A new application of NIR spectroscopy to describe and predict purees quality from the non-destructive apple measurements

Weijie Lan, Benoit Jaillais, Alexandre Leca, Catherine M.G.C. Renard, Sylvie Bureau

HAL Id: hal-02992917
https://hal.inrae.fr/hal-02992917

Submitted on 21 Jul 2022

1 **A new application of NIR spectroscopy to describe and predict purees quality**

2 **from the non-destructive apple measurements**

3 Weijie Lan[a], Benoit Jaillais[b], Alexandre Leca[a], Catherine M.G.C. Renard[a, c], Sylvie

4 Bureau[a]*

5 [a] UMR408 Sécurité et Qualité des Produits d'Origine Végétale, INRA, Avignon

6 University, F-84000 Avignon, France.

7 [b] INRA-ONIRIS Unité Statistiques, Sensométrie, Chimiométrie (Stat SC), F-44322

8 Nantes, France.

9 [c] CEPIA Division INRA, 44000 Nantes, France.

10

11 **Corresponding authors***

12 Sylvie Bureau (sylvie.bureau@inra.fr)

13 INRA, UMR408 SQPOV « Sécurité et Qualité des Produits d'Origine Végétale »

14 228 route de l'Aérodrome

15 CS 40509

16 F-84914 Avignon cedex 9

17 Tel: +33 432722509

18 Fax: +33 432722492

19 **Others authors**

20 Catherine M.G.C Renard: catherine.renard@inra.fr

21 Benoit Jaillais: benoit.jaillais@inra.fr

22 Alexandre Leca: alexandre.leca@inra.fr

23 Weijie Lan: weijie.lan@inra.fr

24

25   **Highlights**

26   Texture and taste of cooked purees could be predicted from the spectra of raw apples.

27   Apples and purees were well classified by NIRS according to the studied factors.

28   NIRS could predict acceptably quality traits of both apples and purees.

29

30   **Abstract**

31   The potential of NIRS was investigated on both apples and purees to (i) examine

32   factors involving quality variability (variety, agricultural practice, cold storage, puree

33   mechanical refining level) and (ii) establish the link between quality traits before and

34   after processing in order to predict the quality characteristics of purees from spectral

35   information of raw apples. Apples and purees were well-classified at over 82% and 88%

36   according to varieties and storage times respectively. The PLS models showed a good

37   ability to estimate puree characteristics from spectra acquired on corresponding apples

38   such as viscosity ($R^2 > 0.82$), cell wall content ($R^2 > 0.81$) and also dry matter ($R^2 >$

39   0.83), soluble solids content ($R^2 > 0.80$) and titratable acidity ($R^2 > 0.80$). NIR

40   technique should be a useful tool for industry insofar as it can give a reliable

41   assessment of texture and taste of the final products based on the non-destructive

42   fresh materials evaluation.

43

44   Key words: *Malus domestica Borkh*, Near infrared spectroscopy, PLS models,

45   discriminant analyses, apples and purees

46

1. Introduction

Apple is one of the most widely cultivated fruits around the world (totally 68.6 million tons in 2018 (USDA, 2018)), consumed both as fresh fruits and processed products. The fruit could be processed into various products to meet consumers' basic nutritious demand. Among them, apple puree has recently been reported to be a good source of polysaccharides (Le Bourvellec, Bouzerzour, Ginies, Regis, Plé, & Renard, 2011) and antioxidant compounds (Loncaric, Dugalic, Mihaljevic, Jakobek, & Pilizota, 2014; Oszmiański, Wolniak, Wojdyło, & Wawer, 2008). Additionally, apple purees can be used in the food industry as the basic ingredient of many fruit-based products such as jams, preserves or compotes, yogurts and pie fillings (Defernez, Kemsley, & Wilson, 1995).

However, the modifications of the initial physical structure, color and composition, which occur during processing, often make difficult for fruit processors to know and predict the quality characteristics of purees according to the raw apples. Qualities of apple purees depend on complex interactions between process conditions and raw material characteristics. These in turn are determined by the genetic diversity (varieties), pedoclimatic conditions, agricultural practices, maturity stages, and storage periods (Espinosa-Muñoz, Symoeaux, Reard, Biau, & Cuvelier, 2012; Espinosa-Muñoz, To, Symoeaux, Reard, Biau, & Cuvelier, 2011; Keenan, Brunton, Butler, Wouters, & Gormley, 2011; Picouet, Landl, Abadias, Castellari, & Viñas, 2009). Apple puree manufacturers therefore encounter difficulties to maintain the expected and constant quality level of the final apple products. Until now, the research

4

69 studies regarding the quality assessment of apple purees have been mainly focused on

70 the changes in polyphenol contents and total antioxidant activity (Loncaric, Dugalic,

71 Mihaljevic, Jakobek, & Pilizota, 2014; Sukhonthara, Kaewka, & Theerakulkait, 2016),

72 color (Oszmiański, Wolniak, Wojdyło, & Wawer, 2008), ascorbic acid (Picouet, Landl,

73 Abadias, Castellari, & Viñas, 2009), organic acids (Bengoechea *et al.*, 1997), sugars

74 (Keenan, Brunton, Butler, Wouters, & Gormley, 2011), polysaccharides (Le

75 Bourvellec, Bouzerzour, Ginies, Regis, Plé, & Renard, 2011), rheological properties

76 (Espinosa-Muñoz, Renard, Symo□eaux, Biau, & Cuvelier, 2013; Espi□osa-Muñoz,

77 Symoneaux, Renard, Biau, & Cuvelier, 2012) and sensory appreciation

78 (Espinosa-Muñoz, To, Symoneaux, Renard, Biau, & Cuvelier, 2011). However,

79 almost all of these quality parameters have been measured through specific laboratory

80 analyses, such as chromatography, which are time-consuming, expensive and not

81 suitable for fast and numerous characterizations. Consequently, the development of

82 rapid, accurate and reliable methods is required to control the quality of the raw

83 apples and processed purees, and meet the ever-increasing demands for consistent and

84 high quality fruit products.

85 Near infrared spectroscopy (NIRS) has been increasingly used for the safety

86 inspection and quality assessment of agricultural products (Nicolai *et al.*, 2007). It has

87 several advantages such as rapid spectrum acquisition, limited preparation

88 requirements and no chemical waste, but it requires an initial calibration step, which

89 is time consuming. Indeed, on a set of samples, representative of the expected

90 variability, both NIRS spectra and their corresponding reference data are needed to

91   establish predictive models using multivariate statistical and mathematical data

92   analyses. Several parameters can thus be evaluated from a single spectrum, with

93   varying precision. Intensive investigations using NIRS have been reported regarding

94   the measurement of apple internal attributes in the past decades (Nicolai *et al.*, 2007).

95   Satisfactory evaluation results are reported for soluble solids contents (Peirs, Tirry,

96   Verlinden, Darius, & Nicolaï, 2003), dry matter (McGlone, Jordan, Seelye, & Clark,

97   2003), titratable acidity (Liu & Ying, 2005), starch index (Menesatti *et al.*, 2009),

98   chlorophyll content (Zude, Truppel, & Herold, 2002), firmness (Zude *et al.*, 2006),

99   individual sugars (Liu, Ying, Yu, & Fu, 2006) and antioxidant capacity (Schmutzler &

100   Huck, 2016). Further, NIRS spectra are shown to classify apples according to varieties

101   (Luo *et al.*, 2011), geographical origins (Bobelyn *et al.*, 2010) and postharvest storage

102   periods (Camps, Guillermin, Mauget, & Bertrand, 2007; Giova�archelli, Siᴄelli, Beghi,

103   Guidetti, & Casiraghi, 2014).

104   Thus, NIRS can assess a diversity of quality traits in raw apples and processed

105   products, while some of the puree characteristics are directly linked to those of the

106   raw fruit. We therefore can suppose that NIRS spectra of the raw fruits could be used

107   to predict the properties of the processed products, here purees, at least given a

108   constant process operation. However, as far as we know, there is no literature related

109   to the feasibility of using NIRS to evaluate the changes of apple puree properties and

110   to trace back to their corresponding raw apple quality. The challenge here was to

111   assess the possibility of predicting the properties of processed fruit products based on

112   the raw fruit material spectral information, and so, to provide practical and suitable

113   strategies to estimate the quality potential of fruits, to monitor their processing, and to

114   control the quality of fruit products.

115   The specific objectives of our current work were to assess the potential of NIRS to 1)

116   detect different factors such as variety, fruit thinning, storage period and mechanical

117   puree refining on apples and/or their corresponding purees, 2) evaluate the quality

118   traits of interest in both apples and the corresponding purees such as

119   textural/rheological properties, soluble solids content, titratable acidity, dry matter

120   content, insoluble solids content 3) and then establish the links between fruit materials

121   before and after processing.

122   **2. Materials and methods**

123   **2.1 Fruit materials**

124   **2.1.1 Apples**

125   The experiment was conducted on three apple varieties: 'Golden Smoothee', 'Golden

126   Delicious', and 'Granny Smith' during two subsequent harvesting seasons, 2016 and

127   2017, that are summarized in a supplementary figure **(Suppl. Figure 1)**. In 2016, 240

128   'Golden Smoothee' apples were harvested from the experimental orchard of INRA

129   (Drôme, France). In 2017, 480 'Golden Delicious' and 240 'Granny Smith' apples

130   were obtained from the experimental orchard at La Pugère (Bouches du Rhône,

131   France).

132   Two fruit thinning levels were also compared during the ripening of 'Golden

133   Delicious' apples in 2017. For this, trees were thinned 40 days after flowering (on

134   June 2nd), and the treatment named Th+ corresponded to 50 to 100 fruits/tree (a

135   standard commercial fruit load) and Th- to 150-200 fruits/tree (highly loaded trees).

