

1 Determination of chemical properties in ‘calçot’ (*Allium cepa* L.) by near
2 infrared spectroscopy and multivariate calibration

3

4 ^(a)**Silvia Sans**, ^(b)**Joan Ferré**, ^(b)**Ricard Boqué**, ^{(a)(c)}**José Sabaté**, ^{(a)(c)}**Joan Casals**, ^{(a)(c)} **Joan Simó**

5 ^(a) Departament d’Enginyeria Agroalimentària i Biotecnologia, Universitat Politècnica de Catalunya,
6 Campus Baix Llobregat, Esteve Terrades 8, 08860 Castelldefels, Spain

7 ^(b) Department of Analytical Chemistry and Organic Chemistry, Universitat Rovira i Virgili, 43007
8 Tarragona, Spain

9 ^(c) Fundació Miquel Agustí, Campus Baix Llobregat, Esteve Terrades 8, 08860 Castelldefels, Spain

10 *Corresponding author. E-mail address: silvia.sans.molins@upc.edu (S. Sans).

11 Declarations of interest: none

12

13

14

15

16

17

18

19

20

21

22 **ABSTRACT**

23 ‘Calçots’, the immature floral stems of second-year onion resprouts, are an economically important
24 traditional crop in Catalonia (Spain). Classical approaches to evaluating the chemical properties of
25 ‘calçots’ are time consuming and expensive; near-infrared spectroscopy (NIRS) may be faster and
26 cheaper. We used NIRS to develop partial least square (PLS) models to predict dry matter, soluble
27 solid content, titratable acidity, and ash content in cooked ‘calçots’. To guarantee the robustness of
28 the models, calibration samples were grown and analyzed in a first season (2014-15) and validation
29 samples in a second season (2015-16). NIRS on puree spectra estimated dry matter and soluble solid
30 content with excellent accuracy ($R^2_{\text{pred}}=0.953, 0.985$ and $\text{RPD}=4.571, 8.068$, respectively). However,
31 good estimation of titratable acidity and ash content required using ground dried puree spectra
32 ($R^2_{\text{pred}}=0.852, 0.820$ and $\text{RPD}=2.590, 1.987$, respectively). NIRS can be a helpful tool for ‘calçots’
33 breeding and quality control.

34 **KEYWORDS** ‘calçot’, multivariate calibration, PLS, NIR, chemometrics, quality

35 **1. INTRODUCTION**

36 ‘Calçots’, the immature floral stems of second-year onion (*Allium cepa* L.) resprouts, are
37 economically important for agriculture and restaurants in Catalonia (Spain), where popular
38 celebrations called ‘calçotades’ bring people to eat cooked ‘calçots’. As ‘calçots’ grow they are
39 covered with soil so that the edible lower part remains white. The European Union has designated the
40 Protected Geographical Indication ‘Calçot de Valls’ (EC No 905/2002) for ‘calçots’ from the ‘Ceba
41 Blanca Tardana de Lleida’ landrace of onions cultivated in four counties.

42 Despite the economic importance of ‘calçots’, research on this crop is relatively recent. To date, only
43 their agronomic performance has been studied, and some tools for ‘calçots’ yield breeding have been
44 developed (Simó, Valero, Plans, Romero del Castillo, & Casañas, 2013). As a result, two new
45 varieties with significantly increased yields have been obtained (Simó, Romero del Castillo, Almirall,
46 & Casañas, 2012a); in parallel, a sensory ideotype has been elaborated, determining that the ideal

47 'calçot' has a high level of sweetness, low fiber perception, and no off flavors (Simó, Romero del
48 Castillo, & Casañas, 2012b). Little is known about the chemical components related with the sensory
49 and nutritional attributes of 'calçots', although some studies have focused on total phenols and
50 antioxidant activity (Santas, Carbó, Gordon, & Almajano, 2008; Zudaire et al., 2017).

51 Breeding programs need to evaluate many accessions or individuals. On the other hand, breeding for
52 organoleptic traits requires sensory analysis with trained panelists who can only assess a limited
53 number of samples per tasting season, thus making it impracticable to deal with a large number of
54 samples (Plans, Simó, Casañas, Romero del Castillo, Rodriguez-Saona, & Sabaté, 2014). Therefore,
55 establishing relationships between chemical composition and sensory traits would facilitate sensory
56 analysis. Such relationships have been established in other crops such as rice, peaches, onions,
57 melons, and apples (Bett-Garber, Lea, McClung, & Chen, 2013; Colaric, Veberic, Stampar, &
58 Hudina, 2005; Crowther et al., 2005; Escribano, Sanchez, & Lazaro, 2010; Harker et al., 2002; Wall
59 & Corgan, 1992). For example, soluble solid content has been widely used to indicate sweetness of
60 fresh and processed horticultural products (Magwaza & Opara, 2015); titratable acidity correlates
61 with acid taste perception (Flores-Rojas, Sánchez, Pérez-Marin, Guerrero, & Garrido-Varo, 2009;
62 García-Martínez et al., 2012; Ncama, Opara, Tesfay, Fawole, & Magwaza, 2017); and ash is a rough
63 measure of the amount of mineral cations, which interact with carboxylic acids and facilitate the
64 cross-linking of uronic acid polymers that can increase the perception of fiber (Plans, Simó, Casañas,
65 & Sabaté, 2012). Nevertheless, standard chemical analyses are expensive, laborious, time-consuming,
66 and sometimes require considerably large samples. Thus, breeding programs involving sensory or
67 chemical measurements face several impediments.

