



HAL
open science

Fruit variability impacts puree quality: Assessment on individually processed apples using the visible and near infrared spectroscopy

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

► To cite this version:

Weijie Lan, Benoit Jaillais, Songchao Chen, Catherine M.G.C. Renard, Alexandre Leca, et al.. Fruit variability impacts puree quality: Assessment on individually processed apples using the visible and near infrared spectroscopy. *Food Chemistry*, 2022, 390, pp.133088. 10.1016/j.foodchem.2022.133088 . hal-03698772

HAL Id: hal-03698772

<https://hal.inrae.fr/hal-03698772v1>

Submitted on 9 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

1 **Fruit variability impacts puree quality: assessment on individually processed**
2 **apples using the visible and near infrared spectroscopy**

3 Weijie Lan^a, Benoit Jaillais^b, Songchao Chen^c, Catherine M.G.C. Renard^{a,d}, Alexandre
4 Leca^a, Sylvie Bureau^{a*}

5
6 ^a INRAE, Avignon University, UMR408 Sécurité et Qualité des Produits d'Origine
7 Végétale, F-84000 Avignon, France.

8 ^b INRAE, ONIRIS, Unité Statistiques, Sensométrie, Chimiométrie (StatSC), F-44322
9 Nantes, France.

10 ^c INRAE, Unité InfoSol, F-45075 Orléans, France.

11 ^d INRAE, TRANSFORM Division, F-44000 Nantes, France.

12
13 **The corresponding author***

14 Sylvie Bureau (E-mail: sylvie.bureau@inrae.fr).

15 INRAE, UMR408 SQPOV « Sécurité et Qualité des Produits d'Origine Végétale »
16 228 route de l'Aérodrome CS 40509

17 F-84914 Avignon cedex 9

18 Tel: +33 432722509

19 **Other authors**

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

21 Benoit Jaillais: benoit.jaillais@inrae.fr

22 Weijie Lan: Weijie.Lan@inrae.fr

23 Alexandre Leca: Alexandre.Leca@inrae.fr

24 Songchao Chen: chensongchao@zju.edu.cn

25

26 **Highlights**

27 Visible-NIR was applied on single apples and their corresponding cooked purees.

28 Apple inter and intra-variability made highly variable cooked purees for viscosity.

29 A strong correlation of spectra was detected between single apples and their purees.

30 The indirect prediction of puree quality from apple spectra was confirmed.

31

32 **Abstract**

33 This study was designed to have the absolute definition of ‘one apple to one puree’,
34 which gave a first insight into the impacts of fruit inter-variability (between varieties)
35 and intra-variability (between individual fruits) on the quality of processed purees. Both
36 the inter-variability of apple varieties and the intra-variability of single apples induced
37 intensive changes of appearance, chemical and textural properties of their
38 corresponding microwave-cooked purees. The intra-variability of cooked purees was
39 different according to apple cultivars. Some strong correlations of visible-near infrared
40 (VIS-NIR) spectra were observed between fresh and cooked apples, particularly in the
41 regions 665-685 nm and 1125-1400 nm. These correlations allowed then the indirect
42 predictions of puree color (a^* and b^* , $RPD \cong 2.1$), viscosity ($RPD \cong 2.3$), soluble
43 solids content (SSC, $RPD = 2.1$), titratable acidity ($RPD = 2.8$), and pH ($RPD = 2.5$)
44 from the non-destructive acquired VIS-NIR spectra of raw apples.

45

46 **Keywords:**

47 *Malus x domestica* Borkh.; Apple variability; Two-dimensional correlation
48 spectroscopy (2D-COS); Partial least square regression; Machine learning regression.

49

50 1. Introduction

51 Apple puree is one of the most popular fruit processed products (over 0.3 million
52 tons consumed per year in France) (FranceAgriMer, 2017) used as a basic ingredient of
53 jams, preserves or compotes and fruit-based baby food (Defernez, Kemsley, & Wilson,
54 1995). The usual industrial conditions to process apple purees are a cooking at 93 - 98 °C
55 for about 4 - 5 min and a pasteurization at 90 °C for around 20 min to obtain a shelf-
56 life of 6 months at room temperature (Oszmiański, Wolniak, Wojdyło, & Wawer, 2008).
57 Such conventional cooking conditions allow the investigation of the ‘inter-variability’
58 among apple cultivars (Buegy, Rolland-Sabaté, Leca, & Renard, 2021; Lan, Bureau,
59 Chen, Leca, Renard, & Jaillais, 2021). In these conditions, the different apple batches
60 of one variety and their cooked purees still presented a high variability due to
61 agricultural practices and storage conditions, affecting the quality characteristics and
62 levels of final products (Lan, Jaillais, Leca, Renard, & Bureau, 2020). However, these
63 experiments did not make possible to address the impact of ‘intra-variability’ between
64 the individual apples on their corresponding cooked purees. Knowing the ‘intra-
65 variability’ between raw fruits and cooked purees can help field growers and industrial
66 manufacturers to sort fruits and produce sustainable and expected final products. From
67 a research point of view, understanding the relationships between raw and processed
68 apples, made possible here by the exact link between each apple and its puree, could
69 contribute to a better management of fruit processing.

70 Microwave processing has the advantage of heating solids such as apples, rapidly
71 and uniformly, inactivating the enzymes and then preserving quality, such as color,
72 texture, polyphenols etc. (Guo, Sun, Cheng, & Han, 2017). It has already been applied
73 on apple batches to produce purees (Oszmiański et al., 2008; Picouet, Landl, Abadias,
74 Castellari, & Viñas, 2009) and also reported to be a mini-processing strategy to process
75 one apple into one puree (Picouet et al., 2009). With our objective to assess the impact
76 of ‘inter’ and ‘intra’ variability of raw fruits on the processed purees, microwave
77 processing gives the possibility to individually cook apples in order to study the direct
78 relationship of quality and properties between one apple and one puree.

79 Visible-near infrared (VIS-NIR) spectroscopy, known as a rapid, relatively cheap,
80 easy-to-use and non-destructive technique for apple online sorting (Huang, Lu & Chen,
81 2020) and quality assessment (Xia, Fan, Li, Tian, Huang, & Chen, 2020). And it has
82 been applied for detecting the different apple species in mixed purees (Lan, Bureau,
83 Chen, Leca, Renard, & Jaillais, 2021) and evaluating apple puree major components
84 (soluble solids, titratable acidity and dry matter, etc.) (Lan, Jaillais, Leca, Renard, &
85 Bureau, 2020). From our previous works, strong correlations of chemical and textural
86 properties have been pointed out between raw apples and their corresponding purees
87 (Lan, Jaillais, Leca, Renard, & Bureau, 2020; Lan, Renard, Jaillais, Leca, & Bureau,
88 2020). These results opened a new possibility to predict the quality of final processed
89 purees from the nondestructive spectral information acquired on a batch of raw apples
90 by developing regression models associating the infrared spectra of raw apples with the
91 reference data of corresponding processed purees (Lan, Jaillais, Leca, Renard, &
92 Bureau, 2020). However, these relationships between fresh and processed apples were
93 obtained using a laboratory-scale cooker-cutter processing system (Roboqbo, Qb8-3,
94 Bentivoglio, Italy) needing at least 2.5 kg of raw fruits. This means around 15 apples
95 were processed in a single puree, ignoring the ‘intra-variability’ brought by each
96 individual apple. Indeed, a strong variability and heterogeneity due to color, chemical
97 and textural properties of raw apples (Lan, Jaillais, Renard, Leca, Chen, Le Bourvellec,
98 & Bureau, 2021; Pissard, Baeten, Romnée, Dupont, Mouteau, & Lateur, 2012) and a
99 large variability of puree characteristics (different cultivars) have been clearly
100 highlighted (Lan, Bureau, Chen, Leca, Renard, & Jaillais, 2021). As far as we know,
101 there has been no attempt to investigate the effect of both ‘inter’ and ‘intra’ variability
102 at the level of single fruit (size, appearance and chemical properties etc.) on the quality
103 of final processed products. Besides, no similar work linking VIS-NIR spectra of
104 individual fruits to their processed products characteristics and spectra had been
105 reported. The challenge here was to know how much the inter- and/or intra-variability
106 of raw apples impacts cooked purees? How VIS-NIR spectral data were affected due to
107 the physical and chemical changes considering the experimental design of ‘one apple
108 to one puree’? The potential of predicting the quality traits of the final cooked purees

109 using the VIS-NIR spectra of intact raw apples was also investigated.

110 Accordingly, VIS-NIR spectroscopy and reference data determination were
111 performed on 120 individual apples of 4 varieties and their corresponding individual
112 processed purees, in order to reach three aims: i) investigating the inter- and intra-
113 variability of both, the individual apples and corresponding purees; ii) exploring the
114 spectral correlations and variations before and after each apple processing; and iii)
115 predicting the textural properties and biochemical composition of cooked purees from
116 the VIS-NIR spectra of individual raw apples using direct modelling methods.

117 **2. Material and methods**

118 **2.1 Apple materials**

119 Apple of four varieties: ‘Golden Delicious’ (GD), ‘Granny Smith’ (GS), ‘Breaburn’
120 (BR) and ‘Royal Gala’ (GA) were harvested at a commercial maturity from La Pugère
121 experimental orchard (Mallemort, Bouches du Rhône, France) (**Fig. 1**). All apples were
122 stored for four months at 4°C before processing. In total, 120 individual apples (4
123 varieties × 10 apples × 3 weeks) were measured with the non-destructive techniques
124 (color, VIS-NIR spectra).