136   The 'Golden Smoothee' in 2016 and 'Granny Smith' in 2017 grew under regular fruit

137   thinning (Th-). After harvesting (Golden Smoothee on September 14th, 2016, 'Golden

138   Delicious' on September 11th, 2017 and 'Granny Smith' on September 25th, 2017),

139   apples were kept in a cold storage chamber at 4°C and at around 90% of humidity

140   during one, three and six months (respectively T1, T3 and T6), except the first group

141   (T0) for which apples were analyzed and processed the day after harvest. These

142   storage durations were chosen in order to increase the fruit variability linked to

143   firmness and biochemical changes, such as demethylation and depolymerization of

144   pectins (Billy, Mehinagic, Royer, Renard, Arvisenet, Prost, et al., 2008). Each set (T0,

145   T1, T3 and T6) was divided into two sub-groups. The first one was dedicated for fresh

146   apple characterization and the second one for processing. For characterization, three

147   replicates of 10 apples, representative of the total apple set, were analyzed after one

148   night temperature equilibration at 23°C.

149   First, nondestructive measurements (NIR, color, ethylene releasing rate, fruit weight),

150   and then texture tests (puncture mean load and puncture linear distance) were

151   performed on each apple. After that, each apple was cored and divided as described by

152   Bureau *et al.* (Bureau, Scibisz, Le Bourvellec, & Renard, 2012) in order to create, for

153   each replicate of ten apples, three batches of 40 pieces representative of each apple.

154   The pieces were immediately put in liquid nitrogen to avoid any oxidization. Finally,

155   one batch was stored at -20 °C and then was subsequently freeze-dried to evaluate the

156   alcohol insoluble solids (AIS) contents. The other two batches were stored at -80°C

157 for biochemical characterization measurement.

158 **2.1.2 Purees**

159 For each raw apple condition, three replicates of apple puree were produced with 4 kg

160 of apples each. After sorting and washing, apples were cored and cut in 8 portions,

161 then processed in a multi-functional processing system (Roboqbo, Qb8-3, Bentivoglio,

162 Italy) following a Hot Break recipe: cooked at 95°C for 5 min at a 1500 rpm grinding

163 speed, then cooled down to 65°C while maintaining the grinding speed. Half of the

164 batch was refined at 0.5 mm using a Robot Coupe C80 automatic refiner (Robot

165 Coupe SNC, Vincennes, France) in order to study two levels of granularity:

166 non-refined (NR) and 0.5 mm refined (Ra). Finally, processed purees were

167 conditioned in hermetically sealed cans, then placed and cooled at 23 °C before

168 measurements, which took place the next day.

169 **2.2 Determination of quality traits**

170 **2.2.1 Color**

171 The skin color (un-blushed and blushed sides) was determined using a CR-400

172 chromameter (Minolta, Osaka, Japan) and expressed in the CIE 1976 L*a*b* color

173 space (illuminant D65, 0° view angle, illumination area diameter 8 mm). The puree

174 color was measured three times through a dedicated glass cuvette using the same

175 method and equipment.

176 **2.2.2 Ethylene production**

177 Each group of ten apples was put in a hermetic jar at 23°C for 1 hour, and ethylene

178 production was analyzed by taking 500 µL of the headspace and injecting it in gas

179   chromatography (Agilent, California, United States) equipped with a porapak Q

180   column and a FID detector and expressed in nmol kg^{-1} h^{-1}.

2.2.3 Fruit texture and puree rheology

182   Fresh apple texture was evaluated by a puncture test using a multipurpose texture

183   analyzer (TAPlus, Lloyd Instruments, Farenham, UK). The puncture tests were

184   operated with a punch probe (diameter 1.2 mm), which could penetrate up to a depth

185   of 17 mm into each peeled section of apple. Firmness was then evaluated as the mean

186   load value calculated by the division of penetration energy by the height of testing.

187   Crunchiness of apple flesh (Gregson & Lee, 2002) was estimated as the linear

188   distance values from the area under the force-distance curve in the range of 10 mm at

189   the load plateau, consisting in summing the lengths between consecutive points.

190   The puree rheological measurements were carried out using a Physica MCR-301

191   controlled stress rheometer (Anton Paar, Graz, Austria) at 22.5 °C. The flow curves

192   were performed after a pre-shearing period of 1 minute at 50 s^{-1} followed by 5

193   minutes at rest. The viscosity was measured at a rate of 1 point every 15 seconds, at a

194   co□trolled shear rate ra□ge of [10; 250] s^{-1} on a logarithmic ramp. The value of the

195   viscosity at 100 s^{-1} η_{100} was kept as an indicator of the puree viscosity. As often used

196   to model fruit purees (Colin-Henrion, Cuvelier, & Renard, 2007), the complete flow

197   curves were fitted with a power-law viscosity model as described by Eq. 1.

198
$$\eta = K \ \dot{\gamma}^{n-1} \qquad\qquad (Eq1)$$

199   where η is the apparent viscosity (Pa.s), $\dot{\gamma}$ the shear rate (s^{-1}), K the consistency

200   parameter, and *n-1* the flow parameter.

2.2.4 Biochemical characterization of apples and purees

Soluble solids content (SSC) was determined with a digital refractometer (PR-101 ATAGO, Norfolk, VA, USA) and expressed in °Brix at 20°C. Titratable acidity (TA) was determined by titration up to pH 8.1 with 0.1 mol/L NaOH and expressed in mmol H^+/kg of fresh weight (FW) using an autotitrator (Methrom, Herisau, Switzerland). Contents of sugars (glucose, fructose, and sucrose) and malic acid were quantified using an enzymatic method with kits for food analysis (Sigma-Aldrich, Deisenhofen, Germany) and expressed in g kg^{-1} FW. These measurements were performed with a SAFAS flx-Xenius XM spectrofluorimeter (SAFAS, Monaco). The dry matter content (DMC) was estimated with the difference between the weight values of fresh samples and of freeze-dried samples upon reaching constant weight (freeze-drier, 5 days). Content of alcohol insoluble solid (AIS) was evaluated using the method proposed by Renard (Renard, 2005) and expressed as the ratio on both fresh weight (FW) and dry matter weight (DW). Three biological replicates were obtained for each biochemical trait and each sample.

2.3 FT-NIR spectrum acquisition

The spectral data of apples and purees were both acquired with a multi-purpose analyzer spectrometer (Bruker Optics®, Wissembourg, France), which provides diffuse reflectance measurements with a spectral resolution of 2 nm from 800 to 2500 nm. For each spectrum, 32 scans were recorded and averaged. The spectral acquisition and instrument adjustments were controlled by OPUS software Version 5.0 (Bruker Optics®). The apples were placed on an automated 30-positions sample wheel, each

223   position corresponding to a measured area of 18 mm diameter. For each intact apple,

224   two spectra were collected (on the blushed and un-blushed sides) though the 18 mm

225   diameter areas at 23°C. Puree were transferred into 10 mL glass vials (5 cm height x

226   18 mm diameter) which were placed on the automated sample wheel of the

227   spectrophotometer. Each puree sample was measured three times on different aliquots.

228   A reference background measurement was automatically activated before each data

229   set acquisition using an internal Spectralon reference.

230   **2.4 Statistical analyses and chemometrics**

231   After ensuring the normal distribution of dataset, the results were presented as mean

232   values and the data dispersion within our experimental dataset expressed as standard

233   deviation values (SD). Analysis of variance (ANOVA) was carried out to determine

234   the significant differences due to the tested factors on both apples and purees using

235   XLSTAT (version 2018.5.52037, Addionsoft SARL, Paris, France) data analysis

236   toolbox. The pairwise comparison between means was performed using Tukey's test

237   at the 95% level of certainty ($p < 0.05(*)$, 0.01 (**) and 0.001 (***)). For apples, a

238   one-way ANOVA was applied to access the effect of storage period on 'Golden

239   Smoothee' in 2016 and 'Granny Smith' i□2017; a two-way ANOVA concerned the

240   effects of storage period and fruit thinning on 'Golden Delicious' in 2017. For purees,

241   a two-way ANOVA accessed the effects of storage periods and refining treatments on

242   'Golden Smoothee' and Granny Smith, and a three- way ANOVA for 'Golden

243   Delicious' in terms of fruit thinning, storage periods and puree refining levels.

244   Pearson's determination coefficients (R^2) were calculated in order to study the

245   significance of the relationship between apples and purees and then output as an heat

246   map using R software (version 3.5.2) (R Core Team, 2018) and the additional package

247   named 'ComplexHeatmap' (Gu, Eils, & Schlesner, 2016).

248   Spectral pre-processing and multivariate data analysis were performed with Matlab

249   7.5 (Mathworks Inc. Natick, MA) software using the SAISIR package (Bertrand &

250   Cordella, 2008). All the NIR data were pre-processed with standard normal variate

251   (SNV) and a derivative transform calculation (Savitzky–Golay method, gap size = 11,

252   21, 31, 41) of first or second order. Each of the preprocessing methods was tested in

253   the discrimination models. As SNV pre-processing had the best performances to

254   correct multiplicative interferences and variations in baseline shift, the results shown

255   are those obtained with the SNV pretreatment. Principal Component Analysis (PCA)

256   and Factor Discriminant Analysis (FDA) were carried out on spectral data to evaluate

257   the possibility to discriminate samples according to the tested factors (cultivars,

258   thinning and storage). The specificity and sensitivity values of FDA discriminations,

259   which help for a better evaluation of the rate of sample differentiation, were

260   calculated by the already reported method of Nargis (Nargis *et al.*, 2019). The Partial

261   least-square (PLS) regression method was used to develop predictive models of the

262   quality traits of interest in apples and purees. The whole spectral dataset included 840

263   spectra of apples and 240 spectra of purees. The dataset was randomly split, two third

264   of dataset (560 spectra of apples and 160 spectra of purees) were used for calibration

265   and one third of dataset (280 spectra of apples and 80 spectra of purees) for validation.

