown in Fig. 3b and 3c, respectively. As seen in Fig. 3b, the main signal peak was
307 observed at 1944 nm, while in Fig. 3c, signal peaks at wavelengths of 1728, 1764, 1884,
308 2312 and 2348 nm were evidenced. The region comprised between 1600–1800 nm
309 predominantly corresponds to the first overtone region of carbohydrates (C–H bands)
310 (Bázár et al., 2016), and the region comprised between 1700 and 2300 nm normally
311 attributes to the first overtones of C-H stretching associated with sugars (Osborne,
312 Fearn, & Hindle 1993).

313 Table 3 contains the results obtained for the PLS models constructed to predict
314 the different analytes. As seen for the PLS models obtained with the optimal pre-
315 treatment (marked in italics), all the PLS models provided satisfactory R^2_C, R^2_{CV} and
316 R^2_P values within ranges 0.949-0.990, 0.931-0.972 and 0.818-0.882, respectively. For
317 methylxanthine content predictions, the RMSEP values were 0.061 and 0.020 and the
318 bias came very close to 0 (-0.027 and 0.006) for theobromine and caffeine, respectively.
319 With flavanols, the models provided RMSEP values of 6.63 and 15.87, and bias values
320 of 1.942 and 3.056 for catechin and epicatechin, respectively. Notwithstanding, the
321 RPD values for all the models were above 3.0, which are acceptable for quantitative
322 predictions according to the literature (Quelal-Vásconez, et al., 2019). The good fit

323 between the different analytes content measured by HPLC and the contents predicted by
324 the PLS models for the evaluation set samples is shown in Fig. 4.
325 Finally, the wavelengths with a better prediction capability according to the b
326 vector profiles are shown in Fig. 5. The most important wavelengths related to variation
327 in theobromine content (see Fig. 5a) are found at 1414, 1536, 1674, 1682, 1710, 1718,
328 1764, 2092, 2308 and 2322 nm, among others, and agree with those described by
329 Álvarez et al. (2012). Among these wavelengths, 1414, 1682, 1718, 1764, 2092 and
330 2308 nm increased as theobromine content rose, while the rest lowered. According to
331 literature (Veselá et al., 2007), these variations are characterised mainly by the
332 stretching of H_2O of weakly bounded water, proteins and aromatics, the first overtone of
333 stretching of CH of aromatics, the first overtone of the symmetric and asymmetric
334 stretching vibration of CH_2, and the stretching and rocking vibrations of CH_2 of
335 polysaccharides and fats.
336 For catechin content predictions, the most important wavelengths were 1414,
337 1436, 1880, 1908, 2252, 2312 and 2360 nm (see Fig. 5b). Among them, 1414, 1908 and
338 2312 nm increased as catechin content rose, while the rest lowered. These variations
339 may be associated with the stretching of H_2O of weakly bounded water, proteins and
340 aromatics and also of non-bounded water, the asymmetric stretching and rocking of
341 H_2O of very weakly bounded water fat, asymmetric stretching and rocking of CH_2 of
342 polysaccharides and the stretching and rocking of CH and CC (Veselá et al., 2007).
343 In the case of caffeine (see Fig. 5c), the most important wavelengths are located
344 at 1434, 1538, 1730, 2060, 2140, 2292 and 2378 nm (which increase as the caffeine
345 content raised) and the wavelengths at 1700, 2082, 2322, 2340 and 2442 nm (which
346 decrease as the caffeine content raised). In this case, the wavelengths corresponded to
347 the stretching of H_2O (non-bounded water), the first overtone of asymmetric stretching

348 vibration of CH_2, proteins, combination of amides, the stretching of CH and CC and the
349 stretching and rocking of CH_2 of fats (Krähmer et al., 2015).

350 Finally, and as shown in Fig. 5d for epicatechin prediction, the most important
351 wavelengths related to the variation of epicatechin are the 1440, 2080, and 2424 nm
352 bands (which increase as the epicatechin content increases), and the 1872 and 2292 nm
353 bands (which decrease as the epicatechin content raises). These bands could be assigned
354 to the stretching of H_2O (non-bounded water), the asymmetric stretching and rocking of
355 H_2O of very weakly bounded water fat, proteins and the stretching of CH and C=C
356 (Veselá et al., 2007). Similar NIR chemical vibrations has been found in the
357 quantification of total phenols and carotenoids in blackberries (Toledo-Martín et al.,
358 2018), and in the determination of the fat, caffeine, theobromine of sun dried cocoa
359 beans (Álvarez et al., 2012)

360

361 **4. Conclusions**

362

363 The HPLC determination of methylxanthines and flavanols of a large collection
364 of cocoa powders provided not only better knowledge of the concentration variability in
365 natural cocoas from different origins, but also the understanding of the effect that
366 industrial alkalisation has on those contents. A wide fluctuation in flavanol content was
367 observed for natural powders, which highlights the natural variability of unprocessed
368 samples given their different origins. Moreover, the effect of cocoa alkalisation on the
369 content of all analytes was evidenced. Despite such evidence, the many analysed
370 samples indicated strong-alkalised powders with higher analyte contents than some
371 natural cocoa powders. This reinforces the importance of measuring the content of these
372 analytes during raw material selection and in all the industrial processing steps when

