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

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

Weijie Lan[a], Benoit Jaillais[b], Songchao Chen[c], Catherine M.G.C. Renard[a,d], Alexandre Leca[a], Sylvie Bureau[a]*

[a] INRAE, Avignon University, UMR408 Sécurité et Qualité des Produits d'Origine Végétale, F-84000 Avignon, France.

[b] INRAE, ONIRIS, Unité Statistiques, Sensométrie, Chimiométrie (StatSC), F-44322 Nantes, France.

[c] INRAE, Unité InfoSol, F-45075 Orléans, France.

[d] INRAE, TRANSFORM Division, F-44000 Nantes, France.

The corresponding author*

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

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

228 route de l'Aérodrome CS 40509

F-84914 Avignon cedex 9

Tel: +33 432722509

Other authors

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

Benoit Jaillais: benoit.jaillais@inrae.fr

Weijie Lan: Weijie.Lan@inrae.fr

Alexandre Leca: Alexandre.Leca@inrae.fr

Songchao Chen: chensongchao@zju.edu.cn

Highlights

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

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

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

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

Abstract

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

Keywords:

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

1. Introduction

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

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

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

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

2. Material and methods

2.1 Apple materials

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

2.2. Nondestructive characterization of individual apples

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

VIS-NIR spectra of raw apples were acquired using two multi-purpose analyzer spectrometers (Bruker Optics®, Wissembourg, France) at 23°C, which provide diffuse reflectance measurements at wavelength from 500-780 nm (VIS) and 780-2500 nm (NIR), with a spectral resolution of 2 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®). For each apple, VIS-NIR spectra were collected on the blushed and un-blushed sides through a 18 mm diameter area of infrared light. Afterwards, the averaged VIS-NIR spectra, corresponding to the blushed and un-blushed sides of each apple were calculated for further analysis. A reference background measurement was automatically activated before each data set acquisition using an internal Spectralon reference. In total, 120 VIS-NIR spectra of different apples (4 varieties × 10 apples × 3 weeks) were treated before and after cooking.

2.3 Individual apple processing

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

2.4 Determination of quality traits of individual purees

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

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

For all purees, soluble solids content (SSC), titratable acidity (TA), pH and dry matter content (DMC) were characterized based on our previous study (Lan, Jaillais, Leca, Renard, & Bureau, 2020). SSC was determined with a digital refractometer (PR-101 ATAGO, Norfolk, VA, USA) and expressed in °Brix at 22.5°C. 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). The pH values were characterized using a pH meter (FE-20, Mettler-Toledo, China). DMC was estimated from the weight of freeze-dried samples upon reaching a constant weight by a freeze-drying machine (Cryonext, Saint Aunes, France) after 5 days. These measurements were performed with three replicates.

2.5 Spectrum acquisition on individual purees

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

2.6 Statistical analyses and chemometrics

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

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

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

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

3. Results and discussion

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

In this study, both the inter-variability of apple cultivars and the intra-variability of individual apples affected the physical (L*, a*, b*), biochemical (SSC, DMC, TA, pH) and viscosity (η_{50} and η_{100}) properties of corresponding purees (**Fig. 2** and **Fig. S1**).

3.1.1 Color parameters of apples and purees

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

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

3.1.2 Viscosity of purees

Concerning the inter-variability due to varieties on puree rheological properties, BR and GS purees presented a significant ($p < 0.0001$) higher viscosity (η_{50} and η_{100}) than GA and GD purees. BR and GS purees were described to have a bigger particle size and a promoted cell adhesion with more branched pectins than GA and GD involving probably their higher viscosity (Buergy, Rolland-Sabaté, Leca, & Renard, 2020; Buergy, Rolland-Sabaté, Leca, & Renard, 2021). Moreover, the viscosity at the share rate of 50 s^{-1} (η_{50}) was similar ($p > 0.05$) in GD and GA purees. This result was different from our previous one giving a higher viscosity of GD than of GA purees (Lan, Bureau, Chen, Leca, Renard, & Jaillais, 2021). This could be due to the different levels of enzyme inactivation such as pectin methyl-esterase (PME) during apple processing, between microwave processing used in this study and the conventional thermal cooking used previously (Arjmandi, Otón, Artés, Artés-Hernández, Gómez, & Aguayo, 2017). It also could be due to the different apple compositions harvested from two different years in France (2019 for (Lan, Bureau, Chen, Leca, Renard, & Jaillais, 2021) and 2020 for this study). The processing conditions provide indeed different kinds of puree viscosity directly in relation to varieties (Dale, Okos, & Nelson, 1982).