266   The procedure was repeated 10 times in order to obtain the suitable dimensions of the

267 PLS models. The latter performance was described by the root mean square error of

268 calibration (RMSEC), the root mean square error of validation (RMSEV), the number

269 of latent variables (LVs), the determination coefficient (R^2) between the predicted and

270 measured parameters and the RPD (Residual Predictive Deviation) value as described

271 by Nicolai (Nicolai *et al.*, 2007).

272 **3. Results and discussion**

273 **3.1 Apple and puree characteristics measured by classical methods**

274 **3.1.1 Fresh apples**

275 In this experiment, three cultivars, two agricultural conditions and a cold storage for 6

276 months provided an interesting apple fruit variability (**Figure 1a and 1b**). Clear

277 discriminations were shown between the different storage periods along the first

278 principal component, and apple varieties and thinning levels along the second

279 principal component (**Figure 1a**). 'Granny Smith' was clearly differentiated from the

280 two 'Golden' cultivars. Remarkably, the two non-thinned 'Golden' samples were

281 close to each other (blue in 2016 and red in 2017), in spite of different growing

282 seasons and locations, while the thinned samples were clearly differentiated. The most

283 discriminant quality traits were: mean load, linear distance, AIS content (FW and

284 DW), TA, malic acid content, ethylene production rate and color changes (L*, a* and

285 b*) on the first principal component, and SSC, DMC and sucrose content on the

286 second principal component (**Figure 1b**). The totality of the acquired data is presented

287 in a supplementary table (**Suppl. Table 1**).

288 During cold storage, mean load, linear distance and AIS content (FW and DW)

289   decreased remarkably ($p < 0.001$) in all apples, indicating an intensive reduction of

290   apple firmness, crunchiness and cell wall material contents (Johnston, Hewett, &

291   Hertog, 2002). Good correlations were observed between AIS and mean load (R^2

292   =0.78 in 'Golden Smoothee', R^2=0.75 in 'Granny Smith', R^2 =0.82 in 'Golden

293   Delicious'). The acidity in all apples decreased significantly with storage ($p < 0.001$)

294   at a large range from 103.4 to 26.5 meq/kg FW for TA and 6.5 to 2.3 g/kg FW for

295   malic acid. The ethylene production rate increased and then decreased during storage

296   with significant changes ($p < 0.001$). All color parameters (L*, a* and b*) increased

297   clearly for all apples, linked to a degreening and a yellowing during the long-term

298   storage. The changes of all individual sugar contents were significant in 'Golden

299   Smoothee' ($p < 0.05$) and 'Golden Delicious' ($p < 0.001$), but not in 'Granny Smith'

300   ($p > 0.05$). For 'Golden Delicious', SSC and DMC from thinned trees (Th +) appeared

301   to be significantly ($p < 0.001$) higher than from non-thinned trees (Th-).

302   **3.1.2 Apple purees**

303   The fresh apple variability described above affected the characteristics of the

304   corresponding non-refined (NR) purees cooked using the same recipe (**Figure 1c and**

305   **1d**). The NR purees were discriminated according to the apple variety, fruit thinning

306   and storage periods (**Figure 1c**). The first principal component was positively

307   correlated to TA, content of malic acid, AIS (DW and FW) and rheological parameters

308   (η_{100}, K), and negatively linked with colors (L*, b*) and fructose content. The storage

309   periods could be well-classified with this component. The second principal

310   component was highly related to DMC, sucrose content, SSC and AIS content (FW)

311   allowing the separation of varieties and fruit thinning conditions.

312   In all NR purees, clear decreases ($p < 0.001$) of TA, malic acid and AIS (in DW) were

313   observed during storage, which were highly consistent with their changes in raw

314   apples. At the same time, the rheological properties (η_{100} and K) decreased, with

315   statistically significant differences ($p < 0.01$) in all NR purees, but not in Ra purees

316   (**Suppl. Table 2**). A good correlation was found between AIS expressed in fresh

317   weight (FW) and the values of η_{100} in 'Golden Smoothee' ($R^2=0.77$), 'Granny Smith'

318   ($R^2=0.73$) and non-fruit thinned (Th-) 'Golden Delicious' ($R^2=0.84$), meaning there

319   was a good relationship between cell wall content and viscosity in NR purees.

320   Visually perceptible differences of color with an increase of L* and b* (with $\Delta E > 2$)

321   were detectable only after 6 months of storage for 'Golden Delicious' and 'Golden

322   Smoothee' (Hunter, & Harold, 1987). Fructose content, as the major individual sugar,

323   increased significantly ($p < 0.001$) in 'Golden Smoothee' and 'Golden Delicious'

324   purees during storage, but not for 'Granny Smith', again in good agreement with the

325   behavior observed in the raw fruits.

326   The changes of SSC, sucrose content and DMC in purees were still the major

327   discriminative contributors for apple varieties and fruit thinning conditions during

328   puree processing. Obvious differentiations were observed for SSC and sucrose

329   content between 'Granny Smith' purees and the other purees, in accordance with their

330   changes in raw apples. In 'Golden Delicious', tree thinning (Th+) led to a significant

331   increase ($p < 0.001$) of DMC both in apples and their corresponding processed purees

332   at each storage period (supplementary information, **Suppl. Table 1 and 2**), in

16

accordance with the fact that the thinned apples, in addition to being larger, also accumulates more cell materials per volume unit (Palmer, Harker, Tustin, & Johnston, 2010). Additionally, the apples of the tree thinning (Th+) gave purees more viscous with significant higher values of η_{100} and K (p < 0.001) than the non-thinning condition (Th-). The tree thinning treatments, by affecting individual apple growth potential, affected physical properties of raw apples and processed purees, including their viscosity. Small fruits from non-thinned trees (Th-) resulted in less viscous purees than large fruits from thinned trees.

Concerning the refined (Ra) purees, an expected clear reduction was obtained for both AIS content (in DW) and viscosity (η_{100}) after refining (**Suppl. Table 2**). That could be due to the loss of insoluble fibers in the removed particle fraction (Colin-Henrion, Mehinagic, Renard, Richomme, & Jourjon, 2009) leading to a loss of puree viscosity (Espinosa-Muñoz, To, Symoneaux, Renard, Biau, & Cuvelier, 2011; Leverrier, Almeida, & Cuvelier, 2016).

3.1.3 Relationship between the fresh apples and puree characteristics

In order to study the link between physical and chemical parameters of raw apples and their processed purees, the coefficients of determination (R^2) were calculated with the dataset including all three varieties under two thinning conditions (Th+ and Th-) and two refining levels and are displayed as heat maps for R^2 values from red ($R^2 > 0.8$) to blue ($R^2 < 0.2$) (**Figure 2**). A clear similarity was observed between the two maps (**Figure 2a and 2b**) with the same blue and red areas. Between all apples and their processed purees, high R^2 values (0.92 in NR and 0.91 in Ra) were obtained for TA.

355   Acceptable correlations were also found for SSC (0.79 in NR and 0.81 in Ra), DMC

356   (0.72 in NR and 0.73 in Ra) and malic acid content (0.65 in NR and 0.61 in Ra). For

357   the AIS (DW) contents, good correlations (R^2) were obtained for each variety between

358   raw apples and NR purees (not for Ra purees): 0.76 in 'Golden Smoothee', 0.83 in

359   'Granny Smith' and 0.77 in 'Golden Delicious', but lower when using all NR purees

360   (R^2= 0.65). Moreover, acceptable correlations (R^2> 0.71) were obtained between

361   texture characteristics (mean load and linear distances) of all apples and rheological

362   parameters (η_{100} and K) of their corresponding purees, whether non-refined (NR)

363   (**Figure 2a**) or refined (Ra) (**Figure 2b**). The rheological variations in processed

364   purees under the effects of different genotypes, storage periods and refining

365   treatments were consistent with the textural changes in apples. However, no

366   significant correlations ($R^2 < 0.44$) were found for individual sugars (glucose, sucrose

367   and fructose) between all apples and their purees (NR and Ra), probably because: i)

368   these concentrations changed less during long-term storage than TA and malic acid

369   content, and ii) water content varied during thermal processing and refining

370   treatments.

371   The good correlations of acidity (TA and malic acid content), SSC, DMC and physical

372   properties (textual and rheological parameters) between apples and their purees were

373   in line with PCA results (**Figure 1b and 1d**). The chemical and physical variations in

374   purees were shown to be potentially linked to the raw apple properties.

375   **3.2 Apple and puree characteristics measured by NIRS**

376   ANOVA was performed on the SNV pre-treated NIR spectra of apples and processed

377   purees (**Figure 3**), in order to point out the wavelengths that varied during processing.