125 **2.2. Nondestructive characterization of individual apples**

126 The color of all apple skins (un-blushed and blushed sides) was determined three
127 times using a CE-400 chromameter (Minolta, Osaka, Japan), and expressed in the CIE
128 1976 L* a* b* color space (illuminant D65, 0° view angle, illumination area diameter
129 8 mm).

130 VIS-NIR spectra of raw apples were acquired using two multi-purpose analyzer
131 spectrometers (Bruker Optics®, Wissembourg, France) at 23°C, which provide diffuse
132 reflectance measurements at wavelength from 500-780 nm (VIS) and 780-2500 nm
133 (NIR), with a spectral resolution of 2 nm. For each spectrum, 32 scans were recorded
134 and averaged. The spectral acquisition and instrument adjustments were controlled by
135 OPUS software Version 5.0 (Bruker Optics®). For each apple, VIS-NIR spectra were
136 collected on the blushed and un-blushed sides through a 18 mm diameter area of

137 infrared light. Afterwards, the averaged VIS-NIR spectra, corresponding to the blushed
138 and un-blushed sides of each apple were calculated for further analysis. A reference
139 background measurement was automatically activated before each data set acquisition
140 using an internal Spectralon reference. In total, 120 VIS-NIR spectra of different apples
141 (4 varieties \times 10 apples \times 3 weeks) were treated before and after cooking.

142 **2.3 Individual apple processing**

143 Individual and intact apples were sealed in a domestic preserving container (the
144 length, width, and height of 20 cm \times 20 cm \times 12 cm) and placed at the center of an
145 experimental microwave oven (CM1529, Samsung, Korea). Microwave processing was
146 conducted at a power of 1.5 kW for 3 min, and then at 0.7 kW for 1 min. Afterwards,
147 each apple was immediately refined with a 0.5 mm sieve using a manual refiner
148 (A45306, Moulinex, France). Finally, each individual puree was conditioned in a
149 hermetically sealed can, and then placed at 23 °C during one day before further analyses.
150 Totally 120 purees (4 varieties \times 10 purees \times 3 weeks) were obtained **during the three**
151 **successive weeks of processing replicates.**

152 **2.4 Determination of quality traits of individual purees**

153 The color of processed purees, put in measuring cells, was determined using the
154 same method as for apples (described in **part 2.2**).

155 The viscosity of the purees was carried out using a Physica MCR-301 controlled
156 stress rheometer (Anton Paar, Graz, Austria) equipped with a Peltier cell (CPTD-200)
157 and a 6-vane geometry (FL100/6W) with a gap of 3.46 mm, at 22.5°C. The
158 measurements were performed after a pre-shearing period of 1 min at a shear rate of 50
159 s^{-1} , followed by 3 min at rest (Lan, Bureau, Chen, Leca, Renard, & Jaillais, 2021). The
160 values of viscosity at 50 s^{-1} and 100 s^{-1} (η_{50} and η_{100} respectively) were taken as
161 indicators of puree viscosity, which are considered representative of the mouth sensory
162 characteristics during consumption (Chen & Engelen, 2012).

163 For all purees, soluble solids content (SSC), titratable acidity (TA), pH and dry
164 matter content (DMC) were characterized based on our previous study (Lan, Jaillais,

165 **Leca, Renard, & Bureau, 2020**). SSC was determined with a digital refractometer (PR-
166 101 ATAGO, Norfolk, VA, USA) and expressed in °Brix at 22.5°C. TA was determined
167 by titration up to pH 8.1 with 0.1 mol/L NaOH and expressed in mmol H⁺/kg of fresh
168 weight (FW) using an autotitrator (Methrom, Herisau, Switzerland). The pH values
169 were characterized using a pH meter (FE-20, Mettler-Toledo, China). DMC was
170 estimated from the weight of freeze-dried samples upon reaching a constant weight by
171 a freeze-drying machine (Cryonext, Saint Aunes, France) after 5 days. **These**
172 **measurements were performed with three replicates.**

173 **2.5 Spectrum acquisition on individual purees**

174 VIS-NIR spectral data of processed purees were acquired using the same
175 conditions as for apples (described in **part 2.2**). Each sample was transferred into a 10
176 mL glass vial (5 cm height × 18 mm diameter) which was placed on the automated
177 sample wheel of the spectrophotometer. Each puree sample was randomly measured
178 three times on different aliquots and the averaged spectrum was calculated for data
179 treatment and chemometrics. The mean spectra of three replicates of each puree were
180 used for further analysis. A reference background measurement was automatically
181 activated before each data set acquisition using an internal Spectralon reference. Finally,
182 the 120 VIS-NIR spectra of processed purees were obtained and correspond one by one
183 to the spectra of raw individual apples.

184 **2.6 Statistical analyses and chemometrics**

185 After checking the normal distribution of the reference data, T-test analysis was
186 carried out to determine the significant differences between varieties considering them
187 two by two (**Fig. 2**) using R software (version 4.0.2) (R Core Team, 2019) with the
188 package of ‘ggpubr’ (Kassambara, 2020). The significant results (*p*-values) were
189 displayed as ‘*ns*’ (*p*-values > 0.05), ‘*’ (*p*-values ≤ 0.05), ‘**’ (*p*-values ≤ 0.01), ‘***’
190 (*p*-values ≤ 0.001) and ‘****’ (*p*-values ≤ 0.0001), respectively. Pearson correlation
191 analysis was performed between the color parameters (L* a* b*) of apples and the
192 different quality traits of their corresponding processed purees using XLSTAT (version

193 2018.5.52037, Addinsoft SARL, Paris, France) data analysis toolbox.

194 Spectral pre-processing and multivariate data analysis were performed with Matlab
195 7.5 (Mathworks Inc. Natick, MA, USA) software using the SAISIR package (Cordella
196 & Bertrand, 2014). Particularly, the VIS-NIR spectra of apples and corresponding
197 purees from 500-2500 nm were preprocessed with several strategies, including
198 smoothing with a window size of 23 variables, standard normal variate (SNV) and the
199 first derivative Savitzky–Golay transformation with the 11 gap sizes. The two-
200 dimensional correlation spectroscopy method (2D-COS) was used to investigate the
201 spectral correlations between raw apples and purees (Noda, 1993).

202 The partial least square (PLS), support vector machine (SVM) and random forest
203 (RF) models were built using R software (version 4.0.2) (R Core Team, 2019) with
204 several packages, including ‘prospectr’ (Stevens & Ramirez-Lopez, 2013), ‘pls’ (Mevik,
205 Wehrens, & Liland, 2011), ‘kernlab’ (Karatzoglou, Smola, Hornik, & Zeileis, 2004),
206 ‘caret’ (Kuhn, 2015) and ‘Boruta’ (Kursa & Rudnicki, 2010). The whole VIS-NIR
207 spectra dataset included 120 spectra of individual apples (4 varieties × 10 apples × 3
208 weeks) and the 120 corresponding puree spectra. The dataset was split using stratified
209 random sampling as follows: two-thirds of the spectral dataset from each variety (4
210 variety × 20 spectra of apples and their related cooked purees) were used for calibration
211 and one-third of the spectral dataset (4 variety × 20 spectra of apples and their related
212 cooked purees) for validation. The procedure was repeated 10 times with the different
213 sets of calibration and validation, and the model performance was described by the
214 averaged values of the determination coefficients of validation (R_v^2), of the root mean
215 square errors of validation (RMSEV), of the numbers of latent variables (LVs) for PLS
216 models and of the residual predictive derivation (RPD) values as described by Nicolai
217 et al. (2007).

218 **3. Results and discussion**

219 **3.1 Effect of the inter- and intra- variability of apples on the corresponding cooked** 220 **purees**

221 In this study, both the inter-variability of apple cultivars and the intra-variability of

222 individual apples affected the physical (L^* , a^* , b^*), biochemical (SSC, DMC, TA, pH)
223 and viscosity (η_{50} and η_{100}) properties of corresponding purees (**Fig. 2** and **Fig. S1**).

224 3.1.1 Color parameters of apples and purees

225 For color parameters, inter-variability was observed according to the four different
226 apple varieties on redness (a^* values) and yellowness (b^* values) of their processed
227 purees (**Fig. S1**). Both for apples and purees, significantly ($p < 0.0001$) higher redness
228 and lower yellowness were characterized in GA and BR than in GD and GS. GD apples
229 and their cooked purees had the highest ($p < 0.0001$) yellowness among the four puree
230 varieties. Moreover, a larger intra-variability of color parameters observed in the set of
231 the 30 different BR ($a^* = 11.2 \pm 10.9$, $b^* = 33.1 \pm 8.3$) and 30 GA ($a^* = 19.6 \pm 14.0$, b^*
232 $= 33.7 \pm 7.0$) apples resulted in a more intensive variation of the redness and yellowness
233 in their corresponding purees (in **Fig. 2**) than in the 30 GD ($a^* = -7.0 \pm 3.9$, $b^* = 47.2$
234 ± 2.3) and 30 GS ($a^* = -15.0 \pm 4.6$, $b^* = 43.6 \pm 2.5$) apples.