373 functional products want to be launched on the market. In line with this, the possibility
374 of predicting the content of these functional analytes by a fast, non-destructive and
375 reliable methodology, such as NIRS, was confirmed. Despite the fact that the clustering
376 analysis did not allowed a sample grouping according to the alkalization degree or the
377 flavanols and/or methylxantines content by applying PLS models, all analytes were
378 satisfactorily predicted (RPD values > 3.0). Among the different families of analytes,
379 methylxanthine predictions led to better R^2_p, RMSEP and bias values than those
380 obtained for flavanols. Therefore, the present results bridge the information gap in the
381 cocoa sector about the variability found in these functional compounds in commercial
382 samples, and also propose a fast reliable methodology to establish the content of these
383 important functional compounds for the cocoa industry.
384

385 **Acknowledgments**

386 The authors would like to acknowledge the financial support of the Spanish
387 Government and European Regional Development Fund (Project RTC-2016-5241-2).
388 M. A. Quelal thanks the Ministry of Higher Education, Science, Technology and
389 Innovation (SENESCYT) of the Republic of Ecuador for her PhD grant. Olam Food
390 Ingredients Company is aknowledged for proving part of the cocoa samples used in the
391 study.
392

393 **Conflicts of interest**

394 The authors declare that they have no conflict of interest.
395

396 **References**

397

398 ADM Cocoa Manual. (2006). The De Zaan® cocoa manual. Archer Daniels Midland
399 Company BV. The Netherlands.

400 Afoakwa, E. O., Paterson, A., Fowler, M., & Ryan, A. (2008). Flavor Formation and
401 Character in Cocoa and Chocolate: A Critical Review. *Critical Reviews in Food*
402 *Science and Nutrition, 48*(9), 840–857.

403 Álvarez, C., Pérez, E., Cros, E., Lares, M., Assemat, S., Boulanger, R., & Davrieux, F.
404 (2012). The use of near infrared spectroscopy to determine the fat, caffeine,
405 theobromine and (–)-epicatechin contents in unfermented and sun-dried beans of
406 Criollo cocoa. *Journal of Near Infrared Spectroscopy, 20*(2), 307-315.

407 Bázár, G., Romvári, R., Szabó, A., Somogyi, T., Éles, V., & Tsenkova, R. (2016). NIR
408 detection of honey adulteration reveals differences in water spectral pattern. *Food*
409 *Chemistry, 194*, 873–880.

410 Bro, R., & Smilde, A. K. (2014). Principal component analysis. *Analytical Methods,*
411 *6*(9), 2812–2831.

412 Brunetto, M. del R., Gutiérrez, L., Delgado, Y., Gallignani, M., Zambrano, A., Gomez,
413 A., & Romero, C. (2007). Determination of theobromine, theophylline and caffeine
414 in cocoa samples by a high-performance liquid chromatographic method with on-
415 line sample cleanup in a switching-column system. *Food Chemistry*, *100*(2), 459-
416 467.

417 Cádiz-Gurrea, M. L., Lozano-Sanchez, J., Contreras-Gámez, M., Legeai-Mallet, L.,
418 Fernández-Arroyo, S., & Segura-Carretero, A. (2014). Isolation, comprehensive
419 characterization and antioxidant activities of Theobroma cacao extract. *Journal of*
420 *Functional Foods, 10*, 485–498.

421 Elwers, S., Zambrano, A., Rohsius, C., & Lieberei, R. (2009). Differences between the
422 content of phenolic compounds in Criollo, Forastero and Trinitario cocoa seed
423 (Theobroma cacao L.). *European Food Research and Technology, 229*(6), 937–
424 948.
425 Fayeulle, N., Vallverdu-Queralt, A., Meudec, E., Hue, C., Boulanger, R., Cheynier, V.,
426 & Sommerer, N. (2018). Characterization of new flavan-3-ol derivatives in
427 fermented cocoa beans. *Food Chemistry, 259*, 207–212.
428 Franco R., Oñatibia-Astibia, A., & Martínez-Pinilla, E. (2013). Health benefits of
429 methylxanthines in cacao and chocolate. *Nutrients, 5*(10), 4159–73.
430 Gottumukkala, R. V. S. S., Nadimpalli, N., Sukala, K., & Subbaraju, G. V. (2014).
431 Determination of Catechin and Epicatechin Content in Chocolates by High-
432 Performance Liquid Chromatography. *International Scholarly Research Notices,*
433 *2014*, 1–15.
434 Hue, C., Gunata, Z., Bergounhou, A., Assemat, S., Boulanger, R., Sauvage, F. X., &
435 Davrieux, F. (2014). Near infrared spectroscopy as a new tool to determine cocoa
436 fermentation levels through ammonia nitrogen quantification. *Food Chemistry,*
437 *148*, 240–245.
438 Humston, E. M., Knowles, J. D., McShea, A., & Synovec, R. E. (2010). Quantitative
439 assessment of moisture damage for cacao bean quality using two-dimensional gas
440 chromatography combined with time-of-flight mass spectrometry and
441 chemometrics. *Journal of Chromatography A, 1217*(12), 1963–1970.
442 International Conference on Harmonization (ICH guidelines). (1996). Validation of
443 analytical procedures: Text and methodology. Geneva: ICH-Q2.
444 Kongor, J. E., Hinneh, M., de Walle, D. Van, Afoakwa, E. O., Boeckx, P., &
445 Dewettinck, K. (2016). Factors influencing quality variation in cocoa (Theobroma