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

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

3.1.3 Biochemical compositions of purees

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

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

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

3.2 Spectral analysis of apples and purees

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

After the pre-processing, the VIS-NIR spectra of all individual raw apples and their related cooked purees with the most of variability could be observed at around 500-700 nm, 1140 nm, 1386-1392 nm, 1880 nm, 1930-2197 nm and 2250-2450 nm (**Fig. 3a, 3b and 3c)**. The specific chlorophyll absorption wavebands at 657-665 nm (Khatiwada, Subedi, Hayes, Carlos, & Walsh, 2016) and 682-689 nm (Mehl, Chen, Kim, & Chan, 2004) were clearly observed in the visible part of spectra for raw apples and their corresponding purees (**Fig. 3a and 3b**), giving information of the color diversity. The second overtone vibrations of C-H bonds between 1100-1250 nm in both, apples and cooked purees (**Fig. 3c**) presented minor changes, indicating a limited effect of processing on sugar contents (Ma, Li, Inagaki, Yang, & Tsuchikawa, 2018). The important bands at 1386-1392 nm, 1880 nm, 1930-2197 nm, and 2250-2450 nm were mainly explained by the combination bands of O-H bonds of waters and C-H bonds of sugars and organic acids in apples (Camps, Guillermin, Mauget, & Bertrand, 2007; Kemps, Leon, Best, De Baerdemaeker, & De Ketelaere, 2010). Compared to raw apples, the lower absorption peaks in their corresponding purees could be due to that apple cooking attenuated the variation of water contents in the processed purees (Lan, Renard, Jaillais, Leca, & Bureau, 2020). Generally, the spectral variability was clearly higher in apples than in their corresponding purees, as already observed in purees prepared from 4 kg of apples (Lan, Renard, Jaillais, Leca, & Bureau, 2020). Besides the possible effect of processing, this difference between apples and purees was probably due to the sample 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 characteristics of apple purees using VIS-NIR spectra acquired on purees (in **Table 1**), or on the corresponding individual raw apples (in **Table 2**).

3.3.1 Prediction of puree characteristics using spectra of purees

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

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

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

3.3.2 Prediction of puree characteristics using spectra of intact apples

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

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

For puree color parameters, VIS-NIR spectra of individual apples coupled with PLS and RF regressions provided an acceptable prediction of redness (a* value) (R_v^2 > 0.77, RPD > 2.1) and yellowness (b* values) (R_v^2 = 0.79, RPD > 2.2) in corresponding purees, but not of lightness (L* values) (R_v^2 < 0.59, RPD < 1.5). The major wavelengths contributing to these models were highly consistent with the models developed using the puree spectra, such as 514 nm and 672 nm. Besides, these good predictions were also in line with their strong internal correlations between apples and purees (**Part 3.1)**. However, these good results need to be interpreted with caution because they concerned the only microwave cooked apples and not the conventional thermal processing (at laboratory scale) in probable relation with a rapid inactivation of enzymes by microwaves and so, a limitation of apple oxidization and color change (Picouet et al., 2009).

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

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

4. Conclusion

This study was designed applying the absolute definition of 'one apple to one puree', which gave a first insight into the impacts of fruit inter- or intra-variability during processing, from the spectroscopic point of view. Importantly, the intra-variability in fruits introduced intensive changes of visual aspects, chemical and textural properties of their corresponding microwave-cooked purees. Taking into account the variability of fruit varieties and intra-variations between apples in each one, could improve the prediction accuracy of regression models. Notably, our results show that simple non-destructive measurement using near infrared spectroscopy applied on apples can provide to processors of apple industry the basic information on the inter- and intra-variability of raw materials and help them to determine the best blend of apples in order to obtain always the same final products such as puree. Further, strong correlations while apple processing obtained from spectral data provided further evidence on such the indirect predictions of color parameters, viscosity and biochemical parameters (SSC, TA and pH) of purees from the non-destructive spectral information acquired on raw apples. Therefore, by systematically scanning all apples, the obtained ~~is could provide~~ objective data on apple quality traits should help to 1) better manage apples according to their quality, 2) predict final product characteristics and 3) reduce fruit wastes at the end.