235 Briefly, the variation of color properties of cooked purees came from both, the
236 inter- and intra- variability of individual apples. It can be assessed based on the good
237 correlation of redness ($R^2 = 0.70$) and yellowness ($R^2 = 0.58$) between apples and purees.

238 3.1.2 Viscosity of purees

239 Concerning the inter-variability due to varieties on puree rheological properties,
240 BR and GS purees presented a significant ($p < 0.0001$) higher viscosity (η_{50} and η_{100})
241 than GA and GD purees. BR and GS purees were described to have a bigger particle
242 size and a promoted cell adhesion with more branched pectins than GA and GD
243 involving probably their higher viscosity (Buergy, Rolland-Sabaté, Leca, & Renard,
244 2020; Buergy, Rolland-Sabaté, Leca, & Renard, 2021). Moreover, the viscosity at the
245 share rate of 50 s^{-1} (η_{50}) was similar ($p > 0.05$) in GD and GA purees. This result was
246 different from our previous one giving a higher viscosity of GD than of GA purees (Lan,
247 Bureau, Chen, Leca, Renard, & Jaillais, 2021). This could be due to the different levels
248 of enzyme inactivation such as pectin methyl-esterase (PME) during apple processing,
249 between microwave processing used in this study and the conventional thermal cooking

250 used previously (Arjmandi, Otón, Artés, Artés-Hernández, Gómez, & Aguayo, 2017).
251 It also could be due to the different apple compositions harvested from two different
252 years in France (2019 for (Lan, Bureau, Chen, Leca, Renard, & Jaillais, 2021) and 2020
253 for this study). The processing conditions provide indeed different kinds of puree
254 viscosity directly in relation to varieties (Dale, Okos, & Nelson, 1982).

255 The intra-variability of puree viscosity (η_{50} and η_{100}) in GS and BR apple sets
256 presented a larger variation than in GA and GD sets (**Fig. 2**). This intra-variability of
257 puree viscosity was not directly related to the appearance of the raw apples. Indeed, for
258 the two kinds of BR apples (the averaged a^* values of 10 apples for each sets), the more
259 ($a^* = 12.2 \pm 6.2$) or less apple redness ($a^* = 9.6 \pm 5.9$) gave a different puree viscosity
260 ($\eta_{50} = 2.36 \pm 0.12$ Pa.s and $\eta_{50} = 1.57 \pm 0.18$ Pa.s). However, this was not the case for
261 the two GA apple sets with different redness ($a^* = 27.2 \pm 5.0$ and $a^* = 14.2 \pm 2.3$)
262 resulting in a similar puree viscosity of $\eta_{50} = 1.26 \pm 0.18$ Pa.s and 1.39 ± 0.30 Pa.s,
263 respectively (**Fig. S1**).

264 Thus, both, inter-variability of apple varieties (BR and GS > GA and GD) and the
265 intra-variability of individual apples (especially for individual GS and BR apples)
266 generated a wide range of puree viscosity. The color properties of single apples will not
267 allow anticipating the viscosity of cooked purees.

268 3.1.3 Biochemical compositions of purees

269 The significant inter-variability ($p < 0.05$) was observed for SSC between the four
270 puree varieties, except between BR and GS purees ($p > 0.05$) (**Fig. 2**). Clearly,
271 individual GD apples introduced the largest intra-variability of SSC in cooked purees
272 compared to the other three varieties. Interestingly, the a^* values of fresh GD apples
273 were positively correlated to the SSC ($R^2 = 0.57$) of their corresponding purees. In
274 addition, the inter-variations of DMC were significantly different ($p < 0.05$) among BR,
275 GA and GD purees, but not between GS purees and these three groups (BR, GA, and
276 GD purees). This result can be explained by the large intra-variability of DMC in GS
277 purees (**Fig. 2**). A significant difference ($p < 0.001$) was observed also for TA and pH
278 among the four puree varieties. For TA, the inter-variability was ranked as GS > BR >

279 GD > GA and it was the contrary for pH, as expected. However, the intra-variability of
280 TA was different and in the following order: BR > GS > GD > GA.

281 Consequently, both, inter and intra-variability of apples induced variations of SSC,
282 DMC, TA and pH in the cooked purees. With a first insight of individual apple
283 processing, the large intra-variability of GD, GS and BR apples resulted in intensive
284 variations of SSC, DMC and TA in cooked purees, respectively.

285 **3.2 Spectral analysis of apples and purees**

286 *3.2.1 The inter and intra variability of apples and purees measured by VIS-NIRS*

287 After the pre-processing, the VIS-NIR spectra of all individual raw apples and their
288 related cooked purees with the most of variability could be observed at around 500-700
289 nm, 1140 nm, 1386-1392 nm, 1880 nm, 1930-2197 nm and 2250-2450 nm (**Fig. 3a, 3b**
290 **and 3c**). The specific chlorophyll absorption wavebands at 657-665 nm (Khatiwada,
291 Subedi, Hayes, Carlos, & Walsh, 2016) and 682-689 nm (Mehl, Chen, Kim, & Chan,
292 2004) were clearly observed in the visible part of spectra for raw apples and their
293 corresponding purees (**Fig. 3a and 3b**), giving information of the color diversity. The
294 second overtone vibrations of C-H bonds between 1100-1250 nm in both, apples and
295 cooked purees (**Fig. 3c**) presented minor changes, indicating a limited effect of
296 processing on sugar contents (Ma, Li, Inagaki, Yang, & Tsuchikawa, 2018). The
297 important bands at 1386-1392 nm, 1880 nm, 1930-2197 nm, and 2250-2450 nm were
298 mainly explained by the combination bands of O-H bonds of waters and C-H bonds of
299 sugars and organic acids in apples (Camps, Guillermin, Mauget, & Bertrand, 2007;
300 Kemps, Leon, Best, De Baerdemaeker, & De Ketelaere, 2010). Compared to raw apples,
301 the lower absorption peaks in their corresponding purees could be due to that apple
302 cooking attenuated the variation of water contents in the processed purees (Lan, Renard,
303 Jaillais, Leca, & Bureau, 2020). Generally, the spectral variability was clearly higher in
304 apples than in their corresponding purees, as already observed in purees prepared from
305 4 kg of apples (Lan, Renard, Jaillais, Leca, & Bureau, 2020). Besides the possible effect
306 of processing, this difference between apples and purees was probably due to the sample
307 structure (solid fruit and liquid purees), affecting the diffuse reflectance.

3.2.2 Correlations between the VIS-NIR spectra of apples and purees

2D-COS was performed on all smoothed and SNV pre-treated VIS (500-780 nm) (Fig. 4a) and NIR (800-2500 nm) spectra (Fig. 4b) to point out the highly correlated wavelengths between apples and their processed purees. The correlations were much higher in VIS than in NIR ranges. Particularly in the VIS range (Fig. 4a), a clear positive relationship was obtained from 665 nm to 685 nm, thus confirming a strong color correlation between apples and purees. It was also in line with the colorimetric measurements previously described (Part 3.1).

In the NIR region, the wavelengths around 1125-1400 nm, 1850-2150 nm, and 2250-2450 nm in apples (X-axis) were positively correlated to the corresponding spectral areas at the same wavelengths of the purees (Y-Axis). The positive correlations at 1125-1400 nm may be due to the high correlation of SSC between apples and purees, whereas the 1850-2150 nm and 2250-2450 nm could be due to the water and major soluble matters (sugars and organic acids) which varied in the same way in apples and purees. Reversely, the wavelengths between 1125-1400 nm in apples were negatively correlated with the spectral regions at 1850-2150 nm and 2250-2450 nm in cooked purees. Some possible reasons might be: i) the decrease of water content during cooking while limited changes of biochemical compounds such as SSC between apples and purees, so their corresponding wavelengths were negatively correlated, or ii) the wavelengths specific to sugars in purees were negatively correlated with those to water in the corresponding apples.

As mentioned in our previous work (Lan, Renard, Jaillais, Leca, & Bureau, 2020), a strong relationship of physical and biochemical properties is observed between raw and processed apples and allows us to predict the puree properties from the spectra collected on apples. The observed spectral correlations in this study, considering the large inter-variability with varieties and their intra-variability with individual apples, support these previous results.

3.3 Prediction of puree quality traits

PLS, SVM and RF models were built to predict color, viscosity and biochemical

337 characteristics of apple purees using VIS-NIR spectra acquired on purees (in **Table 1**),
338 or on the corresponding individual raw apples (in **Table 2**).

339 *3.3.1 Prediction of puree characteristics using spectra of purees*

340 Both, the linear (PLS) and non-linear (SVM and RF) regressions of puree
341 rheological parameters (η_{50} and η_{100}) did not give satisfactory predictions ($R_v^2 < 0.46$,
342 RPD < 1.4). These results were in agreement with the poor PLS predictions of puree
343 viscosity at the shear rate of 100 s^{-1} (η_{100}) using VIS-NIR (500-2500 nm) ($R_v^2 = 0.35$,
344 RPD = 1.2) and NIR (800-2500 nm) techniques ($R_v^2 = 0.39$, RPD = 1.3) (Lan, Renard,
345 Jaillais, Leca, & Bureau, 2020). However, they were much lower than the acceptable
346 VIS-NIR predictions obtained in a previous experiment consisting in studying the
347 preparation of apple puree mixtures with different proportions of variety ($R_v^2 > 0.73$,
348 RPD > 1.9) (Lan, Bureau, Chen, Leca, Renard, & Jaillais, 2021) and presenting less than
349 half as much variability (the SD of η_{50} of 0.12 Pa.s lower than the SD of η_{50} of 0.36 Pa.s
350 in this study). Thus, the VIS-NIR or NIR techniques cannot provide acceptable
351 estimations of the puree viscosity, considering both a large inter- and intra-variability
352 of raw apples.