68 Near-infrared spectroscopy (NIRS) is a well-established technique for determining the chemical
69 components of foods (Blanco & Villaroya, 2002; Williams, 2001). Extensive reviews about its use in
70 meat, dairy products, fruit, and vegetables have been published (Butz, Hofmann, & Tauscher, 2005;
71 Karoui & De Baerdemaeker, 2007; Prieto, Roehe, Lavín, Batten, & Andrés, 2009; Nicolai et al.,
72 2007). Advantages of NIRS include speed and ease of analysis, low cost per test, and the possibility

73 of simultaneously estimating several properties from a single spectrum. Moreover, NIRS does not
74 produce chemical waste. All these advantages make it suitable for breeding, research, and/or quality
75 evaluation when many samples need to be analyzed.

76 Although NIRS or other rapid techniques have not been used to predict chemical attributes of
77 ‘calçots’, they have been employed in related crops such as onions and garlic. The determination of
78 dry matter content of onions (Birth, Dull, Renfro, & Kays, 1985) was an early application of NIRS
79 in horticulture. NIRS has also been used to determine solid soluble content (Wang, Li, & Li, 2013a;
80 Wang, Li, & Wang, 2013b). Jantra, Slaughter, Liang, and Pathaveerat (2017) recently used NIRS to
81 predicted dry matter and solid soluble content in intact dehydrator onions and garlic cloves. In all
82 these studies, spectra were recorded from the intact bulbs. Another rapid technique, Fourier-transform
83 infrared spectroscopy, has been used to determine some nutritional characteristics in onions: total
84 phenolic content, total antioxidant capacity, and quercetin content (Lu et al., 2011; Lu, Ross, Powers,
85 & Rasco, 2011).

86 Foods undergo chemical changes during cooking, so the chemical contents of cooked foods should
87 correlate better both with nutritive value and with organoleptic attributes than those of the raw
88 samples. Although not often used on cooked vegetables, NIRS has been used to determine chemical
89 or sensory parameters in cooked potatoes (López, Arazuri, García, Mangado, & Jarén, 2013), rice
90 (Cho, Choi, & Rhee, 1998) and carrots (Moscetti et al., 2017).

91 This study aimed to investigate the ability of NIRS to predict chemical parameters in cooked ‘calçots’
92 that can affect their sensory and nutritive properties. We developed models to predict dry matter,
93 soluble solid content, titratable acidity, and ash content. To guarantee the robustness of the models,
94 calibration samples were grown and analyzed in a first season and validation samples in a second
95 season.

96 **2. MATERIAL AND METHODS**

97 **2.1 Samples**

98 To develop and validate NIRS models, 240 ‘calçot’ samples were screened to encompass the
99 maximum chemical variation in this crop. Four cultivars were used to represent interpopulation
100 genetic variability in ‘calçots’. Two improved cultivars: Roquerola (R) and Montferri (M) (Simó et
101 al., 2012a), a traditional population (T) and a new in process variety (X). To induce further phenotypic
102 differences among samples, the four cultivars were grown in six locations with different pedo-
103 climatic conditions in Catalonia (north-east Spain) that represent standard ‘calçot’ production areas:
104 La Masó (41°13'41.0"N 1°13'33.6"E, 115 masl), Valls (41°15'11.8"N 1°13'44.9"E, 230 masl), La
105 Juncosa del Montmell (41°18'59.4"N 1°27'07.5"E, 430 masl), Altafulla (41°08'46.7"N 1°22'58.7"E,
106 52 masl), Viladecans (41°17'19.3"N 2°02'42.5"E, 18masl) and Rubí (41°30'07.7"N 2°01'03.6"E, 123
107 masl). Field experiments were conducted using the traditional cultivation techniques, in two
108 consecutive seasons (2014-15 and 2015-16) with two planting dates and two or three harvesting dates
109 (early, intermediate, late harvests), depending on the progress of the fields. In total, 120 samples were
110 harvested in each season.

111 **2.2 Sample preparation**

112 After each harvest samples were prepared as described by Simó et al. (2012b). We collected a set of
113 50 commercial ‘calçots’ (commercial size: ‘calçot’ with a compact white edible base measuring 1.7–
114 2.5 cm in diameter, measured at 5 cm from the root, and 15–25 cm in length) from each sample
115 (combination of variety, year, location, planting date, and harvest date). Leaves were cut 4 cm above
116 the ligule, and roots were removed. Then, the ‘calçots’ were rinsed with tap water to remove adhered
117 soil and roasted at 270 °C for 18 minutes in a convection oven (SALVA Kwik-co). After cooking, the
118 two most external leaves were removed and the edible lower white part of each ‘calçot’ was cut. All
119 ‘calçots’ in each sample were triturated with a mixer (Taurus BAPI 850). Pureed samples were frozen
120 with liquid nitrogen and stored at -20 °C until their analysis.

121 **2.3 Chemical analysis**

122 Soluble solid content was directly determined in the puree with a hand refractometer (Erma, Japan)
123 and is expressed as °Brix. To analyze titratable acidity, 10 g of puree was mixed with 50 mL of
124 distilled water and titrated with 0.1 M sodium hydroxide (NaOH) to pH 8.1; titratable acidity is
125 expressed as g/100 g of malic acid.

126 For dry matter determination, 30 g of puree was dried to a constant weight for 72 h at 60 °C; dry
127 matter is expressed as g/100 g of fresh matter. To enable ash content to be determined according to
128 the AOAC method 923.03 (AOAC, 2005), dried samples were ground to an average particle size <
129 0.4 mm to obtain the ground dried puree; then 1 g of ground dried puree was burned in a muffle at
130 450 °C for 4 hours, cooled to room temperature in a desiccator, and then weighed. Ash is expressed
131 as g/100 g of dry matter. All chemical analyses were carried out in triplicate.