378   According to the F-values, the variability was clearly higher for the spectra of apples

379   (**Figure 3a and 3b**) than those of purees (**Figure 3c and 3d**), as could be expected

380   given that each puree was prepared from 4 kg of fruit. For raw apples, the effect of

381   variety (F-values of 800) was higher than the effect of storage (F-value of 120)

382   (**Figures 3a and 3b**). However, after processing into purees, the opposite conclusion

383   was obtained: the effect of storage (F-values of 110) was almost three times higher

384   than the effect of variety (F-value of 40) (**Figure 3c and 3d**). Combine with their

385   averaged spectral results (not shown), when apples were processed into puree, the

386   peaks at 1930 nm in apples (**Figure 3a**) were not variable in purees (**Figure 3c**),

387   demonstrating that the water contents had very limited variations in the purees.

388   **3.2.1 Discrimination of fresh apples**

389   The wavelength range with the most variability (between 1700 and 2350 nm),

390   identified by the ANOVA (**Figure 3a**), was chosen to discriminate the effects of

391   cultivar and storage. This range was used to perform PCA and FDA (**Figure 4a and**

392   **4b**).

393   The first PCA displayed the discrimination of apples according to the variety (**Figure**

394   **4a**). The first PC-score (PC1) discriminated 'Granny Smith' (GS) on the left and

395   'Golden Delicious' (GD) and 'Golden Smoothee' (GO) on the right, and accounted for

396   83.5% of the total variability. As observed for the reference data, 'Golden Smoothee'

397   spectra were overlapped with those of 'Golden delicious'. The wavelengths at around

398 1880 nm, 1930 nm and 2100-2300 nm were the main contributors of the PC1 (not

399 shown).The two bands at 1880 nm and 1930 nm are explained by the O-H

400 combinations, which have been reported to characterize the water content in apples

401 (Camps, Guillermin, Mauget, & Bertrand, 2007). The broad band at 2100-2300 nm

402 corresponds to the first combination band of C-H bond of sugars or organic acids,

403 already used to determine the concentration of individual sugars in apple juices (León,

404 Kelly, & Dowey, 2005; Liu, Yig, Yu, & Fu, 2006). These fingerprint wavelengths

405 are consistent with the discrimination of apple cultivars harvested in France (Camps,

406 Guillermin, Mauget, & Bertrand, 2007).

407 In a second step, the different storage periods of all apples could be separated (100%

408 of discrimination sensitivity and specificity between T0 and T1 apples, and 98.5% for

409 sensitivity and 99.5% for specificity between T1 and T3 apples) by FDA (**Figure 4b**).

410 It was observed that T3 and T6 apples were overlapped (**Figure 4b**), in line with their

411 changes of mean load and linear distance (PCA could not well-classified storage

412 periods). However, this result was inconsistent with previous report regarding well

413 classification of storage periods of 'Golden Delicious' at 2°C by FDA (Giovanelli,

414 Sinelli, Beghi, Guidetti, & Casiraghi, 2014). In our experiment, the strong variability

415 and heterogeneity from different apple cultivars and fruit-thinning treatments could

416 provide more variations of water contents and carbohydrates, and thus introduced

417 difficulties to well classify the storage stages after 3 months (T3). The use of different

418 storage temperatures might also be involved, with a faster evolution at 4°C than at

419 2°C. The relevant wavelengths were mainly located in the ranges from 1700-1900 nm

420   and 2250 nm (not shown).

421   The wavelengths around 1880 nm, 1930 nm, and 2100-2300 nm could be applied to

422   the discrimination of apple varieties, while those at 1700-1900 nm and 2250 nm could

423   be used for the classification of apple storage periods.

424   **3.2.2 Discrimination of apple purees**

425   For purees, the ANOVA indicated major variations in the following wavelength

426   ranges: 800-1050 nm, 1550-1730 nm, 1870 nm and 2100-2200 nm (**Figures 4c and**

427   **4d**). Thus, the whole wavelength range from 800 to 2500 nm was used for FDA on the

428   spectral dataset of all purees (not well-classified with PCA). The first two factors of

429   the FDA allowed the discrimination of the three cultivars, 'Golden Smoothee' (GO),

430   'Granny Smith' (GS) and 'Golden Delicious' (GD), with the discrimination specificity

431   and sensitivity values of 86.8% and 84.6% in GD and GO apples; 84.0% and 82.2%

432   in GD and GS apples; 88.5% a☐d 91.9% i☐GD a☐d GS apples (**Figure 4c**). The F1

433   and F2 coefficients were both highly correlated with the area between 800 and 1000

434   nm (not shown), which is known as the absorption of apple carbohydrates and water

435   (Giovanelli, Sinelli, Beghi, Guidetti, & Casiraghi, 2014; Zude, Herold, Roger,

436   Bellon-Maurel, & Landahl, 2006), and already used for apple cultivar classification

437   (Bobelyn *et al.*, 2010). Purees could be classified according to the storage periods

438   with a distinct group for T0 and T1 (91.7% of sensitivity and 95.0% of specificity),

439   but a mixed group for T3 and T6 (**Figure 4d**). Besides the aforementioned absorbance

440   region between 800 and 1000 nm, the wavelengths around 1400 nm and between

441   2100 and 2300 nm were also major contributors for discrimination of storage

21

442   durations. These regions have been shown to be related to water loss and SSC

443   variations, and could be regarded as the fingerprint wavelengths of apple storage

444   periods (Camps, Guillermi□, Mauget, & Bertra□d, 2007; Giova□elli, Si□elli, Beghi,

445   Guidetti, & Casiraghi, 2014; Leó□, Kelly, & Dow□ey, 2005).

446   Other interesting results were obtained with the FDA applied on 'Granny Smith'

447   purees taking into account the refining levels and the storage durations (not shown).

448   According to the F1 and F2 axes, the two refining levels (Na and Ra) were separated

449   both at T0 and T1, but not at T3 and T6. This result is highly consistent with the

450   rheological changes of refined and non-refined purees of 'Granny Smith' (**Suppl.**

451   **Table 2**). This refining treatment led to stronger losses of viscosity and cell wall (AIS)

452   content before the first month of apple storage (T1), compared to purees prepared

453   after three months (T3 and T6).

454   **3.3 Prediction of quality traits by NIRS**

455   In this study, we tested the ability of NIR spectra and reference data coupled with PLS

456   to predict the physical and biochemical parameters of: (1) all apples from their NIR

457   spectra (**Suppl. Table 3**); (2) all processed purees (NR and Ra) from their NIR spectra

458   (**Suppl. Table 4**); (3) a□d all purees from the spectral information of apples (**Tables 1**

459   **and 2**).

460   **3.3.1 Prediction of quality traits by NIRS on fresh apples**

461   The prediction models were developed based on the 840 NIR spectra of apples

462   combining 3 varieties, 2 years, 2 fruit thinning practices and 4 storage periods (**Suppl.**

463   **Table 3**). Selected results ($R^2 > 0.8$ obtained for the validation and RPD values > 2)

464   were further discussed.

465   The prediction results of AIS content (R^2=0.85 expressed in dry weight and R^2=0.83

466   in fresh weight) during cold storage stood out in **Suppl. Table 3**. The prediction of

467   AIS content has already been studied but using the destructive mid-infrared technique

468   on freeze-dried apple powder (Canteri, Renard, Le Bourvellec, & Bureau, 2019), and

469   using NIRS to predict AIS content on 'Golden Delicious' apples during seven months

470   storage at 2 ± 0.5°C with a R^2 of 0.96 (Lovász, Merész, & Salgó, 1994). The lower R^2

471   value in our study was probably in relation with the fact that three varieties and fruit

472   thinning conditions were introduced in the same models. For the crunchiness (linear

473   distance), the prediction result was acceptable (R^2=0.82, RPD=2.34), in accordance

474   with previous results obtained on 'Golden Delicious', 'Braeburn' and 'Fuji' apples

475   during a 7 months cold storage (R^2=0.84), but using the crunchiness data from

476   sensory evaluation and the averaged NIR spectra of each group (Mehinagic *et al.*,

477   2003). Good predictions were obtained with DMC (R^2=0.87, RPD= 2.53) and SSC

478   (R^2=0.81, RPD= 2.21), which were in accordance with previous studies (Giovanelli,

479   Si□elli, Beghi, Guidetti, & Casiraghi, 2014; McGlo□e, Jorda□, Seelye, & Clark, 2003).

480   Moreover, acceptable correlation coefficients were obtained for TA (R^2=0.80, RPD=

481   2.09) and for the main organic malic acid (R^2=0.78, RPD=2.03). For the other

482   individual sugar compounds, results were acceptable for fructose content (R^2=0.81,

483   RPD=1.93) and sucrose content (R^2=0.81, RPD=2.14).

484   Consequently, in apples, NIR spectroscopy was a powerful tool to qualify the

485   crunchiness (linear distance), SSC, TA, DMC, content of individual sugars (fructose

486   and sucrose) and AIS. The benefit of AIS content prediction by NIRS was evident

487   because the classical method of extraction and analysis needs a long time and lots of

488   chemical solvents. Our models were also robust, given the large fruit variability used,

489   with factors such as varieties, thinning practices and cold storage periods.

490   **3.3.2 Prediction of quality traits of all purees**

491   For purees, good predictions were observed for global parameters such as DMC

492   (R^2=0.85, RPD=2.42) and SSC (R^2=0.92, RPD=3.12) (**Suppl. Table 4**). The higher R^2

493   regarding SSC in purees than in apples, could possibly due a better homogeneity of

494   the puree samples after processing. Additionally, acceptable results were also be

495   observed for TA (R^2=0.80, RPD=2.22). For individual compounds, results were

496   acceptable only for fructose (R^2=0.83, RPD=2.51). For the physical properties,

497   rheological parameters (η 100, *K* and n) and color (L*, a* and b*), only poor results

498   were obtained in all purees (R^2 < 0.51). However, in 'Granny Smith' purees (not

499   shown), good correlations were obtained between NIRS and η100 (R^2=0.94,

500   RPD=6.53), and *K* (R^2=0.93, RPD=3.52), and n (R^2=0.87, RPD=2.94). It seemed

501   NIRS provided the possibility to access the evolution of rheological parameters in

502   'Granny Smith' purees from different apple storage times, but this relationship was

503   not robust if a large variability of genotypes and agricultural practices were involved.

