A method using near infrared hyperspectral imaging to highlight the internal quality of apple fruit slices

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

A method using near infrared hyperspectral imaging to highlight the internal quality of apple fruit slices

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

[a] INRAE, Avignon Université, UMR 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, TRANSFORM, F-44000 Nantes, France.

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

Corresponding authors*

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

Fax: +33 432722492

Other authors

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

Benoit Jaillais: benoit.jaillais@inrae.fr

Alexandre Leca: Alexandre.Leca@inrae.fr

Songchao Chen: Songchao.Chen@inrae.fr

Carine Le Bourvellec: carine.le-bourvellec@inrae.fr

Weijie Lan: Weijie.Lan@inrae.fr

Highlights

- Near-infrared hyperspectral imaging assessed the apple heterogeneity.
- Principal component analysis allowed to select the region of interest in apples.
- Single apple presented a large heterogeneity of biochemical compositions.
- Models mapped total sugars and dry matter contents in each apple.

Abstract

The heterogeneity of apple fruit was highlighted by near-infrared hyperspectral imaging (NIR-HSI) using a data analysis in two successive steps. First, NIR-HSI images were acquired on the cut surface of six transverse slices per apple, which were then systematically sampled with 5 or 6 cylinders per slice. PCA carried out on the NIR-HSI images allowed to select 141 representative cylinders from the total dataset (1056 samples), in which the contents of dry matter (DMC), total sugars (TSC), fructose, glucose, sucrose, malic acid and polyphenols were quantified by spectrophotometry and chromatography. In a second step, leave-one-out PLS models were developed and successfully used to describe the distribution of DMC (R_{cv}^2 = 0.83, RPD = 2.39) and TSC (R_{cv}^2 = 0.81, RPD =2.20) in each apple slice. A strong heterogeneity of DMC and TSC was detected inside each fruit. Such a simple and rapid method reduced the needs of numerous chemical characterizations to demonstrate the distribution of quality traits within and between fruit and contributed to better manage the fruit quality measurements.

Keywords: *Malus domestica* Borkh.; partial least square regression; random forest regression; apple variability and heterogeneity.

Introduction

An external aesthetic appearance and a sustainable internal quality of fruit are both crucial for consumers (Ma et al., 2018; Zhang et al., 2018). However, genetic diversity (varieties), pedoclimatic conditions and agricultural practices are known to provide variability and heterogeneity of fruit, which limit the precision and prediction of quality using infrared methods (Vis-NIR and NIRS) and thus hinders their widespread applications for online commercial fruit sorting (Barritt et al., 1991; Xia et al., 2020; Zhang et al., 2018). It appears necessary to develop some applications using efficient and rapid technologies to phenotype internal heterogeneity of the fruit, in order to help field growers and industrial manufacturers to improve quality of fruit products.

Apple is one of the most consumed agricultural commodities in the global fruit market (68.6 million tons at 2018) (USDA, 2018). The high heterogeneity of soluble solids content (Fan et al., 2016; Mo et al., 2017; Peiris et al., 1999), starch (Menesatti et al., 2009), polyphenols and vitamin C (Pissard et al., 2012) in a single apple fruit has been proven to truly exist in different directions, from proximal to distal direction (Fan et al., 2016; Peiris et al., 1999), in radial direction from inside to outside (Mo et al., 2017) and along equatorial direction (Mo et al., 2017; Pissard et al., 2012).

As known, conventional chemical analyses (HPLC-DAD, GC-MS and ultraviolet/ visible spectrometry etc.) are costly and time-consuming to determine the heterogeneity occurring at the level of the tissues in a single fruit (Peng et al., 2019; Pissard et al., 2012). To determine the chemical heterogeneity within a fruit, most

previous works encountered difficulties of i) long-periods and intensive labor operations, ii) a large amount of targeted fruit samples and the high requirements for characterization, and iii) the limited stability of fruit samples (highly hydrated, rapid oxidation). In addition, the limited knowledge of apple heterogeneity becomes a barrier to obtain robust predictive models by high-throughput techniques (Vis-NIRS, NIRS, MIRS, NMR) (de Oliveira et al., 2014; Fan et al., 2016; Pissard et al., 2012). Particularly with the non-destructive and localized (around 2 cm^2) NIR measurements on apples, it is essential to know more about the distribution of components in fruit in order to determine where and how many measurements are needed, as well as to access the representative sample portion to be characterized using reference methods for calibration dataset.

Hyperspectral imaging (HSI) is an emerging platform technique that integrates imaging and spectroscopy to provide both spatial and spectral information (Gowen et al., 2007). It is safer than X-ray imaging, more rapid and affordable than FT-IR imaging and Magnetic resonance imaging, and with a better image quality than thermal imaging (Fan et al., 2016; Ma et al., 2018). Until now, applications of HSI in the Visible-NIR (400-1000 nm) or NIR (1000-2400 nm) ranges were carried out to evaluate the variability of apple quality, such as fruit defects (Mehl et al., 2004), firmness (Peng and Lu, 2008), mealiness (Huang and Lu, 2010) and soluble solids content (Mendoza et al., 2011). These studies were applied nondestructively on apple fruit. As the NIRS radiation penetration depth is around 0.2 to 0.3 cm in the spectral area between 900 and 1900 nm (Lammertyn et al., 2000), the non-destructive

97 detection of HSI does not allow to evaluate the entire internal heterogeneity of apple
98 fruit. Thus, the HSI is used destructively by scanning fruit slices and makes possible
99 to describe the distribution of the internal soluble solids content, as shown in apples
100 (Mo et al., 2017) and melons (Sun et al., 2017). However, these studies need a large
101 number of reference data (numbers of samples and limited samples quantity) on all
102 the targeted areas of single fruit, required for model calibration.

103 Consequently, the main objective of this work was to provide a simple and
104 efficient method to reduce the intensive reference measurements (contents of dry
105 matter, total sugars, individual sugars, acids and polyphenols) in order to develop a
106 HSI modelling calibration and to evaluate the apple variability and heterogeneity.

107 **2. Material and methods**

108 2.1 Apple fruit

109 The experiment was conducted on four different apple varieties: 'Golden
110 Delicious' (GD), 'Granny Smith' (GS), 'Braeburn' (BR) and 'Royal Gala' (GA). In
111 2018, all apples were harvested in the experimental orchard at La Pugère (Bouches du
112 Rhône, France). 'Braeburn', 'Granny Smith' and 'Royal Gala' apples were grown
113 under a commercial fruit thinning (Th+, 50-100 fruit/ tree). 'Golden Delicious' apples
114 were grown under two thinning conditions, the commercial fruit thinning (Th+,
115 50-100 fruit/ tree) and without thinning (Th-, 150-200 fruit/ tree). After the
116 commercial harvesting ('Royal Gala' on August 28th, Golden Delicious on September
117 19th, 'Granny Smith' on September 20th, and 'Braeburn' on October 3rd), all apples
118 were stored in a cold chamber at 4 °C and at around 90 % of humidity until their

characterization (November 2018).