132 **2.4 Spectra measurement**

133 A spectrophotometer (Foss NIRSystems 5000, Silver Spring, MD, USA) equipped with a rapid
134 content analyzer module and Vision software version 2.51 were used to register NIRS spectra from
135 puree samples and from the ground dried puree. Spectra were recorded every 2 nm between 1100 nm
136 to 2500 nm and averaged from 32 scans. Puree spectra were registered in reflectance (R) mode and
137 expressed as $\log(1/R)$, and ground dried puree spectra were registered in transfectance mode and
138 expressed as absorbance. Three spectra were registered for each sample, and the mean spectrum was
139 used for computations.

140 **2.5 Chemometric analysis**

141 Since the purpose of this study was to test the use of multivariate models to predict chemical
142 parameters for upcoming seasons, the 120 samples collected during 2014-15 were used as a
143 calibration set, and the 120 samples collected during season 2015-16 were used as a validation set.
144 Spectra from puree and ground dried puree were treated independently, using the reference data
145 expressed over fresh matter or dry matter, respectively. Spectral data were explored by principal
146 components analysis (PCA) to detect clustering of spectra. Quantification was carried out using

147 partial least squares (PLS) regression with the NIPALS algorithm as implemented in PLS_Toolbox
148 v.8.21 software (Eigenvector Research Inc., Wenatchee, USA).

149 To obtain the best PLS models, different spectral pretreatments were tested to reduce unwanted
150 variation due to sources not related to the properties of interest. The preprocessing methods evaluated
151 were standard normal variate (SNV) transformation, multiplicative scatter correction (MSC),
152 Savitzky-Golay (SG) first- and second-order derivatives with second order polynomial approximation
153 and 7-41 point window size, and combinations of these. In all cases, the pretreated spectra and the
154 property values were mean-centered before being submitted to the regression algorithm.

155 PLS regressions for calibration were evaluated using 10-fold cross-validation. The optimal
156 combination of data preprocessing and number of latent variables was selected as the one providing
157 a model with a good enough compromise of a low root mean square error of calibration (RMSEC),
158 low root mean square error of cross-validation (RMSECV), high coefficient of determination (R^2) for
159 calibration and cross-validation, and low bias. An upper limit of optimal PLS terms was set at 1 PLS
160 factor per 10 samples of calibration plus 2 (Ruiz-Jiménez, Priego-Capote, & Luque de Castro, 2006).
161 No outliers were detected.

162 The established PLS models were finally tested with the validation set of samples from the 2015-16
163 season, which had not been used for the calibration. The root mean square error of prediction
164 (RMSEP) evaluated with these samples was used as an estimation of the future performance of the
165 calibration model.

166 RMSEC, RMSECV, and RMSEP were calculated as follows:

167
$$\text{RMSEC} = \sqrt{\frac{\sum_{i=1}^I (y_i - \hat{y}_i)^2}{I - A - 1}}$$

168
$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^{I_t} (y_i - \hat{y}_i)^2}{I_t}}$$

169

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{I_t} (y_i - \hat{y}_i)^2}{I_t}}$$

170 where y_i is the reference value of sample i , \hat{y}_i is the model prediction for sample i , I is the number of
171 calibration samples, I_l is the number of samples left-out in the cross-validation process and I_t is the
172 number of test samples for the test set. A is the number of latent variables used in the PLS model.

173 Ratio of the performance to deviation (RPD) values of the prediction models were calculated as well.
174 RPD is a dimensionless index widely used to evaluate NIRS models in agricultural products (Bellon-
175 Maurel, Fernandez-Ahumada, Palagos, Roger, & McBratney, 2010). RPD was calculated as follows:

176

$$\text{RPD} = \frac{\text{SD}_x}{\text{RMSEP}}$$

177 in which SD_x is the standard deviation of the validation reference data.

178 Data were analyzed with PLS_Toolbox v.8.21 software (Eigenvector Research Inc., Wenatchee,
179 USA) and in-house routines running under MATLAB R2017a (The MathWorks™ Inc., Natick, MA,
180 USA).

181 **3. RESULTS AND DISCUSSION**

182 **3.1 Reference data**

183 Table 1 reports the means, standard deviations, and ranges of the chemical properties determined in
184 the calibration set (season 2014-15) and validation set (season 2015-16). The laboratory repeatability
185 precision of the reference methods for dry matter, soluble solid content, titratable acidity, and ash
186 content was lower than $\pm 5\%$.

187 All parameters varied over a wide enough range to enable suitable calibrations, though in dry matter,
188 soluble solid content, and ash content, the validation ranges are slightly beyond the calibration ranges.
189 The magnitude of the variation is due to environmental and genetic effects.

190 There are no previous reports related to these chemical parameters in ‘calçots’, but the orders of
 191 magnitude of the measured properties were similar to those reported in raw onions (Barzegar, Jabbari,
 192 Rajabi, & Hassandokht, 2008; Petropoulos, Fernandes, Barros, Ferreira, & Ntatsi, 2015) and roasted
 193 onions (Spanish food composition database, 2018).

194 **Table 1.** Chemical composition of the samples in the calibration (season 2014-15) and validation
 195 (season 2015-16) sets.

Parameter	Units	Calibration set			Validation set		
		Range	Mean	SD	Range	Mean	SD
DM	g/100 g FM	12.73-20.16	16.25	1.74	11.79-19.70	15.62	1.87
SSC	°Brix	9.80-17.10	13.10	1.63	8.75-15.95	12.40	1.74
TA	g/100 g FM	0.08-0.22	0.12	0.03	0.08-0.19	0.13	0.03
	g/100 g DM	0.43-1.50	0.74	0.20	0.48-1.36	0.84	0.20
Ash	g/100 g FM	0.64-1.09	0.81	0.10	0.63-0.97	0.80	0.07
	g/100 g DM	3.59-7.72	5.09	0.79	4.01-6.85	5.20	0.66

196 SD: standard deviation; DM: dry matter; SSC: soluble solid content; TA: titratable acidity; FM: fresh matter

197 3.2 Characteristics of NIRS spectra

198 Figure 1 shows the recorded and preprocessed NIRS spectra of both puree and ground dried puree
 199 for all 240 samples studied. Raw spectra had a similar profile but different intensity. Mathematical
 200 preprocessing decreased the baseline differences and highlighted the relevant bands. The two seasons
 201 can be distinguished in the raw spectra (Figures 1A and 1C); however, the differences disappear after
 202 preprocessing (Figures 1B and 1D). PCA of NIRS spectra found no clustering or separation in the
 203 dataset.