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

Acknowledgements

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

References:

Arjmandi, M., Otón, M., Artés, F., Artés-Hernández, F., Gómez, P. A., & Aguayo, E. (2017). Microwave flow and conventional heating effects on the physicochemical properties, bioactive compounds and enzymatic activity of tomato puree. *Journal of the Science of Food and Agriculture, 97*(3), 984-990. https://doi.org/https://doi.org/10.1002/jsfa.7824.

Buergy, A., Rolland-Sabaté, A., Leca, A., & Renard, C. M. G. C. (2020). Pectin modifications in raw fruits alter texture of plant cell dispersions. *Food Hydrocolloids, 107*, 105962. https://doi.org/https://doi.org/10.1016/j.foodhyd.2020.105962.

Buergy, A., Rolland-Sabaté, A., Leca, A., & Renard, C. M. G. C. (2021). Apple puree's texture is independent from fruit firmness. *LWT - Food Science and Technology, 145*, 111324. https://doi.org/https://doi.org/10.1016/j.lwt.2021.111324.

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.

Chen, J., & Engelen, L. (2012). Food oral processing: fundamentals of eating and sensory perception: John Wiley & Sons.

Cordella, C. B. Y., & Bertrand, D. (2014). SAISIR: A new general chemometric toolbox. *TRAC Trends in Analytical Chemistry, 54*, 75-82. https://doi.org/https://doi.org/10.1016/j.trac.2013.10.009.

Dale, M. C., Okos, M. R., & Nelson, P. (1982). Concentration of Tomato Products: Analysis of Energy Saving Process Alternatives. *Journal of Food Science, 47*(6), 1853-1858. https://doi.org/https://doi.org/10.1111/j.1365-2621.1982.tb12898.x.

de Brito, E. S., de Araújo, M. C. P., Lin, L.-Z., & Harnly, J. (2007). Determination of the flavonoid components of cashew apple (Anacardium occidentale) by LC-DAD-ESI/MS. *Food Chemistry, 105*(3), 1112-1118. https://doi.org/10.1016/j.foodchem.2007.02.009.

FranceAgriMer. (2017). La Pomme en 2016-2017. Accessed October 2020, from https://www.rnm.franceagrimer.fr.

Guo, Q., Sun, D.-W., Cheng, J.-H., & Han, Z. (2017). Microwave processing techniques and their recent applications in the food industry. *Trends in Food Science & Technology, 67*, 236-247. https://doi.org/https://doi.org/10.1016/j.tifs.2017.07.007.

Huang, Y., Lu, R., & Chen, K. (2020). Detection of internal defect of apples by a multichannel Vis/NIR spectroscopic system. *Postharvest Biology and Technology, 161,* 111065. https://doi.org/10.1016/j.postharvbio.2019.111065

Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). kernlab-an S4 package for kernel methods in R. Journal of statistical software, 11(9), 1-20. https://doi.org/ 10.18637/jss.v011.i09.

Kassambara, A. (2020). ggpubr: 'ggplot2' Based Publication Ready Plots. https://CRAN.R-project.org/package=ggpubr.

Kemps, B., Leon, L., Best, S., De Baerdemaeker, J., & De Ketelaere, B. (2010). Assessment of the quality parameters in grapes using VIS/NIR spectroscopy.

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

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

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

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

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

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

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

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

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

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

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

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

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. https://doi.org/https://doi.org/10.1016/j.postharvbio.2007.06.024.

Noda, I. (1993). Generalized two-dimensional correlation method applicable to infrared, Raman, and other types of spectroscopy. *Applied Spectroscopy, 47*(9), 1329-1336. https://www.osapublishing.org/as/abstract.cfm?URI=as-47-9-1329.

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. https://doi.org/10.1016/j.foodchem.2007.10.003.

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. https://doi.org/10.1016/j.ifset.2009.05.007

Pissard, A., Baeten, V., Romnée, J. M., Dupont, P., Mouteau, A., & Lateur, M. (2012). Classical and NIR measurements of the quality and nutritional parameters of apples: a methodological study of intra-fruit variability. *BASE.* https://doi.org/https://popups.uliege.be:443/1780-4507/index.php?id=8782.

R Core Team, R. C. (2019). R: A language and environment for statistical computing.

Rao, S. R., Thomas, E. G., & Javalgi, R. G. (1992). Activity Preferences and Trip-planning Behavior of the U.S. Outbound Pleasure Travel Market. *Journal of Travel Research, 30*(3), 3-12. https://doi.org/10.1177/004728759203000301.