446 cacao) bean flavour profile - A review. *Food Research International, 82*, 44–52.
447 Krähmer, A., Engel, A., Kadow, D., Ali, N., Umaharan, P., Kroh, L. W., & Schulz, H.
448 (2015). Fast and neat – Determination of biochemical quality parameters in cocoa
449 using near infrared spectroscopy. *Food Chemistry, 181*, 152–159.
450 Lacueva, C. A., Monagas, M., Khan, N., Izquterdo-Pulido, M., Urpi-Sarda, M.,
451 Permanyer, J., & Lamuela-Raventós, R. M. (2008). Flavanol and flavonol contents
452 of cocoa powder products: Influence of the manufacturing process. *Journal of*
453 *Agricultural and Food Chemistry*, *56*(9), 3111–3117.
454 Langer, S., Marshall, L. J., Day, A. J., & Morgan, M. R. (2011). Flavanols and
455 methylxanthines in commercially available dark chocolate: a study of the
456 correlation with nonfat cocoa solids. *Journal of agricultural and food chemistry*,
457 *59*(15), 8435–8441.
458 Li, Y., Feng, Y., Zhu, S., Luo, C., Ma, J., & Zhong, F. (2012). The effect of alkalization
459 on the bioactive and flavor related components in commercial cocoa powder.
460 *Journal of Food Composition and Analysis*, *25*(1), 17-23.
461 Machonis, P., Jones, M., Schaneberg, B., Kwik-Uribe, C., & Dowell, D. (2014). Method
462 for the determination of catechin and epicatechin enantiomers in cocoa-based
463 ingredients and products by high-performance liquid chromatography: First action
464 2013.04. *Journal of AOAC International, 97*(2), 506–509.
465 Miller, K. B., Hurst, W. J., Payne, M. J., Stuart, D. A., Apgar, J., Sweigart, D. S., & Ou,
466 B. (2008). Impact of alkalization on the antioxidant and flavanol content of
467 commercial cocoa powders. *Journal of Agricultural and Food Chemistry, 56*(18),
468 8527–8533.

469 Oñatibia- Astibia, A., Franco, R., & Martínez- Pinilla, E. (2017). Health benefits of
470 methylxanthines in neurodegenerative diseases. *Molecular Nutrition & Food*
471 *Research*, *61*(6), 1600670.
472 Osborne, B. G., Fearn, T., & Hindle, P.H. (1993). Practical NIR spectroscopy with
473 applications in food and beverage analysis. *Burnt Mill, Harlow, Essex, England:*
474 *Longman Group,* 2nd ed., 123–132.
475 Payne, M. J., Hurst, W. J., Miller, K. B., Rank, C., & Stuart, D. A. (2010). Impact of
476 fermentation, drying, roasting, and dutch processing on epicatechin and catechin
477 content of cacao beans and cocoa ingredients. *Journal of Agricultural and Food*
478 *Chemistry, 58*(19), 10518–10527.
479 Pérez-Esteve, É., Lerma-García, M. J., Fuentes, A., Palomares, C., & Barat, J. M.
480 (2016). Control of undeclared flavoring of cocoa powders by the determination of
481 vanillin and ethyl vanillin by HPLC. *Food Control, 67*, 171–176.
482 Quelal-Vásconez, M. A., Pérez-Esteve, É, Arnau-Bonachera, A., Barat, J. M., & Talens,
483 P. (2018). Rapid fraud detection of cocoa powder with carob flour using near
484 infrared spectroscopy. Food Control, *92*, 183–189.
485 Quelal-Vásconez, M. A., Lerma-García, M. J., Pérez-Esteve, É., Arnau-Bonachera, A.,
486 Barat, J. M., & Talens, P. (2019). Fast detection of cocoa shell in cocoa powders
487 by near infrared spectroscopy and multivariate analysis. *Food Control, 99*, 68–72.
488 Risner, C. H. (2008). Simultaneous Determination of Theobromine, (+)-Catechin,
489 Caffeine, and (-)-Epicatechin in Standard Reference Material Baking Chocolate
490 2384, Cocoa, Cocoa Beans, and Cocoa Butter. *Journal of Chromatographic*
491 *Science*, 46(10), 892–899.
492 Savitzky, A., & Golay, M. (1951). Smoothing and Differentiation of Data by Simplified
493 Least Squares Procedures. *Analytical Chemistry*, *40*(42).