2.2 Samples preparation

A calibration dataset corresponded to the data of 30 apples with similar sizes (6 fruit × 5 apple groups of GD Th-,GD Th+, GS, BR, GA) and scanned using the NIR-HSI imaging system. Each apple was cut with a slicing tool along horizontal direction to produce six apple slices, including five 1.2 cm thick slices (named slices from 'A' to 'E' at the stem, equator and calyx directions) and the one residual piece of varying thickness (named slice 'F' at the calyx positions). Hyperspectral images of 180 apple slices (5 apple groups × 6 fruit × 6 slices) were acquired and six cylindrical 1.6 cm diameter portions were extracted with a cookie cutter (numbered 1 to 6) from each of the apple slices A to E, and five or six cylinders from the residual slice F (**Fig. 1**).

The cylinders were put immediately in liquid nitrogen prior to storage at -20 °C, giving 35 to 36 cylinders per apple, following the previous works of Mo et al. (2017) and Bureau et al. (2013). These cylinders were distributed with a systematic repartition for each apple from the top to the bottom and from the sunny to the shady faces. In total 1056 cylinders (5 apple groups × 6 fruit × 35-36 cylinders) were numbered and stored (**Part 2.4.1**). After the extraction of all the cylinders, RGB photos were taken on each apple slice in order to ensure the correct correspondence between the cylinders and HSI images (**Fig. 1**).

2.3 Hyperspectral Imaging (HSI) System

A pushbroom (a line-scanning type) near infrared hyperspectral imaging system

(SPECIM, Oulu, Finland) was used to acquire the hyperspectral images of apple slices. Particularly, this NIR-HSI system consisted of a SWIR camera (SWIR-CL-400-N25E, SPECIM) covering the spectral range of 1000-2500 nm with a spectral resolution of about 12 nm, an OLES 56 camera lens (SPECIM), an illumination source (halogen lamps) and a translating scanner. All the image acquisition parameters (the exposure time of camera, the scanning speed etc.) were controlled by the LUMO® software from SPECIM. Before measurements, a reflectance calibration was performed by recording a dark current image (0 % reflectance) with an internal shutter and a white image using a reference standard close to 100 % reflectance (Spectralon® 100 %). To reduce the impact of light and noise, the calibrated hyperspectral images could be automatically obtained using the dark and white reference images, with the following equation:

$$R(\lambda) = \frac{R_0(\lambda) - R_d}{R_w - R_d} \times 100\ \% \qquad (1)$$

with R: the calibrated hyperspectral image data, R_0: the raw image data, R_d and R_w: the dark and white reference images, respectively.

All images were acquired in the reflectance mode and the final image size for each kernel is 387 × xdim × 288, the two first values representing pixel dimensions in the x and y directions (field of view of 9.8 × 6.3 cm, with a spatial resolution of 225 µm) and the third value accounting for the number of spectral channels. The xdim values varied according to the dimensions of apple slices. Each image was acquired in about twenty seconds. As the beginning and ending wavelengths contained noise caused by the instrument itself (Sun et al., 2017), the 258 bands from 990 to 2450 nm

were selected for further spectral analysis.

2.4 Imaging pre-processing

The pre-processing of the hyperspectral images and the selection of region of interest (ROIs) were performed with Matlab 7.5 (Mathworks Inc. Natick, MA) software using the SAISIR package (Cordella & Bertrand, 2014). Due to the high volume of data, the processing of all images was not possible using a common computer. In this way, 10,000 spectra were randomly extracted from the HSI images of each apple slice, counting around one third of the total number of spectra in each HSI image. Afterwards, all random selected spectra were gathered into a matrix $\boldsymbol{X}$ (5 apple groups × 6 fruit × 6 slices × 10,000 rows by 258 columns). After pre-tests, matrix $\boldsymbol{X}$ was smoothed by a window size of three pixels. A given value x (i) of index i was replaced by the local average of x (i - 1) + x (i) + x (i + 1). Then it was pre-processed with standard normal variate (SNV) to increase its signal to noise ratio for the selection of ROIs.

2.5 ROI selection and characterization

PCA has been commonly applied on the NIR-HSI of agro-food products for safety and quality assessments (Dale, et al., 2013). It was performed on the pre-processed matrix $\boldsymbol{X}$ to check the major components causing variability in the apples. Afterwards, this model was applied to all pixels of all images, and the major components (PCs) were selected as estimators to refold into PCs images to point out the heterogeneous areas in each HSI image of apple slice. Finally, the ROIs to be analyzed by chemical and biochemical measurements (141 samples) were manually

selected depending on the results of the major principal components and the same location on photographical images (an example of the ROIs marked black circles in **Fig. 1**).

2.6 Chemical and biochemical measurements

All chemical and biochemical characterizations (contents of dry matter, fructose, glucose, sucrose, malic acid and sum of polyphenols) were performed on these ROIs (141 samples) and expressed as the ratio on fresh weight. Particularly, individual sugars (glucose, fructose, and sucrose) and malic acid were quantified on the half of each sample using an enzymatic method with commercial kits for food analysis, following the manufacturer's instructions (R-biopharm, Darmstadt, Germany). The total sugars content were computed by the sum of all individual sugars (fructose, glucose and sucrose). The dry matter content (DMC) was estimated from the weight of freeze-dried samples upon reaching a constant weight (freeze-drier, 3 days). The freeze-dried samples were further used to quantify polyphenols by HPLC-DAD after thioacidolysis as described in Le Bourvellec (Le Bourvellec et al., 2011). Particularly, apple polyphenols were separated in an Agilent 1050 separation system coupled with a (250 mm × 4 mm i.d.) Licrospher PR-18 5 µm column (Merck, Darmstadt, Germany) operated at 30 °C. This data was presented as the sum of individual polyphenols including procyanidins and monomeric flavanols, phenolic acids, dihydrochalcones and flavonols.

2.7 Modelling

After smoothing with a 3-point window and the first order derivative with a 11

point window, the averaged spectra of each ROI (giving 141 spectra) and their related reference data were used for modelling. Leave-one-out partial least squares (LOO-PLS) regession was used to build prediction models with Matlab 7.5 (Mathworks Inc. Natick, MA) software using the SAISIR package (Cordella & Bertrand, 2014). Random forest (RF) regression was also applied to compare the prediction ability of developed models, using R software (version 4.0.2) (R Core Team, 2019) coupled with several packages including 'prospectr' (Stevens and Ramirez-Lopez, 2014), 'Rmatlab' (Bengtsson et al., 2018), 'caret' (Kuhn, 2015) and 'randomForest' (Liaw and Wiener, 2002).

The developed model performance was assessed using the determination coefficient of cross-validation (R_{cv}^2), the root mean square error of cross-validation ($RMSE_{cv}$), the number of latent variables (LVs), the ratio of the standard deviation values (RPD). The interpretations of beta-coefficients were used to determine the relevant spectral regions. The spectral bands related to the maximum and minimum of beta-coefficient values can present the most important wavelengths (Sun et al., 2017).