204 Due to the complex chemical composition of vegetables, it is difficult to assign specific absorption
 205 bands to specific functional groups, and even less to chemical components. Puree spectra are strongly
 206 influenced by water absorption bands (Figure 1A and 1B), as water is a major chemical constituent
 207 of ‘calçots’. The two characteristic water absorption peaks around 1450 nm (stretch of the O–H bonds,

208 first overtone) and 1940 nm (stretch of the O–H bonds and O–H deformation) are clearly seen. These
209 bands are much less relevant in the dried puree spectra.

210 Although the high water content in the samples could limit the use of NIRS due to the strong
211 absorption bands that predominate in the spectrum, NIRS has been successfully used on a wide
212 variety of vegetables with high moisture content (Butz et al., 2005; López et al., 2013). However,
213 Nicaolaï et al. (2007) point out that when water absorption bands dominate the spectrum of fruit and
214 vegetables, it is not likely that minor constituents can be measured well. The ground dried puree
215 spectra show less intense water absorption peaks (Figure 1C and D), so we expected to obtain better
216 prediction models of the parameters that are present in lower concentrations such as titratable acidity
217 and ash content.

218

219 **Figure 1**

220

221 **3.3 PLS calibration models**

222 Multivariate calibration models relating preprocessed NIRS spectra and the values of the chemical
223 parameters were developed by means of PLS regression, using the entire spectral range. Calibration
224 models for titratable acidity and ashes were calculated using spectra from either puree or ground dried
225 puree. Since dry matter and soluble solid content are related to water content, only the model for
226 puree spectra was calculated. Table 2 presents the statistical descriptors for the NIRS determination,
227 including optimal preprocessing and number of latent variables for all modeled properties.

228 **Table 2.** Statistical descriptors for NIRS determinations.

Parameter	Sample	Pre-processing	LVs	RMSEC	R ² _{cal}	RMSECV	R ² _{CV}	RMSEP	R ² _{pred}	RPD
DM	Puree	SG-1D(25)	6	0.273	0.973	0.378	0.950	0.409	0.953	4.571
SSC	Puree	SNV+SG-1D(11)	5	0.246	0.975	0.294	0.964	0.216	0.985	8.068

TA	Puree	SG-1D(17)	8	0.013	0.791	0.018	0.600	0.023	0.433	1.087
	GD puree	SNV+SG-2D(15)	11	0.048	0.940	0.077	0.858	0.078	0.852	2.590
Ash	Puree	SG-1D(41)	7	0.078	0.432	0.093	0.236	0.054	0.435	1.213
	GD puree	SNV+SG-1D(7)	8	0.282	0.872	0.364	0.793	0.331	0.820	1.987

229 DM: dry matter; SSC: soluble solid content; TA: titratable acidity; GD puree: ground dried puree; LVs: number of latent
230 variables; cal: calibration; CV: cross-validation; pred: prediction; SNV: standard normal variate; SG-1D: Savitzky-Golay
231 first-order derivative; SG-2D: Savitzky-Golay second-order derivative; between parentheses: window size.

232 The performance of the models developed with the spectra of the two different sample preparations
233 varied for each attribute. The predictive ability of the PLS models from puree spectra was high for
234 soluble solid content and dry matter. For soluble solid content, the model found $R^2_{\text{pred}} = 0.985$ and
235 $\text{RMSEP} = 0.216$ °Brix, which is an error lower than the 2% of the mean value of the prediction samples
236 set. For dry matter, $R^2_{\text{pred}} = 0.953$ and $\text{RMSEP} = 0.409$ g/100g, representing 2.6% of the mean value.
237 By contrast, PLS models using puree spectra yielded poor results for titratable acidity and ash content,
238 for which spectra from ground dried puree provided better predictions: $R^2_{\text{pred}} = 0.852$ and RMSEP
239 $= 0.078$ g/100g (9% of mean value) for titratable acidity and $R^2_{\text{pred}} = 0.820$ and $\text{RMSEP} = 0.331$ g/100g
240 (6.3 % of mean value) for ash. It makes sense that titratable acidity and ash content were predicted
241 better from ground dried puree spectra, since acids and ash are minor components in cooked ‘calçots’.
242 Figure 2 plots the results of the prediction sample set for the optimal models. The prediction bias was
243 low in all cases.

244 These results are comparable to those reported in intact onions (Table 3); however, our models for
245 predicting dry matter and soluble solid content perform better than those reported in the literature,
246 since the error of prediction is lower.

247 **Table 3.** Comparison of statistics for predictions of dry matter and soluble solid content by NIRS in
248 onions reported in the literature.

Dry matter			Soluble solid content			Spectral range (nm)	Reference
R^2	Error	RPD	R^2	Error	RPD		

0.99 (cal.)	0.79 (SEP)	-	0.94 (cal.)	3.41 (SEP)	-	700-1000	Birth et al., 1985
-	-	-	0.90 (pred.)	1.84 (RMSEP)	3.00	924-1720	Wang et al., 2013a
0.93 (pred.)	1.61 (SEP)	-	0.93 (pred.)	1.46 (SEP)	-	400-1000	Wang et al., 2013b
0.96 (CV)	1.29 (RMSECV)	4.91	0.97 (CV)	1.23 (RMSECV)	5.72	700-1000	Jantra et al., 2017
0.95 (pred.)	0.41 (RMSEP)	4.57	0.99 (pred.)	0.22 (RMSEP)	8.06	1100-2500	The present study

249 All errors expressed as g/100g FM. cal.: calibration; CV: cross-validation; pred.: prediction; SEP: standard error of
250 prediction.