Stevens, A., & Ramirez-Lopez, L. (2013). An introduction to the prospectr packageR package Vignette R package version 0.1. 3. https://CRAN. R-project. org/package= prospectr.

Wang, J., Wang, J., Chen, Z., & Han, D. (2017). Development of multi-cultivar models for predicting the soluble solid content and firmness of European pear (Pyrus communis L.) using portable vis–NIR spectroscopy. *Postharvest Biology and Technology, 129,* 143-151. https://doi.org/https://doi.org/10.1016/j.postharvbio.2017.03.012.

Xia, Y., Fan, S., Li, J., Tian, X., Huang, W., & Chen, L. (2020). Optimization and comparison of models for prediction of soluble solids content in apple by online Vis/NIR transmission coupled with diameter correction method. *Chemometrics and Intelligent Laboratory Systems, 201*, 104017. https://doi.org/10.1016/j.chemolab.2020.104017

Figure captions:

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

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

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

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

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

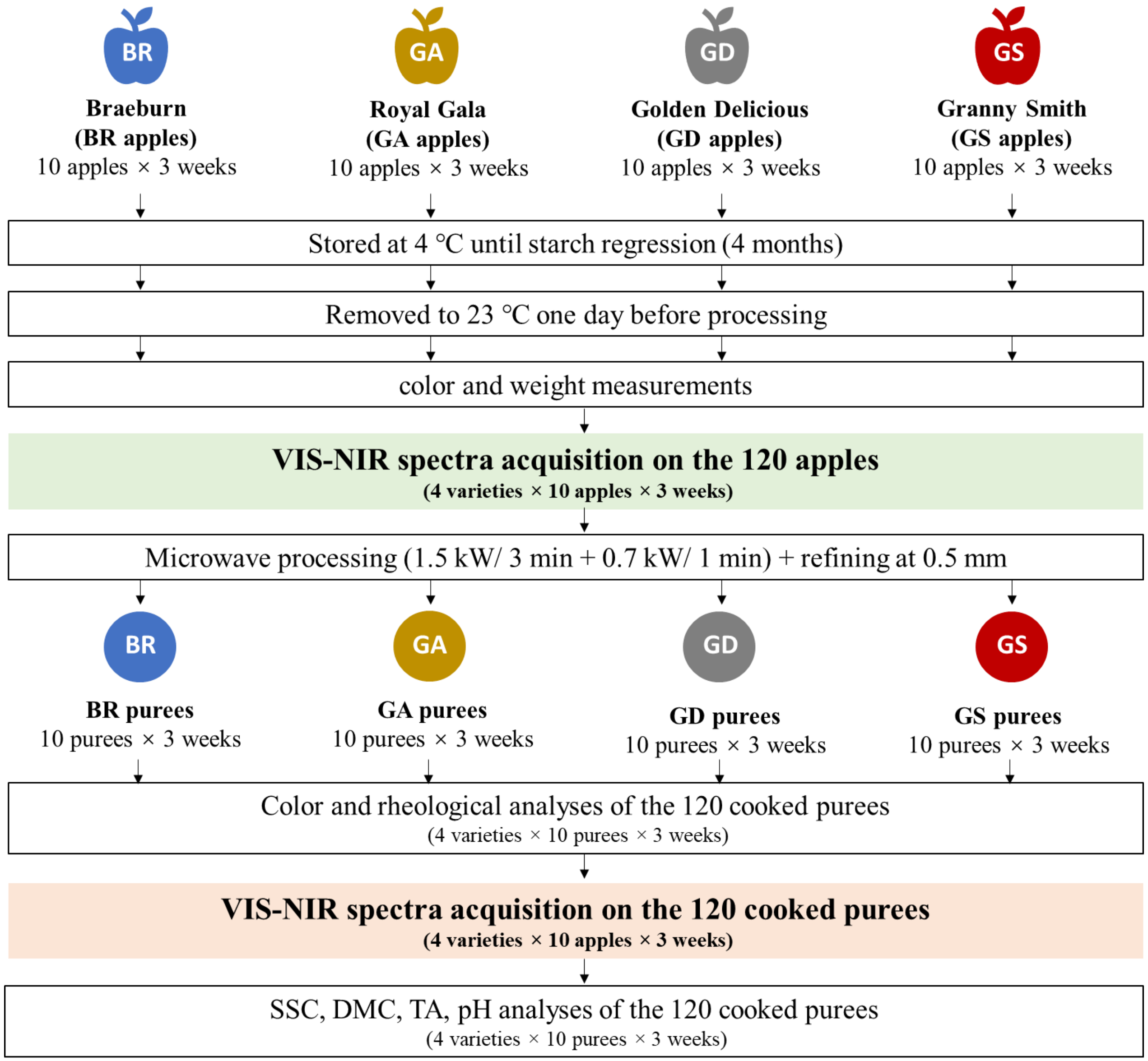
BR
Braeburn
(BR apples)
10 apples × 3 weeks
GA
Royal Gala
(GA apples)
10 apples × 3 weeks
GD
Golden Delicious
(GD apples)
10 apples × 3 weeks
GS
Granny Smith
(GS apples)
10 apples × 3 weeks
Stored at 4 °C until starch regression (4 months)
Removed to 23 °C one day before processing
color and weight measurements
VIS-NIR spectra acquisition on the 120 apples
(4 varieties × 10 apples × 3 weeks)
Microwave processing (1.5 kW/ 3 min + 0.7 kW/ 1 min) + refining at 0.5 mm
BR
BR purees
10 purees × 3 weeks
GA
GA purees
10 purees × 3 weeks
GD
GD purees
10 purees × 3 weeks
GS
GS purees
10 purees × 3 weeks
Color and rheological analyses of the 120 cooked purees
(4 varieties × 10 purees × 3 weeks)
VIS-NIR spectra acquisition on the 120 cooked purees
(4 varieties × 10 apples × 3 weeks)
SSC, DMC, TA, pH analyses of the 120 cooked purees
(4 varieties × 10 purees × 3 weeks)