2.8 Prediction maps of apple quality attributes

After comparison of the modeling results of each apple quality attribute, only the models with RPD values higher than 2.0 allowing a coarse quantitative prediction (Nicolai et al., 2007), were selected to predict fruit quality attributes of all apple slices at the individual pixel level. The prediction values were then visualized under the form of prediction maps, which were used to phenotype the internal distributions of the predicted quality attributes in apples.

3 Results and discussion

3.1 Spectral characteristics

The initial PCA conducted on the random selected spectra of one out of three pixels of all apple slices (matrix $\boldsymbol{X}$) was able to discriminate the variability and heterogeneity of apple fruit between the top (slice A) and the bottom (slice F). The first two principal components represented 68.0 % of the total variability, with the first component (PC1) of 43.7 % and the second component (PC2) of 24.2 %, respectively. For all apple groups, a clear discrimination was shown along the first two principal components (PC1 and PC2) between the middle slices (slices C, D) and the others (top slices A, B and bottom slices E, F). The most contributing wavelengths of PC1 and PC2 were: i) the sharp peak around 1065 nm corresponding to the C-H and O-H stretching in second overtone, which is linked to the sugar variations in fruit (Sun et al., 2017); ii) the absorption region from 1157 - 1364 nm which is associated with the first overtone of O-H band in water (Ignat et al., 2014); and iii) the broad band at 1400-1530 nm which corresponds to the combination of second overtone of C-H stretching and the first overtone of O-H stretching, already used to determine the soluble solids content in apples (Zhang et al., 2019). These fingerprint wavelengths pointed out the variations of water and carbohydrate contents in a single apple, which were consistent with previous results using chemical measurements (Peiris et al., 1999; Pissard et al., 2012).

In a second step, the variability expressed on the dominant PC1 components (43.7 % of total variability) was used for phenotyping all apple slices based on a

correspondence between the different areas described by a color range, according to their hyperspectral spectra. PC1 scores-images have directly pointed out the most variable locations with the color range (**Fig. 1**). ROIs in each apple were targeted at top slice A, middle slice C and bottom slice E, with the most different colored areas (such as the area No. 3 of slice C and the area No. 3 of slice F in **Fig. 1**). Besides, the clear color differences inside the middle slices (area No. 2 and No. 5 of slice D in **Fig. 1**) were also selected, depending on apple cultivars. A total of 141 ROIs was manually selected and characterized by reference chemical measurements to check if these targeted positions really showed variations consistent with the corresponding hyperspectral images, and to identify the chemical components responsible for the heterogeneity observed in PC1 scores images.

3.2 Chemical characteristics of ROIs

The boxplot of chemical reference data (**Fig. 2**) of the 141 selected ROIs showed a large variation of contents of dry matter, total sugars, malic acid and polyphenols in the different apple cultivars.

Royal Gala apples had the most intensive variations of DMC among the five apple groups (**Fig. 2a**). Conversely, the lowest variations of DMC and of TSC (**Fig. 2b**) were observed in the thinned (GD Th+) and non-thinned Golden Delicious (GD Th-), presenting a relatively limited heterogeneity of DMC and TSC in single GD apples. The fructose content of Granny Smith (GS) had the lowest variations among the four cultivars (**Fig. 2c**). Moreover, the contents of polyphenols varied a lot in each apple cultivar (**Fig. 2f**). Golden Delicious (thinned and non-thinned) (0.34 ± 0.14 g/kg

in non-thinned GD and 0.34 ± 0.12 g/kg in thinned GD) and Royal Gala (GA) (0.27 ± 0.14 g/kg) apples presented a large polyphenolic variation compared to GS apples (0.55 ± 0.14 g/kg). This result was different from a previous work showing a small internal heterogeneity of polyphenols in Gala (Vidot et al., 2019). This inconsistent result could be due to the difference in the measured targeted areas in apples, only parts close to the fruit surface (Vidot et al., 2019) versus parts distributed everywhere inside the entire fruit (our experiment).

Concerning the effect of agricultural practices on Golden apple quality, the average contents of total sugars and malic acid were higher in the thinning condition (GD Th+) than in the non-thinning one (GD Th-), which was in line with our previous results observed during the 2017 harvested season (Lan et al., 2020). Interestingly, the tree thinning treatment, by increasing the individual apple growth potential, led to a lower variability of malic acid (**Fig. 2f**) and sucrose (**Fig. 2d**) contents in Golden Delicious apples, with the standard derivation values decreasing from 0.89 to 0.62 g/kg and from 10.9 to 9.3 g/kg, respectively.

Consequently, the most variable regions chosen according to the PC scores-images truly exhibited a large heterogeneity, in agreement with the variations of the reference values of total sugars, dry matter, malic acid and polyphenols. The apple internal heterogeneity should be then considered as an important factor for apple fruit quality characterization and understanding.

3.3 Prediction of apple quality traits based on averaged spectra of ROIs

The chemical composition data obtained on the 141 selected ROIs was used to

build prediction models validated within this selected subset, using the averaged spectra of each ROI. Acceptable predictions of DMC (SD = 21.9 mg/g, R_{cv}^2 = 0.83, $RMSE_{cv}$ = 9.7 mg/g, RPD = 2.39) and TSC (SD = 18.7 g/kg, R_{cv}^2 = 0.81, $RMSE_{cv}$ = 8.4 g/kg, RPD = 2.20) were obtained by LOO-PLS, respectively (**Table 1**). According to Nicolai et al. (2007), a RPD over 2 indicates the possibility to a coarse qualitative prediction of the internal attributes of fruit. The linear models (PLS) were much better than the random forest (RF) (**Table1**), as described by Sun et al. (2017) to predict soluble solids content in melon fruit. The small number of latent variables (LVs) employed in PLS models indicated the robust prediction of DMC (LVs = 7) and TSC (LVs = 5), based on data including different apple varieties and growing agricultural practices. All predicted DMC and TSC on 141 ROIs by LOO-PLS regression were well correlated to the measured values, according to their linearity correlation plots (**Fig. 3a and 3b**). Moreover, the beta-coefficients showed strong positive or negative bands (**Fig. 3c and 3d**) for both, the PLS regressions of DMC and TSC, including informative spectral regions at around 1123 nm, 1208 nm, 1389- 1401 nm, 1474-1480 nm, 1857- 1863 nm and 2319- 2336 nm, which have been widely reported to estimate water and sugar contents in apple fruit (Giovanelli et al., 2014; Lan et al., 2020; Peirs et al., 2003). Particularly, six sharp peaks at 1208 nm, 1123 nm, 1389 nm, 1474 nm, 1857 and 2336 nm were identified as being important wavelengths to predict dry matter content in apples. And the specific wavelengths at 1123 nm, 1401 nm, 1480 nm, 1863 nm and 2319 nm contributed to the determination of total sugars in apple tissues.

However, modelling using the averaged spectra of ROIs showed a limited ability to predict the individual sugars (fructose, glucose and sucrose), malic acid and sum of polyphenols (**Table 1**). This was expected and in agreement with the previous work (Walsh et al., 2020). That could be due to i) their respective lower content in apple tissues compared with DMC and TSC and ii) the limited chemical variations in our studied apple varieties. Concerning polyphenols, a larger variation is observed in the cider apple varieties from 1 to 7 g/kg in apple parenchyma (Sanoner et al., 1999) than in the dessert varieties, such as those of this study, from 0.6 to 0.9 g/kg (Guyot et al., 2002) because of their highest content in procyanidins, the main polyphenols. Thus, a better prediction of these compounds might be obtained taking into account the entire variability within apple varieties.