251 To the best of our knowledge, there are no reports about NIRS prediction of titratable acidity and ash
252 in onions or other related crops, such as garlic or leek. However, NIRS has been used to predict these
253 parameters in other crops: titratable acidity in tomatoes (Flores-Rojas et al., 2009; García-Martínez
254 et al., 2012) and oranges and grapefruits (Ncama et al., 2017) and ash in common beans (Plans et al.,
255 2012) and alfalfa (Halgerson, Sheaffer, Martin, Peterson, & Weston, 2004). The predictive abilities
256 varied widely.

257 According to Nicolai et al. (2007), RPD values obtained for dry matter and soluble solid content
258 models (4.571 and 8.068, respectively) indicate that these parameters can be predicted with excellent
259 accuracy from spectra from puree. In contrast, accurate prediction of titratable acidity (RPD=2.590)
260 and ash (RPD=1.987) requires drying and grinding the samples. Because the amount of titratable
261 acidity and ash in puree is much lower than the amount of dry matter and soluble solid content, strong
262 water absorption bands mask their signals, decreasing the PLS model's ability to predict titratable
263 acidity and ash in puree samples.

264 In summary, these results demonstrate that the PLS-NIRS models can predict properties of interest
265 in 'calçot' samples. Furthermore, since the validation samples were cultivated, harvested, treated,
266 cooked, and analyzed one year later than the calibration set, these models are robust and feasible, and
267 they can be used with confidence for the evaluation of samples in future seasons.

268

269 **Figure 2**

270

271 **4. CONCLUSION**

272 We developed models to predict dry matter, soluble solid content, titratable acidity, and ash content
273 in cooked ‘calçots’ that have proven useful for all the parameters studied. According to the RPD
274 values, models from puree spectra can predict dry matter and soluble solid content with excellent
275 accuracy, but fail to predict titratable acidity and ash content; these parameters, however, can be
276 predicted with good accuracy with models from ground dried puree spectra. The results of this study
277 show that it is feasible to use NIRS to evaluate the chemical properties of cooked ‘calçot’.

278 As validation samples were cultivated and analyzed one year later than the calibration samples, the
279 models are robust and able to predict properties of this crop in coming years. Therefore, NIRS can be
280 a helpful tool for handling large numbers of samples in ‘calçot’ breeding programs or quality
281 evaluation dealing directly or indirectly with chemical composition.

282 **Funding**

283 This research was carried out with the financial support of ACCIÓ (Generalitat of Catalonia, RD14-
284 1-004), ‘Cooperativa de Valls’, ‘Cooperativa de Cambrils’ and PGI ‘Calçot de Valls’. This work has
285 been supported by the Secretaria d’Universitats i Recerca del Departament d’Economia i
286 Coneixement (FI-DGR 2015).

287 **Acknowledgments**

288 The authors gratefully acknowledge dr. Manel Alcalà and his team for letting us use the laboratory
289 equipment.

290 **Conflict of interest**

291 None.

292 **References**

- 293 AOAC (2005). *Official methods of analysis* (18th ed.). Gathersburg MD, USA: Association of
294 Official Analytical Chemists.
- 295 Barzegar, M., Jabbari, A., Rajabi, A., & Hassandokht, M. R. (2008). Chemical composition of
296 different cultivars of onion (*Allium cepa* L.) produced in Iran. *Horticulture, Environment, and*
297 *Biotechnology*, 49 (2), 121–127.
- 298 Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.-M., & McBratney, A. (2010).
299 Critical review of chemometric indicators commonly used for assessing the quality of the
300 prediction of soil attributes by NIR spectroscopy. *TrAC Trends in Analytical Chemistry*, 29(9),
301 1073–1081.
- 302 Bett-Garber, K. L., Lea, J. M., McClung, A. M., & Chen, M.-H. (2013). Correlation of sensory,
303 cooking, physical, and chemical properties of whole grain rice with diverse bran color. *Cereal*
304 *Chemistry*, 90(6), 521–528.
- 305 Birth, G. S., Dull, G. G., Renfroe, W. T., & Kays, S. J. (1985). Nondestructive spectrophotometric
306 determination of dry-matter in onions. *Journal of the American Society for Horticultural*
307 *Science*, 110(2), 297–303.
- 308 Blanco, M., & Villarroya, I. (2002). NIR spectroscopy: a rapid-response analytical tool. *TrAC Trends*
309 *in Analytical Chemistry*, 21(4), 240–250.
- 310 Butz, P., Hofmann, C., & Tauscher, B. (2005). Recent developments in noninvasive techniques for
311 fresh fruit and vegetable internal quality analysis. *Journal of Food Science*, 70(9), R131–R141.
- 312 Cho, S. Y., Choi, S. G., & Rhee, C. (1998). Determination of degree of retrogradation of cooked rice
313 by near infrared reflectance spectroscopy. In A. M. C. Davies, (Ed.), *Journal of Near Infrared*
314 *Spectroscopy* (Vol. 6, pp. A355–A359). England: NIR Publications.