494 Srdjenovic, B., Djordjevic-Milic, V., Grujic, N., Injac, R., & Lepojevic, Z. (2008).
495 Simultaneous HPLC Determination of Caffeine, Theobromine, and Theophylline
496 in Food, Drinks, and Herbal Products. *Journal of Chromatographic Science, 46*(2),
497 144–149.
498 Sunoj, S., Igathinathane, C., & Visvanathan, R. (2016). Nondestructive determination of
499 cocoa bean quality using FT-NIR spectroscopy. *Computers and Electronics in*
500 *Agriculture, 124*, 234–242.
501 Teye, E., Huang, X., Dai, H., & Chen, Q. (2013). Rapid differentiation of Ghana cocoa
502 beans by FT-NIR spectroscopy coupled with multivariate
503 classification. *Spectrochimica Acta Part A: Molecular and Biomolecular*
504 *Spectroscopy*, *114*, 183-189.
505 Toledo-Martín, E. M., García-García, M. C., Font, R., Moreno-Rojas, J. M., Salinas-
506 Navarro, M., Gómez, P., & del Río-Celestino, M. (2018). Quantification of total
507 phenolic and carotenoid content in blackberries *(Rubus fructicosus L.*) using near
508 infrared spectroscopy (NIRS) and multivariate analysis. *Molecules, 23*(12), E3191.
509 Toro-Uribe, S., Montero, L., López-Giraldo, L., Ibáñez, E., & Herrero, M. (2018).
510 Characterization of secondary metabolites from green cocoa beans using focusing-
511 modulated comprehensive two-dimensional liquid chromatography coupled to
512 tandem mass spectrometry. *Analytica Chimica Acta*, *1036*, 204-213.
513 Van Durme, J., Ingels, I., & De Winne, A. (2016). Inline roasting hyphenated with gas
514 chromatography-mass spectrometry as an innovative approach for assessment of
515 cocoa fermentation quality and aroma formation potential. *Food Chemistry, 205*,
516 66–72.
517 Vergara-Barberán, M., Lerma-García, M. J., Herrero-Martínez, J. M., & Simó-Alfonso,
518 E. F. (2015). Cultivar discrimination of Spanish olives by using direct FTIR data

519 combined with linear discriminant analysis. *European Journal of Lipid Science*
520 *and Technology*, *117*(9), 1473–1479.
521 Veselá, A., Barros, A. S., Synytsya, A., Delgadillo, I., Čopíková, J., & Coimbra, M. A.
522 (2007). Infrared spectroscopy and outer product analysis for quantification of fat,
523 nitrogen, and moisture of cocoa powder. *Analytica Chimica Acta, 601*(1), 77–86.
524 Whitacre, E., Oliver, J., Van Den Broek, R., Engelsen, P., Kremers, B., Van Der Horst,
525 B., Stewart, M., Jansen-Beuwink, A. (2003). Predictive Analysis of Cocoa
526 Procyanidins Using Near-Infrared Spectroscopy Techniques. *Food Chemistry and*
527 *Toxicology Predictive, 68*(9), 2618–2622.
528 Wold, S., Antti, H., Lindgren, F., & Öhman, J. (1998). Orthogonal signal correction of
529 near-infrared spectra. *Chemometrics and Intelligent Systems*, *44*, 175–185.
530

531 **Figure captions**

532

533 **Fig 1.** Representative chromatogram of a natural cocoa sample. Peak identification: (1)

534 theobromine, (2) catechin, (3) caffeine and (4) epicatechin peaks.

535

536 **Fig. 2.** Dendogram of the clustering analysis performed with the NIR spectra of the

537 entire sample set (n=86).

538

539 **Fig. 3.** Spectra of the cocoa powders obtained from (a) raw, b) pre-treated with OSC

540 and (c) pre-treated with 2nd derivative S-G + OSC.

541

542 **Fig. 4.** HPLC measured *versus* NIR predicted (a) theobromine (THEO), (b) catechin

543 (CAT), (c) caffeine (CAF) and (d) epicatechin (EPI) contents by PLS in the prediction

544 set.

545

546 **Fig. 5.** B vector profiles of the PLS models constructed to predict (a) theobromine, (b)

547 catechin, (c) caffeine and (d) epicatechin contents.

Table 1

Table 1. Linear ranges, determination coefficient, LOD and LOQ of the determination of methylxanthines and monomeric flavanols by HPLC.

Analyte	t_R (min)	Linear range ($mg\ L^{-1}$)	R^2	LOD ($mg\ L^{-1}$)	LOQ ($mg\ L^{-1}$)
Theobromine	5.1	50-500	0.9998	0.005	0.017
Catechin	9.1	1-50	0.9996	0.1	0.33
Caffeine	12.0	5-250	0.9998	0.05	0.17
Epicatechin	15.8	0.5-100	0.9999	0.1	0.33

t_R: retention time; R^2: determination coefficient; LOD: limit of detection; LOQ: limit of quantification.

Table 2

Table 2. Theobromine, catechin, caffeine and epicatechin content (mean and minimum and maximum values with their standard deviation) found for the different types of cocoa powders employed in this study.