As mentioned in **section 3.1**, the fingerprint wavelengths of apple variability and heterogeneity were mainly related to water and carbohydrates. Thus, for these five apple groups (BR, GA, GS, thinned and non-thinned GD), prediction models based on the averaged HSI spectra of ROIs and their reference values were suitable to estimate intensive variations of water and the dominated soluble contents in apple fruit, such as dry matter and total sugars, but not of individual compounds (fructose, glucose, sucrose and malic acid) or microcomponents (sum of polyphenols).

3.4 Phenotyping apple heterogeneity by HSI

For a more in-depth assessment of the internal composition of each apple, the best PLS models described in the **Part 3.2** were applied to predict the quality traits at each pixel on all hyperspectral images of apple slices. The resulting images were

presented as ‘prediction maps’ for TSC (**Fig. 4**) for each apple slice. In total, 10 colors were used to fit the different intervals of the predicted values and pixels with the similar predicted values appeared in the same color. The prediction results demonstrated a large variability and heterogeneity of total sugars and dry matter contents i) in different apple varieties; ii) between individual apple fruit and iii) inside single fruit.

For the traditional non-destructive NIR analyses on apples, to obtain a robust prediction model, the calibration dataset should be sufficiently rich in variations, particularly taking into account the existing variability with the fruit itself (Zhang et al., 2018). Our prediction results provided advanced knowledge to determine where and how many positions are needed with the non-destructively NIRS measurements on apple surfaces, as well as to access the sample portion to be analyzed by reference methods for the calibration set.

In the literature, NIR predictions of apple quality traits involve taking measurements at up to four points located in the equatorial region (Liu and Ying, 2005; Peirs et al., 2003; Pissard et al., 2012), or along the stem, equator and calyx positions of apples (Fan et al., 2016). However, there was a reverse conclusion to reach the accurate predictions of developed models following each of these two methods. From our results, a specific attention needs to be paid according to the ‘cultivar’, which is the major factor influencing the fruit heterogeneity and the possible reason to explain the aforementioned disagreement result. According to the relative standard deviation (RSD) values of the predicted DMC and TSC of all pixels in single apples, different

levels of internal chemical variations were observed in Braeburn (RSD of DMC = 24.6 % and of TSC = 22.1 %), Royal Gala (RSD of DMC = 26.5 % and of TSC = 27.1 %), Granny Smith (RSD of DMC = 18.9 % and of TSC = 22.0 %) and thinned Golden Delicious (RSD of DMC = 13.2 % and of TSC = 15.7 %). These results indicated the same and limited spectral measurement points for all apples could not present such intensive internal quality variations of different cultivars. From a spectroscopic point of view, an increase of measured positions on apple surfaces therefore is particularly important to improve accuracy in the calibration steps.

In all apples, the large DMC and TSC differences among the middle (the average predicted DMC of all pixels in slice C and D of all cultivars = 136.5 ± 16.2 g/kg and TSC = 115.6 ± 14.3 g/kg), top (the average predicted DMC of all pixels in slice A and B of all cultivars = 117.1 ± 22.4 g/kg and TSC = 79.5 ± 17.1 g/kg) and bottom slices (the average predicted DMC of all pixels in slice E and F of all cultivars = 124.1 ± 25.2 g/kg and TSC = 87.3 ± 20.1 g/kg) demonstrated that four points at the equatorial region might not be enough to provide the representative spectra of the entire apple fruit. NIRS information from top to bottom of apple surfaces therefore needs to be considered for all apple cultivars.

Consequently, the strong variability and heterogeneity of apples were highlighted using our developed models, and probably constitute the major barrier to an accurate NIR modelling. The similar distribution results of TSC and DMC in apple slices were observed in most apple slices of each cultivar (at least 4 over 6 fruits). These results provided an important opportunity to advance our knowledge on the quality

383 measurement: where and how many specific positions need to be measured on apple
384 surfaces with NIRS, in order to develop accurate and robust prediction models.

385 The previous HSI models mainly detected the soluble solids content and firmness
386 changes in single fruit (Mo et al., 2017; Sun et al., 2017), because of the quick and
387 easy reference data quantification of all targeted samples using digital refractometers
388 and hardness detectors. Compared to these studies, our work provided an efficient
389 solution for the HSI modelling calibration step, depending on the reference data
390 measured on 141 representative samples instead of the 1056 prepared samples.
391 Importantly, this method offered a new sight on contents of total sugars (sum of the
392 fructose, glucose and sucrose) and dry matter in apples, with a limited number of
393 complicated (individual sugars measured by spectrometry using enzymatic kits) and
394 time-consuming (at least 24 hours for freeze-drying) analyses for HSI modelling. In
395 future, such a rapid and efficient approach for HSI modelling calibration would be
396 helpful to detect the variations of apple internal quality parameters according to
397 different environmental conditions (crop load, irrigation, light penetration and
398 elevations of regions etc.——…) and growing stages, and then contribute to an
399 improvement of apple quality and production. The objective at the end could be to
400 have a better knowledge of the apple homogeneity in order to manage them better for
401 fresh market and processing taking into account the sustainability of practices.

402 **4. Conclusion**

403 In this study, the power of chemometric methods was harnessed in a two-steps
404 procedure for mapping of apple fruit heterogeneity while minimizing the number of

chemical analyses. PCA of NIR-HSI data was used to scan the heterogeneity of apple slices and to pin-point the best representing areas of the whole spectral variation. A limited number of chemical measurements could then be carried out and exploited by PLS regression to identify the underlying compositional information present in NIR-HSI data at individual pixels. NIR-HSI coupled with PLS regression showed a good ability to phenotype the distribution of dry matter content and total sugars content in apple fruit. The prediction models developed with the reference values of the most variable areas identified by PCA on HSI data were enough to assess the variability and heterogeneity of apple global parameters, with acceptable precisions (range of values). For dry matter and total sugars, the PLS results had a better ability than the random forest ones to estimate their distributions in apple slices. With the rapid scanning of apple slices and a limited number of chemical measurements, this method showed the great advantages of a simple fruit sampling, less experimental deviations caused by rapid oxidation of fruit, and a high efficiency of model developments. This method opens the possibility to more systematically evaluate the fruit variability and heterogeneity in future projects.

Acknowledgements

The authors thank Shouyang Liu, Xiuliang Jin, Jiantao Zhao, Patrice Reling, Barbara Gouble, Line Touloumet, Marielle Boge, Caroline Garcia (INRAE, SQPOV unit) for their technical help. The ‘Interfaces’ project is an Agropolis Fondation Flashship 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

Barritt, B.H., Rom, C.R., Konishi, B.J., Dilley, M.A., 1991. Light level influences spur quality and canopy development and light interception influence fruit production in apple. HortScience 26, 993-999. https://doi.org/10.21273/HORTSCI.26.8.993.