353 For the color parameters, two regression methods, PLS and RF, gave acceptable
354 predictions of L^* , a^* and b^* values, with RPD values reaching 2.0, 2.7 and 2.3,
355 respectively. PLS slightly improved the a^* prediction ($R_v^2 = 0.86$, RMSEV = 0.46, RPD
356 = 2.7) in comparison with RF ($R_v^2 = 0.85$, RMSEV = 0.49, RPD = 2.6), using 514 nm,
357 524 nm and 672 nm as the most contributing wavelengths, all in the visible range. These
358 same wavelengths are already identified in apple purees (Lan, Bureau, Chen, Leca,
359 Renard, & Jaillais, 2021), corresponding probably to the carotenoids (Wang, Wang,
360 Chen, & Han, 2017), anthocyanins (de Brito, de Araújo, Lin, & Harnly, 2007) and
361 chlorophylls (Khatiwada, Subedi, Hayes, Carlos, & Walsh, 2016) in fruits. The specific
362 peak at 672 nm was the major contributor to predict yellowness (b^*) values. A relatively
363 large puree variability for b^* (SD = 4.1) compared to our previous study (SD = 1.7)
364 significantly improved the VIS-NIR prediction results, with RPD values from 1.5 to 2.2
365 (Lan, Renard, Jaillais, Leca, & Bureau, 2020). When variability was large enough,

366 prediction of color parameters by VIS-NIR was possible.

367 For the biochemical parameters, VIS-NIR coupled with PLS regression provided a
368 better prediction of DMC, SSC, TA and pH than the SVM and RF ones (**Table 1**).
369 Particularly, a good prediction of SSC ($R_v^2 = 0.80$, RMSEV = 0.6, RPD = 2.3) was
370 obtained based on the dominant wavelengths corresponding to the absorptions of
371 carbohydrates between 1150-1400 nm, as already described in **Part 3.2**. Although SSC
372 and DMC were highly correlated ($R^2 = 0.65$) in apple purees, VIS-NIR coupled with
373 PLS models showed a better ability to predict SSC than DMC ($R_v^2 = 0.73$, RMSEV =
374 0.01, RPD = 1.9), as previously observed (Lan, Renard, Jaillais, Leca, & Bureau, 2020).
375 However, these predictions of SSC and DMC were relatively lower than our previous
376 prediction by NIR of SSC ($R_v^2 = 0.92$, RPD = 3.1) and DMC ($R_v^2 = 0.85$, RPD = 2.4).
377 The main reason was probably related to the lower variations of SSC and DMC in these
378 four purees varieties at one date (SD of SSC = 1.4 °Brix and of DMC = 0.01 g/g) in
379 comparison with the previous study including two varieties at different dates during a
380 six months cold storage (SD of SSC = 2.1 °Brix and DMC = 0.02 g/g) (Lan, Renard,
381 Jaillais, Leca, & Bureau, 2020). Considering the different expressions of apple puree
382 acidity, TA and pH, VIS-NIR coupled with PLS provided their excellent predictions
383 with $R_v^2 > 0.89$ and RPD > 3.1. Additionally, VIS-NIR gave a better prediction of puree
384 acidity (TA) than NIR in apple purees (Lan, Renard, Jaillais, Leca, & Bureau, 2020),
385 presenting similar ranges of variations with SD values of 21.0 mmol H⁺/kg and 20.2
386 mmol H⁺/kg, respectively. The specific visible wavelengths at around 672 nm were one
387 of the main contributors for the prediction of puree acidity. Consequently, prediction of
388 SSC and DMC in purees needs enough intra-variability from individual apples and
389 inter-variability from both different fruit varieties and experimental conditions
390 (varieties, cold storage periods of raw fruits) to be acceptable by VIS-NIR. However,
391 for acidity, VIS-NIR models integrating the variability of different apples and varieties
392 were enough to give an excellent estimation of TA and pH.

393 *3.3.2 Prediction of puree characteristics using spectra of intact apples*

394 Based on the strong internal VIS-NIR spectral correlations between apples and

395 purees (**Part 3.2.2**), good predictions ($R_v^2 > 0.78$, $RPD > 2.1$) of puree viscosity (η_{50}
396 and η_{100}), a^* , b^* , SSC, TA and pH were obtained using the VIS-NIR spectra of their
397 corresponding individual apples (**Table 2**). Particularly, PLS models provided the best
398 predictions of η_{50} , η_{100} , b^* , SSC and TA than SVM and RF ones. Compare to the PLS
399 models developed from puree spectra (**Table 1**), more spectral latent variables (LVs
400 from 5-11) were required using the VIS-NIR spectra of apples.

401 What stands out in these results was the much better PLS predictions of puree
402 viscosity (η_{50} and η_{100}) using the spectra of apples ($R_v^2 > 0.81$, $RPD > 2.3$) than the
403 spectra of purees directly ($R_v^2 < 0.46$, $RPD < 1.4$). Particularly, specific wavelengths at
404 around 578 nm, 678 nm, 810-835 nm, 1410-1498 nm, 1880 nm and 1940 nm highly
405 contributed to the PLS predictions of puree viscosity, which were located in the spectra
406 regions presenting strong correlations between apples and purees (**Fig. 4**). This result
407 was in line with our previous study, which used the averaged NIR spectra of a set of
408 apples to predict the viscosity (η_{100}) of their related one cooked puree (Lan, Renard,
409 Jaillais, Leca, & Bureau, 2020). A possible explanation might come from the
410 characteristics of purees, resulting from soft and deformable insoluble particles (pulp)
411 in an aqueous medium (serum) (Rao, Thomas, & Javalgi, 1992), that prevent from an
412 efficient light diffusion in comparison with the structure of intact apples that favors the
413 light diffusion and a good signal to noise ratio. Thus, it is possible to hypothesize that
414 VIS-NIRS applied on raw apples give an acceptable prediction of the viscosity of
415 cooked purees.

416 For puree color parameters, VIS-NIR spectra of individual apples coupled with
417 PLS and RF regressions provided an acceptable prediction of redness (a^* value) ($R_v^2 >$
418 0.77 , $RPD > 2.1$) and yellowness (b^* values) ($R_v^2 = 0.79$, $RPD > 2.2$) in corresponding
419 purees, but not of lightness (L^* values) ($R_v^2 < 0.59$, $RPD < 1.5$). The major wavelengths
420 contributing to these models were highly consistent with the models developed using
421 the puree spectra, such as 514 nm and 672 nm. Besides, these good predictions were
422 also in line with their strong internal correlations between apples and purees (**Part 3.1**).
423 However, these good results need to be interpreted with caution because they concerned
424 the only microwave cooked apples and not the conventional thermal processing (at

425 laboratory scale) in probable relation with a rapid inactivation of enzymes by
426 microwaves and so, a limitation of apple oxidization and color change (Picouet et al.,
427 2009).

428 For puree biochemical parameters, PLS regression models had a good ability to
429 estimate SSC, TA, and pH of all purees with acceptable R_v^2 (> 0.78) and RPD (> 2.1),
430 but not DMC ($R_v^2 < 0.71$, RPD < 1.8). Particularly, the SSC prediction in purees was
431 based on the specific wavelengths at 950 nm, 1150 nm, 1400 nm and 1880 nm,
432 corresponding to the sugars and water variations (**Part 3.2**). Impressively, for acidity,
433 both, TA and pH of cooked purees were excellently predicted using the VIS-NIR spectra
434 of related apples, giving RPD values of 2.8 and 2.5, respectively. These results were
435 better than our previous predictions of puree TA using the apple NIR spectra and giving
436 RPD around 2.1-2.3 (Lan, Renard, Jaillais, Leca, & Bureau, 2020). Indeed, some
437 specific peaks in the visible region at 524 nm and 672 nm contributed to the better
438 prediction of puree TA. It can thus be suggested that integrating the visible range of
439 spectra acquired on apples provided a better prediction of puree acidity than just using
440 the NIR range. Concerning DMC, its bad prediction might be explained by the lower
441 variations, here, compared to our previous work (Lan, Renard, Jaillais, Leca, & Bureau,
442 2020), and probably not by a limited potential of the VIS-NIR range.

443 Accordingly, VIS-NIR spectra acquired on raw apples could give satisfactory
444 predictions of color (a^* and b^* values), viscosity (η_{50} and η_{100}), SSC, TA and pH of the
445 individual cooked purees using both, PLS or RF regressions.