Analyte	Natural (pH 5-6)	Alkalization degree*			
		Light (pH 6-7.2)	**Medium (pH 7.2-7.6)**	**Strong (pH >7.6)**	**Black (pH >7.6, low L*)**
Theobromine (g/100g)	2.15[a] (1.53±0.03-2.36±0.03)	2.13[ab] (1.77±0.02–2.393±0.004)	2.04[ab] (1.68±0.04-2.41±0.18)	2.06[ab] (1.76±0.02–2.33±0.15)	1.95[b] (1.768±0.002-2.07±0.02)
Catechin (mg/100g)	78.21[a] (15.2±0.9-167±2)	60.39[a] (13.3±0.3–131.9±2.3)	27.86[b] (4.9±0.3–65.6±0.9)	23.36[b] (4.53±0.12–46.3±0.9)	< LOD
Caffeine (g/100g)	0.29[a] (0.201±0.003–0.412±0.006)	0.24[b] (0.175±0.002–0.393±0.003)	0.21[bc] (0.157±0.003-0.334±0.002)	0.20[c] (0.172±0.004–0.255±0.002)	0.20[bc] (0.1503±0.0006-0.302±0.004)
Epicatechin (mg/100g)	163.17[a] (11.5±0.3-536.59±0.13)	80.24[b] (11.9±0.3–219±2)	33.27[bc] (8.0±0.6-111.00±0.08)	28.13[c] (10.5±0.4-54.9±0.3)	< LOD

Within rows, values bearing different letters are significantly different ($p < 0.05$)

LOD = limit of detection

Table 3

Table 3. Results of the PLS models constructed for the prediction of theobromine, catechin, caffeine and epicatechin contents in cocoa powders.

Compound	Pre-treatment	#LV	Calibration		Cross-validation		Prediction			
			R^2_C	RMSEC	R^2_{CV}	RMSECV	R^2_P	RMSEP	bias	RPD
Theobromine (g/100g)	Non-pretreated	6	0.628	0.115	0.407	0.149	0.602	0.113	-0.017	1.69
	2nd Der. S-G	7	0.989	0.026	0.793	0.088	0.871	0.064	-0.012	2.98
	OSC	3	0.841	0.075	0.766	0.093	0.402	0.138	-0.022	1.38
	2nd Der. S-G + OSC	*3*	*0.990*	*0.019*	*0.955*	*0.041*	*0.882*	*0.061*	*-0.027*	*3.12*
Catechin (mg/100g)	Non-pretreated	5	0.878	5.363	0.260	13.981	0.612	11.18	1.227	1.41
	2nd Der. S-G	2	0.547	10.33	0.436	12.215	0.192	16.135	6.863	0.98
	OSC	2	0.838	6.170	0.726	8.502	0.505	12.632	3.593	1.25
	2nd Der. S-G + OSC	*2*	*0.988*	*2.204*	*0.947*	*4.929*	*0.863*	*6.632*	*1.942*	*3.14*
Caffeine (g/100g)	Non-pretreated	7	0.793	0.029	0.675	0.037	0.539	0.040	0.006	1.60
	2nd Der. S-G	4	0.807	0.028	0.687	0.036	N.A.	0.071	0.030	0.91
	OSC	*2*	*0.949*	*0.015*	*0.931*	*0.017*	*0.882*	*0.020*	*0.006*	*3.20*
	2nd Der. S-G + OSC	3	0.953	0.014	0.757	0.032	0.660	0.035	0.007	1.87
Epicatechin (mg/100g)	*Non-pretreated*	*6*	*0.987*	*7.294*	*0.972*	*11.250*	*0.818*	*15.870*	*3.056*	*4.12*
	2nd Der. S-G	5	0.989	6.709	0.850	25.895	0.406	28.64	22.80	1.33
	OSC	2	0.961	12.52	0.9458	15.576	0.465	27.195	6.555	2.40
	2nd Der. S-G + OSC	1	0.898	20.39	0.8687	24.248	0.336	30.284	11.85	2.15

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; RPD = ratio of prediction deviation; N.A. = non available.