Bengtsson, H., Jacobson, A., Riedy, J., Bengtsson, M.H., LazyLoad, T., ByteCompile, T., 2018. Package 'R. matlab'.

Bureau, S., Ścibisz, I., Le Bourvellec, C., & Renard, C. M. G. C. (2012). Effect of Sample Preparation on the Measurement of Sugars, Organic Acids, and Polyphenols in Apple Fruit by Mid-infrared Spectroscopy. J. Agric. Food Chem. 60(14), 3551-3563. https://doi.org/10.1021/jf204785w.

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

Dale, L. M., Thewis, A., Boudry, C., Rotar, I., Dardenne, P., Baeten, V., & Pierna, J. A. F. (2013). Hyperspectral Imaging Applications in Agriculture and Agro-Food Product Quality and Safety Control: A Review. Applied Spectroscopy Reviews, 48(2), 142-159. https://doi.org/10.1080/05704928.2012.705800.

de Oliveira, G.A., Bureau, S., Renard, C.M.G.C, Pereira-Netto, A.B., de Castilhos, F., 2014. Comparison of NIRS approach for prediction of internal quality traits in three fruit species. Food Chem. 143, 223-230. https://doi.org/10.1016/j.foodchem.2013.07.122.

Fan, S., Zhang, B., Li, J., Huang, W., Wang, C., 2016. Effect of spectrum measurement position variation on the robustness of NIR spectroscopy models for soluble solids content of apple. Biosystems Engineering 143, 9-19. https://doi.org/10.1016/j.biosystemseng.2015.12.012.

Giovanelli, G., Sinelli, N., Beghi, R., Guidetti, R., Casiraghi, E., 2014. NIR spectroscopy for the optimization of postharvest apple management. Postharvest Biol. Technol. 87, 13-20. https://doi.org/10.1016/j.postharvbio.2013.07.041.

Gowen, A.A., O'Donnell, C.P., Cullen, P.J., Downey, G., Frias, J.M., 2007. Hyperspectral imaging – an emerging process analytical tool for food quality and safety control. Trends in Food Science & Technology 18, 590-598. https://doi.org/10.1016/j.tifs.2007.06.001.

Guyot, S., Le Bourvellec, C., Marnet, N., Drilleau, J., 2002. Procyanidins are the most abundant polyphenols in dessert apples at maturity. LWT-Food Science and Technology 35, 289-291. https://doi.org/10.1006/fstl.2001.0843.

Huang, M., Lu, R., 2010. Apple mealiness detection using hyperspectral scattering technique. Postharvest Biol. Technol. 58, 168-175. https://doi.org/10.1016/j.postharvbio.2010.08.002.

Ignat, T., Lurie, S., Nyasordzi, J., Ostrovsky, V., Egozi, H., Hoffman, A., Friedman, H., Weksler, A., Schmilovitch, Z.e., 2014. Forecast of Apple Internal Quality Indices at Harvest and During Storage by VIS-NIR Spectroscopy. Food and Bioprocess Technology 7, 2951-2961. https://doi.org/10.1007/s11947-014-1297-7.

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

Lammertyn, J., Peirs, A., De Baerdemaeker, J., Nicolaï, B., 2000. Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment. Postharvest Biol. Technol. 18, 121-132. https://doi.org/10.1016/S0925-5214(99)00071-X.

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 Chem. 310, 125944. https://doi.org/10.1016/j.foodchem.2019.125944.
Le Bourvellec, C., Bouzerzour, K., Ginies, C., Regis, S., Plé, Y., Renard, C.M.G.C., 2011. Phenolic and polysaccharidic composition of applesauce is close to that of apple flesh. J Food Compost Anal 24, 537-547. https://doi.org/10.1016/j.jfca.2010.12.012.
Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R news 2, 18-22.
Liu, Y., Ying, Y., 2005. Use of FT-NIR spectrometry in non-invasive measurements of internal quality of 'Fuji' apples. Postharvest Biol. Technol. 37, 65-71. http://doi.org/10.1016/j.postharvbio.2005. 02.013.
Ma, T., Li, X., Inagaki, T., Yang, H., Tsuchikawa, S., 2018. Noncontact evaluation of soluble solids content in apples by near-infrared hyperspectral imaging. J. Food Eng. 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. J. Food Eng. 61, 67-81. https://doi.org/10.1016/S0260-8774(03)00188-2.
Mendoza, F., Lu, R., Ariana, D., Cen, H., Bailey, B., 2011. Integrated spectral and image analysis of hyperspectral scattering data for prediction of apple fruit firmness and soluble solids content. Postharvest Biol. Technol. 62, 149-160. https://doi.org/10.1016/j.postharvbio.2011.05.009.
Menesatti, P., Zanella, A., D'Andrea, S., Costa, C., Paglia, G., Pallottino, F., 2009. Supervised Multivariate Analysis of Hyper-spectral NIR Images to Evaluate the Starch Index of Apples. Food and Bioprocess Technology 2, 308-314. https://doi.org/10.1007/s11947-008-0120-8.
Mo, C., Kim, M.S., Kim, G., Lim, J., Delwiche, S.R., Chao, K., Lee, H., Cho, B.-K., 2017. Spatial assessment of soluble solid contents on apple slices using hyperspectral imaging. Biosystems engineering 159, 10-21. https://doi.org/10.1016/j.biosystemseng.2017.03.015.
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 Biol. Technol. 46, 99-118. https://doi.org/10.1016/j.postharvbio. 2007.06.024.
Peiris, K., Dull, G., Leffler, R., Kays, S., 1999. Spatial variability of soluble solids or dry-matter content within individual fruits, bulbs, or tubers: implications for the development and use of NIR spectrometric techniques. HortScience 34, 114-118. https://doi.org/10.21273/HORTSCI. 34.1.114.
Peirs, A., Scheerlinck, N., Nicolaï, B.M., 2003. Temperature compensation for near infrared reflectance measurement of apple fruit soluble solids contents. Postharvest Biol. Technol. 30, 233-248. https://dx.doi.org/10.1016/S0925-5214(03)00118-2.
Peng, J., Bi, J., Yi, J., Wu, X., Zhou, M., & Zhao, Y. (2019). Characteristics of cell wall pectic polysaccharides affect textural properties of instant controlled pressure drop dried carrot chips derived from different tissue zone. Food chem. 293, 358-367. https://doi.org/10.1016/ j.foodchem.2019.05.008.
Peng, Y., Lu, R., 2008. Analysis of spatially resolved hyperspectral scattering images for assessing apple fruit firmness and soluble solids content. Postharvest Biol. Technol. 48, 52-62. https://doi.org/10.1016/j.postharvbio.2007.09.019.
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://popups.uliege.be:443/1780-4507/index.php?id=8782.

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

Sanoner, P., Guyot, S., Marnet, N., Molle, D., Drilleau, J.-F., 1999. Polyphenol profiles of French cider apple varieties (Malus domestica sp.). J. Agric. Food. Chem. 47, 4847-4853. https://doi.org/10.1021/jf990563y.