504   Moreover, the results were surprising for AIS which was well-predicted in the

505   corresponding intact apples (**Suppl. Table 3**), but not in all processed purees (R^2 <

506   0.69, RPD < 1.59). The acceptable prediction of AIS content in apples probably

507   depended on the good correlation between AIS and textural changes (firmness and

508 crunchiness).

3.3.3 Prediction of puree quality traits from NIR spectra of fresh apples

510 In this part, PLS models were developed by combining spectral data acquired on fresh

511 apples and reference data acquired on purees with two approaches: a) use the 48

512 averaged apple spectra (means of spectra of 2 faces x 10 apples by set) and the 48

513 reference data of their corresponding NR or Ra purees (3 replicates x 4 storage

514 periods x 4 puree groups); b) use the 480 averaged spectra of i☐dividual apples

515 (means of faces a and b only) and their 48 reference data of corresponding NR or Ra

516 purees. In this case, the same values of puree characteristics were linked to the 10

517 apples of the same set. These two methods (a and b) obtained similar prediction

518 results and only results of the method b taking into account the apple spectra

519 variability are shown, for both the NR purees (**Table 1**) and Ra purees (**Table 2**).

520 In NR and Ra purees, good predictions were obtained for rheological parameters

521 (η_{100}, *K* and n). Especially for η_{100}, impressive R^2 and RPD values were observed

522 for NR purees (R^2=0.88, RPD=2.31) and Ra purees (R^2=0.82, RPD=2.44). Good

523 results were also obtained for AIS content (expressed in FW and DW) in NR purees

524 (R^2=0.81, RPD=2.23) and Ra purees (R^2=0.84, RPD=2.48). As the AIS content is one

525 of the main contributors of puree viscosity (Leverrier, Almeida, & Cuvelier, 2016),

526 these concomitant results between AIS content and rheological parameters could

527 probably be related to their good correlations in purees. In all studied purees, good

528 correlations were obtained between their AIS and viscosity behaviors (η_{100}) (R^2

529 =0.75), but not for the AIS and SSC values (R^2=0.32). For coloration, acceptable

530   prediction results of b* value were obtained both in NR purees (R^2=0.81, RPD=2.19)

531   and Ra purees (R^2=0.79, RPD=2.12). Moreover, considering the DMC, SSC and TA,

532   the PLS regression models had a good ability to estimate each characteristic for all

533   purees on the basis of acceptable R^2 and RPD values ($R^2 > 0.80$, RPD > 2.11).

534   However, the NIR technique cannot be used to estimate satisfactorily the content of

535   individual sugars (fructose, sucrose, glucose) and of malic acid of purees depending

536   on the spectral information of raw apples.

537   What stands out in these results was the better predictions of some quality traits of

538   puree from fresh apple spectra (**Table 1 and 2**) than from the puree spectra directly

539   (**Suppl. Table 4**). It was the case for rheological parameters, R^2=0.82 from apples and

540    $R^2 < 0.44$ from purees, possibly owing to the acceptable links ($R^2 > 0.71$) between

541   apple texture (mean load and linear distance) and puree rheological properties (**Figure**

542   **2**). In addition, better PLS results to predict AIS content from fresh apple spectra

543   (R^2=0.81 in FW and 0.84 in DW) were obtained than from purees spectra ($R^2 < 0.69$)

544   (**Suppl. Table 4**), probably due to good relationships between puree viscosity and AIS

545   content mentioned above. Besides, the prediction of DMC, SSC and TA were still

546   acceptable in all cases. Therefore, NIR technique showed a potential to directly

547   predict the viscosity properties, b*, AIS content, SSC, DMC and TA of processed

548   purees using their corresponding apple spectral information directly.

549   Compared with the PLS models used to predict the characteristics of purees based on

550   their own spectra (**Suppl. Tables 3 and 4**), more LVs and lower prediction accuracy

551   (RPD values) have generally found when models were built using the spectra of the

552 intact apples to predict the characteristics of the processed purees. This fact has been

553 also observed when other raw materials, e.g. meat (Meullenet, Jonville, Grezes, &

554 Owens, 2004) or whole grain (Windham et al., 1997) were used to predict quality

555 traits of the final cooked food. Such indirect prediction is a challenge as the spectra

556 were not acquired on the material for which prediction was done, and because the

557 chemical and textural traits from the material on which the spectra were acquired (the

558 raw apples) are modified by processing. However, such predictions, albeit only

559 semi-quantitative, are relevant for industrial use. Indeed, these developed models

560 provide a promising solution to evaluate puree viscosity, a primary quality trait of this

561 product, and cell wall contents (AIS), only based on spectra of fresh apples. A

562 remarkable fact was that these predictions could not be done using the NIR spectra of

563 the purees themselves (**Suppl. Table 4**).

564 **4. Conclusion**

565 As far as we know, this was the first report concerning the assessment of quality

566 variation of apple purees depending on NIR spectral information of the corresponding

567 raw apples. Up to now, in apple industry, manufacturers use their experience and

568 knowhow to make blend of apples in order to obtain always the same puree. From our

569 results, NIR had the potential to predict internal quality of apples but also that of their

570 processed products: a reliable assessment of texture and taste of the purees could be

571 obtained based only on spectral data of fresh apples. This opens the possibility to sort

572 or select apples according to the expected purees. By systematically scanning all

573 apples, this could provide some objective data to predict the final product

574 characteristics and thus reduce waste of materials along the processing chain. Further

575 work will be needed to investigate the interaction of the processing conditions

576 (temperature, time, oxygen and so on) with raw apples under various growing and

577 storage conditions, so as to provide guidance for adapted processing procedures to

578 reach stable final puree qualities.

579 **Acknowledgements**

587

References

Bengoechea, M. L., Sancho, A. I., Bartolomé, B., Estrella, I., Gómez-Cordovés, C., Hernández, M. T. (1997). Phenolic composition of industrially manufactured purees and concentrates from peach and apple fruits. *Journal of Agricultural and Food Chemistry, 45*(10), 4071-4075.

Bertrand, D., & Cordella, C. (2008). SAISIR package. Free toolbox for chemometrics in the Matlab, Octave or Scilab environments.

Billy, L., Mehinagic, E., Royer, G., Renard, C. M., Arvisenet, G., Prost, C., & Jourjon, F. (2008). Relationship between texture and pectin composition of two apple cultivars during storage. *Postharvest Biology and Technology, 47(3),* 315-324.

Bobelyn, E., Serban, A. S., Nicu, M., Lammertyn, J., Nicolai, B. M., & Saeys, W. (2010). Postharvest quality of apple predicted by NIR-spectroscopy: Study of the effect of biological variability on spectra and model performance. *Postharvest Biology and Technology, 55*(3), 133-143.

Bureau, S., Scibisz, I., Le Bourvellec, C., & Renard, C. M. (2012). Effect of sample preparation on the measurement of sugars, organic acids, and polyphenols in apple fruit by mid-infrared spectroscopy. *Journal of Agricultural and Food Chemistry, 60*(14), 3551-3563.

Camps, C., Guillermin, P., Mauget, J. C., & Bertrand, D. (2007). Discrimination of Storage Duration of Apples Stored in a Cooled Room and Shelf-Life by Visible-Near Infrared Spectroscopy. *Journal of Near Infrared Spectroscopy, 15*(3), 169-177.

Canteri, M. H., Renard, C. M., Le Bourvellec, C., & Bureau, S. (2019). ATR-FTIR spectroscopy to determine cell wall composition: Application on a large diversity of fruits and vegetables. *Carbohydrate Polymers*, 212, 186-196.

Colin-Henrion, M., Cuvelier, G., & Renard, C. (2007). Texture of pureed fruit and vegetable foods. *Stewart Postharvest Review, 5*, 1-14.

Colin-Henrion, M., Mehinagic, E., Renard, C. M., Richomme, P., & Jourjon, F. (2009). From apple to applesauce: Processing effects on dietary fibres and cell wall polysaccharides. *Food Chemsitry 117*(2), 254-260.

Defernez, M., Kemsley, E. K., & Wilson, R. H. (1995). Use of infrared spectroscopy and chemometrics for the authentication of fruit purees. *Journal of Agricultural and Food Chemistry*, 43(1), 109-113.

Espinosa-Muñoz, L., Renard, C. M. G. C., Symoneaux, R., Biau, N., & Cuvelier, G. (2013). Structural parameters that determine the rheological properties of apple puree. *Journal of Food Engineering, 119*(3), 619-626.

Espinosa-Muñoz, L., Symoneaux, R., Renard, C. M. G. C., Biau, N., & Cuvelier, G. (2012). The significance of structural properties for the development of innovative apple puree textures. *LWT - Food Science and Technology, 49*(2), 221-228.

Espinosa-Muñoz, L., To, N., Symoneaux, R., Renard, C. M. G. C., Biau, N., & Cuvelier, G. (2011). Effect of processing on rheological, structural and sensory properties of apple puree. *Procedia Food Science, 1*, 513-520.