Figure 1

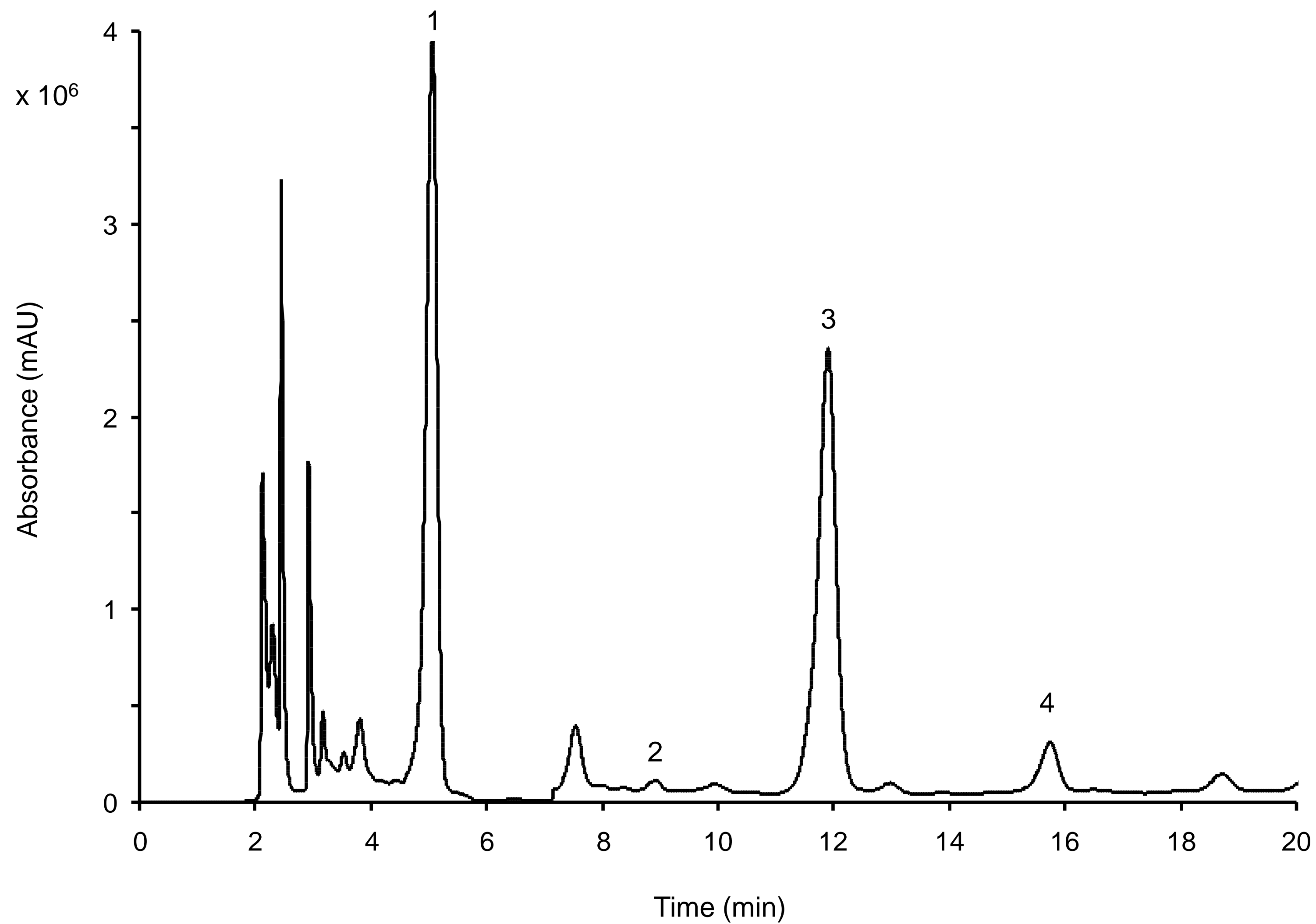

Fig 1. M.A. Quelal-Vásconez et al.

Figure 2

Fig 2. M.A. Quelal-Vásconez et al.