446 **4. Conclusion**

447 This study was designed applying the absolute definition of ‘one apple to one puree’,
448 which gave a first insight into the impacts of fruit inter- or intra-variability during
449 processing, from the spectroscopic point of view. Importantly, the intra-variability in
450 fruits introduced intensive changes of visual aspects, chemical and textural properties
451 of their corresponding microwave-cooked purees. Taking into account the variability of

452 fruit varieties and intra-variations between apples in each one, could improve the
453 prediction accuracy of regression models. **Notably, our results show that simple non-**
454 **destructive measurement using near infrared spectroscopy applied on apples can**
455 **provide to processors of apple industry the basic information on the inter- and intra-**
456 **variability of raw materials and help them to determine the best blend of apples in order**
457 **to obtain always the same final products such as puree.** Further, strong correlations
458 while apple processing obtained from spectral data provided further evidence on such
459 the indirect predictions of color parameters, viscosity and biochemical parameters (SSC,
460 TA and pH) of purees from the non-destructive spectral information acquired on raw
461 apples. **Therefore, by systematically scanning all apples, the ~~obtained is-could-provide~~**
462 **objective data on apple quality traits ~~should help~~ to 1) better manage apples according**
463 **to their quality, 2) predict final product characteristics and 3) reduce fruit wastes at the**
464 **end.**

465 **Future work will be needed to identify the quantitative parameters to describe how**
466 **much change of both, apple inter- and intra-variability according to the processing**
467 **conditions (temperature, time, grinding, oxygen and so on).**

468 **Acknowledgements**

469 The authors thank Patrice Reling, Romain Bott, Barbara Gouble (INRAE, SQPOV
470 unit) for their technical help. The ‘Interfaces’ project is an Agropolis Fondation project
471 publicly funded through the ANR (French Research Agency) under “Investissements
472 d’Avenir” programme (ANR-10-LABX-01-001 Labex Agro, coordinated by Agropolis
473 Fondation). Weijie Lan was supported by a doctoral grant from Chinese Scholarship
474 Council.

475

476 **References:**

- 477 Arjmandi, M., Otón, M., Artés, F., Artés-Hernández, F., Gómez, P. A., & Aguayo, E.
478 (2017). Microwave flow and conventional heating effects on the
479 physicochemical properties, bioactive compounds and enzymatic activity of
480 tomato puree. *Journal of the Science of Food and Agriculture*, 97(3), 984-990.
481 <https://doi.org/https://doi.org/10.1002/jsfa.7824>.
- 482 Buergy, A., Rolland-Sabaté, A., Leca, A., & Renard, C. M. G. C. (2020). Pectin
483 modifications in raw fruits alter texture of plant cell dispersions. *Food*
484 *Hydrocolloids*, 107, 105962.
485 <https://doi.org/https://doi.org/10.1016/j.foodhyd.2020.105962>.
- 486 Buergy, A., Rolland-Sabaté, A., Leca, A., & Renard, C. M. G. C. (2021). Apple puree's
487 texture is independent from fruit firmness. *LWT - Food Science and Technology*,
488 145, 111324. <https://doi.org/https://doi.org/10.1016/j.lwt.2021.111324>.
- 489 Camps, C., Guillermin, P., Mauget, J. C., & Bertrand, D. (2007). Discrimination of
490 storage duration of apples stored in a cooled room and shelf-life by visible-near
491 infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, 15(3), 169-177.
- 492 Chen, J., & Engelen, L. (2012). *Food oral processing: fundamentals of eating and*
493 *sensory perception*: John Wiley & Sons.
- 494 Cordella, C. B. Y., & Bertrand, D. (2014). SAISIR: A new general chemometric toolbox.
495 *TRAC Trends in Analytical Chemistry*, 54, 75-82.
496 <https://doi.org/https://doi.org/10.1016/j.trac.2013.10.009>.
- 497 Dale, M. C., Okos, M. R., & Nelson, P. (1982). Concentration of Tomato Products:
498 Analysis of Energy Saving Process Alternatives. *Journal of Food Science*, 47(6),
499 1853-1858. <https://doi.org/https://doi.org/10.1111/j.1365-2621.1982.tb12898.x>.
- 500 de Brito, E. S., de Araújo, M. C. P., Lin, L.-Z., & Harnly, J. (2007). Determination of
501 the flavonoid components of cashew apple (*Anacardium occidentale*) by LC-
502 DAD-ESI/MS. *Food Chemistry*, 105(3), 1112-1118.
503 <https://doi.org/10.1016/j.foodchem.2007.02.009>.
- 504 FranceAgriMer. (2017). La Pomme en 2016-2017. Accessed October 2020, from
505 <https://www.rnm.franceagrimer.fr>.
- 506 Guo, Q., Sun, D.-W., Cheng, J.-H., & Han, Z. (2017). Microwave processing techniques
507 and their recent applications in the food industry. *Trends in Food Science &*
508 *Technology*, 67, 236-247.
509 <https://doi.org/https://doi.org/10.1016/j.tifs.2017.07.007>.
- 510 Huang, Y., Lu, R., & Chen, K. (2020). Detection of internal defect of apples by a
511 multichannel Vis/NIR spectroscopic system. *Postharvest Biology and*
512 *Technology*, 161, 111065. <https://doi.org/10.1016/j.postharvbio.2019.111065>
- 513 Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). kernlab-an S4 package
514 for kernel methods in R. *Journal of statistical software*, 11(9), 1-20.
515 <https://doi.org/10.18637/jss.v011.i09>.
- 516 Kassambara, A. (2020). ggpubr: 'ggplot2' Based Publication Ready Plots.
517 <https://CRAN.R-project.org/package=ggpubr>.
- 518 Kemps, B., Leon, L., Best, S., De Baerdemaeker, J., & De Ketelaere, B. (2010).
519 Assessment of the quality parameters in grapes using VIS/NIR spectroscopy.

520 *Biosystems engineering*, 105(4), 507-513.
521 <https://doi.org/10.1016/j.biosystemseng.2010.02.002>

522 Khatiwada, B. P., Subedi, P. P., Hayes, C., Carlos, L. C. C., & Walsh, K. B. (2016).
523 Assessment of internal flesh browning in intact apple using visible-short wave
524 near infrared spectroscopy. *Postharvest Biology and Technology*, 120, 103-111.
525 <https://doi.org/https://doi.org/10.1016/j.postharvbio.2016.06.001>.

526 Kuhn, M. (2015). Caret: classification and regression training. Astrophysics Source
527 Code Library.

528 Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package.
529 *Journal of Statistical Software*, 36(11), 1-13.
530 <http://hdl.handle.net/10.18637/jss.v036.i11>.

531 Lan, W., Bureau, S., Chen, S., Leca, A., Renard, C. M. G. C., & Jaillais, B. (2021).
532 Visible, near- and mid-infrared spectroscopy coupled with an innovative
533 chemometric strategy to control apple puree quality. *Food Control*, 120, 107546.
534 <https://doi.org/https://doi.org/10.1016/j.foodcont.2020.107546>.

535 Lan, W., Jaillais, B., Leca, A., Renard, C. M. G. C., & Bureau, S. (2020). A new
536 application of NIR spectroscopy to describe and predict purees quality from the
537 non-destructive apple measurements. *Food Chemistry*, 310, 125944.
538 <https://doi.org/https://doi.org/10.1016/j.foodchem.2019.125944>.

539 Lan, W., Jaillais, B., Renard, C. M. G. C., Leca, A., Chen, S., Le Bourvellec, C., &
540 Bureau, S. (2021). A method using near infrared hyperspectral imaging to
541 highlight the internal quality of apple fruit slices. *Postharvest Biology and
542 Technology*, 175, 111497.
543 <https://doi.org/https://doi.org/10.1016/j.postharvbio.2021.111497>.

544 Lan, W., Renard, C. M. G. C., Jaillais, B., Buergy, A., Leca, A., Chen, S., & Bureau, S.
545 (2021). Mid-infrared technique to forecast cooked puree properties from raw
546 apples: a potential strategy towards sustainability and precision processing.
547 *Food Chemistry*, 129636.
548 <https://doi.org/https://doi.org/10.1016/j.foodchem.2021.129636>.

549 Lan, W., Renard, C. M. G. C., Jaillais, B., Leca, A., & Bureau, S. (2020). Fresh, freeze-
550 dried or cell wall samples: Which is the most appropriate to determine chemical,
551 structural and rheological variations during apple processing using ATR-FTIR
552 spectroscopy? *Food Chemistry*, 330, 127357.
553 <https://doi.org/https://doi.org/10.1016/j.foodchem.2020.127357>.

554 Ma, T., Li, X., Inagaki, T., Yang, H., & Tsuchikawa, S. (2018). Noncontact evaluation
555 of soluble solids content in apples by near-infrared hyperspectral imaging.
556 *Journal of Food Engineering*, 224, 53-61.
557 <https://doi.org/10.1016/j.jfoodeng.2017.12.028>.

558 Mehl, P. M., Chen, Y. R., Kim, M. S., & Chan, D. E. (2004). Development of
559 hyperspectral imaging technique for the detection of apple surface defects and
560 contaminations. *Journal of food engineering*, 61(1), 67-81.
561 [https://doi.org/10.1016/S0260-8774\(03\)00188-2](https://doi.org/10.1016/S0260-8774(03)00188-2).

562 Mevik, B. H., Wehrens, R., & Liland, K. H. (2011). pls: Partial least squares and
563 principal component regression. R package version, 2(3).