- 315 Colaric, M., Veberic, R., Stampar, F., & Hudina, M. (2005). Evaluation of peach and nectarine fruit
316 quality and correlations between sensory and chemical attributes. *Journal of the Science of Food
317 and Agriculture*, 85(15), 2611–2616.
- 318 Crowther, T., Collin, H. A., Smith, B., Tomsett, A. B., O'Connor, D., & Jones, M. G. (2005).
319 Assessment of the flavour of fresh uncooked onions by taste-panels and analysis of flavour
320 precursors, pyruvate and sugars. *Journal of the Science of Food and Agriculture*, 85(1), 112–
321 120.
- 322 EC No 905/2002 (2002). Commission Regulation (EC) No 905/2002 of 30 May 2002 supplementing
323 the Annex to Regulation (EC) No 2400/96 on the entry of certain names in the 'Register of
324 protected designations of origin and protected geographical indications' OJ L 142/2.
- 325 Escribano, S., Sanchez, F. J., & Lazaro, A. (2010). Establishment of a sensory characterization
326 protocol for melon (*Cucumis melo* L.) and its correlation with physical-chemical attributes:
327 indications for future genetic improvements. *European Food Research and Technology*, 231(4),
328 611–621.
- 329 Flores-Rojas, K., Sanchez, M.-T., Perez-Marin, D., Emilio Guerrero, J., & Garrido-Varo, A. (2009).
330 Quantitative assessment of intact green asparagus quality by near infrared spectroscopy.
331 *Postharvest Biology and Technology*, 52(3), 300–306.
- 332 Garcia-Martinez, S., Galvez-Sola, L. N., Alonso, A., Agullo, E., Rubio, F., Ruiz, J. J., & Moral, R.
333 (2012). Quality assessment of tomato landraces and virus-resistant breeding lines: quick
334 estimation by near infrared reflectance spectroscopy. *Journal of the Science of Food and
335 Agriculture*, 92(6), 1178–1185.
- 336 Halgerson, J. L., Sheaffer, C. C., Martin, N. P., Peterson, P. R., & Weston, S. J. (2004). Near-infrared
337 reflectance spectroscopy prediction of leaf and mineral concentrations in alfalfa. *Agronomy
338 Journal*, 96(2), 344–351.

- 339 Harker, F. R., Marsh, K. B., Young, H., Murray, S. H., Gunson, F. A., & Walker, S. B. (2002).
340 Sensory interpretation of instrumental measurements 2: sweet and acid taste of apple fruit.
341 *Postharvest Biology and Technology*, 24(3), 241–250.
- 342 Jantra, C., Slaughter, D. C., Liang, P.-S., & Pathaveerat, S. (2017). Nondestructive determination of
343 dry matter and soluble solids content in dehydrator onions and garlic using a handheld visible
344 and near infrared instrument. *Postharvest Biology and Technology*, 133, 98–103.
- 345 Karoui, R., & De Baerdemaeker, J. (2007). A review of the analytical methods coupled with
346 chemometric tools for the determination of the quality and identity of dairy products. *Food*
347 *Chemistry*, 102(3), 621–640.
- 348 Lopez, A., Arazuri, S., Garcia, I., Mangado, J., & Jaren, C. (2013). A review of the application of
349 near-infrared spectroscopy for the analysis of potatoes. *Journal of Agricultural and Food*
350 *Chemistry*, 61(23), 5413–5424.
- 351 Lu, X., Ross, C. F., Powers, J. R., & Rasco, B. A. (2011). Determination of quercetins in onion
352 (*Allium cepa*) using infrared spectroscopy. *Journal of Agricultural and Food Chemistry*, 59(12),
353 6376–6382.
- 354 Lu, X., Wang, J., Al-Qadiri, H. M., Ross, C. F., Powers, J. R., Tang, J., & Rasco, B. A. (2011).
355 Determination of total phenolic content and antioxidant capacity of onion (*Allium cepa*) and
356 shallot (*Allium oschaninii*) using infrared spectroscopy. *Food Chemistry*, 129(2), 637–644.
- 357 Magwaza, L. S., & Opara, U. L. (2015). Analytical methods for determination of sugars and
358 sweetness of horticultural products: A review. *Scientia Horticulturae*, 184, 179–192.
- 359 Moscetti, R., Haff, R. P., Ferri, S., Raponi, F., Monarca, D., Liang, P., & Massantini, R. (2017). Real-
360 time monitoring of organic carrot (var. Romance) during hot-air drying using near-infrared
361 spectroscopy. *Food and Bioprocess Technology*, 10(11), 2046–2059.

- 362 Ncama, K., Opara, U. L., Tesfay, S. Z., Fawole, O. A., & Magwaza, L. S. (2017). Application of
363 Vis/NIR spectroscopy for predicting sweetness and flavour parameters of 'Valencia' orange
364 (*Citrus sinensis*) and 'Star Ruby' grapefruit (*Citrus x paradisi Macfad*). *Journal of Food*
365 *Engineering*, *193*, 86–94.
- 366 Nicolai, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., & Lammertyn, J. (2007).
367 Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A
368 review. *Postharvest Biology and Technology*, *46*(2), 99–118.
- 369 Petropoulos, S. A., Fernandes, A., Barros, L., Ferreira, I. C. F. R., & Ntatsi, G. (2015). Morphological,
370 nutritional and chemical description of 'Vatikiotiko', an onion local landrace from Greece.
371 *Food Chemistry*, *182*, 156–163.
- 372 Plans, M., Simó, J., Casañas, F., Romero del Castillo, R., Rodriguez-Saona, L. E., & Sabaté, J. (2014).
373 Estimating sensory properties of common beans (*Phaseolus vulgaris* L.) by near infrared
374 spectroscopy. *Food Research International*, *56*(0), 55–62.
- 375 Plans, M., Simó, J., Casañas, F., & Sabaté, J. (2012). Near-infrared spectroscopy analysis of seed
376 coats of common beans (*Phaseolus vulgaris* L.): A potential tool for breeding and quality
377 evaluation. *Journal of Agricultural and Food Chemistry*, *60*(3), 706–712.
- 378 Prieto, N., Roehe, R., Lavin, P., Batten, G., & Andres, S. (2009). Application of near infrared
379 reflectance spectroscopy to predict meat and meat products quality: A review. *Meat Science*,
380 *83*(2), 175–186.
- 381 Ruiz-Jimenez, J., Priego-Capote, F., & de Castro, M. D. L. (2006). FT-midIR determination of fatty
382 acid profiles, including trans fatty acids, in bakery products after focused microwave-assisted
383 Soxhlet extraction. *Analytical and Bioanalytical Chemistry*, *385*(8), 1532–1537.
- 384 Santas, J., Carbó, R., Gordon, M. H., & Almajano, M. P. (2008). Comparison of the antioxidant
385 activity of two Spanish onion varieties. *Food Chemistry*, *107*(3), 1210–1216.