Stevens, A., Ramirez-Lopez, L., 2014. An introduction to the prospectr package. R Package Vignette, Report No.: R Package Version 0.1 3.

Sun, M., Zhang, D., Liu, L., Wang, Z., 2017. How to predict the sugariness and hardness of melons: A near-infrared hyperspectral imaging method. Food Chem. 218, 413-421. https://doi.org/10.1016/j.foodchem.2016.09.023.

USDA, U., 2018. Fresh deciduous Fruit: World markets and trade (apples, grapes, & pears). https://usda.library.cornell.edu/concern/publications/1z40ks800?locale=en

Vidot, K., Devaux, M.-F., Alvarado, C., Guyot, S., Jamme, F., Gaillard, C., Siret, R., Lahaye, M., 2019. Phenolic distribution in apple epidermal and outer cortex tissue by multispectral deep-UV autofluorescence cryo-imaging. Plant Science 283, 51-59. https://doi.org/10.1016/j.plantsci.2019.02.003.

Walsh, K.B., Blasco, J., Zude-Sasse, M., Sun, X., 2020. Visible-NIR 'point' spectroscopy in postharvest fruit and vegetable assessment: The science behind three decades of commercial use. Postharvest Biol. Technol. 168, 111246. https://doi.org/10.1016/j.postharvbio.2020.111246.

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.

Zhang, B., Dai, D., Huang, J., Zhou, J., Gui, Q., Dai, F., 2018. Influence of physical and biological variability and solution methods in fruit and vegetable quality nondestructive inspection by using imaging and near-infrared spectroscopy techniques: A review. Critical Reviews in Food Science and Nutrition 58, 2099-2118. https://doi.org/10.1080/10408398.2017.1300789.

Zhang, D., Xu, Y., Huang, W., Tian, X., Xia, Y., Xu, L., Fan, S., 2019. Nondestructive measurement of soluble solids content in apple using near infrared hyperspectral imaging coupled with wavelength selection algorithm. Infrared Physics & Technology 98, 297-304. https://doi.org/10.1016/j.infrared.2019.03.026.

548 **Figure captions:**

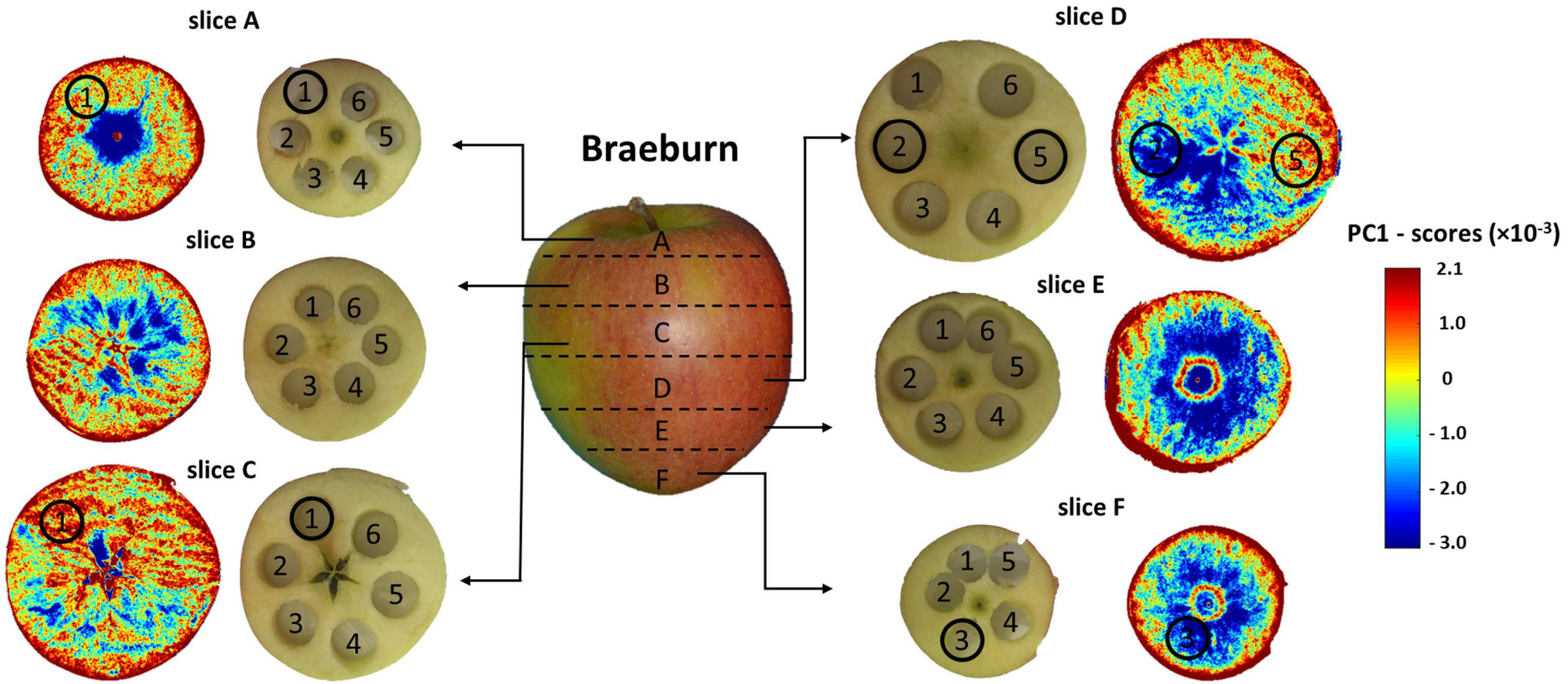
549 **Fig. 1.** The photographs of Braeburn apple slices and the first principal component
550 (PC1) score (from the PCA results on all apple groups) plot of all near-infrared
551 hyperspectral pixels (990- 2450 nm) for each slice (A, B, C, D, E, F). The selected
552 ROIs were labelled with black circles.

553 **Fig. 2.** The boxplots of: (**a**) dry matter, (**b**) total sugars, (**c**) fructose, (**d**) sucrose, (**e**)
554 glucose, (**f**) malic acid, (**g**) sum of polyphenols of 'Braeburn' (BR); 'Granny Smith'
555 (G□); 'Royal Gala' (GA); thinned 'Golden Delicious' (GD Th+) and non-thinned
556 'Golden Delicious' (GD Th-) apples.

557 **Fig. 3.** Comparison of the measured and the full-cross validated **(a)** dry matter content
558 (DMC) and **(b)** total sugars content (T□C) of the 141 ROI samples; and the most
559 contributing wavelengths for (**c**) DMC and (**d**) TSC prediction, using the
560 leave-one-out PLS regression on the ROI averaged spectra.

561 **Fig. 4.** The distribution of total sugars content (TSC) in apple slices predicted by the
562 LOO- PLS models developed based on the ROI averaged spectra.