Fig. 1

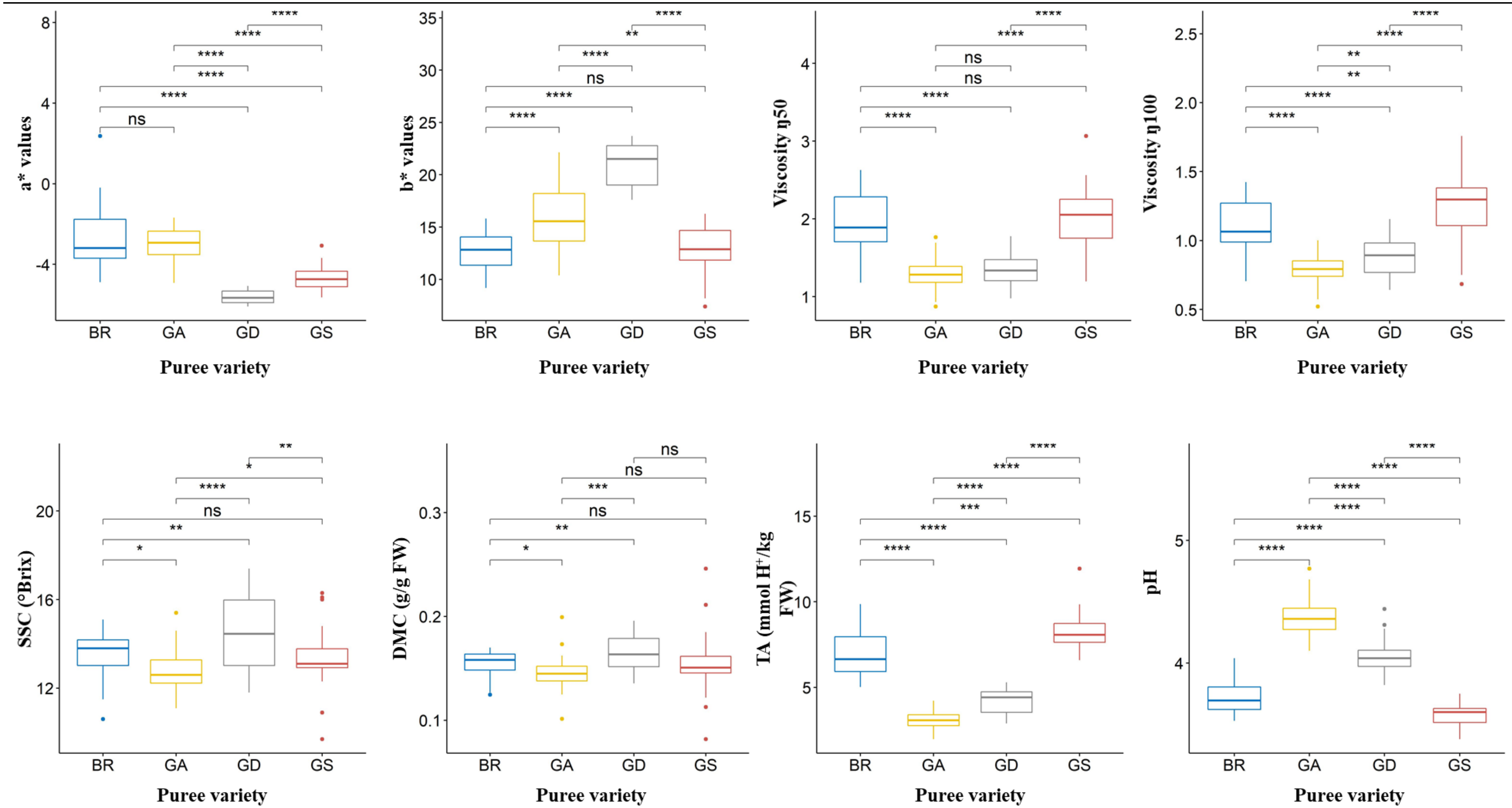
a* values
b* values
Viscosity η50
Viscosity η100
SSC (°Brix)
DMC (g/g FW)
TA (mmol H+/kg FW)
pH
BR
GA
GD
GS
Puree variety
ns
*
**
***
****