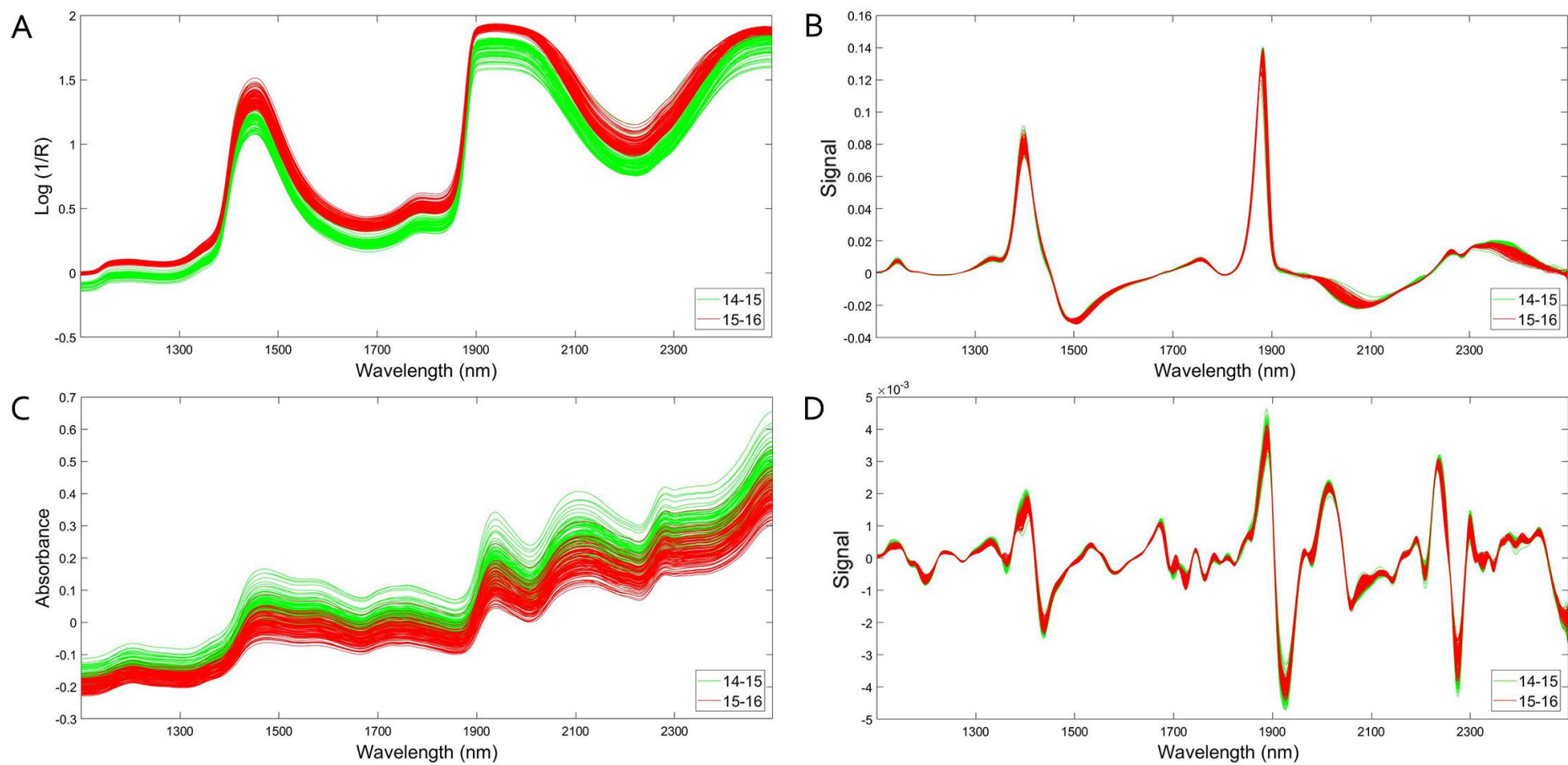
- 386 Simó, J., Romero del Castillo, R., Almirall, A., & Casañas, F. (2012a). “Roquerola” and “Montferri”,
387 first improved onion (*Allium cepa* L.) cultivars for “calçots” production. *Hortscience*, 47(6),
388 801–802.
- 389 Simó, J., Romero del Castillo, R., & Casañas, F. (2012b). Tools for breeding “calçots” (*Allium cepa*
390 L.), an expanding crop. *African Journal of Biotechnology*, 11(50), 11065–11073.
- 391 Simó, J., Valero, J., Plans, M., Romero del Castillo, R., & Casañas, F. (2013). Breeding onions
392 (*Allium cepa* L.) for consumption as “calçots” (second-year resprouts). *Scientia Horticulturae*,
393 152(0), 74–79.
- 394 Spanish food composition database (2018). http://www.bedca.net/bdpub/index_en.php, Accessed 11
395 January 2018.
- 396 Wall, M. M., & Corgan, J. N. (1992). Relationship between pyruvate analysis and flavor perception
397 for onion pungency determination. *Hortscience*, 27(9), 1029–1030.
- 398 Wang, H., Li, C., & Li, M. (2013a). Detection of onion soluble solids content based on the near-
399 infrared reflectance spectra. *Spectroscopy and Spectral Analysis*, 33(9), 2403–2406.
- 400 Wang, H., Li, C., & Wang, M. (2013b). Quantitative determination of onion internal quality using
401 reflectance, interactance, and transmittance modes of hyperspectral imaging. *Transactions of the*
402 *ASABE*, 56(4), 1623–1635.
- 403 Williams, P. C. (2001). Implementation of near-infrared technology. In P. C. Williams, & K. H.
404 Norris (Eds.), *Near-infrared technology in the agricultural and food industries* (2nd ed., pp.145–
405 169). St Paul: American Association of Cereal Chemistry.
- 406 Zudaire, L., Vinas, I., Abadias, M., Simo, J., Echeverria, G., Plaza, L., & Aguilo-Aguayo, I. (2017).
407 Quality and bioaccessibility of total phenols and antioxidant activity of calçots (*Allium cepa* L.)
408 stored under controlled atmosphere conditions (vol 129, pg 118, 2017). *Postharvest Biology and*