Giovanelli, G., Sinelli, N., Beghi, R., Guidetti, R., & Casiraghi, E. (2014). NIR spectroscopy for the optimization of postharvest apple management. *Postharvest Biology and Technology, 87*, 13-20.

Gregson, C. M., & Lee, T. C. (2002). Evaluation of numerical algorithms for the instrumental measurement of bowl-life and changes in texture over time for ready-to-eat breakfast cereals. *Journal of Texture Studies, 33*(6), 505-528.

Gu, Z., Eils, R., & Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics, 32(18)*, 2847-2849.

Hunter, R. S., & Harold, R. W. (1987). *The measurement of appearance*: John Wiley & Sons.

Johnston, J. W., Hewett, E. W., Hertog, M. L. (2002). Postharvest softening of apple (Malus domestica) fruit: a review. *New Zealand Journal of Crop and Horticultural Science*, 30(3), 145-160.

Keenan, D. F., Brunton, N., Butler, F., Wouters, R., & Gormley, R. (2011). Evaluation of thermal and high hydrostatic pressure processed apple purees enriched with prebiotic inclusions. *Innovative Food Science & Emerging Technologies, 12*(3), 261-268.

Le Bourvellec, C., Bouzerzour, K., Ginies, C., Regis, S., Plé, Y., & Renard, C. M. G. C. (2011). Phenolic and polysaccharidic composition of applesauce is close to that of apple flesh. *Journal of Food Composition and Analysis, 24*(4-5), 537-547.

León, L., Kelly, J. D., & Downey, G. (2005). Detection of apple juice adulteration using near-infrared transflectance spectroscopy. *Applied spectroscopy, 59*(5), 593-599.

Leverrier, C., Almeida, G., & Cuvelier, G. (2016). Influence of particle size and concentration on rheological behaviour of reconstituted apple purees. *Food Biophysics, 11*(3), 235-247.

Liu, Y., & Ying, Y. (2005). Use of FT-NIR spectrometry in non-invasive measurements of internal quality of 'Fuji' apples. *Postharvest Biology and Technology, 37*(1), 65-71.

Liu, Y., Ying, Y., Yu, H., & Fu, X. (2006). Comparison of the HPLC method and FT-NIR analysis for quantification of glucose, fructose, and sucrose in intact apple fruits. *Journal of Agricultural and Food Chemistry, 54*(8), 2810-2815.

Loncaric, A., Dugalic, K., Mihaljevic, I., Jakobek, L., & Pilizota, V. (2014). Effects of sugar addition on total polyphenol content and antioxidant activity of frozen and freeze-dried apple puree. *Journal of Agricultural and Food Chemistry, 62*(7), 1674-1682.

Lovász, T., Merész, P., & Salgó, A. (1994). Application of near infrared transmission spectroscopy for the determination of some quality parameters of apples. *Journal of Near Infrared Spectroscopy, 2*(4), 213-221.

Luo, W., Huan, S., Fu, H., Wen, G., Cheng, H., Zhou, J., Wu, H., Shen, G., & Yu, R. (2011). Preliminary study on the application of near infrared spectroscopy and pattern recognition methods to classify different types of apple samples. *Food Chemistry, 128*(2), 555-561.

McGlone, V. A., Jordan, R. B., Seelye, R., & Clark, C. J. (2003). Dry-matter - a better predictor of the post-storage soluble solids in apples? *Postharvest Biology and Technology, 28*(3), 431-435.

Mehinagic, E., Royer, G., Bertrand, D., Symoneaux, R., Laurens, F., Jourjon, F. (2003). Relationship between sensory analysis, penetrometry and visible–NIR spectroscopy of apples belonging to different cultivars. *Food quality and preference, 14*(5-6), 473-484.

Menesatti, P., Zanella, A., D'Andrea, S., Costa, C., Paglia, G., Pallottino, F. (2009). Supervised multivariate analysis of hyper-spectral NIR images to evaluate the starch index of apples. *Food and Bioprocess Technology, 2*(3), 308-314.

Meullenet, J. F., Jonville, E., Grezes, D., & Owens, C. M. (2004). Prediction of the texture of cooked poultry pectoralis major muscles by near-infrared reflectance analysis of raw meat. *Journal of texture studies, 35*(6), 573-585.

Nicolai, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology, 46*(2), 99-118.

Nargis, H. F., Nawaz, H., Ditta, A., Mahmood, T., Majeed, M. I., Rashid, N., Muddassar, M., Bhatti,

H.N., Saleem, M., Jilani, K., Bonnier, F., & Byrne, H.J. (2019). Raman spectroscopy of blood plasma samples from breast cancer patients at different stages. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 222,* 117210.

Oszmiański, J., Wolniak, M., Wojdyło, A., & Wawer, I. (2008). Influence of apple purée preparation and storage on polyphenol contents and antioxidant activity. *Food Chemistry, 107*(4), 1473-1484.

Palmer, J. W., Harker, F. R., Tustin, D. S., & Johnston, J. (2010). Fruit dry matter concentration: a new quality metric for apples. *Journal of the Science of Food and Agriculture, 90*(15), 2586-2594.

Peirs, A., Tirry, J., Verlinden, B., Darius, P., & Nicolaï, B. M. (2003). Effect of biological variability on the robustness of NIR models for soluble solids content of apples. *Postharvest Biology and Technology, 28*(2), 269-280.

Picouet, P. A., Landl, A., Abadias, M., Castellari, M., & Viñas, I. (2009). Minimal processing of a Granny Smith apple purée by microwave heating. *Innovative Food Science & Emerging Technologies, 10*(4), 545-550.

R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.r-project.org/.

Renard, C. M. (2005). Variability in cell wall preparations: quantification and comparison of common methods. *Carbohydrate Polymers, 60*(4), 515-522.

Schmutzler, M., & Huck, C. W. (2016). Simultaneous detection of total antioxidant capacity and total soluble solids content by Fourier transform near-infrared (FT-NIR) spectroscopy: A quick and sensitive method for on-site analyses of apples. *Food Control, 66*, 27-37.

Sukhonthara, S., Kaewka, K., & Theerakulkait, C. (2016). Inhibitory effect of rice bran extracts and its phenolic compounds on polyphenol oxidase activity and browning in potato and apple puree. *Food Chemistry, 190*, 922-927.

USDA. (2018). Fresh deciduous Fruit: World markets and trade (apples, grapes, & pears).

Windham, W. R., Lyon, B. G., Champagne, E. T., Barton, F. E., Webb, B. D., McClung, A. M., Moldenhauer, K. A., Linscombe, S., & McKenzie, K. S. (1997). Prediction of cooked rice texture quality using near-infrared reflectance analysis of whole-grain milled samples. *Cereal Chemistry, 74*(5), 626-632.

Zude, M., Truppel, I., & Herold, B. (2002). An approach to non-destructive apple fruit chlorophyll determination. *Postharvest Biology and Technology, 25*(2), 123-133.

Zude, M., Herold, B., Roger, J. M., Bellon-Maurel, V., & Landahl, S. (2006). Non-destructive tests on the prediction of apple fruit flesh firmness and soluble solids content on tree and in shelf life. *Journal of Food Engineering, 77*(2), 254-260.

711 **Figure captions**

712 **Figure 1.** Principal Component Analysis of physical, physiological and biochemical

713 compositions for: apples **(a)** and processed NR (no refined) purees **(c)** in three

714 varieties (GO: 'Golden Smoothee' , GD: 'Golden Delicious' and GS: Granny Smith)

715 growing under two different thinning conditions (Th+ marked with red solid circle

716 and Th- with red dotted circle) and stored at 4°C from harvest (T0), 1 (T1), 3 (T3) and

717 6 (T6) months. Correlation plot of the first principal components for apples **(b)** and

718 NR purees **(d)**.

719 **Figure 2.** Determination coefficient (R^2) of all physical and biochemical parameters

720 between the 'Golden Smoothee', 'Granny Smith' and 'Golden Delicious' apples

721 (titled with "F") and their processed purees (titled with "P"): **(a)** raw apples and

722 non-refi☐ed (NR) purees; **(b)** raw apples and refined (Ra) purees.