563


slice A
slice B
slice C
slice D
slice E
slice F
Braeburn
A
B
C
D
E
F
PC1 - scores (×10-3)
2.1
1.0
0
- 1.0
- 2.0
- 3.0

564

565 **Fig. 1**

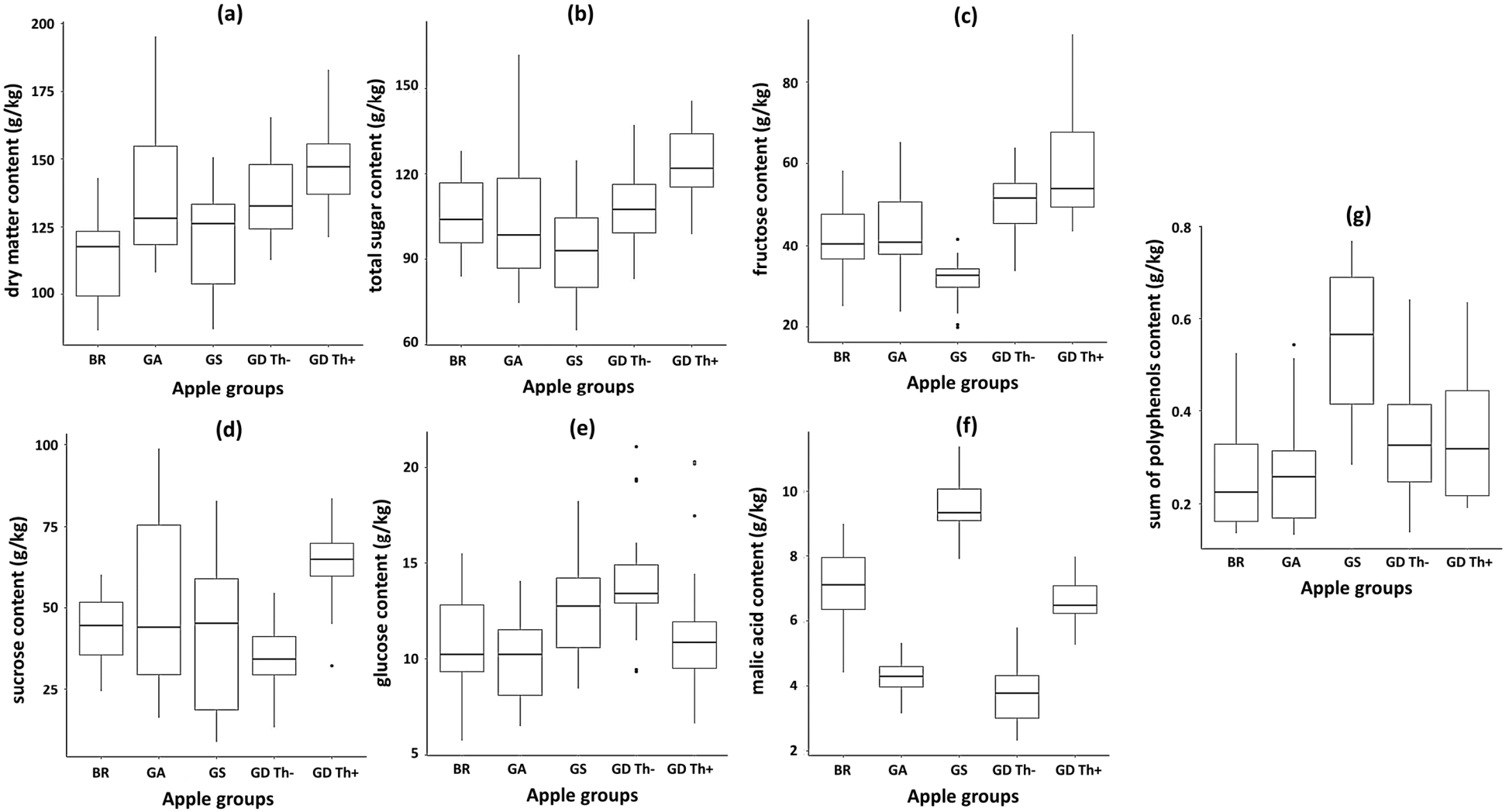
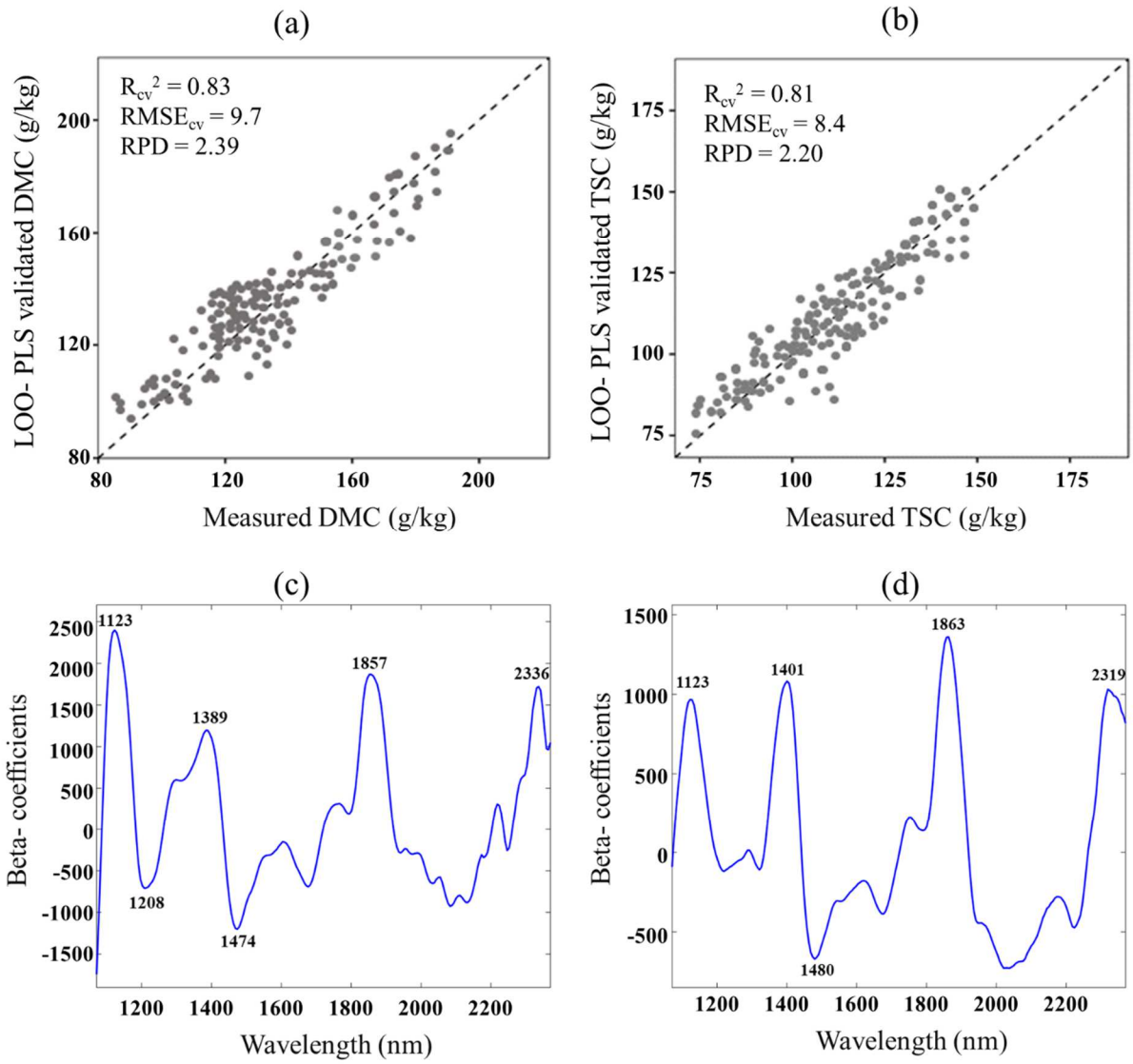
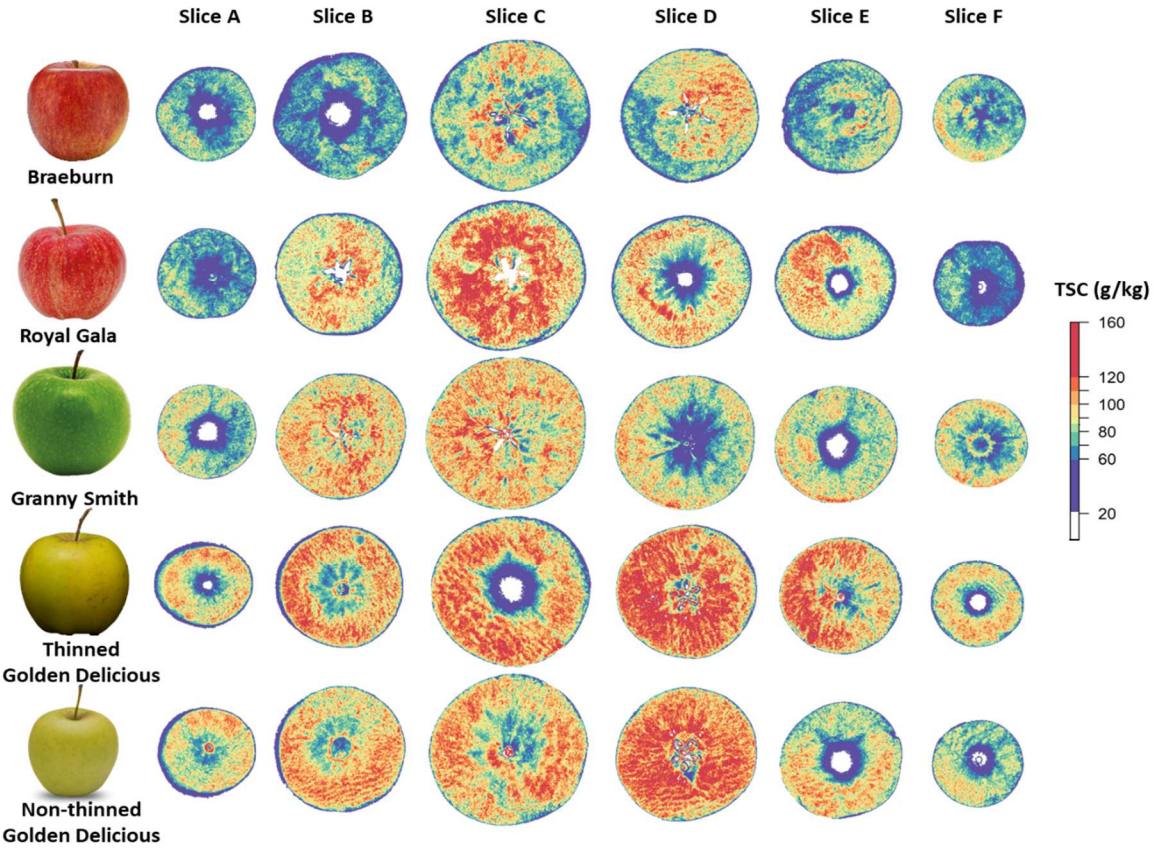