-
- 564 Nicolai, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., &
565 Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable
566 quality by means of NIR spectroscopy: A review. *Postharvest Biology and*
567 *Technology*, 46(2), 99-118.
568 <https://doi.org/https://doi.org/10.1016/j.postharvbio.2007.06.024>.
- 569 Noda, I. (1993). Generalized two-dimensional correlation method applicable to infrared,
570 Raman, and other types of spectroscopy. *Applied Spectroscopy*, 47(9), 1329-
571 1336. <https://www.osapublishing.org/as/abstract.cfm?URI=as-47-9-1329>.
- 572 Oszmiański, J., Wolniak, M., Wojdyło, A., & Wawer, I. (2008). Influence of apple purée
573 preparation and storage on polyphenol contents and antioxidant activity. *Food*
574 *Chemistry*, 107(4), 1473-1484. <https://doi.org/10.1016/j.foodchem.2007.10.003>.
- 575 Picouet, P. A., Landl, A., Abadias, M., Castellari, M., & Viñas, I. (2009). Minimal
576 processing of a Granny Smith apple purée by microwave heating. *Innovative*
577 *Food Science & Emerging Technologies*, 10(4), 545-550.
578 <https://doi.org/10.1016/j.ifset.2009.05.007>
- 579 Pissard, A., Baeten, V., Romnée, J. M., Dupont, P., Mouteau, A., & Lateur, M. (2012).
580 Classical and NIR measurements of the quality and nutritional parameters of
581 apples: a methodological study of intra-fruit variability. *BASE*.
582 <https://doi.org/https://popups.uliege.be:443/1780-4507/index.php?id=8782>.
- 583 R Core Team, R. C. (2019). R: A language and environment for statistical computing.
- 584 Rao, S. R., Thomas, E. G., & Javalgi, R. G. (1992). Activity Preferences and Trip-
585 planning Behavior of the U.S. Outbound Pleasure Travel Market. *Journal of*
586 *Travel Research*, 30(3), 3-12. <https://doi.org/10.1177/004728759203000301>.
- 587 Stevens, A., & Ramirez-Lopez, L. (2013). An introduction to the prospectr packageR
588 package Vignette R package version 0.1. 3. [https://CRAN.R-project.](https://CRAN.R-project.org/package=prospectr)
589 [org/package=prospectr](https://CRAN.R-project.org/package=prospectr).
- 590 Wang, J., Wang, J., Chen, Z., & Han, D. (2017). Development of multi-cultivar models
591 for predicting the soluble solid content and firmness of European pear (*Pyrus*
592 *communis* L.) using portable vis-NIR spectroscopy. *Postharvest Biology and*
593 *Technology*, 129, 143-151.
594 <https://doi.org/https://doi.org/10.1016/j.postharvbio.2017.03.012>.
- 595 Xia, Y., Fan, S., Li, J., Tian, X., Huang, W., & Chen, L. (2020). Optimization and
596 comparison of models for prediction of soluble solids content in apple by online
597 Vis/NIR transmission coupled with diameter correction method. *Chemometrics*
598 *and Intelligent Laboratory Systems*, 201, 104017.
599 <https://doi.org/10.1016/j.chemolab.2020.104017>
- 600

601 **Figure captions:**

602 **Fig. 1.** Experimental design of apple processing, spectral acquisition and quality
603 characterization.

604 **Fig. 2.** The boxplots and the T-test results of color, rheological and biochemical
605 properties of four apple puree varieties. (The significances were displayed as ‘ns’ (p
606 values > 0.05), ‘*’ (p values ≤ 0.05), ‘**’ (p values ≤ 0.01), ‘***’ (p values ≤ 0.001)
607 and ‘****’ (p values ≤ 0.0001)).

608 **Fig. 3.** The pre-processed (smoothing with 13 windows + SNV+ 1st derivation with 11
609 windows) VIS-NIR spectra of (a) individual apples and (b) their related cooked purees,
610 and (c) the averaged pre-processed spectra of all apples (blue line) and cooked purees
611 (red line).

612 **Fig. 4.** The 2D-COS (two-dimensional correlation spectroscopy) plot between the
613 spectra of all individual apples and their related cooked purees in the (a) visible (500-
614 780 nm) and near infrared (800-2500 nm) ranges.

615 **Fig. S1.** The pictures of individual apples and the corresponding microwave cooked
616 purees.