Figure 3

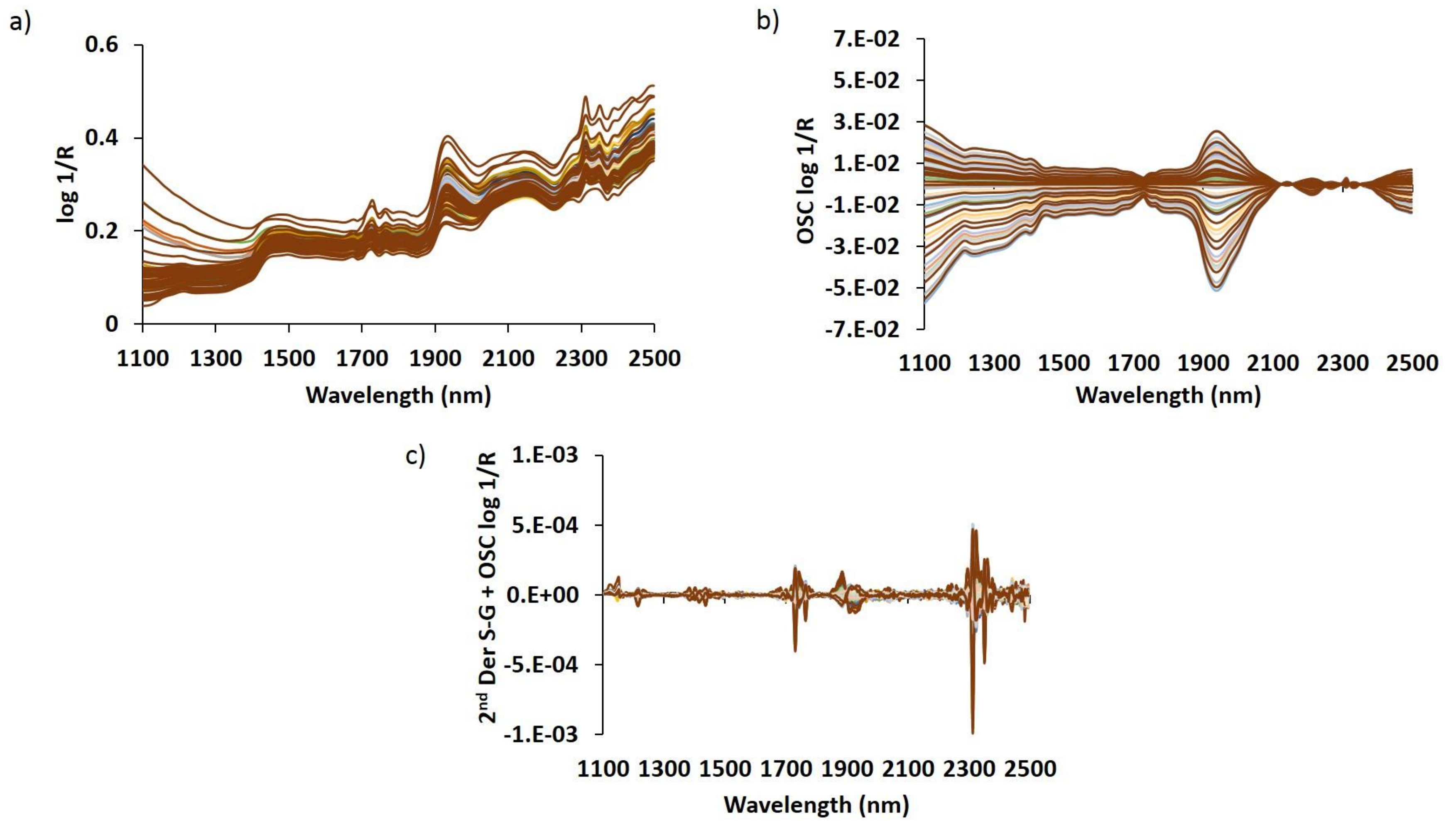

Fig 3. M.A. Quelal-Vásconez et al.

Figure 4

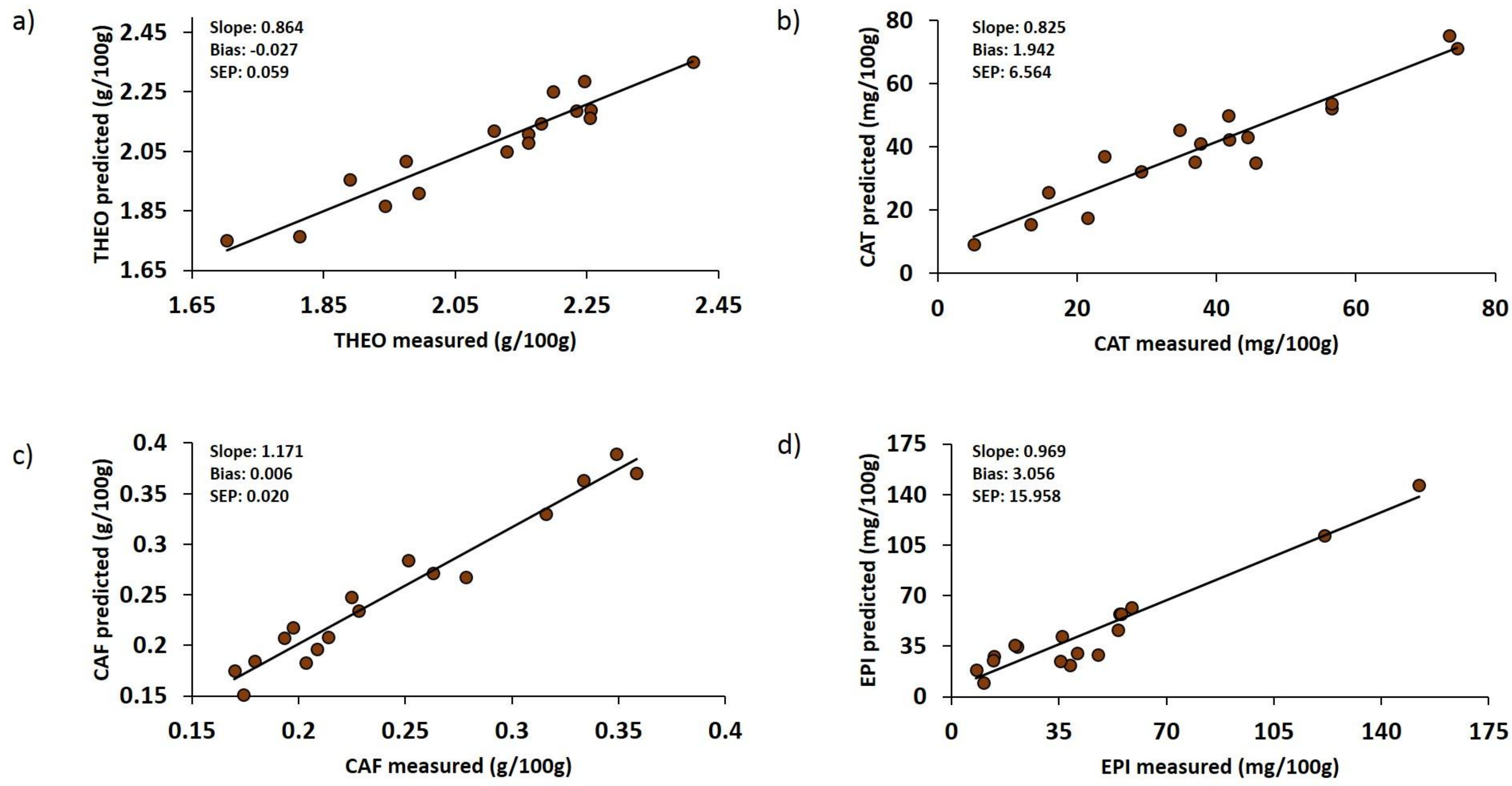

Fig 4. M.A. Quelal-Vásconez et al.