Fig. 2

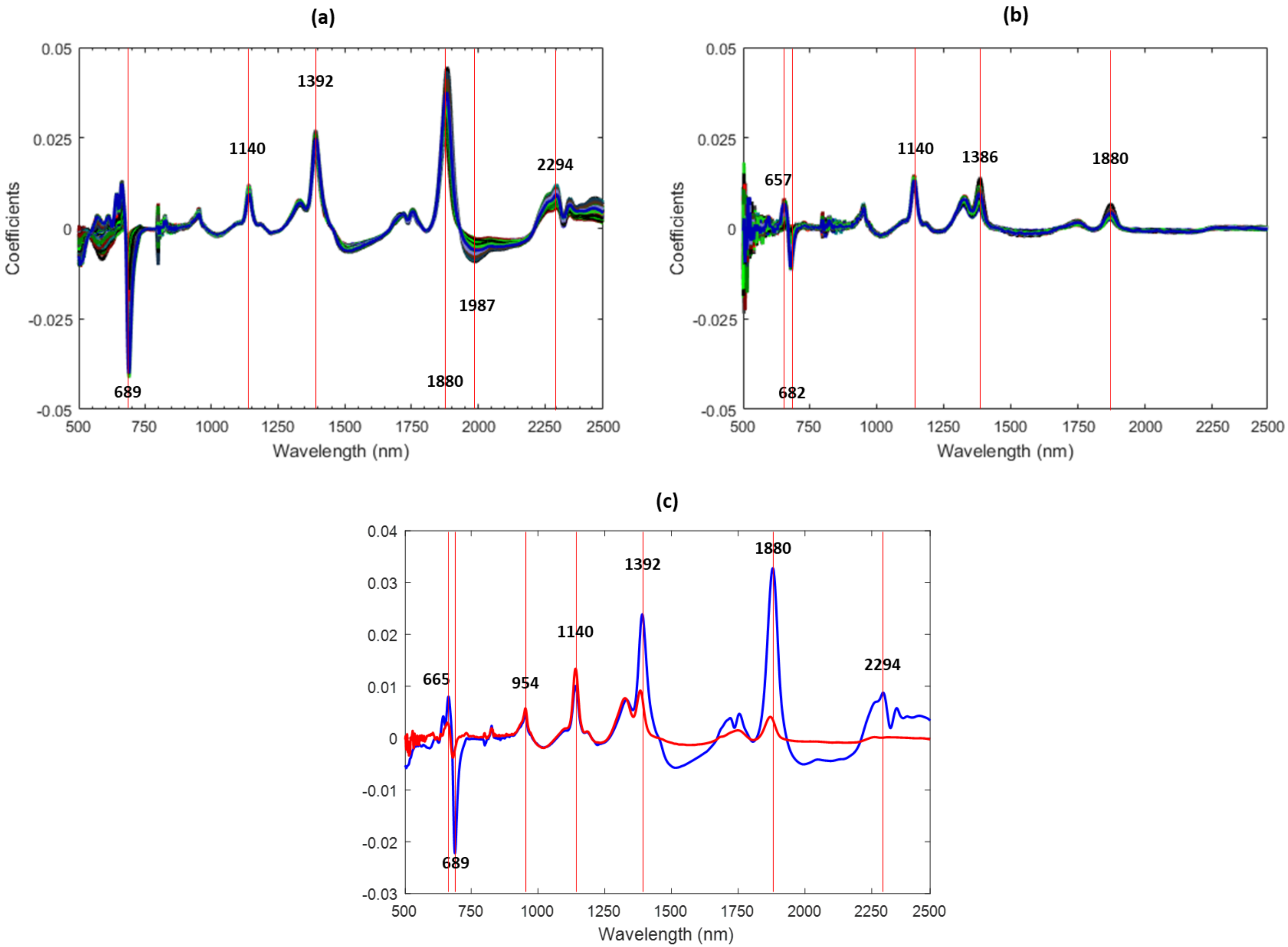
(a)
Coefficients
Wavelength (nm)
689
1140
1392
1880
1987
2294
(b)
Coefficients
Wavelength (nm)
657
682
1140
1386
1880
(c)
Wavelength (nm)
665
689
954
1140
1392
1880
2294