409 *Technology, 134, 122–123.*

410

411

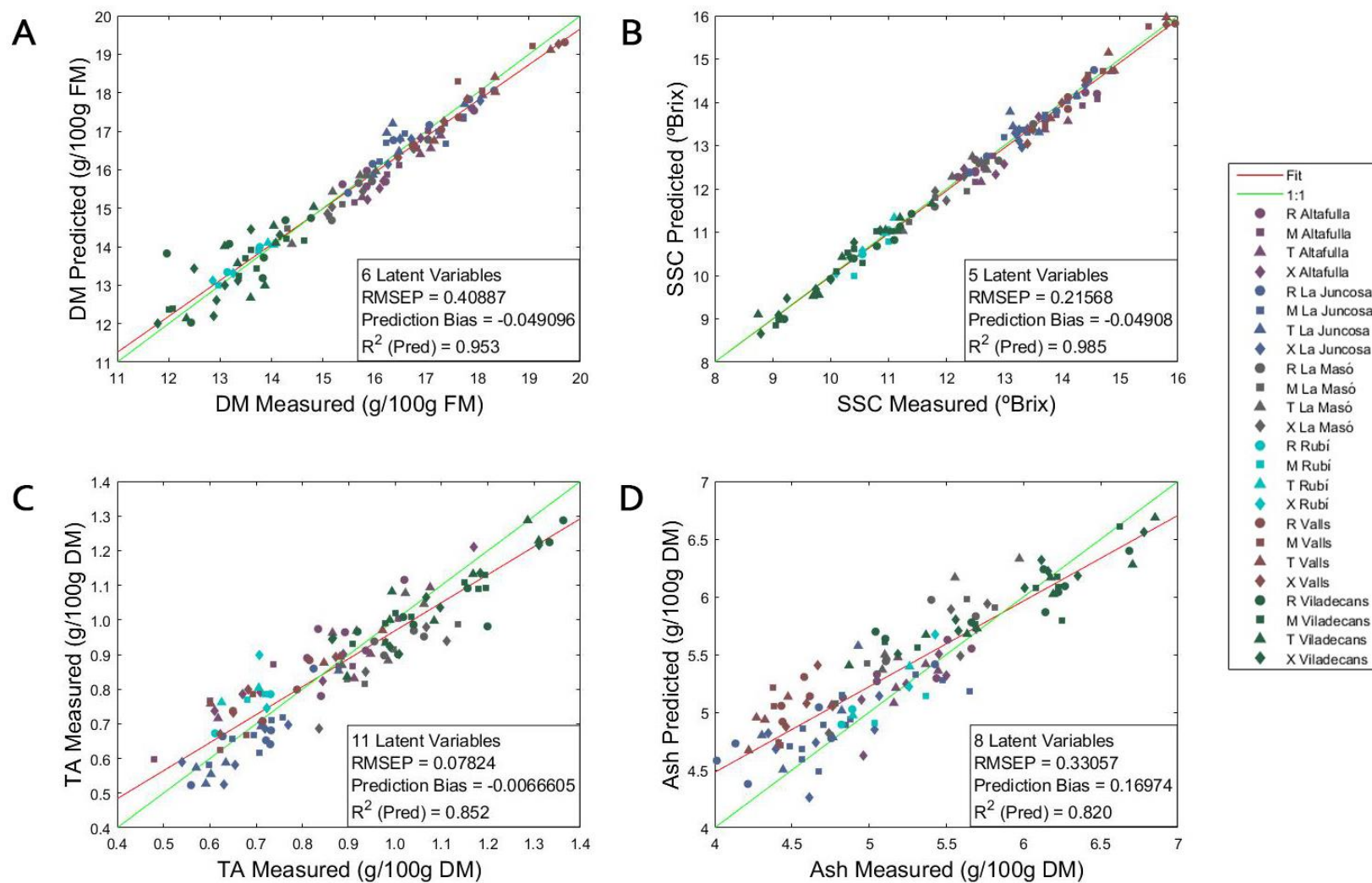
412 **Figure 1.** (A) Raw spectra from puree; (B) pre-processed spectra from puree (SNV + SG 2nd deg. Plym.; window width=11; 1st deriv.) used for soluble
413 solid content determination; (C) raw spectra from ground dried puree; (D) pre-processed spectra from ground dried puree (SNV + SG 2nd deg. Plym.;
414 window width=15; 2nd deriv.) used for titratable acidity determination.



415

416 **Figure 2.** Comparisons between measured laboratory analyses and NIRS predicted values of the prediction set (season 2015-16), for dry matter (DM)

417 (A), soluble solid content (SSC) (B), titratable acidity (TA) (C), and ash content (D).



418