Figure 5

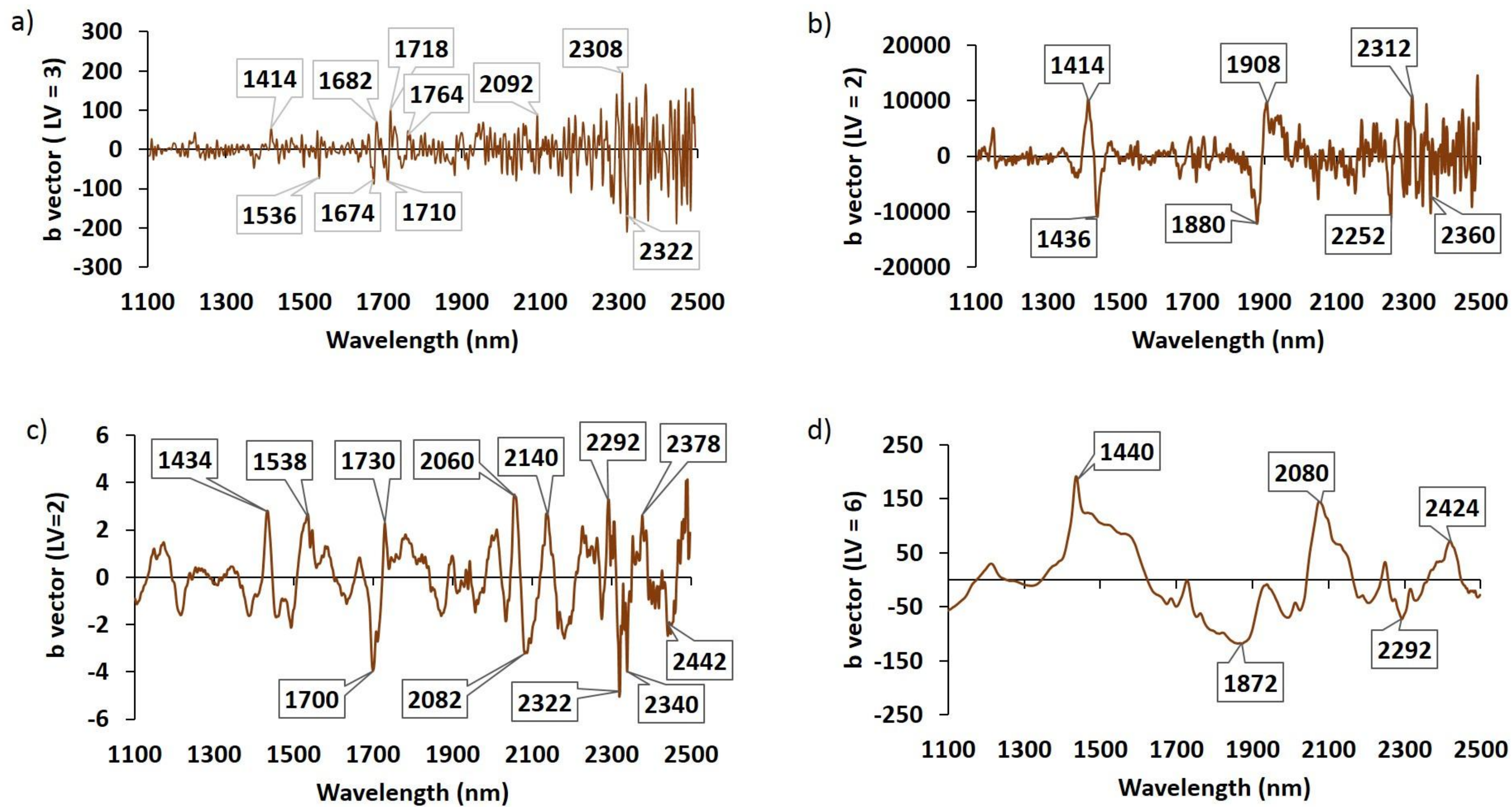

Fig 5. M.A. Quelal-Vásconez et al.

**Conflict of Interest Form

5 June 2019

Dear Editor,

We declare that we do not have any conflict of interest.

Yours faithfully,

Dr. María Jesús Lerma-García
Departamento de Tecnología de Alimentos
Universitat Politècnica de València
Camino de Vera, s/n, 46022 Valencia, Spain
E-mail address: malerga1@tal.upv.es