(a)
dry matter content (g/kg)
200
175
150
125
100
BR
GA
GS
GD Th-
GD Th+
Apple groups
(b)
total sugar content (g/kg)
150
120
90
60
BR
GA
GS
GD Th-
GD Th+
Apple groups
(c)
fructose content (g/kg)
80
60
40
20
BR
GA
GS
GD Th-
GD Th+
Apple groups
(g)
sum of polyphenols content (g/kg)
0.8
0.6
0.4
0.2
BR
GA
GS
GD Th-
GD Th+
Apple groups
(d)
sucrose content (g/kg)
100
75
50
25
BR
GA
GS
GD Th-
GD Th+
Apple groups
(e)
glucose content (g/kg)
20
15
10
5
BR
GA
GS
GD Th-
GD Th+
Apple groups
(f)
malic acid content (g/kg)
10
8
6
4
2
BR
GA
GS
GD Th-
GD Th+
Apple groups

Fig. 2


(a)
R_{cv}^2 = 0.83
$RMSE_{cv}$ = 9.7
RPD = 2.39
LOO- PLS validated DMC (g/kg)
Measured DMC (g/kg)
(b)
R_{cv}^2 = 0.81
$RMSE_{cv}$ = 8.4
RPD = 2.20
LOO- PLS validated TSC (g/kg)
Measured TSC (g/kg)
(c)
Beta- coefficients
1123
1208
1389
1474
1857
2336
Wavelength (nm)
(d)
Beta- coefficients
1123
1401
1480
1863
2319
Wavelength (nm)

568

569 **Fig. 3**


Slice A
Slice B
Slice C
Slice D
Slice E
Slice F
Braeburn
Royal Gala
Granny Smith
Thinned
Golden Delicious
Non-thinned
Golden Delicious
TSC (g/kg)
160
120
100
80
60
20

570

571 **Fig. 4**

572 **Table 1** Leave- one- out partial least square (LOO-PLS) and random forest (RF) results of apple internal quality traits using the averaged
573 spectra of ROIs.

Parameters	Measured range	SD	Models	Full-crossed validation (n = 141)			
				R_{cv}^2	$RMSE_{cv}$	RPD	LVs
dry matter (mg/g)	86.2- 195.3	21.9	PLS	0.83	9.7	2.39	7
			RF	0.67	14.8	1.58	7
total sugars content (g/kg)	58.8- 156.8	18.7	PLS	0.81	8.4	2.20	5
			RF	0.78	9.2	2.11	4
fructose (g/kg)	19.8- 91.6	15.4	PLS	0.38	9.0	1.35	9
			RF	0.32	10.1	1.24	8
sucrose (g/kg)	9.1- 98.7	8.4	PLS	0.67	4.9	1.73	8
			RF	0.65	5.8	1.40	6
glucose (g/kg)	5.7- 21.1	3.0	PLS	0.29	2.5	1.19	6
			RF	0.27	2.5	1.18	6
malic acid (g/kg)	2.3- 11.4	2.2	PLS	0.31	2.1	1.23	7
			RF	0.15	2.3	1.08	8
Sum of polyphenols (g/kg)	0.13- 0.77	0.16	PLS	0.14	0.17	1.01	8
			RF	0.13	0.21	0.85	9

574
575

Graphical abstract