723 **Figure 3.** ANOVA results of the SNV pre-treated NIR spectra between 800 and 2500

724 nm: **(a)** effect of variety o☐all apples spectra; **(b)** effect of storage period on all apples

725 spectra; **(c)** effect of variety o☐ all purees spectra; **(d)** effect of storage period on all

726 purees spectra;

727 **Figure 4.** Factorial maps of the SNV pre-treated NIR spectra of all apples between

728 1700 and 2350 nm or all purees between 800 and 2500 nm: **(a)** Principal Component

729 A☐alysis showi☐g apples cultivars; **(b)** the Factorial Discriminant Analysis of storage

730 periods of apples; **(c)** Factorial Discriminant Analysis of cultivars of all purees; **(d)**

731 Factorial Discriminant Analysis of storage periods of all purees.

732 'Golden Smoothee' (GO), 'Golden Delicious' (GD), 'Granny Smith' (GS); storage

733   duration at 4°C from harvest (T0), 1 (T1), 3 (T3) and 6 (T6) months.
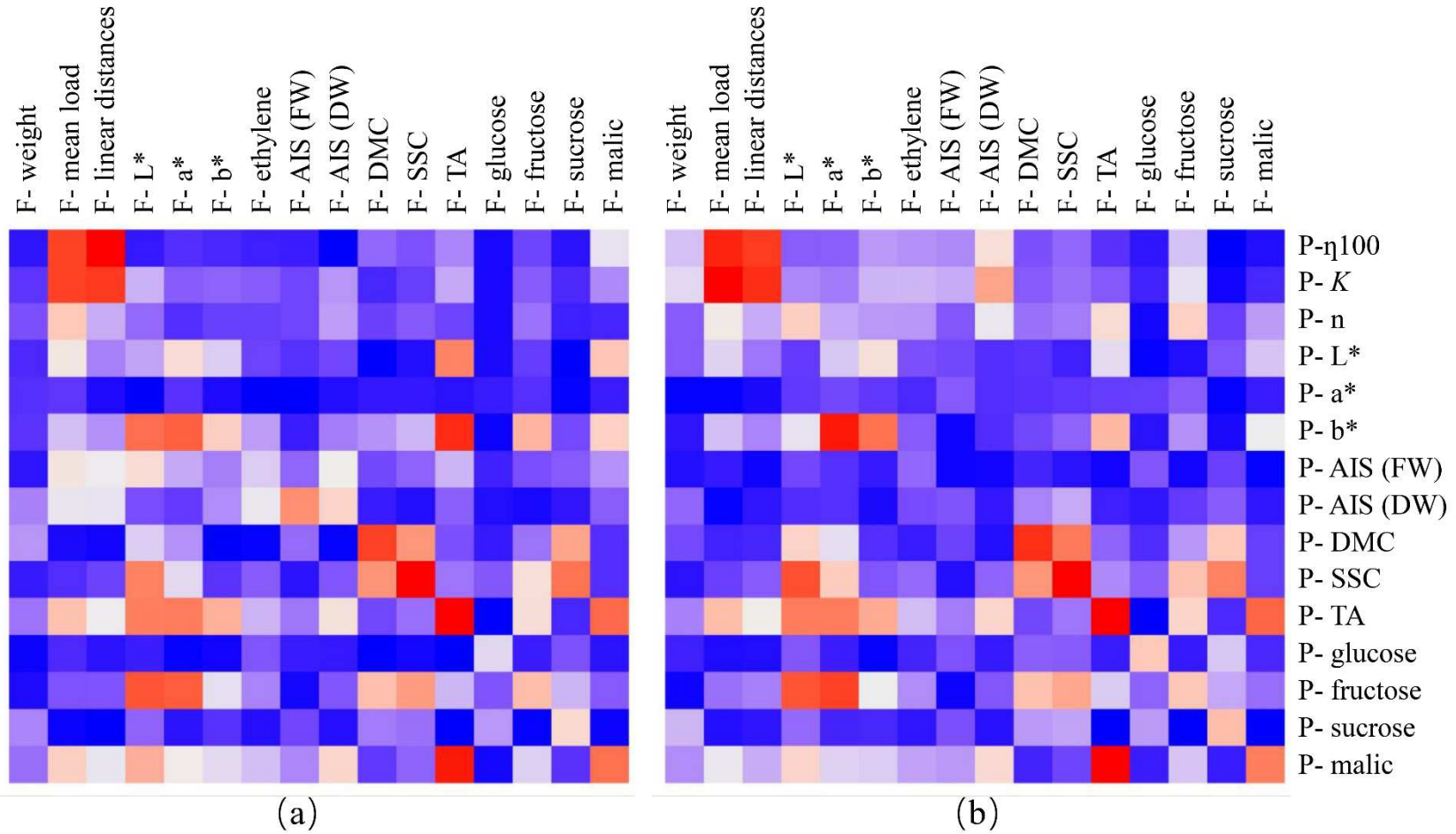
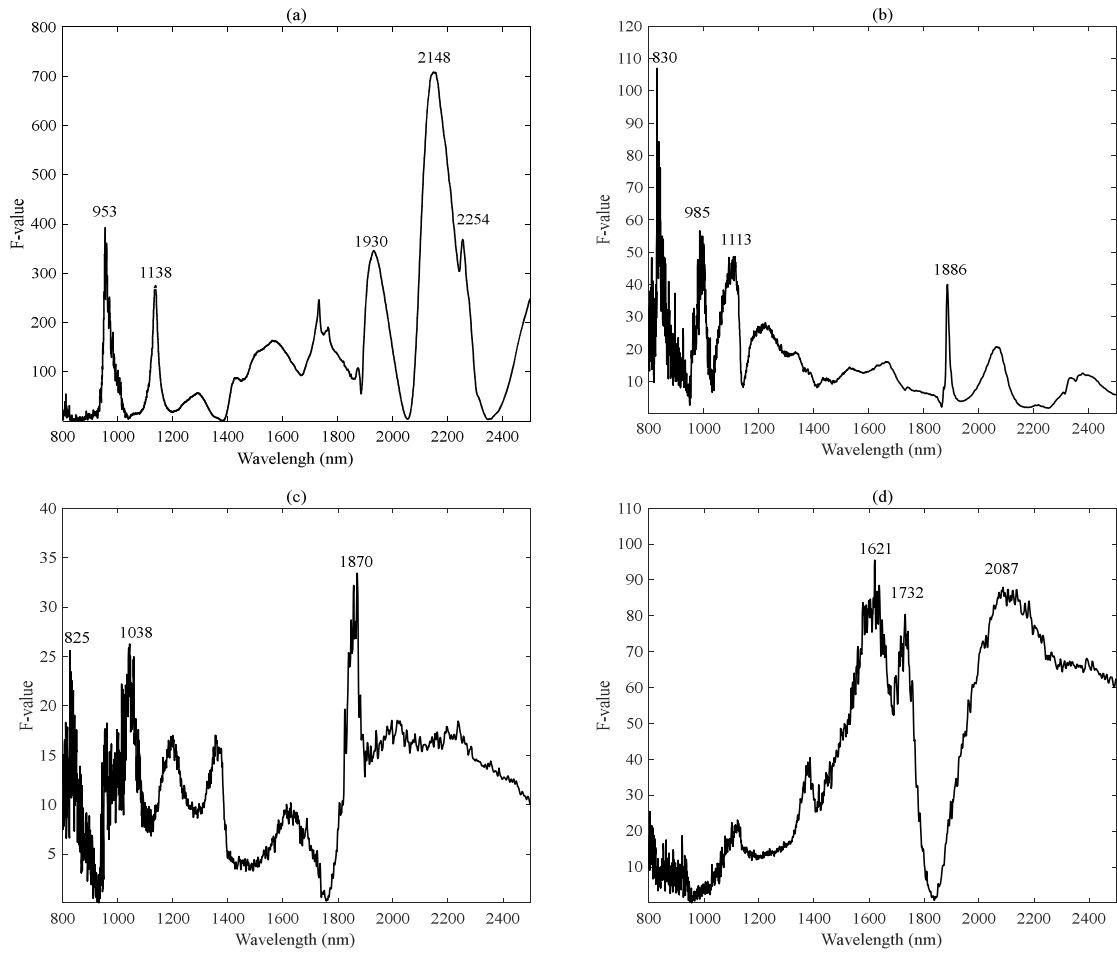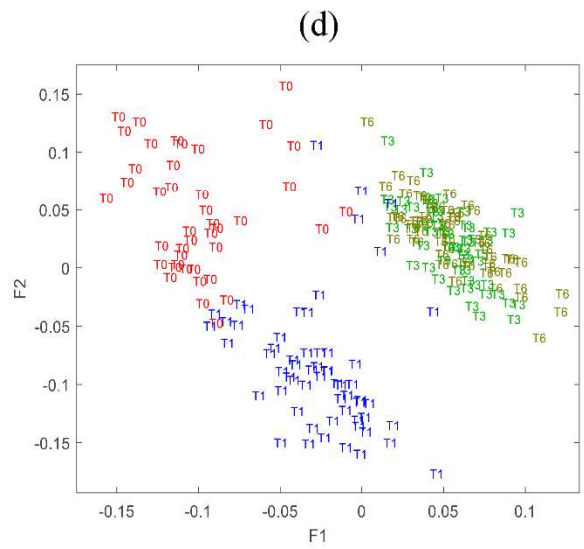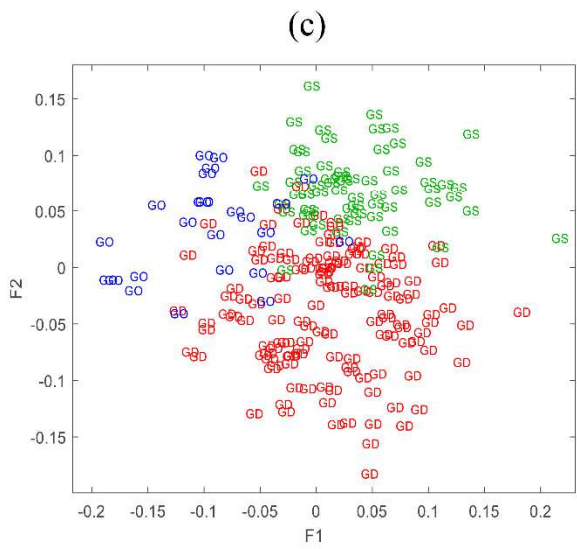734 **Figures**
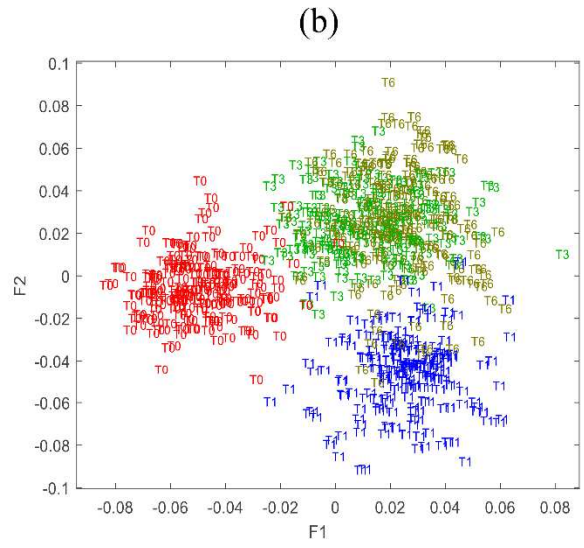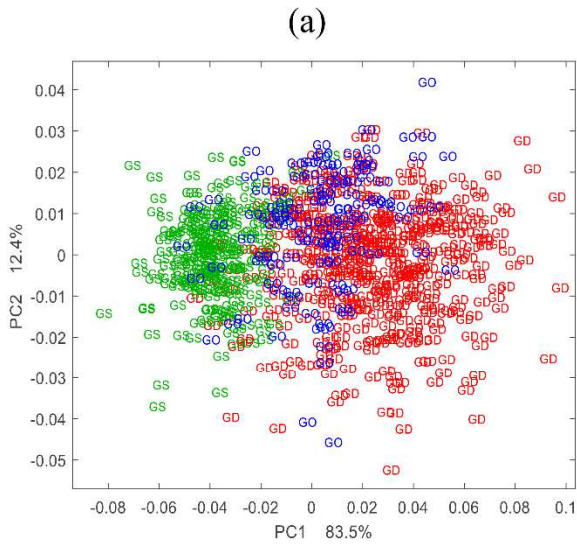