617

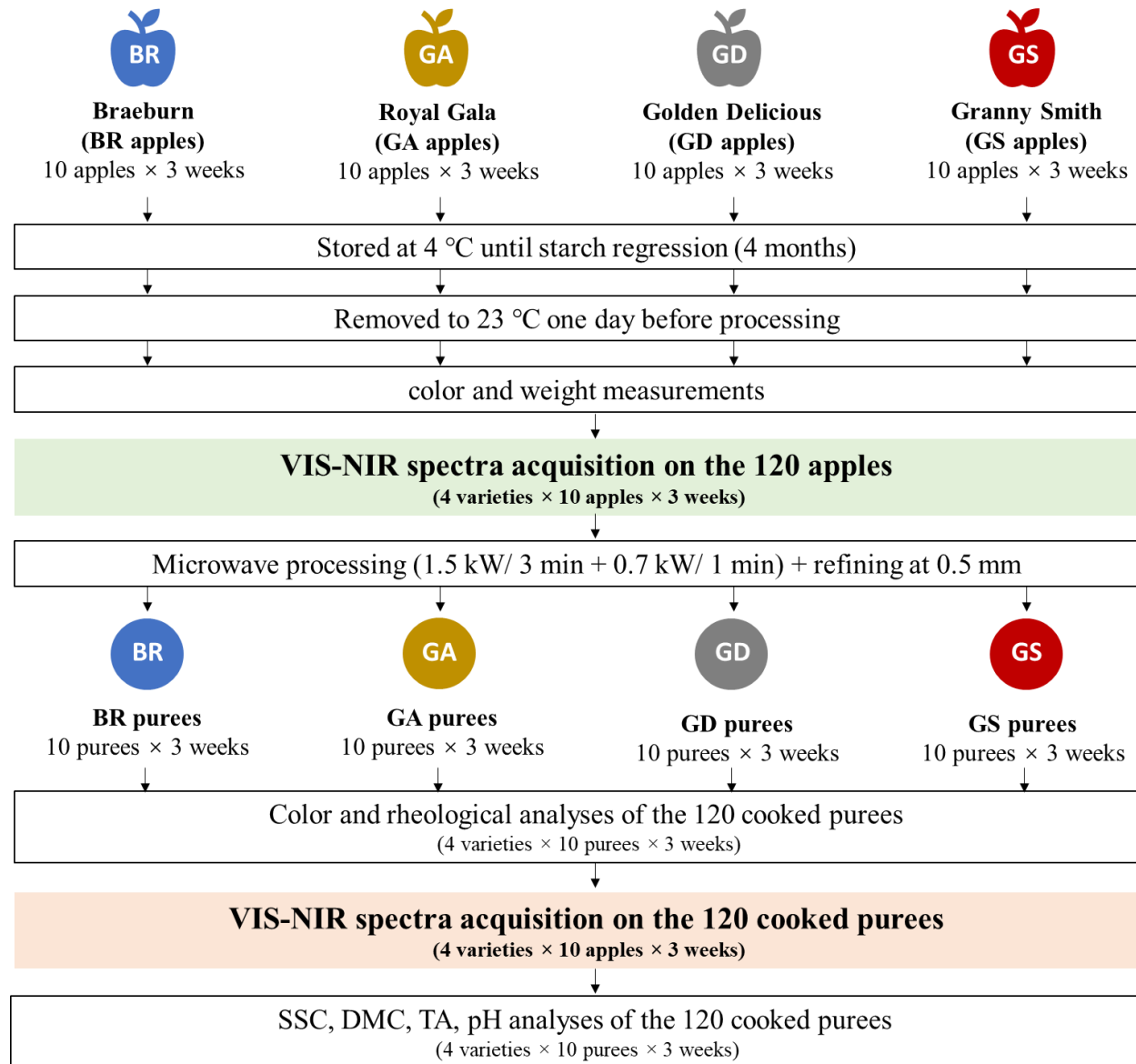


Fig. 1

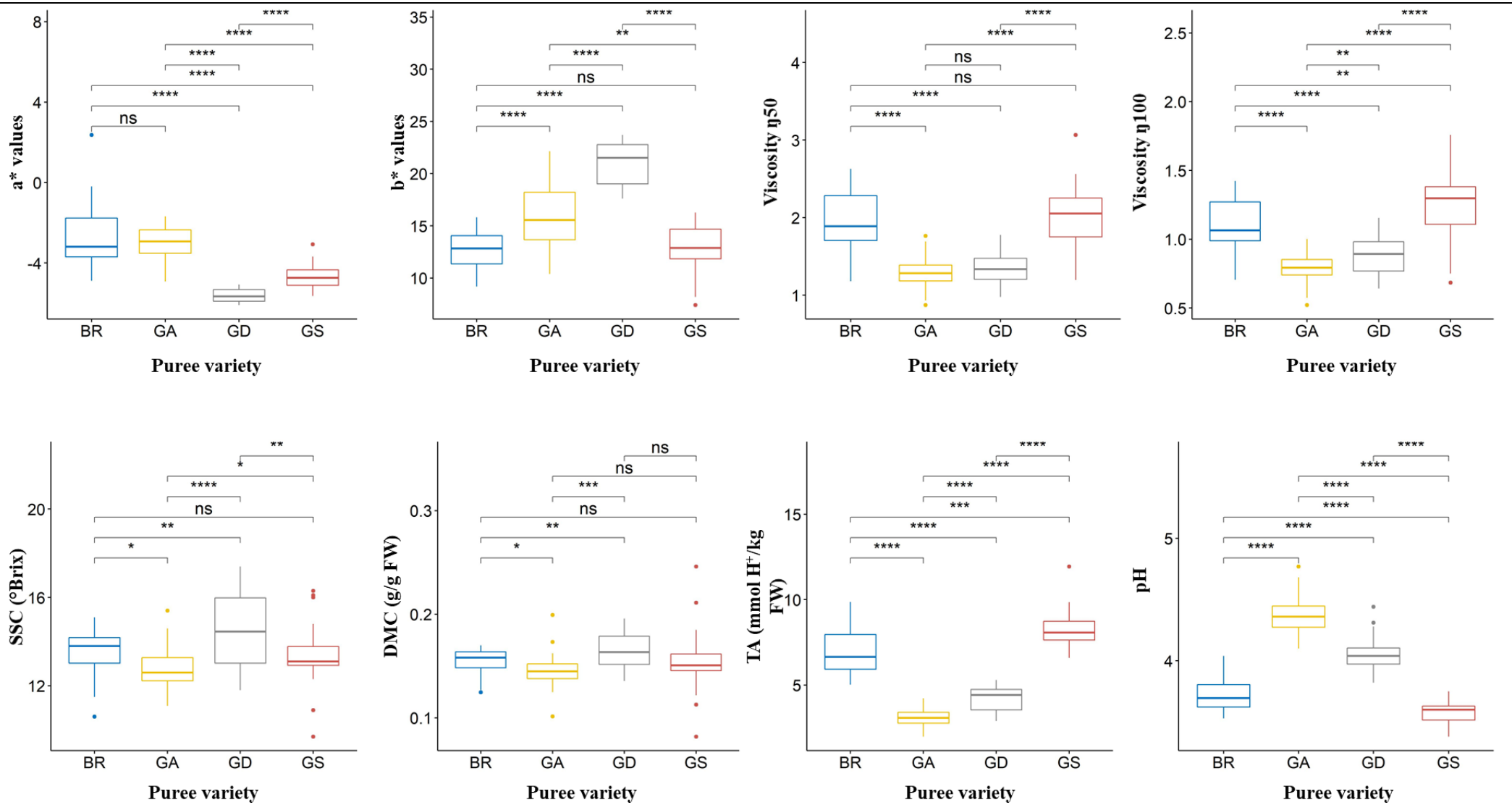


Fig. 2

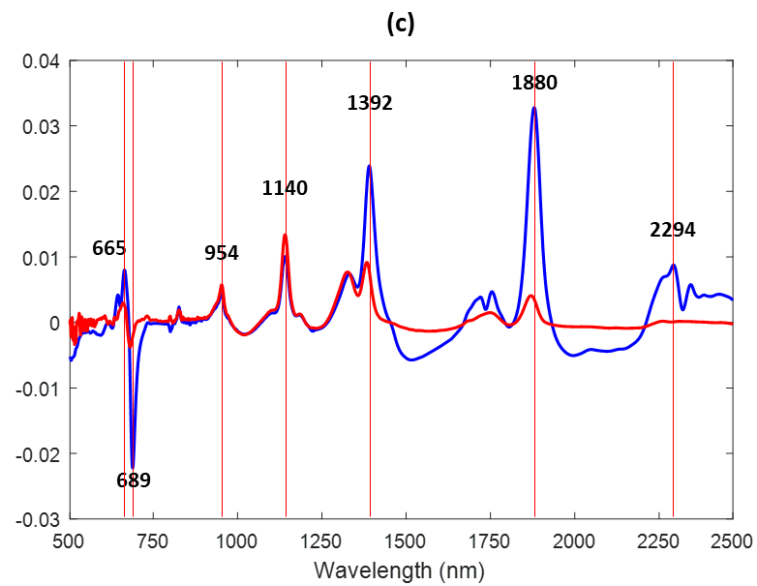
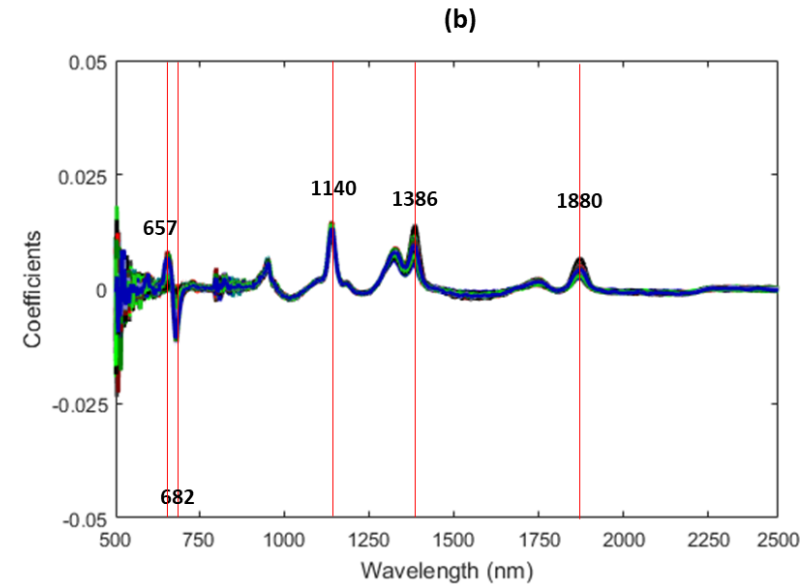
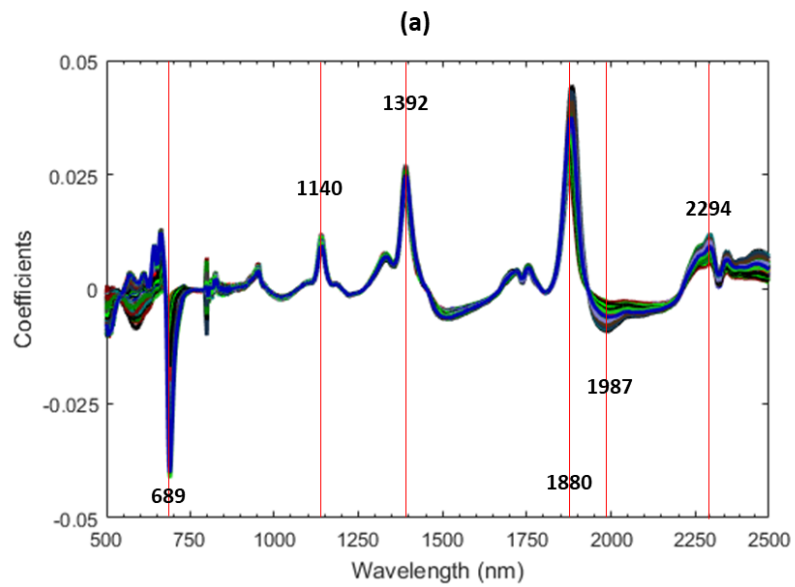