Fig. 3

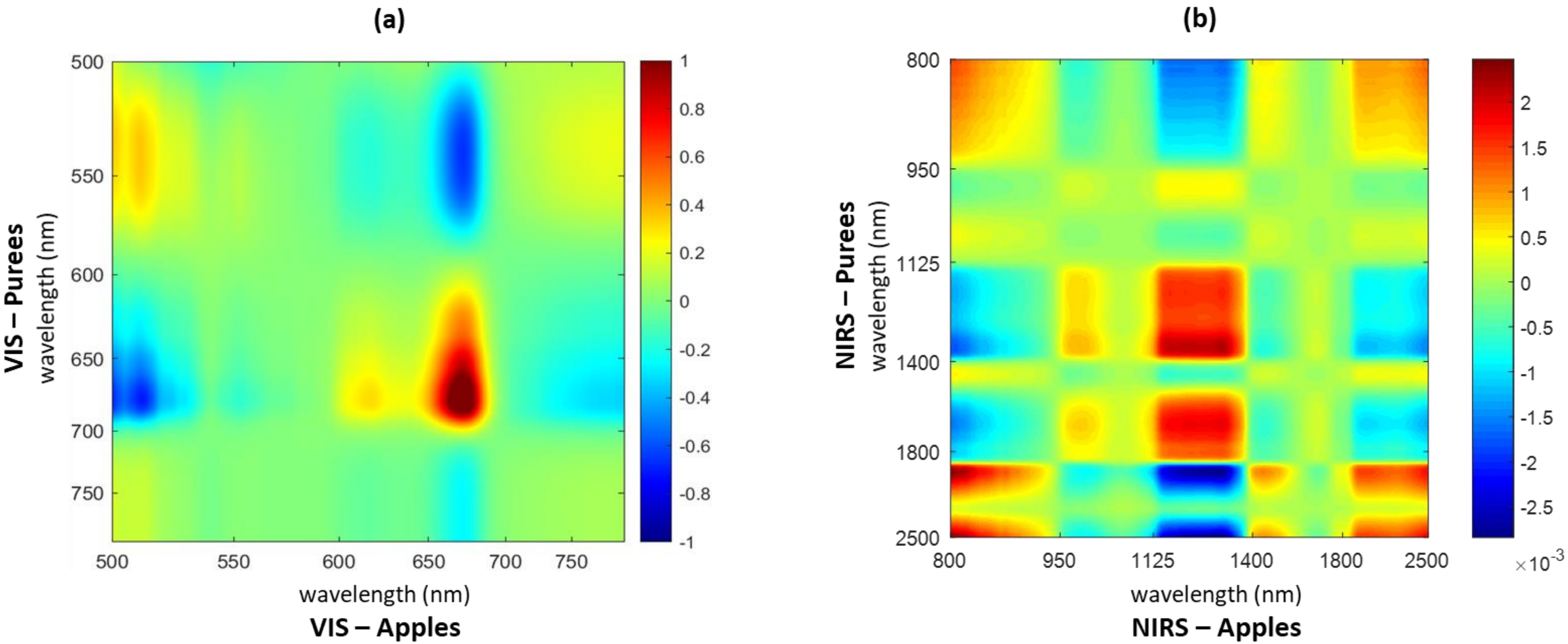
(a)
VIS – Purees
wavelength (nm)
500
550
600
650
700
750
wavelength (nm)
VIS – Apples
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
(b)
NIRS – Purees
wavelength (nm)
800
950
1125
1400
1800
2500
wavelength (nm)
NIRS – Apples
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
×10^-3