735

736 **Figure 1**

(a)　　　　　　　　　　　　　　(b)

Figure 2

740

741 **Figure 3**

742
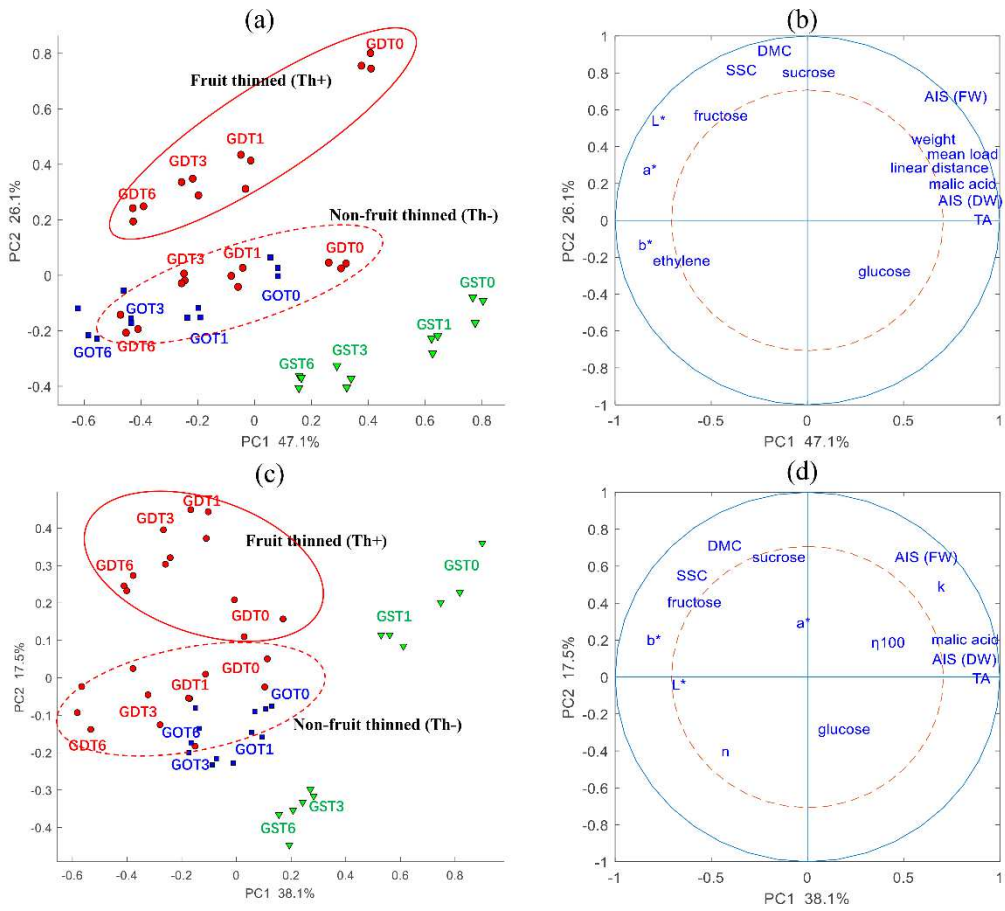
Figure 4

746 **Table 1.** Prediction of puree quality traits from spectral data of fresh apples: PLS (Partial Least Squares) results using NIR spectra of fresh

747 apples from three varieties ('Granny Smith', 'Golden Delicious' and 'Golden Smoothee'), two thinning conditions and two harvest seasons, at 4

748 cold storage durations (0, 1, 3 and 6 months) for prediction of quality traits of non-refined (NR) purees.

| Parameter | range | SD | Calibration n=320 | | Validation n=160 | | LVs | RPD | spectral range(nm) |
			r^2	RMSEC	R^2	RMSEV			
CSR (η_{100})	0.49-1.45	0.19	0.89	0.06	0.88	0.08	10	2.31	900- 2500
viscosity-K	16.8-59.5	11.24	0.79	5.0	0.79	5.2	12	2.15	900- 2500
viscosity-n value	0.06-0.39	0.07	0.83	0.03	0.82	0.03	12	2.06	900- 2500
L*	44.0-53.5	1.9	0.81	0.8	0.81	1.1	10	1.77	900- 2500
a*	-(5.0-3.4)	0.3	0.48	0.3	0.46	0.3	10	1.36	900- 2500
b*	9.2-23.0	3.5	0.84	1.4	0.81	1.6	12	2.19	900- 2500
AIS（mg/g DW）	114.0-171.7	32.6	0.82	16.4	0.80	16.3	11	2.00	900- 2500
AIS（mg/g FW）	19.3-48.9	5.8	0.82	2.5	0.81	2.6	13	2.23	900- 2500
DMC (g/g FW)	0.16-0.23	0.02	0.84	0.01	0.83	0.01	12	2.11	900- 2500
SSC (°Brix)	10.5-18.6	2.2	0.85	0.9	0.80	1.0	11	2.25	900- 2500
TA (meq/kg FW)	25.0-103.9	20.0	0.83	8.3	0.80	9.4	11	2.14	900- 2500
glucose (g/kg FW)	13.5-25.4	3.1	0.65	1.8	0.60	2.0	10	1.55	900- 2500
fructose (g/kg FW)	40.0-98.7	17.2	0.80	7.6	0.73	9.1	11	1.90	900- 2500
sucrose (g/kg FW)	32.2-118.5	22.0	0.80	9.6	0.76	11.1	11	1.98	900- 2500
malic (g/kg FW)	2.4-9.0	1.5	0.77	0.7	0.76	0.8	10	1.91	900- 2500

749
750

751 **Table 2.** Prediction of puree quality traits from spectral data of fresh apples: PLS (Partial Least Squares) results using NIR spectra of fresh

752 apples from three varieties ('Granny Smith', 'Golden Delicious' and 'Golden Smoothee'), two thinning conditions and two harvest seasons, at 4

753 cold storage durations (0, 1, 3 and 6 months) for prediction of quality traits of refined (Ra) purees.

| T | | | Calibration n=320 | | Validation n=160 | | | | |
Parameter	range	SD	r^2	RMSEC	R^2	RMSEV	LVs	RPD	spectral range(nm)
CSR (η_{100})	0.25-0.99	0.18	0.89	0.06	0.82	0.07	11	2.44	900- 2500
viscosity-K	2.8-39.1	8.81	0.87	3.0	0.86	3.7	10	2.40	900- 2500
viscosity-n value	0.17-0.49	0.07	0.85	0.03	0.82	0.03	13	2.11	900- 2500
L*	43.9-53.8	2.2	0.6	1.3	0.61	1.4	10	1.54	900- 2500
a*	-(5.2-3.6)	0.4	0.5	0.3	0.51	0.4	8	1.04	900- 2500
b*	7.8-22.9	3.4	0.8	1.6	0.79	1.6	12	2.12	900- 2500
AIS（mg/g DW）	90.9-189.6	18.3	0.79	8.4	0.76	8.9	12	2.05	900- 2500
AIS（mg/g FW）	14.8-33.3	4.1	0.85	1.6	0.84	1.6	12	2.48	900- 2500
DMC (g/g FW)	0.15-0.24	0.02	0.84	0.01	0.84	0.01	12	2.37	900- 2500
SSC (°Brix)	10.3-17.6	2.0	0.82	0.9	0.80	0.9	11	2.16	900- 2500
TA (meq/kg FW)	25.2-109.1	21.6	0.83	8.9	0.80	9.6	11	2.26	900- 2500
glucose (g/kg FW)	13.9-25.7	3.1	0.65	1.8	0.62	2.1	10	1.50	900- 2500
fructose (g/kg FW)	42.1-99.9	15.6	0.71	9.5	0.70	10.0	10	1.56	900- 2500
sucrose (g/kg FW)	33.4-123.1	23.5	0.78	11.0	0.76	11.0	11	2.14	900- 2500
malic (g/kg FW)	2.9-8.3	1.5	0.79	0.7	0.75	0.8	10	1.93	900- 2500

754