Fig. 3

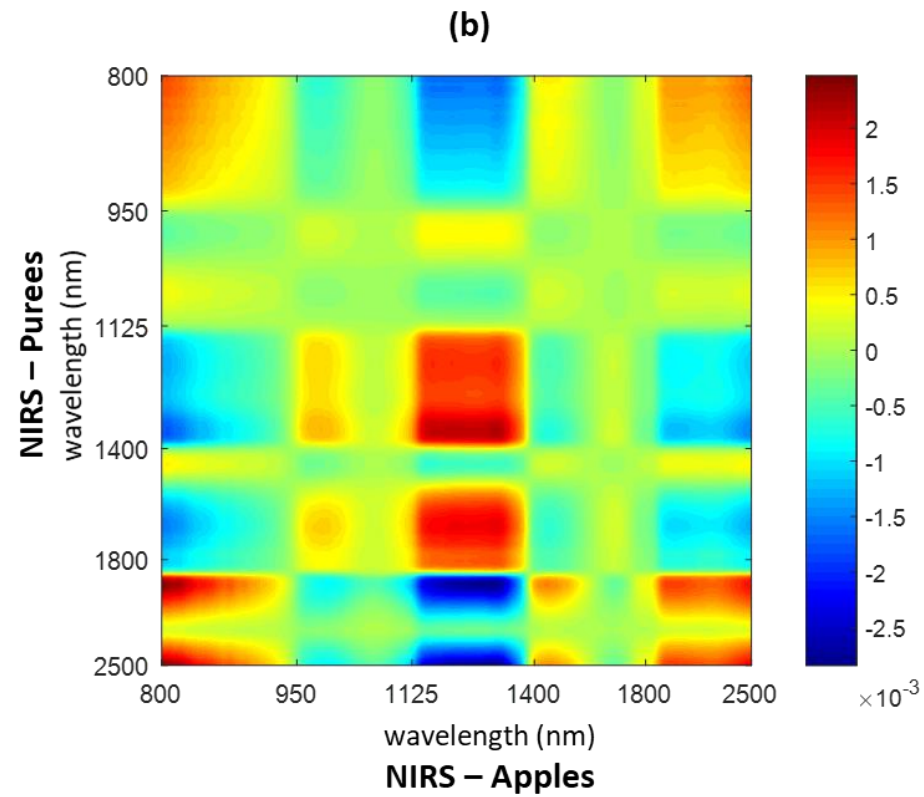
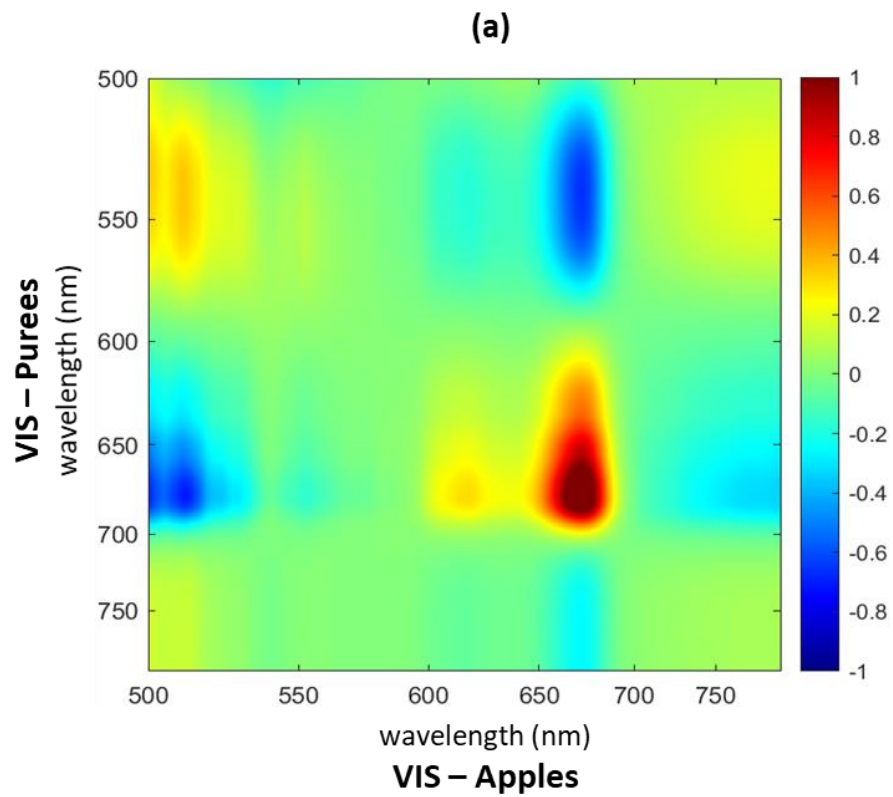


Fig. 4

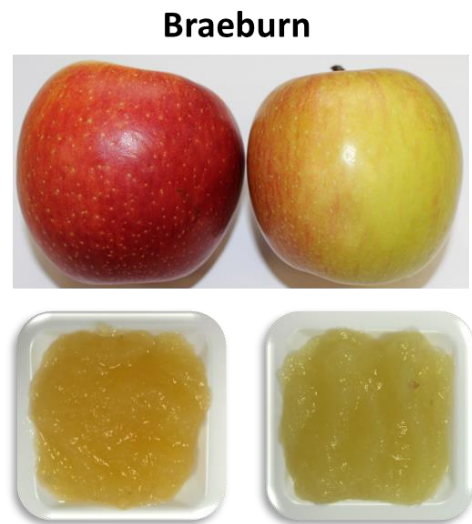
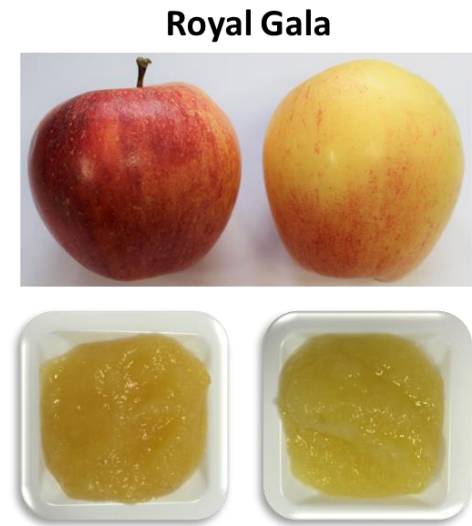
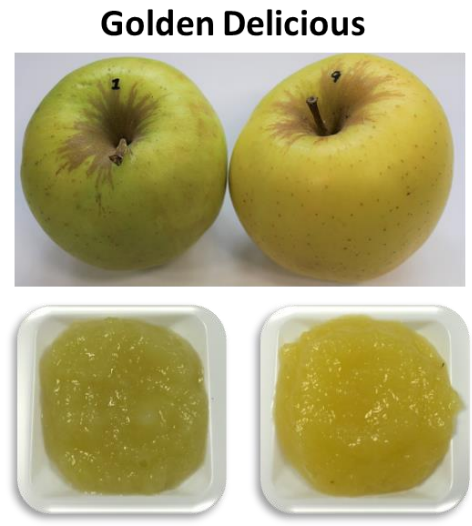


Fig. S1

Table 1. Prediction of puree quality traits from the VIS-NIR spectra of cooked purees

Parameter	Range	SD	R_v^2	PLS-R			SVM-R			RF-R		
				RMSEV	RPD	LVs	R_v^2	RMSEV	RPD	R_v^2	RMSEV	RPD
η_{50}	0.87 - 3.07	0.36	0.34	0.30	1.2	5	0.38	0.29	1.2	0.45	0.26	1.4
η_{100}	0.52 - 1.76	0.20	0.35	0.16	1.2	4	0.36	0.16	1.2	0.46	0.14	1.4
L*	39.0 - 55.2	3.5	0.75	1.7	2.0	4	0.50	2.5	1.3	0.73	1.7	2.0
a*	(-6.1) - 2.4	1.5	0.86	0.5	2.7	6	0.74	0.8	1.9	0.85	0.5	2.6
b*	7.4 - 23.7	4.1	0.81	1.8	2.3	5	0.68	2.6	1.6	0.81	1.8	2.3
DMC (g/g FW)	0.08 - 0.25	0.01	0.73	0.01	1.9	9	0.57	0.01	1.4	0.56	0.01	1.4
SSC (°Brix)	9.7 - 17.4	1.4	0.80	0.6	2.3	8	0.64	1.0	1.5	0.69	0.9	1.5
TA (mmol H ⁺ /kg FW)	19.8- 119.4	21.0	0.89	0.6	3.1	9	0.65	1.5	1.4	0.80	1.0	2.1
pH	3.4 - 4.8	0.3	0.90	0.1	3.3	10	0.65	0.2	1.4	0.83	0.1	2.3

Notes: All regression models based on the smoothed (13 windows) and SNV pre-treated VIS-NIR spectra of purees at 500-2500 nm. PLS-R: partial least square regression; SVM-R: support vector machine regression; RF-R: random forest regression. Totally, 120 puree spectra and reference data from four varieties ('Golden Delicious', 'Braeburn', 'Granny Smith' and 'Royal Gala'). The averaged results of 10 times random calibration (80 samples) and validation (40 samples) tests. R_v^2 : determination coefficient of the validation test; RMSEv: root mean square error of validation test; RPD: the residual predictive deviation of validation test, LVs: the optimal numbers of latent variables.

37

Table 2. Prediction of puree quality traits from the VIS-NIR spectra of corresponding raw apples

Parameter	Range	SD	R_v^2	PLS-R			R_v^2	SVM-R		R_v^2	RF-R	
				RMSEV	RPD	LVs		RMSEV	RPD		RMSEV	RPD
η_{50}	0.87 - 3.07	0.36	0.81	0.15	2.3	8	0.73	0.19	1.9	0.65	0.21	1.7
η_{100}	0.52 - 1.76	0.20	0.85	0.07	2.6	10	0.75	0.10	2.0	0.68	0.11	1.8
L*	39.0 - 55.2	3.5	0.59	2.1	1.6	4	0.53	2.3	1.5	0.58	2.2	1.6
a*	(-6.1) - 2.4	1.5	0.84	0.7	2.5	5	0.67	1.0	1.8	0.81	0.8	2.3
b*	7.4 - 23.7	4.1	0.79	1.9	2.2	7	0.61	2.3	1.8	0.59	1.8	2.3
DMC (g/g FW)	0.08 - 0.25	0.01	0.71	0.01	1.8	11	0.59	0.01	1.4	0.57	0.01	1.3
SSC (°Brix)	9.7 - 17.4	1.4	0.78	0.7	2.1	9	0.60	1.2	1.3	0.59	1.2	1.3
TA (mmol H ⁺ /kg FW)	19.8- 119.4	21.0	0.87	0.8	2.8	10	0.78	1.0	2.1	0.83	0.9	2.5
pH	3.4 - 4.8	0.3	0.84	0.1	2.5	11	0.78	0.2	2.1	0.84	0.1	2.5

38

Notes: All regression models based on the smoothed (13 windows) and SNV pre-treated VIS-NIR spectra of apples at 500-2500 nm. PLS-R: partial least square regression; SVM-R: support vector machine regression; RF-R: random forest regression. Totally, 120 spectra of raw apples and their reference data of cooked purees from four varieties ('Golden Delicious', 'Braeburn', 'Granny Smith' and 'Royal Gala'). The averaged results of 10 times random calibration (80 samples) and validation (40 samples) tests. R_v^2 : determination coefficient of the validation test; RMSEv: root mean square error of validation test; RPD: the residual predictive deviation of validation test, LVs: the optimal numbers of latent variables.

43

44