Fig. 4

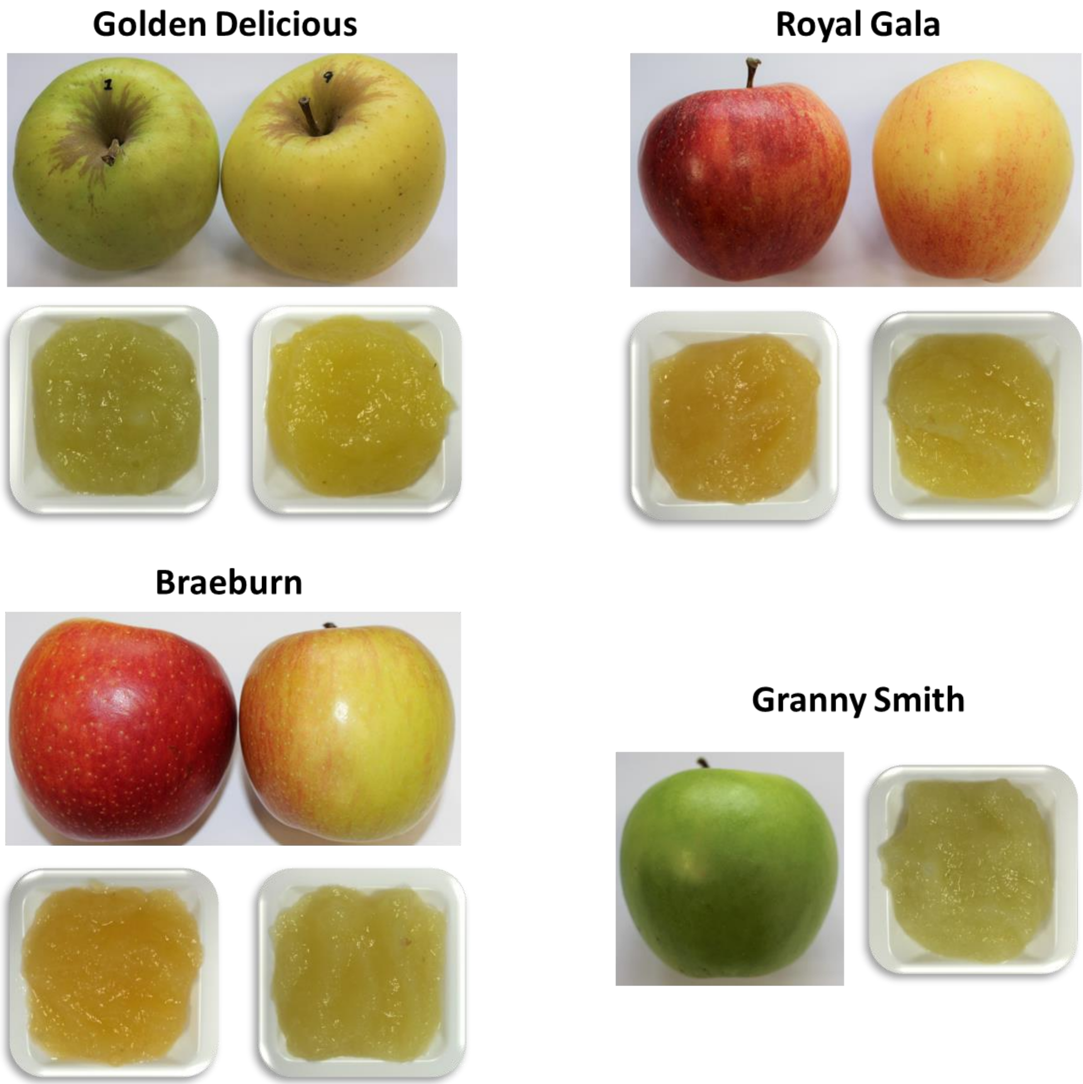

Fig. S1

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

			PLS-R				SVM-R			RF-R		
Parameter	Range	SD	R_v^2	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.

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

			PLS-R				SVM-R			RF-R		
Parameter	Range	SD	R_v^2	RMSEV	RPD	LVs	R_v^2	RMSEV	RPD	R_v^2	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

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.